

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Engineering Sciences

**Development of Automated Bearing Condition Monitoring
Using Artificial Intelligence Techniques**

by

Su Liang Chen

Thesis for the degree of Doctor of Philosophy

September 2009

UNIVERSITY OF SOUTHAMPTON
ABSTRACT
FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF ENGINEERING SCIENCES

Doctor of Philosophy
DEVELOPMENT OF AUTOMATED BEARING CONDITION
MONITORING USING ARTIFICIAL INTELLIGENCE TECHNIQUES
By Su Liang Chen

A recent series of tapered roller bearing tests have been conducted at the University of Southampton to evaluate the effectiveness of using multiple sensing technologies to detect incipient faults. The test rig was instrumented with on-line sensors including vibration, temperature and electrostatic wear and oil-line debris sensors. Off-line techniques were also used such as debris analysis and bearing surface examination. The electrostatic sensors, in particular, have the potential to detect early decay of tribological contacts within rolling element bearings. These sensors have the unique ability to detect surface charge associated with surface phase transformations, material transfer, tribofilm breakdown and debris generation. Thus, they have the capability to detect contact decay before conventional techniques such as vibration and debris monitoring. However, precursor electrostatic events can not always be clearly seen using time and frequency based techniques. Therefore, an intelligent system that can process signals from multiple sensors is needed to enable early and automatic detection of novel events and provide reasoning to these detected anomalies. Operators could then seek corroborative trends between sensors and set robust alarms to ensure safe running. This has been the motivation of this study.

This thesis presents the development of an integrated data fusion and bearing fault detection/diagnostic scheme based on artificial intelligence (AI) methods. The core of this approach is to establish a valid Gaussian Mixture model (GMM) that detects anomalies in new data. To enable this, a large number of features are extracted from the raw vibration and electrostatic signals based on bearing rotation and defect frequencies. These were also used for fault diagnosis at a later stage. A series of baseline tests were conducted on new bearings under various test conditions to generate data for normal model training. A Gaussian Mixture Model (GMM) was used to summarise the probabilistic density of the training data and to build the normal model. During the model training procedure, the random effectiveness of the GMM was reduced by using a systematic evaluation method and by removing externally generated anomalies using novel approaches to eliminate fault masking. Hotelling's T-squared statistic and the associated threshold defining method are then applied to extract prognostic information by measuring the distance between test samples and the origin of the established model. After an anomaly in the test data is detected, the contribution value of each extracted feature is calculated to diagnose fault signatures based on the sensor locations and feature characteristics.

The results from a series of pre-indented bearing tests have demonstrated that: 1) the randomness of the GMM has been significantly reduced and the quality of the model has been improved by increasing the number of candidate models and the initial iterations (up to a certain level); 2) model adaptation methods based on information theory and distance measurement are effective in removing anomalies in the training data to minimise fault masking and rate detections as high as 95% can be achieved with the combination of these methods; 3) the calculated T-squared statistic can detect failure precursors up to 10 hours prior to the fatigue failure which occur after 70 to 100 hours and the new threshold defining method based on extreme value theory can significantly decrease the false alarm rate with approximately 20% lower rates than that of the conventional approaches; 4) The contribution

values of the extracted features can assist the operator in locating faulty bearings and their defective parts as well as indicating the impact ranking of the applied features. These diagnostic results agree well with off-line debris analysis and post-test bearing surface inspection. In addition to the experimental bearing data, multivariate simulated data were generated using a mixture of Gaussian methods and used to further validate the effectiveness of the condition monitoring scheme.

Overall the most significant novelty of this work is that a fully integrated AI based condition monitoring scheme has been developed that is able to extract abnormal events in both training and test datasets, and provide reasoning to the detected anomalies by correlating different sensor features from an instrumented bearing rig.

CONTENTS

Table of Contents	i
List of Tables	vii
List of Figures	ix
Author's Declaration	xvi
Acknowledgements	xviii
Glossary of Key Terms	xix
List of Nomenclature	xxi
1. Introduction	1
1.1. Background and motivation	1
1.2. Research objectives	4
1.3. Thesis outline	5
1.4. Contribution from this study	7
1.5. References	9
2. Literature Review	10
2.1. Rolling element bearings	10
2.1.1. Bearing failure mechanism	11
2.1.2. Rolling element bearing fatigue life	12
2.1.3. Bearing defect frequencies	14
2.2. Machinery maintenance strategies	15
2.2.1. The nature of machinery life and operation	15
2.2.2. Maintenance strategies	16
2.2.3. Concept of condition monitoring	18
2.3. Bearing condition monitoring with multiple sensing technologies.	19
2.3.1. Wear debris monitoring	20
2.3.2. Vibration monitoring	22
2.3.2.1. Statistical parameters in the time domain	23

