

**A DATA ANALYTICS APPROACH TO GAS TURBINE
PROGNOSTICS AND HEALTH MANAGEMENT**

A Thesis
Presented to
The Academic Faculty

by

Ousmane Diallo

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

School of Aerospace Engineering
Georgia Institute of Technology
December 2010

Copyright © 2010 by Ousmane Diallo

**A DATA ANALYTICS APPROACH TO GAS TURBINE
PROGNOSTICS AND HEALTH MANAGEMENT**

Approved by:

Prof. Dimitri Mavris, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Prof. Joseph Saleh
School of Aerospace Engineering
Georgia Institute of Technology

Prof. Vitali Volovoi
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Sameer Vittal
Advanced Technology Operations
General Electric Company

Dr. Xiaomo Jiang
Condition Based Maintenance
General Electric Energy

Dr. Virendra Kumar
Principal Engineer
General Electric Energy

Date Approved: October 8, 2010

To my parents

Amara Diallo

And

Dyoubu Diallo

ACKNOWLEDGEMENTS

I am forever indebted to many people whose help has made this endeavor possible. First and foremost I extend my sincere gratitude to my advisor, Prof. Dimitri Mavris, for giving me the opportunity to join the ASDL and guiding me through his inspiring philosophy of scientific research, with a fatherly patience. I would like to thank my committee members: Prof. Joseph Saleh for his availability, friendship and his willingness to always listen and guide me in the right direction; Prof. Vitali Volovoi for his valuable feedback and for believing in me; Dr. Sameer Vittal for acting as a de-facto second advisor by supporting and guiding me; Dr. Xiaomo Jiang for his help and recommendations with appropriate techniques; and Dr Virendra Kumar for his words of wisdom and encouragements.

Also, I would like to thank many ASDL members who supported and helped during my tenure making this experience somewhat bearable. First, I want to thank Mark Waters who recruited me. I am grateful to Brian Kestner, Scott Duncan and Young Ki Lee for their many hours of time and for challenging me on making the thesis better. My thanks go to Reza Rezvani, Stéphane Dufresne, Irian Ordaz, Léon Phan for their friendship and encouragement. I want to thank my colleagues in room 310 with: Keith Becker, Joseph Iacobucci, and WoongJe Sung for their support. I was privilege to meet friends at Georgia Tech who had a deep impact on my life: Sang Il (Ryan) Lee, Christianna Taylor, Yoko Watanabe, Ogundiran Sormani, and Olivier Deigni, to name a few. My sincere appreciation goes to Prof. Stephen Ruffin and Wanda Pierson, at the Georgia Space Grant Consortium, for their continuous support and encouragement.

There are many people outside of Tech GA who provided inspiration and encouragement; at GE Energy, I would to thank Samar Soliman, Mike Sullivan, Doug Warwick, Erwing

Calleros, and a special thanks to Ted Fisher for indirectly leading me to the problem that became the focus of my research.

Last but not least, I owe a debt gratitude to my siblings: Mariama, Mahamadi, Kalil, Abou, Assiata, Souleymane, Fousseny, Koumba (Matou) and Salimata; my mother and father and to my dear friends Aminata Touré, Jaketa and Langston for their unconditional love, sacrifices, and belief in me, even when I doubted myself.

To all those cited by name here and the many more I did not mention by name, I will forever owe you this achievement. Thank you all very much!

Ousmane, October 2010

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	IV
LIST OF TABLES	X
LIST OF FIGURES	XIV
LIST OF SYMBOLS AND ABBREVIATIONS	XVIII
SUMMARY	XX
1 INTRODUCTION.....	1
1.1 Power plant operation and maintenance trend	3
1.2 Motivation.....	6
1.2.1 The reasons for OEMs to enter in LTSA.....	8
1.2.2 OEMs develop strategy to make LTSA profitable	9
1.2.3 Relevance of Potential impact of research	11
1.3 Research Organization	12
2 BACKGROUND AND RESEARCH OBJECTIVES.....	14
2.1 Context	14
2.2 Background	14
2.2.1 Traditional maintenance policies.....	15
2.2.2 A OEM customized maintenance	19
2.2.3 Remote Monitoring center.....	20
2.3 Observations.....	21
2.3.1 Observation 1	21
2.3.2 Observation 2	21
2.3.3 Observation 3	21
2.4 Research Objectives and questions.....	22
2.4.1 Objective 1:	22
2.4.2 Objective 2:	22
2.4.3 Research question 1:.....	23
2.4.4 Research question 1a:.....	23
2.4.5 Research question 2:.....	23
2.4.6 Research Question 2a:.....	23
2.4.7 Research question 3:.....	24
2.4.8 Research question 4:.....	24

2.5	Previous work on fault diagnostics of time-frequency processing techniques	25
2.6	Literature review on residual life estimation	28
2.6.1	Data-driven:	28
2.6.2	Model-driven:	29
2.6.3	Hybrid of data-driven with Model-driven:	29
2.7	Assumptions.....	29
2.8	Hypothesis.....	30
2.8.1	Research Hypothesis 1:	30
2.8.2	Hypothesis 1a:	31
2.8.3	Research hypothesis 2:	31
2.8.4	Research Hypothesis 2a:	31
2.8.5	Research hypothesis 3:	31
2.8.6	Research hypothesis 4:	31
3	MATHEMATICAL BACKGROUND OF WAVELETS.....	32
3.1	Analysis of Time-frequency of high frequency signals	32
3.1.1	Fourier transforms overview	33
3.1.2	Mathematical overview of wavelet transforms	34
4	PROPOSED APPROACH FOR FAILURE PRECURSOR DETECTION.....	41
4.1	Detection of failure precursors	42
4.2	Steps for failure precursor detection.....	43
4.2.1	Raw sensors data collection	44
4.2.2	Data Pre-Processing	46
4.2.3	De-noising.....	52
4.2.4	Multi-resolution analysis using discrete Wavelet packet decomposition	65
4.2.5	Data Fusion using Probabilistic Principal Component Analysis.....	73
4.2.6	Anomaly detection decision	92
4.2.7	Combining both types of information	107
4.2.8	Effect of sampling interval on the proposed anomaly detection methodology	113
5	PROGNOSTICS AND HEALTH MANAGEMENT APPROACH TO RESIDUAL TIME TO FAILURE MODELING	119
5.1	Importance of Prognostics and Health Management.....	119
5.1.1	Data generation process for PHM analysis	124
5.2	Deterministic approach to modeling and simulation for estimation of residual life	126
5.2.1	Parametric Regression.....	127
5.2.2	Non-parametric Regression.....	129
5.3	Implementation of deterministic modeling for estimation of residual life	137
5.3.1	Residual life modeling using RSE.....	137
5.3.2	Non-parametric RUL estimation	138
5.3.3	WAVELET Neural Network (WNN).....	141
5.4	Empirical or non-parametric approach to residual life estimation: Kaplan-Meier.....	142

5.4.1	Censored data.....	142
5.4.2	Kaplan Meier overview.....	144
5.5	Parametric analysis for assumed distributions.....	149
5.5.1	Exponential Distribution.....	150
5.5.2	Weibull Distribution.....	153
 6 IMPLEMENTATION OF THE PROPOSED METHOD FOR GAS TURBINE COMPRESSOR CATASTROPHIC FAILURE		157
6.1	Implementation of failure precursor detection.....	157
6.1.1	Assumptions made for proposed approach implementation.....	158
6.1.2	Inspection background of General Electric gas turbine.....	158
6.1.3	Failure precursor detection process applied to a gas turbine compressor for data collection...	160
6.1.4	Data collection procedure for RTTF modeling.....	175
6.2	Modeling of Residual Time to Failure.....	179
6.2.1	Implementation of the deterministic approach to residual lifetime modeling.....	180
6.2.2	Implementation of non-parametric data analysis using the Kaplan-Meier estimator.....	191
6.2.3	Parametric Distribution.....	198
6.3	Results Discussion	207
 7 CONCLUDING REMARKS		211
7.1	Revisiting the research objectives and questions.....	211
7.1.1	Revisiting the research objectives.....	211
7.1.2	Revisiting research question 1, 1a and hypothesis 1, 1a.....	212
7.1.3	Revisiting research question 2 and hypothesis 2.....	213
7.1.4	Revisiting research question 3 and hypothesis 3.....	214
7.1.5	Revisiting research question 4 and hypothesis 4.....	215
7.2	Summary of Findings and Research Contributions.....	216
7.3	Conclusion	217
7.4	Recommendations for Further research	219
 APPENDIX A- EXPERIMENTS		221
A.1	Test Unit 1.....	221
A.2	Test Unit 2.....	221
A.3	Test Unit 3.....	227
A.4	Test Unit 4.....	231
A.5	Test Unit 5.....	235
A.6	Test Unit 6.....	239
A.7	Test Unit 7.....	242

A.8	Test Unit 8.....	246
A.9	Test Unit 9.....	250
A.10	Test Unit 10.....	253
A.11	Test Unit 11.....	257
APPENDIX B- MRA LEVEL DECISION.....		261
B.1	Level-2 multi-resolution analysis.....	261
B.2	Level-3 multi-resolution analysis.....	263
B.3	Level-4 multi-resolution analysis.....	265
B.4	Level of decomposition decision for multi-resolution analysis.....	266
APPENDIX C- KAPLAN-MEIER TABLES.....		268
C.1	First Anomaly tabular data with its corresponding confidence interval.....	268
C.2	Second Anomaly tabular data with its corresponding confidence interval.....	268
C.3	Extended tabular data.....	269
C.4	Tabular data for the Kaplan-Meier plot of residual time for extended set.....	271
C.5	Extended tabular data with for confidence interval.....	272
APPENDIX D- RESIDUAL PLOTS FOR PARAMETRIC ANALYSIS.....		273
D.1	Residual plots for limited data set: first detected anomaly.....	273
D.2	Residual plots for extended data set.....	274
REFERENCES.....		276

LIST OF TABLES

	Page
Table 1: Comparison of the different time-frequency method adapted from [24].....	27
Table 2: Summary of Normalization Techniques [88].....	48
Table 3: Categories of Signal Denoising Techniques (adapted from [96]).....	54
Table 4: Level-2 decomposition Node energy content for vibration sensor X1.....	70
Table 5: Level 3 decomposition Node energy content of vibration sensor X1.....	72
Table 6: Fusion techniques for associating passively acquired data to locate and track targets [124].....	80
Table 7: Feature Extraction and Projection Methods [133].....	83
Table 8: Covariance Matrix for eigenvalues calculation.....	89
Table 9: Pareto chart of eigenvavlues contributions.....	90
Table 10: Original twelve eigenvectors.....	90
Table 11: Maximum likelihood weight for principal components.....	91
Table 12: Summary of type I and type II errors in decision making (adapted from [124]).....	103
Table 13: Pareto plot of Eigenvalues contribution for vibration sensors only.....	107
Table 14: two retained principal components from the PPCA step for vibration sensors only.....	108
Table 15: Pareto plot of Eigenvalues contribution for performance sensors only.....	109
Table 16: Two retained principal components from the PPCA step for performance sensors only.....	109
Table 17: Pareto plot of Eigenvalues contribution for all sensors.....	111
Table 18: Three retained PC from the PPCA step for ALL sensors combined.....	111
Table 19: Pareto plot for five minute sampling interval.....	114
Table 20: Pareto plot for one minute sampling interval.....	115
Table 21: Pareto plot for five second sampling interval.....	117
Table 22: Factors Subdivided on the Basis of Level of Evidence Obtained by a Correlation With Actual Survival [153].....	121
Table 23: GE technology specific baseline recommended inspection intervals [9].....	159
Table 24: Monitored Gas turbine compressor sensors.....	161
Table 25: Node energy content of each sensor at level-3 DWPT decomposition.....	166
Table 26: Covariance Matrix for PCA.....	167
Table 27: Pareto chart of eigenvavlues contributions.....	168
Table 28: Eigenvectors or weight matrix for three principal components.....	168
Table 29: Maximum likelihood weight for PPCA principal components.....	169
Table 30: Summary of first detected anomaly.....	177
Table 31: Summary of all detected anomalies.....	178
Table 32: Actual first anomaly to be used for PHM model.....	178

Table 33: Actual extended data set representing all anomalies for PHM model.....	179
Table 34: First anomaly.....	179
Table 35: RSE fit coefficients	180
Table 36: Contribution of each term and cross terms to the RTTF prediction	180
Table 37: Extended data set of ALL the anomalies.....	185
Table 38: Extended data set with 3 inputs.....	188
Table 39: Tabular data for the Kaplan-Meier plot of residual time after the first detected anomaly.....	192
Table 40: Tabular data for the Kaplan-Meier plot of survival time after the second detected anomaly.....	194
Table 41: Summary of parametric survival data analysis.....	210
Table 42: Monitored Gas turbine compressor sensors	221
Table 43: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 2	223
Table 44: Pareto chart of eigenvlues contributions of UNIT2.....	224
Table 45: PCA weight matrix for UNIT2.....	225
Table 46: PPCA Maximum likelihood weight matrix for UNIT2.....	226
Table 47: Actual first anomaly to be used for PHM model for unit 2	227
Table 48: Actual extended data set representing all anomalies for PHM model for unit 2	227
Table 49: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 3	228
Table 50: Pareto chart of eigenvlues contributions of UNIT3.....	228
Table 51: PCA weight matrix for UNIT3.....	229
Table 52: PPCA Maximum likelihood weight matrix for UNIT3.....	229
Table 53: Actual first anomaly to be used for PHM model for unit 3	231
Table 54: Actual extended data set representing all anomalies for PHM model for unit 3	231
Table 55: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 4	231
Table 56: Pareto chart of eigenvlues contributions of UNIT4.....	232
Table 57: PCA weight matrix for UNIT4.....	232
Table 58: PPCA Maximum likelihood weight matrix for UNIT4.....	232
Table 59: Actual first anomaly to be used for PHM model for unit 4	234
Table 60: Actual extended data set representing all anomalies for PHM model for unit 4	234
Table 61: Monitored sensors for unit 5	235
Table 62: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 5	235
Table 63: Pareto chart of eigenvlues contributions of UNIT 5.....	236
Table 64: PCA weight matrix for UNIT 5.....	236
Table 65: PPCA Maximum likelihood weight matrix for UNIT 5.....	237
Table 66: Actual first anomaly to be used for PHM model for unit 5	238
Table 67: Actual extended data set representing all anomalies for PHM model for unit 5	238
Table 68: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 6	239
Table 69: Pareto chart of eigenvlues contributions of UNIT 6.....	239

Table 70: PCA weight matrix for UNIT 6.....	240
Table 71: PPCA Maximum likelihood weight matrix for UNIT 6.....	240
Table 72: Actual first anomaly to be used for PHM model for unit 6.....	242
Table 73: Actual extended data set representing all anomalies for PHM model for unit 6.....	242
Table 74: Monitored sensors for unit 7.....	242
Table 75: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit.....	243
Table 76: Pareto chart of eigenvavlues contributions of UNIT 7.....	243
Table 77: PCA weight matrix for UNIT 7.....	244
Table 78: PPCA Maximum likelihood weight matrix for UNIT 7.....	244
Table 79: Actual first anomaly to be used for PHM model for unit 7.....	246
Table 80: Actual extended data set representing all anomalies for PHM model for unit 7.....	246
Table 81: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 8.....	246
Table 82: Pareto chart of eigenvavlues contributions of UNIT 8.....	247
Table 83: PCA weight matrix for UNIT 8.....	247
Table 84: PPCA Maximum likelihood weight matrix for UNIT 8.....	248
Table 85: Actual first anomaly to be used for PHM model for unit 8.....	249
Table 86: Actual extended data set representing all anomalies for PHM model for unit 8.....	249
Table 87: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 9.....	250
Table 88: Pareto chart of eigenvavlues contributions of UNIT 9.....	250
Table 89: PCA weight matrix for UNIT 9.....	251
Table 90: PPCA Maximum likelihood weight matrix for UNIT 9.....	251
Table 91: Actual first anomaly to be used for PHM model for unit 9.....	253
Table 92: Actual extended data set representing all anomalies for PHM model for unit 9.....	253
Table 93: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 10.....	253
Table 94: Pareto chart of eigenvavlues contributions of UNIT 10.....	254
Table 95: PCA weight matrix for UNIT 10.....	254
Table 96: PPCA Maximum likelihood weight matrix for UNIT 10.....	255
Table 97: Actual first anomaly to be used for PHM model for unit 10.....	257
Table 98: Actual extended data set representing all anomalies for PHM model for unit 10.....	257
Table 99: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 11.....	257
Table 100: Pareto chart of eigenvavlues contributions of UNIT 11.....	258
Table 101: PCA weight matrix for UNIT 11.....	258
Table 102: PPCA Maximum likelihood weight matrix for UNIT 11.....	259
Table 103: Actual first anomaly to be used for PHM model for unit 11.....	260
Table 104: Actual extended data set representing all anomalies for PHM model for unit 11.....	260
Table 105: Node energy content of each sensor at level-2 DWPT decomposition.....	262
Table 106: Node energy content of each sensor at level-3 DWPT decomposition.....	264

Table 107: Node energy content of each sensor at level-4 DWPT decomposition266

LIST OF FIGURES

	Page
Figure 1: Spectrum of OEM risk mitigation adapted from [2]	3
Figure 2: Cost of electricity trend with the O&M expenses [2]	7
Figure 3: GE Monitoring & Diagnostics concept adapted from [2]	9
Figure 4: Paradigm shift in industrial maintenance [16]	18
Figure 5: Illustration of Hard and Soft failures [13]	18
Figure 6: GE gas turbine maintenance need based on counts of starts and hours adapted from [9]	20
Figure 7: Notion of filtering for wavelet decomposition	35
Figure 8: Illustration of the multiple-level decomposition	36
Figure 9: Illustration of Time-Frequency domain of Wavelet Transform	37
Figure 10: Illustration of the difference between Fourier transforms and Wavelet transform	38
Figure 11: Wavelet Packet decomposition [77]	39
Figure 12: Illustration of the proposed approach organization	41
Figure 13: Precursor failure detection steps	44
Figure 14: Scatterplot Matrix of raw data	46
Figure 15: Actual and Normalized Health sensor distributions	51
Figure 16: Actual and Normalized Vibration sensor distributions	52
Figure 17: Bayesian DWPT de-noising procedure for times series [109]	60
Figure 18: Original signal and Noise added signal	64
Figure 19: Denoised signal using the Bayesian DWPT denoising procedure	64
Figure 20: Level-2 Tree DWPT decomposition of vibration sensor X1	70
Figure 21: MRA level-2 DWPT decomposition of vibration sensor X1 signal	71
Figure 22: Level 3 Tree DWPT decomposition of Vibration sensor X1	71
Figure 23: MRA of level-3 DWPT decomposition of Vibration sensor X1 signal	72
Figure 24: Data fusion for detection, classification, and identification algorithms [124]	75
Figure 25: PPCA result principal components	92
Figure 26: Anomaly Detection Decision Flowchart	100
Figure 27: Graphical representation of type II Error [152]	102
Figure 28: Reconstructed 1-dimensional signal from the 3 PPCA	104
Figure 29: Converged statistical parameters	105
Figure 30: Monitoring of the damage indicator SAD	105
Figure 31: Monitoring of the damage indicator SSD	106
Figure 32: H-Function of the hypothesis testing result	106
Figure 33: Reconstructed signal for vibration sensors only	108

Figure 34: H-function for vibration sensors only	108
Figure 35: Reconstructed signal for performance sensors only.....	110
Figure 36: H-Function for vibration sensors only	110
Figure 37: Reconstructed signal for ALL sensors combined	112
Figure 38: H-Function for ALL sensors combined	112
Figure 39: Monitoring of the damage indicator SAD for five minute sampling intervals.....	114
Figure 40: H-Function for five minute sampling interval.....	115
Figure 41: Monitoring of the damage indicator SAD for one minute sampling interval.....	116
Figure 42: H-Function for one minute sampling interval	116
Figure 43: Monitoring of the damage indicator SAD for five second sampling interval	117
Figure 44: Schematic of the RUL architecture	122
Figure 45: Example of PHM System Architecture for a fighter aircraft [165].....	123
Figure 46: PHM operations in military application [165]	124
Figure 47: Illustration of the prognostics parameters	126
Figure 48: Illustration of an artificial neural from [171]	130
Figure 49: Neural Network Conceptual Diagram [170].....	132
Figure 50: Feed-Forward Neural Network	132
Figure 51: Depiction of a dynamic feed-forward network [174].....	133
Figure 52: Recurrent Neural Network [176]	134
Figure 53: General structure of WNN [181]	136
Figure 54: Schematic of conceptual M&S for residual life estimation using RSE.....	137
Figure 55: Schematic of conceptual M&S for residual life estimation using ANN	139
Figure 56: Conceptual representation of Dynamic Wavelet Neural Network [46]	141
Figure 57: Censored data with different beginning of experiment time [186]	144
Figure 58: Illustration of empirical survivor function with no censored observations [186]	147
Figure 59: Illustrative example of Kaplan-Meier plot for censored data.....	148
Figure 60: GE gas turbine inspection scope [9]	158
Figure 61: Layout of monitored sensors for gas turbine compressor anomaly detection	160
Figure 62: Distribution of the 12 sensors pre-processed	163
Figure 63: Noisy and de-noised sensor X3.....	164
Figure 64: Tree decomposition of sensor X3 signal with the wavelet component at node 7	165
Figure 65: DWPT at 3-level MRA of sensor X3	166
Figure 66: Signal obtained from PPCA process.....	170
Figure 67: Reconstructed 1-dimensional Signal.....	171
Figure 68: Statistical parameters for SAD and SSD.....	172
Figure 69: Monitoring of damage indicator SAD.....	173
Figure 70: Monitoring of damage indicator SSD	174

Figure 71: Result of the Bayesian hypothesis testing	175
Figure 72: Illustration of the time to failure parameter	177
Figure 73: Prediction profiler of RTTF (ASI,ADI) using RSE	181
Figure 74: RSE fit Actual vs. Predicted	181
Figure 75: RSE fit Residual vs. Predicted RTTF	182
Figure 76: Neural Network for modeling RTTF	183
Figure 77: Prediction profiler of RTTF (ASI,ADI) using ANN.....	183
Figure 78: Actual vs. Predicted plot of NN regression.....	184
Figure 79: Residual vs. Predicted plot of NN regression	184
Figure 80: Actual vs. Predicted regression of extended data set using RSE	186
Figure 81: Residual vs. Predicted regression of extended data set using RSE	186
Figure 82: Actual vs. Predicted regression of extended data set using ANN	187
Figure 83: Actual vs. Predicted regression of extended data set using ANN.....	187
Figure 84: Pareto plot of the terms contribution for 3 inputs	189
Figure 85: Actual vs. Predicted regression of extended data set with 3 Inputs using RSE	189
Figure 86: Residual vs. Predicted regression of extended data set with 3 Inputs using RSE	190
Figure 87: Actual vs. Predicted regression of extended data set with 3 Inputs using ANN	190
Figure 88: Actual vs. Predicted regression of extended data set with 3 Inputs using ANN	191
Figure 89: Kaplan-Meier plot of the residual time to failure for first detected anomaly.....	193
Figure 90: Kaplan-Meier plot of the RTTF from first detected anomaly with 95% confidence band.....	194
Figure 91: Kaplan-Meier plot of RTTF for second detected anomaly	195
Figure 92: Kaplan-Meier plot of the RTTF from SECOND detected anomaly with 95% confidence.....	195
Figure 93: Comparison of the first and second anomalies.....	196
Figure 94: Kaplan-Meier plot for RTTF with extended data set.....	197
Figure 95: Kaplan-Meier plot of the RTTF for extended data set with 95% confidence	198
Figure 96: pdf of the first Anomaly using Exponential distribution.....	199
Figure 97: Exponential plot of the Kaplan-Meier estimation of the RTTF	200
Figure 98: Kaplan-Meier estimated RTTF of first anomaly and exponential fit.....	201
Figure 99: pdf of extended data set using the exponential distribution	201
Figure 100: Exponential plot of the Kaplan-Meier estimated of the RTTF using the extended data set.....	202
Figure 101: Kaplan-Meier estimated RTTF for extended data set and exponential fit	203
Figure 102: pdf of the first anomaly using the Weibull distribution	203
Figure 103: Weibull plot of the Kaplan-Meier estimated RTTF	204
Figure 104: Kaplan-Meier estimated RTTF of first anomaly and Weibull fit.....	205
Figure 105: pdf of extended data set using the Weibull distribution.....	205
Figure 106: Weibull plot of the Kaplan-Meier estimation of the RTTF using extended data set.....	206
Figure 107: Kaplan-Meier estimated RTTF for extended data set and Weibull fit	207

Figure 108: Comparison of exponential and Weibull distribution fits with the Kaplan-Meier	208
Figure 109: Exponential and Weibull distribution fits with the Kaplan-Meier for extended data set	209
Figure 110: Unit 2-Tree decomposition of sensor X3 signal with the wavelet component at node 7	222
Figure 111: Unit2- DWPT 3-level MRA of sensor x3	223
Figure 112: Reconstructed 1-dimensional Signal for Unit 2	226
Figure 113: Monitoring of damage indicator SAD for unit 2	227
Figure 114: Reconstructed 1-dimensional Signal for Unit 3	230
Figure 115: Monitoring of damage indicator SAD for unit 3	230
Figure 116: Reconstructed 1-dimensional Signal for Unit 4	233
Figure 117: Monitoring of damage indicator SAD for unit 4	234
Figure 118: Reconstructed 1-dimensional Signal for Unit 5	237
Figure 119: Monitoring of damage indicator SAD for unit 5	238
Figure 120: Reconstructed 1-dimensional Signal for Unit 6	241
Figure 121: Monitoring of damage indicator SAD for unit 6	241
Figure 122: Reconstructed 1-dimensional Signal for Unit 7	245
Figure 123: Monitoring of damage indicator SAD for unit 7	245
Figure 124: Reconstructed 1-dimensional Signal for Unit 8	248
Figure 125: Monitoring of damage indicator SAD for unit 8	249
Figure 126: Reconstructed 1-dimensional Signal for Unit 9	252
Figure 127: Monitoring of damage indicator SAD for unit 9	252
Figure 128: Reconstructed 1-dimensional Signal for Unit 10	256
Figure 129: Monitoring of damage indicator SAD for unit 10	256
Figure 130: Reconstructed 1-dimensional Signal for Unit 11	259
Figure 131: Monitoring of damage indicator SAD for unit 11	260
Figure 132: Level-2 tree decomposition of sensor X3 signal with the wavelet component at node 3	261
Figure 133: DWPT at level-2 MRA of sensor X3	262
Figure 134: Level-3 tree decomposition of sensor X3 signal with the wavelet component at node 7	263
Figure 135: DWPT at level-3 MRA of sensor X3	264
Figure 136: Level-4 tree decomposition of sensor X3 signal with the wavelet component at node 15	265
Figure 137: DWPT at level-4 MRA of sensor X3	266
Figure 138: Residual by Predicted for Exponential distribution assumption for limited data set	273
Figure 139: Residual by Predicted for Weibull distribution assumption for limited data set	273
Figure 140: Residual by Predicted for Exponential distribution assumption for extended data set	274
Figure 141: Residual by Predicted for Weibull distribution assumption for extended data set	275

LIST OF SYMBOLS AND ABBREVIATIONS

ADI	Anomaly Duration Index
ANN	artificial neural networks
ASI	Anomaly Severity Index
BOP	balance of plant
CWD	Choi-Williams distribution
CSD	Cone-Shape distribution
CWT	Continuous Wavelet Transform
DWNN	dynamic wavelet neural network
DWPT	discrete wavelet packet transforms
EPRI	electric power research institute
FFT	fast Fourier transforms
FT	Fourier transforms
LTSA	Long term service agreement
M&S	Modeling and Simulation
MRA	Multi-resolution Analysis
NPA	Number of Previous Anomaly
OEM	Original equipment manufacturer
O&M	Operation and maintenance
O&MA	Operation and Maintenance agreement
PCA	Principal Component Analysis
PPCA	Probabilistic Principal Component Analysis
RSE	Response surface equation
RTTF	Residual time to failure
RUL	Remaining Useful Life
STFT	Short Time Fourier Transform
WFT	windowed Fourier transforms
WNN	Wavelet Neural network

WVD

Wigner-Ville distribution

SUMMARY

As a consequence of the recent deregulation in the electrical power production industry, there has been a shift in the traditional ownership of power plants and the way they are operated. Many new private entrepreneurs with no prior experience in power plants operation have entered into the power plant business. Thus, to hedge their business risks, they enter into long-term service agreement (LTSA), which can last as long as twenty years, with third parties for their operation and maintenance (O&M) activities. They share not only the rewards of plant performance surplus, but also the risks of any performance shortfall. As the major LTSA providers, original equipment manufacturers (OEM) have invested huge amounts of money to develop preventive maintenance strategies to minimize the occurrence of costly unplanned outages resulting from failures of the equipments covered under LTSA contracts. As a matter of fact, a recent study by the Electric Power Research Institute (EPRI) estimates the cost benefit of preventing a failure of a General Electric 7FA or 9FA technology compressor at \$10 to \$20 million.

Therefore, in this dissertation, a two phase data analytics approach is proposed that first uses the existing monitoring gas path and vibration sensors data to develop a proactive strategy that systematically detects and validates catastrophic failure precursors so as to avoid the failure; then secondly estimate the residual time to failure of the unhealthy items. To achieve the goal of this research, a step by step methodology is developed for each of the two phases. For the first part of this work, the time-frequency technique of the wavelet packet transforms is used to de-noise the noisy sensor data. Next, the time-series signal of each sensor is decomposed to perform a multi-resolution analysis to extract its features. After that, the probabilistic principal component analysis is applied as a data fusion technique to reduce the number of the potentially correlated multi-sensors measurement into a few uncorrelated principal components. The last step of the failure precursor detection methodology, the anomaly detection decision, is in itself a multi-stage process. The obtained principal components from the data fusion step are first

combined into a one-dimensional reconstructed signal representing the overall health assessment of the monitored systems. Then, two damage indicators of the reconstructed signal are defined and monitored for defect using a statistical process control approach. Finally, the Bayesian evaluation method for hypothesis testing is applied to a computed threshold to test for deviations from the healthy band.

To model the residual time to failure, the anomaly severity index and the anomaly duration index are defined as the defects characteristics. Two modeling techniques are investigated for the prognostication of the survival time after an anomaly is detected: the deterministic regression approach, and parametric approximation of the non-parametric Kaplan-Meier plot estimator. It is established that the deterministic regression provides poor prediction estimation. The non parametric survival data analysis technique of the Kaplan-Meier estimator provides the empirical survivor function of the data set comprised of both non-censored and right censored data. Though powerful because no a-priori predefined lifetime distribution is made, the Kaplan-Meier result lacks the flexibility to be transplanted to other units of a given fleet that were not included in the failure data. The parametric analysis of survival data is performed with two popular failure analysis distributions: the exponential distribution and the Weibull distribution. The conclusion from the parametric analysis of the Kaplan-Meier plot is that the larger the data set, the more accurate is the prognostication ability of the residual time to failure model.

CHAPTER 1

INTRODUCTION

In recent years, a paradigm shift has been occurring in the role played by complex system designers and manufacturers in many different fields. For the most part, major gas turbine designers and manufacturers do not just produce and sell new gas turbines to customers, they are becoming involved in the total life cycle of the products from designing and manufacturing to providing the after sale operation and maintenance of the systems for several years. The trend is valid for civil aviation applications as well as for electrical power production applications. Particularly in the heavy-duty gas turbine for electricity production, the need to outsource plants operation and maintenance (O&M) arose as plant owners sought to hedge some of the normally high financial risk of power plant operation by paying for some type of insurance premium in return to have insurance providers to share the potential risks.

In general, the electricity production industry is a much-deregulated one in many ways with an extensive competition between producers. The reason for that is local and federal governments are very much involved at different stages from the production of electricity with its efficiency requirements and environment restrictions to the reliability and availability of the distribution. The rationale behind the government's involvement is understandable because our modern society depends heavily on reliable and continuous power guarantee for businesses, hospitals, airports, school, public transportation systems, etc. Therefore, municipalities and electrical power regulatory authorities put strict requirements on the minimum acceptable efficiency, reliability and availability, making them ones the highest requirement metrics for power plant ability to be dispatched.

To diverge some of the continuous business risk, plant owners enter into complex and specific long-term contractual agreements that can last for several years with third

partied O&M providers. The contractual agreements work like insurance policies where the manufacturers guarantee a given level of power output and/or efficiency over several years. They also provide repair, replacement, and upgrade parts to the degrading power plant. Overall, it is supposed to be a win-win partnership for both parties by sharing the operation risks and the rewards. On one hand, plant owners don't have to carry all alone the burden of potential investment failures, and also the long-term contractual agreement raises the re-sale value of the plant.

On the other hand, gas turbine manufacturers became the natural fits for new and lucrative insurance business ventures because they are the ones who know and understand their designed products and will be willing to guarantee its operations. In addition to that, by entering into the operational partnership with the plant owners, the manufacturers have unprecedented access to “a live laboratory” that should allow them to learn from eventual design shortcomings made in previous gas turbine designs in order to improve upon future ones, ultimately giving them a competitive advantage. Also, it is important to note that being the plants owners' partners provide the gas turbine manufacturers with a reliable customer base when new parts are needed for upgrade over the life of the contract.

Although there have always been third party players in the gas turbine for electricity power plants long-term contractual agreements, the most important players are the major gas turbine original equipment manufacturers (OEM) such as General Electric Energy (GE Energy), Siemens Power Generation, Alstom Power, Mitsubishi Heavy Industries, and Ansaldo Energia. Together, these major OEMs represented about 94% of the global market for the period of 2000 through 2004 [1]. Each of these main gas turbine original equipment manufacturers has its own definition and the foreseeable benefits to the plant owners of their long-term service agreement (LTSA).

1.1 Power plant operation and maintenance trend

With the deregulation of the electrical power production market, private ownership of power plants occurred, that introduced new attitudes towards the operation and maintenance of those plants, mainly outsourcing. Outsourcing the O&M came as a way to face the unprecedented level of competition in the power plant business. Because third party O&M deem lucrative for the service providers, a consequence was a generation of an environment of competition among independent third party O&M provider and OEM, with the competitive hedge to OEM which can assume bigger risk. Then the role played by OEMs moved from transactional risk management mindset of building and selling gas turbines to end owners, to a long-term contractual risk management or partnership with the equipments buyers, see illustrative figure 1 below.

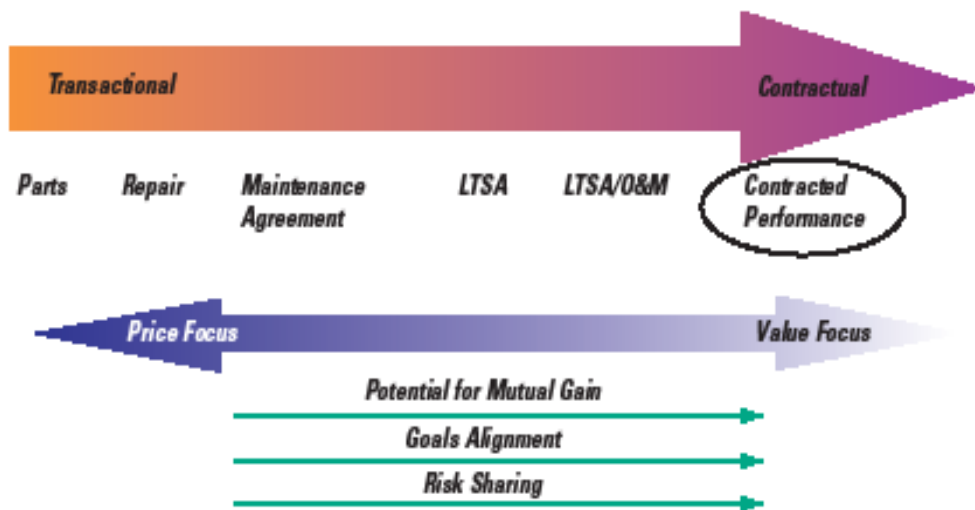


Figure 1: Spectrum of OEM risk mitigation adapted from [2]

In the past, the transactional approach to risk management was the way power plant operation business was done, which is fundamentally different from the now common contractual service. In the context of gas turbine usage in power plants, the transactional approach can be defined as the plant owner with an outage who can shop around and buy the parts, repairs and services as the plant requires. Therefore, the main goal of the plant

owner is to find the lowest price bids for services that makes the transactional approach a very price focus strategy with little to no ownership interest to the vendors [2]. In the contractual approach by contrast, the vendor is a partner who has the same end goal for the plant operation and maintenance as the plant owners do. Though the basics of the long-term service agreement (LTSA) [3] is similar in nature for all majors OEMs, each of them has its own tailored version of definition based on its offerings and the level of commitment it is willing to assume as the provider of operation and maintenance. Thus, the contractual agreement definition for each of the major gas turbine OEM for power plant applications is different and is provided below.

General Electric Energy promotes its Contractual Service Agreement (CSA) as “changing the focus from minimum price to maximum value by aligning the risk/reward goals.”[2] GE offers customized CSA in the forms of LTSA and operation and maintenance agreement (O&MA) to plant owners. GE defines LTSA as fixed-price maintenance contract, based on performance recovery at the key gas turbine inspections, over 3 years or longer for planned and/or unplanned maintenance, including all parts, repairs and servicing of the equipment. While, GE describes O&MA as a fixed-price total O&M and performance guarantee that guarantees the daily performance. The O&M is aimed at reducing the risks to the plant owner for future total plant price uncertainty, plant technology changes and equipment life. In general, GE Energy offerings are flexible and its CSA can be customized to cover what the plan owner deems necessary. At the end, GE and the owner share the common goal profit maximization for each other.

Siemens Power Generation calls its version of the long-term service agreement the Long Term Maintenance Agreement (LTMA). Siemens defines its LTMA as: “offers of power plant owners with turnkey outage services and equipment, long-term surety of pricing, priority emergency support, remote monitoring, and OEM technical support for base load gas plants”. Siemens offers LTMA contracts typically over a period of six to

twelve years and the contract could be extended to cover generators, steam turbine, and balance of plant (BOP) in combined cycle operation [4].

Alstom Power sells its long-term service agreement under the umbrella of its “Full Turnkey capabilities” with the marketing slogan of “You deserve the best” [5]. Alstom offers a customized partnership of total customer solutions, from components to full turnkey power plant [6]. The service available to customers is a complete portfolio of maintenance services, ranging from spare parts, repair and field services to full-range of operation and maintenance packages. The service can also include the refurbishment and modernization of existing plants; all of this flexibility is intended to meet specific customers’ needs. Furthermore, Alstom differentiates itself from other major OEMs with its concept of “Plant Integrator” which covers all aspects of power plant phases from design of plan component beyond providing the gas turbine through the commissioning process to the plant operation. The end goal of the Alstom strategy is to keep the plant owner competitive through the long term O&M by optimizing the capital investment and lowering operating and maintenance costs while guaranteeing plant operation and performance for the length of the long-term contract.

Like other major OEMs, Mitsubishi Heavy Industries (MHI) customizes its available options to meet customer needs from a simple long-term part supply contract (LTPS) which is a 4years to 6 years contract of guarantee parts supply and engineering support to a more complex long-term service agreement (LTSA). MHI provides its LTSA for its gas turbine and other power plant equipments for normally six years to twelve years term. Typically the LTSA of MHI consists of supplying necessary manpower, replacement parts, and maintenance engineering support. Specifically under its LTSA, MHI provides: maintenance management that is planning of scheduled inspection and hot gas parts management; maintenance engineering service that consists of having a site service director for daily support stationed at the customer power plant; a continuous remote monitoring service [7].

Ansaldo Energia acquired its full technological independence in 2004 with the termination of the Siemens license, becoming one of the major OEMs [8]. In its new position, Ansaldo Energia created the concept of OSP (Original Service Provider) to support its non-OEM service portfolio. Thus, it can provide service from on-call daily technical representative field dispatch to long-term turnkey maintenance contract. Under its LTSA, Ansaldo Energia engages in business partnerships with plant owners to share the revenues and the operational risks, to optimize costs of operation, to tailor maintenance policies and to meet the most demanding environmental regulations. Also, Ansaldo will guarantee throughout the duration of the contract the plant output, efficiency, technology upgrades and availability.

Because the competition among OEMs has become fierce with the LTSA offerings becoming comparable in scope, there are qualified third party consultants like Electric Power Research Institute (EPRI) which can help customers choose the right O&M strategy that meet their needs by providing unbiased comparative study of the offerings. EPRI typically assesses competitive offerings using they own tools for neutral decision in the best interest of its customers [3]. Thus, LTSA contracts are becoming more complex, creative, and flexible in their terms and can be individually customized. Therefore, OEMs are much more willing to take higher financial risks by guaranteeing the LTSA stipulations for several more years than their typical offerings to win contracts.

1.2 Motivation

As a result of the growing competition, non-OEMs and OEMs alike are exposing themselves to increasingly high financial risks by guaranteeing plant performance output, efficiency, environmental requirements in some cases, and taking on the risk associated with O&M for several decades. In fact, the maintenance costs and availability are two of

the most significant apprehension a heavy-duty gas turbine that plant owners faces [9] because of the financial consequence of unplanned outages. Thus for plant owners the LTSA is an insurance policy to mitigate their financial risk associated with wear and tear and/or equipment and component failure. Currently, O&M expenditures is a very important part of total life cycle cost consisting of 15% to 20% of total life cost cycle cost, while the equipment maintenance costs account for approximately 10% to 15% of total life cycle cost [2] as illustrated on figure 2 below along with the fuel cost and the cost of running the plant.

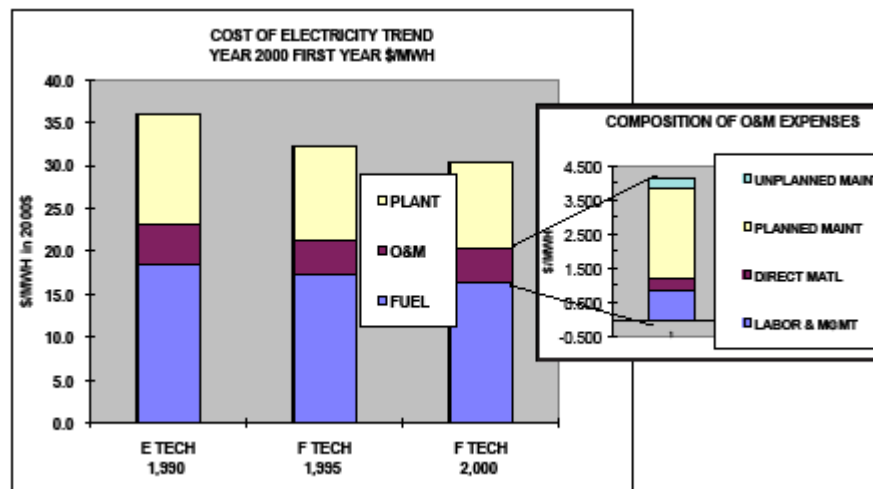


Figure 2: Cost of electricity trend with the O&M expenses [2]

Also a further decomposition of O&M expenses shows that there is always a cost associated with an outage whether it is a planned or unplanned one.

In return, the plant owner enters in a partnership with the LTSA provider in the form of revenue and risk sharing. Typically, the partnership works as a stick and carrot policy where any extra revenue due to savings on O&M or the excess of performance over the guaranteed level after an upgrade is considered a bonus and is shared between both parties. However, any loss of revenue due to an unplanned outage or a performance shortfall from the expected guaranteed level corresponds to a penalty. Furthermore, if the

plant owner is required under contract to provide a set level of electrical power to a district when the forced outage occurs, he or she will be obligated to buy the corresponding electrical power from other power plants at higher rates. Consequently, the LTSA provider has to compensate for the loss revenue to the plant owner and/or the cost to buy from other power plants for the necessary electrical power that the plant was dispatched for as a liquidated damage. It is clear that the liquidated damage can grow quickly and become very expensive for the provider in the case of extended forced outage. In general, a forced outage cost includes: loss production, repair cost, and eventual penalty.

1.2.1 The reasons for OEMs to enter in LTSA

Despite the risks involved with LTSA contracts, they are an enormous revenue stream for OEMs. Over the years, the profit margin on sale of new gas turbines has been shrinking for OEMs, while LTSA for the most part assured OEMs of the sale of upgraded components to sustained degrading gas turbine performance throughout the duration of the LTSA. The market of LTSA can be a very lucrative one as illustrated through the following example:

- On November 25th 2008, Business Wire reported that GE Energy and the Algerian state-owned power company Sonelgaz entered in a contractual service agreement (CSA) worth more than US \$1 billion for a period as long as 18-years [10].

Once a CSA or a LTSA contract is set, the goal of the provider (third party non-OEM or OEM) is to design ways to help reduce and manage to lower the level of maintenance costs which as shown on figure 2 is typically 15 to 20% of power plant life cycle cost.

1.2.2 OEMs develop strategy to make LTSA profitable

Like in any business endeavor, the main goal of OEM in providing LTSA is to make it profitable for both parties. With the steep liquidated damage amounting to multi-millions of dollars associated with not meeting the reliability and durability requirement, LTSA providers need to develop strategies so that the revenues generated from the contracts exceed the costs of the involved risks. Thus, OEMs have been investing huge amounts of money to develop strategies to avoid unplanned plant outages in the first place, or to be as well prepared as possible to mitigate the effects for those unavoidable ones. Thus, OEMs like GE Energy created a Power Answer in Atlanta, GA, where all power plants under the CSA contracts are continuously monitored using installed sensors; the data is recorded and stored for post processing to detect any abnormal trend.

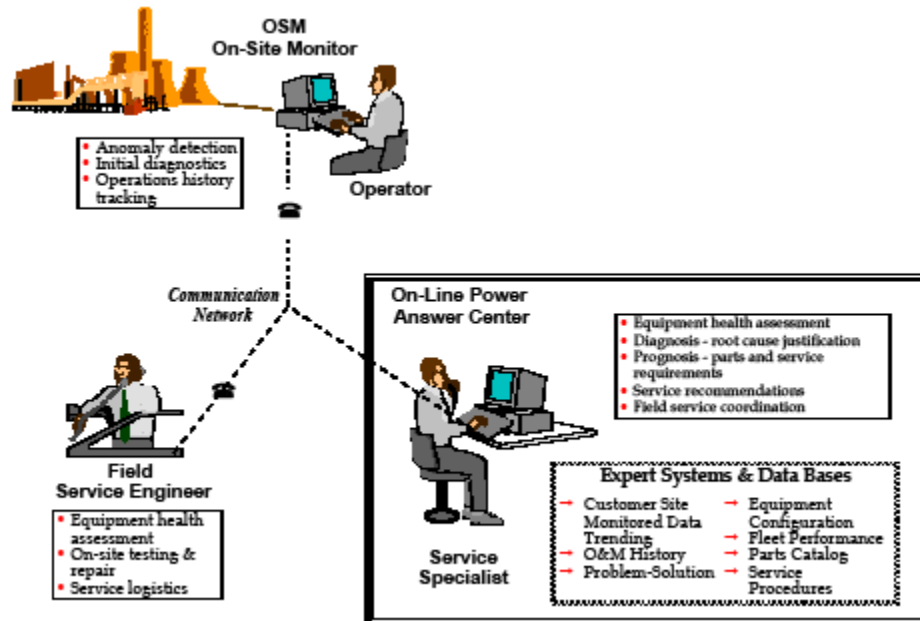


Figure 3: GE Monitoring & Diagnostics concept adapted from [2]

On the illustrative Figure 3 shown, the on-site monitor compares the actual unit performance with baseline predictions and provides the first level of anomaly detection and notification. The field service engineer is on hand to assess the equipment health,

while the engineer at the power answer center reviews the overall performance data on a daily basis and considers all abnormal events. There is an ongoing communication among the three parties for an adequate health parameters assessment of the covered equipments of the plant to identify signs of any equipment malfunction.

However, with the length of the LTSA contracts becoming longer with some them lasting more than twenty (20) years, the task of detecting all potential anomalies of gas turbine health is becoming daunting. There are many explanations for the difficulties in recognizing some of the rare or short lasting fault events that might lead to engine trips such as the machine-to-machine variation, uncertainties of machine degradation and power plant operating conditions (i.e. load condition). In addition, OEMs like GE Energy services provides long-term contractual service agreement to hundreds of utility companies throughout the world based on different power plants design and operating condition, making each CSA contract to be individually written to suite the plant owner interests and needs (e.g. plant Power, plant efficiency, plant emission, etc). On top of that gas turbine designs are becoming more complex in order to meet higher requirements for efficiency, output, and environmental regulations. All these constraints will add to the gas turbine vulnerability to uncommon faults. The strategy of continuously monitoring the plant equipment to detect early indication of faulty behaviors has become widely popular in the industry. However, although many companies have invested huge amounts of money in sensor technology, sensor accuracy, and other monitoring devices, and have significantly improved the storage capacity and access of the collected data in the past several years, advanced analysis of the accumulated monitored data is still in its infancy stage and very limited to simple data post-processing techniques like data trending.

As a consequence of the limited extraction of useful information out of the collected data and due to not taking full advantage of the capability acquired from the monitoring in real time, many failures such as first row (R-0) of both 7FA and 9FA of GE gas turbine compressors are still happening as reported in the EPRI report titled “GE FA Compressor

Dependability (Phase 2)” [11]. Those undetected failures defeat the goal of the power answer which goal is to detect anomalies that may have led to them, so to avoid them in the first place. Thus, in recent years, there have been new and improved techniques such as condition based monitoring (CBM) and prognostics health management (PHM), to detect anomalies in their early stages of development. Currently, the new techniques have not allowed to totally solving the issue of missed early detection of all the anomalies; although the merit of their concept is well accepted, their practical implementation is still not done in real time for the most part. Therefore, as a solution to avoid failure, in this work a twofold systematic approach to take full advantage of the real-time monitoring process is proposed. First, the equipment health in real-time using advance data processing techniques is continuously assessed. Then, once a fatal anomaly is detected, an estimation of the remaining useful life (RUL) is provided.

1.2.3 Relevance of Potential impact of research

The aim of this thesis is to propose a process to systematically detect anomalies at their earliest possible stage and to evaluate their severity. In the case where a detected anomaly is deemed fatal, then the potential survival time of the faulty equipment is estimated as to allow the power plant operators to make the appropriate decision whether to conduct an emergency shut down or to call for another power plan to come online to assume their power production.

The most significant accomplishment of this work will be the ability to capture the thus far undetected precursory events to catastrophic failures and shut down power plant so to avoid them. The ability to avoid or even decrease the number of catastrophic gas turbine failures has an immense economically consequence considering the steep cost of unplanned power plant outages to LTSA providers. It is understandable that detected and

replacing a faulty compressor blade would have cheaper cost in consequence to replace than to letting the blade detached and destroy the entire compressor. Because, the total cost of replacing a whole gas turbine compressor instead of a single compressor blade can add up quickly with a loss of revenue due to an extended outage, the cost of a maintenance crew for a longer period of time, and the cost of entire compressor equipment. As a matter of fact, EPRI estimates the cost benefit from preventing a General Electric gas turbine 7FA and 9FA technology compressor failure at around US \$10-20 millions [11].

Therefore moving the maintenance strategies from a concept of preservation and protection to that of asset management and optimization [8] will be the key for earlier detection of maintenance need. Overall, given that the LTSA contracts are as long as over twenty years, it is important for the O&M providers to become proactive to avoid as much as possible catastrophic failures that would consume their potential profits or even produce a negative return on investment on the long run as repeated costly failures would defeat the capitalist purpose of their prospect as insurance providers.

1.3 Research Organization

As stated previously, this research offers a methodic process for early failure precursor detection and model the residual time to failure once an anomaly is detected as a tool to decrease the cost of power plant outages. The remainder of this thesis is organized in six more chapters.

In chapter 2, the research is put in context with its background and the research goals are defined. The chapter 3 is an overview of the mathematical aspect of the time-frequency techniques, with the emphasis on the Wavelet Transform. Then, the chapter 4 is devoted to the first part of the proposed methodology, where each step of the failure precursor detection is presented.

Next, the chapter 5 is dedicated to the second part of the methodology, which is to model as accurately as possible the prognostication of the residual time to failure of gas turbine compressor after the detection of a failure anomaly. Thus, the deterministic regression, the non-parametric Kaplan-Meier, and the parametric survival analysis techniques are investigated. The chapter 6 is about implementing the proposed methodologies to a fleet of 7FA type technology of the General Electric gas turbines compressor as a proof of concept. Finally, in the chapter 7, a brief recapitulation of the proposed approach is provided with the successes and the challenges, followed by a few recommendations for future work to complement this research.

CHAPTER 2

BACKGROUND AND RESEARCH OBJECTIVES

2.1 Context

As explained in the introductory chapter, as a consequence of the deregulation of the electricity power generation, many new private investors coming into the electricity generation industry had no previous power plant operation management experiences. As a way to circumvent some of the financial risks associated with the power plants O&M, private investors outsource the operation and maintenance activities of their power plants to third parties and focused on the management side of the plants that led to the conception of third party insurance providers.

2.2 Background

Original equipment manufacturers (OEM) are the third parties natural fit to becoming insurance providers to cover power plant equipments like gas turbines. Therefore, the major gas turbine OEMs are also the main long-term service agreement (LTSA) provider. The LTSA contracts can usually last as long as over twenty years. There are expensive financial and reputational consequences for the OEMs in the cases where failures of equipments under coverage are the cause of unplanned outages. Accordingly, OEMs are investing substantial amount of money to develop maintenance strategies to avoid unplanned power outages or in case they were to occur to keep the disruptions as short as possible and the consequences to a minimum. So far the strategies have mainly been to establish rigorous maintenance plans and more proactive ways such as the installation of sensors and other monitoring devices to remotely assess the condition of the covered equipments. Thus, the next few paragraphs will devoted to an overview of the different traditional maintenance plans, then an OEM in-house

maintenance strategy will be presented, and some other proactive ways to improve the overall plants availability.

2.2.1 Traditional maintenance policies

2.2.1.1 Run-to-failure

Run to failure is the oldest and simplest type of maintenance strategies. It requires a repair, a replacement, or a restorative action be performed on a machine or a facility after the occurrence of a failure in order to bring this machine or facility to at least its minimum acceptable condition. Run-to-failure is a corrective maintenance policy also called "crisis maintenance" or "hysterical maintenance" because of its nature of being unplanned. When the run to failure strategy is in place, the equipment is used until failure occurs then the failed component is either repaired or replaced [12] (e.g. light bulb).

The run to failure method has some advantages [13]:

- It maximizes the useful life of the equipments
- It is a low cost strategy
- It requires a minimal management
- It can be useful on small non-integrated plant

On the other hand, the run to failure strategy has many disadvantages:

- It is not applicable in cases of irreversible damage
- It usually has a high downtime because maintenance crew are often not on hand
- It is done on an ad-hoc basis with a high cost of repairs
- It requires an unnecessary crisis management when failure occur
- It requires a large volume of spare inventory, which can be very expensive

2.2.1.2 Standard preventive maintenance

Preventive maintenance can be defined as a schedule of planned maintenance actions aimed at the prevention of breakdowns and failures. The main goal of preventive maintenance is to catch a potential failure of equipment before it actually occurs. There are two main preventive maintenance plans: periodic preventive maintenance and sequential preventive maintenance [14]. Unfortunately, because both types of preventive maintenance rely on schedule planned inspection to observe an imminent fault, there is still a possibility for a catastrophic failure happening between two inspections [12]. Preventive maintenance is an improvement over the run to failure strategy. Thus some of the advantages of the preventive maintenance are:

- Outages can be planned ahead of time with the possibility of logistic planning
- The downtime of maintenance can be reduced
- The maintenance occurs under the management control

But there are still some disadvantages to the preventive maintenance strategy:

- It leaves some unused useful life of the equipment
- It can be difficult to implement for varying failure patterns
- It can be expensive in terms of over maintenance
- It does not eliminate all catastrophic failures because they can still happen between scheduled maintenance

2.2.1.3 Condition-based maintenance

Condition-based maintenance can be defined as a planned maintenance based upon measuring the condition of all machine elements during the normal operation of the machine [13]. The goal of these measurements is to permit the prediction of the time to failure for all elements and thus to allow the maintenance activities to be planned before any elements fail. However, the concept of condition-based maintenance relies on the

recognition of a change in condition and/or performance of an equipment to trigger the implementation of the maintenance [15].

The condition-based maintenance is a more active strategy than the previous two. A number of the condition-based maintenance advantages [13] are:

- It decreases overall maintenance costs because maintenance is performed as needed when equipment deterioration is detected
- It can extend system life
- It has the ability to be machine specific
- It allows management and logistic control by planning maintenance operation ahead of time

There are disadvantages as well to condition-based maintenance:

- It still requires inspection of the equipment to assess potential faulty conditions, so catastrophic failures can still happen between inspections
- It relies on equipment deterioration, thus there is a chance to miss hard failure because of the short deterioration time see figure (4).
- It requires a continuous management effort

For the CBM to be successfully implemented, it is important to know the items failure modes in order to understand whether or not a given maintenance strategy would appropriately prevent the failure from happening.

In essence, the field of maintenance has been an evolving one as illustrated on Figure 4 [16]. Overtime, as systems became more complex, the maintenance strategies followed suit and became more proactive.

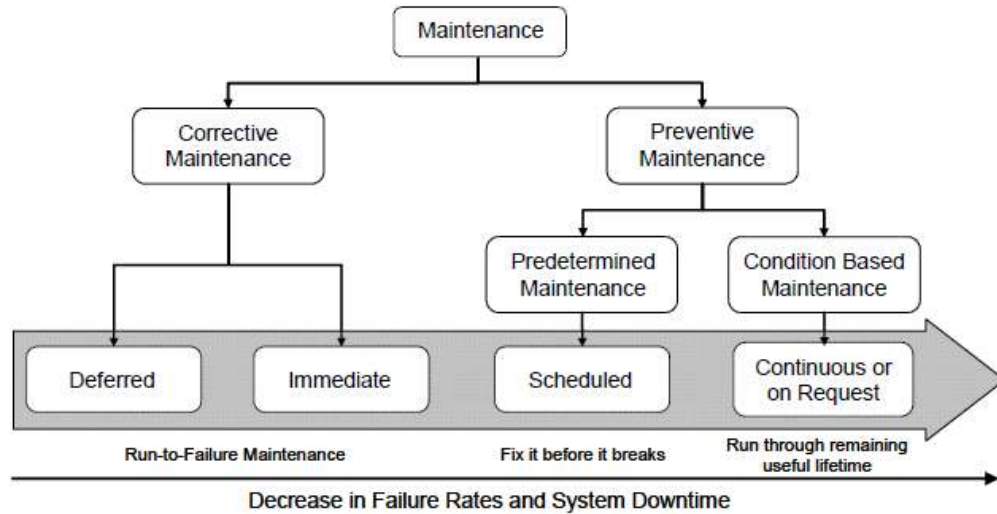


Figure 4: Paradigm shift in industrial maintenance [16]

Hence, advanced maintenance strategies such as CBM+ (developed by the DOD) and PHM are becoming widely accepted and implemented on complex systems because their maintenance policy decision is based on systems a health centric approach. Thus, a brief overview of the failure categories is presented followed by a manufacturer developed specific maintenance plan for its systems.

2.2.1.4 Types of failure [13]

There are two types of failure as illustrated on Figure 5:

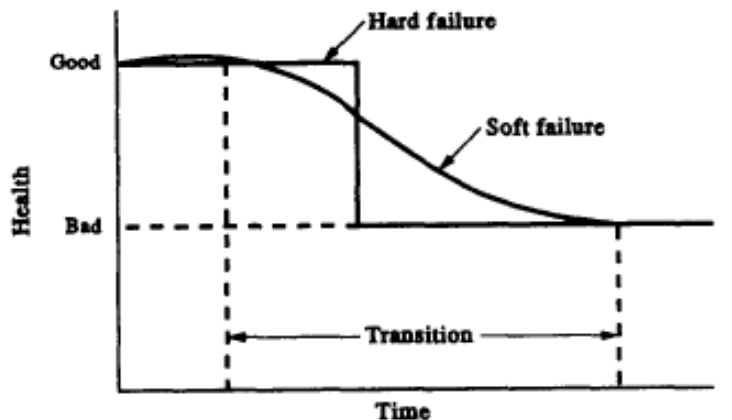


Figure 5: Illustration of Hard and Soft failures [13]

2.2.1.4.1 Soft failure:

It develops gradually over time. It is usually a characteristic of mechanical elements as the wear of the elements is the basis of a gradual degradation of its operation. Its signature is much easier to observe because of the trend in the performance degradation. Thereby, soft failure allows that a fault prediction could be remedy through an adequate implementation of preventive maintenance and the condition-based maintenance.

2.2.1.4.2 Hard failure:

Contrary to soft failure, hard failure tends to happen instantaneously. It is more a characteristic of electrical circuit. But it can occur in mechanical elements when the change in equipment performance is a short lasting one and/or a precursory catastrophic failure goes undetected [13]. Evidently, neither the preventive maintenance nor the condition-based maintenance would be suitable to avoid elements failure in the case of hard failure.

2.2.2 A OEM customized maintenance

Besides the well-known maintenance strategies introduced above, maintenance plans can be customized toward given interests of the users. Thus, in their 2006 paper titled “A Profit-Based Approach for Gas Turbine Power Plant Outage Planning”, Zhao et al. presented a method to plan maintenance actions so to maximize the profit of power plant operation [17]. Also in [14], Zhao et al proposed a way to plan preventive maintenance schedule with economic factors as the decision driver. General Electric, one the major OEMs has developed more tailored preventive maintenance techniques to meet its LTSA contracts commitment as an O&M provider. Therefore, based on its extensive experience acquired through the wide range of gas turbines it has sold and it is currently maintaining worldwide, GE bases its gas turbine maintenance need on the criteria of

independent counts of starts and hours. Whichever criteria limit is first reached determines the maintenance interval as shown on figure (5) [9], instead of the normal performance degradation trending as a tool to make a preventive maintenance decision.

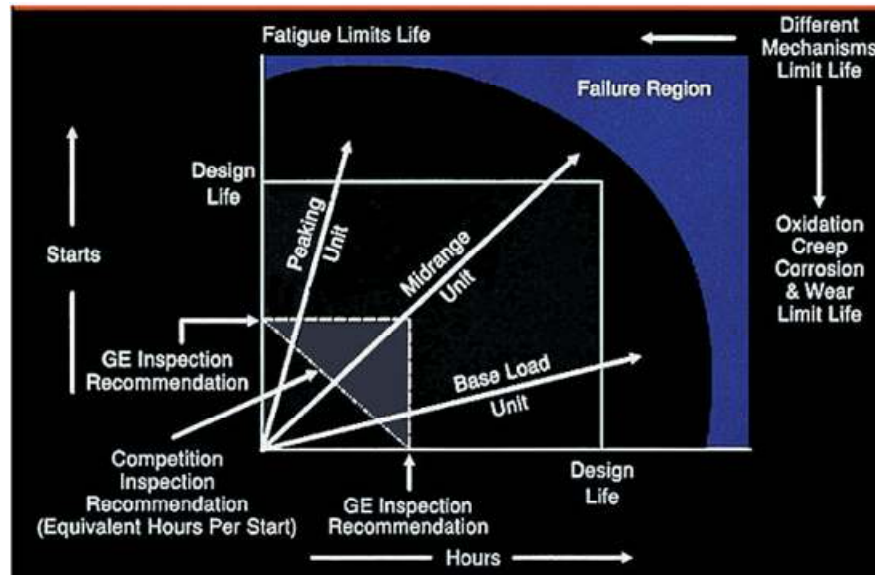


Figure 6:GE gas turbine maintenance need based on counts of starts and hours adapted from [9]

2.2.3 Remote Monitoring center

Thanks to the complex maintenance strategies developed by OEMs like GE, they are able to improve their preventive maintenance for soft equipments failure where there is usually a noticeable fault signature over time. However, those strategies can hardly prevent equipments hard failures. Therefore, there is a need to find more proactive ways to handle that type of failures. Hence as reported in the chapter 1, GE has invested resources in the creation of a monitoring center called “Power answer” to improve its ability to expand the existing maintenance plans to be more proactive by detecting indicative events in real time that might have led to catastrophic hard failure.

2.3 Observations

Through an analysis of the context and background a few observations can be made:

2.3.1 Observation 1

Even though a lot of the efforts and resources have been deployed through the development of complex and customized maintenance policies, and the creation of a monitoring center, there are still failure precursors going undetected. Indeed as reported by the EPRI in its report 1016159 in 2007, the GE gas turbines 7FA and 9FA technology compressor tilted: “blade cracks and several failure incidents prompted an independent root cause investigation” [18]. Furthermore, the fact that there was a second phase documented in report 1016269 [19] in 2008 to further study the same problem of compressor failure implies that the issue has become increasingly troublesome.

2.3.2 Observation 2

Usually all the machines covered under the LTSA contracts are continuously monitored in real time using installed sensors on the gas turbines and the readings have been stored for years in most of the cases. The stored sensors measurements data can be accessed at any time for post-processing of signals. Therefore, full advantage should be taken of the stored data, especially since there is no extra cost to be inquired for the monitoring process (i.e. use of existing sensors).

2.3.3 Observation 3

The sensors are set to make measurements at a regular interval of time. Among the sensors, there are sensors that measure the machine health condition, some sensors measure the vibration level of different subsystem (compressor, combustor, and turbine) of the gas turbine, other sensors measure the overall condition of the gas turbine performance (e.g. power output, efficiency), and some other sensors measure the

operating condition (e.g. load) and the environment condition (e.g. ambient condition measurement). Based on a definition of a time series as a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals, it can be concluded that sensors measurements overtime represent a time series signal [20].

2.4 Research Objectives and questions

Based on the observations, even with the existence of well-established preventive maintenance techniques in its arsenals, including its capability to monitor the gas turbine online monitoring, the GE 7FA and 9FA gas turbine compressor catastrophic hard failure has remained a problem. It is true that it could be difficult to avoid unplanned failure all together given the hundreds of gas turbines worldwide under GE's LTSA contracts. However if there can be a way to make sense of the events that preceded the known compressor failure, an approach could be developed to recognize similar signs in the future in other gas turbines so to avoid the same failure from happening. Thus, the main research objectives can be defined:

2.4.1 Objective 1:

Develop a process to find precursory failure signatures as early as possible in order to avoid systems catastrophic failure.

2.4.2 Objective 2:

After an anomaly is detected, develop a strategy to prognosticate an estimation of the residual time to failure.

As a consequence of the defined research objectives, the following research questions can be formulated:

2.4.3 Research question 1:

How can a precursory anomaly that might lead to a gas turbine catastrophic failure be detected?

2.4.4 Research question 1a:

How can a failure precursor detection method be made robust?

After a failure precursor is detected, it is important to confirm that any detected anomaly is indeed a true failure sign so to reduce or eliminate the number of false alarms. Because, a high number of false alarms can be detrimental to the practical implementation of any diagnostics techniques as the system operators in charge of taking required actions when warning signs are identified, may start to disregard those warning alerts after a given number of false alarms, which may defeat the purpose of the earlier detection of failure precursors. Thus, a second research question can be formulated:

2.4.5 Research question 2:

How can a detected precursory event be statistically validated?

2.4.6 Research Question 2a:

How does the sampling interval impact on the quality of detection?

Once a failure precursor is validated, the much more valuable information to plant owners is the time left before the unhealthy items actually fail. That yields to the third research question

2.4.7 Research question 3:

If a failure precursor event is validated, how can the residual time to failure of the item be estimated?

Since for power plants emergency shut downs can be costly, it is imperative to assess the probability of a health-compromised item ability to survive through the estimated remaining life. That concern lead to the fourth research question:

2.4.8 Research question 4:

How can the confidence level of the estimated residual time to failure be determined?

An attempt to answer these posed research questions leads to organizing the research approach into two main parts:

- 1) Failure precursor detection, followed by the validation of the detection
- 2) Estimation of the residual time to failure of a precursory anomaly, followed by the confidence level of the estimation.

In the contest of industrial and technology, and based on the following formal definitions:

- Prognostics is defined to be the detection of failure precursor followed by the prediction of remaining useful life (RUL) [21],
- Prognosis is defined to be detecting the precursors of a failure, and predicting how much time remains before a likely failure [22],
- Prognosis as the answer of the question: what is the remaining useful life of a machine or a component once an impending failure condition is detected and identified? [23],

The scope of the undertaken work to respond the research questions can be formally defined as a prognostics problem.

2.5 Previous work on fault diagnostics of time-frequency processing techniques

To answer the posed research questions, there is a need to review how time series data have been dealt with in the past to extract useful information about the content of the signals. Signal analysis is one of the most important methods used for condition monitoring and fault diagnostics [24]. In the literature, there exist many signal processing techniques that have been efficiently used to extract information from time series signal. As a way to process the time series data, Carden and Fanning recommend to convert the data from the time domain into the frequency domain using the Fourier transform [25]. Fourier transform is one of the oldest and the most used of the time-frequency domain analysis techniques. However, the Fourier transform is limited as it erases all time dependence once the data is converted in the frequential representation [26]. Accordingly, Peng and Chu provide an extensive review of other signal processing methods in [24]. In the past Fast Fourier Transform (FFT) which is an efficient algorithm to compute the discrete Fourier transform (DFT) and its inverse has been one of the signal analysis methods used for fault diagnostics [24]. But FFT is limited in dealing with non-stationary signals [24]. Because the sensor data are non-stationary, other methods have been contemplated like the Wigner-Ville distribution (WVD) [27, 28] and the short-time Fourier transform (STFT) [29, 30] that are able to transform one-dimensional temporal signal into a two-dimensional function of time and frequency [24, 31]. Though, the WVD and STFT are true-time frequency representations for the sensor data, they have some drawbacks. For instance, the STFT can only provide a constant resolution for all frequencies because it uses the same window for the entire signal

analysis. To address those flaws, other methods were brought about like the Choi-Williams distribution (CWD) [32] and the cone-shape distribution (CSD), but they all had some sort of disadvantages.

Therefore, to deal with the shortcomings enumerated with previous techniques, the wavelet theory has become the signal processing method of choice for time series analysis. Mallat and Hwang successfully used the wavelet to detect signal singularities [33]. Also, Jiang et al used wavelet methodology to detect damage in the thermal protection system panels [34]. Wang and McFaden had a series of publications on the use of wavelet to successfully detect mechanical failure [35, 36]. Boulahbal et al. found that the wavelet transform was appropriate for the detection of vibration transients generated by developing localized faults in gear trains [37]. Also, different modifications of the wavelet method and particular wavelet have been rightfully applied for fault detection. That is, Butler and Dey used the discrete wavelet to detect equipment failures for the purpose of preventive maintenance [38]. Then, Lin and Qu used the Morlet wavelet for mechanical fault diagnosis [39]. Saxena et al also used the complex Morlet wavelet to extract useful features to distinguish between faulty and healthy gear plates [40]. Mufti and Vachtsevanos applied a fuzzy-wavelet for fault detection [41, 42], while Wu and Du used the wavelet packets to extract signal features [43], which were necessary information for fault detection. Sun and Chang [44] also used the wavelet packet for damage assessment. Others like Pittner and Kamarthi used the wavelet coefficients instead of the wavelet transform components to extract signal feature [45]. One of the combinations much used for fault diagnostics and prognostics is the wavelet transform augmented with the neural networks, like in [46, 47].

The literature on the successful use of wavelet and its variants for the purpose of fault detection and prognostics is enormous.

Since the goal in signal processing for fault diagnostics is to extract dominant features so to make the accurate decision, it is important to choose an appropriate technique that

would not arbitrarily truncate some useful information. Because, there are many factors, like changing environment and faults from the machines themselves that make the output signals of the running machines contain non-stationary components, time-frequency analysis is a better choice for non-stationary signal analysis.