2.3.2.2. Frequency domain approach	25
2.3.2.3. Time-frequency domain approach	27
2.3.3. Acoustic emission monitoring	29
2.3.3.1. Time domain approach	30
2.3.3.2. Frequency domain approach	32
2.3.3.3. Time-frequency domain approach	32
2.3.4. Electrostatic charge monitoring	33
2.3.4.1. Electrostatic charging mechanism	34
2.3.4.2. Electrostatic wear site sensor (WSS)	35
2.3.4.3. Electrostatic oil-line sensor (OLS)	36
2.3.4.4. Wear detection using laboratory testing rig	36
2.3.4.5. Wear detection using realistic wear rig	39
2.3.5. Comparison of various condition monitoring techniques	40
2.4. Bearing condition monitoring with AI and statistical modelling techniques.	45
2.4.1. Multi-class fault classification	45
2.4.1.1. Expert system	45
2.4.1.2. Numerical Artificial Intelligent methods	48
2.4.1.3. Statistical pattern recognition	52
2.4.2. Anomaly detection	56
2.4.2.1. Construction of anomaly detection model	58
2.4.2.2. Anomaly detection index	65
2.4.2.3. Fault diagnosis	67
2.4.3. Feature extraction and data filtering	67
2.4.4. Comparison and combination of AI and statistical based methods	69
2.5. References	74
3. Methodology and experiment of bearing wear system	89
3.1. Methodology of the condition monitoring scheme	89
3.2. Experiment and Instrumentation	95
3.2.1. General	95

3.2.2. Sensing technologies	99
3.2.3. Data logging and analysis systems	101
3.3. Feature Extraction	101
3.3.1. Time domain features	101
3.3.2. Frequency domain features	105
3.4. Experimental bearing data (4 baseline tests and 7 RTF tests)	107
3.5. References	109
4. Construction of anomaly detection model.	113
4.1. Introduction to the Application of Mixture-model based clustering in tapered roller bearing condition monitoring	113
4.2. Clustering Model Input Patterns	115
4.3. Clustering Models Topologies	117
4.4. A systematic procedure to construct optimized anomaly detection model	120
4.4.1. Model training programme and evaluation	123
4.4.1.1. Model training programme and matrix	123
4.4.1.2. Evaluation indices	126
4.4.2. Examples of the training result	128
4.4.3. Evaluation of the number of candidate models	131
4.4.3.1. Lowest BIC score and training time	131
4.4.3.2. Similarity between the clustering models	137
4.4.4. Evaluation of the initial iterations	140
4.4.4.1. Lowest BIC score and training time	140
4.4.4.2. Similarity between the clustering models	143
4.4.5. Evaluation of the covariance structure	143
4.4.6. Summary	145
4.5. Adaptation of the anomaly detection model	147
4.5.1. Fundamental methodology	147
4.5.1.1. Method based on the information theory	147
4.5.1.2. Method based on the distance between the clusters	149

4.5.2. Demonstration with two dimensional bearing data	150
4.5.2.1. Result of the information theory	150
4.5.2.2. Result of the distance method	154
4.5.3. Optimised approaches	154
4.5.3.1. Entropy statistic associated with multivariate segmentation approach	155
4.5.3.2. Entropy statistic associated with inspection of density changes	158
4.5.4. Application of the proposed approaches to the baseline data	160
4.5.4.1. Result of entropy statistic associated with multivariate segmentation approach	161
4.5.4.2. Result of entropy statistic associated with inspection of density changes	164
4.5.4.3. Result of the distance method	166
4.5.5. Effect of training parameters on the adaptation of the anomaly detection models	166
4.5.6. Summary	169
4.6. Updating of the anomaly detection model	170
4.6.1. Approaches to discover the difference between the two groups of data	170
4.6.1.1. Static clustering	171
4.6.1.2. Dynamic comparison process	175
4.6.2. Approaches to select appropriate dataset to update anomaly detection model	183
4.6.3. Anomaly detection model updating and merging processes	187
4.6.3.1. Concatenating process	187
4.6.3.2. Merging/simplifying process	188
4.7. References	191
5. Anomaly detection	193
5.1. Anomaly detection index	193

5.2. Threshold set-up	194
5.2.1. Approach of using Extreme value theory (EVT) to set threshold for anomaly detection	194
5.2.2. Threshold set-up for distance measure (T-squared distance)	195
5.2.3. Probabilistic boundary set-up	198
5.2.4. Effect of model adaptation on the threshold set-up	199
5.2.5. Comparison of the developed method with the conventional methods	200
5.3. Results of anomaly detection	205
5.3.1. Analysis procedure of the anomaly detection	206
5.3.2. Overview of the anomaly detection results	208
5.3.3. Tests 16 and 18: detection of significant abnormal trends	213
5.3.4. Tests 20 and 13: extraction of information not seen in the original features	222
5.3.5. Summary of the anomaly detection	228
5.4. References	230
6. Diagnosis of bearing faults using statistical techniques	232
6.1. Introduction	232
6.2. Background and method	233
6.2.1. PCA algorithm	233
6.2.2. Fault detection indices	233
6.2.3. Contribution plots	233
6.2.4. Steps of bearing fault diagnosis with PCA algorithm	234
6.3. Results and discussion	235
6.3.1. The usage of applied features for fault diagnosis	235
6.3.2. Establishment of the normal subspace using PCA	236
6.3.3. Case 1 (Expected failure to the inner race of bearing 2, test 16)	238
6.3.4. Case 2 (Unexpected failure to the rolling element of bearing 2, test 18)	242
6.3.5. Case 3 (Unexpected failure to the rolling elements of bearing 4,	245

test 20)	
6.3.6. Case 4 (Almost normal at the end of the test, test 13)	250
6.3.7. Case 5 (Correlation to the AE amplitude, test 21)	252
6.4. Summary	256
6.5. References	264
7. Examination of the developed scheme with simulated data	265
7.1. Characteristic of the simulated data and simulation plan	265
7.2. Approaches to simulate multivariate data with mixture of Gaussians	269
7.3. Summary of the simulated data	272
7.4. Results of the scheme evaluation with simulated data	276
7.5. References	284
8. Conclusions and Future work.	285
8.1. Conclusions	285
8.1.1. Summary of the completed work	285
8.1.2. Conclusions	287
8.2. Future work	290
Appendix	
A. Key elements and flow chart for each main development	
B. Developed approaches and their corresponding developing programs	
C. Method of selecting appropriate data to update model (bearing data)	
D. The Principal Component Algorithm (PCA)	

List of Tables

Table 2.1 Characteristics of the analyzed sensing technologies	44
Table 2.2 Comparison of the analysed AI/statistical methods	73
Table 3.1 Key elements in the developed scheme	94
Table 3.2 Experimental conditions	95
Table 3.3 Calculated frequencies for bearing elements	105
Table 3.4 Extracted features for the data analysis	107
Table 3.5 Datasets utilized within the thesis	112
Table 4.1 Pre-identification for the baseline data	117
Table 4.2 Testing matrix under the diagonal covariance structure	125
Table 4.3 Testing matrix under the full covariance structure	125
Table 4.4 Evaluated parameter – number of candidate models and its evaluation index	131
Table 4.5 Evaluated parameter – initial iteration and its evaluation index	140
Table 4.6 Structure of the context in section 4.5	147
Table 4.7 Posterior probability of the trained clusters and their entropy score with different methods	160
Table 4.8 Summary of the proposed approach to adapt anomaly detection model	169
Table 4.9 Result of cluster purity exploration	175
Table 4.10 Similarity values between the clusters	189
Table 5.1 Summary of the threshold-set results	202
Table 5.2 Summary of the anomaly detection results	211
Table 5.3 Comparison of the anomaly detection of test 16 with different anomaly detection model and threshold set-up strategies	214
Table 5.4 Comparison of the anomaly detection of test 18 with different anomaly detection model and threshold set-up strategies	219
Table 5.5 Comparison of the anomaly detection of tests 20 and 13 with different anomaly detection model and threshold set-up strategies	222
Table 5.6 Summary of the detected anomalies and the issues need to be confirmed or diagnosed	230

List of Tables

Table 6.1 Selected features and their relevant functions	236
Table 6.2 Obtained principal components and its parameters	237
Table 6.3 Significant findings of the two cases relative to original plots	257
Table 6.4 Feature number indication for the colour plot	258
Table 7.1 Data simulation plan and the characteristic features	268
Table 7.2 Summary of the simulation data	275
Table 7.3 Summary of the anomaly detection rates based on different models and threshold set-up strategies	282

List of Figures

Figure 1.1 Detection, diagnostic and prognostic of rolling element bearings	2
Figure 1.2 Thesis structure and layout	7
Figure 2.1 Rolling element bearing components and main geometry	10
Figure 2.2 Rolling element bearings failure mechanisms [5]	12
Figure 2.3 (a) bearing sub-surface crack [14], (b) a bearing surface pit [15]	13
Figure 2.4 Machine life ‘bath-tub’ curve	16
Figure 2.5 Three maintenance strategies: (a) run-to-failure maintenance, (b) scheduled based maintenance, (c) condition based maintenance	17
Figure 2.6 Three key steps in the condition monitoring program	19
Figure 2.7 Histograms of RMS of healthy and faulty bearings	24
Figure 2.8 The process of envelope analysis	26
Figure 2.9 Acoustic emission signal: schematic representation [80]	29
Figure 2.10 Electrostatic charging sensing system [37]	34
Figure 2.11 Possible electrostatic charge sources in tribology [37]	35
Figure 2.12 Precursors of the lubricated sliding contact [100]	37
Figure 2.13 Test results from the long-distance dry sliding test [37].	38
Figure 2.14 Machine operation life time with detection property of the sensors	42
Figure 2.15 Schematic structure of expert system [109]	47
Figure 2.16 Standard supervised training feed forward neural network	49
Figure 2.17 Statistical pattern recognition results: (a) linear discriminant classifier, (b) support vector machine, (c) Bayesian classifier	54
Figure 2.18 Concept of the anomaly detection	57
Figure 2.19 Example of the fault detection based on PCA subspace	66
Figure 3.1 Flow chart of the developed scheme	93
Figure 3.2 Bearing test chamber	96
Figure 3.3 (a) optical image of indent; (b) profile from 2-D Talysurf profilometer with 3 μm diamond tip stylus; (c) optical image of inner raceway and indent location	97
Figure 3.4 Schematics of the whole bearing monitoring system	98

Figure 3.5 Installation positions of the WSS sensors	99
Figure 3.6 RMS values of the 5 applied sensors for all the 14 tests	104
Figure 3.7 Waterfall plots for the frequency spectrum, (a) shows the plot of vibration signal, (b) shows the plot of WSS signals	106
Figure 3.8 Data matrix of the 14 sets of bearing data	111
Figure 4.1 The procedure of using the clustering approach to construct anomaly detection model	114
Figure 4.2 RMS values of the 6 applied sensors for the baseline data	116
Figure 4.3 Histogram of the vibration RMS values of the baseline data	116
Figure 4.4 The programme for finding the best performance of the clustering model	122
Figure 4.5 Concept of the training scheme	124
Figure 4.6 Gaussian mixture models with different candidate models under the full structure	129
Figure 4.7 Gaussian mixture models with different candidate models under the diagonal structure	130
Figure 4.8 Training result of the clustering models under the diagonal covariance and different initial iterations (10-250 (a-f))	134
Figure 4.9 Training result of the clustering models under the full covariance and different initial iterations (10-250 (a-f))	137
Figure 4.10 Similarity measures between clustering models under the diagonal covariance and different initial iterations (10-250)	138
Figure 4.11 Similarity measures between clustering models under the full covariance and different initial iterations (10-250)	139
Figure 4.12 Training result of the clustering models under the diagonal covariance and different candidate models (1-35 (a-f))	142
Figure 4.13 Training result of the clustering models under the full covariance and different candidate models (1-35 (a-f))	142
Figure 4.14 Similarity measures between clustering models under the diagonal covariance and different candidate models (1-35)	143
Figure 4.15 Similarity measures between clustering models under the full covariance and different candidate models (1-35)	143