As a summary of the literature review for fault detection, it can be safe to say that though in the past dynamic systems based on vibration signatures and other health monitoring sensors data have generally relied on Fourier transforms type of analysis as a means of transforming time series signals from the time domain into the frequency domain, Fourier analysis proved to be limited as it provided poor representation of signals well localized in time. Therefore, it is difficult to detect and identify the signal pattern from their expansion coefficients. As a remedy, the wavelet methods in general and the wavelet packet transforms in particular appear to be a successful alternative means of extracting time-frequency information from vibration signatures and health sensors [48].

Table 1: Comparison of the different time-frequency method adapted from [24]

Methods	Resolution	Interference term	Speed
CWT	⊗	●	●
STFT	⊗	●	⊗
WVD	●	⊗	●
CWD	●	●	○
CSD	●	○	○

● Excellent ⊗ Good ● Fair ○ Poor

The Table 1 above compares the different time-frequency methods for the adaptive time-frequency analysis skill. The wavelet performs better than other methods, such as the

STFT, WVD, CWD and CSD. That explains the fast expansion of the wavelet techniques for fault diagnostics [24].

2.6 Literature review on residual life estimation

There are a fair amount of techniques in the literature used to estimate the remaining life of component after a failure precursor is detected. Many entities have in-house proprietary tools to assess the remaining life of their systems like the life extension analysis and prognostics (LEAP) developed by the Pacific Northwest National Laboratory [49]. Otherwise in general the prognostics techniques can be organized into three major groups [22]:

1. data-driven
2. Model-driven
3. Hybrid of data-driven with model-driven

2.6.1 Data-driven:

In a 2005 paper entitle “A survey on data-driven prognostics”, Schwabacher reports a list of fairly recent data-driven studies [22]. Vachtsevanos et al [50] report that artificial neural networks (ANNs) is a good candidate for prediction because the ANNs are self-adaptive and make few assumptions about the models. However, a major drawback is that standard statistical method for confidence estimation is not applicable to ANN. Thus, Vachtsevanos et al [50] and also Zhang [23] present the dynamic wavelet neural network (DWNN) as a tool that successfully accomplishes the prediction of the remaining useful life. Fuzzy logic has been selected as a way to make a component failure prediction [23, 51, 52]. Byington et al [53] put the neural networks and fuzzy logic as viable methods. And in [54] Byington et al maintain the neural network as a data-drive methodology for remaining life predictions for aircraft actuator components.

2.6.2 Model-driven:

In his PhD thesis, G. Zhang [23] defines the model-based approach as a physical model based that requires detailed and thorough understanding of the system as opposed to the data-driven approach. Also, Vachtsevanos et al define model-based as physics based [50]. Therefore in this research a model-based prognostics techniques is considered to be necessary a physics-based model. A list of recent works on artificial intelligence that is a model-based method can be found in [55].

2.6.3 Hybrid of data-driven with Model-driven:

The mixture of the data-driven and the model-driven approach can be very powerful as it makes use of the best of both worlds. Thus, Gebraeel [56] and Gebraeel et al [57] use a statistical approach with a preset threshold to compute residual life distribution. Peng and Vachtsevanos chose DWNN as the tool to address the problem of assessing the remaining useful life of critical components [46]. Suarez et al use the notion of damage accumulated to calculate the remaining life estimate of the components [58].

2.7 Assumptions

In this research, the following assumptions are considered:

- The required sensors (vibration, health, and operating condition) to continuously monitor the system are installed and working as intended.
- The required sensors are installed at appropriate position

- The proposed process to be hypothesized is based on the gas turbine current operating profile (e.g. gas turbine load)
- For a relatively short period of time, the number of starts and stops has no impacts on the remaining life of the machine.
- For the remaining life estimation, the gas turbine operating conditions remain unchanged after the estimated remaining life (e.g. the gas turbine continues to operate after the remaining useful life is estimated as it did before the failure precursor was detected during its healthy period)

This work is not intended to decide on the number or the position of monitoring sensors as other works that have focused on the optimal number of sensors needed and their location like in [23], or in the paper [59] that was devoted to the sensor location for the purpose of fault diagnostics study.

2.8 Hypothesis

Based on the literature review for time series signal analysis, the Wavelet transform methods are clearly the best choice for time-frequency analysis. As Percival and Walden stated in [60] wavelet is a powerful technique for signal processing, which was reiterated by Jiang and Adeli in [61]. Therefore, the following hypothesis is stated:

2.8.1 Research Hypothesis 1:

A multi-step process that relies on the Wavelet transforms theory can be used to detect a precursory anomaly that might lead to the catastrophic failure of a gas turbine compressor.

As the result of the literature review on the estimation of residual time to failure and remaining useful life, a method that relies on the artificial neural networks will be used. Therefore a second hypothesis can be formulated:

2.8.2 Hypothesis 1a:

Combining the information from the two types of sensors could decrease the number of missed precursory anomalies

2.8.3 Research hypothesis 2:

The X-bar control chart type of statistical process control can be utilized to establish the anomaly threshold, and thus distinguish the faulty events from the non-defective ones.

2.8.4 Research Hypothesis 2a:

The smaller the time step (the higher the frequency) is, the more accurate the detection quality will be.

2.8.5 Research hypothesis 3:

Survival analysis techniques such as deterministic regression, non-parametric and parametric analysis of failure data can be used to build models that estimate the residual time to failure.

2.8.6 Research hypothesis 4:

Existing mathematical techniques can be used to compute the confidence interval of the survival time estimation.

CHAPTER 3

MATHEMATICAL BACKGROUND OF WAVELETS

The goal of this research going forward is to answer the research questions and to validate the proposed hypothesis. Thus, the first hypothesis, which states that the wavelet transform theory could be used for the detection of a precursory anomaly that might lead to a gas turbine compressor catastrophic failure, has to be confirmed. First an overview of the time-frequency techniques for signal analysis is given, followed by a step by step explanation of the proposed approach to the detection failure precursor.

3.1 Analysis of Time-frequency of high frequency signals

The monitoring sensors types dictate the signal frequencies to being dealt with in this work. In general, the sensor measurements can be classified in two different groups:

- The static or process-related measurements which are typically used for temperature measurements, pressure measurements, flow rate measurements
- The sensors characterized by their high bandwidth used for vibration measurements, ultrasonic measurement [50].

The measurements illustrated by their high bandwidth are particularly high-frequency devices. That is, the vibration sensors measurements are necessary high-frequency measurements. Therefore to encompass all the different types of sensor measurements, the high-frequency signal processing techniques are retained. Among them, the Fourier transforms is one of the well known, though in recent years the wavelet transforms has been taking over as the premier choice, which coincidentally is based on the underlining idea of Fourier transforms.

3.1.1 Fourier transforms overview

Fourier transforms (FT) is the most popular frequency domain analysis technique because of its ability to decompose an energy limited signal $f(t)$ by its Fourier transforms $F(\omega)$ so to analyze the signal in the time domain $f(t)$ for its frequency contents $F(\omega)$ as defined by the following equation below:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \quad (1)$$

$$\text{Where, } F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (2)$$

However, the Fourier transforms gives global information on the frequencies of the investigated signal, it can not give local information if the spectra composition of a signal changes rapidly with time [62]. To address some of the limitations of the Fourier transforms, some particular cases have been developed like the Fast Fourier Transform [63] which is an efficient algorithm to compute the discrete Fourier transform (DFT) and its inverse. Whereas the short time Fourier transform (STFT) or windowed Fourier transform (WFT) is a true time-frequency analysis method [31]. The STFT concept consists of multiplying the signal to be analyzed $f(t)$ by an analysis window $\gamma^*(t-b)$, then calculate the Fourier transform of the windowed signal as followed:

$$F_f^\gamma(b, \omega) = \int_{-\infty}^{\infty} [f(t) \gamma^*(t-b)] e^{-i\omega t} dt \quad (3)$$

The main weakness of the STFT is that it can only provide a constant resolution for all frequencies because it uses the same window for the entire signal analysis [24, 64]. Although, any time-frequency analysis techniques like the FT and STFT could theoretically be used for the signal content investigation, the need to know the frequency content along with the location in time of where there exists an anomaly at different

resolutions make both the FT and the STFT unappealing. Consequently, those limitations of the Fourier transform and its special cases make the wavelet transform particularly an attractive time-frequency analysis technique.

3.1.2 Mathematical overview of wavelet transforms

There is a panoply of literature that explain the theory of wavelet transforms and its applications, going from the most basic presentations [65] to the more mathematically rigorous ones discussed in [66-68]. In a similar manner as for the Fourier transforms, the wavelet transform can be defined [43] for any function square-integrable function $L^2(\mathfrak{R})$. But instead of using the harmonics $e^{i\omega t}$, the wavelet transforms use wavelet basis ψ defined in equation (4) and called a mother wavelet function is used for the decomposition:

Error! Objects cannot be created from editing field codes.

(4)

Where: a is the dilation or scaling parameter and b is the time location or translation parameter. Thus a signal $f(t)$ can be decomposed as [67]:

$$f(t) = \frac{2}{C\psi} \int_{-\infty}^{+\infty} \int_0^{+\infty} (W_{\psi} f)(b, a) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \frac{dadt}{a^2} \quad (5)$$

The wavelet transform $W_f(a, b)$ of a signal $f(t)$ is computed as shown in equation

(6)

Error! Objects cannot be created from editing field codes.

(6)

Contrary to the Fourier transforms, the wavelet transforms is suited for approximating data with sharp discontinuities [69]. One of the main disadvantages of the Fourier

transforms compared to the wavelet transforms is that once a signal is Fourier transformed from the time domain to its frequency domain, all the information related to the time are lost, while the wavelet transforms preserve both the frequency-domain as well as the time-domain information. In other words, the wavelet transforms conserve the location and time of anomalies. Though the wavelet transforms were established on the model of Fourier transform, they decompose a signal into a wide range of basis as opposed to Fourier which decomposes signal into basis generated by the sine and cosines functions only; making the wavelet transforms such an improvement over the Fourier transforms for time series analysis. Wavelet transforms decomposes time series signal into two parts: a low frequency part that trends and smoothes the original signal (approximations) and a high frequency part that shows local properties such as anomalies (details). The wavelet transform decomposes signals by taking the original signal through two complementary filters introduced by Mallat [70]: a low pass filter that gives the approximation component (or scale) and a high-pass filter that yields the detail components as illustrated on figure (7) below:

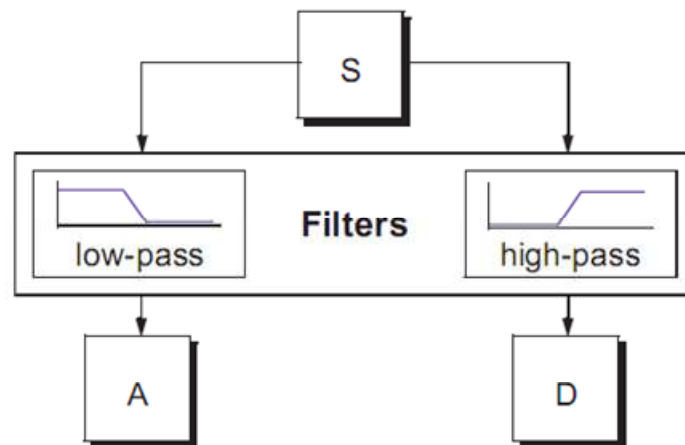


Figure 7: Notion of filtering for wavelet decomposition

One of the strengths of the wavelet transform is that the decomposition process can be iterated to obtain multi-resolution, with the approximations being successively

decomposed in turn, so that one signal can be broken down into many lower resolution components. Figure (8) is called the wavelet decomposition tree.

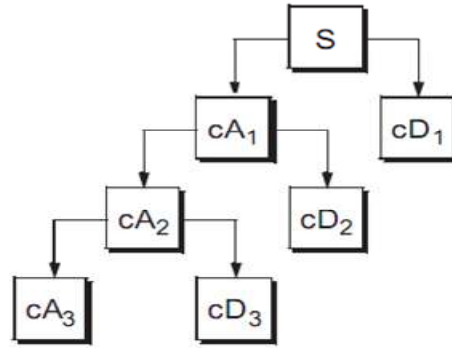


Figure 8: Illustration of the multiple-level decomposition

The multi-resolution decomposition is also referred to multi-resolution analysis (MRA) and can be formulated mathematically as well. Thus, first let j be a given decomposition level such that all details of the data series on scales smaller than 2^{-j} are suppressed at resolution j . The MRA decomposes the function space into a sequence of scaling function subspace V_j (subspace of functions that contains the information all the way down to 2^{-j}). By definition of the MRA, V_j is contained in the higher subspace V_{j+1} (e.g. $V_j \subset V_{j+1}$) so that the information at resolution j is included within the information at any higher resolution [71]. Let W_j be the wavelet subspace at level j and orthogonal to V_j . Then the following relation is established:

$$V_{j+1} = V_j \oplus W_j \tag{7}$$

That is V_{j+1} is the sum of V_j and W_j at the immediate lower level

For $j=1$: the preceding equation becomes:

$$V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus W_1 \quad (8)$$

This equation can be generalized as:

$$V_{j+1} = V_j \oplus W_j = V_{j-1} \oplus W_{j-1} \oplus W_j = \dots = V_{j-J} \oplus W_{j-J} \oplus \dots \oplus W_{j-1} \oplus W_j \quad (9)$$

Thus a nested relationship [72] can be written:

$$V_{j-J} \subset \dots \subset V_{j-1} \subset V_j \subset V_{j+1} \subset L^2(R) \quad (10)$$

Also, a charm of the wavelets is its ability to act as a “mathematical microscope”, where through the multi-resolution “the big picture” or “the context” is seen at coarse resolution or scale (large window) while the fine details are observed at the finest resolution (higher frequency or small window) [65] as illustrated in Figure 9. Graps refers to this wavelet features as the ability to see both the “forest” and the “trees” [69]. A result of the wavelet capability to gradually increase the resolution from coarse to fine enables many direct applications like in the pattern recognition algorithms [70].

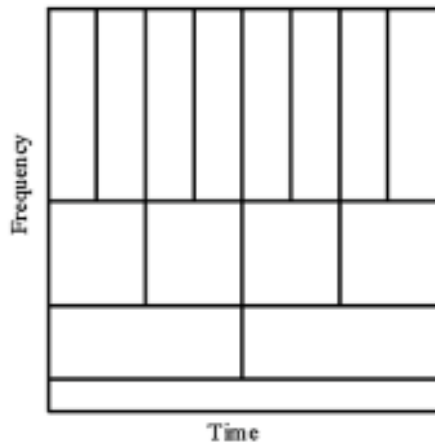


Figure 9: Illustration of Time-Frequency domain of Wavelet Transform

Like for the Fourier transforms, there are many specific wavelet transforms like the discrete wavelet transform (DWT), fast wavelet transform, and the wavelet packet, each of which is suitable for different applications.

3.1.2.1 Comparison between wavelet transform and Fourier transform techniques

There are some similarities as well as differences between the wavelet transform and the Fourier transform. As far as similarities, both the fast Fourier transform (FFT) and the discrete wavelet transform (DWT) are linear operations that generate a data structure that contains $\log_2 n$ segments of various lengths. Also, the matrices used for the transforms have similar mathematical properties, while the transpose of the original matrix is the inverse transforms of the FFT and DWT.

However, the dissimilarities are many. The most obvious one is that the individual wavelet functions are localized in space contrary to the Fourier sine and cosine functions. Another difference between the two transforms is that while the window size is constant for the Fourier transforms it varies for the wavelet transform [73] as illustrated on figure (10). This specific feature of the wavelet transform allows to have on one hand short basis functions to be used on short (i.e. over time) high-frequency to isolate signal discontinuities, and on the other hand to have long basis functions to be used on long low-frequency to get more detailed frequency analysis.

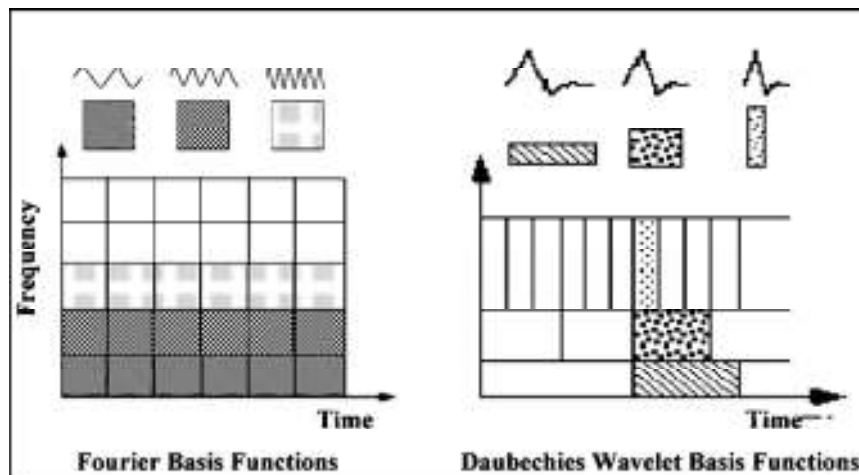


Figure 10: Illustration of the difference between Fourier transforms and Wavelet transform

3.1.2.2 Wavelet packet

One of the limitations of the standard wavelet is that only the approximations are decomposed into the subsequent subspaces (see Figure 8). Thus, Coifman and Wickerhauser [74] introduced the notion of wavelet packet to address that limitation in order to allow the details as well as the approximations to be further decomposed for permitting finer resolutions for both of them [61]. As such, the wavelet transform materializes as a subset of the wavelet packet [75], because wavelet transform is just a part of the wavelet packet (see Figure 11). The wavelet packets are particular linear combinations or superpositions of wavelets [76].

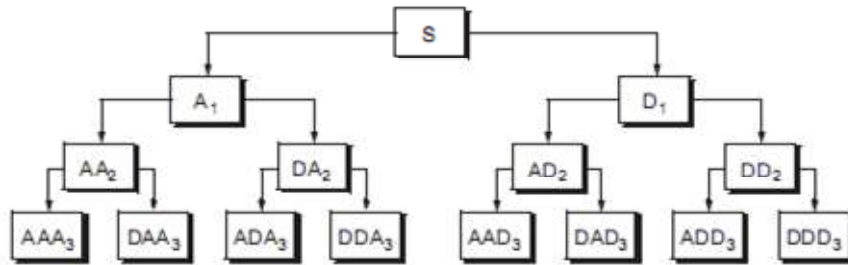


Figure 11: Wavelet Packet decomposition [77]

A Signal S can be decomposed at the level 3 as: $S=A1+ +AAD3+DAD3+ADD3+DDD3$.

Similarly to the MRA using wavelet, the MRA can be performed using the wavelet packet and can be written as the sum of the components it is decomposed into. That is for $j=1$: the following equation is obtained:

$$V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus W_{10} \oplus W_{11} \tag{11}$$

Where $V_1 = V_0 \oplus W_0$ and $W_1 = W_{10} \oplus W_{11}$

This process of simultaneously splitting both the scaling function V_j and the wavelet subspace W_j can theoretically be repeated indefinitely. But in practice the decomposition is repeated until a desirable level of resolution is attained. Thereby, the ability of the

wavelet packets to decompose both the approximations and the details parts provide more details about the signal than the standard wavelet transform can. Just like for its standard wavelet counterpart, which has the DWT, the wavelet packet also has the discrete wavelet packet transform (DWPT) [76].

CHAPTER 4

PROPOSED APPROACH FOR FAILURE PRECURSOR DETECTION

To efficiently tackle the posed problem, the proposed approach will be organized along two major axes: the identification of failure precursor and the prediction of the remaining life as illustrated in Figure 12. In the first step the inputs are the data from the system monitoring sensors, while the second step is initiated once there is validated information about the existence of a failure precursor. The output of the proposed process is the remaining useful life of an unhealthy item with its corresponding probability of successful survival through the entire estimated remaining life. Therefore the worth of the output is linked to the accuracy of step 2, which in turn depends on the quality of step 1

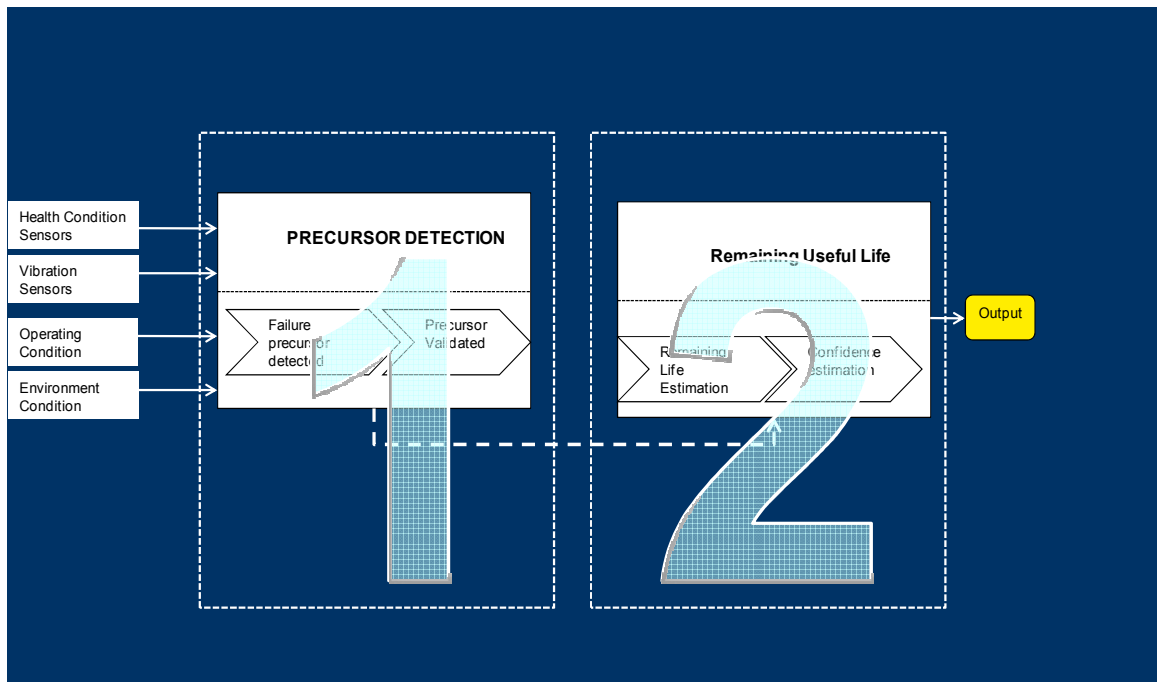


Figure 12: Illustration of the proposed approach organization

In this thesis, a detailed methodology is proposed to detect the precursors of catastrophic failure. The proposed process is presented following a step-by-step approach, where each step has been conceived to make the approach robust and efficient in terms of both cost and ease of use. Most damage monitoring techniques rely either on vibration-based damage detection [78-80] or health performance soft failure [81], or on some type of limited combination of both [82-84]. This research gets its strength from the combination of both the vibration sensors and the performance sensors measurements.

4.1 Detection of failure precursors

Though in recent years there have been new and improved techniques such as condition-based monitoring (CBM) to help detect anomalies in their early stages of development, the new techniques have not totally resolved the issue of missed detections for all the anomalies. Although the techniques' merit is well accepted, their practical implementation is still inefficient because these techniques tend to be theoretical, difficult, and/or expensive to apply to real world problems. Therefore, the method proposed herein intends to take advantage of the monitoring sensors to capture catastrophic failure precursors.

In general, the health and condition of power plants are monitored using two types of sensors:

- Static or process-related sensors (used to measure temperature, pressure, and flow);
- Sensors, characterized by their high-bandwidth, are used for high-frequency signals like the vibration measurements.

4.2 Steps for failure precursor detection

The approach presented in this thesis has been well thought out and methodically organized. Each step is a natural progression of the previous one, which enables the approach to handle precursor failure detection in complex systems with enough installed monitoring sensors. As mentioned above, the proposed approach intends to take advantage of monitoring sensors to capture catastrophic failure precursors. Therefore, the proposed process does not require any new financial investment in monitoring sensors; the existing ones will be used.

A step-by-step explanation of each block in the flowchart is presented in the subsections below. Also, the different steps of the proposed approach have been summarized in [85].

The process of failure precursor detection can be implemented as a systematic method, as shown on Figure 13, which starts with the identification and collection of the raw data from the sensors of interest using the monitoring procedure. Then, the selected time series sensors go through a pre-processing step where the raw data is transformed into a useful form of data. Then, the time series is de-noised using a discrete wavelet packets transform de-noising scheme. Next, the multi-resolution analysis (MRA), which is decomposed at an appropriate pre-defined level for maximum useful signal features extraction, is applied to the de-noised signal. After that, the multi variables sensor measurements are fused into fewer uncorrelated variables. The reduced number of variables resulting from the data fusion process goes through the anomaly detection procedure.

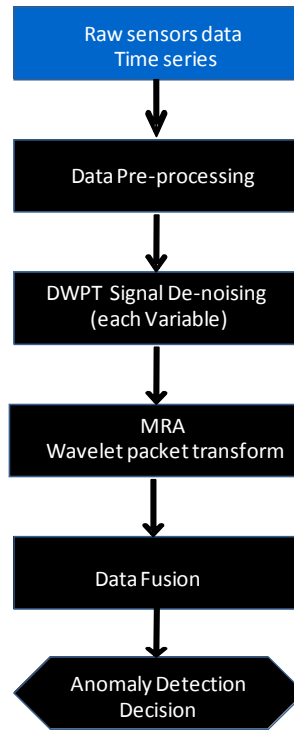


Figure 13: Precursor failure detection steps

The next few sections provide a detailed explanation of each of the failure precursor detection steps.

4.2.1 Raw sensors data collection

Modern systems have become more complex and integrated than ever before. It is very common to see failure-critical systems that have as high as several hundred sensors installed to monitor all facets of their health and operation conditions. Thus, one of the very first things to do in any practical health monitoring implementation is to identify the list of relevant sensors to be observed. Typically systems or subsystems with large numbers of monitoring sensors may require that some type of screening process be performed. When there are sensors that represent some type of response, techniques whose outcome is the identifications of main effects can be considered (e.g. Pareto plot,

multivariable analysis, scatterplot matrix, etc). This thesis will be applied to heavy-duty gas turbines that typically have two types of installed monitoring sensors: static or process-related sensors and high-bandwidth sensors used to measure high-frequency measurements. Furthermore, this study is limited to base load operation only of heavy-duty gas turbines. Also, the assumption is made that all the necessary sensors needed to perform the diagnostic and prognostic appropriately are installed at the appropriate location, as there are many other projects that have centered on the optimization of number and localization of monitoring sensors [23].

To illustrate how one decides on the different sensors to be included in a given study, let us consider a GE 7FA technology heavy gas turbine compressor as the subsystem of interest. Thus, the sensors that may be directly affected by anomalies in the subsystem of interest are selected. A quick screening of the sensor data, performed using a scatterplot matrix such as the one shown in Figure 14, reveals that two of the sensors did not correlate to any other sensors. Therefore, those two sensors may be eliminated, as they don't provide any useful information about monitored condition of interest (e.g. base load operation), and any changes in the other sensors that may indicate an anomaly will not have a signature in those sensors. Typically, a scatterplot matrix can be used as a very fast tool to look at trends and correlations between parameters, providing valuable information for making decision about which sensors will be monitored. The systems health and operating condition parameters are continuously monitored and collected using installed sensors and stored for potential post-processing.

The results of the screening process will determine which sensors will be included in the study.



Figure 14: Scatterplot Matrix of raw data

4.2.2 Data Pre-Processing

The pre-processing of the raw data is a necessary step for a study that relies on any type of data mining or data driven technique. Therefore, in the context of this work, which is a data driven methodology, the pre-processing stage is an essential one for a couple of important reasons. First, the OEMs will not want to share their proprietary data on equipment malfunctions because that may affect their competitive advantages due to

the risk of reverse engineering. More importantly the pre-processing is needed to avoid undesirable scaling effects [86]. The approach outlined in this paper intends to detect anomalies based on different types of sensor measurements, such as health monitoring sensors, which are static, process-related sensors (e.g. pressure, temperature, flow, etc), system condition sensors, which are also static (e.g. gas turbine output, gas turbine efficiency), and high bandwidth sensors (e.g. vibration sensors). However, these sensor values are recorded in different units and with different orders of magnitude.

For instance, a typical normal base load operation of GE's 7FA+e gas turbine technology can have a compressor discharge temperature measurement in the range of 600 to 800 degrees Fahrenheit, while the vibration sensor measurements could be on the order of 1/10 of an inch per second. Therefore, an analysis with the raw measurement could be artificially skewed towards the variables with higher absolute values. Care should be taken when deciding the type of data normalization techniques, because data normalization has a very visible effect on the results of given experiments and the validity of the conclusions drawn [87]. While there are many data normalizing techniques, whose corresponding robustness and efficiency are summarized in the Table 2 below [88], the pre-processing step used in this work consisted of normalizing each measured parameter value by the mean value of that variable measurement. That approach put all the different types of sensors readings in the same order of magnitude, and thus eliminated the potential outlier measurements that would misrepresent the findings and affect the accuracy of the conclusion.

Table 2: Summary of Normalization Techniques [88]

Normalization Technique	Robustness	Efficiency
Min-Max	No	N/A
Decimal scaling	No	N/A
z-score	No	High
Median and MAD	Yes	Moderate
Double sigmoid	Yes	High
tanh-estimators	Yes	High
Biweight estimators	Yes	High

The min-max normalization, which is the simplex normalization widely used, is defined as:

$$S'_k = \frac{S_k - \min}{\max - \min} \quad (12)$$

Where S'_k is the normalized value of a value S_k of attribute A and for $k=1, 2, \dots, n$

The min-max normalization technique performs linear transformation of the original data; therefore, it can only shift the minimum and maximum to 0 and 1 [88].

As for the decimal scaling normalization, it normalizes a value v of an attribute A by moving the decimal point, computing normalized v' as: $v' = (v / 10^n)$ [89].

In practice, the decimal scaling normalization can be applied when the scores of different matchers are on a logarithmic scale.

The z-score is the most commonly used score normalization technique, and it is calculated using the arithmetic mean and standard deviation of the given data. It can perform well if prior knowledge about the average score and the score variations of the matcher is available. However, to work well, it relies on prior knowledge of the nature of the matching algorithm. The normalized value v' of a z-score normalization of a value v of an attribute A is calculated as: $v' = (v - \mu) / \sigma$. The z-score method of normalization is useful when the actual minimum and maximum of attribute A are unknown. However, the z-score normalization does not guarantee a common numerical range for the normalized scores of the different matchers because if the input scores are not

Gaussian distributed, z-score normalization does not retain the input distribution at the output [88].

The median and MAD (Median Absolute Deviation) normalization is more robust (has more insensitivity to outliers and to the points in the extreme tails of the distribution) than the previous methods. The disadvantage is that the median and the MAD estimators have a low efficiency compared to the mean and the standard deviation estimators, and when the score distribution is not Gaussian, median and MAD are poor estimators of the location and scale parameters [88]. This method normalizes a value v of a given attribute A by computing v' as follows: $v' = (v - \text{median}) / \text{MAD}$, where $\text{MAD} = \text{median}(|v - \text{median}|)$ [90]. The major drawback of this normalization technique in the context of the study is that it does not retain the input distribution and does not transform the scores into a common numerical range.

The double sigmoid normalization is based on the use of the double sigmoid function [91]. It transforms the normalized scores into intervals with values between 0 and 1. The double sigmoid normalization is not as straightforward and simple as the previous methods, as it requires a careful tuning of the parameters needed to calculate the normalized value in order to achieve a good efficiency. It works by providing a linear transformation of the scores in the region of overlap, while the scores outside this region are transformed non-linearly.

The tanh-estimators is one of the robust normalization techniques [92]. It was introduced by Hampel et al. and it relies on the influence function (ψ) [93]. It is a highly efficient technique; the influence of the points at the tails of the distribution during the estimation of the location and scale parameters makes this method insensitive to outliers. However, the parameters needed to estimate the normalized value must be carefully chosen in order to ensure the amount of robustness required, which in turn depends on the estimate of the amount of noise in the available training data. Consequently, because there are so many conditions needed to have a good normalization, this technique is not

appropriate for the proposed study, as there is little knowledge of the parameters. Furthermore, one of the goals of this thesis is to achieve a methodology that can easily be implemented on other systems even with little background knowledge, so the ability to apply the characteristics of the raw data to the normalized data is paramount.

The Biweight estimator was introduced by Mosteller and Tukey [94] and is based on the biweight location and scale estimators, which are robust and efficient. The Biweight estimator normalization techniques are iterative in nature; that is, an initial estimate of the biweight location and scale parameters is chosen, and this estimate is updated based on the training scores. Because the major drawback is that it is only applicable to Gaussian data, the Biweight estimator technique is not appropriate for the current study.

Although all the normalization techniques presented above have some features that make them attractive, a much simpler normalization scheme is used in this study. In this thesis a defect will depend on derivatives and the scaling of the variables so that having them over the same range ensures that the derivatives will not be biased by the magnitude of a particular variable. Therefore, each variable will be normalized using the following very simple equation, one that shifts the mean value of each of the sensors to be monitored to a value of 1:

$$S_N = \frac{S_M}{S_\mu} \quad (13)$$

Where : S_N : is the normalized sensor value

S_M : is the actual sensor measurement

S_μ : is the mean value of the actual sensor measurements

As a result of the normalization, the following figures show that the characteristic distributions of the actual sensor measurements are conserved after normalization.

That is, Figure 15 shows a health sensor that has preserved the same distribution for the actual measurement (Figure 15-a) and the normalized measurement (Figure 15-b)

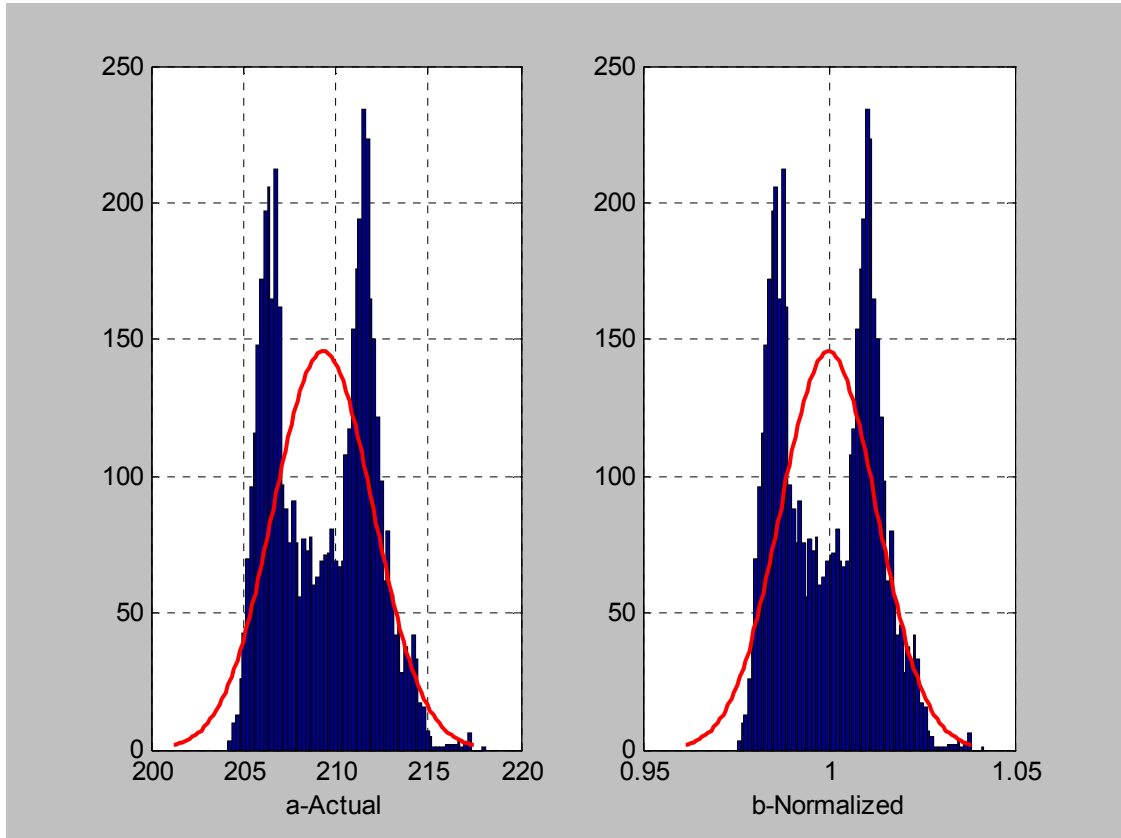


Figure 15: Actual and Normalized Health sensor distributions

Similarly, Figure 16 shows a vibration sensor that has preserved the same distribution for the actual measurement (Figure 16-a) and the normalized measurement (Figure 16-b)

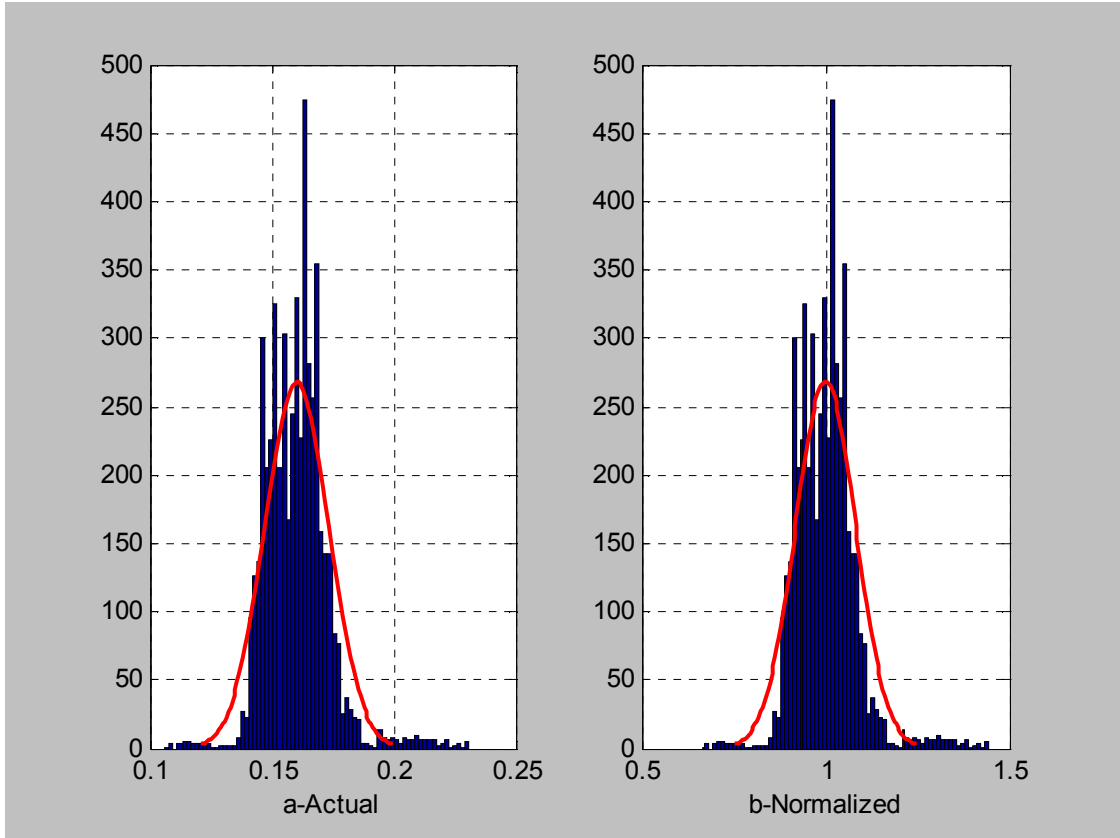


Figure 16: Actual and Normalized Vibration sensor distributions

4.2.3 De-noising

After making the decision about the different sensors to be included in the study, and after each sensor measurement has been normalized following the scheme presented in the previous section, each sensor normalized value must be de-noised.

The de-noising step is an essential one because it is not possible to know with certainty that the signal of a sensor measurement is not tainted by noise [50]. As a matter of fact, in [95] the authors argue that sensor signals are always inevitably corrupted by noise.

4.2.3.1 Time series de-noising techniques

Research in signal de-noising is a very active area within the signal processing field. Sriram et al. present a list of de-noising techniques for signal processing in [96] that is summarized in

Table 3 below. Each of the signal denoising techniques in this table has its own advantages, and some limitations, depending on the type of application.

The first category of techniques in the table is the Fast Fourier Transform (FFT)-based technique; this category can be implemented using either a constant threshold or a frequency-dependent threshold. The use of the FFT expedites the denoising process because the FFT is a discrete Fourier transform (DFT) algorithm that reduces the number of computations for N points from $2N^2$ to $2N \cdot \log_2(N)$. The FFT-based denoising works the following way for a given signal:

- First, the FFT of the signal is taken, and the FFT coefficients that fall above a determined threshold are dropped;

- Then, the signal is reconstructed using the inverse FFT.

The difference between the two FFT-based techniques is that, in the first, the threshold remains constant for all frequencies, while in the second the threshold varies according to the principle: $\text{threshold} = (j-1)^2 * 60 + 10$ ($j=1, \dots, n$, with n: the n-point FFT is taken).

This method is mainly used to remove the sinusoidal noise [96], which has pronounced peaks in the frequency domain. Thus, this method will not be used in our study, because the noise type is not predetermined.

Table 3: Categories of Signal Denoising Techniques (adapted from [96])

Category	Signal Denoising Techniques
1	Fast Fourier Transform (FFT) -Constant Threshold
1	Fast Fourier Transform (FFT) - Frequency-dependent threshold
2	Low-pass Filtering (Butterworth filter)
2	Low-pass Filtering (Chesbyshev filter)
2	Low-pass Filtering (Inverse Chesbyshev filter)
2	Low-pass Filtering (Elliptic filter)
3	Wigner-Ville Distribution (WVD)
4	Short-Time Fourier Transform (STFT)
5	Least Mean Square (LMS)
5	Leaky LMS
5	Sign-error LMS
5	Sign-data LMS
5	Sign-sign LMS
5	Mormalised LMS
5	Kurtosis-driven LMS
5	Adaptive Recursive LMS
5	Cascade adaptive filtering
6	Frequency-Domain Adaptive Filtering (FDAF) using DFT
6	Frequency-Domain Adaptive Filtering (FDAF) using DCT
6	Frequency-Domain Adaptive Filtering (FDAF) using DWT
7	Recursive Least Squares (RLS)
7	Exponentially-weighted Recursive Least Squares (EWRLS)
8	Matched Filtering
9	Notch Filtering, direct implementation
9	Notch Filtering, lattice filter implementation
9	Notch Filtering, direct implementation
10	Wavelet-based (Thresholding)
10	Wavelet-based (Mallat's algorithm)

The next category is low-pass filtering methods. In general, low levels of noise are concentrated in the low frequency region, while the sinusoidal components are usually in the high frequency region. For this reason, when dealing with low levels of white noise, a low pass filter can be used to extract the noise efficiently. The “Digital Filter Design” by Parks et al. [97] contains details about digital low pass filters. Typically, increasing the filter order improves the performance of the filter. However, the low pass filtering methods will not be used as the de-noising technique in this thesis.

The Wigner-Ville Distribution (WVD) of a given signal is a time-frequency representation of that signal [98]. The basis for using WVD as a de-noising technique is that WVD is quadratic in nature, so a signal is divided into blocks, and the WVD of each block is taken and processed. At every time step t the block is integrated in a small range around the center frequency by adding (for upper value) and subtracting (for lower value) the step frequency; if the value exceeds the predetermine threshold, it means the presence of a signal at time t . The other alternative to this WVD-de-noising technique is based on masking the time-frequency [99]. Like in the first method, the signal is divided into blocks, and the blocks are multiplied by a 2-D matrix with ones in the frequency range of the signal and zeroes elsewhere, which masks the noise components [96]. Finally, the desired de-noised signal can be reconstructed by the synthesis algorithm presented in [99].

The Short-Time Fourier Transform (STFT)-based de-noising relies on the STFT of a signal, which is another time-frequency representation of the signal. In the STFT, a given preset length window is used, where the length of the window determines the time and frequency resolutions of the STFT. Thus, the first of the two STFT de-noising techniques is based on the calculation of the STFT of the input signal, multiplying it by a 2-D array of ones and zeroes (masking) and then reconstructing the signal from the modified STFT as outlined in [100]. The second alternative for STFT de-noising is based on the estimation of the signal from the modified STFT magnitude. After the calculation

of the STFT magnitude of the input signal, a first estimate of the actual signal x is computed. Then that STFT magnitude is replaced by the available STFT magnitude. The signal is estimated, and that value is used as the next estimate of the signal. The process is then repeated.

Though the STFT de-noising technique is a true time-frequency method, it relies on a fixed value of window size. Consequently, it will not be retained as the de-noising technique of choice in this thesis.

The next technique is the Least Mean Squares (LMS) de-noising approach which has many variants. All the LMS-based de-noising techniques are based on the least mean square process which is an adaptive iterative gradient search method [96]. It works by using a filter coefficient vector, an error vector, and an input vector. The desired de-noised signal is the input itself or the input delayed by one sample. Its first alternative in the above table is the leaky LMS, which modifies the main LMS approach by using a constant leakage factor that has a value between 0 and 1. As consequence, the leakage allows the impact on the filter coefficient vector of any single input sample to decay with time. The remaining alternatives of the LMS-based de-noising techniques have the advantage of reducing the computational burden however, they are not nearly as efficient [101]. Thus, the sign-error LMS, the Sign-data LMS, the Sign-sign LMS, the Normalized LMS, and the Kurtosis-driven LMS all have a similar expression but different filter length and mean value. The Normalized LMS, on the other hand, is used to prevent instability in the LMS algorithm, while the two Kurtosis-driven LMS alternatives use the ‘kurtosis’ of the error. The Adaptive recursive LMS relies on a set of equations. In the case of the Cascade adaptive filtering, the convergence can be improved by cascading a number of adaptive filters.

The next de-noising technique is the Frequency Domain Adaptive Filtering (FDAF). The FDAF is explained in [101, 102]. When the adaptive filtering is done in the frequency domain FDAF it is computationally more efficient than when it is done

using the LMS algorithm [96]. The FDAF relies on an equation based on input from the discrete Fourier Transform (DFT), discrete cosine transforms (DCT), or discrete wavelet transforms (DWT) concerning the corresponding quantities in the time-domain for a given filter length. A main advantage of the FDAF over the LMS algorithm is that the FDAF (using FFT) necessitates $3*N*\log_2(N/2)$ whereas the LMS requires $2N^2 + N$ multiplication.

The Recursive Least Square (RLS) and the Exponentially Weighted Recursive Least Square (EWRLS) methods are used to alleviate the weaknesses of the LMS methods [96, 101] because they are faster and have better uniform convergence. However, these de-noising techniques rely on cumbersome recursion procedures. Also their computation burdens are much higher.

The Matched Filtering technique uses a matched filter. This technique works with a matched filter arrangement, where the impulse response of the filter is matched to the signal input. The underlying decision is based on the fact that the output signal-to-noise ratio is maximal for a matched filter [96].

The notch filtering-based de-noising techniques rely upon the fact that a notch filter can be used to remove the sinusoidal noise present in the signal [96]. The notch filtering has to meet a set of constraints presented in [103]. The variation between the different notch filtering-based techniques is the use of two different algorithms; one uses the filter directly, and the other uses lattice filters.

The wavelet-based de-noising techniques are among the newest and most elegant time-frequency de-noising techniques because they address many of the limitations of the traditional signal de-noising methods. One of the reasons wavelet-based de-noising techniques have gained popularity very quickly is that they perform so much better than other methods like the Fourier-based ones. The Fourier-based techniques do pretty well when the signal and the noise are located in different bands of the spectrum; however, they do not work well when the time series is chaotic [61]. Furthermore, in the context of

this work, the Fourier-based de-noising approach is not appropriate because there are quick changes in the frequency domain of a signal; they are spread out over the entire spectrum. Thus, the use of Wavelet-based techniques for de-noising is a fairly active area of research because of the ability of wavelet techniques to address the limitation of traditional de-noising techniques. Chapter 3 of this thesis has an extensive explanation of wavelet theory. The first of the two de-noising methods based on wavelets uses thresholding of wavelet coefficients, and the second one uses Mallat’s algorithm presented in [104].

The wavelet-based de-noising method that uses thresholding is based on a scheme called “Wavelet Shrinkage and Thresholding Methods,” proposed by Donoho [105, 106].

Thus, Donoho further shows that wavelet can be used to optimally de-noise a noisy signal using a thresholding in a three-step process [107]:

1) Apply the wavelet decomposition of the noisy signal to obtain wavelet coefficients.

2) Apply:

- The soft thresholding nonlinearity

$$\text{Threshold value } \eta_i(t) = \begin{cases} \text{sgn}(y_i(t)) \cdot (|y_i(t)| - \delta), & \text{if } |y_i(t)| > \delta \\ 0, & \text{otherwise} \end{cases}$$

- The hard thresholding [96, 108]

$$\eta_i(t) = \begin{cases} y_i(t), & \text{if } |y_i(t)| > \delta \\ 0, & \text{otherwise} \end{cases}$$

Where δ is the threshold parameter for de-noising.

3) Reconstruct the signal using the inverse wavelet decomposition with the thresholded coefficients.

The de-noising technique using Mallat’s algorithm works by decomposing the signal into several scales (wavelet decomposition). Then, the detail coefficients are

neglected and the signal is reconstructed using only the approximation coefficients. The reasoning behind this approach is that the major part of the noise components is contained in the detail coefficients; therefore, by eliminating the detail coefficients, the noise will be eliminated. The problem with this approach is that the detail coefficients can contain more than noise alone.

Based on the advantages and disadvantages of the different de-noising techniques, the wavelet thresholding-based approach will be appropriate in this work because little is known of the form of time series signal. In fact, the wavelet shrinkage and thresholding methods have been successfully applied to de-noising noisy data [69, 105, 107]. Also because it is known that the two types of monitoring sensors are static and high bandwidth, wavelet-based de-noising can remove the low amplitude and high frequency noise effectively [61]. In the previous chapter on wavelet theories it was demonstrated that, for the standard discrete wavelet transform, only the scaling functions are decomposed into subspaces. The resulting time-frequency resolution has narrow bandwidths in the low frequencies and wide bandwidth in the high frequencies. Consequently, it is not sufficient for de-noising time series that contain a low signal-to-noise ratio [109]. Thus, the wavelet packet decomposition introduced in [74] allows a finer and more adjustable resolution that yields a better level of detail about the noisy signal. Therefore, instead of decomposing the noisy signal using the wavelet as proposed by Donoho, the discrete wavelet packet decomposition should provide extra details. Thus, Jiang et al. have used the discrete wavelet packet transform (DWPT) instead of the conventional discrete wavelet transform (DWT) [61, 109]. Also, in [110] the authors present a signal de-noising method based on wavelet packet transform.

4.2.3.2 Signal De-noising using DWPT and Bayesian Thresholding

In [109] Jiang et al. took the idea of using the wavelet thresholding approach proposed by Donoho even further by making use of prior knowledge by replacing the conventional discrete wavelet transform with discrete wavelet packet transform and soft thresholding with Bayesian thresholding.

The result of combining the discrete wavelet packet transform (DWPT) and Bayesian thresholding is the adequate removal of the noise, whether it is in the low or high frequency domains. Thus, Jiang et al. proposed applying Bayesian DWPT de-noising procedure for time series signal $f(t)$ in a five step process as shown on Figure 17. The interested reader can see a detailed explanation of the Bayesian wavelet theory in [109].

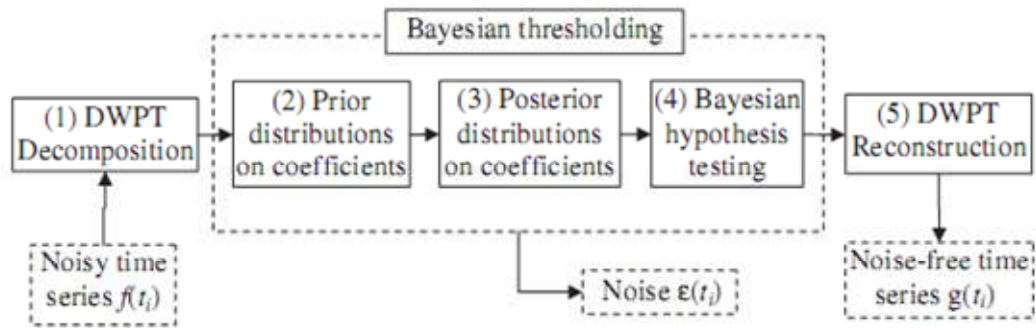


Figure 17: Bayesian DWPT de-noising procedure for times series [109]

A brief account of the five steps of Bayesian DWPT de-noising techniques follows.

4.2.3.2.1 Step 1: the noisy signal is decomposed into series wavelet coefficients by a DWPT

Let's assume a signal representing a time series is contaminated by an additive white Gaussian noise

$$f(t_i) = g(t_i) + \varepsilon(t_i) , \text{ with } i=1, 2, \dots, N$$

Where: $f(t_i)$ is the noisy experimental signal, $g(t_i)$ is the noise-free signal.

The noise $\varepsilon(t_i)$ being assumed as a Gaussian noise, it has the distribution $N(0, \sigma^2)$ of mean 0 and standard deviation σ .

It is interesting to note that $f(t_i)$, $g(t_i)$ and $\varepsilon(t_i)$ are all time series data of their own.

Jiang et al. argue that the de-noising problem becomes a univariate non-parametric regression problem, in which the goal of the Bayesian DWPT de-noising approach is to recover the underlying noise-free time series $g(t_i)$ from the noisy $f(t_i)$.

The noisy signal is decomposed using the DWPT and the wavelet coefficients are obtained using the decomposition scheme presented in Chapter 3.

4.2.3.2.2 Step2: Prior distributions on the decomposed coefficients

Because the noise $\varepsilon(t_i)$ is also a time series, it can be looked at as having a noise level ε_{jk} so that it can be decomposed in turn into coefficients \hat{d}_{jk}

$$\hat{d}_{jk} = d_{jk} + \sigma_j \varepsilon_{jk}, \text{ With } j=0, \dots, J-1; \text{ and } k=0, 1, \dots, 2^j-1$$

Now ε_{jk} can be considered in turn. The distribution $N(0,1)$ and d_{jk} are the noise-free coefficients. Also, the decomposed coefficients conditionally on d_{jk} and σ_j^2 can be:

$$\hat{d}_{jk} | d_{jk}, \sigma_j^2 \text{ is a Gaussian distribution } N(d_{jk}, \tau_j^2)$$

Furthermore, the assumption that a non-informative prior distribution of d_{jk} is:

$$d_{jk} | \gamma_{jk} \text{ is a Gaussian distribution } N(0, \gamma_{jk} \tau_j^2)$$

Where γ_{jk} is a binary random variable with independent Bernoulli distribution π_j ; that is, the probability $P(\gamma_{jk} = 1) = 1 - P(\gamma_{jk} = 0) = \pi_j$. Consequently, the determination of whether the coefficients d_{jk} are zero ($\gamma_{jk} = 0$) or non-zero, ($\gamma_{jk} = 1$) and the variance τ_j^2 , represents the magnitude of d_{jk} at the j^{th} decomposition level.

4.2.3.2.3 Step 3: Posterior distributions of the coefficients using the Bayes' theorem

The third step is based on the use of the Bayes' rule and the combination of these two:

- $\hat{d}_{jk} | d_{jk}, \sigma_j^2$ is a Gaussian distribution $N(d_{jk}, \tau_j^2)$
- $d_{jk} | \gamma_{jk}$ is a Gaussian distribution $N(0, \gamma_{jk} \tau_j^2)$

To evaluate the posterior distribution:

$$d_{jk} | \gamma_{jk}, \hat{d}_{jk}, \sigma_j^2 \text{ is a distribution } N\left(\gamma_{jk} \frac{\tau_j^2}{\sigma_j^2 + \tau_j^2} \hat{d}_{jk}, \gamma_{jk} \frac{\sigma_j^2 \tau_j^2}{\sigma_j^2 + \tau_j^2}\right)$$

4.2.3.2.4 Step 4: Remove the noise from the coefficients using a Bayesian hypothesis testing

The decomposed coefficients are thresholded by testing the following two hypotheses:

$$\text{Null hypothesis } H_0: d_{jk} = 0$$

$$\text{Alternative hypothesis } H_1: d_{jk} \neq 0$$

The hypothesis is tested using a thresholding rule to threshold the wavelet coefficients as follows: $\tilde{d}_{jk} = \hat{d}_{jk} I(\eta_{jk} < 1)$, called the Bayes factor thresholding [111].

Where: $I(\cdot)$ is an indicator function, which is equal to 1 when $\eta_{jk} < 1$ (η_{jk} is defined as the posterior odds ratio of $\gamma_{jk} = 0$ versus $\gamma_{jk} = 1$).

The results of the hypothesis testing are:

- 1) H_0 is rejected and the coefficients d_{jk} is estimated by \hat{d}_{jk}
- 2) Otherwise, H_0 is not rejected $d_{jk} = 0$ and \hat{d}_{jk} is thresholded

4.2.3.2.5 Step 5: The de-noised data is reconstructed through an inverse wavelet transform

Finally, after the result of the hypothesis is established that the thresholded coefficients \tilde{d}_{jk} are the wavelet coefficients that survived thresholding [112], the coefficients are used to reconstruct the noise-free signal by using the inverse discrete wavelet packet transform. Finally, the signal-to-noise ratio can be used to assess the goodness of the de-noising process.

To illustrate the step of the Bayesian DWPT de-noising procedure for times series, let's assume a function $g(t)$ is of the Mexican family defined as:

$$g(t) = c \cdot \exp(-t^2 / 2) \cdot (1 - t^2), \text{ with } t = [-5, 5];$$

Where the constant $c = \frac{2}{\sqrt{3} \cdot \pi^{1/4}}$

Then, let's add some noise with variance $\varepsilon(t) = 0.1 * randn$

Let us define the noisy data as: $f(t) = g(t) + \varepsilon(t)$

The application of the Bayesian de-noising process has been applied to $f(t)$, with the Daubechies 10 discrete wavelet decomposition.

Figure 18-a and Figure 18-b below represent the noise-free Mexican hat signal corresponding to $g(t)$ and the noisy signal (noise added) $f(t)$ respectively.

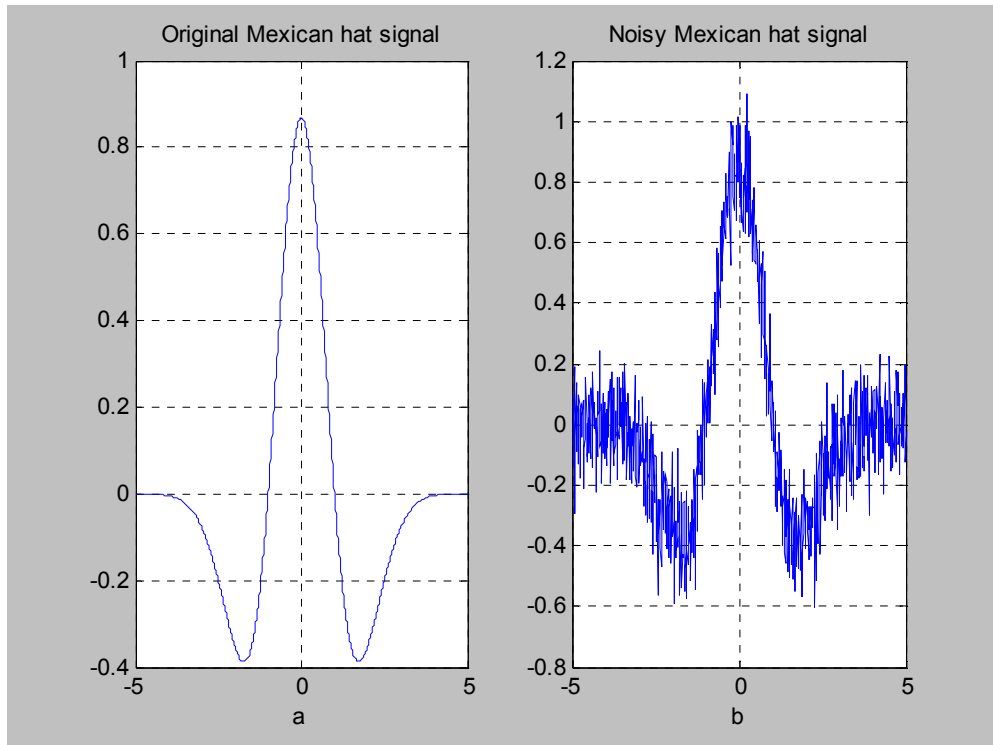


Figure 18: Original signal and Noise added signal

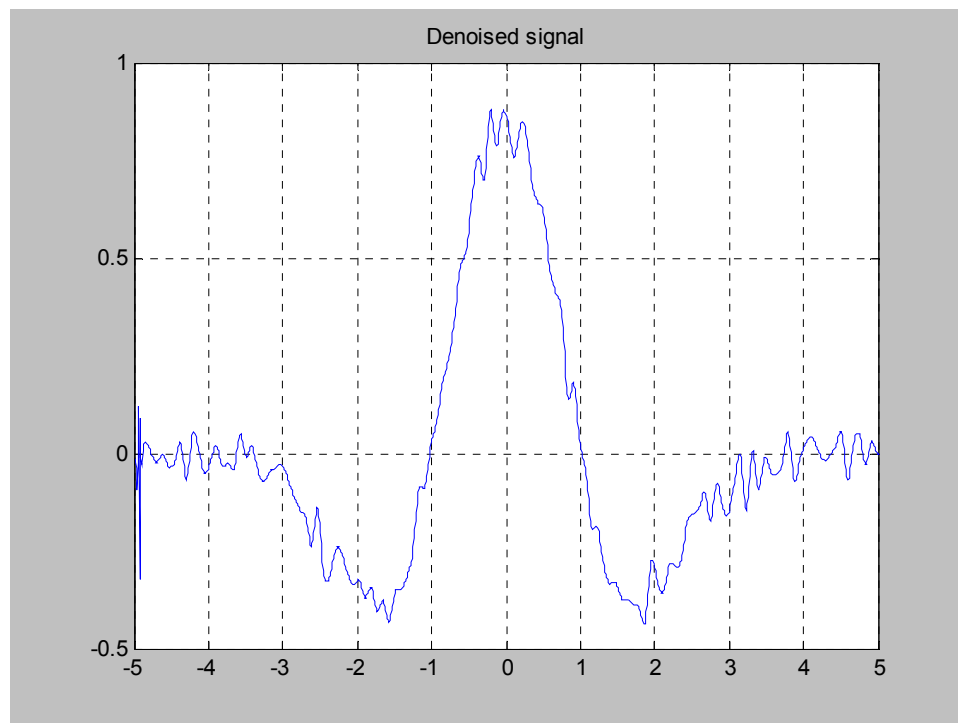


Figure 19: Denoised signal using the Bayesian DWPT denoising procedure

As illustrative Figure 19 shows, the Bayesian DWPT denoising method has successfully removed the noise from the original data. Therefore, the proposed Bayesian DWPT denoising procedure is a sufficient noise removal. The issue with many denoising algorithms is that they either remove too much useful information while taking out the noise, or they remove too little while leaving some noise in the signal. For this illustrative example, the signal-to-noise ratio is about 10.99 decibel units.

4.2.4 Multi-resolution analysis using discrete Wavelet packet decomposition

The goal of this step is to extract the feature of the signal to be used in the subsequent step. In the literature there are many feature extraction techniques for the time series signals. Among the most popular are the discrete Fourier transform (DFT) [113], the discrete wavelet transform (DWT), and the discrete Gabor transform (DGT) [114]. A short description of the popular feature extraction techniques is as follows

4.2.4.1 Discrete Gabor Transform

The Gabor distribution was introduced by Gabor, and works by mapping a time domain signal $s(t)$ into the joint time and frequency domain [115] as follows:

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{m,n} h_{m,n}(t) \quad \text{Where } h_{m,n}(t) = h(t - mT) e^{jn\Omega t}$$

With T and Ω representing respectively the time and frequency sampling intervals. Also, the synthesis function $h(t)$ has the constraint energy of unity. The DGT has the powerful ability to represent the time history of the frequency content of a signal, using the data as a superposition of uniform and frequency translates of a window function. Although that ability allows the DGT to be an effective technique in signal detection, feature extraction, and identification [116], it has limitations. The major issue with the DGT is that its coefficients $C_{m,n}$ are uniquely defined. Due to the difficulty of computing its unique

coefficients [114], the DGT has been very limited in its usefulness. Therefore, the DGT will not be used as the feature extraction technique in this thesis.

4.2.4.2 Discrete Fourier Transform

The Discrete Fourier Transform (DFT) was briefly explained in the previous chapter. The DFT has been successful and widely used as a signal feature extraction technique. The DFT is the projection of a signal from the time domain into the frequency domain.

Recall that the Fourier transform of a signal $f(t)$ is defined as:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt, \text{ where } f(t) \text{ can be calculated in turn as } f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega$$

The FT can be generalized to the case of discrete function, $f(t) \rightarrow f(t_k)$ by allowing $f_k \equiv f(t_k)$, where $t_k \equiv k\Delta$ (time step), with $k = 0, \dots, N-1$. Therefore, the discrete Fourier transform [117] can be derived as:

$$F_n = \sum_{k=0}^{N-1} f_k e^{-i2\pi \cdot k \cdot n / N}$$

The energy content of a sampled signal $f(t)$ can be easily evaluated:

$$E(f(t)) = \sum_{t=1}^n (f(t))^2$$

One of the main attractions of the DFT in general and its application as a feature extraction method is the fact that its coefficients can be calculated very fast using the Fast Fourier Transform (FFT) [118]. Thus, there is extensive literature on the different ways the DFT has been used as a feature extraction tool [113, 118].

However, with the introduction of the wavelet transform, which addresses some of the limitation of the Fourier transform, the DFT is no longer the technique of choice in many data mining problems. In fact, a good feature extraction technique is very important in any mining problem. The DFT can only project a signal into the frequency domain, while

the DWT can project the same signal into a tiling of the time-frequency plane [113]. Furthermore, there are many studies that have qualitatively compared DFT and DWT [119], and which emphasize the advantages of DWT over DFT. Thus, the DWT-based techniques are the natural choice because they enable an advanced level of feature extraction to be accomplished.

4.2.4.3 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a particular form of the family of the wavelet transform. Similar to the DFT, the DWT is a discretized version of the continuous wavelet transform (CWT). However, the DWT provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time, which is one reason it is so attractive for signal feature extraction applications. The DWT of any integrable real function $f(t)$ can be represented as:

$$f(t) = \sum_{j,k} c_{j,k} \psi_{j,k}(t), \text{ where the basis functions are } \psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k),$$

with ψ as the mother wavelet function and $c_{j,k}$ as the coefficients of the DWT. Or, the DWT coefficients can be calculated as the inner product $c_{j,k} = \langle \psi_{j,k}(t), f(t) \rangle$.

The value of the DWT is that it is considerably easier to implement than is the CWT. As a consequence of its capability to measure frequency at different time resolutions and locations, the DWT has enabled the MRA.

Furthermore, the advantage of using the discrete wavelet packet transform (DWPT) instead of the conventional discrete wavelet transform DWT has been explained in the chapter on the wavelet overview; therefore, the DWPT will be used for the MRA

4.2.4.4 The multi-resolution analysis (MRA)

The multi-resolution analysis (MRA) was thoroughly explained in the chapter on the wavelet overview. This multi-resolution analysis (MRA) using the wavelet transforms was introduced by Mallat [70]. The MRA, when it uses signal decomposition, is a very powerful technique in providing unprecedented understanding of time series sensor signals because of its “zoom-in, zoom-out” property [120]. Since its introduction, the MRA, when it uses wavelet packet distribution, has been used extensively in research that focuses on sensor-based techniques such as diagnostic and prognostic.

In this step, called “the multi-resolution analysis using the discrete wavelet packet decomposition”, each of the de-noised sensor signals obtained as the result of the denoising step will be decomposed at different levels in terms of the scaling function (approximation) and wavelet functions (details).

Then, the content of each the components resulting from the decomposition can be analyzed. The first necessary step in the MRA is to determine the appropriate level of decomposition. Although in theory a MRA can be decomposed indefinitely, in practice the number of levels of decomposition is defined based on the type of signal and the goal of the analysis to be achieved, such that the signal representation minimizes some cost function [72]. That is, the appropriate level of decomposition will be set when further decomposition cannot yield any further details. In [74] Coifman and Wickerhauser developed entropy-based algorithms to determine the best basis.

In this work the energy content of the scale function (approximations) and wavelet function (details) of each node of the decomposed tree will be utilized to determine a suitable level of decomposition. Once the decomposed tree is obtained using the wavelet packet (approximation and details), the energy content at each node is calculated [48] as:

$$E_{j,n} = \int_{-\infty}^{\infty} (W_{j,n,k}(dt))^2 dt = \sum_k W_{j,n,k}^2 \quad (14)$$

Where $W_{j,n,k}$ is the wavelet packet coefficient, j is the level, k is the translation, n is the modulation parameter (approximation or detail). The energy content of each node will then be used as the signal feature.

As applied to this thesis, the de-noised signal is decomposed at an appropriate level of resolution to get the approximation and the detail components. The resulting content of each component can be analyzed using the calculated energy content. Furthermore, there are readily available MATLAB® source codes such as “wenergy” that implement Coifman and Wickerhauser’s algorithm for given a time series. As previously stated, the main advantage of using the wavelet packet for the MRA instead of the standard wavelet transform is that the wavelet packet analysis allows “for a finer and adjustable resolution of frequencies at high frequency (details)” [61].

As an illustrative MRA decomposition, let us decompose the de-noised signal of a vibration sensor X1 from the list of the monitored sensors at different levels of decomposition:

- 1) When we apply level-2 decomposition using the Daubechies 7 (db7) wavelet function, we get the tree decomposition shown on Figure 20 with the signal of the first detail node (2,1).

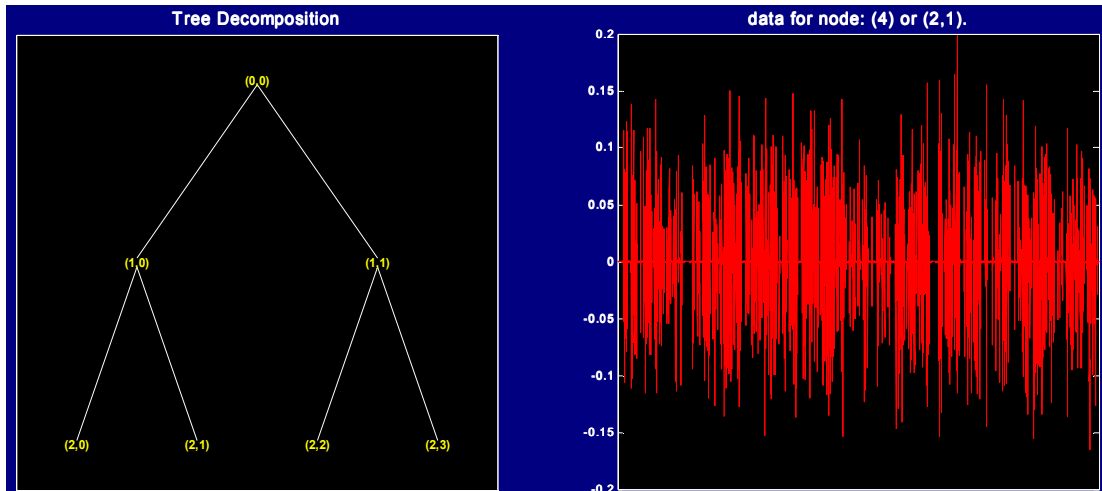


Figure 20: Level-2 Tree DWPT decomposition of vibration sensor X1

The corresponding energy calculated at each node is shown in Table 4

Table 4: Level-2 decomposition Node energy content for vibration sensor X1

Decomposed tree Node	(2,0)	(2,1)	(2,2)	(2,3)
Energy Content (%)	99.945	0.0269	0.011	0.0168

Figure 21 shows the corresponding MRA figures of the original de-noised signal (black). The second figure is the approximation signal (red), and the remaining three signals are the details signals (blue).