Figure 4.16 Training time of clustering models with different parameters	144
Figure 4.17 BIC score of clustering models with different parameters	144
Figure 4.18 similarity measures between clustering models with different parameters	145
Figure 4.19 Steps of the entropy based approach to remove anomalies in the training data	148
Figure 4.20 Selected two dimensional electrostatic sensing data (a) and the trained clustering model (b)	150
Figure 4.21 (a) Consisted probabilities of the clusters, (b) Calculated entropy score for each cluster	151
Figure 4.22 Selected two dimensional electrostatic sensing data after the removal of clusters 9 and 11 (a) and the adapted clustering model (b)	152
Figure 4.23 (a) Hotelling's T-squared statistic of the data based on the virtual space, (b) calculated entropy for each cluster with the segmentation in the 'y' direction	153
Figure 4.24 Bearing data after the removal of clusters 7, 9, 10 and 11(a) and their adapted clustering model (b)	153
Figure 4.25 Hotelling's T-squared distance between the clusters (a) and the distance after the removal of clusters 9 and 11 (b)	154
Figure 4.26 Result of the multivariate conditional clustering for the bearing data	156
Figure 4.27 Two dimensional bearing data with 9 assigned categories	156
Figure 4.28 Cluster entropy score based on the multivariate segmentation result	157
Figure 4.29 (a) clustering result of the T-squared Statistic based on virtual space, (b) normal probability plot of the T-squared statistic	158
Figure 4.30 Cluster entropy score based on the density inspection result	159
Figure 4.31 Results of the multivariate segmentation for the baseline data	162
Figure 4.32 Cluster entropy score and the probabilities of the consisted components	162
Figure 4.33 Original model with 30 clusters	163
Figure 4.34 Adapted model with 20 remained clusters	163
Figure 4.35 T-squared statistic of the baseline data (a) and the trained 11 clusters (b)	164
Figure 4.36 Entropy scores of the clusters based on the density inspection result	165
Figure 4.37 Adapted model with 18 remained clusters	165

Figure 4.38 Hotelling's T-squared distance between the clusters (a) And the distance after the removal of cluster 23 (b)	166
Figure 4.39 Anomaly detection rate versus false alarm rate with diagonal covariance structure	167
Figure 4.40 Anomaly detection rate versus false alarm rate with full covariance structure	168
Figure 4.41 Result of vibration clustering analysis	172
Figure 4.42 Result of WSS2 clustering analysis	173
Figure 4.43 Working procedure of clustering based wear detection	176
Figure 4.44 Dynamic clustering process of vibration features (Figures 5(a-i) illustrate the clustering examples of vibration features at particular sampling period)	177
Figure 4.45 Dynamic clustering process of WSS2 features (Figures 6(a-i) illustrate the clustering examples of WSS2 features at particular sampling period)	178
Figure 4.46 WSS2 precursor signal of the inner race energy	179
Figure 4.47 Samples of OLS dynamic clustering process	180
Figure 4.48 Occupation percentage of the new data in the cluster that newly input data occupies most	181
Figure 4.49 Number of the trained cluster	182
Figure 4.50 Inner race examination of the four bearings	183
Figure 4.51 Simulated dataset: (a) Simulated testing data. (b) Simulated training data with 4 Gaussians. (c) Scatter plot of the GMM space and testing data. (d) Hotelling's T-squared statistic of the testing data based on 4 components GMM.	184
Figure 4.52 Samples of clustering result with two neighbouring datasets.	186
Figure 4.53 Occupation probabilities (a) cluster that current data occupies most; (b) cluster that previous data occupies most.	187
Figure 4.54 Anomaly detection results after the model updating with the group of 5 dataset (a), groups of 5 and 11 datasets (b) and groups of 5, 11 and 19 datasets (c).	188

Figure 4.55 Results of the clustering merging: (a) clustering space for the entire training data; (b) Clustering space after the merging; (c) BIC scores for different clustering spaces; (d) T-squared statistic based on the merging space	189
Figure 5.1 Statistical analysis of the T-squared distance from the training data	196
Figure 5.2 Threshold set-up using Extreme value statistic	197
Figure 5.3 Probabilistic boundary after the model adaptation	198
Figure 5.4 Threshold set-up before the adaptation of the model	199
Figure 5.5 Probabilistic boundary before the model adaptation	200
Figure 5.6 Comparison between the EVT and Gaussian PDF	201
Figure 5.7 Flowchart of the threshold and probabilistic boundary set-up	201
Figure 5.8 Threshold set-up before and after the adaptation of the model	203
Figure 5.9 Probabilistic boundary set-up before and after the adaptation of the model	204
Figure 5.10 Approach to select appropriate Gaussian component for anomaly detection	205
Figure 5.11 Example of the T-squared statistic and its detected abnormal features (test 20)	207
Figure 5.12 Evaluation strategy of different models and thresholds	208
Figure 5.13 (a) Weight loss of the failed components, (b) Maximum T-squared value	212
Figure 5.14 Anomaly detection results of the test 16	216
Figure 5.15 Condition of inner race of bearing 2 after the testing 16	217
Figure 5.16 Anomaly detection results of the test 18	220
Figure 5.17 Condition of inner race and rolling elements of bearing 2 after the testing 18	221
Figure 5.18 Anomaly detection results of the test 20	224
Figure 5.19 Condition of inner race, outer race and rolling elements of bearing 4 after the testing 20	225
Figure 5.20 Condition of inner race of bearing after the testing 13	226
Figure 5.21 Anomaly detection results of the test 13	227
Figure 6.1 Eigenvalues of the principal components	237
Figure 6.2 Contribution values of inner race energy of four applied sensors	238

Figure 6.3 Contribution values of the bearing element energies at 3 periods of interest (test 16)	241
Figure 6.4 Contribution values of the OLS features (test 16)	242
Figure 6.5 Contribution values of the bearing element energies at 3 periods of interest (test 18)	244
Figure 6.6 Contribution values of the OLS features (test 18)	245
Figure 6.7 Contribution values of the bearing element energies at 3 periods of interest (test 20)	248
Figure 6.8 Combined contribution values of the different sensors in the precursor area	249
Figure 6.9 Contribution values of the OLS features (test 20)	249
Figure 6.10 Contribution values of the bearing element energies at 3 periods of interest (test 13)	251
Figure 6.11 Contribution values of the OLS features (test 13)	252
Figure 6.12 T-squared statistic of the test 21 data (a) and AE amplitude (b)	253
Figure 6.13 Contribution values of the bearing element energies at 3 periods of interest (test 21)	255
Figure 6.14 Contribution values of the OLS features (test 21)	256
Figure 6.15 Colour maps of the contribution values	263
Figure 7.1 Typical experimental bearing data with characteristic features	267
Figure 7.2 Approach to simulate noise throughout the dataset	270
Figure 7.3 Approach to simulate abnormal trend	271
Figure 7.4 Approach to simulate discrete trend	271
Figure 7.5 Approach to simulate step changes	272
Figure 7.6 Anomaly detection model built by baseline data and GMM	277
Figure 7.7 Multivariate segmentation for the baseline data	278
Figure 7.8 Results of the entropy examination	279
Figure 7.9 Density changes inspection for the T-squared statistic of the baseline data	279
Figure 7.10 Results of the density changes inspection	280
Figure 7.11 Results of the anomaly detection based on adapted model and	283

GMM-EVT threshold

Figure 7.12 Results of the contribution value for the fault diagnosis	284
Figure 8.1 Knowledge structure of the proposed future study	293

DECLARATION OF AUTHORSHIP

I, Su Liang Chen, declare that the thesis entitled

Development of Automated Bearing Condition Monitoring Using Artificial Intelligence Techniques

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Wear detection of rolling element bearings using multiple-sensing technologies and mixture-model-based clustering method. Proc. IMechE Vol. 222 Part O: J. Risk and Reliability, (2008) (highly commended by the editor)

Craig, M., Chen, S. L., Harvey, T.J., Wood, R. J. K., Masuda, K, Kwabata M. and Powrie, H. E. G.: Advanced condition monitoring of tapered roller bearings, part I. 35th Leeds-Lyon Symposium on Tribology, Leeds, UK, (2009), accepted by Tribology International.

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Tapered Roller Bearing Condition Monitoring Using Data Fusion Techniques. 4th World Tribology Congress, Kyoto, Japan (2009)

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: An integrated scheme for adaptation and updating of anomaly detection model. 3rd World Congress on Engineering Asset Management, Beijing, China, (2008)

Chen, S. L., Craig, M., Callan, R., and Powrie, H. E. G. and Wood, R. J. K.: Use of artificial intelligence methods for advanced bearing health diagnostics and prognostics. 2008 IEEE Aerospace Conference, Big Sky, Montana, USA, (2008)

Chen, S. L., Craig, M., and Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Bearing condition monitoring using multiple sensors and integrated data fusion techniques. 9th International Conference on vibration in rotating machinery, pages586–600, Exeter, UK, (2008)

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Wear detection of rolling element bearings using multiple-sensing technologies and mixture-model-based clustering method. 17th Advances in Risk and Reliability Technology Symposium, Loughborough, UK, (2007)

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: An Integrated Scheme for the Machinery Health Management. 13th Materials, Surface Engineering and Bioengineering Sciences Postgraduate Conference, New Forrest, UK (2007)

- parts of this work have been submitted as:

Chen, S. L., Craig, M., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Advanced condition monitoring of tapered roller bearings, part II: with integrated data fusion techniques. 35th Leeds-Lyon Symposium on Tribology, Leeds, UK, (2009), submitted to Mechanical Systems and Signal Processing.

- parts of this work are in progress for submission as:

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Adaptation of the anomaly detection model using hybrid approach. Propose to submit to Engineering Applications of Artificial Intelligence.

Chen, S. L., Wood, R. J. K., Wang, L., Callan, R., and Powrie, H. E. G.: Anomaly detection and fault diagnosis of tapered roller bearings using statistical data-driven approach. Propose to submit Mechanical Systems and Signal Processing.