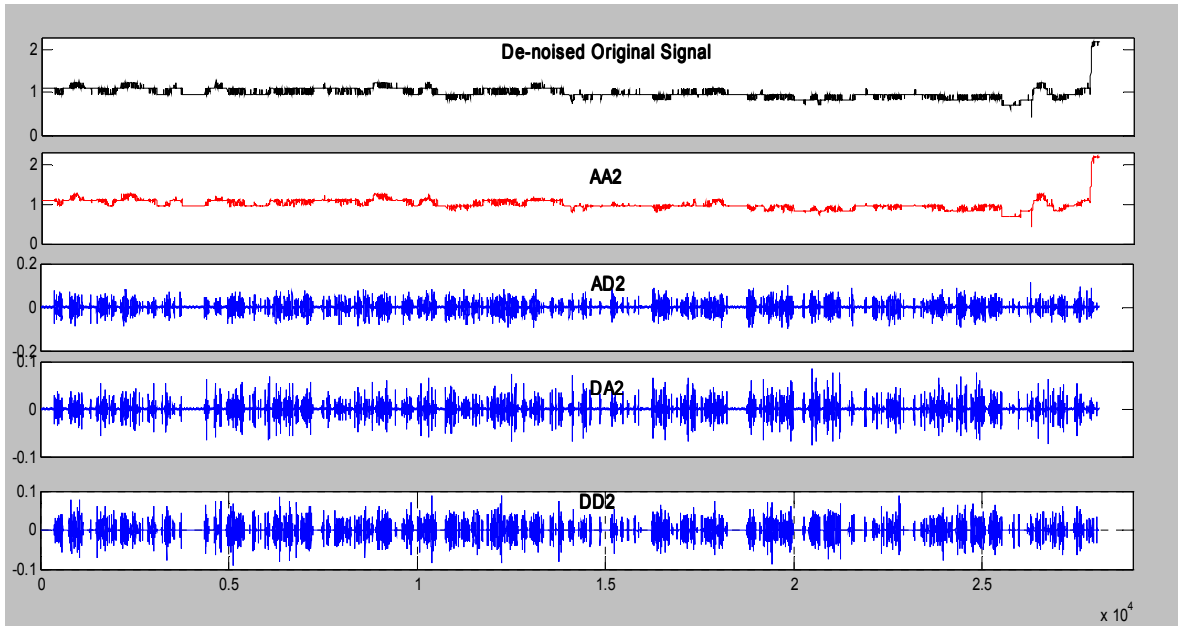


Figure 21: MRA level-2 DWPT decomposition of vibration sensor X1 signal

- 2) Similarly, the level-3 decomposition is done using the Daubechies 7 (db7) wavelet function. The decomposed tree is shown in Figure 22 with the first detail signal of the node (3,1)

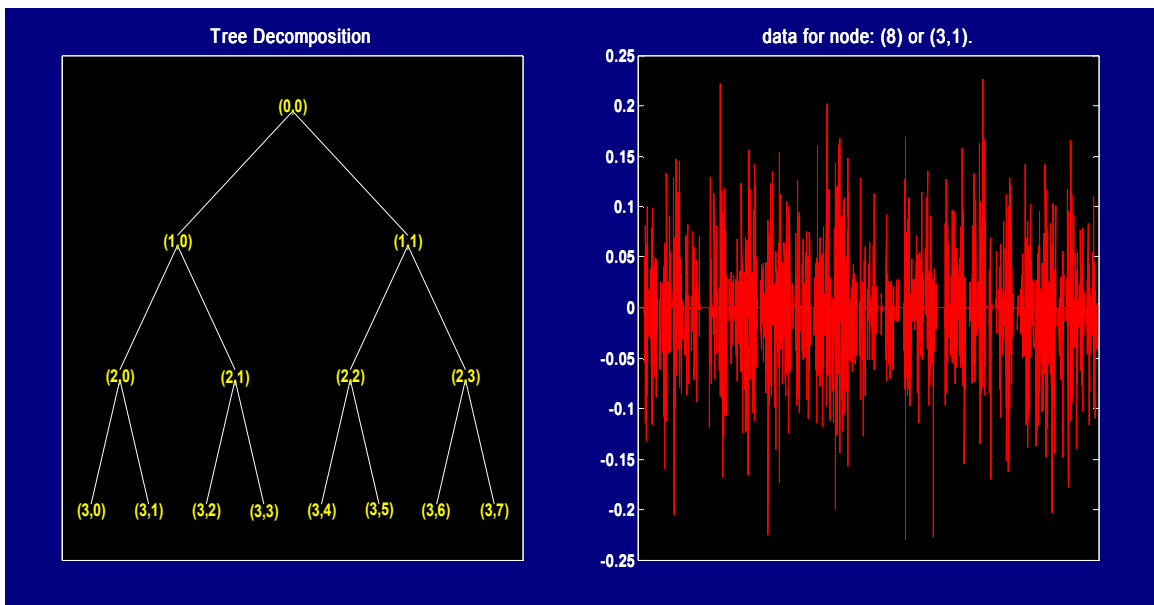


Figure 22: Level 3 Tree DWPT decomposition of Vibration sensor X1

The corresponding energy calculated at each node of level-3 decomposition is shown in Table 5.

Table 5: Level 3 decomposition Node energy content of vibration sensor X1

Decomposed tree Node Energy Content (%)	(3,0)	(3,1)	(3,2)	(3,3)	(3,4)	(3,7)	(3,6)	(3,7)
	99.92384	0.021543	0.011609	0.015248	0.004727	0.006283	0.00965	0.007101

Figure 23 shows the corresponding MRA figures of the original de-noised signal (black), The second figure is the approximation signal (red), and the remaining three signals are the details signals (blue).

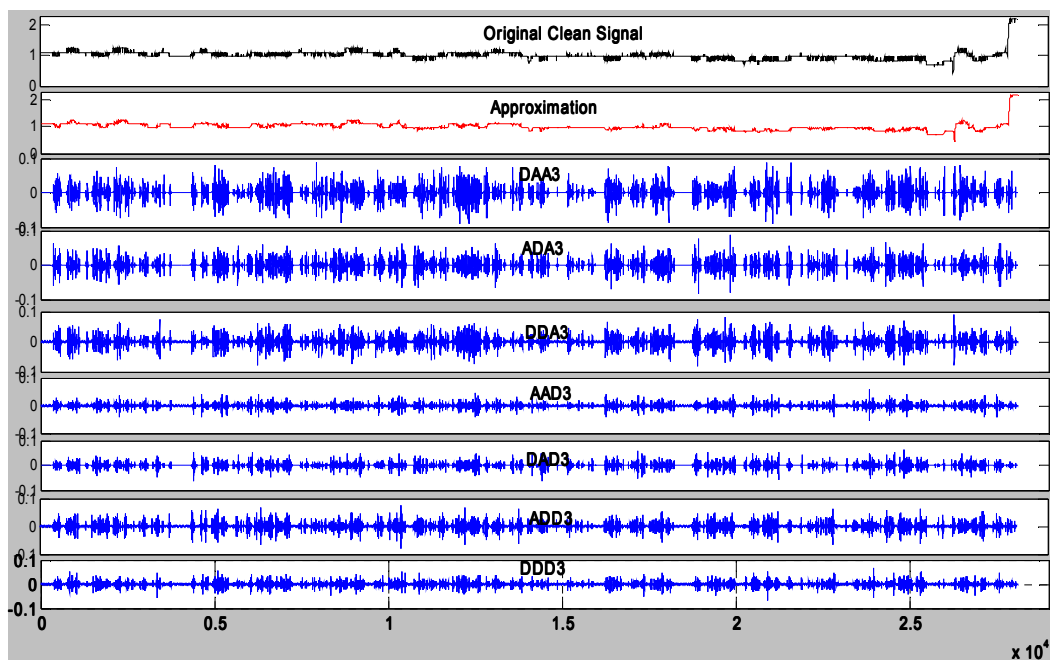


Figure 23: MRA of level-3 DWPT decomposition of Vibration sensor X1 signal

Thus, the level-2 decomposition gives three details nodes while the level-3 decomposition yields seven details nodes. That is for the level-2 DWPT decomposition,

the original signal is represented by four different wavelet components, while in the level-3 DWPT decomposition the original signal is represented by eight wavelet components.

Also, the energy contained at the approximation node of the level-3 decomposition (energy at node (3,0) is 99.924%) is smaller than that contained at the approximation node in the case of the level-2 decomposition (energy at node (2,0) is 99.945%).

In this step, each of selected sensors to be included in the study will go through this MRA process once the appropriate level of DWPT decomposition for MRA is set. Also, the level of decomposition needs to be chosen carefully, as a higher level of decomposition may not necessarily bring in a better understanding of the signal, whereas decomposing on too few levels may not yield the necessary level of detail.

4.2.5 Data Fusion using Probabilistic Principal Component Analysis

The step of data fusion is an important one, because if it is not done or not done right, the anomaly detection may not be accurate. In general, when analyzing data mining problems, where there are many different types of sensors used to extract usable and useful information, the appropriate data fusion technique is crucial. Indeed, not only does the fusion of multi-sensor data provide major advantages over individual source data, the use of multiple sensor sources may increase the accuracy with which a phenomenon can be observed [121].

Therefore, the success of this proposed methodology is dependent on the effective accomplishment of this step. The field of data fusion has become a very important and active area in Research I, so much so that the U.S. Department of Defense created a Joint Directors Laboratory subpanel to assemble a uniform set of technical definitions for use

among researchers. In fact, J. Llinas et al. argue that the lack of unified terminology has been one of the historical barriers to technology transfer in data fusion [122]. For example, data fusion is defined by the U.S. Department of Defense Joint Directors of Laboratories Data Fusion Subpanel as: “a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats and their significance” [123]. Therefore, J. Llinas et al. attempted and defined fusion as: “the integration of information from multiple sources to produce the most specific and comprehensive unified data about an entity”, which goes along the same line as the U.S. Department Joint Directors’ definition. The terms “data fusion” and “sensor fusion” are used interchangeably. The data fusion step remains a vital one if the goal is to make the best possible judgment. In fact, Yan and Goebel assert that most single sensors cannot be relied on to deliver accurate information all the time [57]. The goal of the data fusion step is to combine pieces of information from a multi-sensory data set system into fewer variables, which allows researchers to draw a more accurate conclusion than one could get from an individual sensor. Thus, the data fusion process produces knowledge that is not otherwise obtainable from a narrow wavelength-band sensor, or is more accurate than information gathered from single sensor systems [124]. The literature extensively discusses data fusion techniques; this can be seen in [122, 124]. Data fusion can be done at different stage of data analysis. Thus, in [125] Carvalho et al. propose a three-level data fusion framework based on both data and variables, where the data fusion can be classified as low level, high level, or mixture level. When the data fusion is performed before analysis, it is classified as low level. When the data fusion is performed after some data analysis, it is classified as high level variable fusion. In situations where the data and variables can be fused, it is defined as mixture level data fusion.

4.2.5.1 Overview data Fusion Techniques

There are many data fusion techniques in the literature that have been used for different applications. Thus, some data fusion techniques are more mathematically intensive than others. The reader interested in some mathematical background for data fusion can see [126, 127], among many other papers.

Ultimately, the key to any successful study where data fusion is needed is to find a technique that is suitable for one's problem.

The data fusion architecture can be divided into three main groups: detection, classification, and identification algorithms [124]. The main algorithms can either be physical models, feature-based inference techniques, or cognitive-based models. Each of these three models can cover a different application field, as summarized on Figure 24.

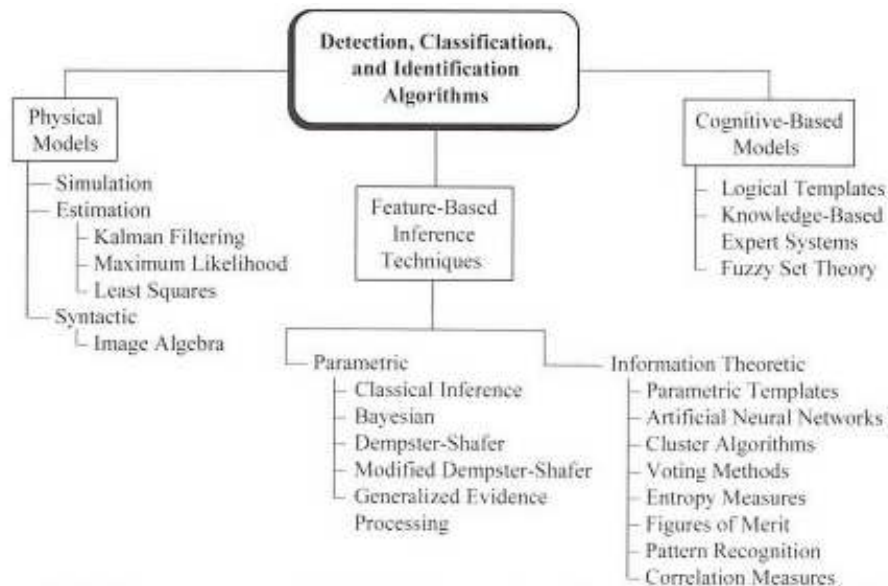


Figure 24: Data fusion for detection, classification, and identification algorithms [124]

Besides the three groups mentioned above, there are other mathematically-based techniques for data fusion, such as the random set theory, conditional algebra, and relational event algebra [127].

The following section is a brief literature review of some of the most commonly used data fusion techniques used to analyze multi-sensors systems, where the classical inference, Bayesian inference, and Dempster-Shafer inference are referred to as parametric techniques.

4.2.5.1.1 The classical inference

The classical inference is mainly used when only a small representative random sample of the population can be obtained to estimate the statistical characteristics of a large population. That is, the classical inference calculates the probability that an observation can be attributed to the presence of an object or event, given an assumed hypothesis. However, its applications are limited due to many disadvantages such as:

- The difficulty of obtaining the density function that describes the observable
- The complexity of the application of multivariate data the classical inference theory becomes complex,
- It can only assess two hypotheses at a given time
- It cannot utilize the knowledge of a priori and likelihood probabilities

Furthermore, another important limitation of the classical inference is that it cannot be used when the data are collected with bias of unknown size. Outliers have to be identified and removed. However, in the study of anomaly detections, the anomalies may well remain within the outliers.

4.2.5.1.2 The Bayesian inference fusion

The Bayesian inference comes from the Bayes theory, so it is a probability-based reasoning technique that relies on a priori knowledge about events. Ultimately, the Bayesian inference provides a method for calculating the conditional a posteriori probability of a hypothesis, given supporting evidence. Although the Bayesian inference has the ability to address some of the difficulties with classical inference (it updates for

example, the a priori given a previous likelihood estimate and additional observations, unlike other hypotheses), the Bayesian inference has its own set of disadvantages:

- It is difficult to define the prior probability and likelihood function
- The computation scheme becomes complex when multiple potential hypotheses and multiple conditionality-dependent events are evaluated
- There is a requirement that competing hypotheses be mutually exclusive
- The Bayesian inference cannot account for general uncertainty

Thus, the Bayesian inference can be applied only outside of its limiting factors.

4.2.5.1.3 The Dempster-Shafer evidential theory:

This theory is a probability-based data fusion; therefore, the most advantageous application is when the sensors or the information sources contributing information cannot associate a 100 percent probability of certainty to their output decisions [124]. The Dempster-Shafer generalizes Bayesian inference, and both methods produce identical results when all singleton propositions are mutually exclusive and there is no support assigned to uncertainty. The Dempster-Shafer method must define processes in each sensor that assigns the degree of support for a proposition. Despite this, the Dempster-Shafer ameliorates some of the limitation of the classical inference. Its main disadvantage is that it is unable to make direct use of prior probability when this probability is known. Because of that limitation, many techniques have been introduced to improve the Demspter-Shafer.

4.2.5.1.4 The Voting logic data fusion technique

This technique combines detection and classification from multiple sources (e.g. sensors) by treating each sensor's declaration as a vote, from which the majority, plurality, or decision-tree is used. Its approach to voting is based on Boolean algebra. Some of the main advantages of the voting logic are:

- It handles false alarms better than the other techniques

- It may be an appropriate data fusion technique to apply when a multiple sensor system is used to detect, classify, and track objects

However, voting logic has its share of disadvantages as well:

- It has a poor rejection of cluster-induced false alarms when sensors are assembled in parallel and their functions are independent of each other
- It has poor detection of suppressed targets when the sensors are in series and the system output is dependent on output from each sensor
- It can increase the complexity of signal processing when the system output is dependent on combinations of multiple sensor outputs

4.2.5.1.5 The fuzzy logic data fusion technique

The fuzzy logic data fusion technique is based on fuzzy logic [128]. Contrary to the other methods, the fuzzy logic technique offers the ability to represent analog processes in a digital framework. Hence, the fuzzy logic technique has the capacity to solve the problems of imprecise knowledge or indistinct boundary definitions of mathematical treatment. It permits uncertainties in knowledge or identity boundaries to be applied to diverse applications. However, the fuzzy logic data fusion technique also has some drawbacks:

- It may not be easily separated into discrete segments and may be difficult to model with conventional mathematics
- It has to be applied within boundary sets of values that have not been sharply defined.

The data fusion techniques presented in this overview are far from being comprehensive, as there are many combinations of the main techniques that can be used to meet researchers' goals. As a matter of fact, there are many other techniques being developed currently by different researchers throughout the world; for example, the AUG (Airborne

Underwater Geophysical) Signals has a toolbox containing algorithms with the following data fusion techniques [129]:

- Statistical methods
- Markov Random Fields
- Dempster-Shafer Theory
- Neural Networks
- Fuzzy Logic
- Wavelets
- Super-resolution
- AUG Signals' Super-resolution

In fact, the data fusion techniques can be as specific as one's specific field of interest. For example, the passive data associated with unambiguous location of targets, which is very specific to military applications, is shown in Table 6 below.

Table 6: Fusion techniques for associating passively acquired data to locate and track targets [124]

Fusion Level	Data Fusion Technique	Advantages	Disadvantages
Received Signal (pixel-level fusion)	Coherent processing of data received from two types of sensors: a scanning surveillance radar and a passive receiver with high-directivity multibeam antenna	All available sensor information used	Large bandwidth communications channel required
		Unambiguous target location obtained	Auxiliary sensor required
		data are processed in real time	One coherent processor for each beam in the multibeam antenna required
Angle Data (feature-level fusion)	Maximum likelihood or relaxation algorithm using direction angle measurements to the target	3D position of target obtained	Ghosts created that have to be removed through increased data processing
		Communication channel bandwidth reduced	
Target Track (decision-level or sensor-level fusion)	Combining of distributed target tracks obtained from each surveillance radar	3D position of target obtained	Many tracks must be created stored, and compared to eliminate false tracks
		Communication channel bandwidth reduced even further	

Despite the panoply of the data fusion techniques, each technique has advantages as well as drawbacks. That is, some of them have very cumbersome implementation, when their application is not limited by restrictions. Since the main goal of this thesis is to have mathematical computations that can be done fairly quickly, to ensure the quick detection of anomalies during the monitoring of a system, the chosen data fusion technique will be one that is simple and transparent to implement. The number of sensors to monitored may be large, so it is important to be able to easily exclude or add given sensors based on their contribution. Thus, the Probabilistic Principal Component Analysis, which is an extension of the conventional Principal Component Analysis, altered by the incorporation of the fact that sensor measurements have uncertainties, will be used in this work.

4.2.5.2 Overview of Principal Component Analysis

The Principal Component Analysis (PCA) is based on a mathematical process whose main objective is to reduce the dimensionality of a data set that has a large number

of potentially correlated (or interrelated) variables by condensing these variables into a smaller number of uncorrelated principal components. In this process, the first principal component accounts for as much of the variability of the data set as possible, and each subsequent component accounts for as much of the remaining variability as possible. The PCA achieves this goal by identifying patterns in the data set and expressing the data in a way that highlights their resemblances and differences [130]. Thus, the main attractiveness of the PCA lies in the fact that once the patterns in the data are found, the data can be compressed. The PCA is also called the discrete Karhunen-Loève expansion (or transform), the Hotelling transform or Proper orthogonal decomposition depending on the area of application. There is quite a large amount of literature on PCA and its application. In [131], I. Jolliffe explains the underlying concept of the PCA technique and its application to time series data. The popularity of the PCA as an information fusion technique may be explained in part by the fact it is a fairly simple technique that is based on simple linear algebra. Moreover, it is a non-parametric method for extracting relevant information from confusing data sets [132]. That is, it allows the identification of patterns in data, and then expresses the data in a way that highlights their similarities and differences. Once the patterns in the data are found, the data can be compressed (the number of dimensions can be reduced) while conserving much of the original information. It is easier to infer characteristics from the reduced data set than from the original.

Some of the tutorials of PCA [130, 132] are among the simplest ones to follow, where the PCA is presented as follows:

- Step 1: Get data composed of multiple variables.
- Step 2: Calculate the mean of each variable and subtract it from the entry element of the variable (Row data Adjust)
- Step 3: Calculate the covariance matrix, which is an n -by- n matrix (for n -dimensional data set)

- Step 4: Calculate the eigenvalues and eigenvectors of the covariance matrix, where the eigenvectors represent the axes of the principal components. Also, the eigenvectors are perpendicular to each other (i.e. forming a basis).
- Step 5: Choose the components and form a feature vector, which in turn decides the number of eigenvectors (k , with $n \geq k$) to be kept. Therefore, the number of disregarded eigenvectors represents the number of eliminated variables

Feature Vector = (eig1 eig2 ... eigk)

- Step 6: Derive the new data set. The new data set is generated using simple linear algebra:

Final data = Row Feature Vector * Row Data Adjust

When engineering practical problems, programming software is utilized to write algorithms to perform the PCA.

Although data set dimensionality reduction is the principal application of the PCA, it has many other applications, including exploratory data analysis (e.g. detection of outliers), data compression and data reconstruction, image processing, visualization, pattern recognition, time series prediction, and data feature extraction. In fact, Table 7 shows a summary of PCA variants used as feature extraction and projection techniques.

Table 7: Feature Extraction and Projection Methods [133]

Method	Property	Comments
Principal Components Analysis (PCA)	Linear map; fast; eigenvector-based	Traditional, eigenvector based method, also known as Karhunen-Loeve expansion; good for Gaussian data
Linear Discriminant Analysis	Supervised linear map; fast, eigenvector-based	Better than PCA for classification; limited to (c-1) components with non-zero eigenvalues
Projection Pursuit	Linear map; iterative; non-Gaussian	Mainly used for interactive exploratory data-analysis
Independent Component Analysis (ICA)	Linear map; iterative; non-Gaussian	Blind source separation, used for de-mixing non-Gaussian distributed sources (features)
Kernel PCA	Nonlinear map; eigenvector-based	PCA-based method, using a kernel to replace inner products of pattern vectors
PCA Network	Linear map; iterative	Auto-associative neural network with linear transfer functions and just one hidden layer
Nonlinear PCA	Linear map; non-Gaussian criterion; usually iterative	Neural network approach, possibly used for ICA
Nonlinear auto-associative network	Nonlinear map; non-Gaussian criterion, iterative	Bottleneck network with several hidden layers; the nonlinear map is optimized by a nonlinear
Multidimensional scaling (MDS), and Sammon's projection	Nonlinear map; iterative	Often poor generation; sample size limited; noise sensitive; mainly used for 2-dimensional visualization
Self-Organization Map (SOM)	Nonlinear, iterative	Based on a grid of neurons in the feature space; suitable for extracting spaces of low dimensionality

The choice of one technique over the others in Table 7 is primarily based on the goal and speed of execution of each method.

Recall that the fundamental goal of the data fusion step is to combine pieces of information from a potentially correlated multi-sensory data set into fewer uncorrelated variables that allow for drawing a more adequate conclusion than one could get from each individual sensor. Therefore, based on the simplicity of the PCA on one hand, and its ability to reduce the dimensionality of a data set on the other, the PCA is a good choice for data fusion. Despite all the qualities the PCA mentioned above, it has some limitations that must be improved upon. In fact, the following are some of the main limitations of the PCA [134]:

- It assumes approximate normality of the input space distribution, which implies that the PCA may still be able to produce a good low dimensional data set even though the data may not be normally distributed.

- It can fail for data that lies on a complicated manifold.
- Its makes the assumption that the input data is real and continuous.
- It is non-parametric (i.e. no probabilistic model for observed data).
- It doesn't handle uncertainty in sensor data.

Although the PCA performs fairly well with many types of data set, because of its last stated limitation, a technique that can handle uncertainty must be explored. The fact that the PCA cannot handle data uncertainty because it doesn't take into account the probabilistic model for the observed data [135, 136] cannot be overlooked in this thesis. It is a necessity for data uncertainty to be taken into account in the study, since power plants monitoring sensor measurements would most likely have uncertainties due to many intrinsic phenomena of power plant operations such as different operating regimes (a wide range of sensor readings at the different load conditions), operating environment (summer vs. winter), sensor built-in uncertainty, machine-to-machine variation, etc. Thus, the Probabilistic Principal Component Analysis (PPCA), which was introduced in order to deal with sensor data uncertainty, emerged as an improvement over the standard PCA, as it has the advantage of taking into account data uncertainty [136].

4.2.5.3 Probabilistic Principal Component Analysis overview

The primary goal of the PPCA in the dimensional reduction application is to make up for the absence of a probabilistic model in the PCA methods. To achieve that objective, the PPCA utilizes the notion of maximum likelihood and the variance of the reduced data matrix, which means that it uses only the most significant weights obtained from the conventional PCA. The PPCA is definitely an improved version of the standard PCA, as it has the advantage of taking into account data uncertainty [136]. Some of the advantages of the PPCA can be summarized as follows:

- It addresses limitations of the regular PCA.

- It can be used as a general Gaussian density model in addition to reducing dimensions.
- The maximum-likelihood estimates can be computed for elements associated with principal components.
- It captures dominant correlations with few parameters.
- It permits multiple PCA models to be combined as a probabilistic mixture.
- It can be used as a base for a Bayesian PCA.

Because of all these capabilities, the PPCA is a good candidate for many applications in practice as reported by M. E. Tipping et al. in [136]. Some possible applications are:

- The PPCA can be used to estimate the principal axes for missing data by making use of the standard methodology of the maximum likelihood of a Gaussian model as presented in the very popular book by R. J. Little and D. B. Rubin [137].
- Through the use an example in [138], the author demonstrates a successful use of the standard PCA to handle missing data.
- The PPCA can be used through the mixtures of PPCA models. Since the standard PCA consists of a single linear projection, more complex models that combine many PCA models have been used in practice in fields such as image compression [139] and visualization [140].
- The PPCA can be used as a constrained Gaussian density model, with the benefit that the maximum likelihood estimates for the parameters associated with the covariance matrix can be efficiently computed from the principal data components.

Nevertheless, in this work the PPCA will be used for its dimensionality reduction ability, especially when the speed of the execution of the analysis is the foremost important factor, because the early detection of an anomaly can avoid potential catastrophic failure down the road. Therefore, a reduced number of variables will lead to decrease of the computational burden.

The underlying mathematical background of PPCA is based on a few considerations, which will be briefly presented in the next section. The interested reader can also access a more detailed approach in [135, 136].

4.2.5.4 Mathematical Background of PPCA (with all the equations)

The very first step of the PPCA consists of performing the computation of the conventional PCA on a given data set. That is, the PCA involves a matrix analysis technique called eigenvalues decomposition. The PPCA mathematically derives from the notion of the Gaussian latent variable model, which in turn is related to statistical factor analysis [135].

4.2.5.4.1 *Factor analysis*

Factor analysis is a mathematical approach used to reduce the number of observed vectors t (d -dimensional) to a corresponding number of vectors (q -dimensional) of latent (unobserved) variable x , where $q < d$ is defined as [136, 141]:

$$t = Wx + \mu + \varepsilon \quad (15)$$

Where $y(x, W) = Wx + \mu$ is a function of the latent variable x with parameters W .

W is the $d \times q$ matrix

The parameter vector μ allows the model to have a non-zero mean.

By definition $x \sim N(0, I)$, with the restriction that the latent variables be independent and Gaussian with unit variance. The noise or error $\varepsilon \sim N(0, \psi)$ is also Gaussian distribution with zero mean, and a variance of ψ .

Consequently, the corresponding observation $t \sim N(\mu, WW^T + \psi)$ is Gaussian as well, and the model can be determined by the maximum likelihood. M.E. Tipping et al. [136] argue that “the distinction made between variance and covariance in the standard factor analysis model, the subspace defined by the maximum likelihood estimates of the

columns of W will generally not correspond to the principal subspace of the observed data.” Furthermore, they establish that the isotropic noise covariance and the variance be equal ($\sigma^2 I = \psi$) with the residual variance σ^2 assumed known. Then, they show that the observation covariance model $W^T W + \sigma^2 I$ can be calculated exactly, allowing W and σ^2 to be determined analytically.

The latent variable x is conventionally defined as independently distributed Gaussian with zero mean and unit variance ($x \sim N(0, I)$). Then, the observable t can be written in Gaussian distribution form as follows from the equation

$$(16)$$

$$t | x \sim N(Wx + \mu, \psi) \tag{16}$$

Where $\psi = \Lambda + \sigma^2 I$, which is the combination of the measurement (or prediction) error Λ that is unique to the response, and the variability σ^2 that is unique to x (i.e. the isotropic noise covariance).

Λ is a diagonal matrix with each diagonal element representing the data uncertainty of the corresponding variable. Also, Λ is the variance of the additional error which has zero mean ($\varepsilon^* \sim N(0, \Lambda)$).

In [135], Jiang et al. stress that the latent variable x in the PPCA is intended to explained the correlations between observed variable t , while the error variable ε represents the variability unique to x , which is different from the standard PCA, which treats covariance and variance identically.

The unadjusted conditional distribution of the latent variable x can be calculated given observation t using the Bayes’ rule. It is also Gaussian:

$$x | t \sim N(M^{-1} W^T t, \sigma^2 M^{-1}) \tag{17}$$

Where $M = W^T W + \sigma^2 I$

M.E. Tipping et al. show that the likelihood is maximized when the weight matrix is:

$$W_{ML} = U_q (\Lambda_q - \sigma^2 I)^{1/2} R \quad (18)$$

U_q is the matrix of the principal eigenvectors with λ_j the corresponding eigenvalues, and

R is an arbitrary $q \times q$ orthogonal rotation matrix; $R = I$ can be assumed in practice.

Jiang et al suggest calculating the q -dimensional data matrix as follows:

$$\Sigma_{ML}^{-1} = I + W_{ML}^T \Psi_{ML}^{-1} W_{ML} \quad (19)$$

Where $\Psi_{ML} = \Lambda + \sigma_{ML}^2 I$, and

$$\sigma_{ML}^2 = \frac{1}{d - q} \sum_{j=q+1}^d \lambda_j \text{ is the maximum likelihood estimate of } \sigma^2$$

Thus, the unadjusted data in the lower dimensional latent space is calculated as:

$$\langle x_n | t_n \rangle = M_{ML}^{-1} W_{ML}^T t_n \quad (20)$$

Where $M_{ML} = W_{ML}^T W_{ML} + \sigma_{ML}^2 I$. Let us rename $\langle x_n | t_n \rangle$ as Φ^*

Thus, the parameters calculated for the PPCA will be applied in the model.

The parameters $\langle x_n | t_n \rangle$ or Φ^* , called the data matrix, represent the entry of the PPCA's principal components.

The data fusion step involves fusing the signal features of each variable obtained from the MRA using the PPCA technique, which is an improvement over the standard principal component analysis (PCA). Thus, the probabilistic principal component analysis (PPCA) is used to merge the information from the sensors of interest. To perform the PPCA, the steps of the principal components analysis are executed, and then the notion of maximum likelihood and the variance of the reduced data are calculated using matrices, where only the most significant weights obtained from the standard PCA are used as entries in a maximum likelihood matrix. The next section presents an application of the PPCA data fusion technique to an illustrative example of sensor data set.

4.2.5.5 Illustration of Data Fusion PPCA (with all figures & tables)

The original data set of twelve sensors (x_1, \dots, x_{12}) has been normalized, de-noised and the MRA step has been completed.

1) First, the conventional PCA steps are done:

The conventional PCA is performed on the retained signal feature (approximation) to determine the principal components (PC).

The covariance matrix of the processed 12 variables is calculated, as shown in Table 8

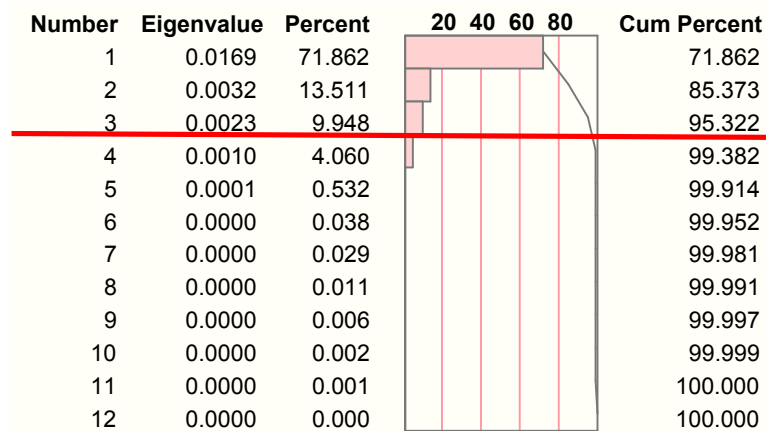
Table 8: Covariance Matrix for eigenvalues calculation

	DWATT	TNH	BB1	BB2	BB4	CPD	CTD	CPR	CSGV	CTIM	AFQ	AFPAP
DWATT	0.00039	0.00000	0.00094	0.00064	0.00107	0.00031	-0.00011	0.00027	0.00000	-0.00103	0.00027	0.00002
TNH	0.00000	0.00000	0.00001	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000
BB1	0.00094	0.00001	0.00598	0.00411	0.00487	0.00080	-0.00022	0.00073	-0.00000	-0.00234	0.00065	0.00001
BB2	0.00064	0.00000	0.00411	0.00450	0.00355	0.00051	-0.00012	0.00049	-0.00000	-0.00146	0.00043	-0.00001
BB4	0.00107	0.00001	0.00487	0.00355	0.00876	0.00089	-0.00030	0.00089	-0.00000	-0.00288	0.00069	-0.00007
CPD	0.00031	0.00000	0.00080	0.00051	0.00089	0.00027	-0.00008	0.00023	-0.00000	-0.00080	0.00022	0.00003
CTD	-0.00011	0.00000	-0.00022	-0.00012	-0.00030	-0.00008	0.00004	-0.00008	0.00000	0.00035	-0.00007	-0.00000
CPR	0.00027	0.00000	0.00073	0.00049	0.00089	0.00023	-0.00008	0.00020	-0.00000	-0.00073	0.00019	0.00001
CSGV	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000
CTIM	-0.00103	0.00000	-0.00234	-0.00146	-0.00288	-0.00080	0.00035	-0.00073	0.00000	0.00315	-0.00069	-0.00002
AFQ	0.00027	0.00000	0.00065	0.00043	0.00069	0.00022	-0.00007	0.00019	-0.00000	-0.00069	0.00019	0.00002
AFPAP	0.00002	-0.00000	0.00001	-0.00001	-0.00007	0.00003	-0.00000	0.00001	0.00000	-0.00002	0.00002	0.00001

Then, the eigenvalues, their respective contributions, and the corresponding eigenvectors for the covariance matrix are obtained and ranked from the highest value to the lowest value, as illustrated on

Table 9. Thus, the highest value of the eigenvalues λ_1 represents 71.862% of the 12-variable combination signal; the second highest λ_2 represents 13.511%; and so on.

Table 9: Pareto chart of eigenvalues contributions



The original twelve eigenvectors and their corresponding eigenvalues are shown in Table 10. However, to retain at least 95% of the information contents (i.e. 95% confidence level), only the first three eigenvalues, representing 95.32% of the information content, are needed. Consequently, the eigenvectors corresponding to those three eigenvalues, which represent the principal components, are chosen as shown in the red rectangle in Table 10.

Table 10: Original twelve eigenvectors

0.11163	-0.04960	-0.24283	-0.05772	0.41572	0.78553	0.04222	-0.32684	0.15429	-0.00937	0.00485	-0.01971
0.00063	0.00001	0.00212	0.00228	0.00651	0.05804	0.04774	0.16821	0.09238	0.97844	0.00953	-0.00521
0.53611	0.44731	-0.07759	0.70935	-0.05629	0.01080	-0.00300	0.00006	-0.00331	-0.00165	0.00024	0.00005
0.41406	0.59490	0.14483	-0.67302	-0.01381	-0.02075	-0.00714	-0.00504	0.00213	0.00332	0.00037	0.00004
0.64797	-0.61745	0.44181	-0.05191	0.00375	0.00264	-0.03088	0.00364	0.00118	-0.00066	0.00022	0.00010
0.09165	-0.03798	-0.18695	-0.01204	0.56166	-0.49463	0.03764	-0.18124	0.25345	0.03669	-0.54319	0.00569
-0.02942	0.03416	0.08944	0.02006	0.10971	0.13193	0.08150	0.69772	0.65425	-0.19476	0.01949	-0.01760
0.08733	-0.04653	-0.14856	-0.02495	0.27663	-0.23984	0.69296	-0.00752	-0.07239	-0.01878	0.58946	-0.01287
-0.00002	0.00007	0.00010	0.00003	0.00082	0.01022	0.00364	-0.01495	0.04026	0.00304	0.02618	0.99867
-0.29391	0.23822	0.78966	0.18877	0.39979	0.07501	0.08759	-0.10941	-0.11084	0.01599	-0.00519	0.00140
0.07437	-0.02580	-0.17164	-0.02906	0.43791	0.07016	-0.22926	0.55784	-0.63215	-0.03118	-0.03573	0.03464
-0.00159	0.01089	-0.02809	0.01353	0.26104	-0.22088	-0.66815	-0.14749	0.23790	0.04096	0.59582	-0.02305

Thus, the original twelve sensor data sets can be reduced into three principal components.

2) Steps for PPCA:

Once the PCA steps are completed, the PPCA will add the probabilistic component.

The entries of the eigenvectors are the weights that will be applied to their corresponding variables in order to obtain the principal component that is the combination of the twelve variables. At this point, the PPCA procedure will be implemented. The major weight contributors are identified in order to calculate the maximum weight that will be allowed as an entry value of an eigenvector; let's say less than 0.1 is set to 0, while any value greater than 0.1 is kept. This is to say that the maximum likelihood depends on the meaningful weights. Consequently, for the PPCA, the following principal components are obtained:

Table 11: Maximum likelihood weight for principal components

Variables	PC1	PC2	PC3
DWATT	0.11163	0	-0.24283
TNH	0	0	0
BB1	0.53611	0.44731	0
BB2	0.41406	0.5949	0.14483
BB4	0.64797	-0.61745	0.44181
CPD	0.09165	0	-0.18695
CTD	0	0	0
CPR	0	0	-0.14856
CSGV	0	0	0
CTIM	-0.29391	0.23822	0.78966
AFQ	0	0	-0.17164
AFPAP	0	0	0

After obtaining the maximum likelihood weight matrix and calculating the remaining parameters for the PPCA using a developed MATLAB® code, the result of the three principal components kept using the PPCA steps are shown in Figure 25a, b, and c representing the first, second and third principal components respectively.

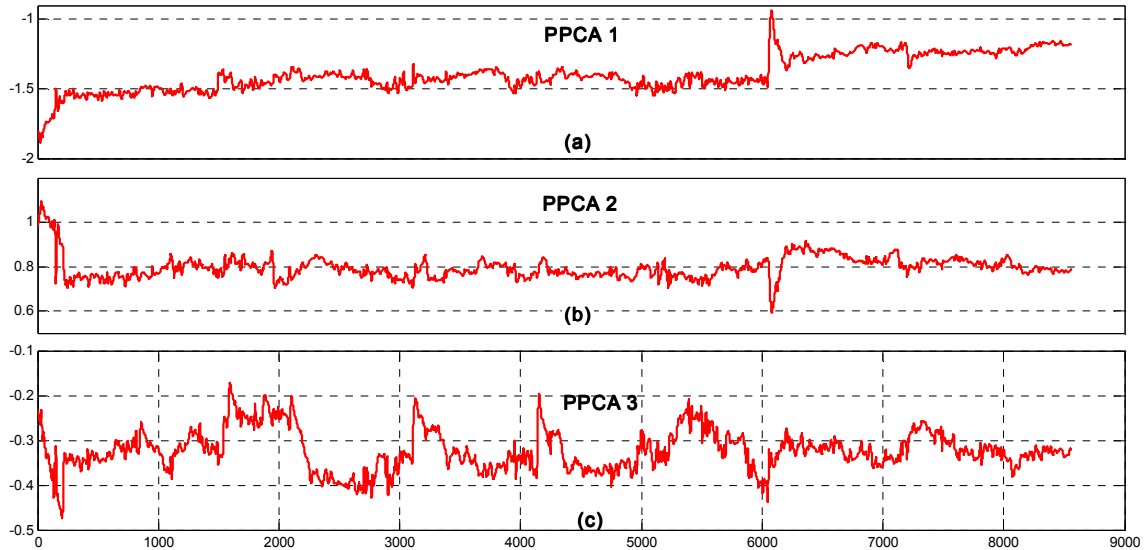


Figure 25: PPCA result principal components

The result of the PPCA data fusion is that the study will go from twelve potentially correlated sensor values (the original number of sensors of interest) to a much more manageable three uncorrelated variables, which are now called the principal components. In general, there exist rules to determine the number of components to be kept. Usually this is dictated by the level of confidence that needs to be achieved in the study; that is, the ratio of the original information to be kept. In this example, a confidence level of 95% was used.

4.2.6 Anomaly detection decision

The last step of the proposed methodology for early precursor detection is to make the correct assessment of the anomaly identification in the appropriate time series. Anomaly detection has been a fascinating and active field of research in the research community for many years. A. Lazarevic et al. provide the following definition:

“anomaly is a pattern in the data that does not conform to the expected behavior.” Also, anomalies can be considered outliers in some cases [142]. Thus, based on the anomaly definition, it is understandable why the anomaly detection is of such interest to so many fields of research. In fact, almost every complex systems developer or operator allocates resources to anomaly detection in order to avoid the possibility that a missed anomaly lead to a catastrophic systems failure. As a consequence of the interest in anomaly detection, there exists plenty of literature [143, 144] on the subject. In [143], V. Chandola et al. present a fairly broad survey of the current anomaly detection decision approaches. Among some of the most common in the engineering field are the accuracy-based metrics, precision-based metrics, and robustness-based metrics [145]. Some of the current active areas of anomaly detection research are engineering, statistical, economic, health care, etc. There are many fault decision metrics; some of them are summarized in [145]. In general, anomaly detection methods are faced with a set of challenges that must be overcome in order to achieve any meaningful type of results. Those challenges may include but are not limited to [142]:

- Defining a representative normal region can be difficult.
- The boundary between normal and outlying behaviour can be imprecise.
- The exact notion of an outlier is different for different application domains.
- The availability of labelled data for training/validation can be limited.
- One might have malicious adversaries.
- Data might contain noise.
- Normal behaviour keeps evolving.

There are three major types of anomaly [142]:

- Point anomalies, which occur when individual data points are anomalous with the data set.
- Contextual anomalies, which occur when an individual data instance is anomalous within a context; the context needs to be understood.

- Collective Anomalies, which occur when a collection of related data points is anomalous. The relationship among data points must be known. The individual points within a collective anomaly may not be anomalous by themselves.

For the purpose of this study, which relies on a data-driven approach, a precision-based metrics type [21, 145] will be utilized in the proposed process to make the decision of whether a found outlier is an anomaly or not. The advantages of using a precision-based metric are that it is simple to implement within the context of a statistical approach based on historical knowledge of given systems, and it is suitable for time series types of study. Furthermore, a statistically-based technique for anomaly (i.e. outlier) detection will first be utilized, followed by the precision based metric for the decision.

Typically the statistically-based technique is a very suitable anomaly detection technique for time series. It works by modeling data points using stochastic distribution, where the points are determined to be outliers depending on their relationship with an established threshold model. Although statistically-based techniques have many advantages, they have some drawbacks that one must be mindful of when using these techniques. Statistically-based techniques have the following characteristics:

- Advantage: they utilize existing statistical modelling techniques to model various types of distributions.
- Drawbacks: it is difficult to estimate distributions with high dimensions, and their underlining parametric assumptions often do not hold for real data sets.

4.2.6.1 Statistically-based approach to anomaly detection

There are other practical advantages of using a statistically-based technique in this thesis. For example, a statistically-based method does not require sophisticated modeling of the data set, such as a neurally-based network would require. The calculations and

recalculations of statistical parameters needed to set the threshold are fairly simple and fast, which is very important for implementation in systems monitoring. In fact, any time there is an exterior intervention to the system (such as water watch for a heavy gas turbine, installation of a new part, completion of scheduled maintenance, or completion of a recovery of degradation action); the statistical parameters will need to be recalculated.

4.2.6.1.1 Reconstruction of representative signal

The procedure for anomaly detection decision is a multi-steps approach. After the completion of the PPCA step, when the different principal components of the signal have been determined, the reduced dimensionality data set is in turn combined into a one-dimensional signal calculated as follows:

$$RS = \sum_{i=1}^n \lambda_k \Phi_k^*(i) \quad (21)$$

Where: n: is the number of retained principal components (PC)

λ_k : is the contribution of Eigenvalues (k)

$\Phi_k^*(i)$: is the signal corresponding to PC(k) of the data matrix

RS: is the reconstructed signal

That is, the original N-dimensional data set was first reduced to a q-dimensional data set using the PPCA data fusion technique, and it is now reduced to a one-dimensional reconstructed signal RS. The eigenvalues corresponding to each principal component are used as weight to consolidate the q-dimensional system into a one-dimensional system.

4.2.6.1.2 Statistical Process Control for threshold setting

The statistical process control, which relies on quality control theories [146], is used to set up the threshold level through a calculation as follows:

Step1:

The obtained reconstructed signal RS is decomposed using the discrete Wavelet packet decomposition (similar to the MRA step). Then, the energy content of each node is calculated (similar to the multi-resolution analysis step) using the equation to calculate the energy content as follows at each time step t:

$$E_{j,n}(t) = \int_{-\infty}^{\infty} (W_{j,n,k}(dt))^2 dt = \sum_k W_{j,n,k}^2(t)$$

Step 2:

Once a system representative signal RS is calculated, the two damage indicators SAD and SSD are calculated to be monitored instead of directly monitoring the change of the energy content [147]. SAD and SSD are defined as follows:

Sum of Absolute Difference (SAD) and computed as:

$$SAD(k) = |E(k) - E_{ref}| \quad (22)$$

Square Sum of Difference (SSD) and computed as:

$$SSD(k) = (E(k) - E_{ref})^2 \quad (23)$$

Where:

E_{ref} is the reference signal energy content, calculated over a healthy period before the monitoring for potential anomaly starts.

$E(k)$: is the RS energy content obtained from the subsequent measurement.

The monitoring of SAD and SSD is a popular approach in practice because these sums permit the easy capture of deviations from target monitoring bands. In fact, an important part of accomplishing any good anomaly detection method is to figure out the best feature to be tracked; thus, SSD is constantly recommended as one such trackable feature [148].

Since the selection of the right feature is key to successful monitoring in general, there are many algorithms developed to track signal or image features [148, 149]. Thus, the two damage indicator parameters SAD and SSD chosen to be track fit that requirement.

It is important to note that the calculations of the damage indicators SAD and SSD have been modified from the ones proposed [44] by Sun, because in the case of a larger size data set, the required computation time becomes a limit to the monitoring. Thus, Sun

proposes:
$$SAD = \sum_{i=1}^m |E(k) - E_{ref}| \quad (24)$$

and

$$SSD = \sum_{i=1}^m (E(k) - E_{ref})^2 \quad (25)$$

Step 3: Conventional Apply SPC (Statistical Process Control)

The X-bar control chart concept [150] is used to establish the threshold of damage indication. Thus Ang et al. suggest the following threshold calculation for a one-sided upper $(1-\alpha)$ upper confidence limit for the damage indicator SAD (a similar formula is calculated for SSD) [151]:

$$UL_{SAD}^{\alpha} = \mu_{SAD} + Z_{\alpha} \left(\frac{\sigma_{SAD}}{\sqrt{q}} \right) \quad (26)$$

Where:

UL_{SAD}^{α} : Upper Confidence Limit

μ_{SAD} : is the value toward which the mean value of the parameter SAD converges

Z_{α} : is the value of standard normal distribution with zero mean and unit variance, so that the cumulative probability is $100*(1-\alpha)$

σ_{SAD} : is the value toward which the standard deviation of the parameter SAD converges

q: interval of monitoring time

Then, the X-bar control chart upper limit is used to monitor the damage indicators over a given period of time.

4.2.6.1.3 Modified threshold calculation

In general, the control charts are effective in defect prevention [150] when used in the context of manufacturing, for example. However, because of the constraint of the execution speed, the choice has been made in this work to modify the conventional SAD and SSD formulas. Therefore, instead of directly using the conventional statistical process control where the mean value of the SAD (μ_{SAD}) is used to calculate the upper specific limit, it is proposed that one uses a modified version of the SPC formula. The reasoning behind the proposed change is simple; because the conventional SPC is a sum of absolute values, it converges to a maximum value of the mean value of the SAD ($\mu_{SAD} = \mu_{SAD_MAX}$). Thus, it is proposed that one uses the same thought process applied to the maximum mean value to calculate the upper specific limit, which will be used as the threshold value, because before the mean values of SAD (and SSD) converge toward a given value, there is an overshoot in the value of SAD before it stabilizes. Therefore, for the determination of the upper specification limit, the original formula must be modified to address the modifications proposed in the calculation of the damage indicators, SAD and SSD. By doing so, the number of false-positive alarms, which would have been arbitrarily high, decreases. That appears to reduce considerably the number of false alarms. Therefore, the modified upper confidence limit for the damage indicator SAD is defined as:

$$UL_{SAD_MX}^{\alpha} = \mu_{SAD_MX} + Z_{\alpha} \left(\frac{\sigma_{SAD}}{\sqrt{q}} \right) \quad (27)$$

In sum, the proposed approach is one step better than the current industry standard practice because it has a better execution (i.e. run faster) time.

The proposed threshold calculation is also machine specific, because different machines in a given fleet can have many differences, such as machine-to-machine variations, operating condition variations, and environmental condition variations. Therefore, the threshold is established from the test machine during its healthy operation time and conditions. This allows for the avoidance of some of the practical issues of fleet management.

4.2.6.1.4 Bayesian Hypothesis for monitoring time

Finally, the Bayesian evaluation method is applied to the modified threshold value $\varepsilon = UL_{SAD_MX}^\alpha$. Thus, the Bayesian evaluation method for hypothesis testing can be conducted. The anomaly function is defined as $H(t)$, which is the vector of the Bayesian hypothesis testing result with null and alternative hypotheses defined as follows.

Once the threshold is set as a positive value, the decision criterion is based on a Bayesian binary type of hypothesis assessment [34] at each time step:

- The null hypothesis H_0 is defined as:

$$SAD(t) \leq \varepsilon, H(t) = 1 \quad (28)$$

That is, at any monitoring time step, if the sum of the absolute difference SAD (or SSD) is less than or equal to the calculated threshold, then the measurement at that point is considered good (i.e. the point is within the preset acceptable band).

- The alternative hypothesis H_a is defined as:

$$SAD(t) > \varepsilon, H(t) = 0 \quad (29)$$

That is, at any monitoring time step, if the sum of the absolute difference SAD (or SSD) is greater than the calculated threshold (i.e. outside of the pre-set band), then the measurement is rejected as a defect.

Hence, at each time step a defect judgment is made, so that all the defective data points are identified.

As a result of the hypothesis testing, a binary type of function is obtained as outcome over a giving period of monitoring time. Thus, the function $H(t)$ has values of 1 or 0 and can be plotted over time for visualization. An $H(t)$ value of 1 is a healthy state, and an $H(t)$ value of 0 is an abnormal one. Therefore, the appearance of the value of $H(t) = 0$ can be considered a failure precursor. The function $H(t)$ can be visualized and tracked over time because it is also a time series function.

Thus, the steps of the anomaly detection decision can be summarized and organized as shown on Figure 26

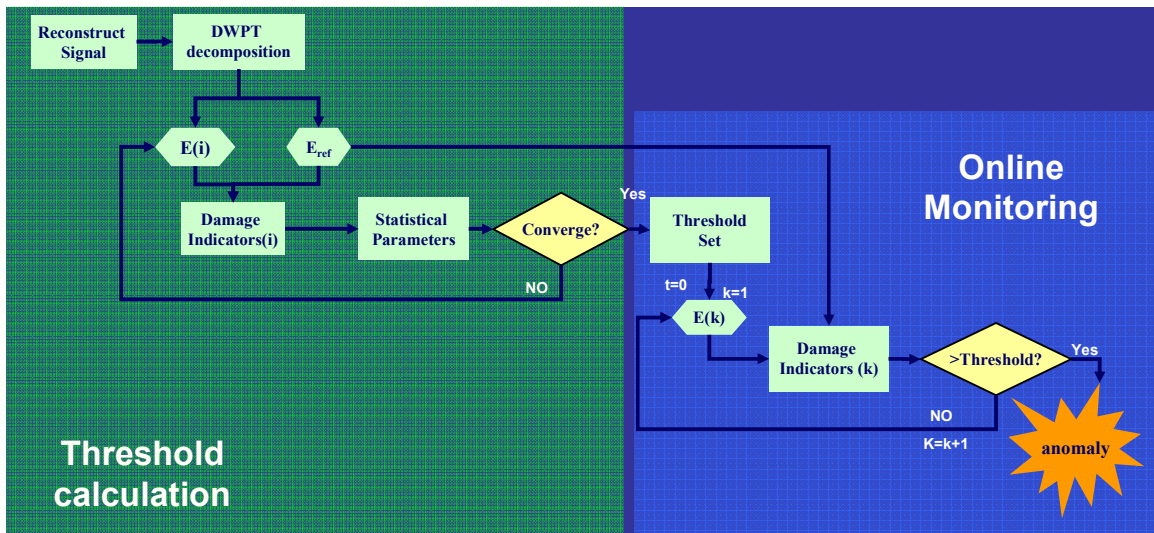


Figure 26: Anomaly Detection Decision Flowchart

4.2.6.2 Different error types (or confidence level) calculations

Any decision making process needs to provide the confidence level associated with its conclusions. Thus, an immediate advantage of the statistically-based techniques is that the confidence levels are easily determined. Moreover, the proposed approach provides an easy, fast, and robust calculation of the different types of errors.

4.2.6.2.1 *Calculation of the Confidence Level for type I error*

Recall that the probability of type I error or false-positive is defined as:

$\alpha = P\{\text{reject } H_0 | H_0 \text{ is true}\}$; that is, the probability of rejecting H_0 while is indeed true.

In other word, type I error is the probability of detecting a failure precursor while there is no defect.

In the proposed process, the statistical confidence level is an input from the system monitor, calculated as $100*(1-\alpha)\%$. In other words, the system monitor decides the level of type I error acceptable to him/her; then, that value is input in the threshold calculation. Consequently, the type I error is a controlled parameter in the proposed process.

4.2.6.2.2 *Type II errors calculation*

Conversely, the type II error or false-Negative is defined as:

$\beta = P\{\text{fail to reject } H_0 | H_0 \text{ is false}\}$; that is, the probability of not rejecting H_0 , while H_0 is false. In other words, the type II error is the probability of missing a defect when one is present. In fact, determining the type II error is crucial in systems where failures have huge consequences.

To estimate the type II error, let us assume the following conceptual approach to determine the probability type II error.

Let null hypothesis H_0 and the alternative hypothesis H_1 be defined as:

$$H_0: \mu = \mu_0 \tag{30}$$

$$H_1: \mu \neq \mu_0 \tag{31}$$

The procedure for testing is to take a random sample of n observations on the random variable x ; the test statistic is then calculated as:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \text{ and reject } H_0 \text{ if } |Z_0| > Z_{\alpha/2}$$

Where σ : is the standard deviation and $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

By recalling the central limit theorem, $\bar{x} \sim N(\mu, \sigma^2/n)$, we can see that for cases $H_0: \mu = \mu_0$ is true $Z_0 \sim N(0,1)$ and the probability of the successful H_0 , $100*(1-\alpha)$ of the values of Z_0 must fall between $-Z_{\alpha/2}$ and $Z_{\alpha/2}$

Then, the probability type II error assumes $H_0: \mu = \mu_0$ is false, then the corresponding Z_0 is :

$$Z_0 \sim N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right), \text{ where } \mu_1 = \mu_0 + \delta, \text{ therefore } H_1: \mu \neq \mu_0 \text{ is true.}$$

The type II error is the probability that the test statistic will fall between $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ when H_1 is true, as illustrated in Figure 27. A more detailed explanation of the concept of determination can be seen in [152]. Finally, the probability can be evaluated as

$$F(Z_{\alpha/2}) - F(-Z_{\alpha/2}),$$

F is the cumulative distribution function of the distribution $N\left(\frac{\delta\sqrt{n}}{\sigma}, 1\right)$.

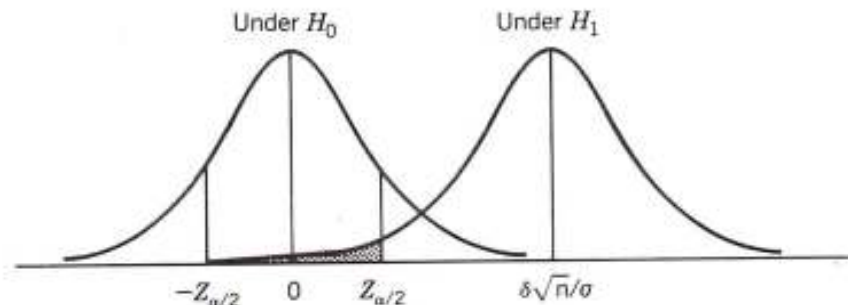


Figure 27: Graphical representation of type II Error [152]

Then, by changing from the cumulative distribution function to the standard normal cumulative function, the type II error can be deduced as:

$$\beta = \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}\right) \quad (32)$$

Where:

Φ is the cumulative standard normal distribution

δ is the difference between the mean value used to calculate the threshold value and the mean value of the monitored interval of time of damage indicators SAD and SSD.

σ is the standard deviation

n is the sample size

Table 12: Summary of type I and type II errors in decision making (adapted from [124])

Decision	Truth about the population	
	H0 True	H1 True
Reject H0	Type I error Probability = α	Correct decision Probability = $1-\beta$
Accept H0	Correct decision Probability = $1-\alpha$	Type II error Probability = β

Another interesting statistical parameter is the process power defined as:

Power = $1-\beta = P\{\text{reject } H_0 | H_0 \text{ is false}\}$; it is the probability of correctly rejecting H_0 when H_0 is false. It is desirable to have a higher value of Power. In other words, the power is the probability of making the right decision.

4.2.6.3 Illustration of anomaly detection step

As it done with previous step, let use an illustrative example to show the anomaly detection steps. Thus the input to the anomaly detection decision is the PPCA data fusion step. By using the same data set as an example as in the data fusion step, the result of the PPCA was three principal components.

Step 1:

The signal representing the system is reconstructed using the three principal components shown on Figure 25 (a), (b) and (c). Thus the reconstructed signal is shown on Figure 28 below

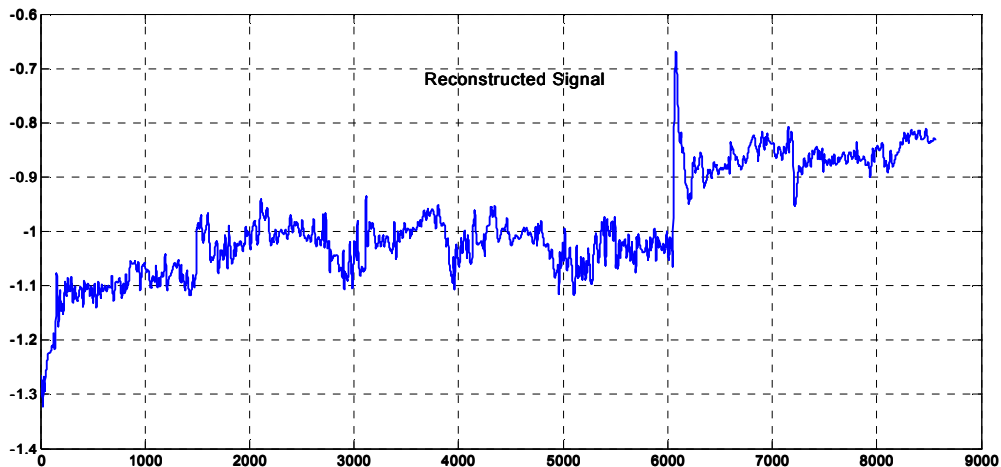


Figure 28: Reconstructed 1-dimensional signal from the 3 PPCA

Step 2:

The convergence of the statistical parameters for SAD and SSD is obtained as shown in Figure 29.

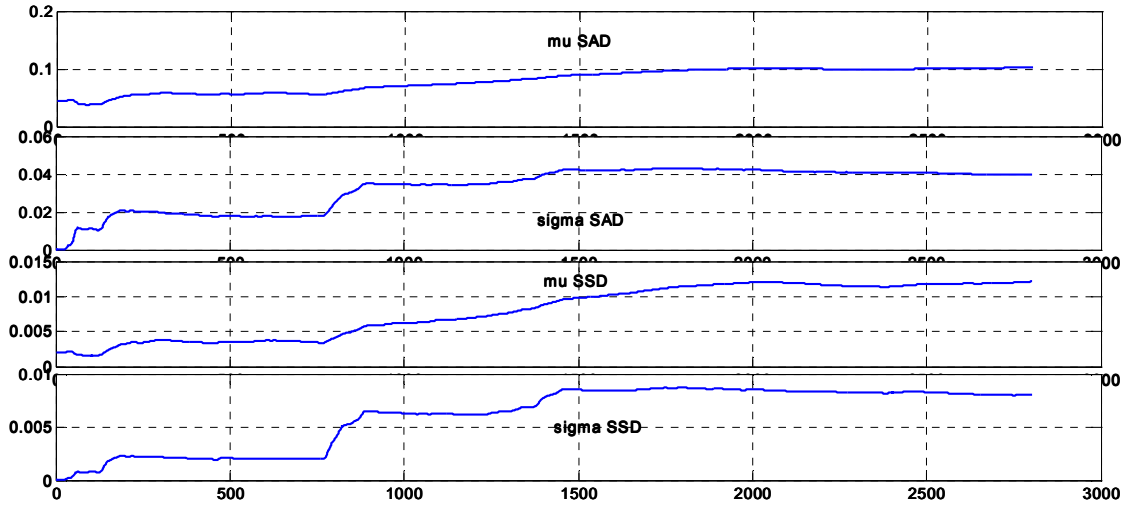


Figure 29: Converged statistical parameters

Step 3:

The upper specific limits for SAD and SSD are computed for a value of α of 2% that serves as the threshold as well: UL_{SAD}^α is 0.20126 and UL_{SSD}^α is 0.04295. Thus, Figure 30 shows the reference time series signal used to determine the statistical parameters (blue), monitor time series (green) and threshold (red dash) for SAD damage indicator. Also, Figure 30 shows the magnitude of the anomalies.

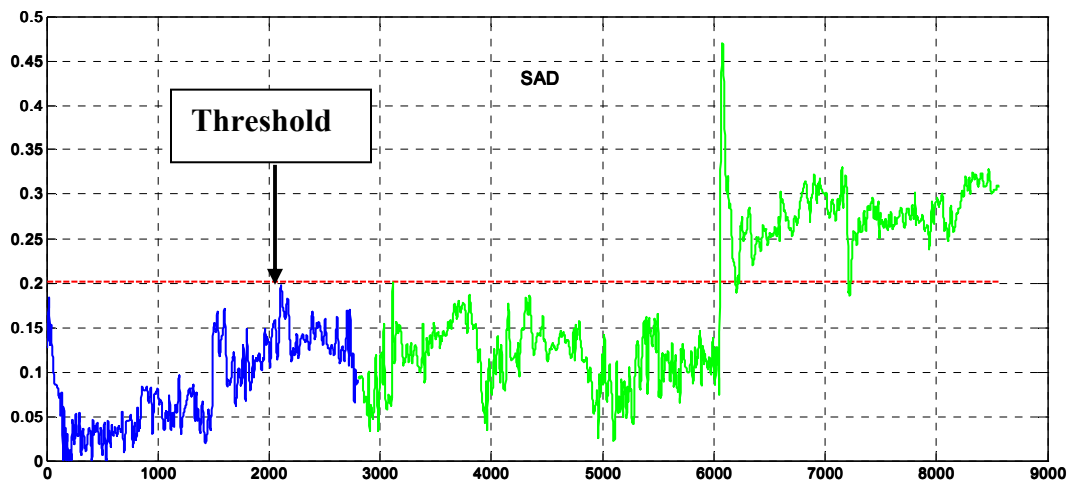


Figure 30: Monitoring of the damage indicator SAD

Similarly, Figure 31 shows the reference time series signal used to determine the statistical parameters (blue), monitor time series (green) and threshold (red dash) for the SSD damage indicator.

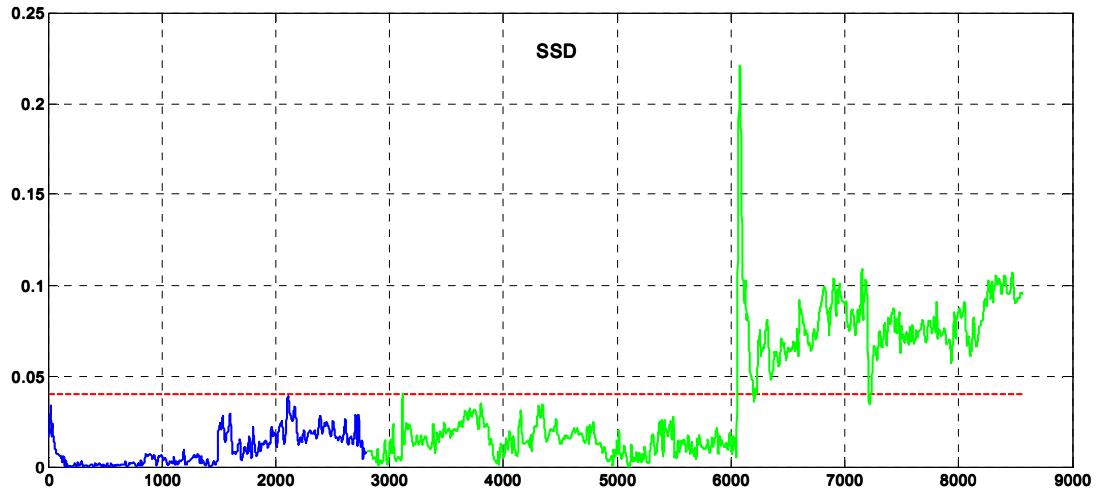


Figure 31: Monitoring of the damage indicator SSD

After setting the threshold, the result of the hypothesis testing is shown in Figure 32

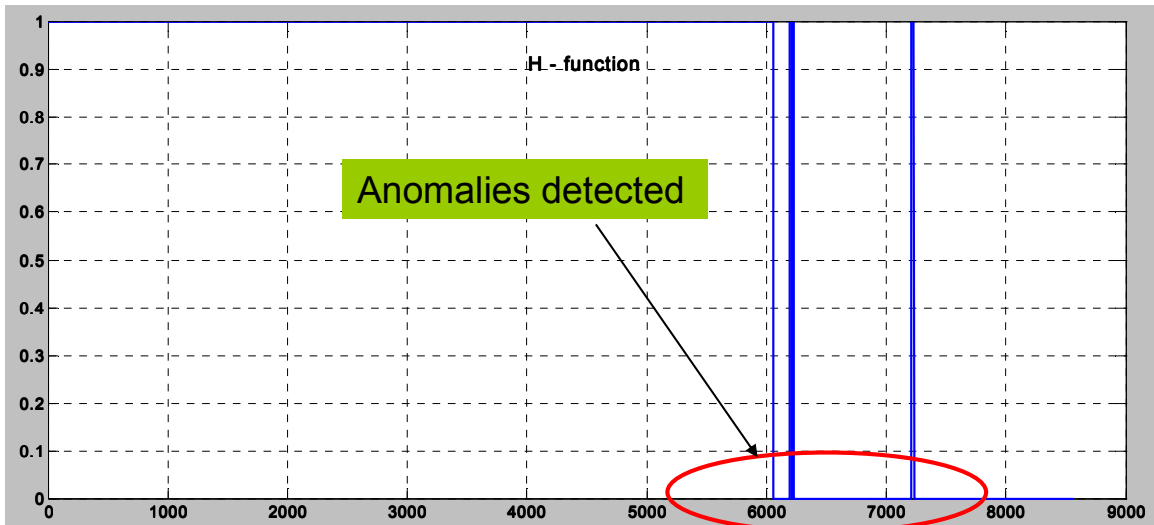


Figure 32: H-Function of the hypothesis testing result

4.2.7 Combining both types of information

After completing all the steps of the precursor detection, let us address the research question 1a. Recall that the research question 1a stated: “*How can a failure precursor detection method be made robust?*”

- Why is the current industry standard missing many early precursory signs of catastrophic failure?
- Can combining both vibration sensors and performance sensors decrease the number of missed detections?

To address this research question, let’s use an example with eleven monitored sensors (three vibration sensors and eight performance sensors).

4.2.7.1 Vibration sensors only

After the preprocessing and de-noising of each of the three vibration sensors, the PPCA step is done. The eigenvalues contribution is shown in Table 13.

Table 13: Pareto plot of Eigenvalues contribution for vibration sensors only

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.4254	83.850		83.850
2	0.0784	15.461		99.311
3	0.0035	0.689		100.000

After the conventional PCA, the PPCA is done, using the maximum likelihood for weight matrix and with entry ($W \geq 0.1$), for the two retained principal components (99.31% of information content). This is shown in Table 14

Table 14: two retained principal components from the PPCA step for vibration sensors only

BB1	0.67079	-0.10819
BB2	0.71260	-0.18010
BB4	0.20550	0.97768

The reconstructed signal is shown in Figure 33.

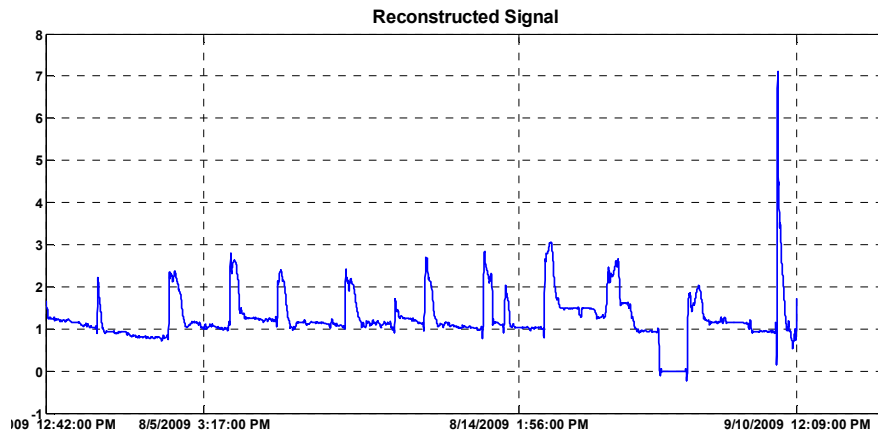


Figure 33: Reconstructed signal for vibration sensors only

Finally, the hypothesis testing is performed and show in Figure 34.

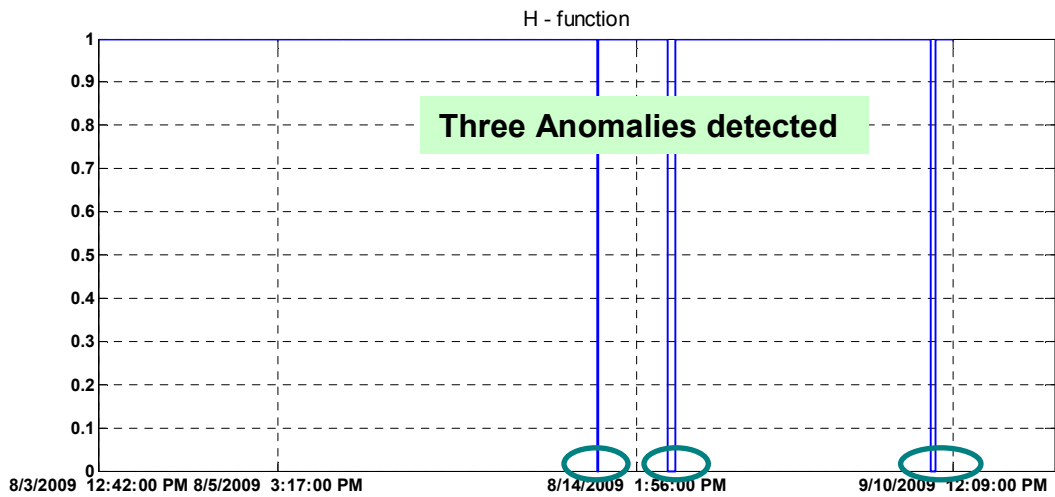


Figure 34: H-function for vibration sensors only

4.2.7.2 Performance sensors only

In a similar fashion, all the steps are repeated with the eight performance sensors only.

The eigenvalues contribution is shown for performance only in Table 15

Table 15: Pareto plot of Eigenvalues contribution for performance sensors only

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.1717	94.989		94.989
2	0.0089	4.928		99.917
3	0.0001	0.069		99.986
4	0.0000	0.006		99.992
5	0.0000	0.004		99.996
6	0.0000	0.003		99.999
7	0.0000	0.001		100.000
8	0.0000	0.000		100.000

After the conventional PCA, the PPCA is done, using the maximum likelihood for weight matrix and with entry ($W \geq 0.1$), for the two retained principal components (99.917% of information content). This is shown in Table 16.

Table 16: Two retained principal components from the PPCA step for performance sensors only

	PC1	PC2
DWATT	0.4769	-0.13152
TNH	0.40805	0
CPD	0.47494	0
CTD	0.4299	0.18129
CPR	0.44293	0
CSGV	0	0
CTIM	0	0.96825
AFPAP	0	0

The reconstructed signal is shown for performance only in Figure 35.

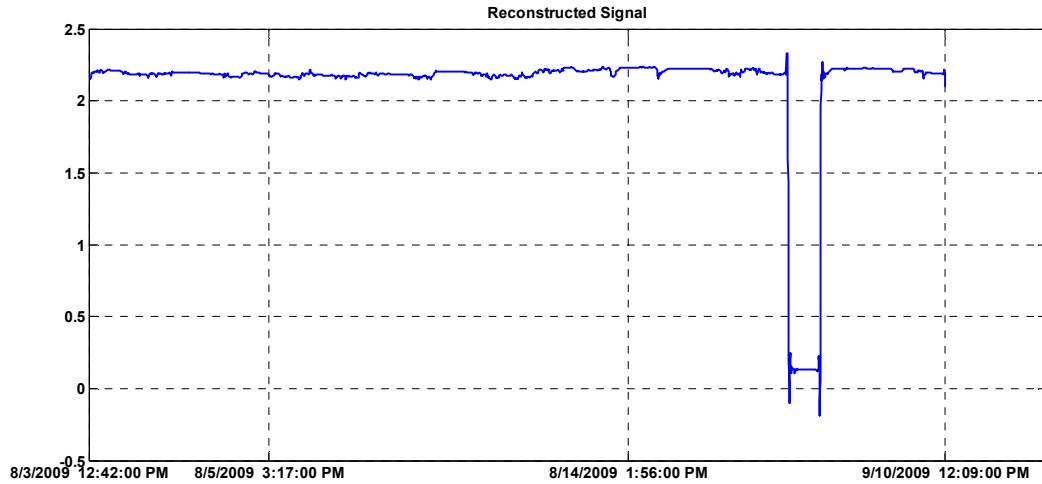


Figure 35: Reconstructed signal for performance sensors only

Finally, the hypothesis testing is done for performance sensors and shown in Figure 36.

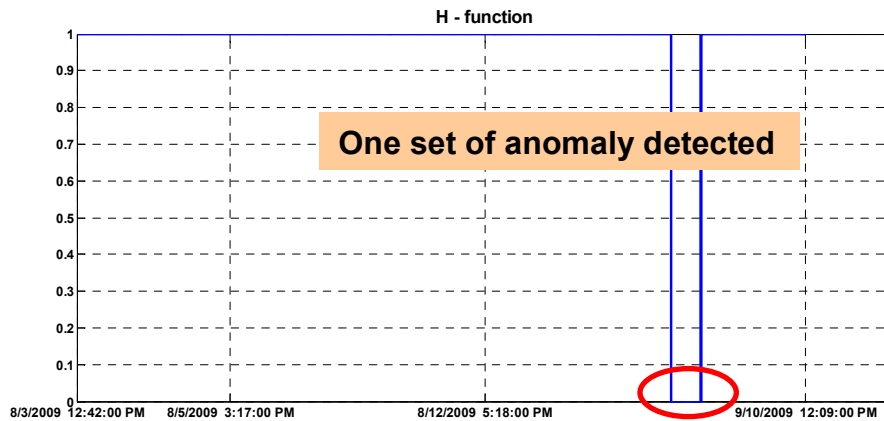


Figure 36: H-Function for vibration sensors only

4.2.7.3 All sensors (vibration and performance) combined

Now, let us combine all the sensors and perform all the steps of the anomaly detection process. The eigenvalues contribution is shown for the combination of all the sensors together in Table 17.

Table 17: Pareto plot of Eigenvalues contribution for all sensors

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.4846	70.425		70.425
2	0.1679	24.402		94.827
3	0.0238	3.456		98.283
4	0.0082	1.188		99.471
5	0.0035	0.508		99.979
6	0.0001	0.017		99.997
7	0.0000	0.001		99.998
8	0.0000	0.001		99.999
9	0.0000	0.001		100.000
10	0.0000	0.000		100.000
11	0.0000	0.000		100.000

After the conventional PCA, the PPCA is done, using the maximum likelihood for weight matrix and with entry ($W \geq 0.1$), for the three retained principal components (98.283% of information content). This is shown in

Table 18.

Table 18: Three retained PC from the PPCA step for ALL sensors combined

	PC1	PC2	PC3
DWATT	0.1897	0.34279	0.28134
TNH	0.1664	0.28797	0.22198
BB1	0.60303	-0.30941	0
BB2	0.63522	0.36886	0
BB4	0.26122	0.5137	-0.80535
CPD	0.19144	0.33751	0.28019
CTD	0.17847	0.29981	0.21548
CPR	0.1778	0.31576	0.26381
CSGV	0	0	0
CTIM	0	0	0
AFFAP	0	0	0

The reconstructed signal is shown for all the sensors combined in Figure 37

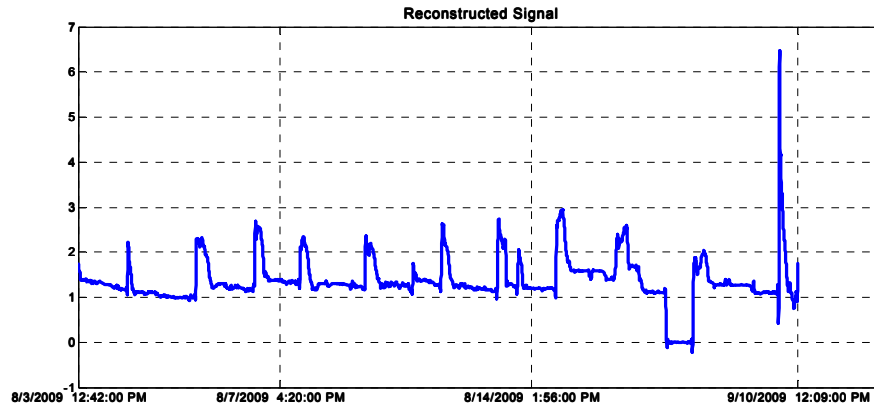


Figure 37: Reconstructed signal for ALL sensors combined

Finally, the hypothesis testing is done for all the sensors combined and shown in Figure 38.

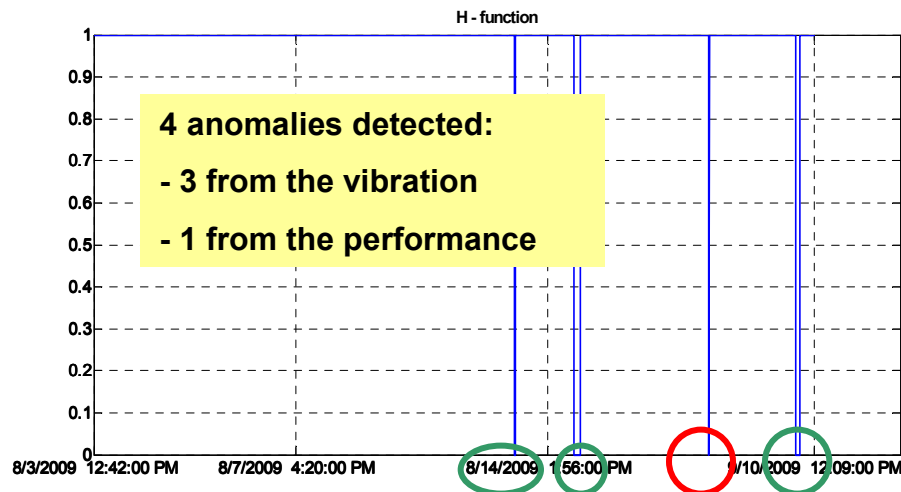


Figure 38: H-Function for ALL sensors combined

Figure 38 above shows that the anomaly signature that was not in the vibration-based analysis appears in the analysis based on the combination of all sensors. Similarly, the three anomalies detected during the vibration-based analysis, whose signatures were not

visible in the performance-based analysis, were detected in the combination of all sensors.

Recall hypothesis 1a: *“Combining the information from the two types of sensors could decrease the number of missed precursory anomalies”*

Thus, hypothesis that stated: “the combination of the two type of sensors through the proposed approach can decrease the number of missed precursory anomaly” has been verified.

Therefore, although there are damage diagnostic method that rely solely on vibration sensors [80], the combination of the vibration with the performance yields a much mmore robust anomaly detection methodology.

4.2.8 Effect of sampling interval on the proposed anomaly detection methodology

Recall research question 2a: *“How does the sampling interval impact on the quality of detection?”*

- Does the sampling time have any impact on the quality of detection?

To study the effect of the sensor sampling interval on the proposed anomaly detection technique, let us consider the following three sampling intervals: five seconds, one minute, and five minutes. Furthermore, let us perform the study through the use of an example. In this example, there are thirteen sensors that are being monitored. In each of the three cases, the normal steps of data preprocessing, de-noising, fusion, and anomaly detection are applied.

4.2.8.1 Five minute sampling interval

In this case, the data set is collected using sensors reading at five minute intervals.

Table 19 shows the Pareto plot representing the contribution of each of the eigenvalues.

Table 19: Pareto plot for five minute sampling interval

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.0843	64.645		64.645
2	0.0428	32.802		97.447
3	0.0018	1.344		98.791
4	0.0015	1.135		99.927
5	0.0001	0.050		99.977
6	0.0000	0.012		99.988
7	0.0000	0.005		99.993
8	0.0000	0.003		99.996
9	0.0000	0.002		99.998
10	0.0000	0.001		99.999
11	0.0000	0.001		100.000
12	0.0000	0.000		100.000
13	0.0000	0.000		100.000

The Figure 39 below shows the magnitude of the anomaly in the five minute sampling interval of the SAD damage indicator.

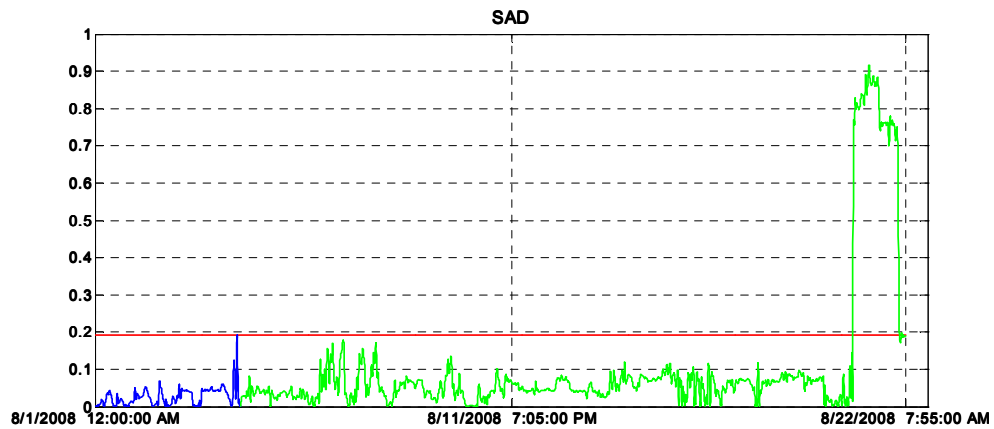


Figure 39: Monitoring of the damage indicator SAD for five minute sampling intervals

The result of the hypothesis shown in Figure 40 indicates, that the detected anomaly started on 008/21/2008 at 12:05AM.

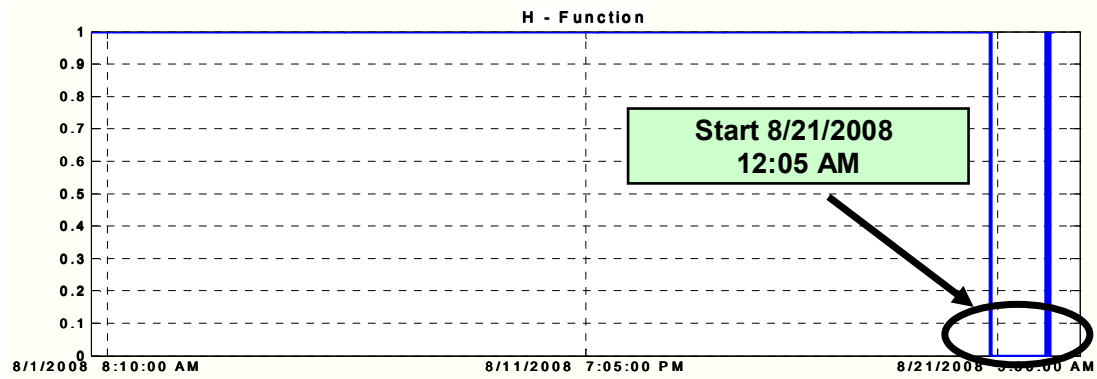


Figure 40: H-Function for five minute sampling interval

4.2.8.2 One minute sampling time

In an approach similar to the one minute sampling interval case, the one minute interval is performed. Table 20 shows the Pareto plot representing the contribution of each of the eigenvalues

Table 20: Pareto plot for one minute sampling interval

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.0848	64.686		64.686
2	0.0430	32.775		97.461
3	0.0018	1.341		98.802
4	0.0015	1.126		99.927
5	0.0001	0.049		99.977
6	0.0000	0.012		99.988
7	0.0000	0.005		99.993
8	0.0000	0.003		99.996
9	0.0000	0.002		99.998
10	0.0000	0.001		99.999
11	0.0000	0.001		100.000
12	0.0000	0.000		100.000
13	0.0000	0.000		100.000

Figure 41 below shows the magnitude of the two anomalies in the one minute sampling interval of the SAD damage indicator.

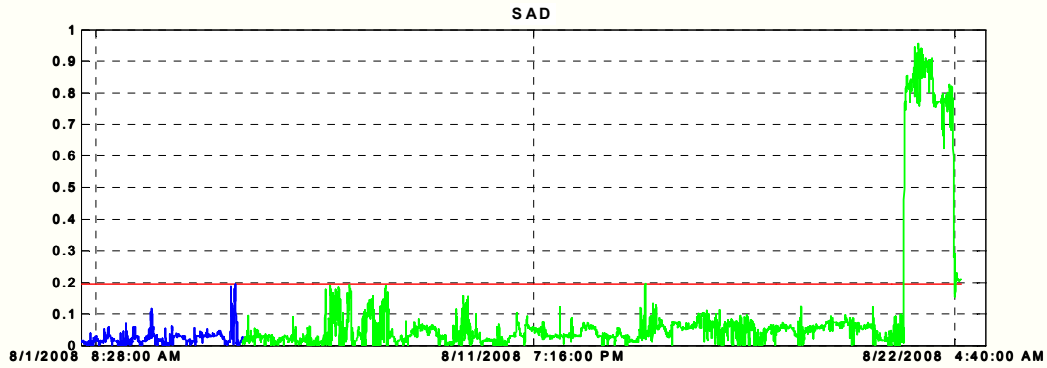


Figure 41: Monitoring of the damage indicator SAD for one minute sampling interval

In the case of the one minute sampling interval, the result of the hypothesis testing shown in Figure 42 indicates that an anomaly was detected that started on 008/21/2008 at 12:05AM (same as in the five minute sampling interval case), but a short anomaly was also detected on 08/14/2008 from 23:05 to 23:06 (lasting only one minute).

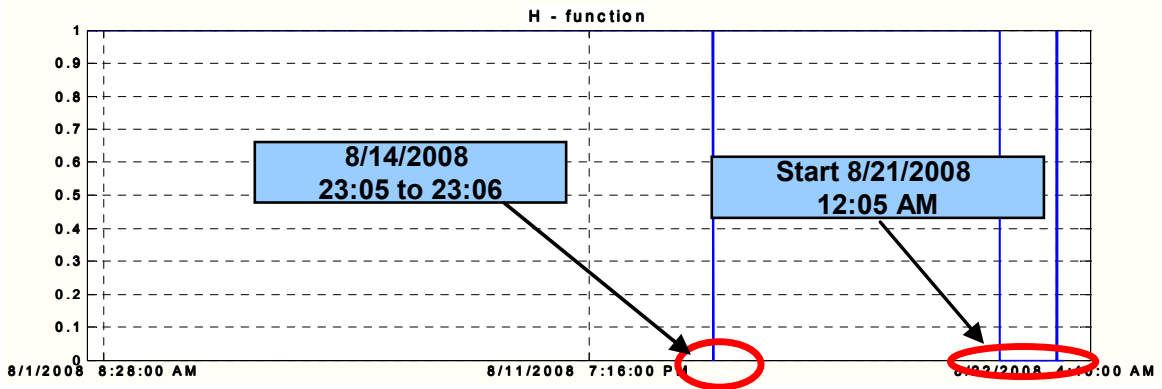


Figure 42: H-Function for one minute sampling interval

4.2.8.3 Five second sampling time

Finally, the five second sampling interval was performed in a similar fashion to the previous tests.

Table 21 shows the Pareto plot representing the contribution of each of the eigenvalues.

Table 21: Pareto plot for five second sampling interval

Number	Eigenvalue	Percent	20	40	60	80	Cum Percent
1	0.0389	53.781					53.781
2	0.0305	42.128					95.909
3	0.0018	2.437					98.346
4	0.0011	1.565					99.911
5	0.0000	0.046					99.957
6	0.0000	0.020					99.977
7	0.0000	0.009					99.987
8	0.0000	0.005					99.992
9	0.0000	0.004					99.996
10	0.0000	0.003					99.999
11	0.0000	0.001					100.000
12	0.0000	0.000					100.000
13	0.0000	0.000					100.000

Figure 43 below shows the magnitudes of the anomalies in the five second sampling interval of the SAD damage indicator. It is seen in the figure that there are many more anomalies than in either the five or one minute sampling interval case.

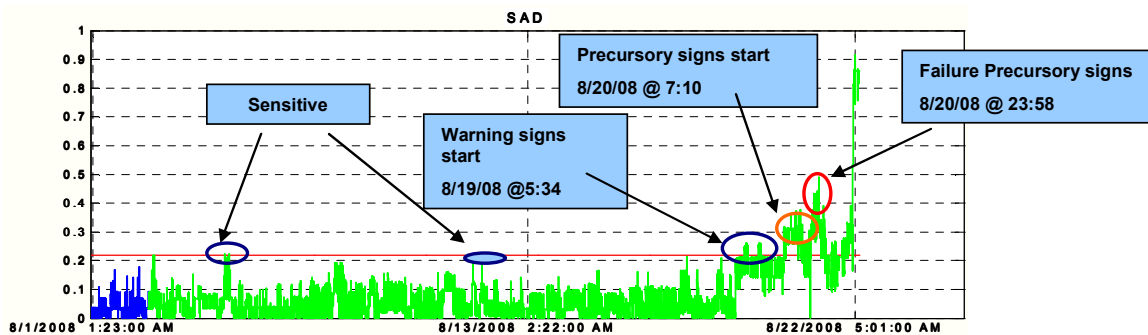


Figure 43: Monitoring of the damage indicator SAD for five second sampling interval

Recall the stated research hypothesis 2a: “*The smaller the time step (the higher the frequency) is, the more accurate the detection quality will be.*”

This research question investigated whether or not the sampling interval had an impact on the anomaly detection, and the hypothesis stipulated that: “the smaller time step (the higher frequency) will have better detection capability.” Through the study of the three different time intervals, it is clear that the smaller sampling interval leads to better detection capabilities, we see that the finer the interval, the more accurate the anomaly detection decision. Also, we see that the smaller the sampling interval, the higher the sensitivity to anomaly detection. However, the smaller the sampling interval, the longer the analysis execution time will be. More importantly, the higher the sensitivity, the more the detection is prone to false alarms.

In sum, the stated research hypothesis 2a is justified through the above experiment of assessing the sampling interval effect on the quality of the detection.

CHAPTER 5

PROGNOSTICS AND HEALTH MANAGEMENT APPROACH TO RESIDUAL TIME TO FAILURE MODELING

After developing the process of anomaly detection, the next interesting question that arises is whether or not the residual time to failure may have been predicted using an understanding of the detected anomaly signatures. Thus, this chapter is devoted to addressing that thought.

5.1 Importance of Prognostics and Health Management

The ability to predict the residual life of a system after the observation of a certain anomaly has always been of interest to the research community. Thus, whenever there has been the opportunity to monitor systems and collect characteristics of those systems, the question of what can be done with that knowledge to avoid surprise system failures has been raised. In other words, in the unfortunate situation where an irreversible failure precursor is observed, how can the acquired knowledge allow the prediction of the remaining life of that system?

The medical field has been the pioneer in the prognostication of remaining lifespan. These estimates are made when certain potentially life-ending diseases like severe acute strokes or “advanced cancers,” known to the general public as “terminal cancers,” are diagnosed in patients. In fact, for a century the medical field has committed a lot of studies to the accurate prediction of the length of time remaining for certain patients with terminal illness, because it is important “for clinical, organizational, and ethical reasons, especially in helping to avoid harm, discomfort, and inappropriate therapies in vulnerable patients and, conversely, in planning specific care strategies” [153]. Moreover, knowledge of survival time can influence many important personal

decisions the patients and his/her love ones might make to enhance the end-of-life experience. In order to improve the accuracy of the prognostication, there are many studies that have particularly focused on determining the factors or health predictors that may be used to correctly predict survival time after the detection of fatal diseases [153-155].

However, as a result of those studies, it was learned that the predictors that may be useful in predicting survival time when a fatal illness is caught at an early stages may be totally different from those that are useful when the discovery is made at a later stage. For instance, in the case of advanced cancers, the prognostication is typically based on tumor development at the earlier stages, whereas at advanced stages (defined as median survival time of no more than 90 days), prognostication is based on many more factors. In [153], Maltoni et al. report that the remaining life estimation is still feasible, but it depends on many other health parameters, which are summarized in Table 22.

Table 22: Factors Subdivided on the Basis of Level of Evidence Obtained by a Correlation With Actual Survival [153]

Factor for which a definite correlation with prognosis has been identified	Factors for which a correlation has been indicated but not confirmed or for which a statistical significance has been identified in patient populations with less advanced disease or for which contradictory data have emerged	Factors with controversial indications
Clinical prediction of survival Performance status	Pain Nausea	Multidimensional quality-of-life questionnaire; it is possible that their prognostic capacity is a result of the identifying components of physical
Signs and symptoms of cancer anorexia-cachexia syndrome anorexia, weight loss, dysphagia, and xerostomia	Tachycardia Fever	
delirium	Neoplastic pattern (primary and secondary sites)	
Dyspnea	Comorbidity	
Some biologic factors (leukocytosis,	Anemia	
Prognostic scores	Hypoalbuminemia	
	Prehypoalbuminemia	
	Proteinuria	
	Serum calcium level	
	Serum sodium level	
	Lactate dehydrogenase and other enzymes	
	Patient characteristics (age, sex, and marital status)	

In a similar fashion, as designed systems became more complex and expensive, the ability to detect abnormalities that may lead to their failure arises as a necessity. For example, in an industrial application, for systems in which an emergency shutdown may create disruption or may be costly (e.g. heavy duty gas turbine for electrical production), it is important to know the level of risk of continuous safe operation, until it becomes less disruptive to shut the system down. The ability to prognosticate the remaining useful life (RUL), or the amount of time left before system health will deteriorate below a defined failure threshold, is important for critical systems. RUL prediction becomes especially valuable in cases of expensive systems such as gas turbines where the failure threshold could be an irreversible catastrophic mechanical failure. Additionally, RUL estimation is necessarily a prediction or forecasting endeavor; that is, it is an extrapolation for the future process [21] from what is believed to be the system current health, as illustrated in

Figure 44 below. For that reason, Vachtsevanos stated that RUL prediction is the most challenging part of the prognostic development [50], because the future is never known with certainty. Therefore, any prediction would be essentially based on models that would have architecture similar to the one shown in Figure 44, and would be built to represent the system behavior based on past knowledge of causes and their resulting effects.

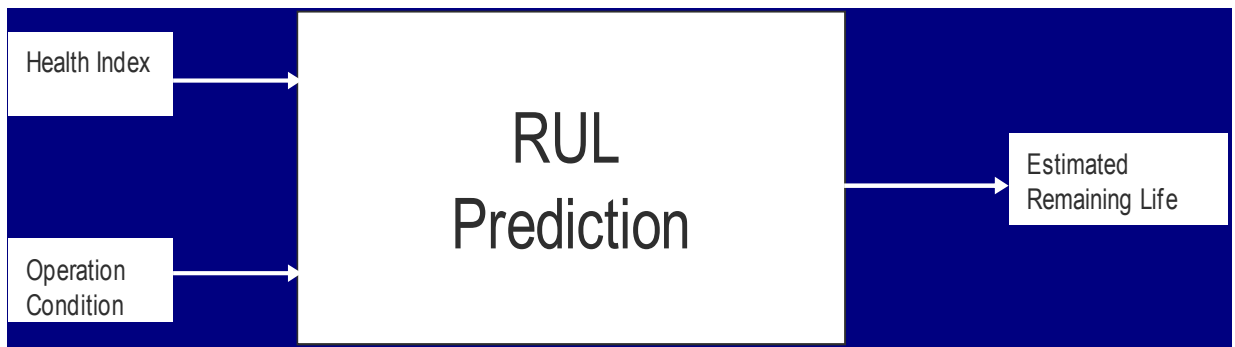


Figure 44: Schematic of the RUL architecture

There are many ways to predict the residual time to failure or to foresee how a system or component may behave during its remaining useful life, as one can see from the literature [156-158]. Furthermore, there are regression models that can serve as templates or starting points for further development [159].

In the modern literature, the field of prognostics that is dedicated solely to studying the residual time to failure of systems or components, based on their intrinsic health from monitoring, is referred to in the general sense as Prognostics and Health Management, or PHM. The PHM field is so vibrant and active that organizations have been created exclusively for its advancement, such as the National Society of PHM. Journals and magazines like the the International Journal of Prognostics and Health Management (IJPHM), and professional meetings and conferences such as the annual conference of the PHM or the IEEE International Conference on Prognostics and Health Management, just

to cite a few, are also signs of the vibrancy of the field. PHM applications and research are thriving in many different fields, ranging from big, general areas such as aircraft systems[160], to some of the narrower and more specific fields such as avionics systems [161] and electronic systems [162]. Current research in the field of PHM is being done both at the systems level, as illustrated in Figure 45, and at the component level [163, 164]. Generally, “PHM includes every method used to assess the health of systems elements beginning with the manufacturing process and continuing through all levels and elements of operation” [165].

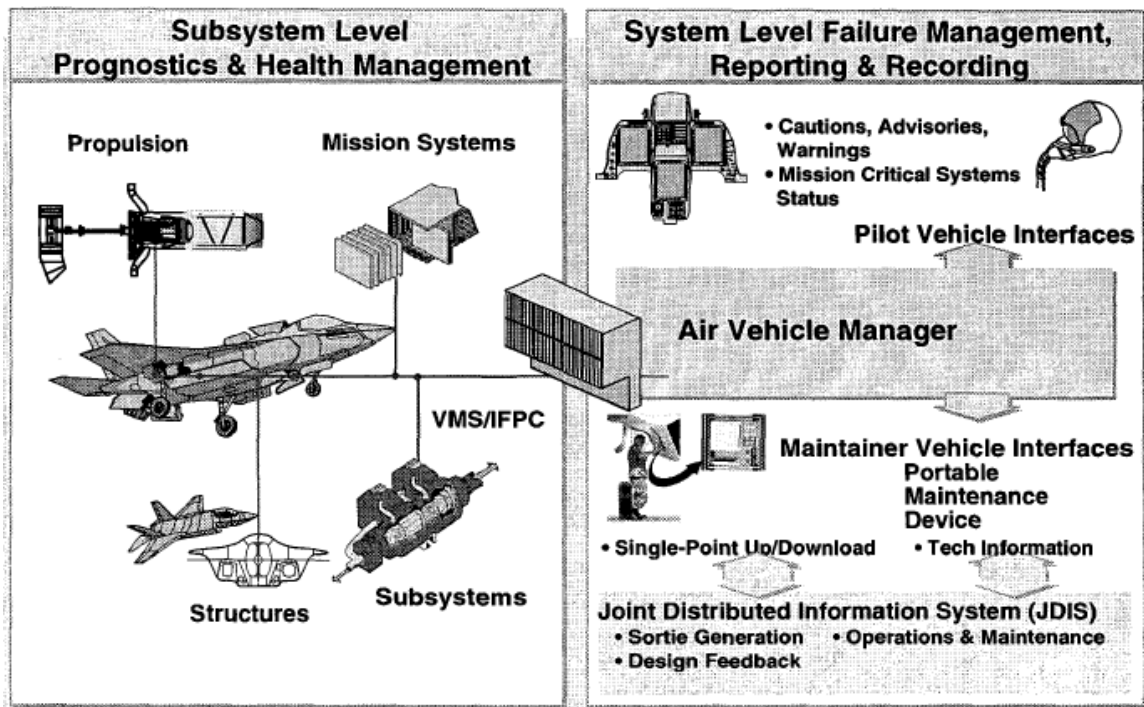


Figure 45: Example of PHM System Architecture for a fighter aircraft [165]

In an example illustrating the broad need for and application of the PHM, shown in Figure 46, B. Ferrell states that PHM is a system function that touches every element of the weapon system in some way.

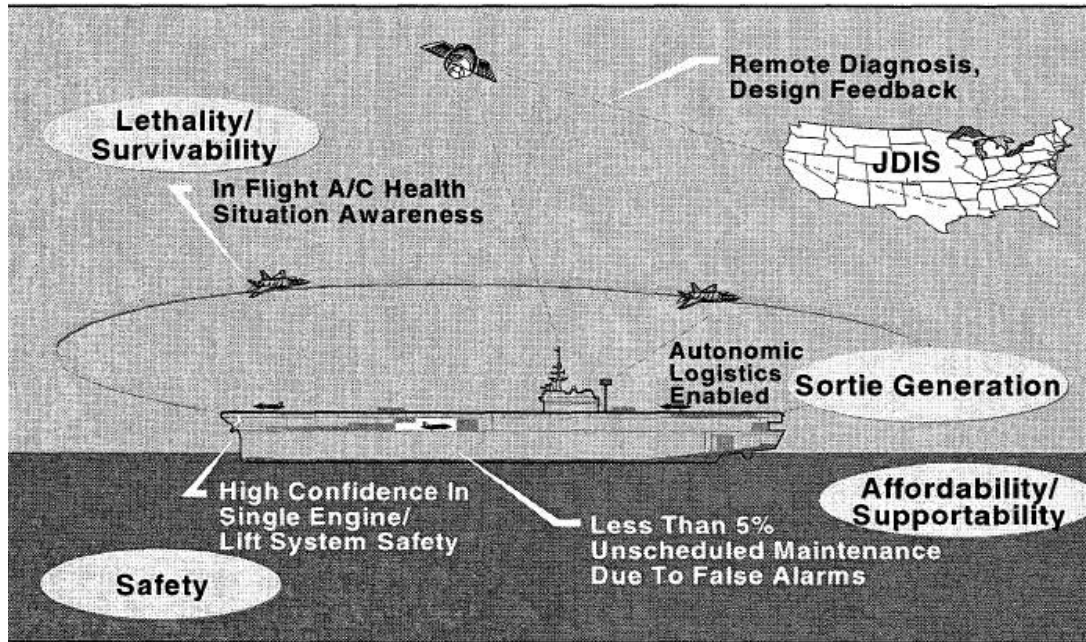


Figure 46: PHM operations in military application [165]

Regardless of the way PHM is implemented or studied, it remains strategically based, in one form or another, on detecting and understanding the signature of anomalies.

One thing is for certain: the more complex a system is, and the more critical its safety and reliability are, the greater the need is to implement PHM, and the more the related cost can be justified.

5.1.1 Data generation process for PHM analysis

The very first issue in any regression building model is to identify the predictors and their corresponding responses. In the case of deterministic prediction, only after the data set of paired inputs and outputs is defined and generated can a regression model be built. The output of interest in this work is the residual time to failure after a failure precursor is detected. The preliminary problem remains in defining a system's health index so that it can be used as an input or predictor factor.

The underlining assumption made here is that the signature of the residual time to failure can be observed in the health indexes. Therefore, the identification of the appropriate

health indexes should allow the building of a residual life estimation regression model, which would map the health indexes after a diagnosed precursory event in order to determine remaining time to failure.

The following health indexes were identified and observed:

- The Anomaly Severity Index (ASI), which records highest value reached during an anomalistic period. This index is referred to as the severity index because it provides an idea of how bad the damage is. It is defined as:

$$ASI(k) = \frac{SAD(k)}{USL} \quad (33)$$

Where:

- $SAD(k)$ is defined as the sum of the absolute difference between the reconstructed signal (representing the system) energy content at the k^{th} time step after the threshold value is set, and the reference energy E_{ref} .
- USL is the calculated upper specific limit that corresponds to the anomaly threshold.
- The Anomaly Duration Index (ADI)

ADI (k) is k^{th} the length of time the monitored signal (reconstructed signal) stays above the anomaly threshold after any crossing of the threshold. That is, the amount of time between the point of detection of an anomaly and the end of that anomaly.

That is:

$$ADI(k) = \Delta T, \text{ while } SAD(k) > USL$$

- The Residual Time to Failure, RTTF (k), is the time between the detection of the k^{th} anomaly and the time of the system failure.

Figure 47 illustrates all the parameters defined above for the residual life analysis.

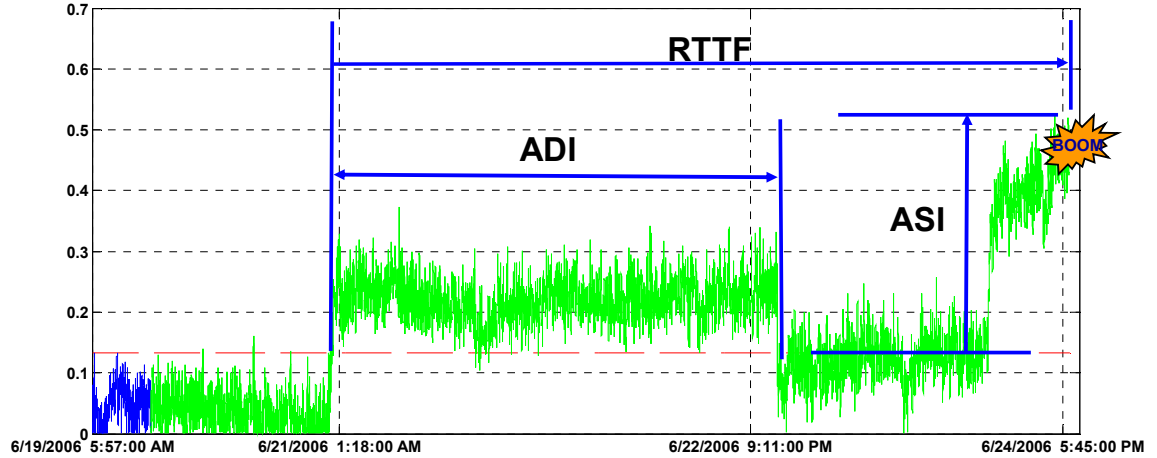


Figure 47: Illustration of the prognostics parameters

So the defined health indexes (ASI, ADI) are mapped into the output or response. The output target is defined as the interval of time between the moment the first failure precursor is detected, and the time of the system failure. In other word, the problem in hand can be represented as follows:

$$RTTF(k) = f(ASI(k), ADI(k)) \quad (34)$$

Thus, the first attempt to resolve the problem posed in equation (34) is the deterministic approach, which is presented in the next section.

5.2 Deterministic approach to modeling and simulation for estimation of residual life

Ideally, it is desirable to be able to deterministically estimate the residual life of a system or component based on an assessment of its current level of health. In practice, this means the development of the capability to model responses giving a set of determined corresponding predictors. Although the idea of developing regression models for residual life estimation is not new, there have always been many challenges to overcome, like in any regression problem, including knowing all the predictors that affect a given response.

The regression modeling works by mapping a set of causes and their corresponding effects. Typically, models to represent causes and deduced effects are constructed by applying statistical regression techniques to experimental data. In general, regression techniques can be subdivided into two main families: parametric regressions and non-parametric regressions. Among the most commonly used regression techniques are the Response Surface Equation (RSE), the Gaussian Process, and the Artificial Neural Networks (ANN). Thus, a brief overview of each of the three regression techniques will be given in the following sections.

5.2.1 Parametric Regression

5.2.1.1 Response Surface Equations (RSE)

Response Surface Equations are simplified polynomial equations used to model the behavior of complex systems. RSE are obtained through multivariate regression techniques known as Response Surface Methodology (RSM) [166, 167]. RSM is a collection of statistical or mathematical techniques useful for developing, improving, and optimizing processes [166]. Typically, RSM is used to relate the factors (or predictors) to the measured responses over some specified region of interest [168]. The underlying principle of the RSM is that it is an experimental strategy that consists of exploring the factors, empirical statistical modeling to develop an appropriate approximation between the yield and the process variables, and optimization methods for finding the values of the process variables that produce desirable response values [169]. The RSE technique models equations through the usage of design of experiment (DOE). The most popularly used RSE is the second-order Taylor series approximation because it requires a minimal computational investment and is as follows:

$$R = b_o + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k b_{ij} x_i x_j + \varepsilon \quad (35)$$

Where,

R is the dependent parameter (response) of interest.

b_o is the intercept term.

b_i are regression coefficients for the first order terms.

b_{ii} are coefficients for the pure quadratic terms.

b_{ij} are the coefficients for the cross-product terms.

$x_i; x_j$ are the independent variables.

k is the number of factors.

ε is the error associated with neglecting higher order effects

RSE assumes that for any model the error, ε , should be distributed as $N(0,1)$.

The RSE approach has many advantages, such as:

- It provides a rapid correlation between design variables or technology metrics and the system level impacts.
- It uses a simplified equation to represent a complex system.
- Its sensitivities are easily obtained.
- Its optimization process is easily obtained.
- It is rapid and efficient.
- It provides instantaneous evaluations.

However, RSE has a series of limitations as well:

- Large variations in the factors can be misleading and generate error or bias
- Critical factors may not be correctly defined or specified.

- If the range of levels of factors is too narrow or too wide, the optimum cannot be defined.
- There is a lack of use of good statistical principles.
- There is an over reliance on computers (e.g. one can only make sure the regression results make sense using a computer).

In order to achieve a satisfactory result when using RSE, the following assumptions need to be verified:

- The critical factors are known.
- The region of interest is known.
- The factors vary continuously throughout the experimental range tested.
- There is a mathematical function that relates the factors to the response.
- The response defined by the function is a smooth curve.

Besides the listed drawback, one of the chief limitations of the RSE method is that it relies on linear requirements, so it can't handle non-linear behaviors of the system parameters, which non-parametric techniques like the artificial neural networks can. In sum, because of the many advantage of the RSE, it should always be considered during regression problems. In fact, because RSM approximates the behavior of a response with respect to certain specific design parameters, it has the capability to quickly give a system analyst insightful knowledge into a problem when the proprietary models are not available for use.

5.2.2 Non-parametric Regression

5.2.2.1 Artificial Neural Networks

The Artificial Neural Networks (ANN) is a non-parametric regression technique that can be used for highly non-linear parameters or discrete problems. Thus, ANN has

the capability to address one of the major limitations of the response surface equation: the inability to handle non-linear problems.

Typically, ANN works by trying to mimic the way a brain functions. Artificial neurons are usually simplified versions of a brain cell. The reason for trying to create a technique that imitates the functional processing approach of the brain is that the replication of the structure of the brain is the best way to achieve artificial intelligence. A human's brain has several billion neurons. Each neuron can have thousands of connections, and each has an axon that serves to transmit a signal, and dendrites that receive signals from other neurons. The signals that are received through the dendrites of a neuron pass a set threshold value; the neuron in turn fires a signal through its axon. Although each connection is simple alone, when a large number of them are put together it leads to fairly complex behavior [170]. Though the structure of a neural network can be complex, Figure 48 shows a simple artificial neuron that is the nuclei of the most complex structures:

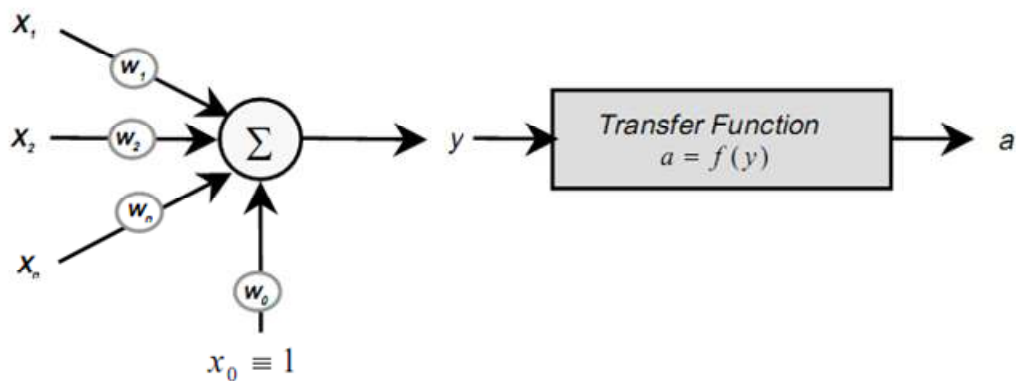


Figure 48: Illustration of an artificial neural from [171]

In the illustrative Figure 48, the x_i are the network inputs, the w_i are the weights, y is the weighted sum of inputs, and also the input to the transfer function f , while a is the network response. That is:

$$y = \sum_{i=0}^n x_i w_i = \bar{x} \cdot \bar{w} \quad (36)$$

Where: $\bar{x} = \{1, x_1, x_2, \dots, x_x\}$ is the vector of inputs and $\bar{w} = \{w_0, w_1, w_2, \dots, w_x\}$ is the vector of weights.

Several neurons can be grouped in layers to construct structures as complex as needed. Typically, functions can be represented as the weighted sum of orthogonal basis functions [47]. The weights used to model a neural network are obtained during the learning (or training) process. Ultimately, the structure and dynamics of the network and the learning (training) process are the major components defining a neural network. The power of neural networks lies in their ability to combine logical parallel computations with serial operations [172]. All the neuron layers can run in the same direction (i.e. feed-forward) or they can feedback from layers to preceding layers (dynamic or recurrent). In general, ANN requires a training of the neural nets with examples of input-output pairs of data [51]. After an adequate training, neural networks can be readily used for process parameter (or state) assessment without requiring any knowledge of the underlying system. This capability makes the neural networks a very powerful tool for pattern recognition [45]. Usually, the accuracy of the regression can be improved by increasing the number of neurons, which comes with the cost of an increase in the computational burden. A conceptual diagram showing the different components and connections between neural network layers is shown in Figure 49.

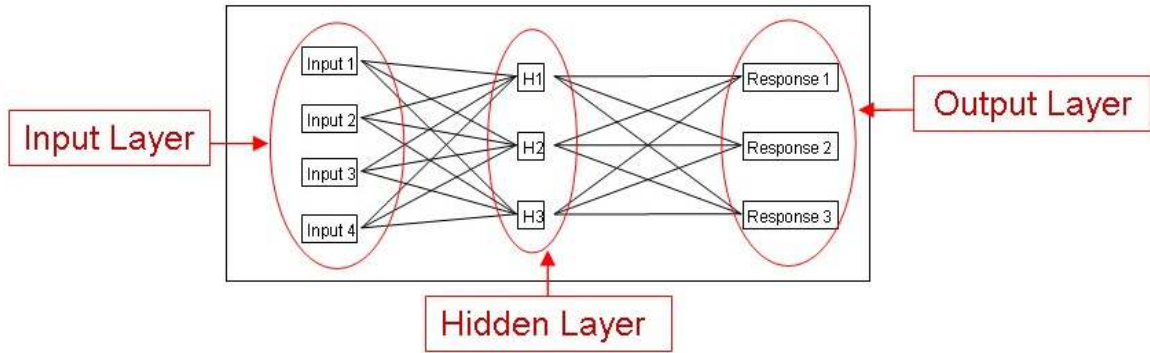


Figure 49: Neural Network Conceptual Diagram [170]

Although the conceptual neural network in Figure 49 has a simple structure, neural networks can be classified into two major categories: the static neural network and the dynamic neural network [173].

- 1) Static neural networks, also called feed-forward neural networks, have no feedback elements and contain no time delays, as illustrated in Figure 50

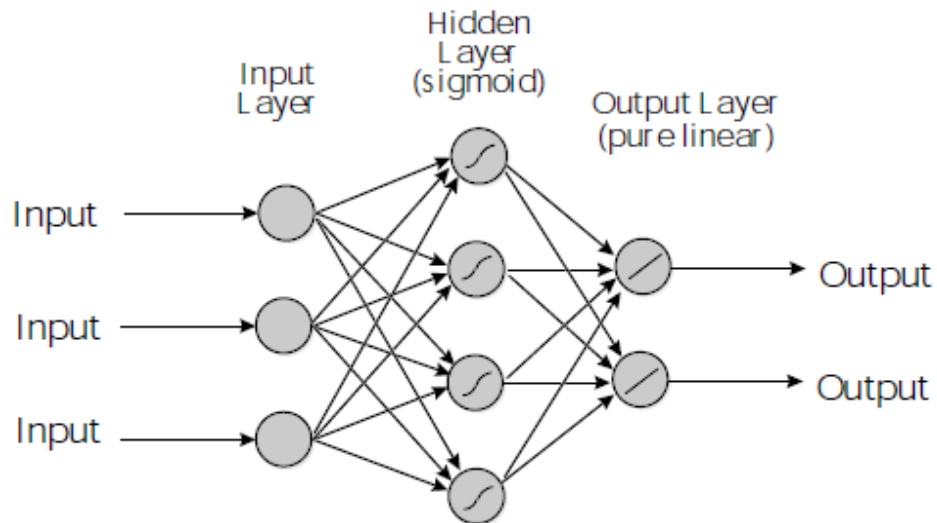


Figure 50: Feed-Forward Neural Network

The static network is trained through a scheme called backpropagation, which relies on determined training algorithms such as the Levenberg-Marquardt to obtain the weights.

2) Contrary to the static neural network, the outputs of dynamic networks depend on the current or previous inputs, outputs, or states of the network. Furthermore, the dynamic network can in turn be divided into two subgroups:

- Dynamic networks with feed-forward connections only. In this case, the output response depends not only on the inputs, but also on time. Moreover, the application is more common when the inputs also depend on time. A depiction of a dynamic feed-forward network is shown in Figure 51.

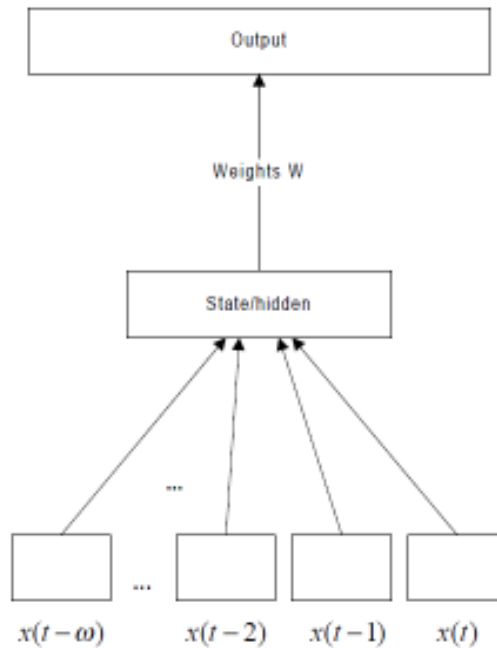


Figure 51: Depiction of a dynamic feed-forward network [174]

- Dynamic networks with feedback or recurrent connections. In this case, the output state depends on the previous states and the time. In [175], the author explains that recurrent neural networks (RNN) are neural networks with one or more feedback loops, where the feedback can be of a local or global kind. The RNN has the

ability to store memory of preceding states, which allows the RNN to record the characteristics of a system it represents over time [174]. Figure 52 shows an illustrative representation of the general structure of a recurrent network. Obviously, that makes the RNN the structure of choice when modeling a dynamic system in which current output depends on past as well as present information. This unique ability makes the RNN very powerful and popular.

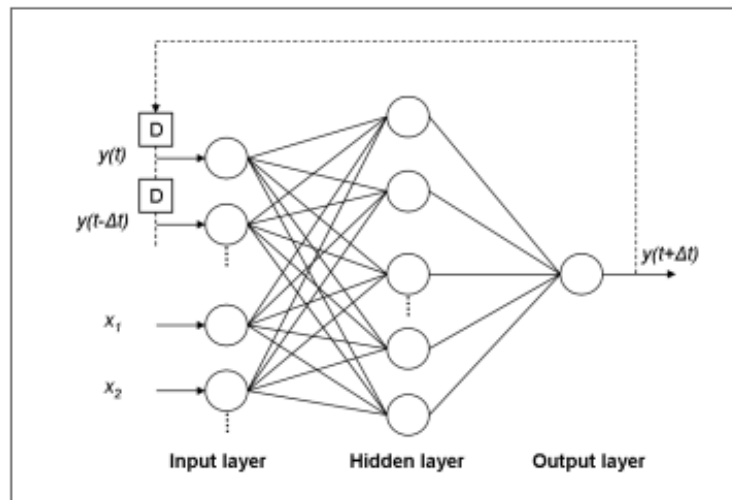


Figure 52: Recurrent Neural Network [176]

There are many training algorithms for the dynamic neural network. A list of them and their implementations can be seen in [177].

It is worth noting that even though the simple structure of neural networks is shown in previous figures, the two most popular types of neural network are the multi-layer perceptron (MLP) and the radial basis function (RBF) [51]. Therefore, it is possible to create structures as complex as necessary using multiple layers.

Regardless of the type of neural network needed for a system modeling, there are two types of learning procedures: supervised and unsupervised learning [171].

- 1) In case of supervised learning, the input values which have known target outputs are used. Then, the network error is computed using the difference between the target values and the network-predicted output values. Backpropagation is most frequently used for supervised training where network error is propagated back through the different layers of the network. Finally, different optimization methods can be used to minimize the network error.
- 2) In unsupervised learning, the weights are adjusted depending on the value of the network-predicted value because there is no use of target values. Another particularity of the unsupervised learning case is that the learning process never stops.

Commonly classical orthogonal functions such as the sinusoids function and the Walsh function are used as the activation functions of the neural network. Though they have been popular, they are global approximators (i.e. use global functions), which can be disadvantageous. A specific field of artificial neural network research that is getting a lot of attention is the use of wavelet family functions as activation functions. In contrast to the classical activation functions, the wavelet functions aren't limited by global approximators because they have good localization properties and orthonormal bases. When wavelets are utilized as the activation functions of a neural network, it is called a Wavelet Neural Network (WNN) [46, 47]. WNN theory was developed as an alternative to the standard feed-forward neural networks for modeling nonlinear functions [178]. The wavelet neural network has been used in a fairly high number of research projects recently, particularly in the field of fault diagnosis [179, 180]. A general illustration of the WNN is shown in Figure 53, where the structure is similar to the conventional static neural network, only the activation function is replaced with a wavelet family function.

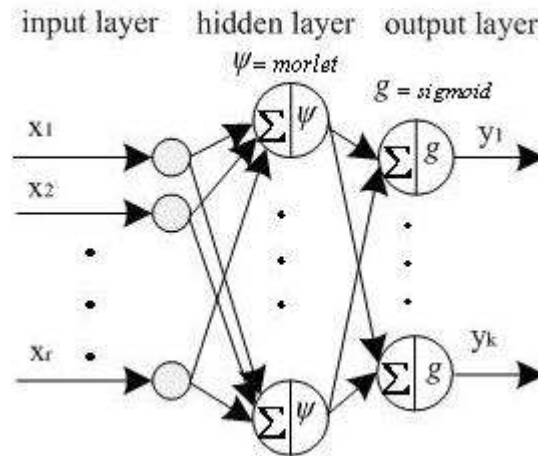


Figure 53: General structure of WNN [181]

Just like in the conventional neural network case, when a wavelet family is used as an activator in a dynamic network, the network is called a Dynamic Wavelet Neural Network (DWNN).

5.2.2.2 Gaussian process

The Gaussian process is a stochastic process which has the following formal definition: “A Gaussian process is a collection of random variables, any finite number of which have (consistent) joint Gaussian distribution” [182]. Training in a Gaussian process involves maximizing the logarithmic likelihood that the predicted process matches the training data [183]. The Gaussian process has several advantages:

- Its models can be used to formulate a Bayesian framework for regression
- It can handle non-linearity
- It can be easier to work with in practice than ANN

However, it has disadvantages as well, as it assumes a “Gaussian distribution”, an assumption which is not valid for a typical power plant operation because periodic shut-downs and turn-ons of a gas turbine [9] are a regular occurrence, which keeps its operation from being a continuous process.

5.3 Implementation of deterministic modeling for estimation of residual life

Based on the literature review and an analysis of the regression techniques (RSE, ANN, and the Gaussian Process), the artificial neural network appears to be the most suitable non-parametric regression technique for the modeling and simulation necessary for the remaining useful life estimation. However, for the sake of completeness, both the RSE (parametric regression) and ANN (non-parametric regression) techniques will be investigated.

5.3.1 Residual life modeling using RSE

Once the RSE is selected as the parametric regression technique, the modeling and simulation tool is built. The inputs will be the system health characteristics (health indexes) and the system operating conditions at the time the first failure precursor is detected and validated, while the responses will be the interval of time between the precursor detection point and the system failure point. A diagram of the conceptual M&S for residual estimation using RSE is shown in Figure 54.

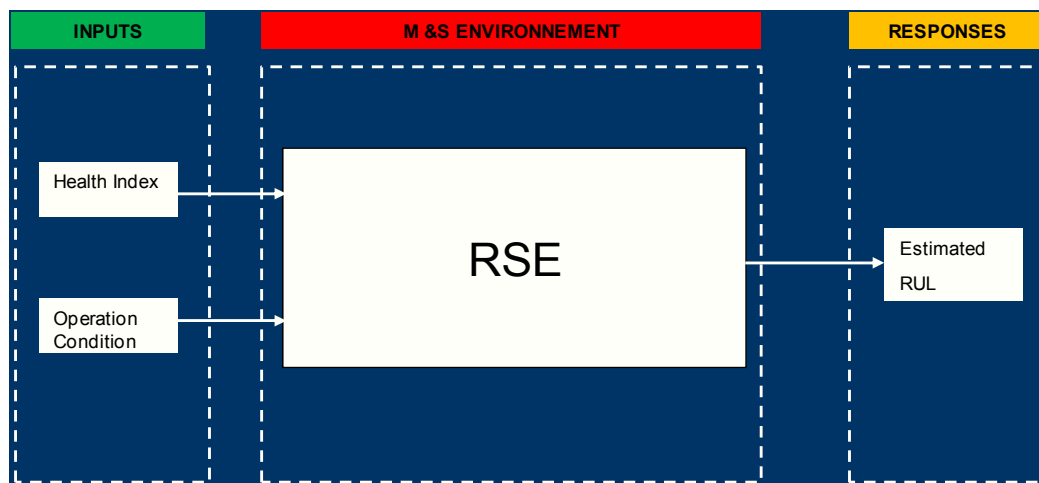


Figure 54: Schematic of conceptual M&S for residual life estimation using RSE

As indicated above, even though the RSE can be a polynomial of any order, it is typically of the second order.

The equation that will be used to approximate the target output is the second-order Taylor series approximation of the following form:

$$R = b_o + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k b_{ij} x_i x_j \quad (37)$$

The estimated coefficients of the response equation will be determined via an Analysis of Variance (ANOVA) approach, in which an appropriate Design of Experiments table is utilized to analyze the different variable settings. The statistical software package of the JMP by SAS® will be used in conjunction with other software in order to achieve the residual life modeling using the RSE regression method.

5.3.2 Non-parametric RUL estimation

In a similar fashion, a modeling and simulation tool is built for the residual life estimation using artificial neural networks as the regression technique. The ANN will be trained using a training data set collected while the item is healthy. The inputs will be the system health characteristics (health indexes) and the system operating conditions at the time the first failure precursor is detected and validated while the responses will be the interval of time between the precursor detection point and the system failure point. A diagram of the conceptual modeling and simulation for residual life estimation using ANN is shown in Figure 55.

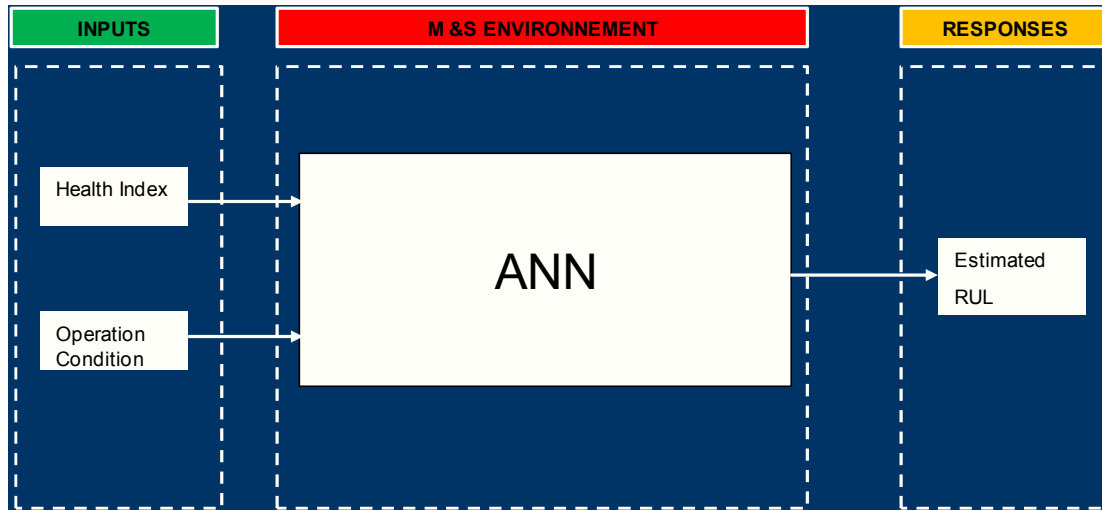


Figure 55: Schematic of conceptual M&S for residual life estimation using ANN

Because the neural network approach has been so commonly used, there exists much customized, ready-made software developed by organizations to meet their specific needs. Thus, for the ANN use, a software called BRAINN [170], developed in the Aerospace Systems Design Laboratory (ASDL) at the Georgia Institute of Technology, will be used. The software BRAINN is capable of creating regression formulas for a set of input-output pairs under certain conditions. It can handle the two types of neural networks for different response types [170], which are: function approximation (used for continuous response, highly non-linear) and classification (used for discrete responses). The sigmoid transfer function defined below in equation

(38) is one of the most commonly used in the neural network, because it can approximate any continuous function so long as the number of neurons in the hidden layer is sufficient and the weights of the neuron connection are adjusted. Therefore, the sigmoid transfer function is utilized by BRAINN.

$$S(z) = \frac{1}{1 + e^{-z}} \quad (38)$$

Then, the sigmoid transfer function is used to compute the characteristic of each value for hidden nodes using equation (39)

below:

$$H_j = S\left(a_j + \sum_{i=1}^N (b_{ij} X_i)\right) \quad (39)$$

Where: a_j is the intercept term for the j^{th} hidden node.

b_{ij} is the coefficient for the i^{th} design variable.

X_i is the value of the i^{th} design variable.

N is the number of input variables.

H_j is the value of the j^{th} hidden node.

If the regression problem is for a function approximation, the value of the response is computed using equation (40)

$$R_k = e_k + \sum_{j=1}^{N_H} (f_{jk} H_j) \quad (40)$$

Where: e_k is the intercept term for the k^{th} response.

f_{jk} is the coefficient for the j^{th} hidden node and k^{th} response

N_H is the number of hidden nodes

Otherwise, if the regression problem is a classification problem, the value of the response is computed using equation (41).

$$O_k = S\left[c_k + \sum_{j=1}^{N_H} d_{jk} \left(S\left(a_j + \sum_{i=1}^N (b_{ij} X_i)\right)\right)\right] \quad (41)$$

Finally, it must be kept in mind that the function approximation requires no post-processing, while the classification needs post-processing.

5.3.3 WAVELET Neural Network (WNN)

Though the wavelet neural network is not applied in this thesis, it has become a good alternative when dealing with systems for which the conventional activation function does not work well. This is because the conventional back-propagation (BP) training that has been the most used in practice has low learning speed and easily falls into local minima as reported in [181]. The Wavelet NN handles some of those deficiencies by using wavelet function as the activation function of the neurons. Also, one major advantage of the WNN is that for systems that are dynamic (time dependent) in nature, the dynamic wavelet network provides much better results than does the conventional dynamic neural network. A conceptual representation of the Dynamic Wavelet Neural Network (DWNN) is shown in Figure 56.

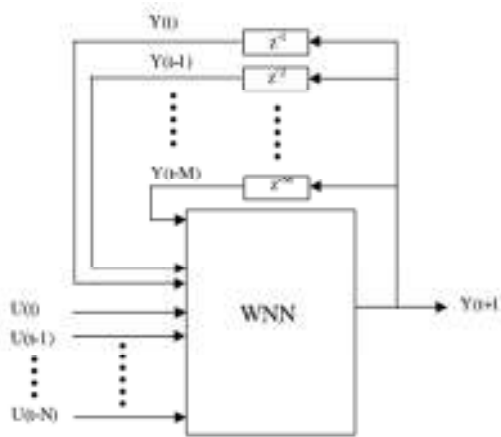


Figure 56: Conceptual representation of Dynamic Wavelet Neural Network [46]

After the models for both the parametric and non-parametric regressions are created, they will be assessed to see if they provide an acceptable way to estimate the residual life of the system. In the case that they are deemed acceptable based on the accuracy of the prediction, the two type of regression will be compared in order to decide which of the regression techniques yields a better prediction model.

On the other hand, if the two regression techniques fail to provide acceptable prognostication estimation, either an empirical cumulative distribution function or the Kaplan-Meier data analysis technique will be used to establish the distribution of the collected residual time to failure data. In effect, all time experimental and empirical failure data are available, and so a priori specific parametric distribution is assumed. The non-parametric and Kaplan-Meier are appropriate approaches to estimating the survival function. Accordingly, the next section is committed to an overview of the Kaplan-Meier method.

5.4 Empirical or non-parametric approach to residual life estimation: Kaplan-Meier

The Kaplan-Meier estimator is one of the most popular statistical data analysis methods for many reasons. In fact, the merit of the Kaplan–Meier estimator of the survival function (also known as reliability) is that it is an empirical or non-parametric method of estimating the survival probability for either non-censored or right-censored data. Because of its ability to handle right-censored data, the Kaplan-Meier is used in this work. This is because in the field currently under study, when an anomaly is detected in a timely fashion, the system’s operator must automatically shut it down to avoid the expensive risk associated with catastrophic failure.

5.4.1 Censored data

Before continuing to an overview of the Kaplan-Meier estimator, let us briefly review the notion of censoring data. Censoring can be defined as the fact that the life data of an observed item or system is not complete. In other words, the residual time to the item failure is not known because the observation was stopped. Censoring occurs predominantly in the engineering and medical fields.

There are many types of censoring.

- Right-censored: an observation is right-censored if it is known that the failure time occurred after time t , but the actual time is not known. Most survival time data are right-censored, because when the decision is made to stop the operation of an item before failure occurs, the cut off is made at the right side [184]. Right-censoring is common in medical applications.
- Left-censored: a survival time is left-censored when at a given time t , it is known that the failure has already occurred, but the actual failure time is not known. In other words, a data point is below a certain value, but it is unknown by how much.
- Type-I censored: data is type-I censored when n items being tested are stopped at a predefined fixed time t_0 , at which point any subjects remaining are right-censored. An application of type-I censoring is conventional preventive maintenance, where, after a predefined lifetime, any remaining operating components are taken off and replaced.
- Type-II censored: data is Type II censored when right-censoring is performed on the remaining number of items $(n-k)$ from a set of n items (original number of tests at start time) after a predetermined number of items $(k \leq n)$ have failed.
- Interval-censored: data is interval-censored when a failure time T cannot be observed, and the only known information is that the survival time is between X and Y ($X < Y$) [185].
- Random (or non-informative) censored: data is random-censored when each item has a censoring time C_i that is statistically independent of its failure time T_i . The observed value $Y_i = \min\{C_i, T_i\}$; any item whose failure time is greater than its censoring time is right-censored.

Based on all the definitions of the different types of censoring, we can see that the survival time data modeled in this thesis belongs to the random censoring category, because each test item is either right censored or it failed. Also, it is worth noting the following points regarding the items or units being considered in this work: 1) they have different anomaly detection times (calendar times); 2) the censoring occurs because a failure signature was detected, the unit was shut down, and the anomaly was fixed. An illustrative concept of the censored survival data is shown in Figure 57

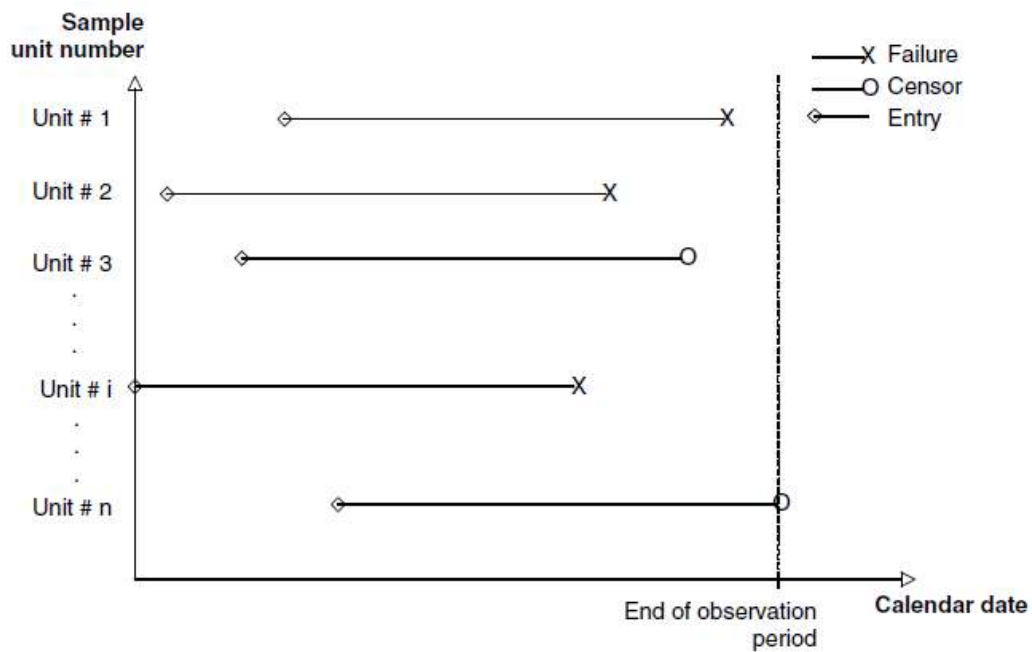


Figure 57: Censored data with different beginning of experiment time [186]

It is important to handle a censored data set with greater care, because deriving a survival function from censored data is not straightforward [186]. Therefore, since the Kaplan-Meier has been used successfully with the same data type, it is an appropriate technique.

5.4.2 Kaplan Meier overview

Edward Kaplan and Paul Meier developed the non-parametric estimation for right-censored lifetime data analysis [157]. It is crucial to have an acceptable level of

confidence in the estimation of the residual life or survival time of critical components because over-estimating the survival time could lead to catastrophic failure. For the given fleet of systems for which the RUL estimation was developed using a non-parametric model, the Kaplan-Meier estimator can be utilized to assess the likelihood of that estimation [112]. The goal of the Kaplan-Meier estimator is to evaluate the maximum likelihood of an estimated survival time. In general, the Kaplan-Meier's plot is dynamic and can be updated with both the right-censored data (units whose survival time was truncated before failure and therefore they did not fail) as well as units that failed. Once a Kaplan-Meier plot is established, the probability that a unit will survive through its estimated RUL can be easily determined, based on the historical behavior of similar units in the fleet. The Kaplan-Meier plot is used for a population with similar characteristics (e.g. for the same type of diseases in the medical field).

5.4.2.1 Estimation of the survivor function

Before proceeding to some mathematical background of the Kaplan-Meier, let's make a few assumptions:

- It is assumed that the time to failure of the different units (n total) after an anomaly is detected can be sorted and rearranged in an ascending fashion:

$$t_1 < t_2 < \dots < t_{n-1} < t_n \quad (42)$$

- The residual time to failure (RTTF) is assumed to be independent and identically distributed.
- It is assumed that there are no ties in the data set.

Therefore, the probability of survival or the empirical survivor function (ESF) can be estimated as follows:

- 1) When there are no censored observations,

$$\hat{p}_i = \frac{n_i - 1}{n_i} \quad (43)$$

Where:

n_i is the number of units at risk just prior to t_i .

\hat{p}_i is the probability of exactly one item failing right after time t_i .

Equation (43) is defined by Kaplan and Meier as the conditional probability of surviving just past time t_i ; mathematically that is written:

$$\hat{p}_i \text{ estimate of } P(T_F > t_i + \delta t | T_F > t_i) \quad (44)$$

Where δt is an arbitrary small time interval in which no censoring occurred.