Signed:

Su Liang Chen

Date:

Acknowledgements

I would like to express my greatest feeling of gratitude to all those who gave help for my research work from the University of Southampton and GE Aviation, Southampton. I would especially like to thank **Dr. R Callan** and **Dr. H Powrie** from GE Aviation, Southampton for their regular technical support and supervision; Dr. T Harvey and Mr. M Craig for their help in providing the experimental data; Professor Philippa Reed from the SES and Dr. Craig Saunders from the ECS for providing valuable advice; my colleagues and friends, Jennifer, Lily, Fitri, Jia, Xiao-Guang, Stephen, Zheng-Fei, Harrison, Sunny, Yue-Ping and Shiva for their advise and help on a whole host of matters. I am grateful to the University of Southampton for the financial support to this project. Huge thanks should be given to my supervisors **Professor R J Wood** and **Dr. L Wang** who have supported me from the beginning right to the end of my study with their tremendous guidance, supervision, encouragement and extreme patience.

Special thanks also go to my dearest **parents** and **girlfriend, Hsiao-Fang** for their psychological and financial support throughout duration of my studies. They were always available at times when I got bit discouraged and helped me cope with all the difficulties that I had encountered during the past three and half years.

Glossary of Key Terms

A/D – Analogue/Digital

AC – Alternating Current

A-D – Anderson-Darling

Adapted model – The model after the model adaptation

AE – Acoustic emission

AI – Artificial Intelligence

AIC – Akaike's information criterion

ANN – Artificial neural network

Anomaly detection rate (true positive rate) – The rate of abnormal events correctly detected to the total abnormal events

Anomalies – Events which are obtained when the components are running under the abnormal conditions

Base line test – The data obtained from the bearing tests with healthy bearings

BIC – Bayesian information criterion

BPFI – Bearing ball pass frequency of the inner race

BPFO – Bearing ball pass frequency of the outer race

BSF – Bearing spin frequency

Candidate models – A series of models for the selection of specific clustering process

Case – Name of a single vector with multiple dimensions with data matrix

CBM – Condition Based Maintenance

CDF – Cumulative Distribution Function

Clusters/Gaussian component – Individual cluster in the clustering model

CPD – Contact Potential Difference

C-S – Chi-square

DC – Direct Current

EA – Evolutionary algorithm

EDF – Empirical Distribution Function

EDMS – Engine Debris Monitoring System

EM – Expectation maximization algorithm

ES – Electrostatic

EVT – Extreme value theory

False alarm rate (false positive rate) – The rate of normal events incorrectly detected to the total normal events

Feature – Name of the physical parameter in the multivariate data matrix

FFBP – Feed forward back propagation

FFT – Fast Fourier Transformation

FLS – Fuzzy logic system

FTF – Bearing fundamental train frequency

GMM – Gaussian Mixture Model

Gumbel distribution – Third form of the EVT distribution

HFRT – High frequency resonance technique

ICA – Independent component analysis

IDMS – Ingested Debris Monitoring System

II – Initial iteration

Initial Iterations – The training iteration for the candidate models

KBS – Knowledge based system

K-S – Kolmogorov-Smirnov

LSM – Least Square Method

Mahalanobis distance – The distance used to find the nearest clusters

MLE – Maximum Likelihood Estimation

MLP – Multi-layer perception network

Multiple clustering – A series of clustering processes with different number of clusters

NC – Number of the trained clusters in the clustering model

NCM – Number of candidate models

Normalities – Events which are obtained when the components are running under the healthy conditions

OLS – Electrostatic oil line sensor

OP – Occupation Probability of the data in the

Optimised model – The model after the training parameters justification

PC – Personal Computer

PCA – Principal component analysis

PDF – Probability Density Function

PODS – Portable oil diagnostic system

RBF – Radial basis function

RCF – Rolling contact fatigue

RMS – Root Mean Square

Run to failure test – The data obtained from the bearing tests with defect bearings

SI – Severity index

Similarity measure – Measure of the similarity between different clustering models

SOM – Self-organising map

SPC – Statistical process control

SPE – Squared prediction error

STFT – Short time Fourier transform

SVD – Singular value decomposition

SVM – Support vector machines

Testing data – The data that are used to test against the anomaly detection model