Recall the conditional probability for two independent events A and B:

$$P(A \cap B) = P(A) \cdot P(B | A) = P(A) \cdot P(B) \quad (45)$$

Now, if the definition of conditional probability is applied to the time t_i ,

where $t_1 < t_2 < \dots < t_i < \dots < t_n$

$$\text{then: } P(T_F > t_i) = P(T_F > \delta t) \cdot P(T_F > t_1 + \delta t | T_F > t_1) \cdot P(T_F > t_2 + \delta t | T_F > t_2) \cdot \dots \cdot P(T_F > t_i + \delta t | T_F > t_i) \quad (46)$$

Recall that the survival function or the reliability function is the probability in equation (46).

Then, by replacing each factor in equation (46) with its expression in equation (43), we find the survivor function

that is the Kaplan-Meier estimate:

$$\hat{S}(t) = \prod_{i, t_i \leq t} \hat{p}_i = \prod_{i, t_i \leq t} \frac{n_i - 1}{n_i} \quad (47)$$

In other word, the Kaplan-Meier estimate is

$$\hat{S}(t) = \frac{\text{Number of observations} \geq t}{n} \quad (48)$$

, for $t \geq 0$

Thus, equation (48) represents a step function that decreases by $1/n$ just after each observed lifetime [187], as illustrated in Figure 58.

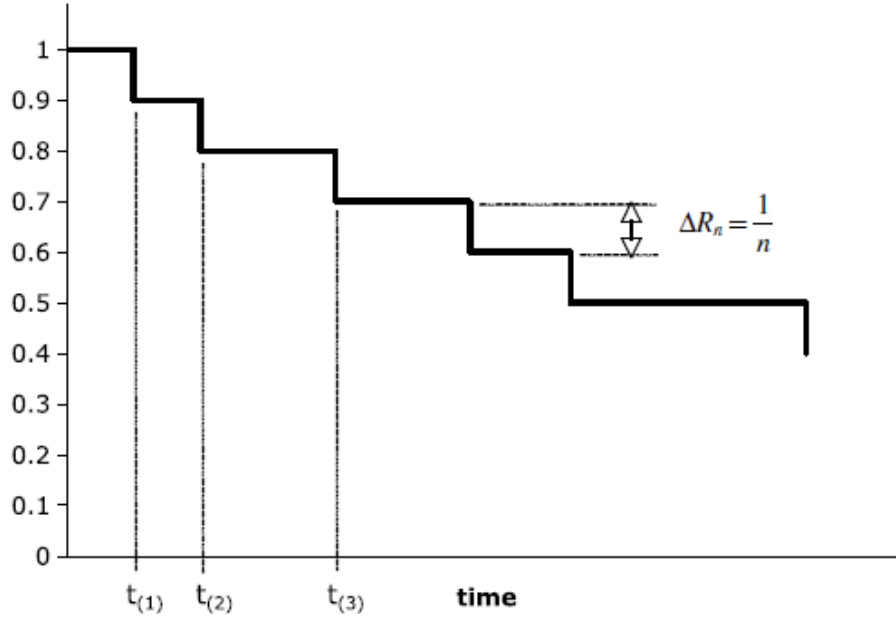


Figure 58: Illustration of empirical survivor function with no censored observations [186]

However, if there are d lifetimes equal to t , then the step function decreases by d/n just past time t .

2) When there are censored observations with d_i failed items at the time t_i , the

Kaplan-Meier estimate becomes:

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i} \quad (49)$$

Where

$\hat{S}(t)$ is the estimated function curve.

n_i is the number of units at risk just prior to t_i .

d_i is the number of units failed at t_i .

An illustrative example showing the Kaplan-Meier plot for censored survival data is shown in Figure 59

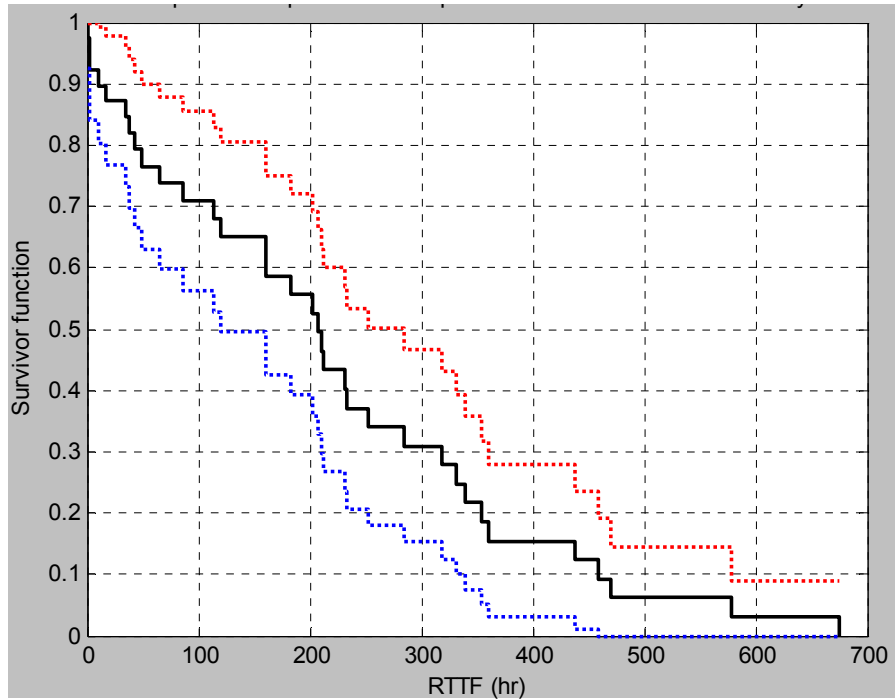


Figure 59: Illustrative example of Kaplan-Meier plot for censored data

It is important to note that in the case of non-censored observations, the Kaplan-Meier estimator reduces to a regular Maximum Likelihood Estimate (MLE). The difference between the Kaplan-Meier and the MLE is that when there is censored observation, the Kaplan-Meier takes the weight which would have been normally assigned to that observation and distributes it evenly among all observed values to the right of that observation [112].

5.4.2.2 Estimation of Kaplan-Meier Confidence Interval

Any rigorous scientific approach requires that, when one is dealing with an approximation or estimation, one must provide the confidence level of that estimation. Since the Kaplan-Meier estimator is simply an estimation of the actual survivor function, a confidence level calculation is necessary.

The problem is that, in the case of right-censoring, the variance does not have a simple form when the Kaplan-Meier estimator is used. Fortunately, a formula developed by Greenwood, and known as Greenwood's formula, can be used to determine the $(1 - \alpha)$ confidence interval [112].

Recall that the survival function is $S = 1 - F$ where F is the cumulative distribution function. Therefore, under the right-censoring, the confidence interval (upper and lower) becomes:

$$S(t_i) \pm Z_{\alpha/2} \cdot \hat{\sigma}(t_i)$$

$$\text{Where } \hat{\sigma}^2(t_i) = \hat{\sigma}^2(S(t_i)) = S(t_i)^2 \sum_{j \leq i} \frac{d_j}{n_j(n_j - d_j)} \quad (50)$$

A 95% confidence interval can be determined as shown in equation (51):

$$S_{95\%}(t_i) = \hat{S}(t_i) \pm 1.96\sigma(t_i) \quad (51)$$

Although this is a good confidence interval calculation method, it is important to note that the confidence interval is a pointwise confidence interval based on fixed values of time t in the CDF $F(t)$ [112].

5.5 **Parametric analysis for assumed distributions**

In cases where the regression model does not yield a satisfactory residual life estimation model, and the Kaplan-Meier provides only a non-parametric empirical survivor estimator, it is interesting to investigate conventional survival data distribution. This is because, although the non-parametric estimator is not constrained to any

particular predetermined lifetime distribution, it is not flexible enough to be used for another purpose. A parametric analysis, on the other hand, allows for the use of a finding from a given data set and for its implementation on other units within a given fleet of units.

However, because there are no previous studies on the type of residual time to failure being undertaken in this thesis, nor is there any knowledge of the underlying distribution form of that residual time to failure, few of the most commonly used distributions in the field of survival function parametric analysis will be investigated.

5.5.1 Exponential Distribution

5.5.1.1 Exponential distribution overview

The first distribution to be considered when there is no previous knowledge is exponential distribution, because it is simple and can give a first level of understanding. The probability density function (PDF) of an exponential distribution is defined as:

$$f(t, \lambda) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (52),$$

where λ is the hazard function.

The major assumption of the exponential function is that the hazard function or hazard rate is constant over time and can be defined as:

$$h(t) = \lambda(t) = \lambda \quad (53)$$

For non censoring cases with n units, the maximum likelihood estimate (MLE) of the hazard function can be determined as follows:

One assumes that n units X with identical independent distribution (iid),

Sample $X = \{X_1, X_2, \dots, X_n\}$. The maximum likelihood estimate is:

$$L(\lambda) = \prod_{i=1}^n \lambda \cdot \exp(-\lambda x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) = \lambda^n \exp(-\lambda \cdot n \cdot \bar{x}) \quad (54)$$

Where
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (55)$$

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{d}{d\lambda} (n \ln(\lambda) - \lambda \cdot n \cdot \bar{x}) = \frac{n}{\lambda} - n \cdot \bar{x} = \begin{cases} > 0, \text{ if } 0 < \lambda < 1/\bar{x} \\ = 0, \text{ if } \lambda = 1/\bar{x} \\ < 0, \text{ if } \lambda > 1/\bar{x} \end{cases} \quad (56)$$

Since the MLE is reached when the derivative vanishes, the MLE is:

$$\hat{\lambda} = \frac{1}{\bar{x}} \quad (57)$$

Recall that \bar{x} is the average of the n units X . Therefore, \bar{x} is also the mean time to failure (MTTF) in the context of time to failure, T_F being the variable of interest.

$$\hat{\lambda} = \frac{1}{MTTF} \quad (58)$$

However, since there are censored units among the data to be analyzed, the use of $\hat{\lambda}$ as determined in equation (58) is not appropriate.

Instead, the approach will assume the general form of an exponential distribution. When the exponential plot of the Kaplan-Meier estimated survival data is complete, the best

linear least squares fit to the data set is found, and the $\hat{\lambda}_{estimated}$ from the curve fit is

deduced. If the exponential distribution assumption was a good one, then the $\hat{\lambda}_{estimated}$ will

be close to the $\hat{\lambda}$ calculated from the maximum likelihood analysis. To plot the survivor or reliability function, one must then perform an exponentially distributed function:

$$S(t) = e^{-\lambda t} \quad (59)$$

Taking the natural logarithm of both sides of equation (59) yields the following:

$$\ln(S(t)) = \ln(e^{-\lambda t}) \quad (60)$$

Then, the equation (60) becomes:

$$\ln(S(t)) = -\lambda t \quad (61)$$

The following variable changes are then made:

$$x = t \quad (62)$$

$$y = -\lambda x \quad (63)$$

It is important to note that the intercept in equation (63) is 0 [188].

Equation (63) shows the best linear least squares fit of the exponential plot of the Kaplan-Meier estimated survivor function. When this is determined, the goodness of fit can be estimated. The coefficient of determination, R^2 , will be used as a metric for goodness of fit.

5.5.1.2 Exponential Confidence Interval

Depending on the size of the data set, the confidence level can vary. Also, it is obvious that the confidence level for a small data set will be lower than that for a larger data set. The $100(1 - \alpha)\%$ confidence interval can be calculated for exponential distribution estimation as:

$$\frac{1}{\hat{\lambda}} \cdot \frac{2n}{\chi_{2n; \alpha/2}^2} < \frac{1}{\lambda} < \frac{1}{\hat{\lambda}} \cdot \frac{2n}{\chi_{2n; 1-\alpha/2}^2} \quad (64)$$

Where $\hat{\lambda}$ is the MLE estimate, λ is the true value of the parameter, and $\chi_{k;x}^2$ is the value of the chi squared distribution with k degrees of freedom, which gives x cumulative probability. Fortunately, the confidence interval can be readily obtained using a Matlab® built-in function `expfit`.

5.5.2 Weibull Distribution

5.5.2.1 Weibull distribution overview

The next distribution to be considered is the Weibull distribution, because it is one of the two most frequently used distributions in the reliability and survival analysis field. Additionally, exponential distribution has a constant failure rate or hazard function, while Weibull has a time-dependent hazard rate.

The Weibull distribution has the following probability density function:

$$f(t, \theta, \beta) = \begin{cases} \lambda(t) e^{-\left(\frac{t}{\theta}\right)^\beta}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (65)$$

Where:

θ is defined as the scale parameter (expressed in units of time).

β is the shape parameter (dimensionless).

The hazard function is defined as:

$$\lambda(t, \theta, \beta) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1}, \text{ with } \theta > 0, \beta > 0, \text{ and } t \geq 0 \quad (66)$$

The survivor function or reliability is defined as:

$$S(t) = e^{-(t/\theta)^\beta} \quad (67)$$

To get the Weibull plot, one must take a logarithm of both side of equation (67), as follows:

$$\ln(S(t)) = \ln\left(e^{-(t/\theta)^\beta}\right) \quad (68)$$

Then, equation (68) becomes equation (69):

$$\ln(S(t)) = -\left(\frac{t}{\theta}\right)^\beta \quad (69)$$

By taking the logarithm of both sides of the equation (69) one more time, it becomes:

$$\ln[-\ln(S(t))] = \beta \ln(t) - \beta \ln(\theta) \quad (70)$$

Then, by making the following variable changes, shown in equation (71),

$$\begin{cases} x = \ln(t) \\ y = \beta \cdot x - \beta \cdot \ln(\theta) \end{cases} \quad (71)$$

Equation (70) becomes:

$$y = \beta x - \beta \ln(\theta) \quad (72)$$

In equation (72), which is the equation of a straight line, β represents the slope, and θ can be deduced from the intercept of equation (72) using the following equation:

$$\theta = e^{\frac{\text{intercept}}{\beta}} \quad (73)$$

Therefore, the Weibull distribution estimate that corresponds to the Kaplan-Meier estimate, or the RTTF, can be obtained.

Similarly, equation (72) represents the best linear least squares fit of the Weibull distribution plot of the Kaplan-Meier estimated survivor function. Once this is determined, the goodness of fit can be estimated. The coefficient of determination, R^2 , will be used as a metric for the goodness of fit.

5.5.2.2 Weibull Confidence Interval

As in the exponential distribution case, the quality of the Weibull distribution estimation is dependent on the size of the data set being studied. Consequently, the 100 $(1 - \alpha)\%$ confidence interval needs to be calculated for the Weibull distribution estimation. The determination of the confidence interval for the Weibull estimate is not trivial. The two sides $(1 - \alpha)$ of the scale and shape parameters can be calculated as

$$\theta = \exp(\mu) \text{ and } \beta = 1/\sigma, \text{ where } \mu \text{ is the extreme value location parameter and}$$

σ is the extreme value scale parameter.

Then, the lower and upper bounds of the Weibull parameters can be obtained, as shown in equations (74) and (75):

$$[\theta_L, \theta_U] = [\exp(\mu_L), \exp(\mu_U)] \quad (74)$$

$$[\beta_L, \beta_U] = [1/\sigma_U, 1/\sigma_L] \quad (75)$$

Likely, the confidence interval can be readily obtained using a MATLAB® built-in function, “wblfit” in MATLAB version 7.8.0 (R2009).

CHAPTER 6

IMPLEMENTATION OF THE PROPOSED METHOD FOR GAS TURBINE COMPRESSOR CATASTROPHIC FAILURE

This chapter concerns the application of the proposed methodology, as described in chapter 4 and chapter 5, to a fleet of eleven failed or right-censored 170MW rated heavy duty gas turbines, in order to demonstrate the viability of the proposed methodology to efficiently handle compressor failure problems. Each unit in the considered fleet has either failed or has displayed failure signatures due to a compressor issue.

The failure precursor detection methodology explained in chapter 4 is first applied to each unit in order to identify the time of the precursory failure events. The characteristic signature of each of those failure precursors is computed and tabulated. Then, the accumulated survival data is analyzed using deterministic regression techniques, non-parametric techniques, and parametric techniques, which produces models for the residual time to failure, as elucidated in the chapter 5. In short, this chapter is divided into three parts: 1) implementation of catastrophic failure precursor detection on the eleven units of a fleet; 2) modeling of the residual time to failure, as obtained from the failure anomaly characteristics; and 3) discussion of the findings and the key challenges of the proposed methodology and its implementation.

6.1 Implementation of failure precursor detection

The proposed approach for failure precursor detection is applied to General Electric 7FA gas turbine compressor failures. There were a fleet of eleven units that were monitored and analyzed.

6.1.1 Assumptions made for proposed approach implementation

It is important to enumerate a list of implicit assumptions made when conducting this implementation:

- 1) The test units are not picked in any particular order.
- 2) The failure precursors independent and identically distributed (iid), which means there is no distinction made between the type and mode of failure.
- 3) Each unit is a test subject.
- 4) Units are operated in a similar manner (mostly at based-load operation).
- 5) Units' owners perform the appropriate scheduled maintenance.
- 6) Monitoring sensors are performing as planned.

6.1.2 Inspection background of General Electric gas turbine

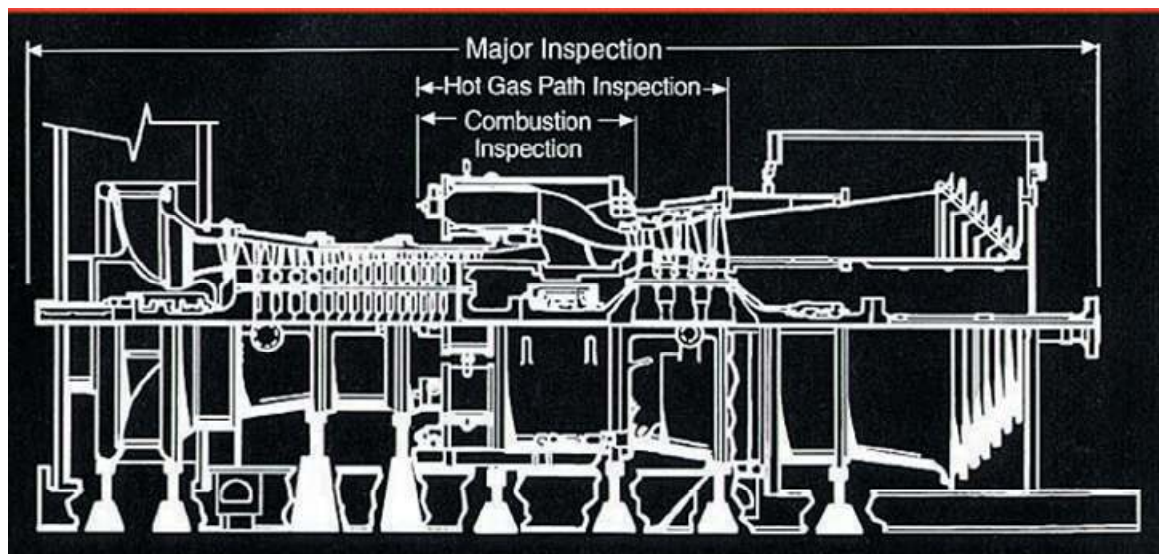


Figure 60: GE gas turbine inspection scope [9]

GE has different inspection scopes for its gas turbines: combustion inspection, hot gas path inspection, and major inspection. Each inspection covers specific parts of the gas turbine, as shown in Figure 60.

Each category of inspection is performed at predefined intervals of time. The combustion inspection is the most frequently done, the hot gas inspection is done the second most frequently, and the major inspection is the least frequently done. Table 23 below shows the recommended inspection intervals for the GE 7FA gas turbine, given the following inferences:

- Combustion inspection should be done every 8000 factored hours (machine running time) or 450 factored starts, whichever occurs first.
- Hot gas path inspection is recommended every 24000 factored hours or 900 factored starts, whichever occurs first.
- Major inspection is recommended every 48000 factored hours or 2400 factored starts, whichever occurs first.

Table 23: GE technology specific baseline recommended inspection intervals [9]

Type of Inspection	Combustion System	Factored Hours / Factored Starts										
		MS3002K	MS5001PA/ MS5002C,D	MS6B	MS7E/EA	MS9E	MS6FA	MS7F/FAFA+	MS7FA+e	MS9F/FAFA+	MS9FA+e	MS7FB
Combustion	Non-DLN	24,000/400	12,000/800 ⁽¹⁾⁽²⁾	12,000/1,200 ⁽²⁾⁽³⁾	8,000/900 ⁽³⁾	8,000/900 ⁽³⁾	-	-	-	-	-	-
	DLN	-	8,000/400	12,000/450	12,000/450	12,000/450	8,000/450	8,000/450	12,000/450	8,000/450	8,000/450	8,000/450
Hot Gas Path		24,000/1,200	Eliminated/1,200	24,000/1,200	24,000/1,200	24,000/900	24,000/900	24,000/900	24,000/900	24,000/900	24,000/900	24,000/900
Major		48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400	48,000/2,400

Factors That Can Reduce Maintenance Intervals

- Fuel
- Load Setting
- Steam/water injection
- Peak Load TF Operation
- Trips
- Start Cycle
- Hardware Design

(1) Units with Lean Head End liners have a 400 starts combustion inspection interval.
 (2) Machines with 6581 and 6BeV combustion hardware have a 12,000/600 combustion inspection interval.
 (3) Multiple Non-DLN configurations exist (Standard, MNQC, IGCC). The most limiting case is shown, however different quoting limits may exist on a machine and hardware basis. Contact a GE Energy representative for further information.

NOTE: Factored Hours/Starts intervals include an allowance for nominal trip maintenance factor effects.
 Hours/Starts intervals for Major Inspection are quoted in Actual Hours and Actual Starts.

As one can observe from Table 23, the compressor, which is normally covered under a major inspection, could develop anomalies that might lead to component failure during a major inspection interval of 48000 factored hours or 2400 factored starts.

Therefore, the proposed approach could help detect compressor failure precursors.

6.1.3 Failure precursor detection process applied to a gas turbine compressor for data collection

Each of the test units is monitored using several sensors. However, because the compressor is the subsystem of interest, only sensors that are affected by the compressor operation are selected. The locations of the monitored sensors are shown in Figure 61. Based on the answer to a previous research question, the sensors measurements are recorded at a one minute sampling time. Because all eleven units will go through the same process, the steps of the proposed failure precursor detection method are applied to the first test unit, as seen below. The application of the failure detection to the other ten units can be seen in **Appendix A**.

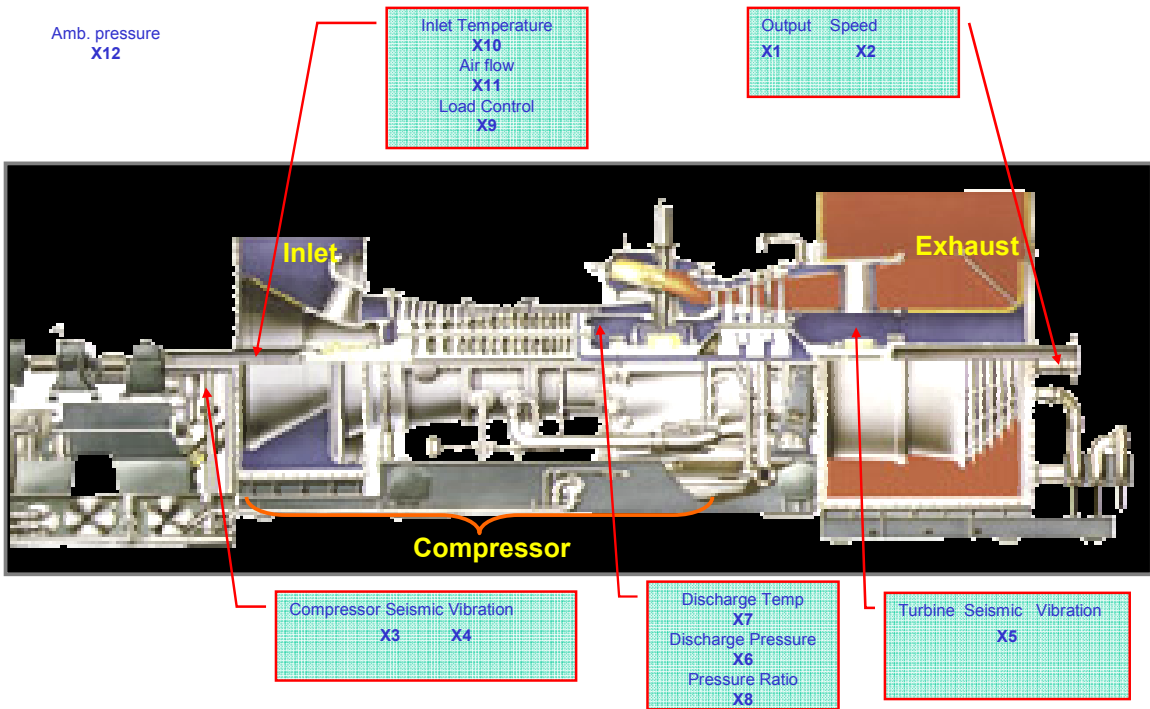


Figure 61: Layout of monitored sensors for gas turbine compressor anomaly detection

6.1.3.1 Backgrounds of the failed units

Test unit 1 had a hot gas path inspection in July 2005. The unit had 34638 factored fired hours (much less than the recommended 48000 FFH for the next major inspection) and 140 factored starts (much less than the recommended 2400 FS required for the next major inspection). In other words, the machine was not close to it required major inspection (i.e. machine has not gone through any major inspection) at the time of the compressor failure. The machine had a compressor catastrophic failure on July 4th 2009 at 2:55 AM.

6.1.3.2 Steps of the anomaly detection methodology

6.1.3.2.1 *Step 1: Query data*

The sensors that measure the health of the gas turbine compressor, the overall health of the gas turbine, and the operation conditions are identified and summarized in Table 24.

Table 24: Monitored Gas turbine compressor sensors

Variable	Variables Description
X1	Overall system health parameter 1
X2	Overall system health parameter 1
X3	Compressor seismic vibration 2
X4	Compressor seismic vibration 2
X5	Turbine seismic vibration 3
X6	Compressor health parameter 1
X7	Compressor health parameter 2
X8	Compressor effectiveness health parameter 1
X9	operating condition (load)
X10	operation condition 1 (environment)
X11	Compressor effectiveness health parameter 2
X12	operation condition 2 (environment)

After the sensors are identified, a query is made to the stored database for information from the sensors of interest at a one minute sampling interval from the period of 6/01/2009 at 12:00 AM to 7/04/2009 at 3:00 AM, soon after the time of failure.

6.1.3.2.2 Step 2: Pre-processing

At the pre-processing step, the raw data is normalized using the mean value of each variable. The normalized sensor readings are within the same value range. It is observed that the normalized data set has the same characteristics as the raw data, and the normalized data has a mean value of 1 for all the sensors.

Then, the data set is filtered to eliminate outliers.

By observing the distribution of the twelve pre-processed sensors, shown in Figure 62, it can be noticed that the distributions are almost normally distributed and centered on the value 1, with the exception of the operating conditions that have a uniform distribution, which indicate that the analysis is done during constant operating conditions.

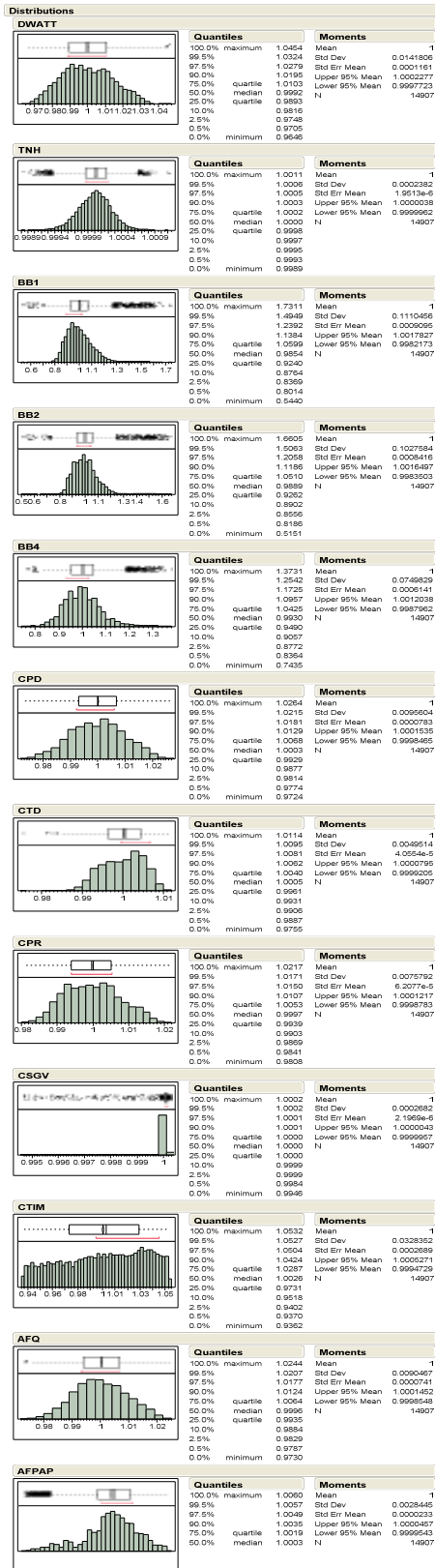


Figure 62: Distribution of the 12 sensors pre-processed

6.1.3.2.3 Step 3: Signal de-noising using the DWPT

At this step, each of the twelve normalized sensor data sets is de-noised using the discrete wavelet packet transforms (DWPT) and the Daubechies 10 Wavelet.

As an illustrative example, Figure 63 shows the noisy and noise-free sensor X3 signals, which will be used in the subsequent steps, with a signal to noise ratio of 13.445 decibels

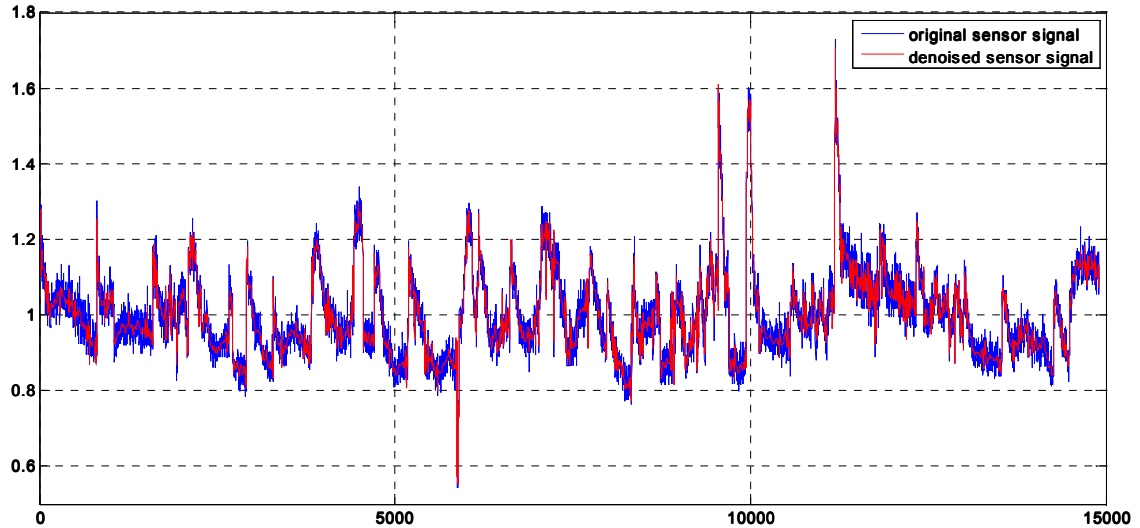


Figure 63: Noisy and de-noised sensor X3

6.1.3.2.4 Step 4: Multi-resolution Analysis

At the MRA step, each sensor signal is decomposed into a three-level tree using the DWPT with the “Daubechies 7”, which leads to eight different wavelet components (1 approximation, and 7 details) as shown on the tree decomposition of the Figure 64:

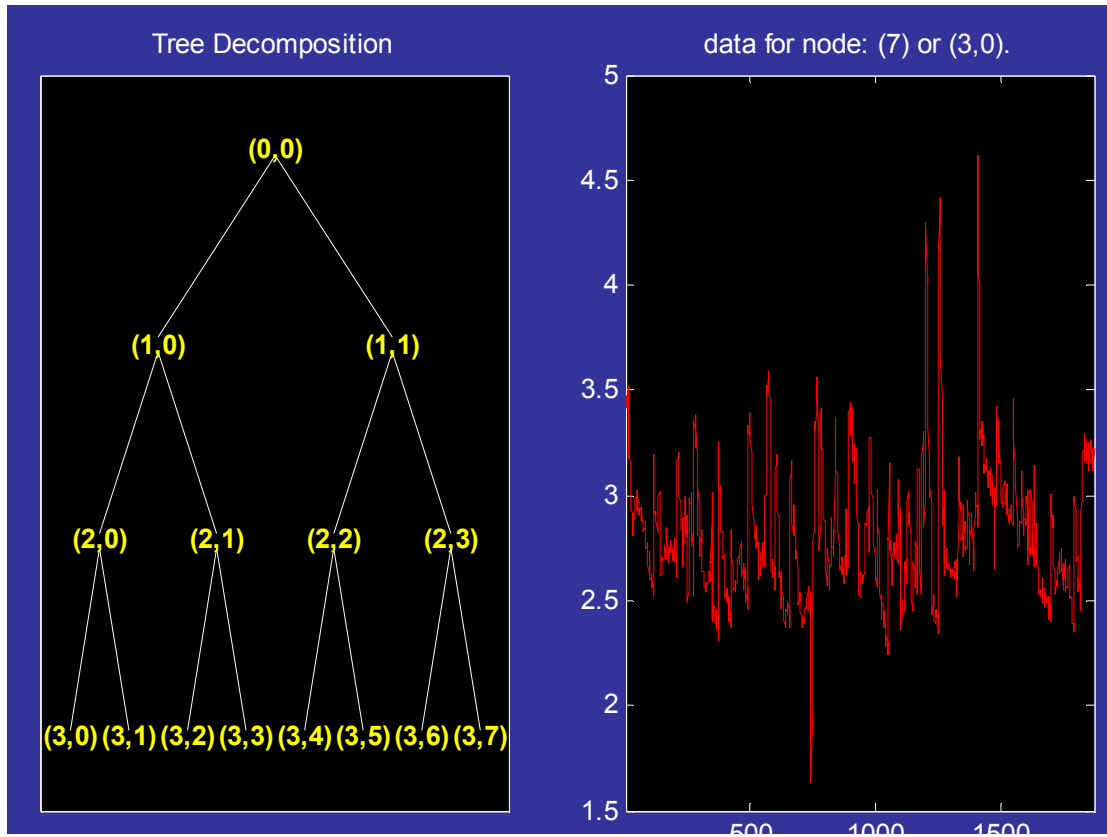


Figure 64: Tree decomposition of sensor X3 signal with the wavelet component at node 7

Figure 65 below shows the original de-noised signal and each of the eight wavelet components (original de-noised sensor signal shown in black, approximation in red, and the details in blue).

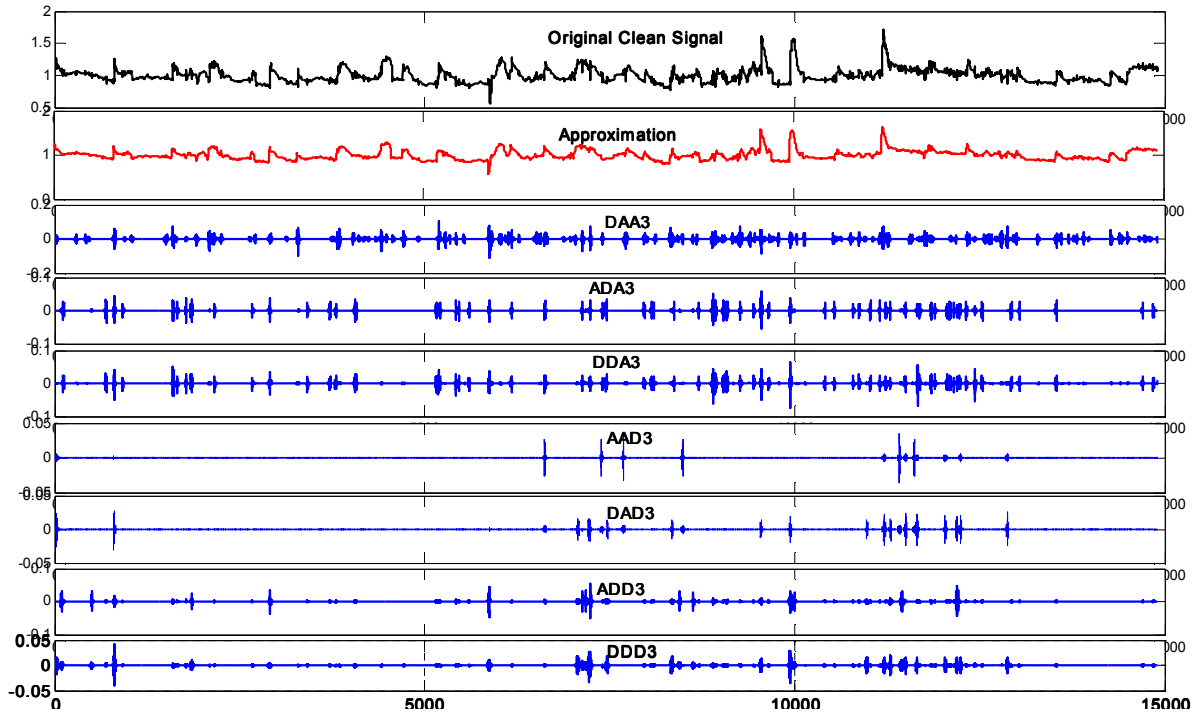


Figure 65: DWPT at 3-level MRA of sensor X3

Then, the energy content at each node of level 3 is evaluated and grouped into two groups: approximation (node 7 or (3,0)) and details (sum of the details of node 9 through 14). This is shown in Table 25 below.

Table 25: Node energy content of each sensor at level-3 DWPT decomposition

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	100	100	99.983	99.988	99.99	100	100	100	100	99.999	100	100
Sum of details (% of Original)	2.33E-04	1.49E-06	0.0165	0.011548	0.010464	5.93E-05	2.72E-05	5.10E-05	4.86E-06	7.59E-04	3.33E-05	1.92E-06

For each of the twelve sensors, the energy contained within the approximation node is over 99.9% of the total energy of the original signal. Therefore, the approximation signal (or scale function) will be used as the representative of the actual signal in the subsequent steps. That is, the energy content in each of the wavelet components will serve as the signal feature characteristic.

6.1.3.2.5 Step 5: PPCA data fusion

First, the standard PCA is performed on the retained signal feature (approximation) to determine the principal components (PC).

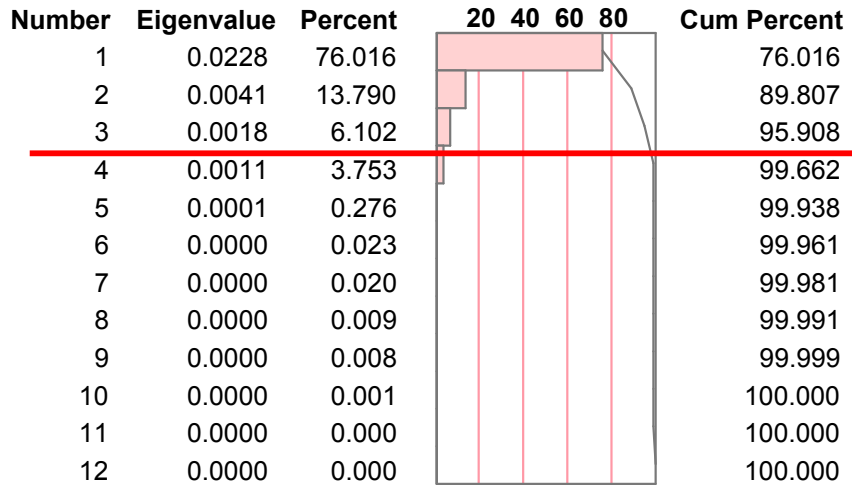
The covariance matrix of the processed twelve variables is calculated, as shown in Table 26.

Table 26: Covariance Matrix for PCA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	0.00020	-0.00000	-0.00000	0.00021	-0.00015	0.00013	-0.00006	0.00010	-0.00000	-0.00038	0.00012	0.00002
X2	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000
X3	-0.00000	-0.00000	0.01233	0.00982	0.00362	-0.00002	0.00003	-0.00004	0.00000	-0.00011	-0.00000	0.00002
X4	0.00021	-0.00000	0.00982	0.01056	0.00365	0.00010	-0.00003	0.00006	-0.00000	-0.00048	0.00012	0.00003
X5	-0.00015	-0.00000	0.00362	0.00365	0.00562	-0.00015	0.00000	-0.00010	0.00000	-0.00029	-0.00010	-0.00004
X6	0.00013	-0.00000	-0.00002	0.00010	-0.00015	0.00009	-0.00004	0.00007	-0.00000	-0.00024	0.00008	0.00002
X7	-0.00006	0.00000	0.00003	-0.00003	0.00000	-0.00004	0.00002	-0.00003	0.00000	0.00015	-0.00003	-0.00000
X8	0.00010	0.00000	-0.00004	0.00006	-0.00010	0.00007	-0.00003	0.00006	-0.00000	-0.00021	0.00006	0.00001
X9	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000
X10	-0.00038	0.00000	-0.00011	-0.00048	-0.00029	-0.00024	0.00015	-0.00021	0.00000	0.00108	-0.00023	-0.00001
X11	0.00012	-0.00000	-0.00000	0.00012	-0.00010	0.00008	-0.00003	0.00006	-0.00000	-0.00023	0.00008	0.00002
X12	0.00002	-0.00000	0.00002	0.00003	-0.00004	0.00002	-0.00000	0.00001	-0.00000	-0.00001	0.00002	0.00001

Then, the eigenvalues with their respective contributions and the corresponding eigenvectors for the covariance matrix are obtained and ranked from highest value to lowest value, as shown in Table 27.

Table 27: Pareto chart of eigenvalues contributions



To retain at least 95% of the information content (i.e. 95% of the confidence level, which is shown with a red bar), only the first three eigenvalues representing 95.908%, are needed. Consequently, the eigenvectors corresponding to those three eigenvalues, which represent the principal components, are chosen, as shown in the Table 28:

Table 28: Eigenvectors or weight matrix for three principal components

Variables	PC1 (76.01%)	PC2 (13.79%)	PC3 (6.10%)
X1	0.00441	-0.04005	0.23462
X2	-0.00016	-0.00003	-0.00048
X3	0.70473	-0.27312	-0.52123
X4	0.64907	-0.12295	0.58098
X5	0.2856	0.95059	-0.07837
X6	0.00059	-0.03636	0.14895
X7	0	-0.00058	-0.07432
X8	-0.00039	-0.02236	0.1228
X9	0.00001	0.00013	-0.00033
X10	-0.02169	-0.04877	-0.51587
X11	0.00232	-0.02716	0.14269
X12	0.00094	-0.01109	0.018

The entries of the eigenvectors are the weights that will be applied to their corresponding variables in order to obtain the principal component; that is, the combination of the twelve variables.

At this point, the Probabilistic Principal Component Analysis (PPCA) procedure will be implemented. First, the major weight contributors are identified. This establishes the maximum weight likelihood; that is, any entry value of an eigenvector less than 0.1 is set to 0, while any value greater than 0.1 is kept. Consequently, the following principal components are obtained for the PPCA and shown in Table 29.

Table 29: Maximum likelihood weight for PPCA principal components

Variables	PC1 (76.01%)	PC2 (13.79%)	PC3 (6.10%)
X1	0	0	0.23462
X2	0	0	0
X3	0.70473	-0.27312	-0.52123
X4	0.64907	-0.12295	0.58098
X5	0.2856	0.95059	0
X6	0	0	0.14895
X7	0	0	0
X8	0	0	0.1228
X9	0	0	0
X10	0	0	-0.51587
X11	0	0	0.14269
X12	0	0	0

As can be noticed from Table 29, the major effects occur along the vibration sensors (X1,X2, and X3) for the two heavily contributing principal components, while the performance parameters have some impact as well along the third component.

Next, the remaining PPCA parameters are computed (i.e. the isentropic noise covariance, the prediction error unique to response, the data matrix and the variance of reduced

dimension). An algorithm that will implement the PPCA steps is created using the MATLAB® software, and the signal along the three principal components is obtained.

The three signals obtained from the PPCA process are shown in Figure 66

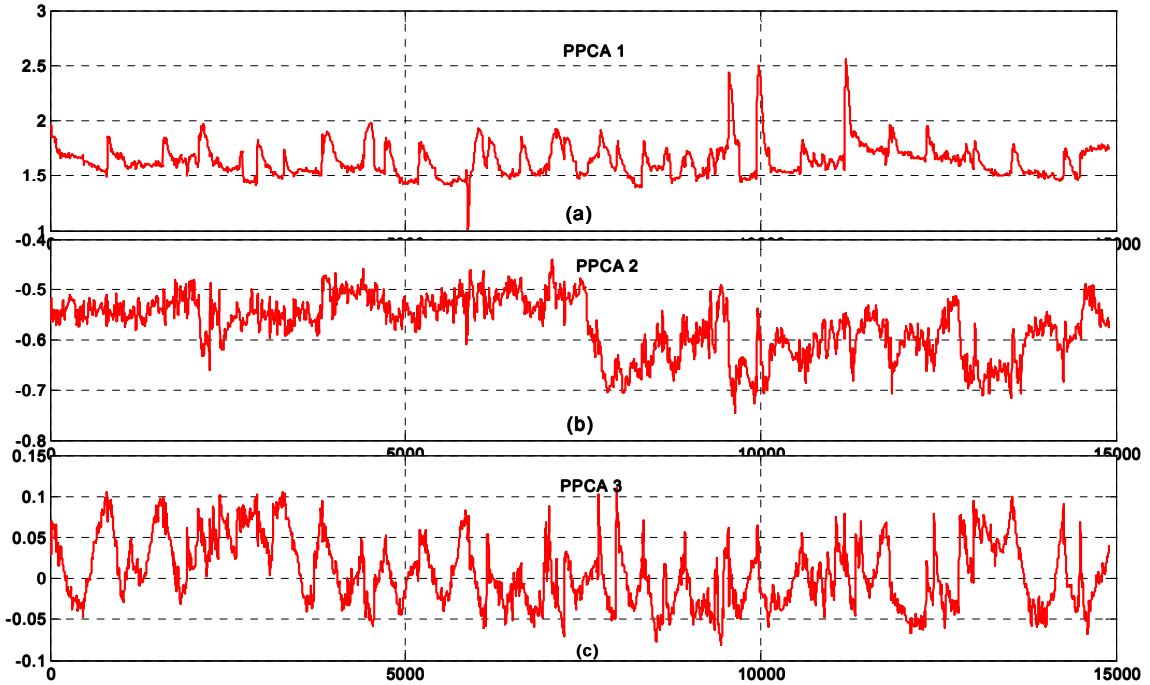


Figure 66: Signal obtained from PPCA process

6.1.3.2.6 Step 6: Anomaly detection decision

Anomaly detection decision is a multi-stage statistical process. The first stage is the reconstruction of a one-dimensional signal, followed by the threshold calculation and so on.

- Computation of reconstructed signal

Since the three most important principal components (shown in Figure 66) are kept, the reconstructed one-dimensional signal is obtained using a linear combination:

$$RS(t) = \lambda_1 \Phi_1^*(t) + \lambda_2 \Phi_2^*(t) + \lambda_3 \Phi_3^*(t) \quad (76)$$

With $\lambda_1=76.01\%$, $\lambda_2=13.79\%$ and $\lambda_3=6.10\%$. These percentages are the total information content in each of the three major eigenvalues. Figure 67 is the reconstruction of the three principal components, which are, in themselves, the representation of the original twelve sensor signals.

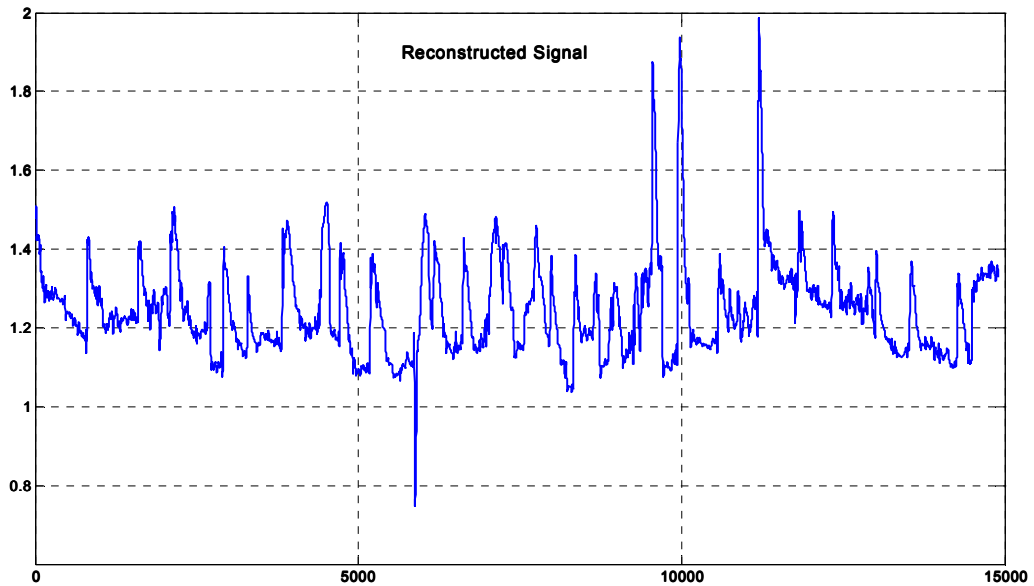


Figure 67: Reconstructed 1-dimensional Signal

- Threshold calculation

To compute the damage indicators SAD and SSD, E_{ref} needs to be established. To that end, the following steps are taken:

- 1) The reconstructed signal is decomposed using DWPT and the Daubechies 6 (db6) wavelet.

- 2) The reference energy, E_{ref} , is calculated as the average energy content of the reconstructed signal for the first ten hours of healthy operation. Thus, a value of $E_{ref}=1.2943$ is obtained.
- 3) Next, to calculate the threshold a converged value of the statistical parameters (μ_{SAD}, σ_{SAD} and μ_{SSD}, μ_{SSD}) need to be determined. The convergence criteria are strictly enforced; that is, the parameter values need to remain static for three hours of operation for each parameter, and then converge after more than eighty hours of operation. Figure 68 shows the point at which all four parameters converge.

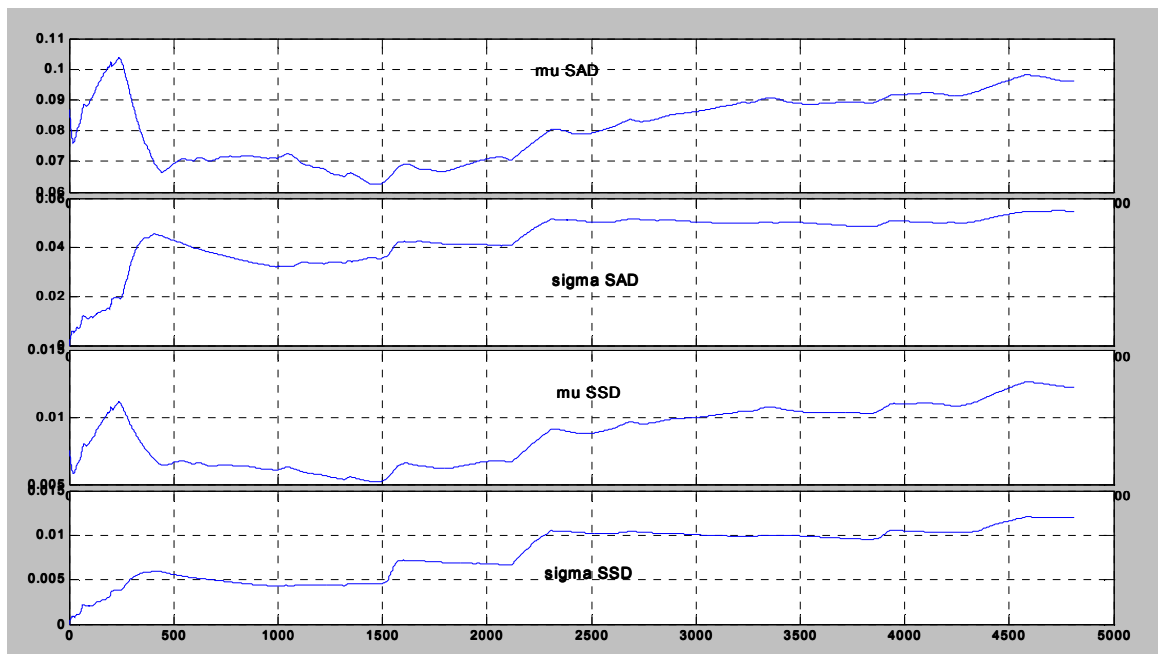


Figure 68: Statistical parameters for SAD and SSD

Then, the modified statistical parameters are obtained and used to calculate the threshold. A value of $\alpha = 0.02$ for a confidence level of $100*(1- \alpha)$ (i.e. corresponding to 98% confidence level) is assumed, and monitored until the failure occurs.

That corresponds to the SAD threshold values: a UL_{SAD}^α of 0.22664 and a UL_{SSD}^α of 0.051105 for the SAD and SSD respectively.

- 4) After the threshold value is set, the statistical process control kicks in, where the damage indicators are monitored over time. Figure 69 below is a graph presenting the monitoring of the damage indicator SAD, and Figure 70 shows the data from the damage indicator SSD. In both figures, the red dash is a visual representation of the threshold, while the signal is in red.

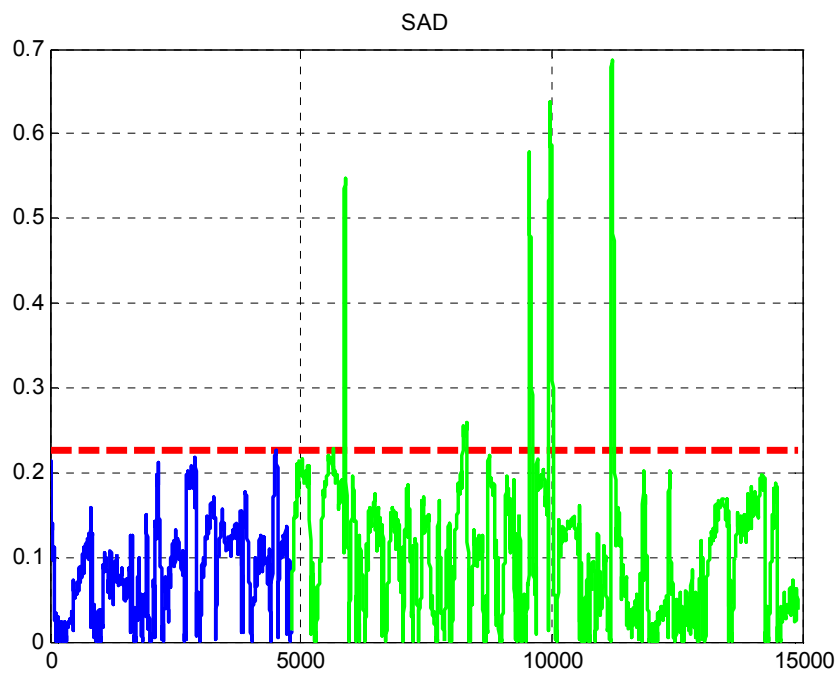


Figure 69: Monitoring of damage indicator SAD

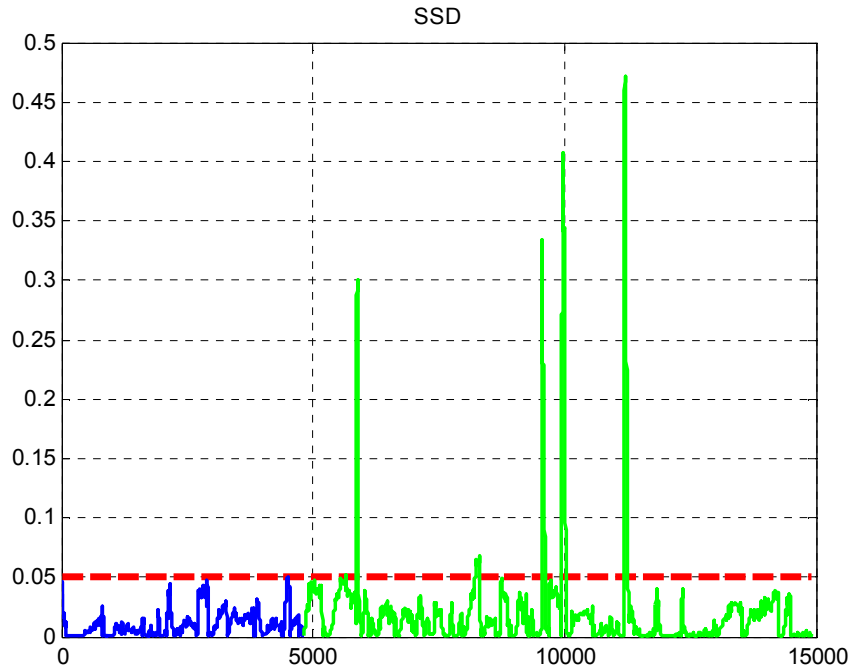


Figure 70: Monitoring of damage indicator SSD

Both Figure 69 and Figure 70 show the threshold, magnitude length of the anomalies.

5) Bayesian Hypothesis Testing

Finally, Bayesian hypothesis testing is performed using the following two equations:

$$SAD(t) \leq 0.22664, H(t) = 1 \quad (77)$$

$$SAD(t) > 0.22664, H(t) = 0 \quad (78)$$

The result of the Bayesian hypothesis test is the binary function $H(t)$, with entry values of “1” and “0” obtained and shown in Figure 71. As shown in the Figure 71, there were five abnormal events during the monitoring periods. This should have allowed the gas turbine operator enough time to avoid a catastrophic failure.

In a post-processing analysis, the gas turbine manufacturer established that there was a failure precursory event on 06/17/2009 at 19:45. The proposed approach has successfully identified at least one previous defect event, as early as 06/12/2009 at 23:09. Furthermore, the proposed process has captured three other defects, missed by the gas

turbine manufacturer's procedure, which should have alerted the manufacturer. Each defect was severe, with the final defect being the most severe.

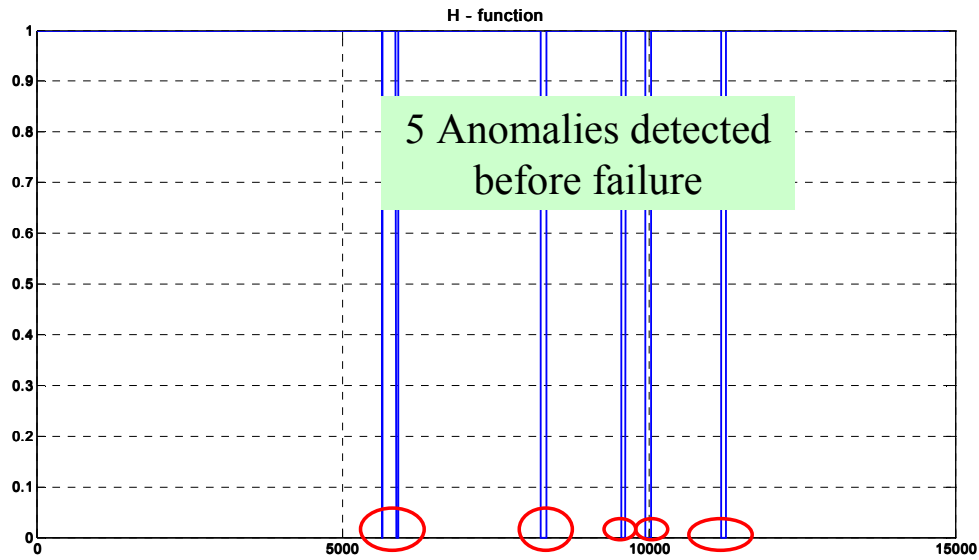


Figure 71: Result of the Bayesian hypothesis testing

Finally, the errors associated with the statistical calculation of the threshold are computed as well. Thus, the probability of a false-positive is an input that is decided by the analysis. It represents 2% in this example. As for the probability of a false-negative or type II error, it is less than $10e-4$ for both damage indicators SAD and SSD, which is very small. The corresponding powers are around 1.0. It can be concluded that type II errors are insignificant compared to type I errors, which is desirable for any statistical process.

6.1.4 Data collection procedure for RTTF modeling

After the implementation of the precursor detection, the failure data needs to be collected in order to model the residual time to failure. Recall that the plan is to map the defect characteristics (the inputs) into the observed length of time (the response) between detection and time of failure.

Thus, using the same test unit 1 as an example, the health characteristics, as defined in the previous chapter, are determined.

The Anomaly Severity Index (ASI) for the SAD damage indicator is computed as:

$$ASI(k) = \frac{SAD(k)}{USL} = \frac{SAD(k)}{0.22664}, \text{ where } k=1 \text{ is the first time step of the}$$

monitoring period (shown as a green curve in Figure 69).

The Anomaly Duration Index (ADI) for the SAD damage indicator is the length of time a defect is continuous. It can be measured using either set of damage indicator figures, or using the results of the Bayesian hypothesis testing figure.

Finally, the response parameter, or the time to failure T_f , is determined for each detected anomaly. The failure time for test unit 1 is illustrated in Figure 72.

The use of both type of figures-Figure 69 for the monitoring of the damage indicator and Figure 71 for the result of Bayesian testing- provides a better visual representation of the anomaly magnitudes (severity and duration). Additionally, the two health indexes can be merged in order to look at as the intensity of the severity.

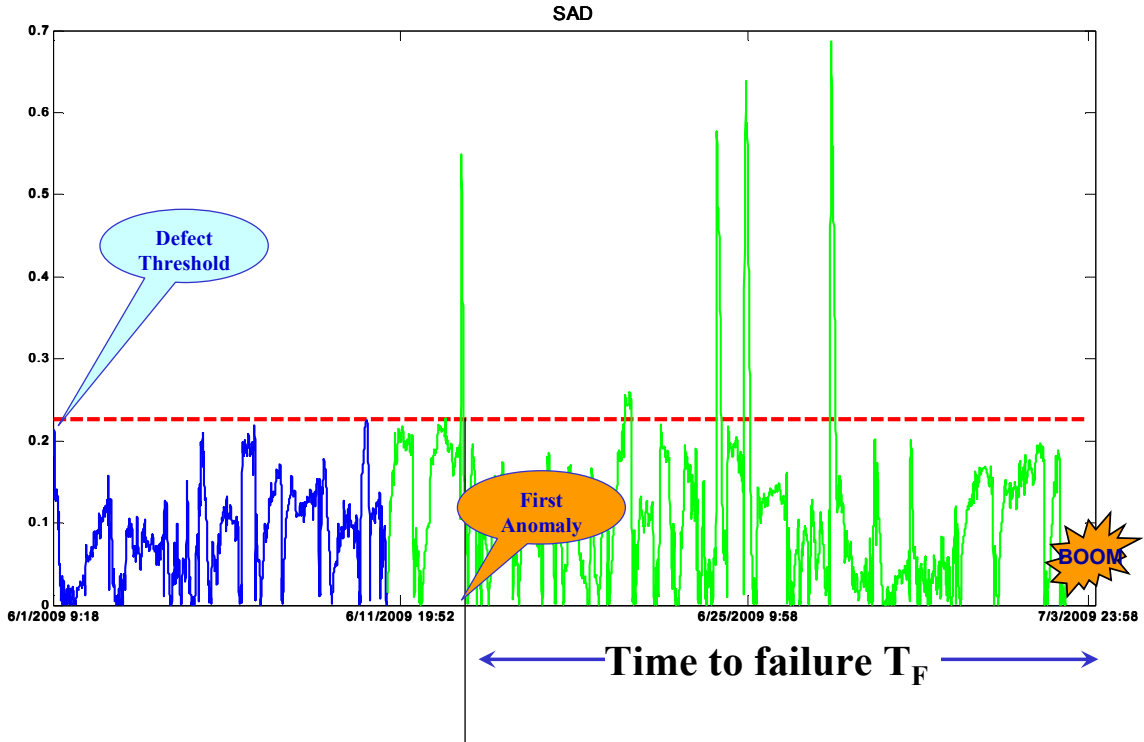


Figure 72: Illustration of the time to failure parameter

The inputs for and response to a detected anomaly are summarized in Table 30. The ASI is dimensionless, while ADI and RTTF are in hours.

Table 30: Summary of first detected anomaly

Anomaly	ASI	ADI	RTTF
1	2.42	0.7	507.76

The process applied to unit 1 to make Table 30, which consists of the first detected anomaly, can be repeated (see Appendix A) for the other ten units in order to produce a table of first anomalies.

If all the detected anomalies were to be considered, an extended data set could be obtained, since there are five anomalies for unit 1, as shown in Table 31.

Table 31: Summary of all detected anomalies

Anomaly	ASI	ADI	RTTF
1	2.42	0.7	507.76
2	1.149	1.583333	342.6667
3	2.582	1.183333	233.0667
4	2.814	1.433333	209.85
5	3.0298	1.366667	161.1667

In a similar way, by considering not only the first anomaly and its real time to failure, but also each subsequent anomaly, the table can be extended as well.

Finally, the table can be altered by doing some tweaking, and disregarding the marginal health indexes in cases where there are many anomalies detected. Practically, this would eliminate the anomalies with small ASI and ADI values (when there are other anomalies detected within the same unit), which would base the prognostic model on the more pronounced health indexes.

By making the proposed changes to the test unit, Table 30, which represents the first anomaly, changes so that the data summarized in Table 32 now represents the first anomaly.

Table 32: Actual first anomaly to be used for PHM model

Anomaly	ASI	ADI	RTTF
3	2.582	1.183333	233.0667

In a similar fashion, Table 31 is changed into and replaced by Table 33, which will be used as the extended data set for prognostication purposes.

Table 33: Actual extended data set representing all anomalies for PHM model

Anomaly	ASI	ADI	RTTF
3	2.582	1.183333	233.0667
4	2.814	1.433333	209.85
5	3.0298	1.366667	161.1667

6.2 Modeling of Residual Time to Failure

After repeating the data generation process for all eleven units, as underlined above, with the health indexes and time to failure (see Appendix A) the data in Table 34 is compiled, representing the failure data for the modified first anomaly detected. However, test units 10 and 11 were censored after failure signatures were identified. Thus, the entries in the column labeled “censored” are either “0,” representing an observed failure, or “1,” representing a right-censored anomaly.

Table 34: First anomaly

Data point	ASI	ADI (hr)	RTTF (hr)	Censored
1	2.548228	1.183333	233.0667	0
2	1.036524	0.216667	332.15	0
3	1.282236	3.6	232.1167	0
4	1.165099	1	339.6833	0
5	1.184042	1.033333	578.2833	0
6	1.062287	0.066667	458.6833	0
7	1.551326	0.833333	35.2	0
8	1.12047	0.183333	119.5167	0
9	1.130492	0.233333	3.183333	0
10	2.335767	2.366667	85.18333	1
11	1.006843	0.033333	124.4	1

6.2.1 Implementation of the deterministic approach to residual lifetime modeling

As presented in the previous chapter, the response surface equation will be used as the parametric regression technique, while the artificial neural network will be used as the non-parametric regression technique.

6.2.1.1 Parametric regression: RSE

For the RSE regression, let us assume a second order polynomial as follows:

$$R = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k b_{ij} x_i x_j \tag{79}$$

In the case of only two inputs, X1 and X2, for ASI and ADI, equation (79) becomes:

$$R = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \tag{80}$$

The goal is to solve for the different regression coefficients $b_0, b_1, b_2, b_{11}, b_{22}, b_{12}$. Then, the JMP software is used to determine the coefficients, as shown in Table 35.

Table 35: RSE fit coefficients

b0	b1	b2	b3	b4	b5
2526.01	-3103.46	675.8822	810.1705	-87.9684	-129.022

Table 36: Contribution of each term and cross terms to the RTTF prediction

Term	Estimate	Std Error	t Ratio	Prob> t
ADI	675.88224	352.0278	1.92	0.1129
ASI	-3103.455	1680.665	-1.85	0.1241
ADI*ADI	-129.022	71.66407	-1.80	0.1317
ASI*ASI	810.1705	450.568	1.80	0.1321
ASI*ADI	-87.96837	168.4539	-0.52	0.6238

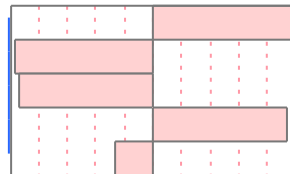


Figure 73 represents the profiler, and shows how RTTF varies with each of the predictors ASI and ADI.

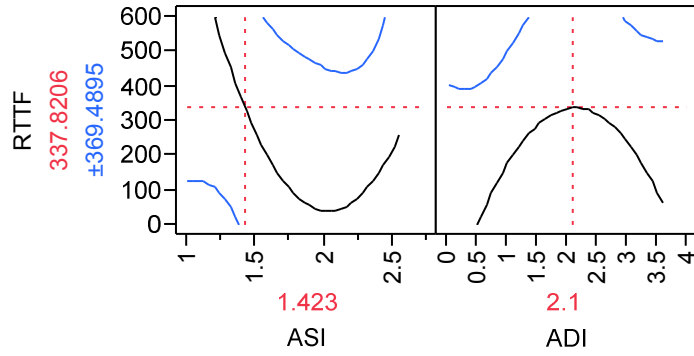


Figure 73: Prediction profiler of RTTF (ASI,ADI) using RSE

The next figures, Figure 74 and Figure 75, are the “actual versus predicted” and “residual versus predicted” plots, respectively. They show the quality of the fit of the eleven point data set regression, performed using the response surface equation.

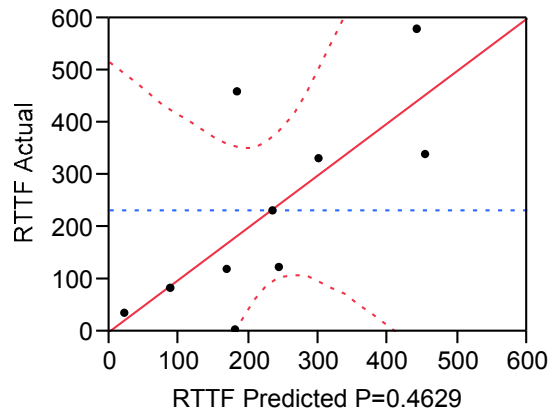


Figure 74: RSE fit Actual vs. Predicted

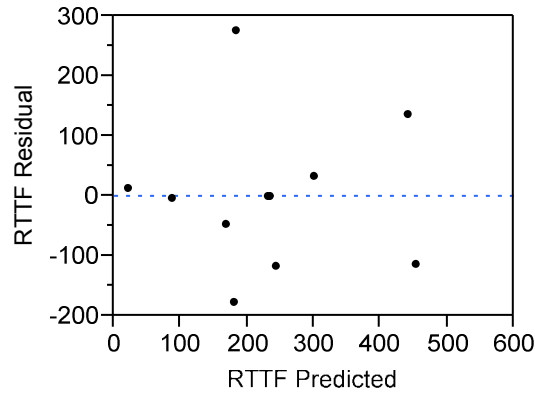


Figure 75: RSE fit Residual vs. Predicted RTTF

The coefficient of determination that corresponds to the fit quality is $R^2=0.52$, which is a poor fit, as can be seen from Figure 74 and Figure 75.

Therefore, it is safe to say that the RSE modeling of the RTTF is not acceptable. Next the neural network, which is a non-parametric regression technique, is used to predict the fit of the RTTF for given pairs of ASI and ADI.

6.2.1.2 Non-parametric regression technique: neural network

The non-parametric regression technique to be implemented is the artificial neural network (ANN). For the prediction of the RTTF response, using ASI and ADI as predictors, a network with one hidden layer, comprising six nodes, is built as shown in Figure 76.

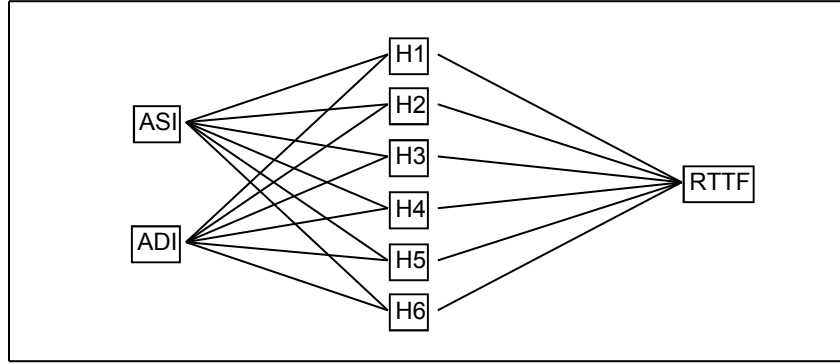


Figure 76: Neural Network for modeling RTTF

Figure 77 represents the profiler that shows how RTTF varies as ASI and ADI change.

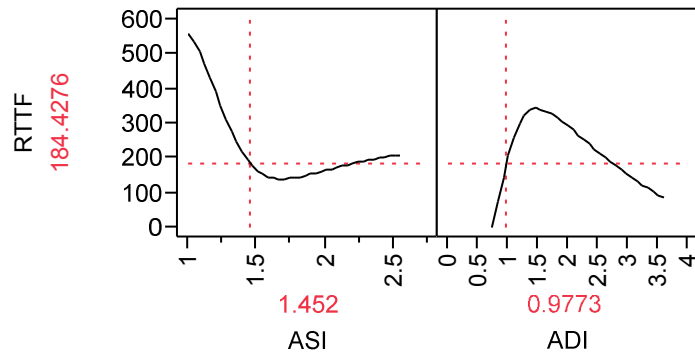


Figure 77: Prediction profiler of RTTF (ASI,ADI) using ANN

Again, the software JMP is used to perform the regression. Figure 78 and Figure 79 show the quality of fit for the eleven point data set regression performed using the artificial neural network.

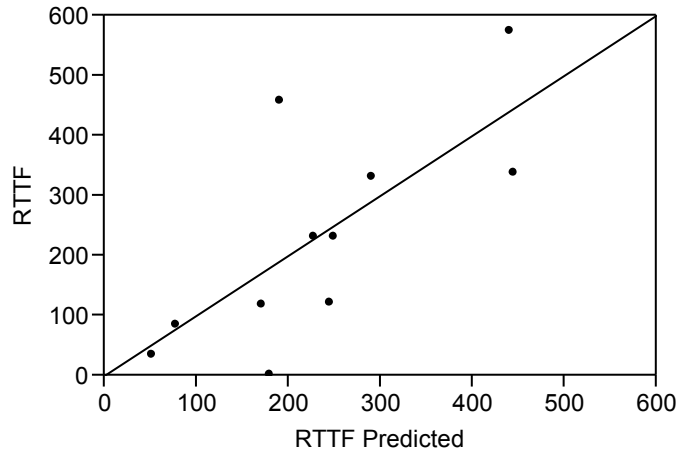


Figure 78: Actual vs. Predicted plot of NN regression

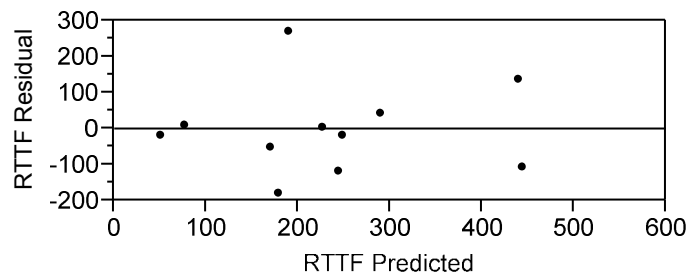


Figure 79: Residual vs. Predicted plot of NN regression

In the case of the ANN, the coefficient of determination that corresponds to the fit quality is $R^2=0.59$. Although slightly better than the regression fit, the fit is still a poor one, as can be seen from Figure 78 and Figure 79. Consequently, the ANN modeling of the RTTF is not acceptable. It must be noted that there was an attempt to utilize other readily available neural network regression tools such as the BRAINN [170] and the MATLAB® GUI “nftool”. Both tools failed to provide adequate fitting results due to the limited size of the data set.

6.2.1.3 Implementation of the extended data set

After the realization that the limited data set size is a constraint when performing acceptable deterministic regressions, the question arises whether or not a larger data size would provide a better fit. For the sole purpose of checking the impact of the limited data set, let us build a table where not just the actual time to failure of the first detected anomaly is considered, but all of the anomalies are included, as described above in Table 33, where the extended data set for test unit 1 is summarized. By adding the extended data for the other ten test units, Table 37 is obtained.

Table 37: Extended data set of ALL the anomalies

Data point	ASI	ADI (hr)	RTTF (hr)	Censored	Data point	ASI	ADI (hr)	RTTF (hr)	Censored
1	2.548228	1.183333	233.0667	0	21	1.165099	1	339.6833	0
2	2.81397	1.433333	209.85	0	22	1.065159	0.6	318.9833	0
3	3.029187	1.366667	161.1667	0	23	1.036482	0.083333	675.4333	0
4	1.036524	0.216667	332.15	0	24	1.184042	1.033333	578.2833	0
5	1.18051	0.166667	469.4833	0	25	3.623064	0.716667	37.31667	0
6	1.132413	0.2	438.5667	0	26	1.062287	0.066667	458.6833	0
7	1.137	0.116667	353.2667	0	27	2.240944	1.016667	1	0
8	1.183305	0.133333	284.4167	0	28	1.106504	0.166667	85.98333	0
9	1.258377	0.366667	252.65	0	29	1.551326	0.833333	35.2	0
10	1.282236	3.6	232.1167	0	30	1.580428	1.066667	11.05	0
11	1.496119	2.133333	212.0167	0	31	1.12047	0.183333	119.5167	0
12	1.246279	3.766667	208.0167	0	32	1.016824	0.05	49.38333	0
13	1.369278	2.266667	201.6333	0	33	1.130492	0.233333	3.183333	0
14	1.252208	2.016667	182.45	0	34	1.18496	2.8	2.783333	0
15	1.151164	2.15	160.9333	0	35	2.335767	2.366667	85.18333	1
16	1.208265	4.65	113.6667	0	36	1.637981	16.38333	78.46667	1
17	1.279143	4.766667	64.81667	0	37	1.626708	22.18333	22.16667	1
18	1.218381	2.283333	41.98333	0	38	1.006843	0.033333	124.4	1
19	1.140144	1.6	17.11667	0	39	1.23587	0.683333	48.36667	1
20	1.359802	0.3	359.9	0					

Let us repeat the deterministic regression techniques using the extended data set.

6.2.1.3.1 Parametric regression of extended data set using RSE

Again using the JMP statistical software package, this time to determine the RSE regression coefficients for the extended data set, Figure 80 and Figure 81 illustrate the quality of the regression.

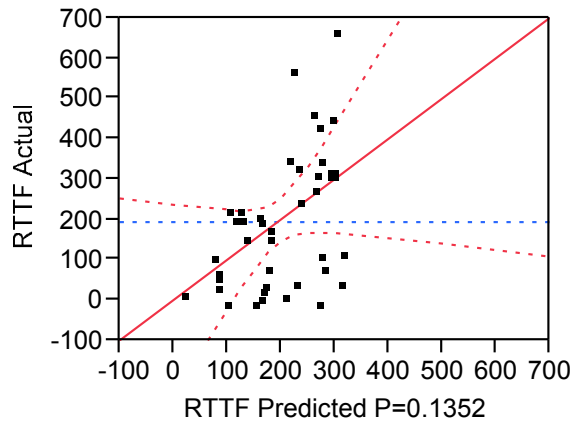


Figure 80: Actual vs. Predicted regression of extended data set using RSE

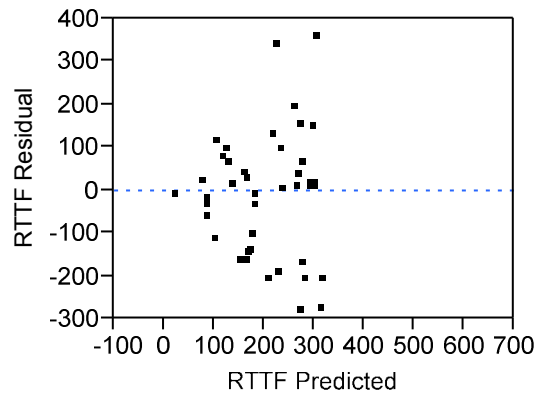


Figure 81: Residual vs. Predicted regression of extended data set using RSE

The corresponding coefficient of determination of the fit is $R^2=0.22$, which is an extremely poor fit of the response using the two inputs ASI and ADI.

6.2.1.3.2 Non parametric regression of extended data set using ANN

Similarly, the JMP statistical software package is used to perform the regression of the extended data set; Figure 82 and Figure 83 illustrate the quality of the regression using the ANN.

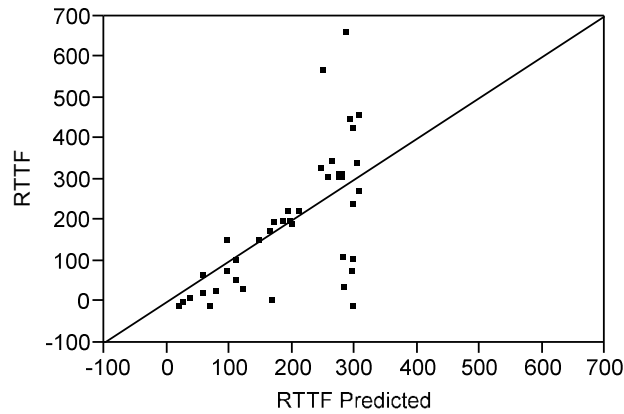


Figure 82: Actual vs. Predicted regression of extended data set using ANN

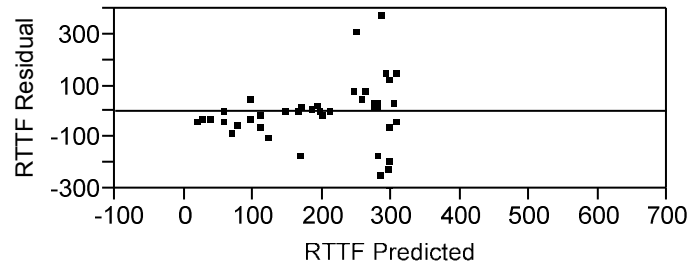


Figure 83: Actual vs. Predicted regression of extended data set using ANN

The corresponding coefficient of determination of the fit is $R^2=0.36$, which is an even worse fit of the response RTTF by the two inputs ASI and ADI than limited data set produced.

However, an implicit assumption was made to build the extended data set: that each anomaly and the subsequent anomalies of a given machine were independent. The

resulting regression is even worse. That means that the assumption of independence is not legitimate. Now, let's assume that an anomaly and its subsequent anomalies are not independent. Thereby, let's define a new predictor that is only valid in the case of the extended data set, called the Number of Previous Anomaly" (NPA). NPA is a metric that gives more weight to an anomaly than to its predecessors if they occur on the same test unit. Table 38 is produced with the addition of the third input (NPA).

Table 38: Extended data set with 3 inputs

Data point	ASI	ADI (hr)	NPA	RTTF (hr)	Data point	ASI	ADI (hr)	NPA	RTTF (hr)
1	2.548228	1.183333	1	233.0667	21	1.165099	1	2	339.6833
2	2.81397	1.433333	2	209.85	22	1.065159	0.6	3	318.9833
3	3.029187	1.366667	3	161.1667	23	1.036482	0.083333	1	675.4333
4	1.036524	0.216667	1	332.15	24	1.184042	1.033333	2	578.2833
5	1.18051	0.166667	1	469.4833	25	3.623064	0.716667	3	37.31667
6	1.132413	0.2	2	438.5667	26	1.062287	0.066667	1	458.6833
7	1.137	0.116667	3	353.2667	27	2.240944	1.016667	2	1
8	1.183305	0.133333	4	284.4167	28	1.106504	0.166667	1	85.98333
9	1.258377	0.366667	5	252.65	29	1.551326	0.833333	2	35.2
10	1.282236	3.6	6	232.1167	30	1.580428	1.066667	3	11.05
11	1.496119	2.133333	7	212.0167	31	1.12047	0.183333	1	119.5167
12	1.246279	3.766667	8	208.0167	32	1.016824	0.05	2	49.38333
13	1.369278	2.266667	9	201.6333	33	1.130492	0.233333	2	3.183333
14	1.252208	2.016667	10	182.45	34	1.18496	2.8	3	2.783333
15	1.151164	2.15	11	160.9333	35	2.335767	2.366667	1	85.18333
16	1.208265	4.65	12	113.6667	36	1.637981	16.38333	2	78.46667
17	1.279143	4.766667	13	64.81667	37	1.626708	22.18333	3	22.16667
18	1.218381	2.283333	14	41.98333	38	1.006843	0.033333	1	124.4
19	1.140144	1.6	15	17.11667	39	1.23587	0.683333	2	48.36667
20	1.359802	0.3	1	359.9					

The entire study with the two main predictors (ASI, ADI) is repeated by adding a third parameter, NPA, to the original two parameters.

6.2.1.3.3 Parametric regression of the extended data set with three inputs using RSE

The JMP statistical software package is used to determine the RSE regression coefficients of the extended data set with the three inputs. Figure 84 shows the contribution of each of the terms of the RSE.

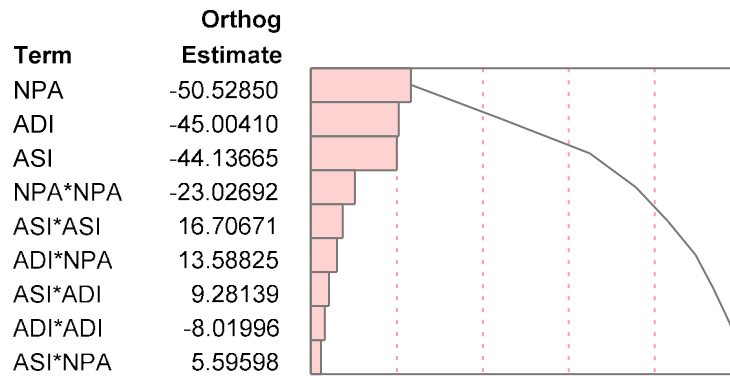


Figure 84: Pareto plot of the terms contribution for 3 inputs

A visual representation of the quality of the extended data set regression using three inputs is shown in Figure 85 and Figure 86.

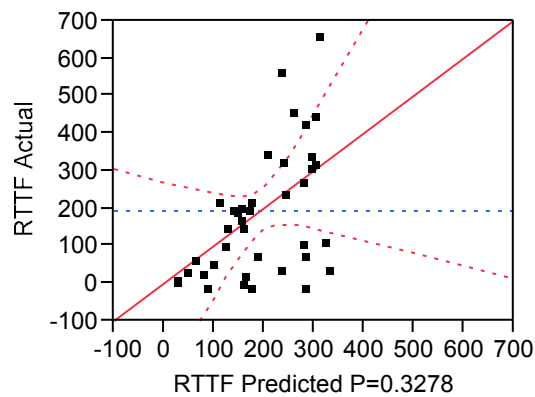


Figure 85: Actual vs. Predicted regression of extended data set with 3 Inputs using RSE

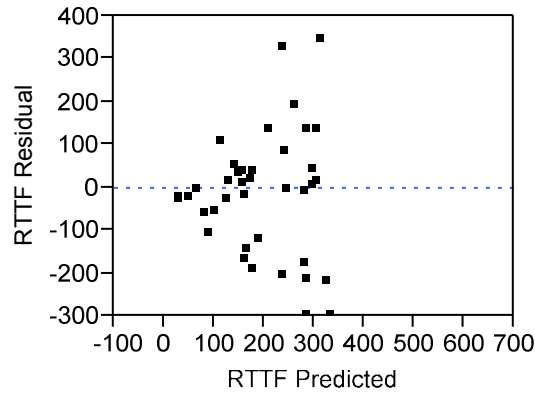


Figure 86: Residual vs. Predicted regression of extended data set with 3 Inputs using RSE

The corresponding coefficient of determination for the regression fit of the extended data set using, the three inputs (ASI, ADI, NPA) is $R^2=0.27$, which is a very poor fit.

6.2.1.3.4 Non-parametric regression of extended data set using ANN

A similar approach is taken for the extended data set with three inputs using the ANN regression. Figure 87 and Figure 88 show the quality of the regression.

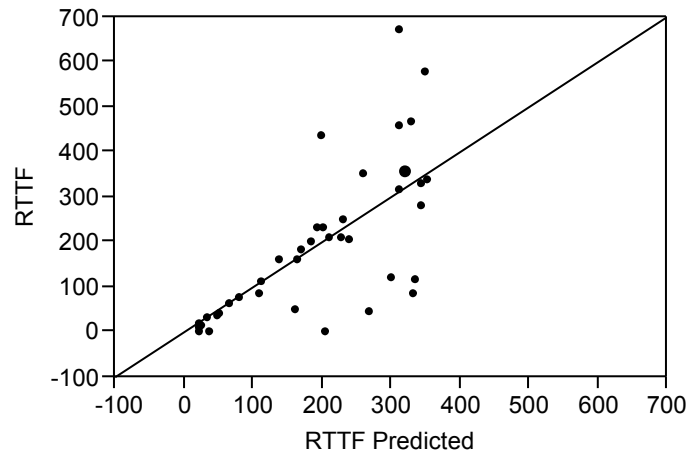


Figure 87: Actual vs. Predicted regression of extended data set with 3 Inputs using ANN

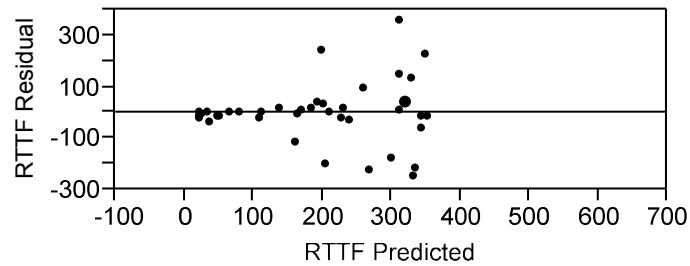


Figure 88: Actual vs. Predicted regression of extended data set with 3 Inputs using ANN

The corresponding coefficient of determination for the regression fit with three inputs using the ANN is $R^2=0.51$, which, though better than the coefficient of determination using two inputs, is still not an acceptable quality of fit.

Ultimately, the deterministic regression approach to the residual time to failure is not appropriate. Although the non-parametric regression (ANN) is slightly better than the parametric (RSE) one, the qualities of the respective fits ironically get worse for both regression techniques when the data size gets larger.

Therefore, as predicted in the previous chapter, the non-parametric data analysis is appropriate for failure data, as it can provide the empirical cumulative distribution of any survival or reliability data.

6.2.2 Implementation of non-parametric data analysis using the Kaplan-Meier estimator

Though the deterministic regression approach not appropriate, the Kaplan-Meier is appropriate to provide the estimate of the survival function plot because the data set is composed of both non-censored and censored data.

Thus, let us compute the Kaplan-Meier estimator with the obtained failure data. The actual failure data is referred to as the table of first anomaly data. The tabular data for the Kaplan-Meier plot of survival time, obtained using the application of the Kaplan-Meier estimation calculation, is shown in Table 39.

Table 39: Tabular data for the Kaplan-Meier plot of residual time after the first detected anomaly

Failure Time (RTTF)	Number failed	Number censored	At Risk	Survival ($S_{\hat{}}$)
0	0	0	11	1
3.183	1	0	11	0.909091
35.2	1	0	10	0.818182
85.183	0	1	9	0.818182
119.517	1	0	8	0.715909
124.4	0	1	7	0.715909
232.117	1	0	6	0.596591
233.067	1	0	5	0.477273
332.15	1	0	4	0.357955
339.683	1	0	3	0.238636
458.683	1	0	2	0.119318
578.283	1	0	1	0

Figure 89 shows the Kaplan-Meier plot of residual time to failure after the first detected anomaly.

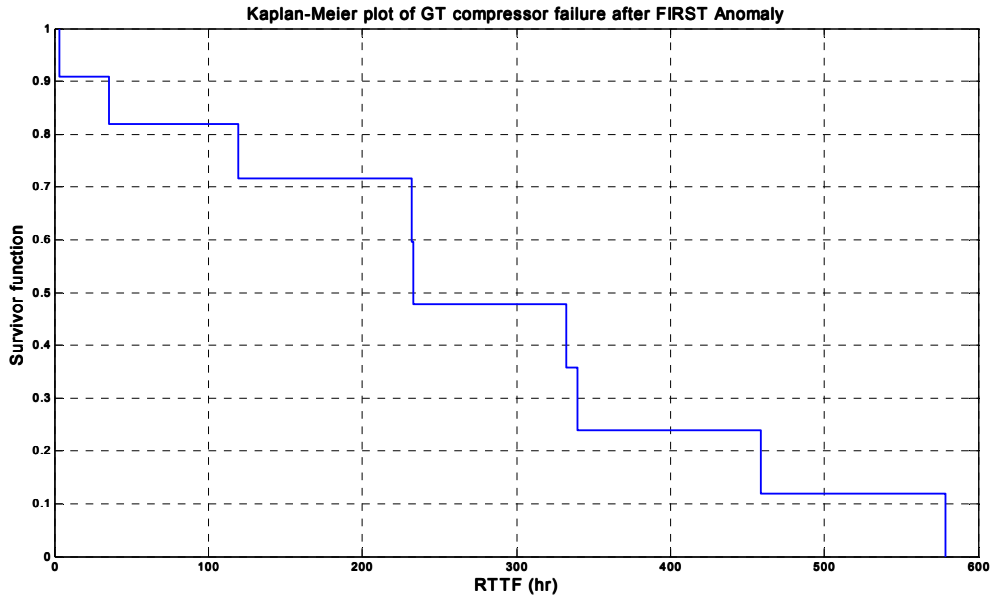


Figure 89: Kaplan-Meier plot of the residual time to failure for first detected anomaly

However, as stated in the previous chapter, Figure 89 is an estimate; therefore, the confidence level of that estimation needs to be established. In order to achieve a 95% percent confidence interval, section C.1 in the appendix is calculated. Figure 90 shows the Kaplan-Meier plot, shown in Figure 89, with the addition of a 95% confidence band. As one can see, the uncertainty is large.

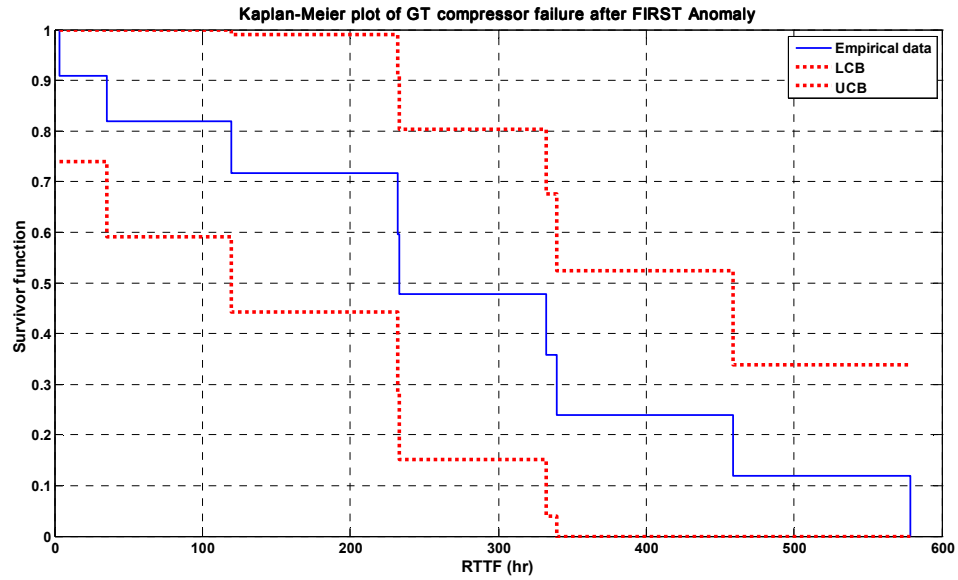


Figure 90: Kaplan-Meier plot of the RTTF from first detected anomaly with 95% confidence band

Since the anomaly detection analysis showed that most of the units had more than one anomaly before failure, it is interesting to see how the Kaplan-Meier plot for the second detected anomaly compares to that for the first detected anomaly. Table 40 shows the tabular data for the Kaplan-Meier plot of residual time after the second detected anomaly.

Table 40: Tabular data for the Kaplan-Meier plot of survival time after the second detected anomaly

Failure Time (RTTF)	Number failed	Number censored	At Risk	Survival (\hat{S})
0	0	0	11	1
1	1	0	11	0.909091
2.783	1	0	10	0.818182
11.05	1	0	9	0.727273
37.317	1	0	8	0.636364
48.367	0	1	7	0.636364
49.383	1	0	6	0.530303
78.467	0	1	5	0.530303
161.167	1	0	4	0.397727
201.633	1	0	3	0.265152
318.983	1	0	2	0.132576
332.15	1	0	1	0

Similarly, Figure 91 shows the Kaplan-Meier plot of the residual time to failure after the second detected anomaly.

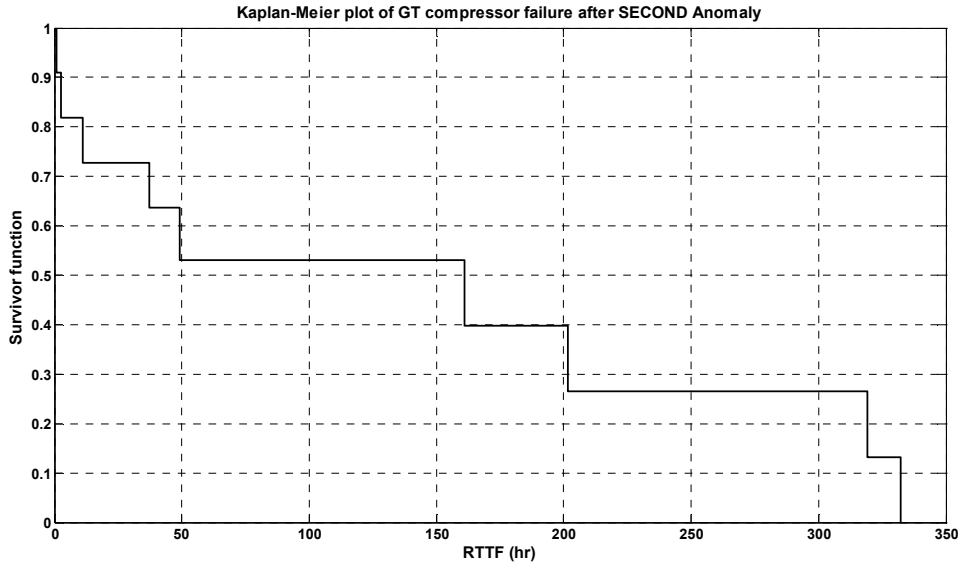


Figure 91: Kaplan-Meier plot of RTTF for second detected anomaly

By proceeding as we did for the first anomaly, the 95% confidence interval for the Kaplan-Meier plot in Figure 91 is tabulated in the section C.2 in the appendix C.

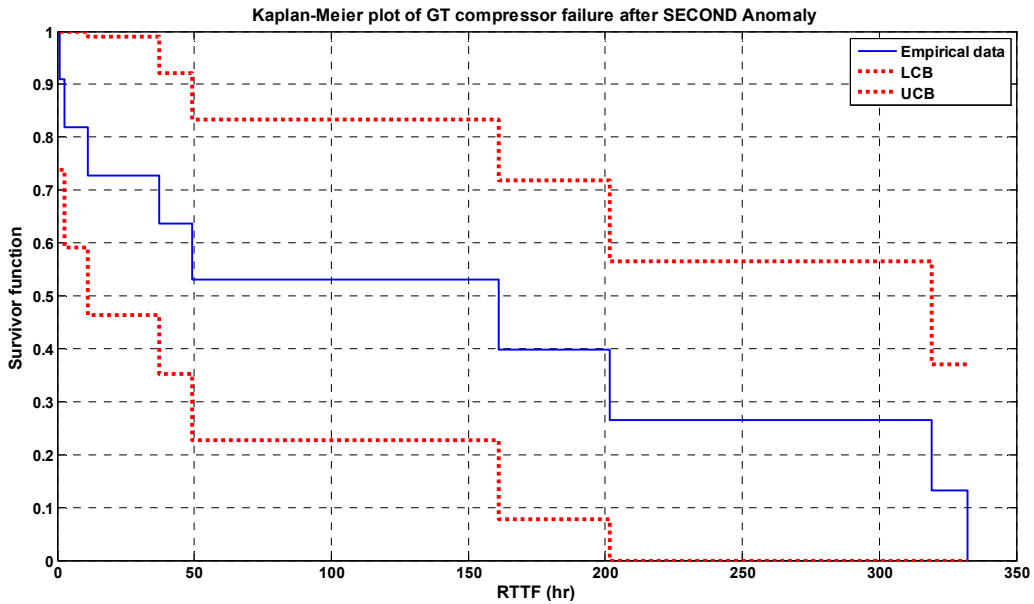


Figure 92: Kaplan-Meier plot of the RTTF from SECOND detected anomaly with 95% confidence

Figure 92 shows the Kaplan-Meier plot from Figure 91 with the addition of its corresponding point-wise 95% confidence interval. There is still a large uncertainty interval.

Although the first and second detected anomalies have a similar shape, plotting them together will provide a visual comparison, as shown in Figure 93.

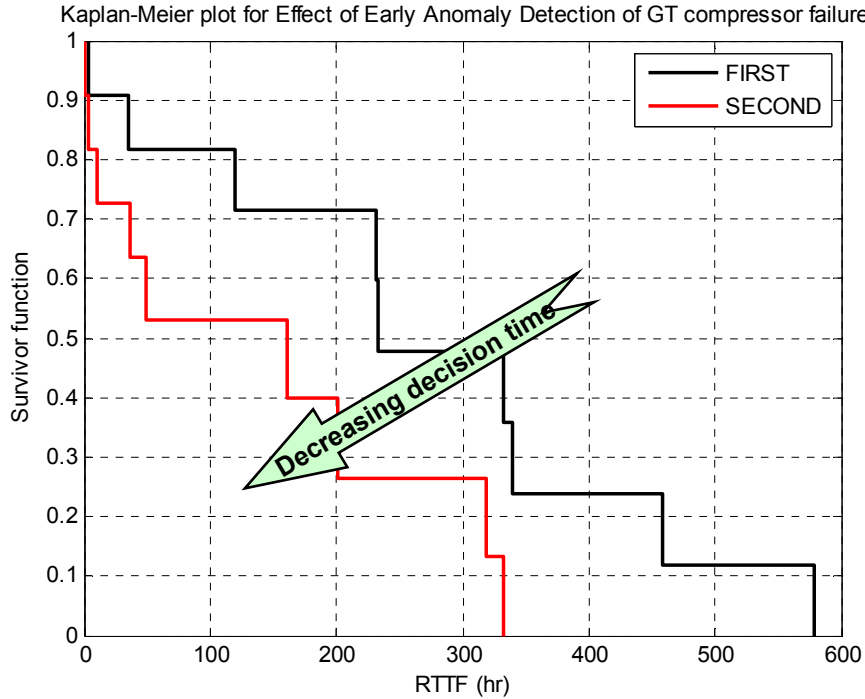


Figure 93: Comparison of the first and second anomalies

As it may be intuitive to many, Figure 93 shows that fact that the sooner an anomaly is detected the more time is given back to the decision makers to shut down the unit so as to avoid catastrophic failure.

As in the case of the deterministic regression, let's investigate the impact of the extended data set on the Kaplan-Meier plot for RTTF. The tabular data for the Kaplan-Meier plot of the extended residual time to failure data set is compiled in section C.3 of Appendix C. Also, Figure 94 is the graphical representation of the Kaplan-Meier plot in section C.3.

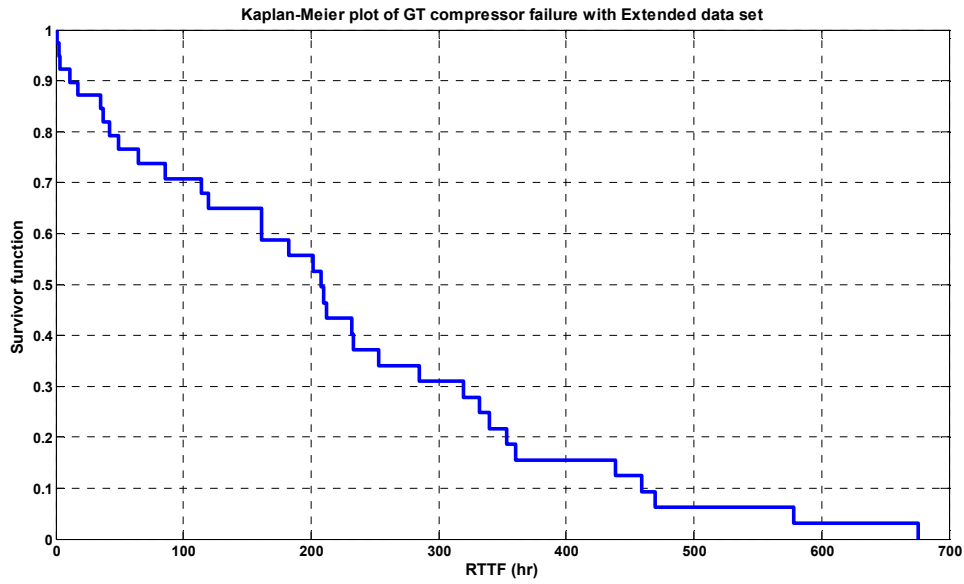


Figure 94: Kaplan-Meier plot for RTTF with extended data set

Again, because Figure 94 represents an estimate, the confidence level needs to be established. The tabular extended data set for the 95% confidence interval is shown in section C.4 of Appendix C. Figure 95 shows the Kaplan-Meier plot of the extended data set with its corresponding 95% confidence interval.

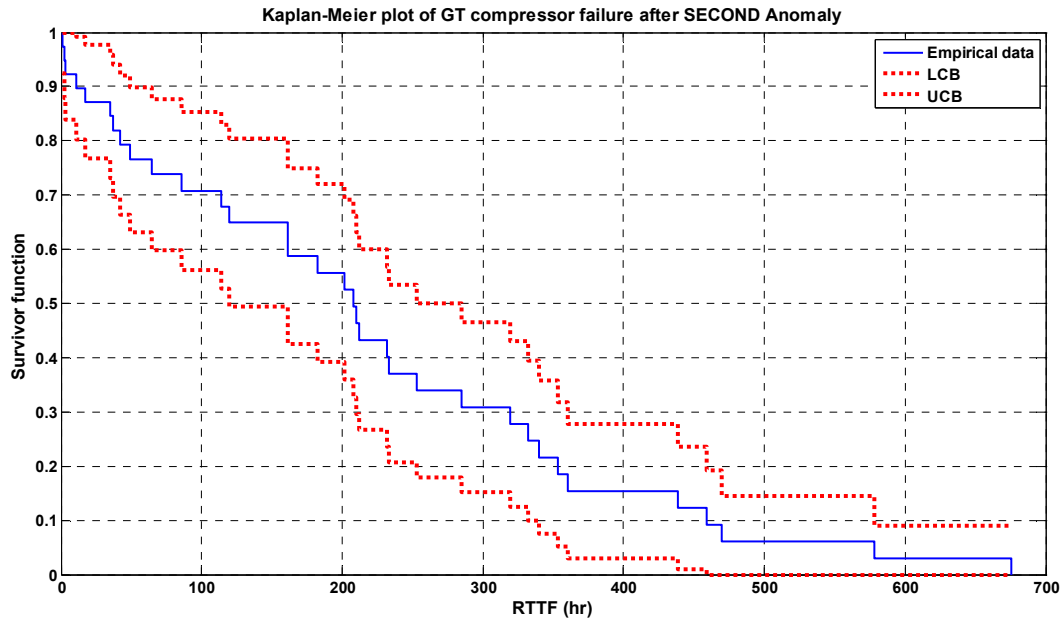


Figure 95: Kaplan-Meier plot of the RTTF for extended data set with 95% confidence

Contrary to the deterministic case, the nonparametric approach to the extended data has a tighter uncertainty band, clearly indicating that the larger, the data set, the more accurately the Kaplan-Meier plot can be estimated.

6.2.3 Parametric Distribution

Although the non-parametric approach provides an empirical plot, which is the best representation of any failure data, the non-parametric approach is not flexible enough to be applied to other units of a given fleet not included in that failure data set. Thus, it is inappropriate for the purpose of this thesis, which is to find a way to guide the gas turbine manufacturers and unit owners to avoid catastrophic failure after a failure precursor is detected. A parametric analysis using different distributions than the ones presented in the previous chapter will be implemented in this section.

6.2.3.1 Exponential Distribution

When there is no knowledge of a time-based distribution, as is the case here with the residual time to failure, the use of exponential distribution as a first estimate is always appropriate. The goal is to try to fit the Kaplan-Meier estimate data using the exponential distribution. That is, if it is assumed that the Kaplan-Meier data is exponentially distributed, then the probability density distribution of the first anomaly is shown in Figure 96.

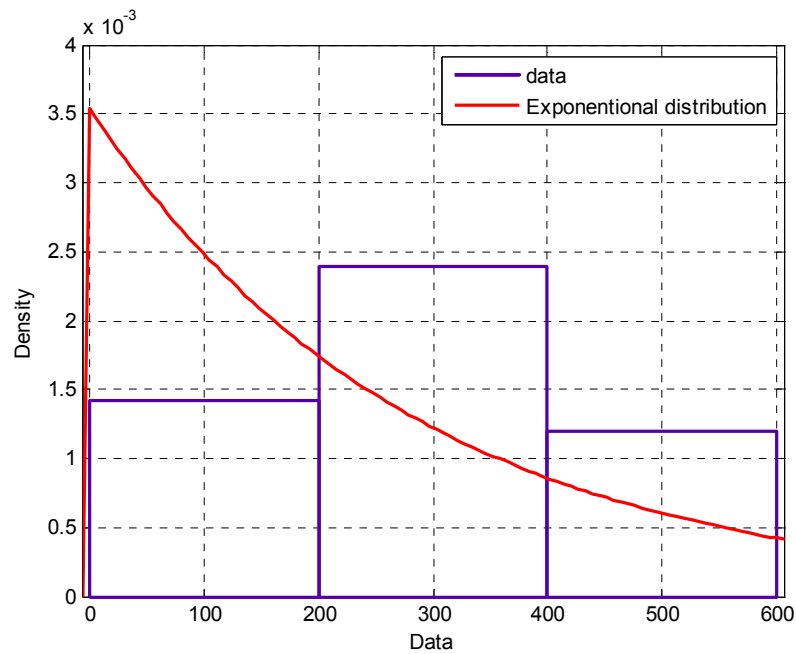


Figure 96: pdf of the first Anomaly using Exponential distribution

By plotting the equation $y = -\lambda x$ with $x = t$, one can create an Kaplan-Meier estimated exponential plot. This is shown in Figure 97.

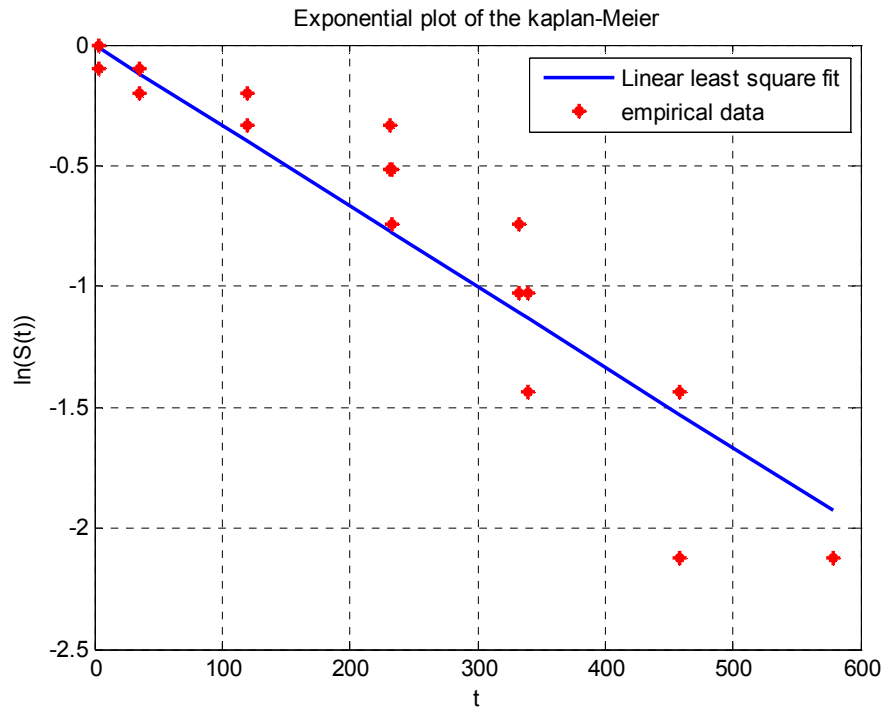


Figure 97: Exponential plot of the Kaplan-Meier estimation of the RTTF

The linear least square fit of the Kaplan-Meier estimated exponential plot of the RTTF has the equation:

$$y = -0.0033x \text{ with a corresponding } R^2 = 0.869 .$$

Although the coefficient of determination is not as bad as in the deterministic regression case, it is not perfect. Figure 98 also provides another visualization of the Kaplan-Meier estimate and the exponential fit.

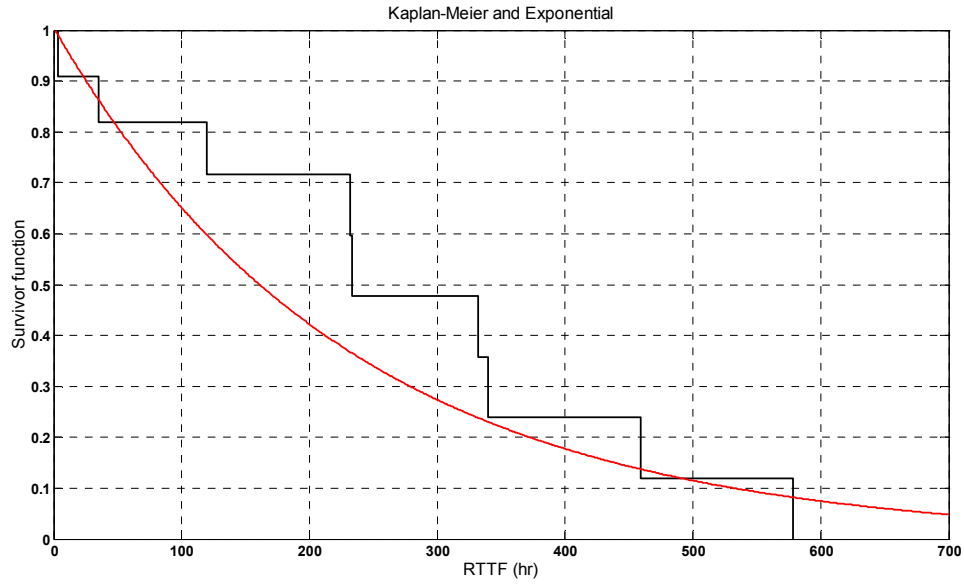


Figure 98: Kaplan-Meier estimated RTTF of first anomaly and exponential fit

Let's analyze the impact of the exponential distribution assumption using the extended data set. First, the pdf of the exponential assumption better matches the extended data set than does the limited (first anomaly only) data set, as illustrated in Figure 99.

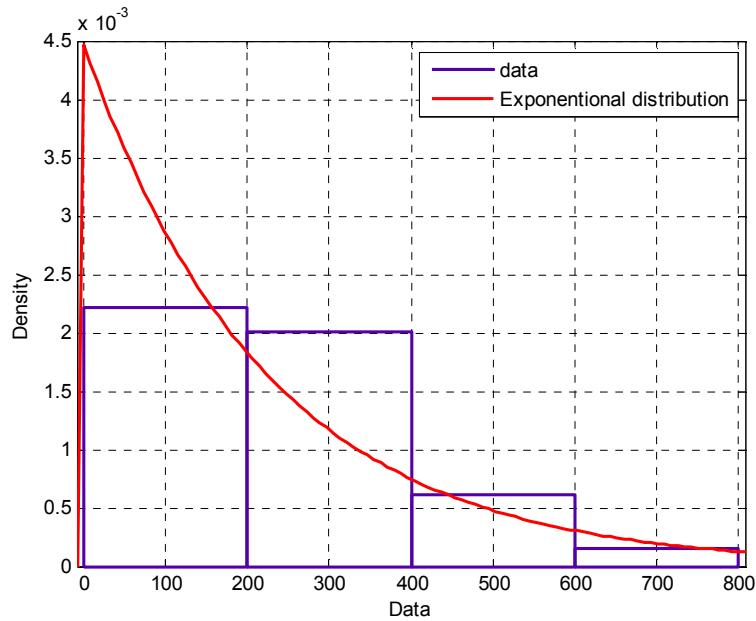


Figure 99: pdf of extended data set using the exponential distribution

We then plot the equation $y = -\lambda x$ with $x = t$. The Kaplan-Meier estimated exponential plot of the extended data set is shown in Figure 100

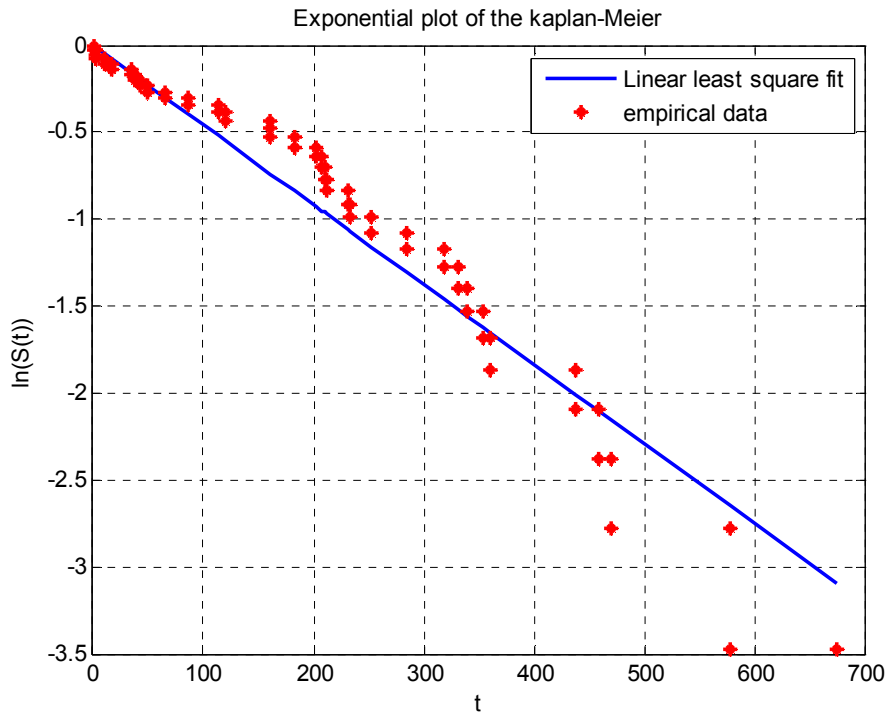


Figure 100: Exponential plot of the Kaplan-Meier estimated of the RTTF using the extended data set

The linear least square fit for the exponential plot of the Kaplan-Meier estimated of the RTTF has an equation $y = -0.0046x$ with a corresponding $R^2 = 0.9386$.

The coefficient of determination in the extended case is much better than in the case of the limited data set. This stresses the fact that for a larger data set, exponential distribution fits the Kaplan-Meier estimate better.

Again, the visual representation of the Kaplan-Meier estimate and the exponential fit shown in Figure 101 is better.

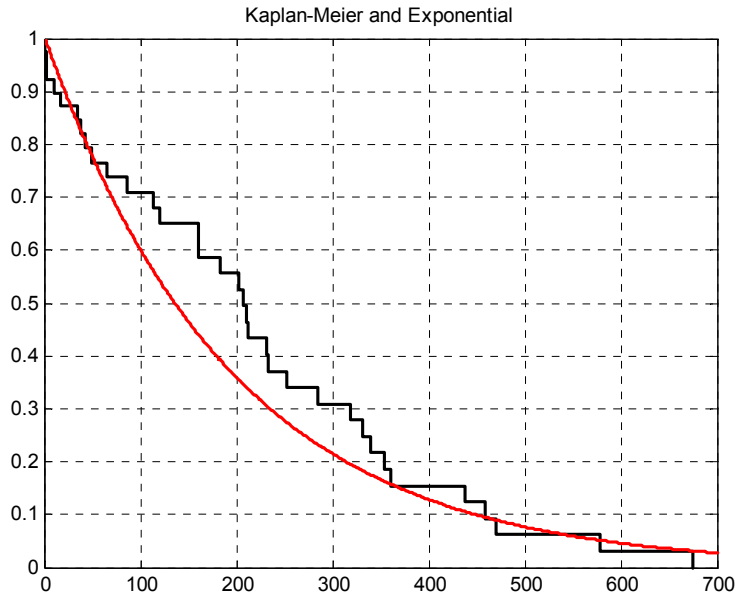


Figure 101: Kaplan-Meier estimated RTTF for extended data set and exponential fit

6.2.3.2 Weibull Distribution

The Weibull is the second distribution to be implemented. Figure 102 shows the pdf of the first detected anomaly.

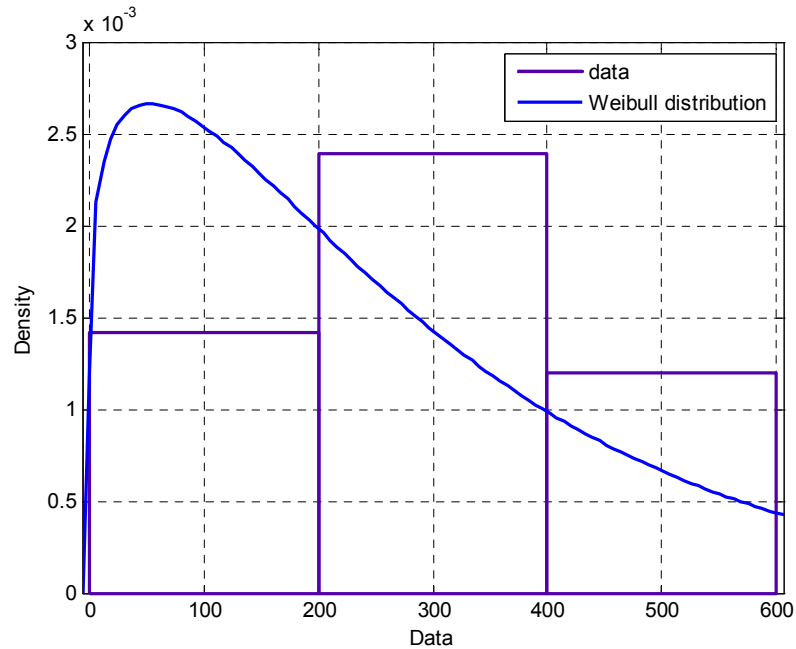


Figure 102: pdf of the first anomaly using the Weibull distribution

In the case of the Weibull plot, the linear least square fit of the Kaplan-Meier estimation of the RTTF has an equation in the form:

$$\begin{cases} x = \ln(t) \\ y = \beta \cdot x - \beta \cdot \ln(\theta) \end{cases}$$

Therefore, the Weibull plot of the Kaplan-Meier is shown in Figure 103.

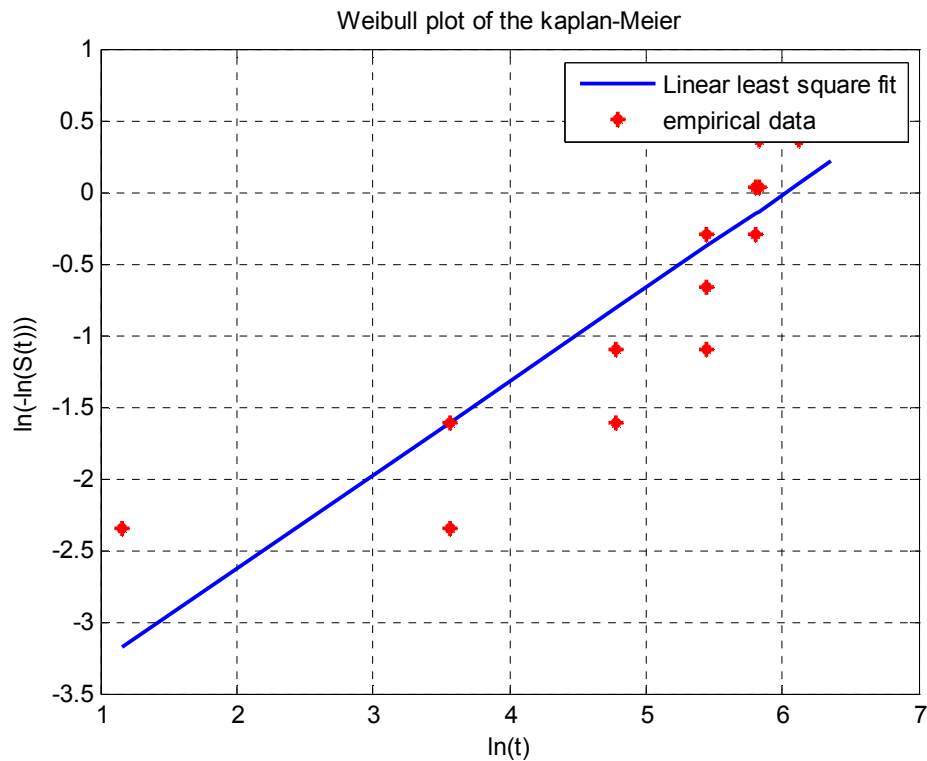


Figure 103: Weibull plot of the Kaplan-Meier estimated RTTF

The equation of the linear fit is:

$$y = 0.6512x - 3.9275 \text{ with a corresponding } R^2 = 0.7492$$

A plot of the Kaplan-Meier estimated RTTF of the first anomaly and the Weibull fit that serves as a visual, is shown in Figure 104

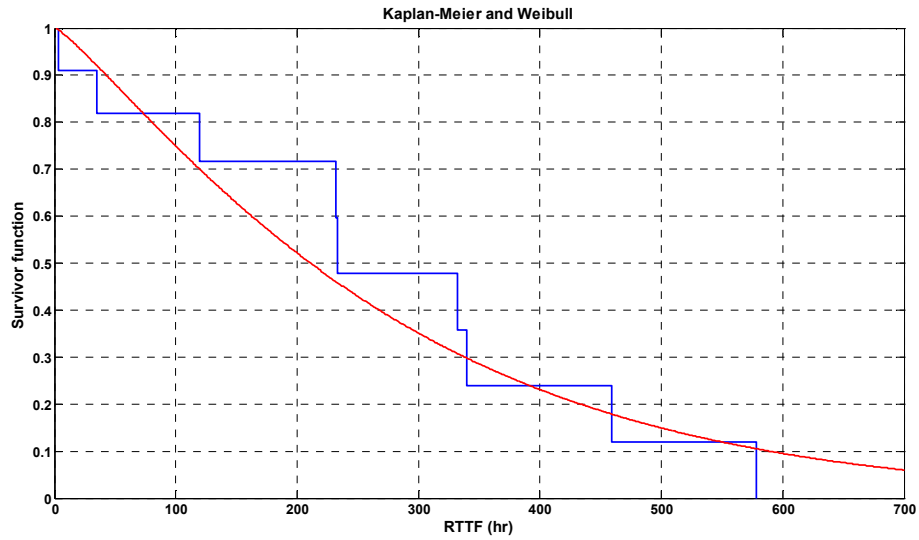


Figure 104: Kaplan-Meier estimated RTTF of first anomaly and Weibull fit

As we did in the exponential case, let's investigate the impact of the extended and Weibull distribution assumptions. The pdf of the Weibull assumption, shown in Figure 105, better matches the extended data set than it does the limited (first anomaly only) data set..

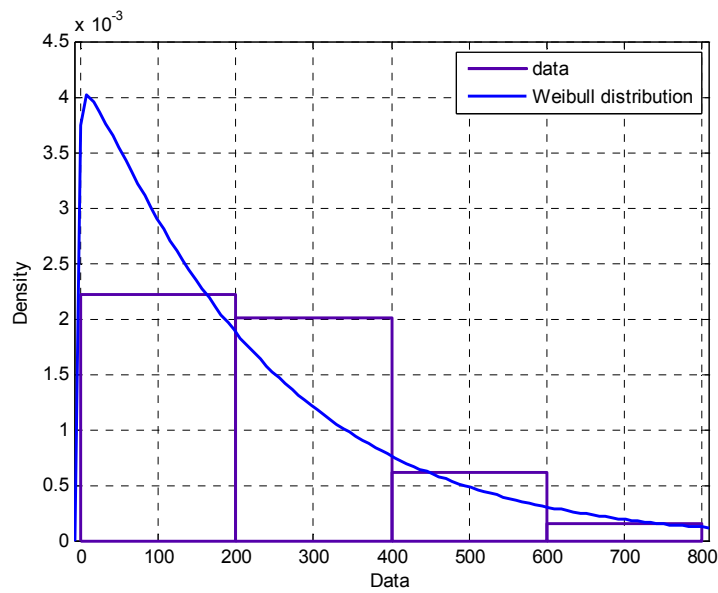


Figure 105: pdf of extended data set using the Weibull distribution

One must then plot the equation.

$$\begin{cases} x = \ln(t) \\ y = \beta \cdot x - \beta \cdot \ln(\theta) \end{cases}$$

The Weibull plot of the Kaplan-Meier estimation of the extended data set is shown in Figure 106.

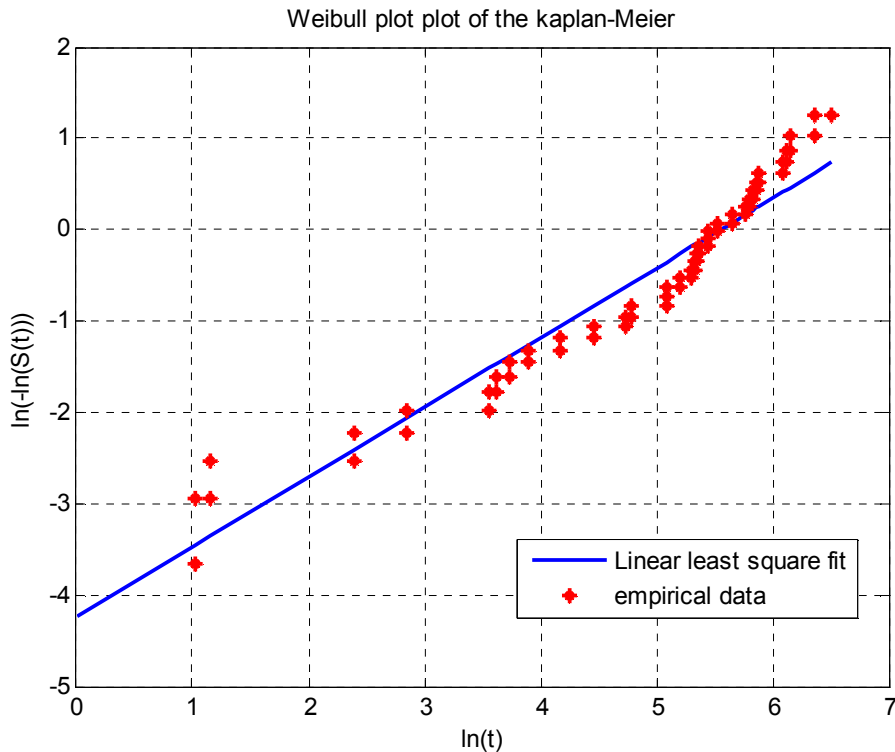


Figure 106: Weibull plot of the Kaplan-Meier estimation of the RTTF using extended data set

The best linear least square fit of the Weibull plot of the Kaplan-Meier estimation of the RTTF has an equation: $y = 0.7624x - 4.233$, with a corresponding $R^2 = 0.9313$.

Again, the coefficient of determination in the extended case is much better than it was in the case of the limited data set for the Weibull distribution. This shows that, in the case of a larger data set, a Weibull distribution fits the Kaplan-Meier estimate better.

Again, the visual representation of the Kaplan-Meier estimate and the Weibull fit, as shown in Figure 107, is better.

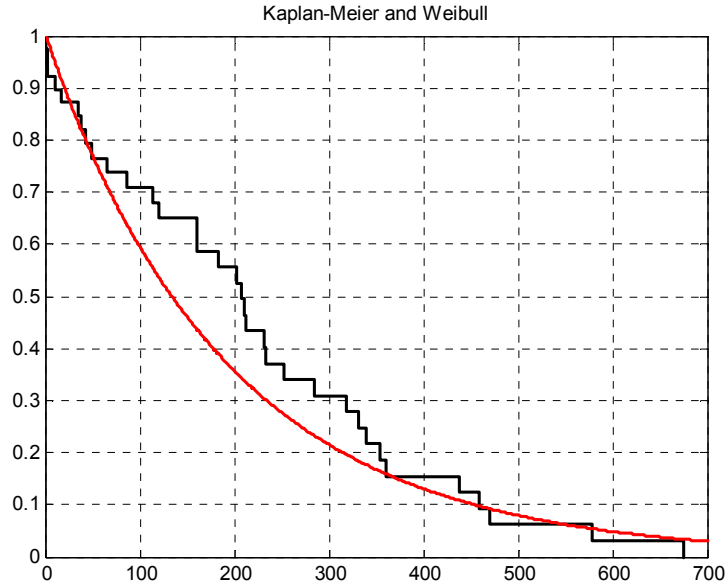


Figure 107: Kaplan-Meier estimated RTTF for extended data set and Weibull fit

6.3 Results Discussion

This section serves to recapitulate the chapter on the implementation of the proposed methodology. It is shown that the deterministic regression approaches, both parametric and non-parametric, are not adequate for accurately predicting the residual time to failure; even when the number of predictors is increased to three, by taking into account the fact that a detected anomaly is not an event independent of its predecessors, the regression quality worsens. One way this observation could be explained is the challenge of knowing with certainty all the predictors that affect a response.

Secondly, the non-parametric analysis using the Kaplan-Meier estimate, which is an empirical plot of the residual time to failure, was conducted. As any estimation requires, a 95% confidence level was computed for both the limited data set and the extended data set. As expected, the larger data set had a reduced uncertainty band. Also,

the first and second detected anomalies were compared to each other. It was revealed that the impact of early anomaly detection is that the gas turbine operation decision maker could take less disruptive actions. In other words, the sooner an anomaly is detected, the more time is given to the decision maker to avoid a catastrophic failure.

Although the non-parametric analysis using the Kaplan-Meier provides a great result because the computation of the residual time to failure does not make any predefined distribution assumption, it cannot be used as an effective prediction tool for other units in a given fleet. This does not fulfill all the self-imposed goals of the thesis, one of which is to predict the residual time to failure for other fleet units based on the knowledge gathered from the analysis of previous failures. Thus, it was imperative from that point of view to investigate some parametric distributions, such as the exponential distribution and the Weibull distribution. The implementation of the parametric distributions was promising but did not fit well enough with the limited data set for either the exponential or Weibull distributions, as shown in Figure 108.

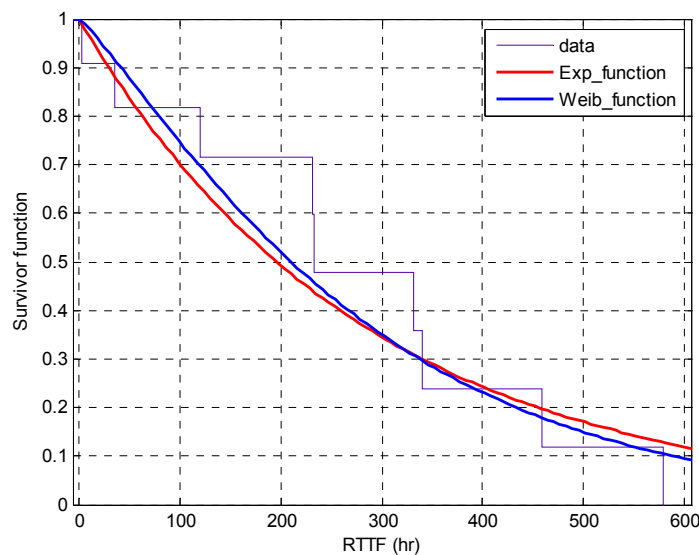


Figure 108: Comparison of exponential and Weibull distribution fits with the Kaplan-Meier

Both the exponential and the Weibull distribution have a much better fit to the non-parametric extended data set, as illustrated in Figure 109. The goodness of fits was assessed using the coefficient of determination R^2 , which yield larger values with the extended data set. Another metric that was used to check how the fit quality gets better with the size of the data set was the Residual vs. Predicted plots shown in the Appendix D for the both data set. Also, a fact worth noting is that both distributions become very similar with larger distributions.

The takeaway is that the larger the data set used for the Kaplan-Meier, the better the parametric distributions fit the Kaplan-Meier.

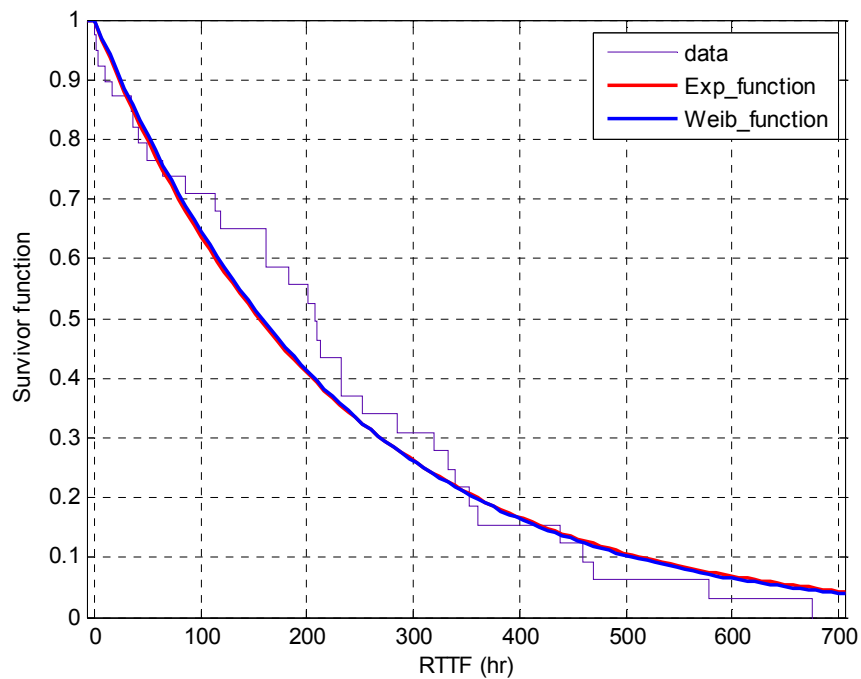


Figure 109: Exponential and Weibull distribution fits with the Kaplan-Meier for extended data set

The parametric analysis using the exponential and Weibull distributions for both the residual time to failure data set and the extended data set version is summarized in Table

41

Table 41: Summary of parametric survival data analysis

Sampling Size	Coefficient of determination		Exponential Parameter	Weibull Parameters	
	Exponential	Weibull	lambda	Beta	Theta(hr)
Limited data set	0.869	0.7492	0.0033	0.6512	416.16
Larger data set	0.9386	0.9313	0.0046	0.7624	257.86

CHAPTER 7

CONCLUDING REMARKS

7.1 Revisiting the research objectives and questions

7.1.1 Revisiting the research objectives

The foremost goal of this dissertation is to develop an approach to detect precursory anomalies as early as possible in order to avoid the catastrophic failure of gas turbines. To focus the thesis on that goal and to meet the requirement for scientific approach research set by the institution, two major objectives are defined:

7.1.1.1.1 *Objective 1:*

Develop a process to find precursory failure signatures as early as possible in order to avoid systems catastrophic failure.

7.1.1.1.2 *Objective 2:*

After an anomaly is detected, develop a strategy to prognosticate an estimation of the residual time to failure.

The first research objective is the main driver of this research, and the ability to achieve it gauges the success of this work. The second objective provides the necessary information to make a system shut down as undistruptive as possible.

In order to look for a way to meet the imposed objectives, a series of research questions are asked and using the literature review and industry practices, potential hypotheses are formulated.

7.1.2 Revisiting research question 1, 1a and hypothesis 1, 1a

The research question 1 and its sub-question 1a are about investigating ways to meet the first objective of this thesis. They are recalled here with their corresponding hypotheses:

7.1.2.1.1 Research question 1:

How can a precursory anomaly that might lead to a gas turbine catastrophic failure be detected?

Though the ability to detect an anomaly as early as possible (i.e. almost real time) is in itself an advancement of the current industry practice, the ability to reduce the number of false alarms is also desirable. Thus the research question 1a is necessary.

7.1.2.1.2 Research Hypothesis 1:

A multi-step process that relies on the Wavelet transforms theory can be used to detect a precursory anomaly that might lead to the catastrophic failure of a gas turbine compressor.

7.1.2.1.3 Research question 1a:

How can a failure precursor detection method be made robust?

7.1.2.1.4 Hypothesis 1a:

Combining the information from the two types of sensors could decrease the number of missed precursory anomalies.

The proposed methodology presented in chapter 4 is a multi-step process that has successfully captured known anomalies detected by a team of gas turbine manufacturers in a post-failure analysis, and has also captured other less severe anomalies that they missed. Specifically, the proposed methodology is a six step approach: 1) raw data acquisition of appropriate sensors, 2) sensor normalization, 3) de-noising of each sensor signal, 4) multi-resolution analysis using the discrete wavelet transform, 5) sensor data fusion into a few uncorrelated principal components, 6) anomaly detection decision.

It is shown in chapter 4 that the combination of the two types of monitoring sensors measurement (vibration and performance) improves the capability to detect failure precursors. Though a study of a group of eleven failed turbines has revealed that most of the failure precursor anomalies have signatures in the vibration sensors measurement, there are few cases where specific failure signatures were detected entirely by the performance sensors. In the case of the latter, any process solely based on the sensor measurements would miss those failure signs.

7.1.3 Revisiting research question 2 and hypothesis 2

Research question 2 and sub-question 2a refer to the ability to determine the level of confidence of the statistical technique used to make the anomaly detection decision. The two questions and hypotheses are as follows:

7.1.3.1.1 Research question 2:

How can a detected precursory event be statistically validated?

7.1.3.1.2 Research hypothesis 2:

The X-bar control chart type of statistical process control can be utilized to establish the anomaly threshold, and thus distinguish the faulty events from the non-defective ones.

7.1.3.1.3 Research Question 2a:

How does the sampling interval impact on the quality of detection?

7.1.3.1.4 Hypothesis 2a

The smaller the time step (the higher the frequency) is, the more accurate the detection quality will be.

The research question 2 and sub-question 2a are important ones in any event detection process because the ability to keep the number of false alerts low is the key to the successful implementation of any new process. Thus, to reduce the risk of the conflicting interpretation of many sensors signals, a process is underlined in chapter 4, to reconstruct a one-dimensional signal representing the overall health of the system, monitor two damage indicators calculated from the reconstructed signal are monitored and calculate the defect threshold using the x-bar chart of the statistical process control. Also, in chapter 4 it is explained that the proposed methodology is designed so that the calculated defect threshold depends on the risk a system analysis is willing to run (as type I error) of interrupting the system operation when anomalies may not be present. The type II error is then deduced.