Training data – The data that are used to build the anomaly detection model

Updated model – The model after the model updating

WSS – Electrostatic wear site sensor

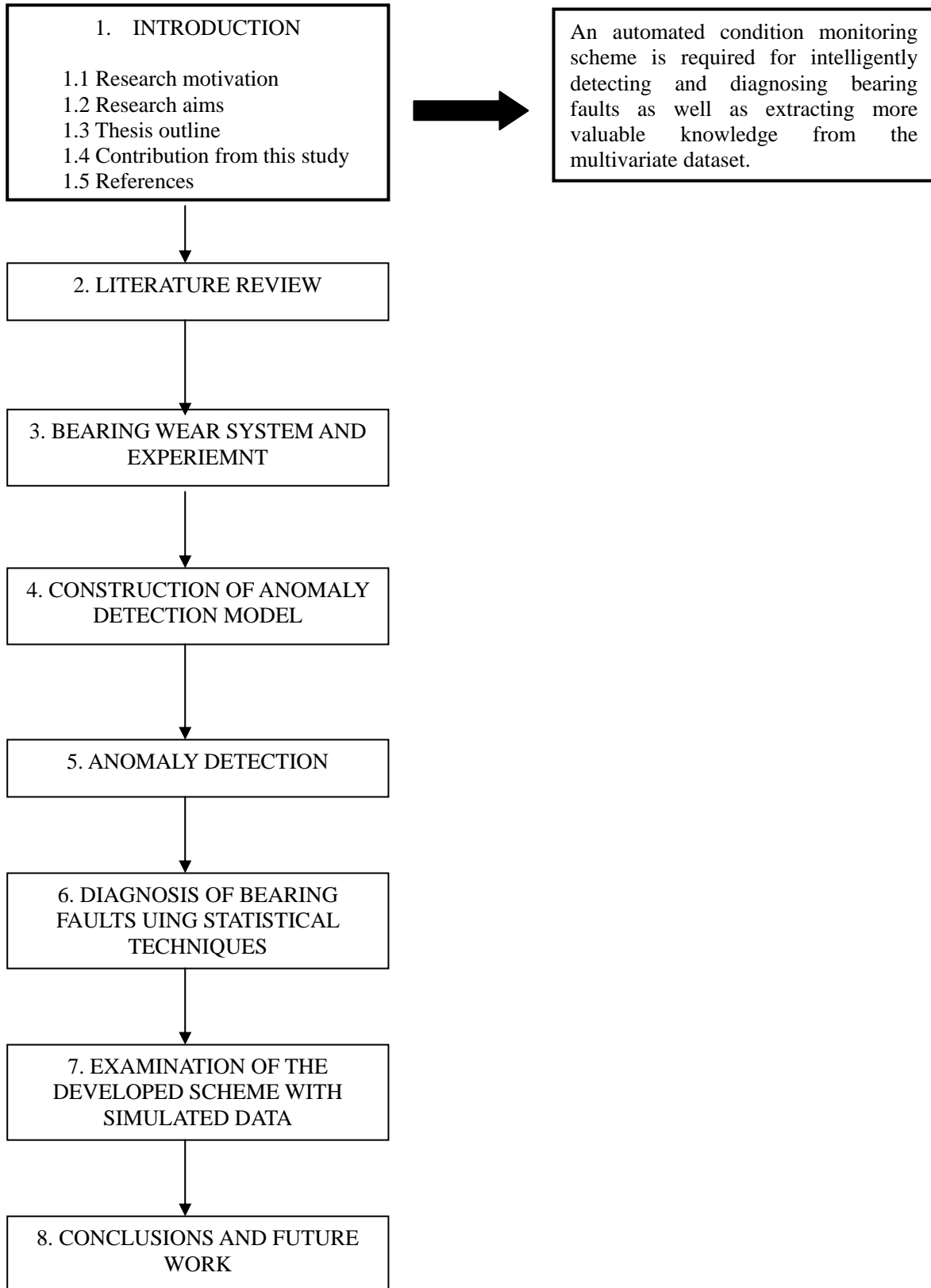
List of Nomenclature

- α Contact angle within the bearings
- $A_i(t_n)$ Membership function of the time related points
- c_j Contribution value of the j^{th} variable to the i^{th} principle component
- d Diameter of the roller within the bearings
- D Pitch diameter within the bearings
- $E(x = v)$ Entropy value for each cluster
- $E(\mu, \sigma)$ Extreme Value distribution function
- $f_k(x_i | \mu_k, \Sigma_k)$ Gaussian probability density function of object i from component k
- $f(x_i)$ Unconditional probability function of the object i .
- k Each Gaussian Component
- L Log-likelihood function.
- m Number of the observed data points which are used to fit the EVT model
- n_c Number of the candidate models
- p Number of the attributes (features)
- $P(k)$ Mixing coefficient of component k .
- $P(k|i)$ Posterior probability of the component k given the object i .
- $P(x; \Theta)_{can}$ A set of candidate models
- P Principal component loading vector
- Σ_k Covariance matrix of component k
- t_n Time ID related variable
- t_i Principal component score vector
- T Fixed time interval
- T^2 Hotelling's T-squared statistic
- μ_k Mean vector of component k
- ω_s Shaft rotational speed in rad /s.
- ω The weight of the Gaussian component
- x_i The i^{th} object in the time series of the training data
- y_i The i^{th} object in the time series of the testing data

- Z Number of rolling elements within the bearings
- Θ Summary of the Gaussian component parameters
- M Location parameter of the Gumbel distribution
- σ Scale parameter of the Gumbel distribution
- λ_i Eigenvalues of the principle component matrix

THESIS STRUCTURE

KEY OUTCOME



Chapter 1. Introduction

1.1 Background and motivation

The main motivation for this PhD project is to improve the reliability of rolling element bearings in complex running environments by developing state-of-the-art multi-sensing data analysis algorithms – detection and diagnostic – and by improving the understanding of the bearing fatigue propagation process in the lubricated environment.

As one of the two main bearing types (another type is journal bearing), rolling element bearings exist in a broad range of applications across almost all industries and they play an effective and important role in all aspects of our modern life. One can expect to come across them in almost every industry, and without any doubt they are the key element of various rotating machines such as driving motors and gas turbine engines.

Rolling element bearings are prone to wear, either in terms of which material is removed from the surfaces of the components, or by which of these surfaces are seriously damaged [1]. These wear phenomena might lead to some modes of failure during the operation, and failures represent a high percentage of the breakdowns in rotating machinery and results in catastrophic failures on many occasions. Such an example appears in the survey of [2], which indicated that 25% of machine shutdowns in the US Naval aircraft are due to bearing failure. It comes naturally then that considerable attention and a lot of research continues to be directed towards bearings, and in particular to their modes of failure, fault detection/diagnosis, fault propagation and remaining life prediction. One significant area, which has helped to detect, diagnose and predict bearing failure, is the implementation of condition monitoring systems, which use data from various measurement techniques to assess the state of the bearings. Condition monitoring is an essential part of Predictive Maintenance, which has been gaining acceptance over the past 30 years. It generally consists of three main stages: fault detection, diagnosis and prognosis. Figure 1.1

illustrates the detection, diagnostic and prognostic parts of the condition monitoring process related to rolling element bearings.