As for the impact of the sampling time, the decision is made to use the 1 minute time interval because it has been proven adequate. In fact, there is an important trade-off to be made between the accuracy and the speed of execution of the anomaly detection; the 5 minute time interval is not accurate enough because it misses short lasting anomalies, while the 5 second interval is more accurate but is also more sensitive and prone to false alarms.

7.1.4 Revisiting research question 3 and hypothesis 3

The research questions 3 and 4 and their respective corresponding hypotheses 3 and 4 are intended to meet the second defined thesis objective.

7.1.4.1.1 Research question 3:

If a failure precursor event is validated, how can the residual time to failure be estimated?

7.1.4.1.2 Research hypothesis 3:

Survival analysis techniques such as deterministic regression, non-parametric and parametric analysis of failure data can be used to build models that estimate the residual time to failure.

The ability to predict the residual time to failure after an anomaly is detected puts more power in the decision markers' hands. In fact, based on the presented methodology in chapter 5, and the application in chapter 6, not all the anomalies are equal. Indeed, the two defined health indexes (anomaly severity and anomaly duration indexes) showed that for the eleven units covered under this study, the failure always happened after a fairly important defect intensity (either high severity or long duration).

However, the deterministic regression techniques implemented in chapter 5 did not yield acceptable residual life estimations. While the Kaplan-Meier provides a true picture of the residual time to failure of the units studied, it is not flexible enough to provide a prediction for the residual time for future units with a detected anomaly, so the parametric analysis of exponential and Weibull distributions provides fairly decent fits. The quality of the prediction model using exponential and Weibull distribution as parametric data analysis techniques gets much higher with the larger data set. Therefore, it can be concluded that for a larger data set, the parametric failure data analysis can be used.

7.1.5 Revisiting research question 4 and hypothesis 4

Research question 4 and hypothesis 4 are necessary for any estimation. In fact, the value of any estimation lies in the confidence interval associated with it.

7.1.5.1.1 Research question 4:

How can the confidence level of the estimated residual time to failure be determined?

7.1.5.1.2 Research hypothesis 4:

Existing mathematical techniques can be used to compute the confidence interval of the survival time estimation.

Although the non-parametric Kaplan-Meier estimator does not have a close form to compute the confidence interval, the Greenwood formula can be used to estimate the variance so that the confidence interval can be calculated. The parametric survival distributions of exponential and Weibull have readily available confidence interval computation forms.

Thus, it can be stated that based on the revisions of the research questions and hypotheses, each stated hypothesis verified has verified.

7.2 Summary of Findings and Research Contributions

This research has produced a number of contributions to the current industry standard. The first contribution is the creation of a systematic methodology, following a step by step approach that can be applied and made easily transferable to other sensor- monitored systems. The second contribution is that the proposed methodology provides a way to combine information from different type of sensors to create a much more robust precursory failure events detection of systems.

The third contribution of this dissertation is that, contrary to other early defect detection techniques that rely on sophisticated mathematical underlying approaches, such as the neural network, this proposed method uses a statistical process control that is more flexible, faster, and easier to implement. Furthermore, the proposed detection scheme is machine specific, which addresses the issues linked to machine-machine variation.

The fourth contribution of this thesis is the idea that the deterministic approach is not adequate to model the residual time to failure, even with a large data set, because it typically does not have a hazard function.

The fifth contribution of the study is that the residual time to failure prognostication can be modeled fairly well with a large data set.

7.3 Conclusion

Despite the risks involved with the LTSA contracts, they can be an enormous revenue stream for OEMs. Over the years, the profit margin for the sale of new gas turbines has been shrinking for OEMs, though LTSA for the most part assured OEMs of the new sale of upgraded components to sustain degrading gas turbine performance throughout the duration of the LTSA. The market for LTSA can be a very lucrative one if the level of liquidated damages is held to a minimum.

In this thesis, a systematic multi-step data analytics prognostic and health management approach is presented, which could be used to detect precursory anomalies that could lead to catastrophic equipment failure. This could reduce or even eliminate the unplanned power plant outages. The approach is divided into two parts: failure precursor detection and residual life estimation.

In general power plants have two main types of monitoring sensors: the process-related measurements and high bandwidth measurement sensors which are high frequency sensors. This study required a time-frequency technique that could process high-frequency signals. Therefore, the proposed approach started with the pre-processing of the raw data from the system-monitoring sensors, which ensured that the analysis outcome would not be skewed toward the variables with high absolute value. Then the discrete wavelet packet transform was used to both remove the noise from the data and to decompose the signal in order to extract the signal features. After that, the multi-sensor

time series variable was fused into a few uncorrelated principal components using the probabilistic principal component analysis technique. Finally, the anomaly detection decision was made based on the Bayesian evaluation method, applied to a precision-related metrics, for hypothesis testing to successfully detect failure precursors.

The proposed process has shown promising results in its application to a GE 7FA gas turbine compressor failure because it has captured known failure precursors.

The proposed approach uses an elaborate statistical process control technique to set the defect threshold. As a result of this utilization of the statistical process control, the proposed methodology can be easily implemented, unlike techniques that rely on the use of a neural network, in which a high fidelity mathematical model is required. Furthermore, the process is robust with few false alarm and a much lower false-negative for given false-positive probability. Additionally, the use of the statistical approach allows the handling of practical issues such machine-to-machine variation, both in the design and the operation, because the abnormality threshold is machine specific. Importantly, the proposed methodology has the ability to not just detect an anomaly, but also to determine its severity through two defined anomaly indexes. Moreover, the statistical approach to threshold determination allows the computation of both the type I and the type II error associated with the detection process.

After a failure precursor is detected, different techniques are used to model the residual time to failure. First, the deterministic regression, which used the response surface methodology and the artificial neural network, was investigated; however, that method did not yield an acceptable accuracy for the residual life prediction. Then, the non-parametric survival data analysis using the Kaplan-Meier estimator provided the empirical fit, and its corresponding confidence interval was determined. Also, by using the Kaplan-Meier to compare multiple anomalies detected on the same unit, in the cases in which there were multiple defects before the catastrophic failure of the unit, the method showed that the earlier an anomaly was detected, the more time was given back

to the decision makers to limit the impact of disruption associated with the unplanned outage of a power plant. Despite the fact that the Kaplan-Meier is not constrained by any predefined residual lifetime distribution, it does not provide an indication of the survival time of other fatally defective units not included in the data set. Therefore, two very popular parametric survival analyses, the exponential and the Weibull distribution, were used. For a smaller data set, both the exponential and Weibull distributions provided a better fit to the empirical Kaplan-Meier than did the deterministic regression techniques. Furthermore, for a larger data set, the two parametric failure data analyses provided even better fits to the empirical data, thereby providing the ability to model the residual time to failure after a failure anomaly is detected.

7.4 Recommendations for Further research

Incorporation of physical aspect of failure and failure modes to data driven to improve the quality of PHM:

Although the proposed methodology has proven promising, the assumption that the failures are independent and identical distribution (iid) may not be adequate. Indeed, there are many different failure modes within a gas turbine compressor. The incorporation of information from physic-based models is desirable in order to improve the estimation of the residual time to failure.

Investigate Ways to reduce the current number of required steps of the proposed method:

It would be desirable to find ways to accelerate the failure precursor detection approach. It would be interesting to investigate, for example, how to go from the multi-resolution analysis to the reconstructed one-dimensional signal of overall system health without

losing the added capability to carry over the data uncertainty provided by the Probabilistic Principal Component Analysis step. One suggestion for such a study is the use of eigenvalues from the covariance matrix as the weight for each variable.

Pattern recognition:

It would be very interesting to expand the proposed approach to another larger data set as a way to perform pattern recognition. One such an application would be the early detection of natural disasters such as earthquakes, in which the response decision is made based on the recognition of a certain recognizable signature.

Life Extension:

This work could be used as a starting point for life extension strategies, in which events that might affect a system's lifetime could be detected as early as possible, and action could be taken to reverse an undesirable course. It is worth noting that life extension is very crucial for critical and often expensive systems. Currently, major life extension research is conducted for the US Nuclear Weapon systems.

Semi-Parametric approach for residual time to failure modeling

Subsequent researchers could repeat the study presented in this thesis for other gas turbine fleets like the 9FA. One could use the semi-parametric approaches of the proportional hazard and the accelerated failure time model to investigate how the residual time to failure of different fleets of gas turbines compare to each other.

APPENDIX A- EXPERIMENTS

The list of variable is the same as shown on Table 42 for each unit unless specified otherwise

Table 42: Monitored Gas turbine compressor sensors

Variable	Variables Description
X1	Overall system health parameter 1
X2	Overall system health parameter 1
X3	Compressor seismic vibration 2
X4	Compressor seismic vibration 2
X5	Turbine seismic vibration 3
X6	Compressor health parameter 1
X7	Compressor health parameter 2
X8	Compressor effectiveness health parameter 1
X9	operating condition (load)
X10	operation condition 1 (environment)
X11	Compressor effectiveness health parameter 2
X12	operation condition 2 (environment)

A.1- Test unit 1

Test unit is as used as a demonstration unit in chapter 6.

A.2- Test unit 2

At the MRA step, each sensor signal is decomposed into 3-level using the DWPT with the Daubechies 4, that lead to 8 different wavelet components (1 approximation, and 7 details as shown on the tree decomposition of the Figure 110: Unit 2-Tree decomposition of sensor X3 signal with the wavelet component at node 7:

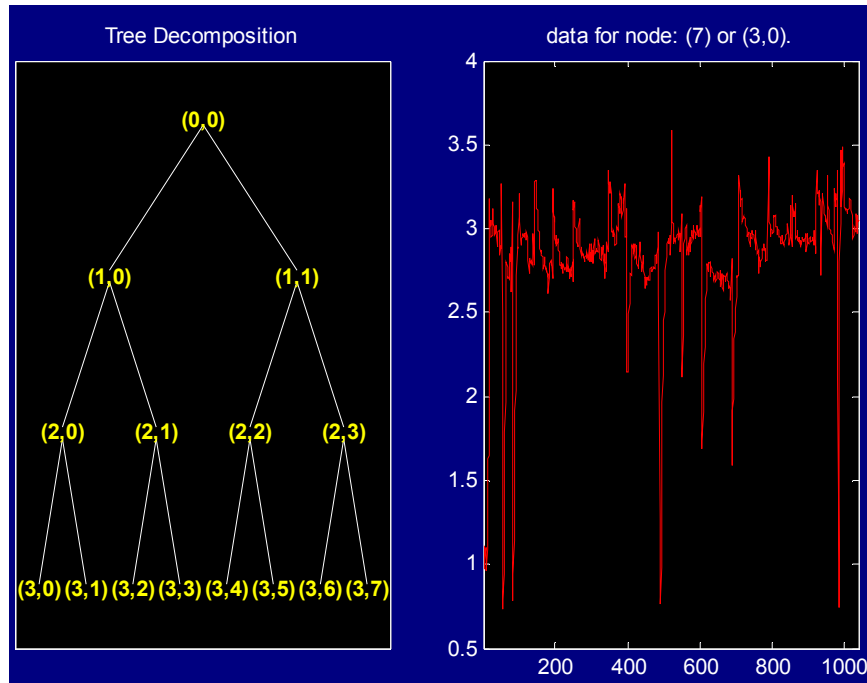


Figure 110: Unit 2-Tree decomposition of sensor X3 signal with the wavelet component at node 7

The Figure 111 below shows the original de-noised signal and each of the 8 wavelet components (original signal in black, approximation in red, details in blue)

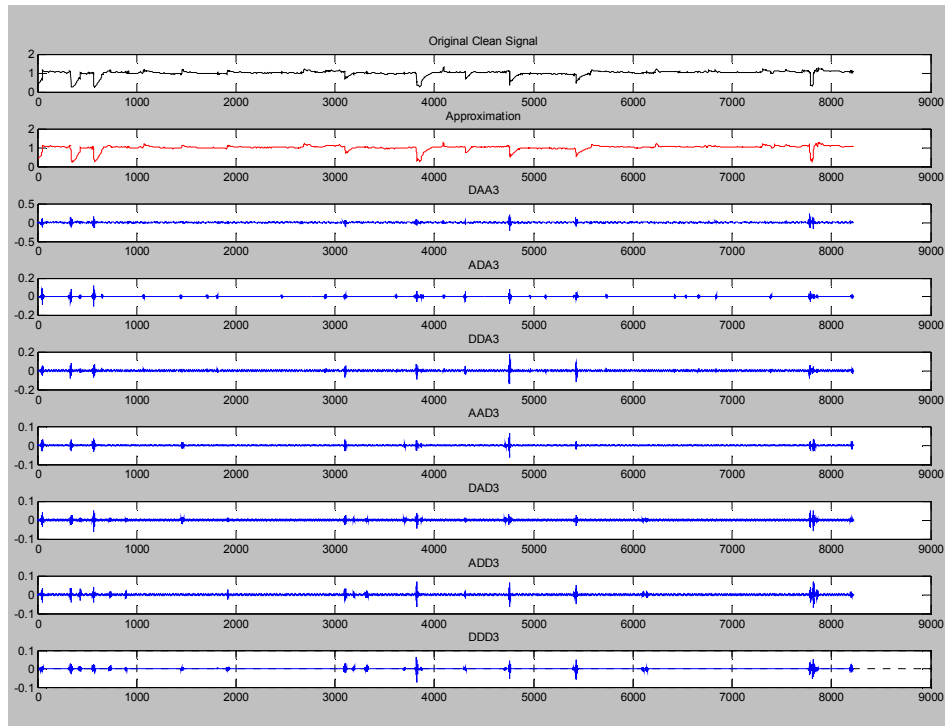


Figure 111: Unit2- DWPT 3-level MRA of sensor x3

Then the energy content at each of the node at the level 3 is evaluated and grouped into two groups: approximation (node 7) and the details (sum of the details of node 9 through 14) is shown in the table below:

Table 43: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 2

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.963	99.962	99.982	99.999	100	100	100	99.994	100	100
Sum of details (% of Original)	1.00E-03	3.27E-06	3.60E-03	3.70E-03	1.80E-03	4.81E-01	1.64E-04	2.74E-04	1.47E-07	6.30E-03	2.44E-01	2.99E-06

For each of the 8 sensors, the energy contained within the approximation node is over 99.9% of the total energy of the original signal. Therefore, the approximation signal (or scale function) will be used as the representative of the actual signal in the subsequent

steps. That is, the energy content in each of the wavelet components will serve as the signal feature characteristic.

Step 5: PPCA data fusion

Then the eigenvalues and their respective contributions and the corresponding eigenvectors for the covariance matrix are obtained and ranked from the highest value to the lowest value:

Table 44: Pareto chart of eigenvalues contributions of UNIT2

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.0399	86.580		86.580
2	0.0037	8.023		94.603
3	0.0017	3.665		98.269
4	0.0007	1.570		99.839
5	0.0000	0.093		99.932
6	0.0000	0.035		99.967
7	0.0000	0.016		99.983
8	0.0000	0.009		99.991
9	0.0000	0.006		99.997
10	0.0000	0.003		100.000
11	0.0000	0.000		100.000
12	0.0000	0.000		100.000

To retain at least 95% of the information contents (i.e. 95% of confidence level), only the first 3 eigenvalues that represent 98.269% is needed. Consequently, the eigenvectors corresponding to those three eigenvalues represent the principal components are chosen as shown on this Table 45

Table 45: PCA weight matrix for UNIT2

Variables	PC1 (85.58%)	PC2 (8.023%)	PC3 (3.665%)
X1	-0.00053	0.05	-0.29211
X2	0.00011	0.00008	-0.00018
X3	0.68666	-0.13455	-0.01526
X4	0.68685	-0.19962	-0.02309
X5	0.23808	0.96099	0.13268
X6	0.00228	0.01872	-0.16403
X7	-0.00275	-0.02153	0.11613
X8	0.00304	0.02187	-0.16519
X9	0	0	-0.00006
X10	-0.00473	-0.12068	0.8959
X11	-0.00031	0.01389	-0.16063
X12	-0.0009	-0.00222	0.01349

The entries of the eigenvectors are called the weights that will be applied to their corresponding variables to obtain the principal component that is the combination of the 8 variables.

At this point, the PPCA procedure will be implemented. First, the major weight contributors are identified to get the maximum likelihood of weight that is any entry value of an eigenvector less than 0.1 is set to 0 ($W_{ML}=0.1$), while any value greater than 0.1 is kept. Consequently, for the PPCA the following principal components are obtained:

Table 46: PPCA Maximum likelihood weight matrix for UNIT2

Variables	PC1 (85.58%)	PC2 (8.023%)	PC3 (3.665%)
X1	0	0	-0.29211
X2	0	0	0
X3	0.68666	-0.13455	0
X4	0.68685	-0.19962	0
X5	0.23808	0.96099	0.13268
X6	0	0	-0.16403
X7	0	0	0.11613
X8	0	0	-0.16519
X9	0	0	0
X10	0	-0.12068	0.8959
X11	0	0	-0.16063
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

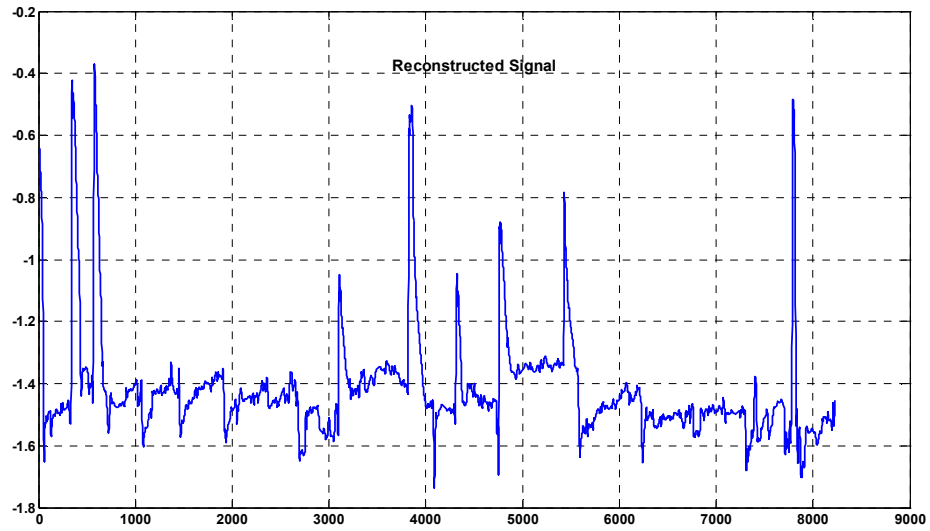


Figure 112: Reconstructed 1-dimensional Signal for Unit 2

- Monitoring of Damage indicator

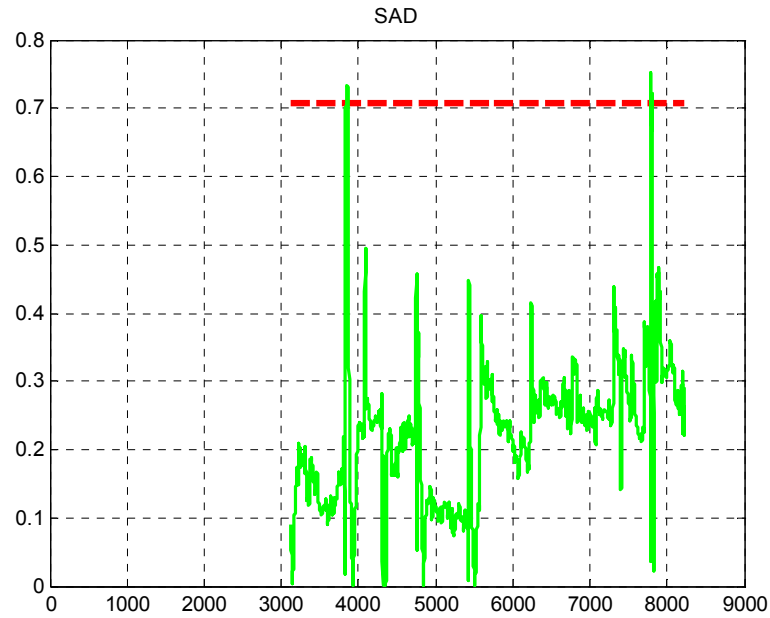


Figure 113: Monitoring of damage indicator SAD for unit 2

- Residual time to failure data table

Table 47: Actual first anomaly to be used for PHM model for unit 2

Anomaly	ASI	ADI	RTTF
1	1.036524	0.216667	332.15

- Residual time to failure extended data table

Table 48: Actual extended data set representing all anomalies for PHM model for unit 2

Anomaly	ASI	ADI	RTTF
1	1.036524	0.216667	332.15

A.3- Test unit 3

Multi-resolution Analysis: Energy Content of each Sensor

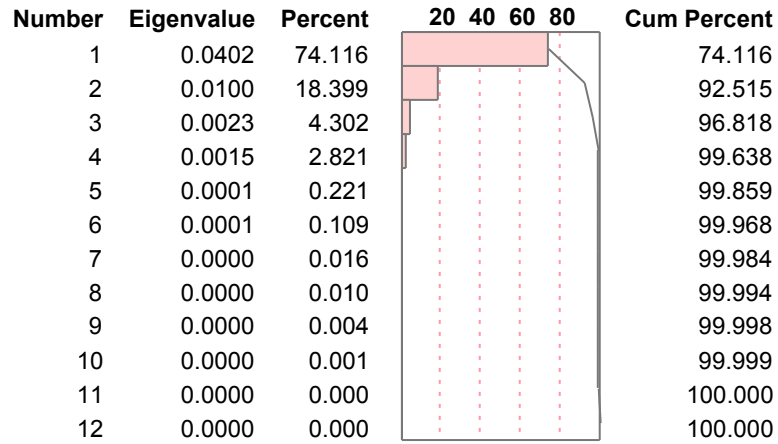
Table 49: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 3

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.837	99.995	99.953	100	100	100	100	99.998	100	100
Sum of details (% of Original)	7.52E-04	4.92E-06	1.63E-01	5.00E-03	4.70E-03	8.07E-05	5.21E-05	6.88E-05	9.88E-06	1.20E-03	5.86E-05	3.32E-07

PPCA data fusion

Eigenvalues

Table 50: Pareto chart of eigenvalues contributions of UNIT3



PCA: Eigenvectors

Table 51: PCA weight matrix for UNIT3

Variables	PC1 (74.11%)	PC2 (18.39%)	PC3 (4.302%)
X1	0.00292	0.04306	-0.11075
X2	-0.00061	0.00009	0.00271
X3	0.98394	0.15459	-0.06909
X4	0.09968	-0.07233	0.84134
X5	-0.1467	0.97058	0.16285
X6	-0.00316	0.03165	-0.09364
X7	-0.00405	-0.01844	0.044
X8	0.00203	0.03152	-0.07455
X9	0.00014	0.00016	0.00027
X10	0.01488	-0.15314	0.46741
X11	-0.01147	0.03482	-0.11652
X12	-0.00403	-0.00046	-0.01627

- PPCA (W_{ML} ≥ 0.1)

Table 52: PPCA Maximum likelihood weight matrix for UNIT3

Variables	PC1 (74.11%)	PC2 (18.39%)	PC3 (4.302%)
X1	0	0	-0.11075
X2	0	0	0
X3	0.98394	0.15459	0
X4	0	0	0.84134
X5	-0.1467	0.97058	0.16285
X6	0	0	0
X7	0	0	0
X8	0	0	0
X9	0	0	0
X10	0	-0.15314	0.46741
X11	0	0	-0.11652
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

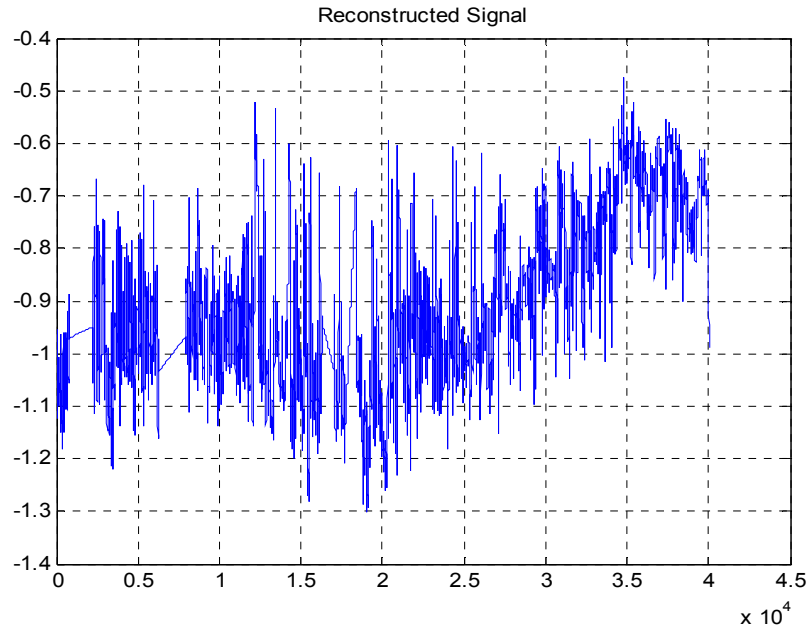


Figure 114: Reconstructed 1-dimensional Signal for Unit 3

- Monitoring of Damage indicator

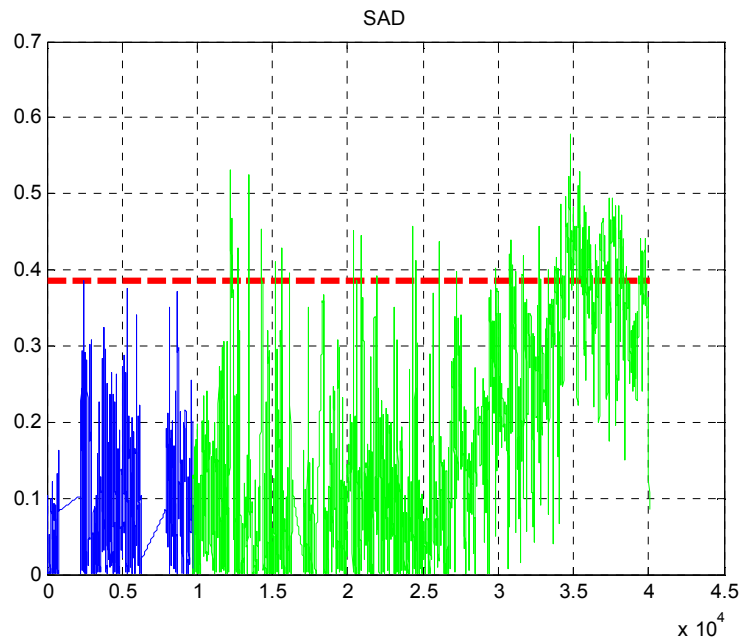


Figure 115: Monitoring of damage indicator SAD for unit 3

- Residual time to failure data table

Table 53: Actual first anomaly to be used for PHM model for unit 3

Anomaly	ASI	ADI	RTTF
	1.282236	3.6	232.1167

- Residual time to failure extended data set table

Table 54: Actual extended data set representing all anomalies for PHM model for unit 3

Anomaly	ASI	ADI	RTTF
	1.18051	0.166667	469.4833
	1.132413	0.2	438.5667
	1.137	0.116667	353.2667
	1.183305	0.133333	284.4167
	1.258377	0.366667	252.65
1	1.282236	3.6	232.1167
	1.496119	2.133333	212.0167
	1.246279	3.766667	208.0167
2	1.369278	2.266667	201.6333
	1.252208	2.016667	182.45
	1.151164	2.15	160.9333
	1.208265	4.65	113.6667
	1.279143	4.766667	64.81667
	1.218381	2.283333	41.98333
	1.140144	1.6	17.11667

A.4- Test unit 4

Multi-resolution Analysis: Energy Content of each Sensor

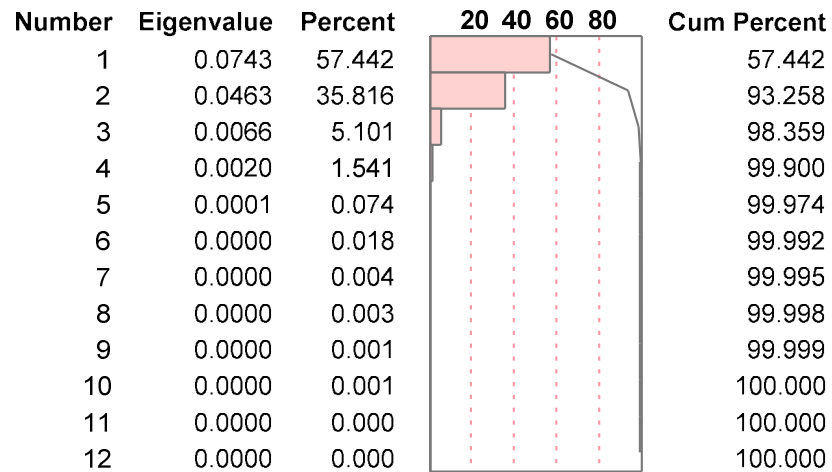
Table 55: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 4

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.92	99.93	99.86	99.999	99.99	99.999	100	99.99	99.99	100
Sum of details (% of Original)	6.56E-04	1.55E-07	8.02E-02	6.91E-02	1.35E-01	2.52E-04	1.40E-04	2.35E-04	5.14E-05	9.30E-03	2.34E-04	5.98E-06

PPCA data fusion

- Eigenvalues table:

Table 56: Pareto chart of eigenvalues contributions of UNIT4



- PCA: Eigenvectors

Table 57: PCA weight matrix for UNIT4

Variables	PC1 (57.44%)	PC2 (35.82%)	PC3 (5.101%)
X1	-0.02597	-0.00802	-0.23687
X2	0.00006	-0.00007	0.00014
X3	0.18902	0.733	0.05655
X4	0.14329	0.63699	-0.10909
X5	0.96947	-0.23804	-0.05664
X6	-0.01425	0.00014	-0.15419
X7	0.00987	0.00767	0.11073
X8	-0.0085	-0.00033	-0.13352
X9	0.00013	0.00008	0.00003
X10	0.05136	0.01252	0.92201
X11	-0.0127	-0.00168	-0.14664
X12	-0.00527	0.00051	-0.01109

- PPCA (abs(W_ML) \geq 0.1)

Table 58: PPCA Maximum likelihood weight matrix for UNIT4

Variables	PC1 (57.44%)	PC2 (35.82%)	PC3 (5.101%)
X1	0	0	-0.23687
X2	0	0	0
X3	0.18902	0.733	0
X4	0.14329	0.63699	-0.10909
X5	0.96947	-0.23804	0
X6	0	0	-0.15419
X7	0	0	0.11073
X8	0	0	-0.13352
X9	0	0	0
X10	0	0	0.92201
X11	0	0	-0.14664
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

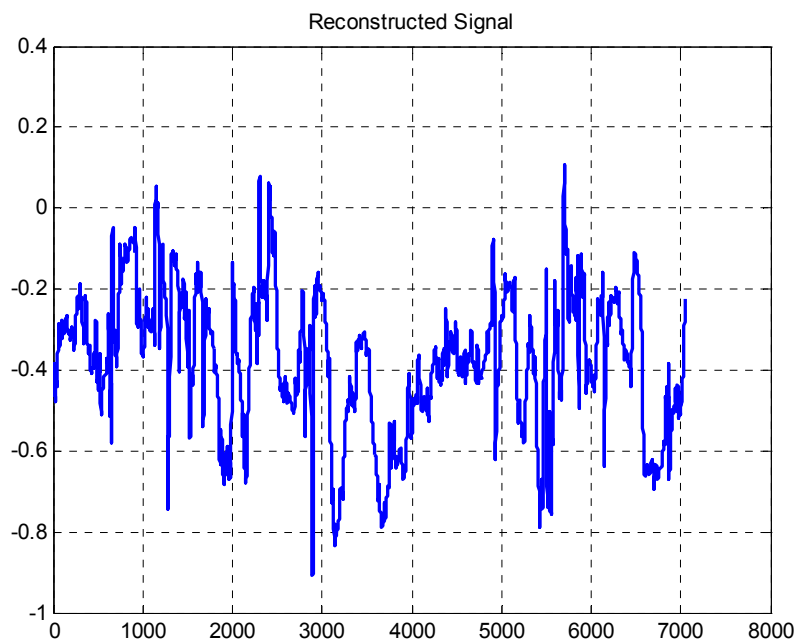


Figure 116: Reconstructed 1-dimensional Signal for Unit 4

- Monitoring of Damage indicator

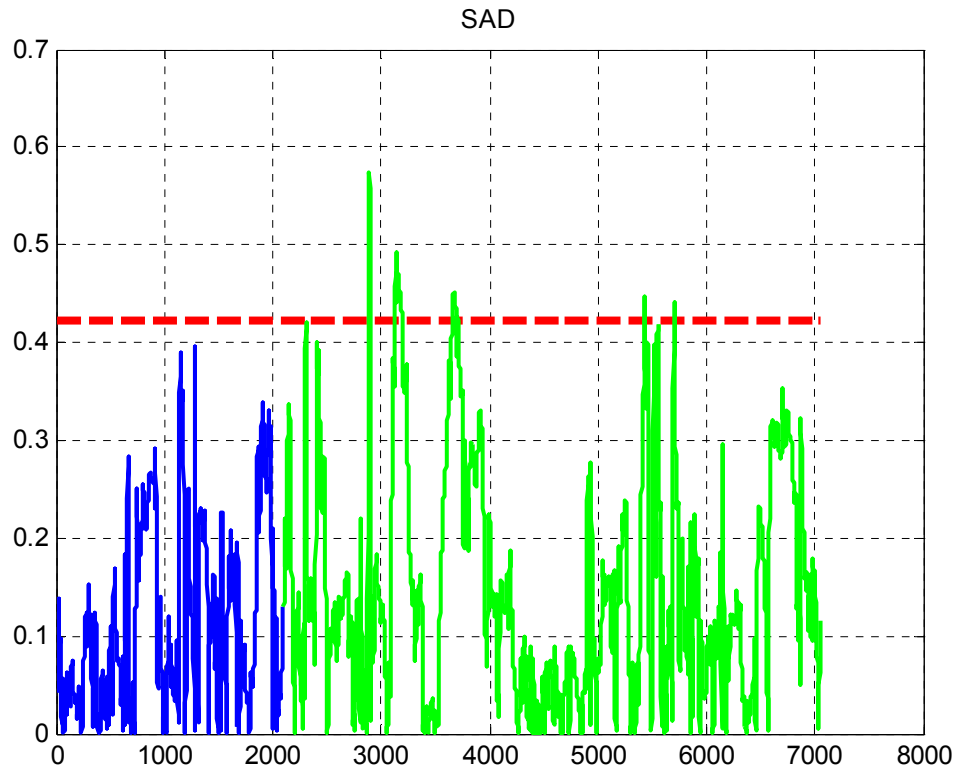


Figure 117: Monitoring of damage indicator SAD for unit 4

- Residual time to failure data table

Table 59: Actual first anomaly to be used for PHM model for unit 4

Anomaly	ASI	ADI	RTTF
	1.165099	1	339.6833

- Residual time to failure extended data set table

Table 60: Actual extended data set representing all anomalies for PHM model for unit 4

Anomaly	ASI	ADI	RTTF
3	1.359802	0.3	359.9
4	1.165099	1	339.6833
5	1.065159	0.6	318.9833

A.5- Test unit 5

Unit 5 has eleven monitored sensors. The sensor X11 (Compressor effectiveness health parameter 2) did not record properly.

Table 61: Monitored sensors for unit 5

Variable	Variables Description
X1	Overall system health parameter 1
X2	Overall system health parameter 1
X3	Compressor seismic vibration 2
X4	Compressor seismic vibration 2
X5	Turbine seismic vibration 3
X6	Compressor health parameter 1
X7	Compressor health parameter 2
X8	Compressor effectiveness health parameter 1
X9	operating condition (load)
X10	operation condition 1 (environment)
X12	operation condition 2 (environment)

Multi-resolution Analysis: Energy Content of each Sensor

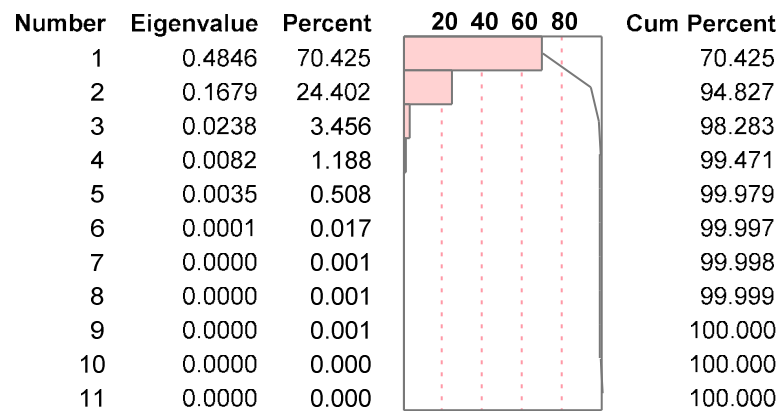
Table 62: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 5

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X12
Approximation (% of Original)	99.976	99.984	99.491	99.417	99.131	99.979	99.982	99.982	100	99.988	100
Sum of details (% of Original)	2.41E-02	1.57E-02	5.09E-01	5.83E+03	8.69E+03	2.08E-02	1.78E-02	1.81E-02	2.56E-05	1.17E-02	7.87E-06

PPCA data fusion

- Eigenvalues table:

Table 63: Pareto chart of eigenvalues contributions of UNIT 5



- PCA: Eigenvectors

Table 64: PCA weight matrix for UNIT 5

Variables	PC1 (70.43%)	PC2 (24.40%)	PC3 (3.456%)
X1	0.1897	0.34279	0.28134
X2	0.16644	0.28797	0.22198
X3	0.60303	-0.30941	-0.03406
X4	0.63552	-0.36886	0.0079
X5	0.26122	0.5137	-0.80535
X6	0.19144	0.33751	0.28019
X7	0.17847	0.29981	0.21548
X8	0.1778	0.31576	0.26381
X9	-0.00004	0.00007	-0.00046
X10	0.02073	-0.01968	-0.16502
X12	0.00086	-0.00089	-0.00246

- PPCA (abs(W_ML)>=0.1)

Table 65: PPCA Maximum likelihood weight matrix for UNIT 5

Variables	PC1 (70.43%)	PC2 (24.40%)	PC3 (3.456%)
X1	0.1897	0.34279	0.28134
X2	0.16644	0.28797	0.22198
X3	0.60303	-0.30941	0
X4	0.63552	-0.36886	0
X5	0.26122	0.5137	-0.80535
X6	0.19144	0.33751	0.28019
X7	0.17847	0.29981	0.21548
X8	0.1778	0.31576	0.26381
X9	0	0	0
X10	0	0	-0.16502
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

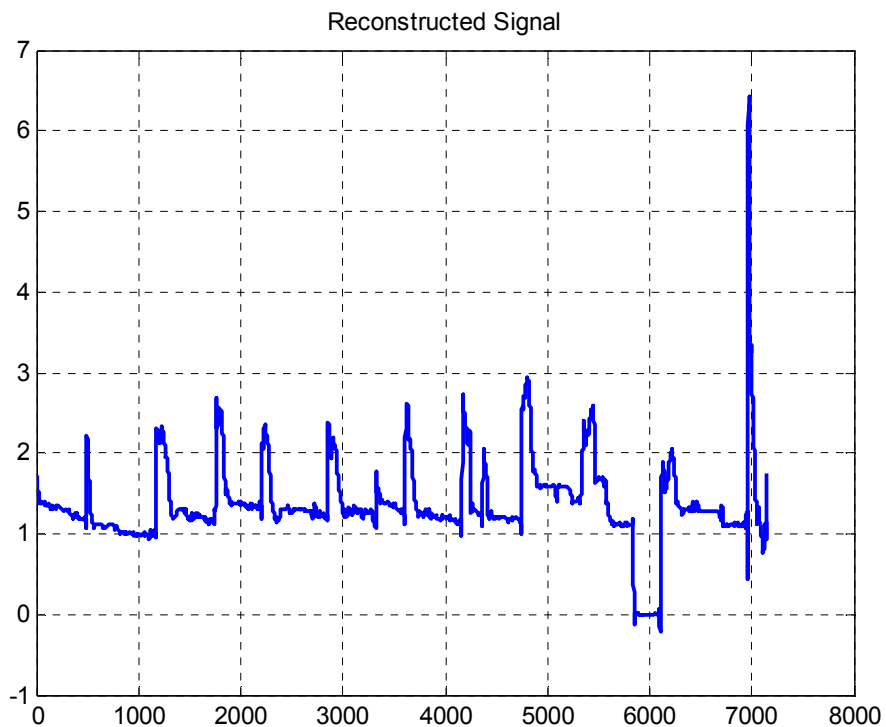


Figure 118: Reconstructed 1-dimensional Signal for Unit 5

- Monitoring of Damage indicator

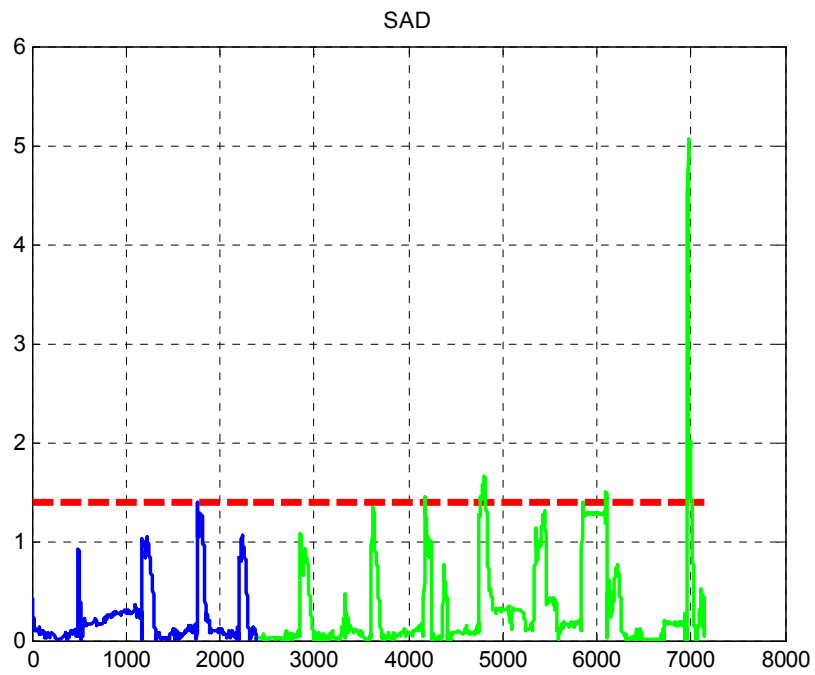


Figure 119: Monitoring of damage indicator SAD for unit 5

- Residual time to failure data table

Table 66: Actual first anomaly to be used for PHM model for unit 5

Anomaly	ASI	ADI	RTTF
	1.184042	1.033333	578.2833

- Residual time to failure extended data set table

Table 67: Actual extended data set representing all anomalies for PHM model for unit 5

Anomaly	ASI	ADI	RTTF
3	1.036482	0.083333	675.4333
4	1.184042	1.033333	578.2833
5	3.623064	0.716667	37.31667

A.6- Test unit 6

Multi-resolution Analysis: Energy Content of each Sensor

Table 68: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 6

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.096	99.2714	99.944	99.999	99.999	99.994	100	99.991	99.999	100
Sum of details (% of Original)	1.10E-03	4.94E-06	9.04E-01	7.29E-01	5.60E-02	6.45E-04	1.92E-04	5.30E-03	1.71E-05	9.10E-03	4.49E-04	2.67E-05

PPCA data fusion

- Eigenvalues table:

Table 69: Pareto chart of eigenavlues contributions of UNIT 6

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.2796	88.304		88.304
2	0.0281	8.876		97.179
3	0.0058	1.834		99.013
4	0.0030	0.938		99.951
5	0.0001	0.036		99.987
6	0.0000	0.006		99.993
7	0.0000	0.004		99.997
8	0.0000	0.001		99.998
9	0.0000	0.001		99.999
10	0.0000	0.001		100.000
11	0.0000	0.000		100.000
12	0.0000	0.000		100.000

- PCA: Eigenvectors

Table 70: PCA weight matrix for UNIT 6

Variables	PC1 (88.30%)	PC2 (8.87%)
X1	-0.00777	-0.0137
X2	0.00004	-0.0001
X3	0.75108	-0.06791
X4	0.65259	-0.07354
X5	0.09936	0.99447
X6	-0.00455	-0.00323
X7	0.00258	0.01013
X8	-0.00364	-0.00504
X9	0	-0.00005
X10	0.0004	0.02559
X11	-0.00421	-0.00395
X12	-0.00215	0.00225

- PPCA (abs(W_ML) \geq 0.1)

Table 71: PPCA Maximum likelihood weight matrix for UNIT 6

Variables	PC1 (88.30%)	PC2 (8.87%)
X1	0	0
X2	0	0
X3	0.75108	0
X4	0.65259	0
X5	0	0.99447
X6	0	0
X7	0	0
X8	0	0
X9	0	0
X10	0	0
X11	0	0
X12	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

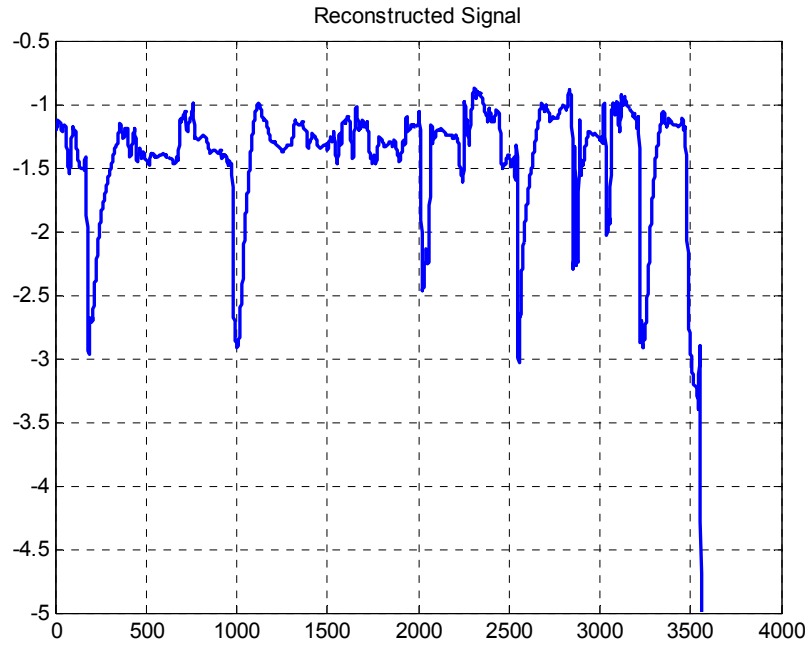


Figure 120: Reconstructed 1-dimensional Signal for Unit 6

- Monitoring of Damage indicator

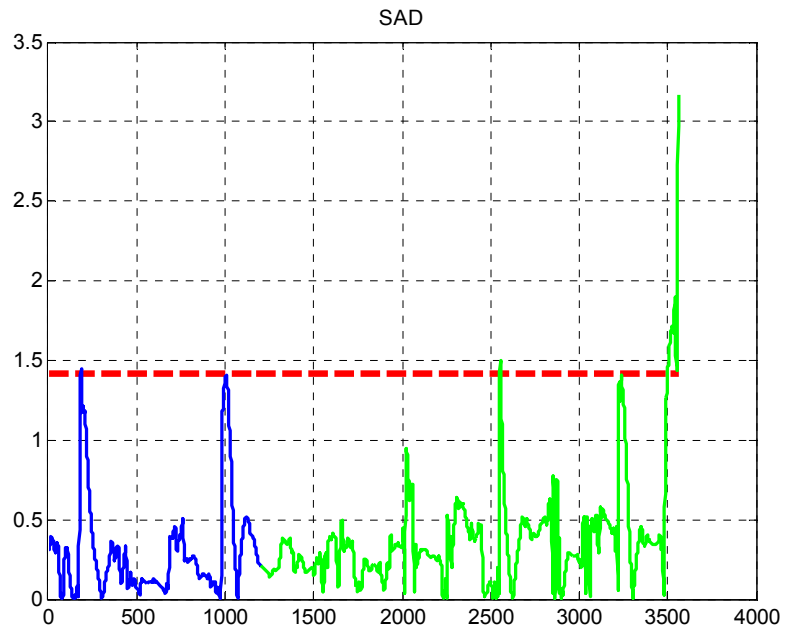


Figure 121: Monitoring of damage indicator SAD for unit 6

- Residual time to failure data table

Table 72: Actual first anomaly to be used for PHM model for unit 6

Anomaly	ASI	ADI	RTTF
	1.062287	0.066667	458.6833

- Residual time to failure extended data set table

Table 73: Actual extended data set representing all anomalies for PHM model for unit 6

Anomaly	ASI	ADI	RTTF
	1.062287	0.066667	458.6833
	2.240944	1.016667	1

A.7- Test unit 7

Unit 7 has eleven monitored sensors. The sensor X11 (Compressor effectiveness health parameter 2) did not record properly.

Table 74: Monitored sensors for unit 7

Variable	Variables Description
X1	Overall system health parameter 1
X2	Overall system health parameter 1
X3	Compressor seismic vibration 2
X4	Compressor seismic vibration 2
X5	Turbine seismic vibration 3
X6	Compressor health parameter 1
X7	Compressor health parameter 2
X8	Compressor effectiveness health parameter 1
X9	operating condition (load)
X10	operation condition 1 (environment)
X12	operation condition 2 (environment)

Multi-resolution Analysis: Energy Content of each Sensor

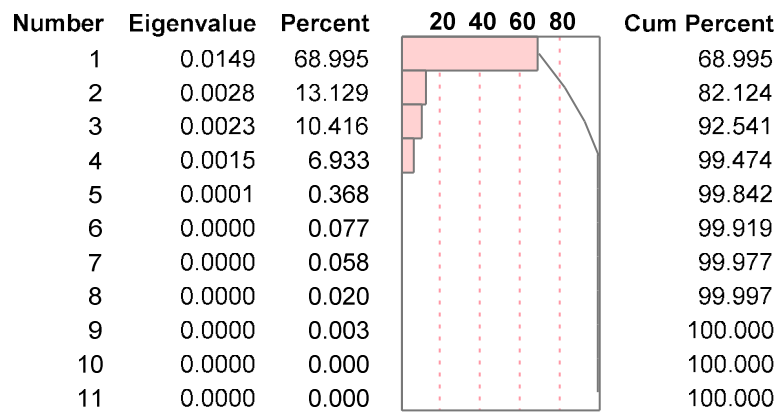
Table 75: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X12
Approximation (% of Original)	99.999	100	99.951	99.949	99.979	99.999	99.999	99.999	100	99.996	100
Sum of details (% of Original)	6.78E-04	1.08E-06	4.87E-02	5.02E-02	2.08E-02	2.34E-04	1.73E-04	2.20E-04	1.59E-06	4.00E-03	2.40E-06

PPCA data fusion

- Eigenvalues table:

Table 76: Pareto chart of eigenvalues contributions of UNIT 7



- PCA: Eigenvectors

Table 77: PCA weight matrix for UNIT 7

Variables	PC1 (68.99%)	PC2 (13.13%)	PC3 (10.42%)	PC4 (6.933%)
X1	0.02585	-0.11531	-0.07147	-0.21823
X2	-0.00037	0.00018	0.0004	-0.00016
X3	0.65133	0.71098	-0.12597	-0.23312
X4	0.75635	-0.59349	0.07415	0.26235
X5	0.03799	0.04331	0.95795	-0.27949
X6	0.01954	-0.07014	-0.05275	-0.12858
X7	-0.01022	0.04424	0.03495	0.10793
X8	0.00888	-0.04667	-0.04843	-0.10464
X9	-0.00003	-0.00001	0.00007	0.00002
X10	-0.03041	0.34311	0.22248	0.84368
X12	0.01037	-0.01821	0.00033	-0.01271

- PPCA (abs(W_ML)>=0.1)

Table 78: PPCA Maximum likelihood weight matrix for UNIT 7

Variables	PC1 (68.99%)	PC2 (13.13%)	PC3 (10.42%)	PC4 (6.933%)
X1	0	-0.11531	0	-0.21823
X2	0	0	0	0
X3	0.65133	0.71098	-0.12597	-0.23312
X4	0.75635	-0.59349	0	0.26235
X5	0	0	0.95795	-0.27949
X6	0	0	0	-0.12858
X7	0	0	0	0.10793
X8	0	0	0	-0.10464
X9	0	0	0	0
X10	0	0.34311	0.22248	0.84368
X12	0	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

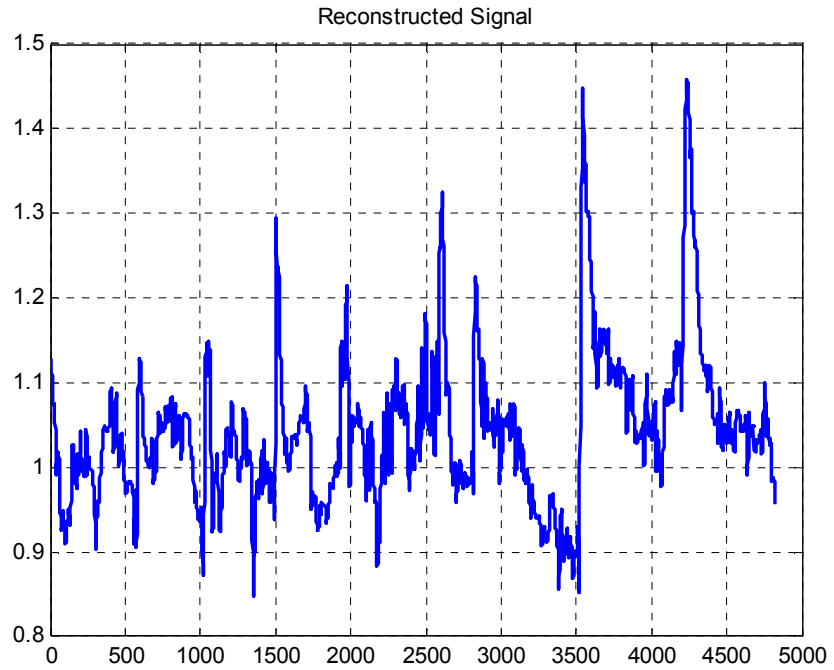


Figure 122: Reconstructed 1-dimensional Signal for Unit 7

- Monitoring of Damage indicator

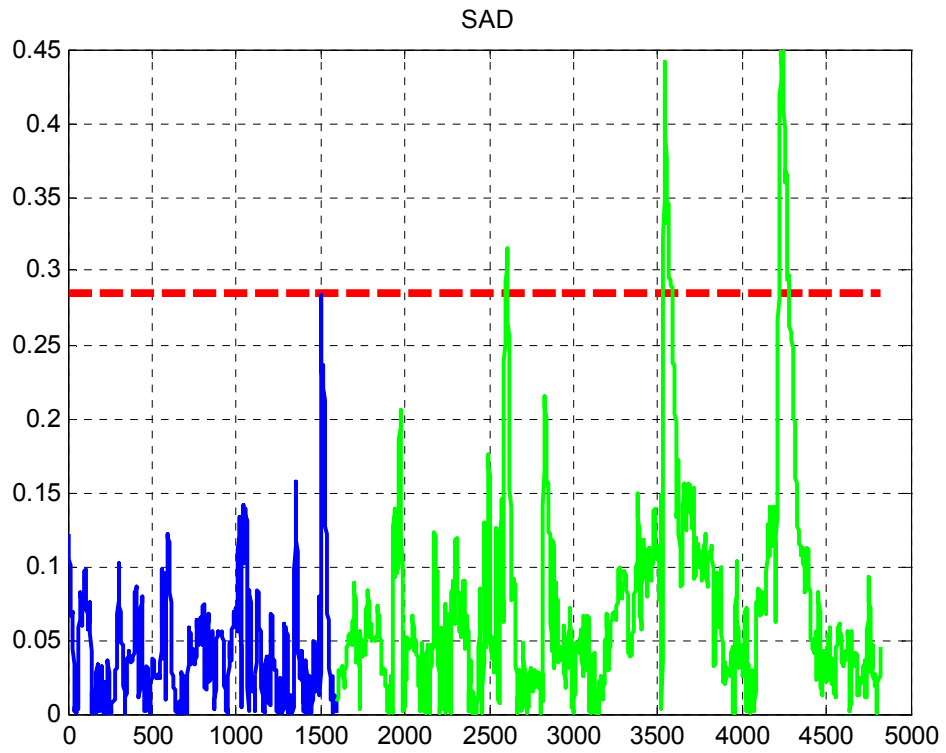


Figure 123: Monitoring of damage indicator SAD for unit 7

- Residual time to failure data table

Table 79: Actual first anomaly to be used for PHM model for unit 7

Anomaly	ASI	ADI	RTTF
	1.551326	0.833333	35.2

- Residual time to failure extended data set table

Table 80: Actual extended data set representing all anomalies for PHM model for unit 7

Anomaly	ASI	ADI	RTTF
	1.106504	0.166667	85.98333
	1.551326	0.833333	35.2
	1.580428	1.066667	11.05

A.8- Test unit 8

Multi-resolution Analysis: Energy Content of each Sensor

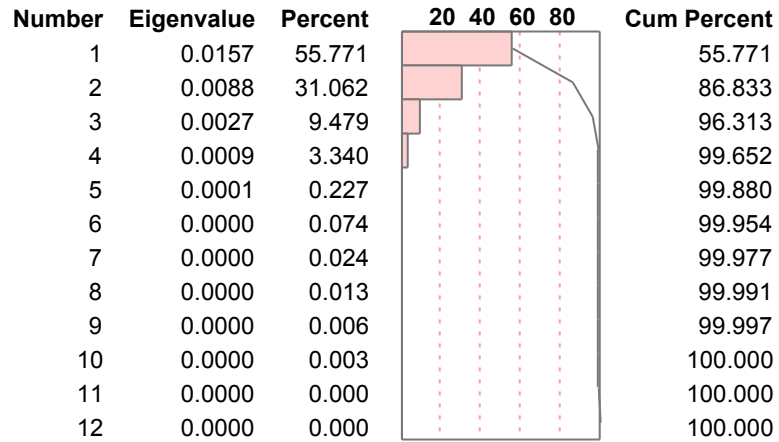
Table 81: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 8

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.953	99.974	99.966	100	100	100	100	99.998	100	100
Sum of details (% of Original)	2.88E-04	1.26E-06	4.72E-02	2.56E-02	3.44E-02	9.58E-05	8.33E-05	1.08E-04	4.68E-06	2.20E-03	1.05E-04	4.10E-04

PPCA data fusion

- Eigenvalues table:

Table 82: Pareto chart of eigenvalues contributions of UNIT 8



- PCA: Eigenvectors

Table 83: PCA weight matrix for UNIT 8

Variables	PC1 (55.77%)	PC2 (31.06%)	PC23 (9.48%)
X1	-0.0238	0.01697	-0.00237
X2	-0.00016	-0.00011	-0.00009
X3	0.64911	0.36963	-0.66389
X4	0.58148	0.31655	0.74687
X5	-0.48381	0.87318	0.01024
X6	-0.01524	0.00621	-0.00467
X7	0.01185	-0.00543	0.01085
X8	-0.02172	0.01416	0.01098
X9	-0.00001	-0.00003	0.00011
X10	0.06941	-0.00756	0.02525
X11	-0.01085	0.00326	-0.00453
X12	0.01092	-0.01033	-0.02003

- PPCA (abs(W_ML)>=0.1)

Table 84: PPCA Maximum likelihood weight matrix for UNIT 8

Variables	PC1 (55.77%)	PC2 (31.06%)	PC23 (9.48%)
X1	0	0	0
X2	0	0	0
X3	0.64911	0.36963	-0.66389
X4	0.58148	0.31655	0.74687
X5	-0.48381	0.87318	0
X6	0	0	0
X7	0	0	0
X8	0	0	0
X9	0	0	0
X10	0	0	0
X11	0	0	0
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

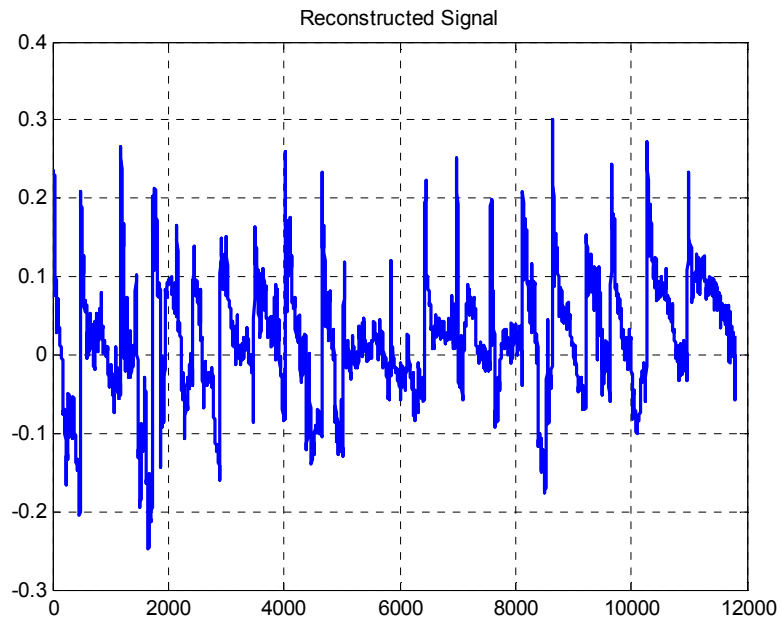


Figure 124: Reconstructed 1-dimensional Signal for Unit 8

- Monitoring of Damage indicator

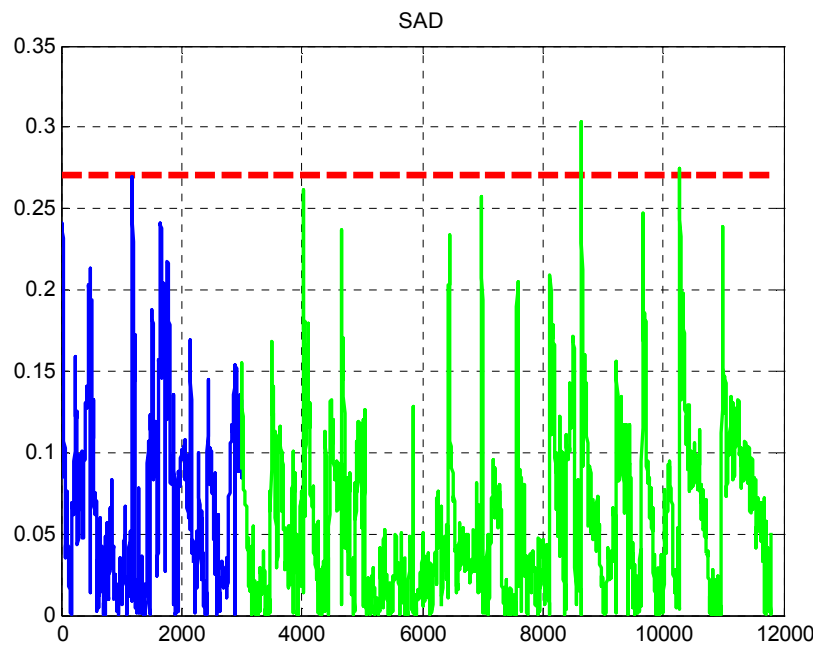


Figure 125: Monitoring of damage indicator SAD for unit 8

- Residual time to failure data table

Table 85: Actual first anomaly to be used for PHM model for unit 8

Anomaly	ASI	ADI	RTTF
	1.12047	0.183333	119.5167

- Residual time to failure extended data set table

Table 86: Actual extended data set representing all anomalies for PHM model for unit 8

Anomaly	ASI	ADI	RTTF
	1.12047	0.183333	119.5167
	1.016824	0.05	49.38333

A.9- Test unit 9

Multi-resolution Analysis: Energy Content of each Sensor

Table 87: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 9

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	100	100	99.92	99.946	99.885	100	100	100	100	99.999	100	100
Sum of details (% of Original)	4.52E-05	1.65E-06	7.52E-02	5.40E-02	1.15E-01	1.81E-05	1.45E-05	1.63E-05	3.15E-06	1.10E-03	9.89E-06	3.75E-05

PPCA data fusion

- Eigenvalues table:

Table 88: Pareto chart of eigenvalues contributions of UNIT 9

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	0.0848	64.734		64.734
2	0.0430	32.786		97.520
3	0.0018	1.339		98.859
4	0.0014	1.091		99.950
5	0.0001	0.038		99.988
6	0.0000	0.005		99.993
7	0.0000	0.003		99.996
8	0.0000	0.002		99.998
9	0.0000	0.001		99.999
10	0.0000	0.001		100.000
11	0.0000	0.000		100.000
12	0.0000	0.000		100.000

- PCA: Eigenvectors

Table 89: PCA weight matrix for UNIT 9

Variables	PC1 (64.73%)	PC2 (32.79%)
X1	-0.00443	-0.03771
X2	-0.00001	0.0001
X3	-0.10883	0.70891
X4	-0.08136	0.68477
X5	0.99044	0.13562
X6	-0.00191	-0.02272
X7	-0.00088	0.01143
X8	-0.00183	-0.02298
X9	-0.00001	0
X10	-0.02222	0.08342
X11	-0.00617	-0.02416
X12	0.00033	0.00141

- PPCA (abs(W_ML)>=0.1)

Table 90: PPCA Maximum likelihood weight matrix for UNIT 9

Variables	PC1 (64.73%)	PC2 (32.79%)
X1	0	0
X2	0	0
X3	-0.10883	0.70891
X4	0	0.68477
X5	0.99044	0.13562
X6	0	0
X7	0	0
X8	0	0
X9	0	0
X10	0	0
X11	0	0
X12	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

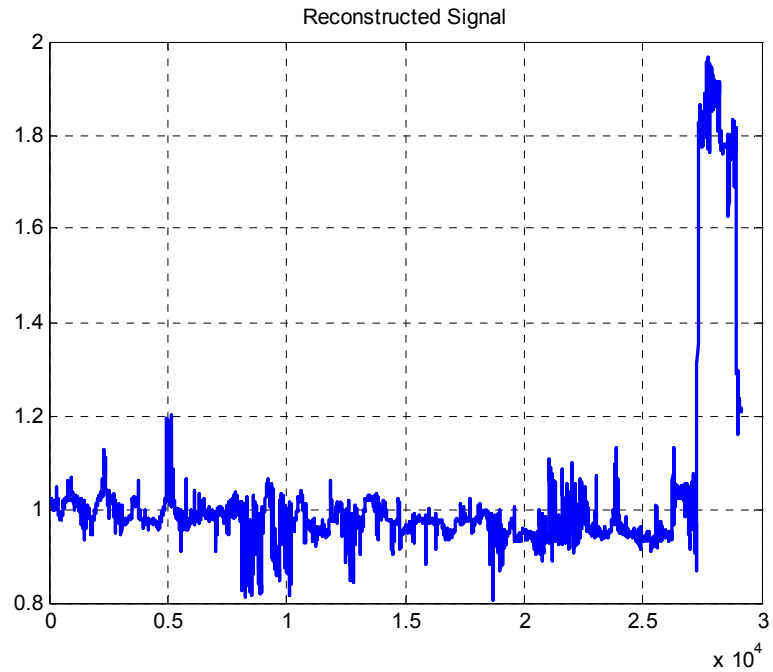


Figure 126: Reconstructed 1-dimensional Signal for Unit 9

- Monitoring of Damage indicator

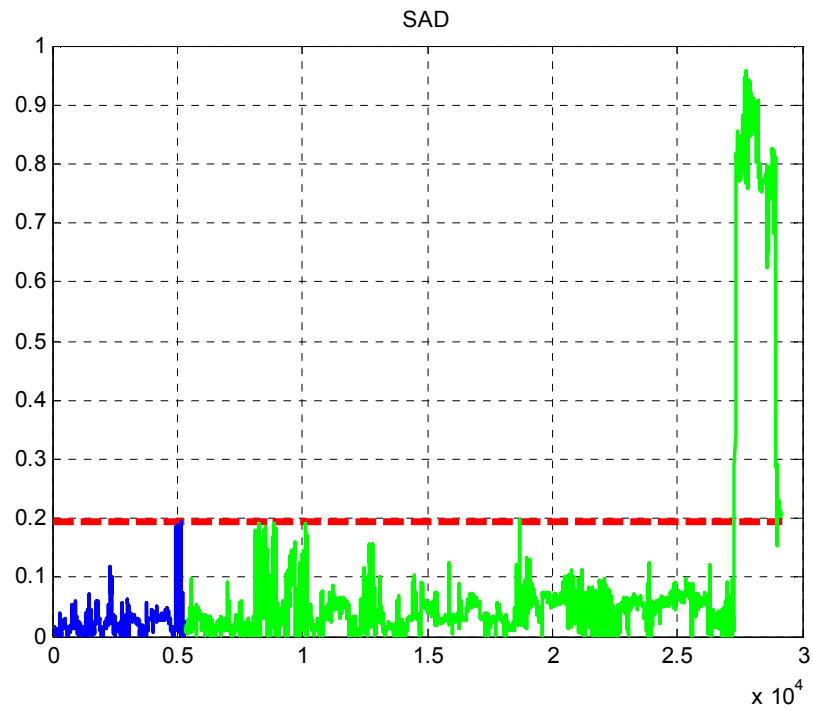


Figure 127: Monitoring of damage indicator SAD for unit 9

- Residual time to failure data table

Table 91: Actual first anomaly to be used for PHM model for unit 9

Anomaly	ASI	ADI	RTTF
	1.130492	0.233333	3.183333

- Residual time to failure extended data set table

Table 92: Actual extended data set representing all anomalies for PHM model for unit 9

Anomaly	ASI	ADI	RTTF
	1.130492	0.233333	3.183333
	1.18496	2.8	2.783333

A.10- Test unit 10

Multi-resolution Analysis: Energy Content of each Sensor

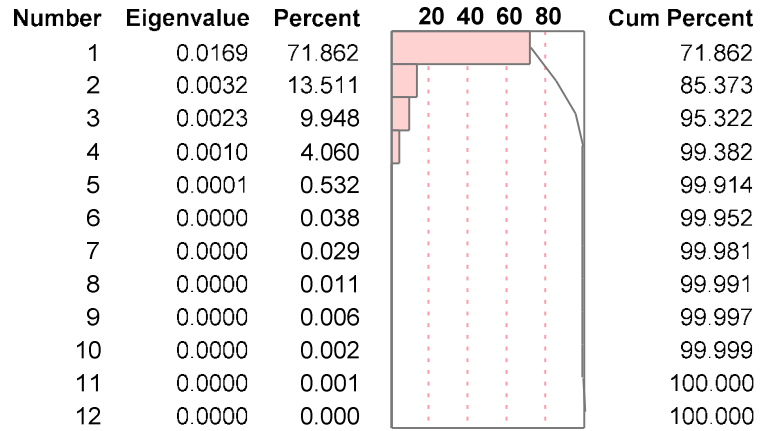
Table 93: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 10

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.991	99.994	99.99	99.999	100	99.999	100	99.998	99.999	100
Sum of details (% of Original)	1.66E-04	7.91E-06	9.10E-03	5.60E-03	1.00E-02	9.08E-05	2.81E-05	7.16E-05	2.06E-06	1.50E-03	6.91E-05	9.80E-07

PPCA data fusion

- Eigenvalues table:

Table 94: Pareto chart of eigenavlues contributions of UNIT 10



- PCA: Eigenvectors

Table 95: PCA weight matrix for UNIT 10

Variables	PC1 (71.86%)	PC2 (13.51%)	PC23 (9.95%)
X1	0.11163	-0.0496	-0.24283
X2	0.00063	0.00001	0.00212
X3	0.53611	0.44731	-0.07759
X4	0.41406	0.5949	0.14483
X5	0.64797	-0.61745	0.44181
X6	0.09165	-0.03798	-0.18695
X7	-0.02942	0.03416	0.08944
X8	0.08733	-0.04653	-0.14856
X9	-0.00002	0.00007	0.0001
X10	-0.29391	0.23822	0.78966
X11	0.07437	-0.0258	-0.17164
X12	-0.00159	0.01089	-0.02809

- PPCA (abs(W_ML)>=0.1)

Table 96: PPCA Maximum likelihood weight matrix for UNIT 10

Variables	PC1 (71.86%)	PC2 (13.51%)	PC23 (9.95%)
X1	0.11163	0	-0.24283
X2	0	0	0
X3	0.53611	0.44731	0
X4	0.41406	0.5949	0.14483
X5	0.64797	-0.61745	0.44181
X6	0	0	-0.18695
X7	0	0	0
X8	0	0	-0.14856
X9	0	0	0
X10	-0.29391	0.23822	0.78966
X11	0	0	-0.17164
X12	0	0	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

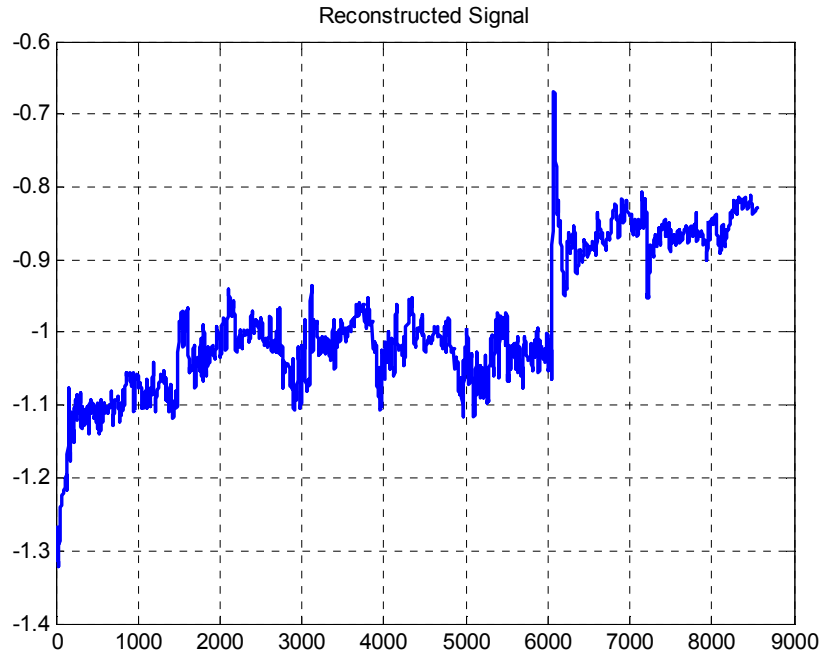


Figure 128: Reconstructed 1-dimensional Signal for Unit 10

- Monitoring of Damage indicator

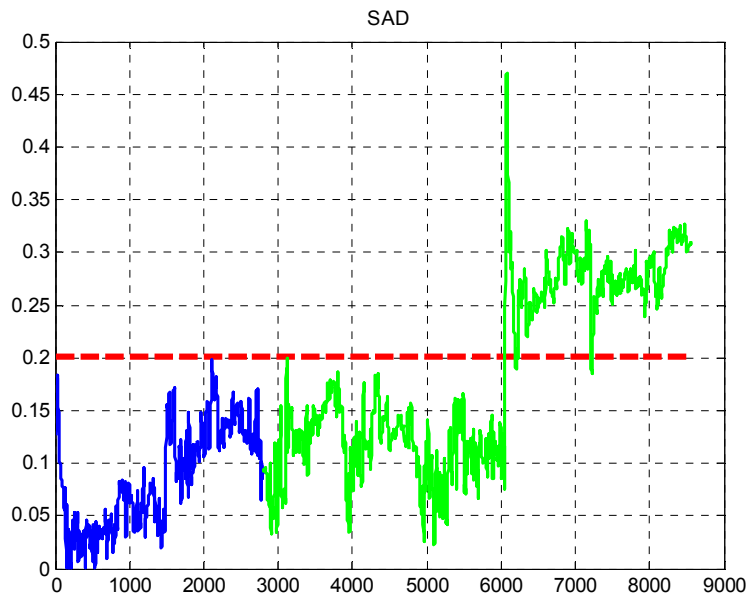


Figure 129: Monitoring of damage indicator SAD for unit 10

- Residual time to failure data table.

Table 97: Actual first anomaly to be used for PHM model for unit 10

Anomaly	ASI	ADI	RTTF
	2.335767	2.366667	85.18333

- Residual time to failure extended data set table

Table 98: Actual extended data set representing all anomalies for PHM model for unit 10

Anomaly	ASI	ADI	RTTF
	2.335767	2.366667	85.18333
	1.637981	16.38333	78.46667
	1.626708	22.18333	22.16667

Note: Unit 10 is considered censored

A.11- Test unit 11

Multi-resolution Analysis: Energy Content of each Sensor

Table 99: Nodes energy content of each sensor at level-3 DWPT decomposition of Unit 11

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of Original)	99.999	100	99.559	99.653	99.7691	99.999	100	99.999	100	99.996	99.999	100
Sum of details (% of Original)	1.74E-04	1.96E-04	4.41E-01	3.47E-01	2.31E-01	8.65E-05	2.53E-05	8.16E-05	1.50E-07	3.90E-03	6.84E-05	9.15E-06

PPCA data fusion

- Eigenvalues table:

Table 100: Pareto chart of eigenvalues contributions of UNIT 11

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	1.1500	97.211		97.211
2	0.0249	2.108		99.319
3	0.0064	0.539		99.858
4	0.0015	0.127		99.984
5	0.0002	0.014		99.998
6	0.0000	0.001		99.999
7	0.0000	0.000		100.000
8	0.0000	0.000		100.000
9	0.0000	0.000		100.000
10	0.0000	0.000		100.000
11	0.0000	0.000		100.000
12	0.0000	0.000		100.000

- PCA: Eigenvectors

Table 101: PCA weight matrix for UNIT 11

Variables	PC1 (97.21%)
X1	0.00247
X2	-0.00001
X3	0.76441
X4	0.64436
X5	0.02129
X6	0.0027
X7	0.00103
X8	0.00181
X9	-0.00002
X10	-0.00092
X11	0.00225
X12	0.00079

- PPCA (abs(W_{ML}) ≥ 0.1)

Table 102: PPCA Maximum likelihood weight matrix for UNIT 11

Variables	PC1 (97.21%)
X1	0
X2	0
X3	0.76441
X4	0.64436
X5	0
X6	0
X7	0
X8	0
X9	0
X10	0
X11	0
X12	0

Step 6: Anomaly detection decision

- 1-d Reconstructed signal

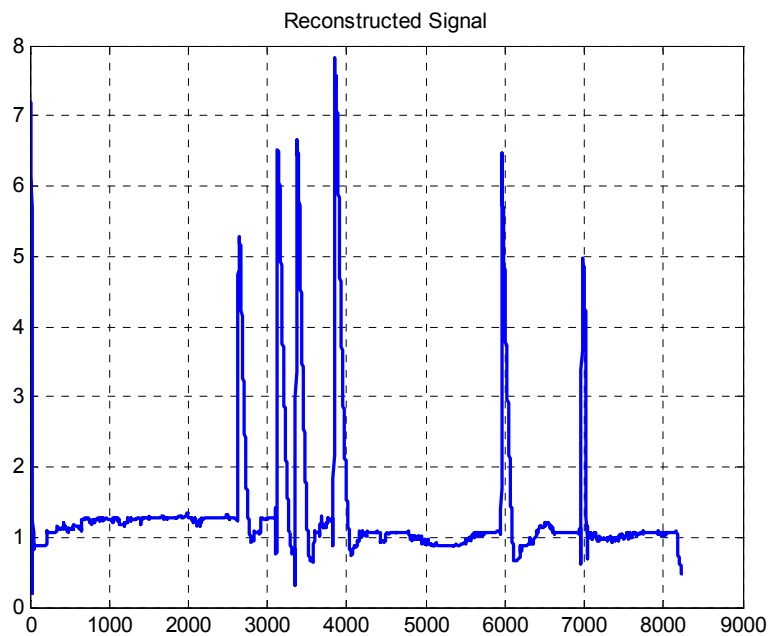


Figure 130: Reconstructed 1-dimensional Signal for Unit 11

- Monitoring of Damage indicator

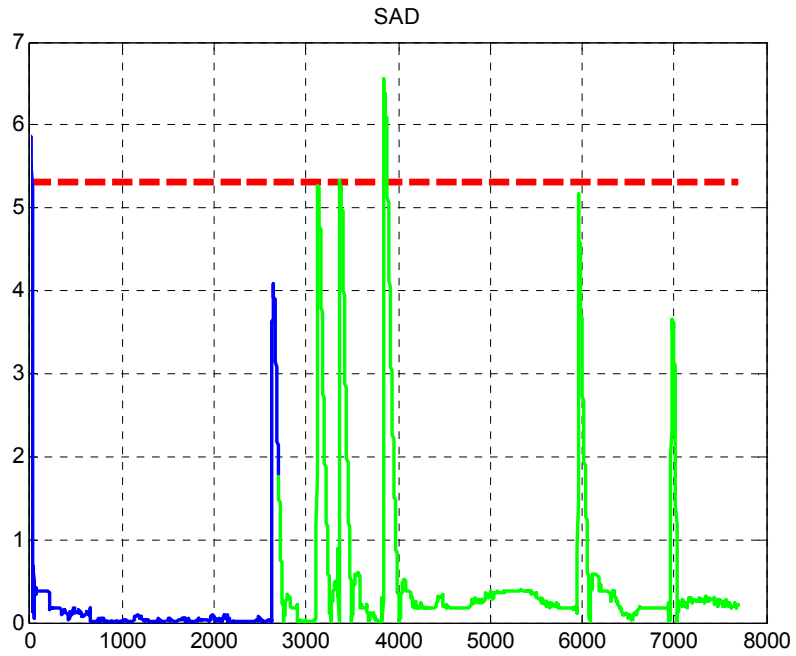


Figure 131: Monitoring of damage indicator SAD for unit 11

- Residual time to failure data table.

Table 103: Actual first anomaly to be used for PHM model for unit 11

Anomaly	ASI	ADI	RTTF
	1.006843	0.033333	124.4

- Residual time to failure extended data set table

Table 104: Actual extended data set representing all anomalies for PHM model for unit 11

Anomaly	ASI	ADI	RTTF
	1.006843	0.033333	124.4
	1.23587	0.683333	48.36667

Note: Unit 11 is right-censored

APPENDIX B- MRA LEVEL DECISION

The test unit 1 is used as an example to assess the effect of decomposition level on the multi-resolution analysis to determine the appropriate level of decomposition. The same de-noised signal used in Chapter 6 is used and decomposed into a two-level, three-level and four-level tree using the DWPT and the “Daubechies 7”. The energy content of the approximation is calculated on one hand and the sum of the energy content of the details nodes is calculated on another hand.

B.1- Level-2 multi-resolution analysis

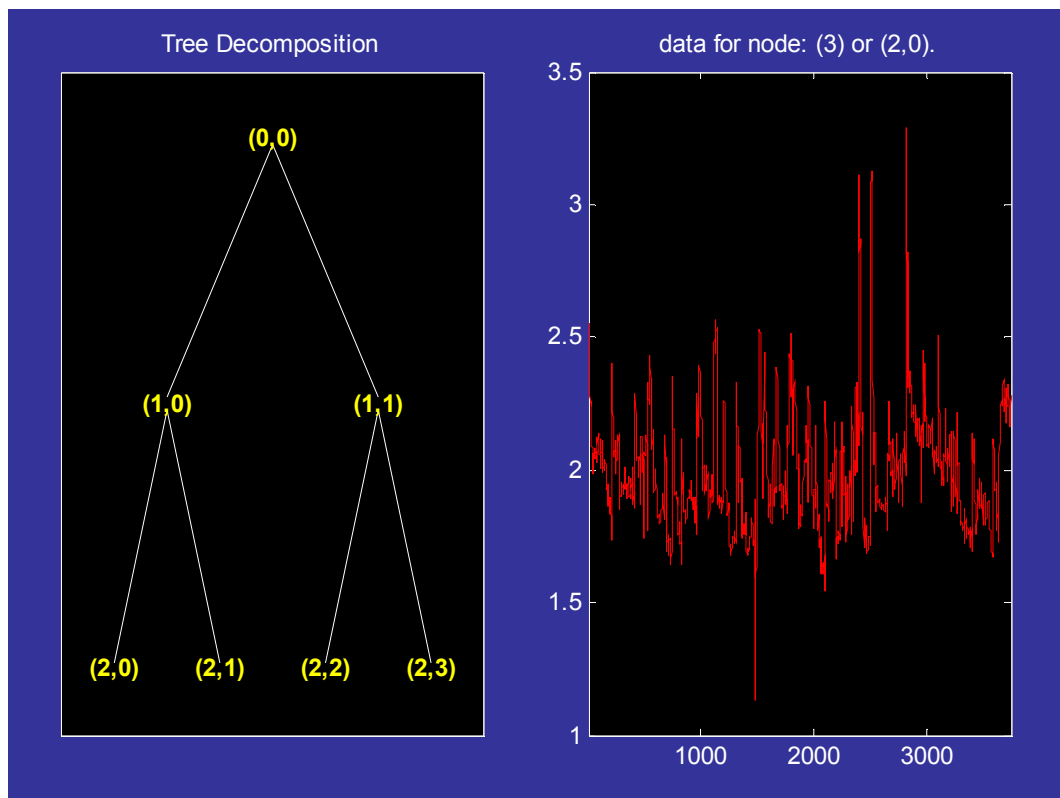


Figure 132: Level-2 tree decomposition of sensor X3 signal with the wavelet component at node 3

Figure 132 is the two level tree decomposition whereas Figure 133 is the multi-resolution analysis plot of the sensor X3 with the original de-noised signal (black), the approximation node signal (red) and the three details (blue).

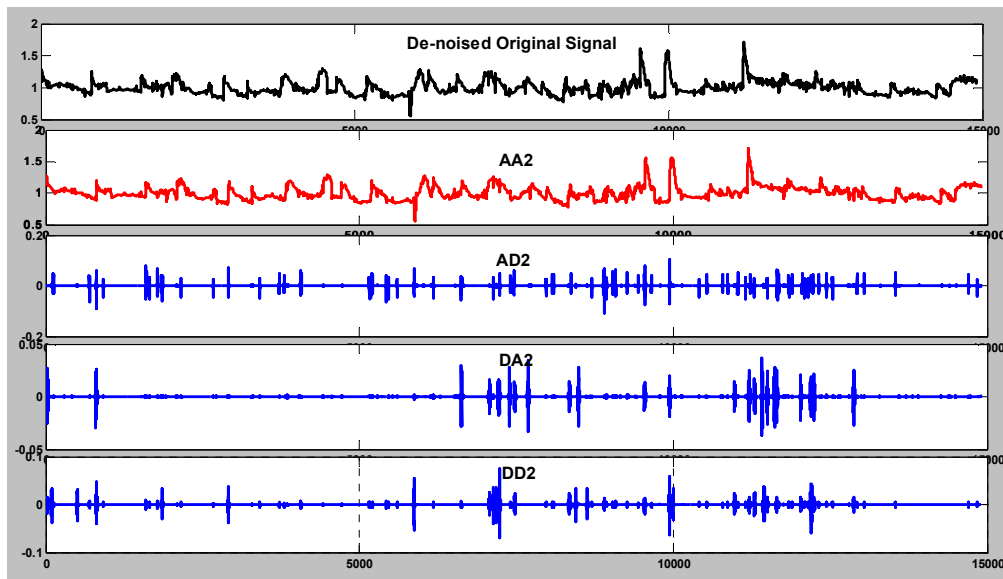


Figure 133: DWPT at level-2 MRA of sensor X3

Table 105: Node energy content of each sensor at level-2 DWPT decomposition

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Aproximation (% of original)	100	100	99.994	99.996	99.994	100	100	100	100	99.9997	100	100
Sum of details (% of original)	1.08E-04	4.23E-07	0.0065	0.0045	0.0062	2.38E-05	8.56E-06	2.10E-05	3.55E-06	2.61E-04	1.19E-05	7.64E-07

Table 105 shows the energy content of the approximation node as well as the sum of the detail nodes of each of the twelve monitored sensors at two-level decomposition.

B.2- Level-3 multi-resolution analysis

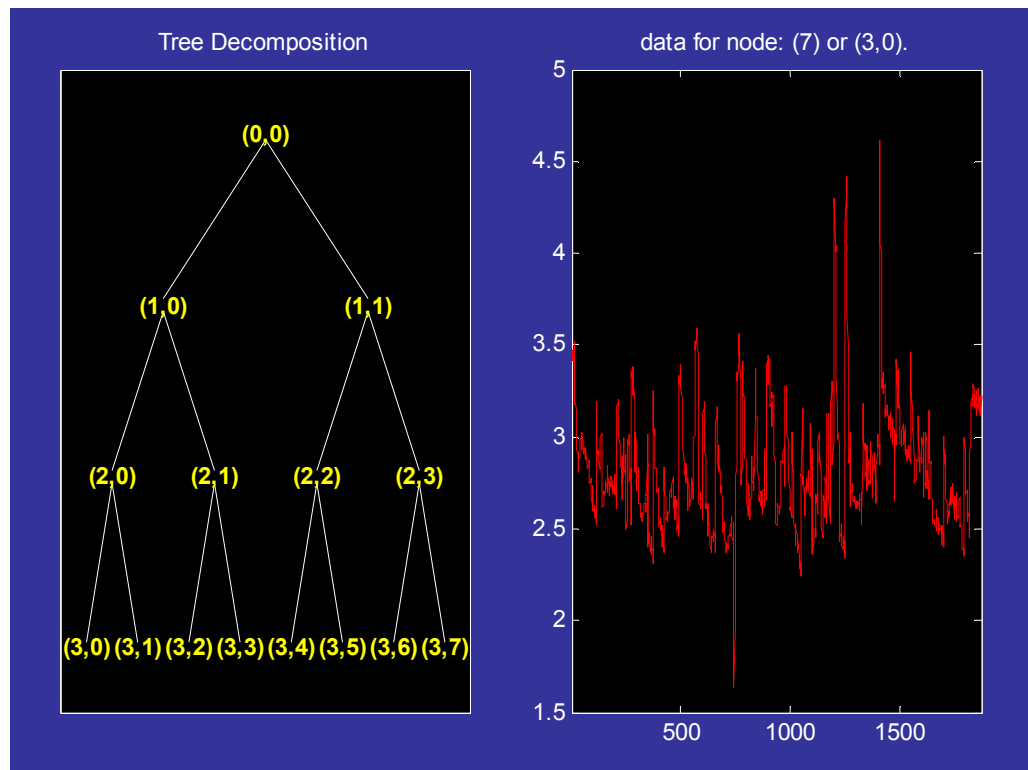


Figure 134: Level-3 tree decomposition of sensor X3 signal with the wavelet component at node 7

Figure 134 is the three level tree decomposition whereas Figure 135 is the multi-resolution analysis plot of the sensor X3 with the original de-noised signal (black), the approximation node signal (red) and the seven details (blue).

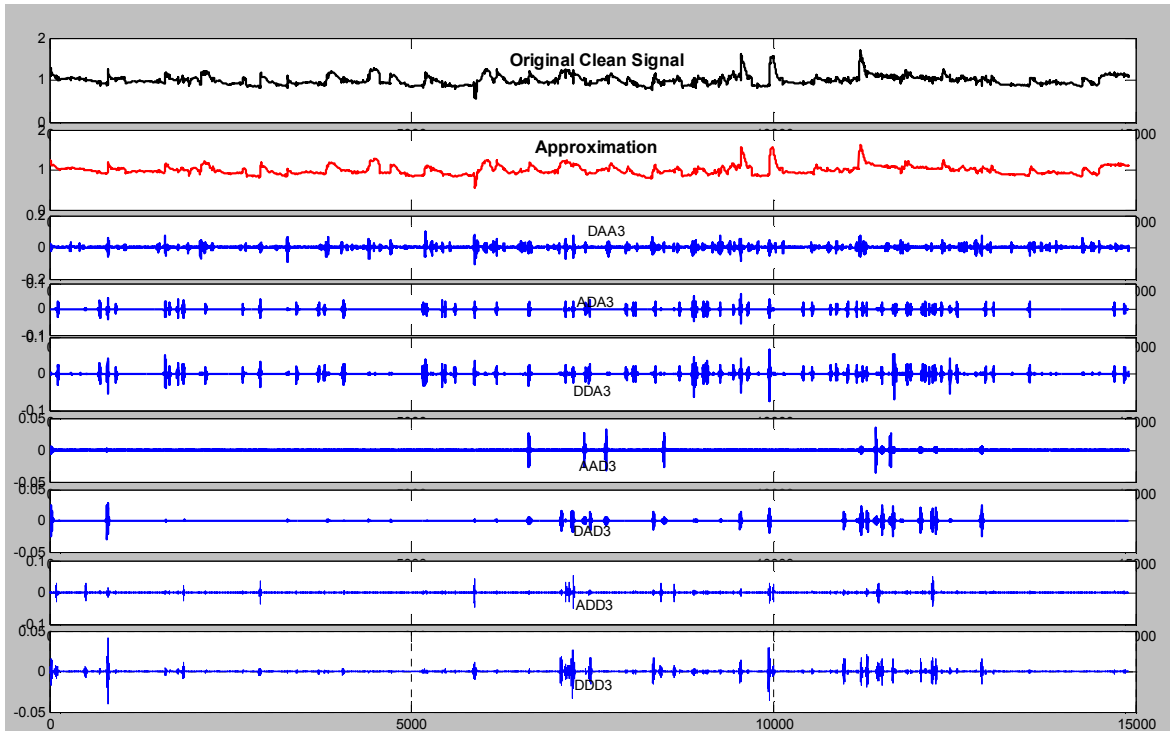


Figure 135: DWPT at level-3 MRA of sensor X3

Table 106: Node energy content of each sensor at level-3 DWPT decomposition

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of original)	99.9998	100	99.9835	99.9885	99.9895	99.9999	100	99.9999	100	99.9992	100	100
Sum of details (% of original)	2.33E-04	1.50E-06	0.0165	0.0115	0.0105	5.93E-05	2.72E-05	5.09E-05	4.86E-06	7.58E-04	3.33E-05	1.92E-06

Similarly, Table 106 shows the energy content of the approximation node as well as the sum of the detail nodes of each of the twelve monitored sensors at three-level decomposition.

B.3- Level-4 multi-resolution analysis

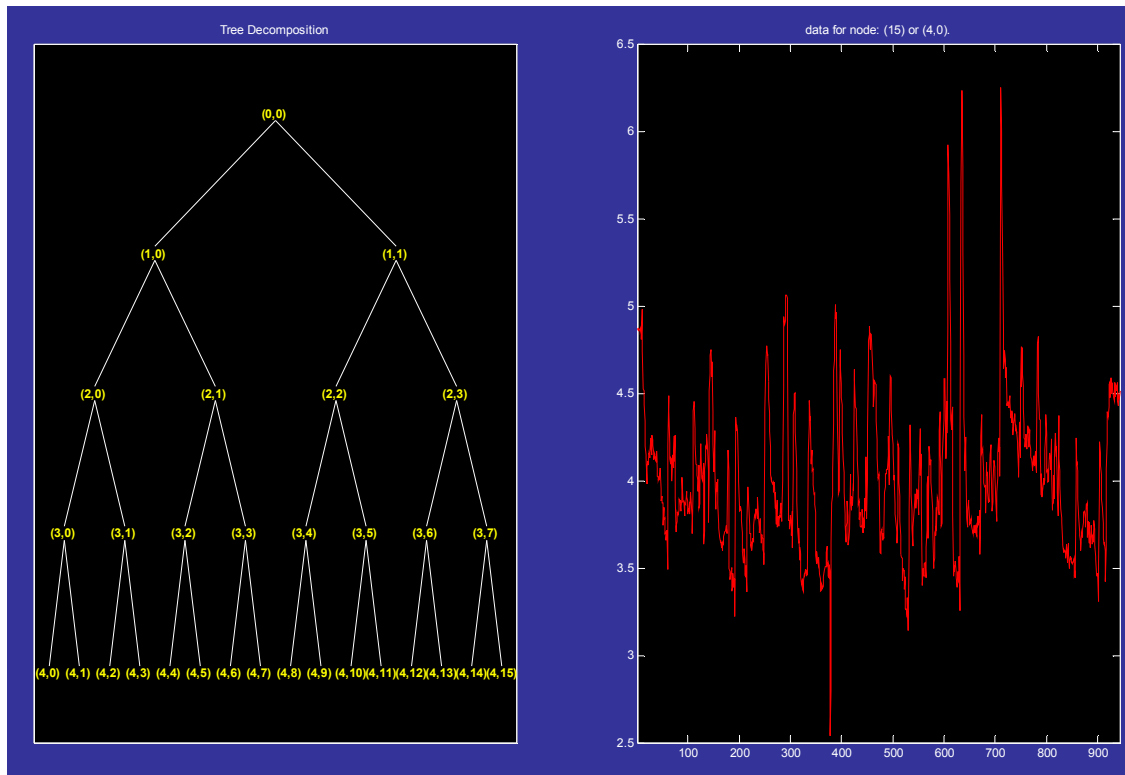


Figure 136: Level-4 tree decomposition of sensor X3 signal with the wavelet component at node 15

The four-level tree decomposition is shown on Figure 136 and Figure 137 represents the multi-resolution analysis plot of the sensor X3 with the original de-noised signal (black), the approximation node signal (red) and the seven details (blue).

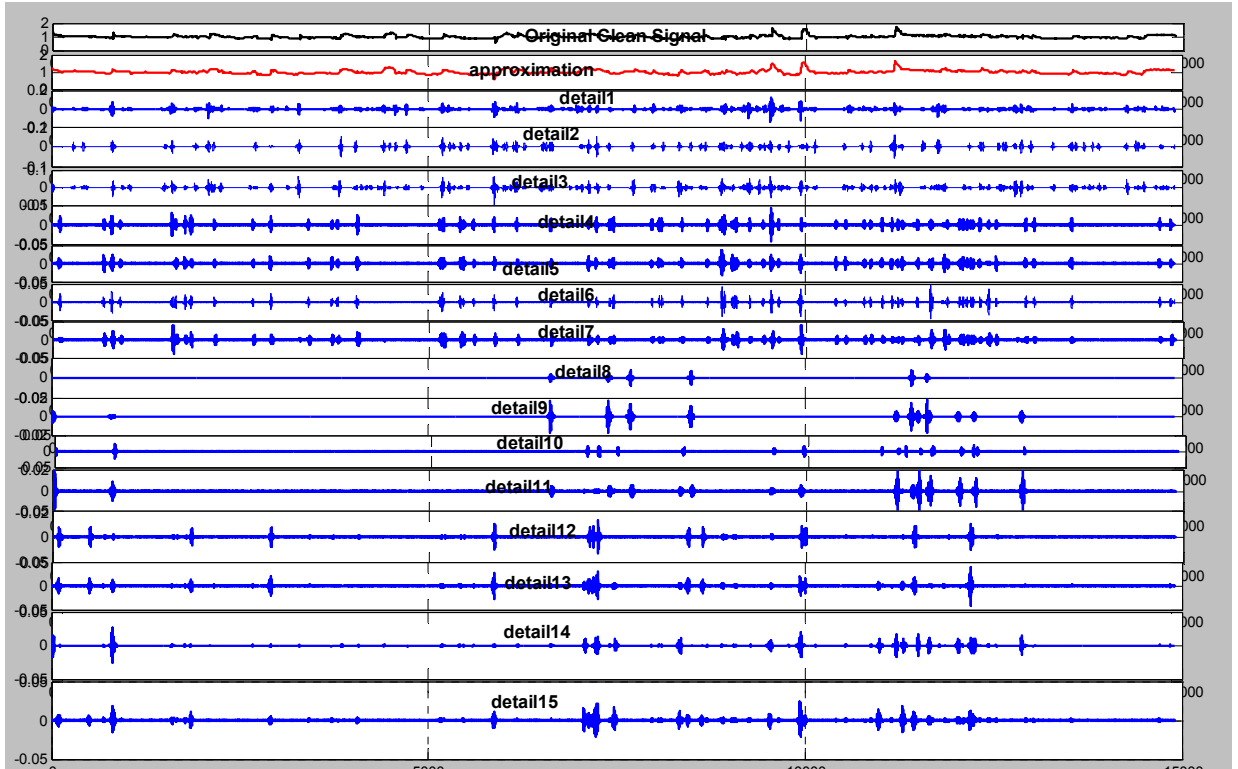


Figure 137: DWPT at level-4 MRA of sensor X3

Table 107: Node energy content of each sensor at level-4 DWPT decomposition

Sensor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Approximation (% of original)	99.9996	100	99.9699	99.976	99.986	99.9999	99.9999	99.9999	100	99.9984	99.9999	100
Sum of details (% of original)	4.32E-04	2.08E-06	0.0301	0.024	0.014	1.20E-04	5.44E-05	1.03E-04	5.97E-06	1.60E-03	7.37E-05	3.65E-06

Table 107 shows the energy content of the approximation node as well as the sum of the detail nodes of each of the twelve monitored sensors at four-level decomposition.

B.4- Level of decomposition decision for multi-resolution analysis

After the decomposition of each of the twelve sensors to two-level, three-level, four-level and the calculation of the energy content each node, it is noticed that the approximation node conserves over 99.9% of the total energy of the original signal in each case.

However, for sensitive sensors such as vibration sensors, the higher the level of decomposition (finer details), the more energy is contained at the details nodes is. Theoretically, a sensor signal could be decomposed a level much higher, but there is a need for a trade-off between the amount of information gained at higher decomposition level and the complexity of high decomposition level.

For example, the sensor vibration sensor X3 has the following:

- Approximation node energy: 99.994 and detail nodes energy: 0.0045 for level-2 DWPT
- Approximation node energy: 99.9835 and detail nodes energy: 0.0165 for level-3 DWPT
- Approximation node energy: 99.9699 and detail nodes energy: 0.0301 for level-4 DWPT

Though the level-4 yields a higher details energy content value than the level-3, which in turn has a higher value than level-2, as it can be inferred by comparing Figure 135 and Figure 137, the higher the level of decomposition, the more cumbersome the multi-resolution analysis becomes. Moreover, based on the value of the energy content at the approximation node that is over 99.9% in each of the three case, it is concluded that level-four decomposition does not provide much more information compared to the level-three for its associated complexity to justify its use. Consequently, the decision is made to use the level-three decomposition, which yields more information about the signal content than the level-two.

APPENDIX C- KAPLAN-MEIER TABLES

C.1- First Anomaly tabular data with its corresponding confidence interval

Failure Time (RTTF)	Survival (S_hat)	Lower bound	Upper Bound
3.183	0.909090909	0.739204	1
35.2	0.818181818	0.590255	1
85.183	0.818181818	0.590255	1
119.517	0.715909091	0.442171	0.9896474
124.4	0.715909091	0.442171	0.9896474
232.117	0.596590909	0.284162	0.9090197
233.067	0.477272727	0.151353	0.8031924
332.15	0.357954545	0.040514	0.675395
339.683	0.238636364	0	0.5236734
458.683	0.119318182	0	0.3376221
578.283	0	0	0.3376221

C.2- Second Anomaly tabular data with it corresponding confidence interval

Failure Time (RTTF)	Survival (S_hat)	Lower bound	Upper Bound
1	0.909091	0.739204	1
2.783	0.818182	0.590255	1
11.05	0.727273	0.464086	0.99046
37.317	0.636364	0.352089	0.920638
48.367	0.636364	0.352089	0.920638
49.383	0.530303	0.226775	0.833831
78.467	0.530303	0.226775	0.833831
161.167	0.397727	0.077631	0.717824
201.633	0.265152	0	0.566069
318.983	0.132576	0	0.370057
332.15	0	0	0.370057

C.3- Extended tabular data

Data point	ASI	ADI (hr)	RTTF (hr)	Censored
1	2.548228	1.183333	233.0667	0
2	2.81397	1.433333	209.85	0
3	3.029187	1.366667	161.1667	0
4	1.036524	0.216667	332.15	0
5	1.18051	0.166667	469.4833	0
6	1.132413	0.2	438.5667	0
7	1.137	0.116667	353.2667	0
8	1.183305	0.133333	284.4167	0
9	1.258377	0.366667	252.65	0
10	1.282236	3.6	232.1167	0
11	1.496119	2.133333	212.0167	0
12	1.246279	3.766667	208.0167	0
13	1.369278	2.266667	201.6333	0
14	1.252208	2.016667	182.45	0
15	1.151164	2.15	160.9333	0
16	1.208265	4.65	113.6667	0
17	1.279143	4.766667	64.81667	0
18	1.218381	2.283333	41.98333	0
19	1.140144	1.6	17.11667	0
20	1.359802	0.3	359.9	0
21	1.165099	1	339.6833	0
22	1.065159	0.6	318.9833	0
23	1.036482	0.083333	675.4333	0
24	1.184042	1.033333	578.2833	0
25	3.623064	0.716667	37.31667	0
26	1.062287	0.066667	458.6833	0
27	2.240944	1.016667	1	0
28	1.106504	0.166667	85.98333	0
29	1.551326	0.833333	35.2	0
30	1.580428	1.066667	11.05	0
31	1.12047	0.183333	119.5167	0
32	1.016824	0.05	49.38333	0
33	1.130492	0.233333	3.183333	0
34	1.18496	2.8	2.783333	0
35	2.335767	2.366667	85.18333	1
36	1.637981	16.38333	78.46667	1
37	1.626708	22.18333	22.16667	1
38	1.006843	0.033333	124.4	1
39	1.23587	0.683333	48.36667	1

In the “censored column, a value of “0” corresponds to an observed failure events, while a value of “1” corresponds to a censored events.

C.4- Tabular data for the Kaplan-Meier plot of residual time for extended data set

Failure Time (RTTF)	Number failed	Number censored	At Risk	Survival (S_hat)
0	0	0	39	1
1	1	0	39	0.9744
2.783	1	0	38	0.9487
3.183	1	0	37	0.9231
11.05	1	0	36	0.8974
17.117	1	0	35	0.8718
22.167	0	1	34	0.8718
35.2	1	0	33	0.8454
37.317	1	0	32	0.819
41.983	1	0	31	0.7925
48.367	0	1	30	0.7925
49.383	1	0	29	0.7652
64.817	1	0	28	0.7379
78.467	0	1	27	0.7379
85.183	0	1	26	0.7379
85.983	1	0	25	0.7084
113.667	1	0	24	0.6789
119.517	1	0	23	0.6493
124.4	0	1	22	0.6493
160.933	1	0	21	0.6184
161.167	1	0	20	0.5875
182.45	1	0	19	0.5566
201.633	1	0	18	0.5257
208.017	1	0	17	0.4947
209.85	1	0	16	0.4638
212.017	1	0	15	0.4329
232.117	1	0	14	0.402
233.067	1	0	13	0.371
252.65	1	0	12	0.3401
284.417	1	0	11	0.3092
318.983	1	0	10	0.2783
332.15	1	0	9	0.2474
339.683	1	0	8	0.2164
353.267	1	0	7	0.1855
359.9	1	0	6	0.1546
438.567	1	0	5	0.1237
458.683	1	0	4	0.0928
469.483	1	0	3	0.0618
578.283	1	0	2	0.0309
675.433	1	0	1	0

C.5- Extended tabular data for confidence interval

Failure Time (RTTF)	Survival (S_hat)	Lower bound	Upper Bound
1	0.9744	0.924752	1
2.783	0.9487	0.879492	1
3.183	0.9231	0.839447	1
11.05	0.8974	0.802219	0.992653
17.117	0.8718	0.766871	0.976719
22.167	0.8718	0.766871	0.976719
35.2	0.8454	0.731571	0.959183
37.317	0.819	0.697501	0.940417
41.983	0.7925	0.664438	0.920643
48.367	0.7925	0.664438	0.920643
49.383	0.7652	0.630794	0.89963
64.817	0.7379	0.598	0.877766
78.467	0.7379	0.598	0.877766
85.183	0.7379	0.598	0.877766
85.983	0.7084	0.562608	0.854127
113.667	0.6789	0.528123	0.829581
119.517	0.6493	0.494457	0.804216
124.4	0.6493	0.494457	0.804216
160.933	0.6184	0.459497	0.777336
161.167	0.5875	0.425378	0.749613
182.45	0.5566	0.392051	0.721098
201.633	0.5257	0.359483	0.691824
208.017	0.4947	0.327651	0.661815
209.85	0.4638	0.296542	0.631082
212.017	0.4329	0.266155	0.599627
232.117	0.402	0.236496	0.567445
233.067	0.371	0.207583	0.534516
252.65	0.3401	0.179442	0.500815
284.417	0.3092	0.152116	0.4663
318.983	0.2783	0.125662	0.430913
332.15	0.2474	0.100158	0.394575
339.683	0.2164	0.075715	0.357176
353.267	0.1855	0.052488	0.318562
359.9	0.1546	0.030703	0.278505
438.567	0.1237	0.010709	0.236657
458.683	0.0928	0	0.192431
469.483	0.0618	0	0.144688
578.283	0.0309	0	0.090522
675.433	0	0	0.090522

APPENDIX D- RESIDUAL PLOTS FOR PARAMETRIC ANALYSIS

D.1- Residual plots for limited data set: first detected anomaly

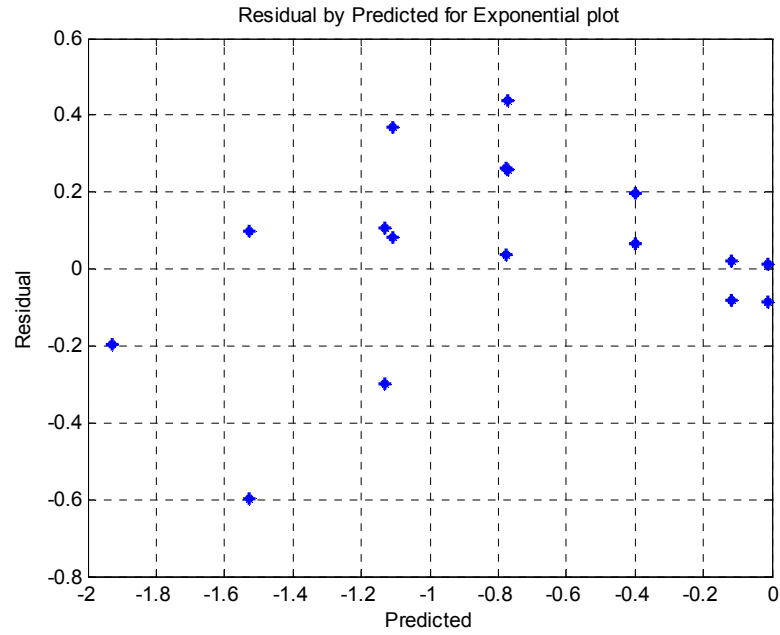


Figure 138: Residual by Predicted for Exponential distribution assumption for limited data set

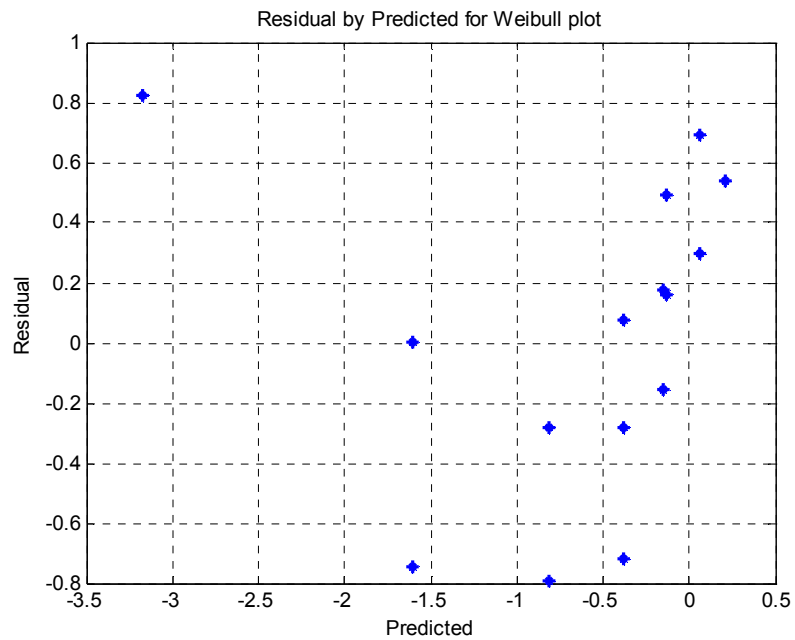


Figure 139: Residual by Predicted for Weibull distribution assumption for limited data set

D.2- Residual plots for extended data set

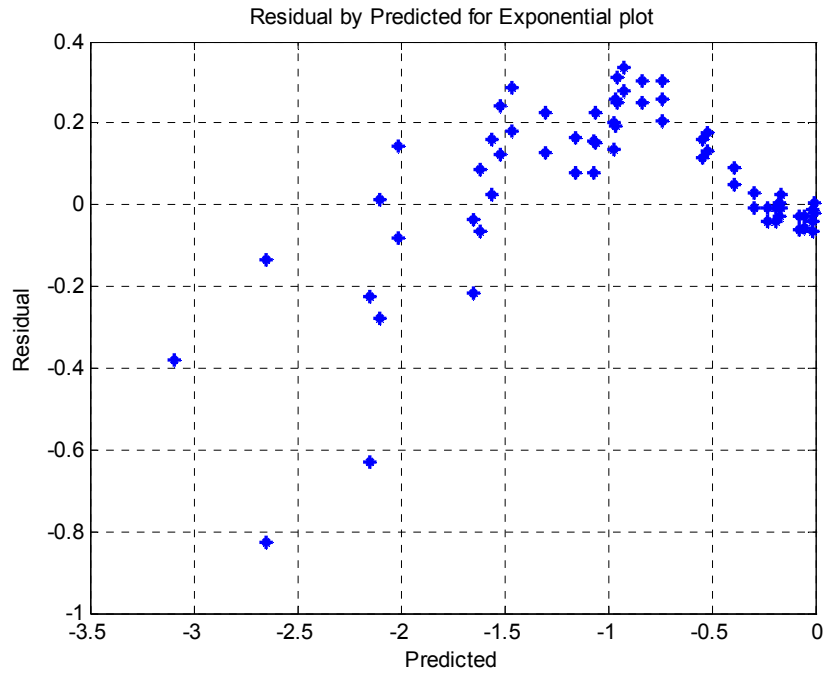


Figure 140: Residual by Predicted for Exponential distribution assumption for extended data set

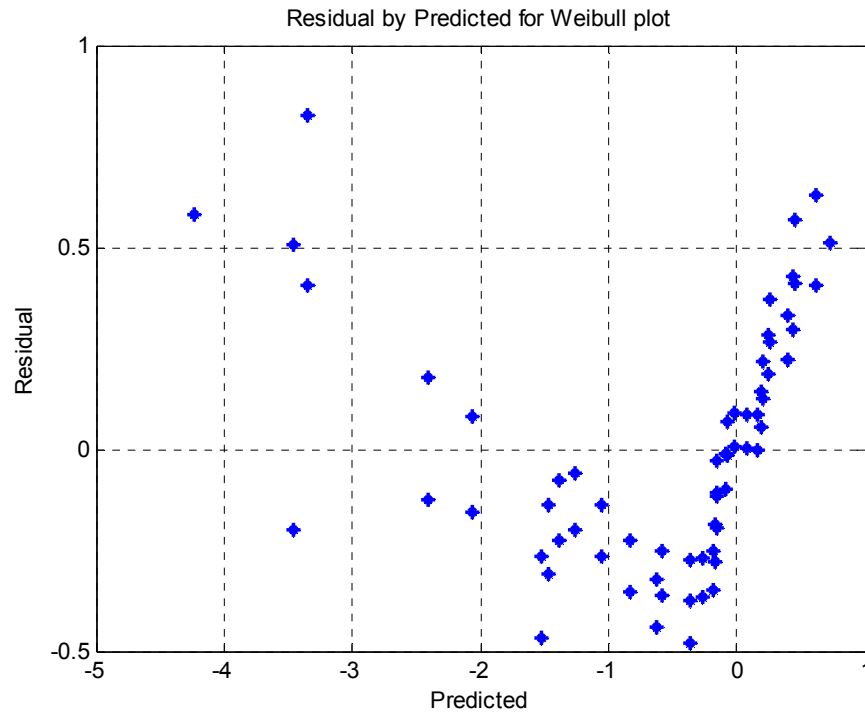


Figure 141: Residual by Predicted for Weibull distribution assumption for extended data set

REFERENCES

1. Thaler, H. *Global Gas Turbine Markets and Trends - What does the future hold?* in *The Future of Gas Turbine Technology - 3rd International Conference*. 2006. Brussels.
2. Stoll, H., *Creating owner's competitive advantage through contractual services*. GER-4208, GE Power Systems.
3. EPRI, *Gas Turbine Maintenance Strategy Decision Support*. 2006, Electric Power Research Institute (EPRI).
4. Siemens, P.G.U.P.P.S.-. *Gas Turbine Long-term Maintenance Agreements*.
5. Alstom-Power, *Full Turnkey capability*. 2009.
6. Alstom-Power, *The Full-Service Provider in Power Generation*. 2004.
7. MHI. *Mitsubishi Heavy Industries Long Term Service Agreement ("LTSA")*. 2009; Available from: <http://www.mhi.co.jp/en/power/service/gas/index2.html>.
8. Ansaldo-Energia, *Services on Ansaldo Energia Technologies*. 2007.
9. Balevic, D., R. Burger, and D. Forry, *Heavy-Duty Gas Turbine Operating and Maintenance Considerations*. GE Gas Turbine Reference Library GER-3620K, 2004: p. 30-32.
10. BusinessWire (2008) *Sonelgaz Signs \$1 Billion in Service Agreements with GE Energy, Targeting Improved Efficiency, Output for Algerian Power Plants*. Business Wire.
11. EPRI, *GE FA Compressor Dependability*. 2008, Electric Power Research Institute (EPRI).
12. Swanson, L., *Linking maintenance strategies to performance*. International Journal of Production Economics, 2001. **70**(3): p. 237-244.

13. Martin, K., *A review by discussion of condition monitoring and fault diagnosis in machine tools*. International journal of machine tools & manufacture, 1994. **34**(4): p. 527-551.
14. Zhao, Y., et al., *A sequential approach for gas turbine power plant preventative maintenance scheduling*. Journal of Engineering for Gas Turbines and Power, 2006. **128**: p. 796.
15. Horner, R., M. El-Haram, and A. Munns, *Building maintenance strategy: a new management approach*. Journal of Quality in Maintenance Engineering, 1997. **3**(4): p. 273-280.
16. Saxena, A., *Knowledge-Based Architecture for Integrated Condition Based Maintenance of Engineering Systems*. 2007, Citeseer.
17. Zhao, Y., et al., *A Profit-Based Approach for Gas Turbine Power Plant Outage Planning*. Journal of Engineering for Gas Turbines and Power, 2006. **128**: p. 806.
18. EPRI, *Gas Turbine Compressor Dependability: Investigation of the MS7001F/900F R0 Compressor Stage Blade Designs and Failures: 2007 Update* 2007, Electric Power Research Institute (EPRI).
19. EPRI, *GE FA Compressor Dependability (phase 2)*. 2008, Electric Power Research Institute (EPRI).
20. Box, G., G. Jenkins, and G. Reinsel, *Time series analysis: forecasting and control*. 1976: Holden-day San Francisco.
21. Saxena, A., et al. *Metrics for Evaluating Performance of Prognostic Techniques*: Citeseer.
22. Schwabacher, M. *A survey of data-driven prognostics*. 2005: Citeseer.
23. Zhang, G., *Optimum Sensor Localization/Selection in A Diagnostic/Prognostic Architecture*, in *Electrical Engineering*. 2005, Georgia Institute of Technology: Atlanta.

24. Peng, Z. and F. Chu, *Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography*. Mechanical Systems and Signal Processing, 2004. **18**(2): p. 199-221.
25. Carden, E. and P. Fanning, *Vibration based condition monitoring: a review*. Structural Health Monitoring, 2004. **3**(4): p. 355.
26. Flandrin, P., *Time-frequency/time-scale analysis*. 1999: Academic Pr.
27. Boashash, B., *Note on the use of the Wigner distribution for time-frequency signal analysis*. IEEE Transactions on Acoustics, Speech and Signal Processing, 1988. **36**(9): p. 1518-1521.
28. Koo, I. and W. Kim, *The development of reactor coolant pump vibration monitoring and a diagnostic system in the nuclear power plant*. ISA transactions, 2000. **39**(3): p. 309-316.
29. Auger, F. and P. Flandrin, *Improving the readability of time-frequency and time-scale representations by the reassignment method*. IEEE Transactions on Signal Processing, 1995. **43**(5): p. 1068-1089.
30. Allen, J. and L. Rabiner, *A unified approach to short-time Fourier analysis and synthesis*. Proc. IEEE, 1977. **65**(11): p. 1558-1564.
31. Mertins, A., *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. 1999: John Wiley & Sons, Inc. New York, NY, USA.
32. Choi, H. and W. Williams, *Improved time-frequency representation of multicomponent signals using exponential kernels*. IEEE Transactions on Acoustics, Speech and Signal Processing, 1989. **37**(6): p. 862-871.
33. Mallat, S. and W. Hwang, *Singularity detection and processing with wavelets*. IEEE Transactions on Information Theory, 1992. **38**(2): p. 617-643.
34. Jiang, X., S. Mahadevan, and R. Guratzsch, *Bayesian Wavelet Methodology for Damage Detection of Thermal Protection System Panels*. AIAA JOURNAL, 2009. **47**(4).

35. Wang, W. and P. McFadden, *Application of orthogonal wavelets to early gear damage detection*. Mechanical Systems and Signal Processing, 1995. **9**(5): p. 497-507.
36. Wang, W. and P. McFadden, *Application of wavelets to gearbox vibration signals for fault detection*. Journal of Sound and Vibration, 1996. **192**(5): p. 927-939.
37. Boulahbal, D., G. MFARID, and F. Ismail, *Amplitude and phase wavelet maps for the detection of cracks in geared systems*. Mechanical Systems and Signal Processing, 1999. **13**(3): p. 423-436.
38. Butler, K. and S. Dey, *Using Discrete Wavelet Transforms to Characterize Equipment Failures for Maintaining Distribution System Reliability*.
39. Lin, J. and L. Qu, *Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis*. Journal of Sound and Vibration, 2000. **234**(1): p. 135-148.
40. Saxena, A., B. Wu, and G. Vachtsevanos. *A methodology for analyzing vibration data from planetary gear systems using complex Morlet wavelets*. 2005.
41. Mufti, M. and G. Vachtsevanos. *Automated fault detection and identification using a fuzzy-waveletanalysis technique*. 1995.
42. Mufti, M., *Fault detection and identification using fuzzy wavelets*. 1995, Georgia Institute of Technology.
43. Wu, Y. and R. Du, *Feature extraction and assessment using wavelet packets for monitoring of machining processes*. Mechanical systems and signal processing, 1996. **10**(1): p. 29-53.
44. Sun, Z. and C. Chang, *Statistical wavelet-based method for structural health monitoring*. Journal of structural engineering, 2004. **130**: p. 1055.
45. Pittner, S. and S. Kamarthi, *Feature extraction from wavelet coefficients for patternrecognition tasks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999. **21**(1): p. 83-88.

46. Wang, P. and G. Vachtsevanos, *Fault prognostics using dynamic wavelet neural networks*. AI EDAM, 2002. **15**(04): p. 349-365.
47. Berenji, H. and Y. Wang. *Wavelet neural networks for fault diagnosis and prognosis*.
48. Yen, G. and K. Lin. *Conditional health monitoring using vibration signatures*. 1999.
49. Greitzer, F. and T. Ferryman. *Predicting Remaining Life of Mechanical Systems*. 2001.
50. Vachtsevanos, G., et al., *Intelligent fault diagnosis and prognosis for engineering systems*. 2006: Wiley.
51. Brotherton, T., et al. *Prognosis of faults in gas turbine engines*. 2000: Citeseer.
52. Adeline, R., R. Gouriveau, and N. Zerhouni. *PRONOSTIC DE DÉFAILLANCES: MAÎTRISE DE L'ERREUR DE PRÉDICTION*.
53. Byington, C., et al., *Prognostic enhancements to gas turbine diagnostic systems*. 2003, IMPACT TECHNOLOGIES LLC STATE COLLEGE PA.
54. Byington, C., M. Watson, and D. Edwards. *Data-driven neural network methodology to remaining life predictions for aircraft actuator components*. 2004.
55. Schwabacher, M. and K. Goebel. *A survey of artificial intelligence for prognostics*.
56. Gebraeel, N., *Sensory-updated residual life distributions for components with exponential degradation patterns*. IEEE transactions on automation science and engineering, 2006. **3**(4): p. 382-393.
57. Gebraeel, N., et al., *Residual-life distributions from component degradation signals: A Bayesian approach*. IIE Transactions, 2005. **37**(6): p. 543-557.
58. Suarez, E., et al. *Jet engine life prediction systems integrated with prognostics health management*. 2004.

59. Bhushan, M. and R. Rengaswamy, *Design of sensor location based on various fault diagnostic observability and reliability criteria*. Computers and Chemical Engineering, 2000. **24**(2-7): p. 735-741.
60. Percival, D. and A. Walden, *Wavelet methods for time series analysis*. 2006: Cambridge Univ Pr.
61. Jiang, X. and H. Adeli, *Wavelet packet-autocorrelation function method for traffic flow pattern analysis*. Computer-aided civil and infrastructure engineering, 2004. **19**(5): p. 324-337.
62. Misrikhanov, A., *Wavelet transform methods: Application in electroenergetics*. Automation and Remote Control, 2006. **67**(5): p. 682-697.
63. Walker, J., *Fast fourier transforms*. 1996: CRC.
64. Chen, H., *A Multiscale Forecasting Methodology for Power Plant Fleet Management*. 2005, Citeseer.
65. Hubbard, B. and F. Meyer, *The world according to wavelets: the story of a mathematical technique in the making*. 1996: AK Peters Wellesley.
66. Daubechies, I., *Ten lectures on wavelets*. 1992: Society for Industrial Mathematics.
67. Chui, C., *An introduction to wavelets*. 1992: Academic Pr.
68. Nievergelt, Y., *Wavelets made easy*. 1999: Birkhauser.
69. Graps, A., *An introduction to wavelets*. IEEE Computational Science & Engineering, 1995. **2**(2): p. 50-61.
70. Mallat, S., *A theory for multiresolution signal decomposition: The wavelet representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989. **11**(7): p. 674-693.
71. Holschneider, M., *Wavelets: an analysis tool*. 1998: Oxford University Press Oxford.

72. Rao, R., A. Bopardikar, and T. Boros, *Wavelet transforms: Introduction to theory and applications*. Journal of Electronic Imaging, 1999. **8**: p. 478.
73. Vetterli, M. and C. Herley, *Wavelets and filter banks: Theory and design*. IEEE Transactions on Signal Processing, 1992. **40**(9): p. 2207-2232.
74. Coifman, R. and M. Wickerhauser, *Entropy-based algorithms for best basis selection*. IEEE Transactions on Information Theory, 1992. **38**(2): p. 713-718.
75. Cody, M., *The wavelet packet transform: Extending the wavelet transform*. Dr. Dobb's Journal, 1994. **19**: p. 44-46.
76. Wickerhauser, M., *Adapted wavelet analysis from theory to software*. 1994: AK Peters Ltd.
77. Matlab, *Wavelet Toolbox 4 -User's Guide*, in *Matlab 7.8.0 (R2009a)*. 1999-2009.
78. Sohn, H., et al., *Structural health monitoring using statistical pattern recognition techniques*. Journal of dynamic systems, measurement, and control, 2001. **123**: p. 706.
79. Hou, Z., M. Noori, and R. Amand, *Wavelet-based approach for structural damage detection*. Journal of Engineering Mechanics, 1999. **126**(7): p. 677.
80. Sohn, H. and C. Farrar, *Damage diagnosis using time series analysis of vibration signals*. Smart Materials and Structures, 2001. **10**: p. 446.
81. Dong, D., J. Hopfield, and K. Unnikrishnan, *Neural networks for engine fault diagnostics*. Neural Networks for Signal Processing VII, New York, 1997: p. 636-644.
82. Aktan, A., A. Helmicki, and V. Hunt, *Issues in health monitoring for intelligent infrastructure*. Smart Materials and Structures, 1998. **7**: p. 674.
83. Ogaji, S. and R. Singh, *Advanced engine diagnostics using artificial neural networks*. Applied Soft Computing, 2003. **3**(3): p. 259-271.

84. Ogaji, S. and R. Singh, *Gas path fault diagnosis framework for a three-shaft gas turbine*. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 2003. **217**(2): p. 149-157.
85. Diallo, O. and D. Mavris. *A Concept for Intelligent Fault Detection Using a Multi-resolution Analysis*. in *Aerospace Systems Conference*. 2010. Los Angeles, CA.
86. Park, B. and H. Kargupta, *Distributed data mining: Algorithms, systems, and applications*. Data Mining Handbook, 2002: p. 341–358.
87. Quackenbush, J., *Microarray data normalization and transformation*. nature genetics, 2002. **32**: p. 496-501.
88. Jain, A., K. Nandakumar, and A. Ross, *Score normalization in multimodal biometric systems*. Pattern recognition, 2005. **38**(12): p. 2270-2285.
89. Al Shalabi, L., Z. Shaaban, and B. Kasasbeh, *Data mining: A preprocessing engine*. Journal of Computer Science, 2006. **2**(9): p. 735-739.
90. Yang, Y., et al., *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation*. Nucleic acids research, 2002. **30**(4): p. e15.
91. Cappelli, R., D. Maio, and D. Maltoni, *Combining fingerprint classifiers*. Multiple Classifier Systems, 2000: p. 351-361.
92. Snelick, R., et al., *Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems*. IEEE Transactions on pattern analysis and machine intelligence, 2005: p. 450-455.
93. Hampel, F., et al., *Robust statistics: the approach based on influence functions*. 1986: John Wiley & Sons New York.
94. Mosteller, F. and J. Tukey, *Data analysis and regression: a second course in statistics*. 1977: Addison-Wesley.
95. Yan, W. and K. Goebel. *Sensor validation and fusion for gas turbine vibration monitoring*. 2003.

96. Sriram, S., et al., *Signal denoising techniques for partial discharge measurements*. IEEE Transactions on Dielectrics and Electrical Insulation, 2005. **12**(6): p. 1182-1191.
97. Parks, T. and C. Burrus, *Digital filter design*. 1987: Wiley-Interscience New York, NY, USA.
98. Hlawatsch, F. and G. Boudreaux-Bartels, *Linear and quadratic time-frequency signal representations*. IEEE Signal Processing Magazine, 1992. **9**(2): p. 21-67.
99. BOUDREAUX-BARTELS, G. and T. PARKS, *Time-varying filtering and signal estimation using Wigner distribution synthesis techniques*.
100. Griffin, D. and J. Lim, *Signal estimation from modified short-time Fourier transform*. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984. **32**(2): p. 236-243.
101. Clarkson, P., *Optimal and adaptive signal processing*. 1993: CRC.
102. Treichler, J., C. Johnson Jr, and M. Larimore, *Theory and design of adaptive filters*. 1987: Wiley-Interscience New York, NY, USA.
103. Cho, N., C. Choi, and S. Lee, *Adaptive line enhancement by using in IIR lattice notch filter*. IEEE transactions on acoustics, speech, and signal processing, 1989. **37**(4): p. 585-589.
104. Wang, H., K. Tan, and D. Zhu, *Extraction of partial discharge signals using wavelet transform*. QINGHUA DAXUE XUEBAO, 1998. **38**(6): p. 119-122.
105. Donoho, D., *Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data*. American Mathematical Society, 1993. **47**: p. 173-205.
106. Donoho, D. and J. JOHNSTONE, *Ideal spatial adaptation by wavelet shrinkage*. Biometrika, 1994. **81**(3): p. 425.
107. Donoho, D., *De-noising by soft-thresholding*. IEEE Transactions on Information Theory, 1995. **41**(3): p. 613-627.

108. Coifman, R. and D. Donoho, *Translation-invariant de-noising*. LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-, 1995: p. 125-125.
109. Jiang, X., S. Mahadevan, and H. Adeli, *Bayesian wavelet packet denoising for structural system identification*. Progress in Structural Engineering and Materials, 2006. **14**(2): p. 333-356.
110. Jian-lin, Z., et al., *De-noising of Sign in Machinery Fault Diagnosis Based on Wavelet Packet [J]*. Coal Mine Machinery, 2005. **9**.
111. Vidakovic, B., *Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors*. Journal of the American Statistical Association, 1998. **93**(441).
112. Kvam, P. and B. Vidakovic, *Nonparametric statistics with applications to science and engineering*. 2007: Wiley-Interscience.
113. Mörchen, F., *Time series feature extraction for data mining using DWT and DFT*. 2003: Citeseer.
114. Simone, G., et al., *Feature extraction techniques for ultrasonic signal classification*. International Journal of Applied Electromagnetics and Mechanics, 2002. **15**(1): p. 291-294.
115. Qian, S. and D. Chen, *Discrete gabor transform*. IEEE transactions on signal processing, 1993. **41**(7): p. 2429-2438.
116. Orr, R., *Derivation of the finite discrete Gabor transform by periodization and sampling*. Signal Processing, 1993. **34**(1): p. 85-97.
117. Weisstein, E., *Discrete Fourier Transform*. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/DiscreteFourierTransform.html>.
118. Agrawal, R., C. Faloutsos, and A. Swami, *Efficient similarity search in sequence databases*. Foundations of Data Organization and Algorithms, 1993: p. 69-84.
119. Wu, Y., D. Agrawal, and A. El Abbadi. *A comparison of DFT and DWT based similarity search in time-series databases*. 2000: ACM.

120. Ogden, R., *Essential wavelets for statistical applications and data analysis*. 1997: Birkhauser.
121. Coué, C., et al. *Using Bayesian programming for multi-sensor multi-target tracking in automotive applications*. 2003.
122. Llinas, J. and D. Hall. *An introduction to multi-sensor data fusion*. 1998: INSTITUTE OF ELECTRICAL ENGINEERS INC (IEEE).
123. White Jr, F. *Joint directors of Laboratories data fusion sub-panel report, "SIGINT sessions"; Tech*.
124. Klein, L., *Sensor and data fusion: a tool for information assessment and decision making*. 2004: SPIE Press.
125. Carvalho, H., et al. *A general data fusion architecture*. 2003: Citeseer.
126. Hall, D. and S. McMullen, *Mathematical techniques in multisensor data fusion*. 2004: Artech House Publishers.
127. Goodman, I., R. Mahler, and H. Nguyen, *Mathematics of data fusion*. 1997: Springer.
128. Ding, H., et al. *Data fusion algorithm based on fuzzy logic*. 2004.
129. AUG-Signals, *Data Fusion Techniques*.
130. Smith, L., *A tutorial on principal components analysis*. Cornell University, USA, 2002. **51**: p. 52.
131. Jolliffe, I., *Principal component analysis*. 2002: Springer verlag.
132. Shlens, J., *A tutorial on principal component analysis*. Copy retrieved [10-27-2004] from: <http://www.sn1.salk.edu/shlens/notes.html>, 2003.
133. Jain, A. and R. Mao, *Statistical pattern recognition: A review*. IEEE Transactions on pattern analysis and machine intelligence, 2000. **22**(1): p. 4-37.

134. Wood, F., K. Esbensen, and P. Geladi, *Principal component analysis*. Chemometr. Intel. Lab. Syst, 1987. **2**: p. 37–52.
135. Jiang, X., et al., *Bayesian Probabilistic Approach for Model Validation of Dynamic Systems*. SAE International, 2009.
136. Tipping, M. and C. Bishop, *Probabilistic principal component analysis*. Journal of the Royal Statistical Society. Series B, Statistical Methodology, 1999: p. 611-622.
137. Little, R. and D. Rubin, *Statistical analysis with missing data*. 1987.
138. Ripley, B., *Pattern recognition and neural networks*. 2008: Cambridge Univ Pr.
139. Dony, R. and S. Haykin, *Optimally adaptive transform coding*. IEEE Transactions on Image Processing, 1995. **4**(10): p. 1358-1370.
140. Bishop, C. and M. Tipping, *A hierarchical latent variable model for data visualization*. IEEE Transactions on pattern analysis and machine intelligence, 1998. **20**(3): p. 281-293.
141. Bishop, C., *Latent variable models*. Learning in graphical models, 1998.
142. Lazarevic, A., et al., *Data Mining for Anomaly Detection*.
143. Chandola, V., A. Banerjee, and V. Kumar, *Anomaly detection: A survey*. ACM Computing Surveys (CSUR), 2009. **41**(3): p. 15.
144. Lazarevic, A., et al. *A comparative study of anomaly detection schemes in network intrusion detection*. 2003.
145. Saxena, A., et al. *Metrics for evaluating performance of prognostic techniques*: Citeseer.
146. Besterfield, D., *Quality control*. 1994: Prentice Hall Career & Technology, Englewood Cliffs, NJ.

147. Sun, Z., *Structural damage assessment based on wavelet packet transform*. Journal of structural engineering, 2002. **128**: p. 1354.
148. Shi, J. and C. Tomasi. *Good features to track*. 1994: Citeseer.
149. Tommasini, T., et al. *Making good features track better*. 1998: Citeseer.
150. Montgomery, D., *Introduction to statistical quality control*. Vol. 3rd. 1996: Wiley- New York.
151. Ang, A. and W. Tang, *Probability concepts in engineering planning and design*. 1975.
152. Montgomery, D., *Introduction to statistical quality control*. 1985.
153. Maltoni, M., et al., *Prognostic factors in advanced cancer patients: evidence-based clinical recommendations--a study by the Steering Committee of the European Association for Palliative Care*. Journal of Clinical Oncology, 2005. **23**(25): p. 6240.
154. Motzer, R., et al., *Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma*. Journal of Clinical Oncology, 1999. **17**(8): p. 2530.
155. Lam, P., M. Leung, and C. Tse, *Identifying prognostic factors for survival in advanced cancer patients: a prospective study*. Hong Kong Medical Journal, 2007. **13**(6): p. 453.
156. Liao, H., et al. *A predictive tool for remaining useful life estimation of rotating machinery components*. 2005: ASME.
157. Kaplan, E. and P. Meier, *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, 1958. **53**(282): p. 457-481.
158. Mazhar, M., S. Kara, and H. Kaebnick, *Remaining life estimation of used components in consumer products: life cycle data analysis by Weibull and*

- artificial neural networks*. Journal of Operations Management, 2007. **25**(6): p. 1184-1193.
159. Cox, D., *Regression models and life-tables*. Journal of the Royal Statistical Society. Series B (Methodological), 1972. **34**(2): p. 187-220.
 160. Litt, J., et al., *A survey of intelligent control and health management technologies for aircraft propulsion systems*. Journal of Aerospace Computing, Information, and Communication, 2004. **1**(12): p. 543-563.
 161. Wilkinson, C., et al. *Prognostic and health management for avionics*. 2004.
 162. Vichare, N. and M. Pecht, *Prognostics and health management of electronics*. IEEE Transactions on Components and Packaging Technologies, 2006. **29**(1): p. 223.
 163. Hess, A., G. Calvello, and T. Dabney. *PHM a key enabler for the JSF autonomic logistics support concept*. 2004.
 164. Roemer, M., et al., *Development of diagnostic and prognostic technologies for aerospace health management applications*. 2001.
 165. Ferrell, B., L. Syst, and F. Worth. *Air vehicle prognostics and health management*. 2000.
 166. Myers, R., D. Montgomery, and C. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. 2009: Wiley.
 167. Montgomery, D., *Design and analysis of experiments*. 1991.
 168. Breyfogle, F., *Implementing Six Sigma: Smarter solutions using statistical methods*. 2003: Productivity Press.
 169. Carley, K., N. Kamneva, and J. Reminga, *Response Surface Methodology1*. 2004.
 170. Johnson, C. and J. Schutte, *Basic Regression Analysis for Integrated Neural Networks (BRAINN) Documentation*. 2009.

171. Daberkow, D. and D. Mavris, *New approaches to conceptual and preliminary aircraft design: A comparative assessment of a neural network formulation and a response surface methodology*. 1998.
172. Harston, C., A. Maren, and R. Pap, *Handbook of neural computing applications*. 1990, Academic Press San Diego, CA.
173. Demuth, H. and M. Beale, *Neural network toolbox*. MathWorks Version, 2001. 4.
174. Bodén, M., *A guide to recurrent neural networks and back propagation*. The DALLAS project. Report from the NUTEK-supported project AIS-8, SICS. Holst: Application of data analysis with learning systems, 2001.
175. Veitch, D., *Wavelet Neural Networks*. 2005.
176. Medsker, L. and L. Jain, *Recurrent neural networks*. 2000: CRC Press.
177. Williams, R. and D. Zipser, *A learning algorithm for continually running fully recurrent neural networks*. Neural computation, 1989. 1(2): p. 270-280.
178. Zhang, Q., A. Benveniste, and L. Hogskola, *Wavelet networks*. IEEE transactions on Neural Networks, 1992. 3(6): p. 889-898.
179. Wenji, P. and L. Xingqi, *Research on vibration fault diagnosis of hydro-turbine generating unit based on wavelet neural network [J]*. Journal of Hydroelectric Engineering, 2007. 1.
180. Guo, Q., H. Yu, and A. Xu. *Modified Morlet wavelet neural networks for fault detection*. 2005.
181. Xu, Q., et al. *Gas turbine fault diagnosis based on wavelet neural network*. 2007.
182. Rasmussen, C. and C. Williams, *Gaussian processes for machine learning*. 2006: Springer.
183. Hollingsworth, P. and D. Mavris, *Gaussian process meta-modeling: Comparison of gaussian process training methods*. AIAA 3rd Annual Aviation Technology, Integration and Operations (ATIO) Tech, 2003.

184. Kleinbaum, D. and M. Klein, *Survival analysis: a self-learning text*. 2005: Springer Verlag.
185. Huang, J. and J. Wellner, *Interval censored survival data: a review of recent progress*. LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-, 1997: p. 123-170.
186. Castet, J. and J. Saleh, *Satellite reliability: statistical data analysis and modeling*. Journal of Spacecraft and Rockets, 2009. **46**(5): p. 1065–1076.
187. Lawless, J., *Statistical models and methods for lifetime data*. 1982.
188. Tobias, P. and D. Trindade, *Applied reliability*. 1995: Chapman & Hall/CRC.

VITA

OUSMANE N. DIALLO

Ousmane Diallo was born in Abidjan, Ivory Coast. He attended the Ivory Coast schools systems where he spent 6 years in primary school, followed by 4 years in middle school before the final 3 years at the high school of “Lycée Technique d’Abidjan”. He started his advanced studies at the Université d’Abidjan-Cocody where he received a “Maitrise 1 des Sciences et Techniques”. In January 2000, he moved to New York City where he attended the City University of New York – College of Staten Island. There, he graduated with two Bachelors of Science one in Mechanical Engineering and the other one in Applied Mathematics with the University system highest award and the “The Giovanni S. Impedulgia Memorial Award for Best Engineering Student” of the year in 2003.

He moved to Atlanta, Georgia to attend graduate school at the Georgia Institute of Technology where he has earned a Master’s degree in Aerospace Engineering (December 2004). He is currently working towards his PhD at the Georgia Institute of Technology in the Aerospace Systems Design Laboratory of the School of Aerospace Engineering.