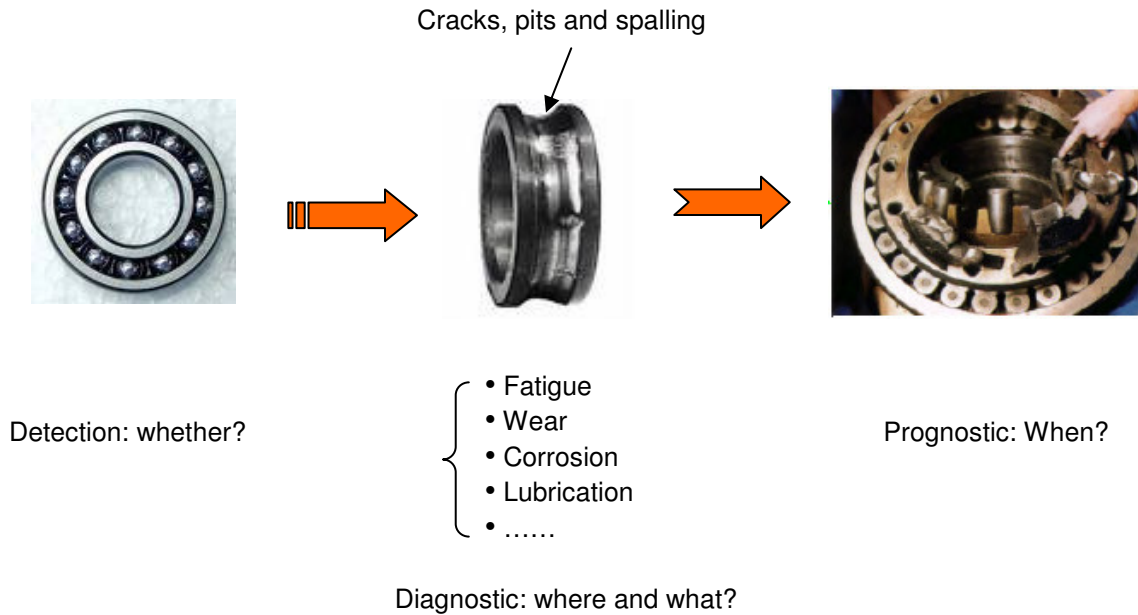


Figure 1.1 Detection, diagnostic and prognostic of rolling element bearings

Measurements can be achieved using various sensing techniques. Traditionally, in the area of condition-based maintenance it is evidenced that the two most significant groups of condition monitoring methodologies used are vibration analysis, and oil and wear debris analysis [3, 4]. Nowadays, more and more new sensing techniques are used to achieve the condition monitoring functions. Acoustic emission (AE) and electrostatic (ES) monitoring methodologies are such examples both showing their unique advantages over the traditional methods. Apart from sensing techniques, signal processing analysis has also been focused on and widely applied to the raw sensing data to extract fault signatures. Research over the past 50 years has established the effectiveness of using multiple sensing technologies for bearing fault detection and fault diagnosis.

However, in modern condition monitoring, the amount of data collected from the sensors is vast with more than one dimensional feature, and it always needs a human expert to analyse the results from the various signal plots and diagnose faults of the monitored components. This leads to difficulties in interpretation and time consuming. Therefore,

the motivation for developing an automated condition monitoring scheme that could aid operators to make the judgment on the health of the monitored bearing components is clear. During the past 20 years, systems developed with artificial intelligence (AI) and statistical modelling algorithms have been playing an increasingly important role in assisting people to solve problems such as minimising uncertainty and achieving automation in both social and industrial applications. On the other hand, recently, at the University of Southampton, a series of tapered roller bearing fatigue life tests have been conducted to evaluate the effectiveness of using multiple sensing technologies to monitor the process of the bearing fatigue and detect incipient faults. The test rig is instrumented with both on-line sensors, including vibration and electrostatic sensors, and off-line analysis such as debris analysis and surface examination. The advantages of using multiple sensing techniques have been demonstrated from conventional data analysis results [5, 6]. However, univariate signal plots only exhibit significant changes when bearings are entering the severe wear stage, and small anomalies that lead to catastrophic failure and do not appear at the extreme tail of a single variable's distribution might be overlooked; this results in the expected precursors becoming indistinguishable from the conventional plots. Thus, it would be extremely valuable if an automatic condition monitoring scheme could be developed so that bearing failure signature could be detected or enhanced at an earlier stage, e.g. bearing fatigue failure initiation. This has been the second motivation of this study.

GE Aviation, as a leading global provider of innovative solutions to manufacturers and operators of military and civil aircraft engines, has recently developed an in-house tool containing several kinds of AI and statistical machine learning algorithms for data mining and fusion. This tool has formed part of the toolset used in the project to develop an automated condition monitoring scheme. The experimental data used in the model development and evaluation in this study are from bearing tests carried out by Dr. Terry Harvey and Mr. Mark Craig in the national Centre of Advanced Tribology at Southampton.

1.2 Research aims and objectives

Ultimately, the research aims to use the numerical data currently collected at a bearing wear rig to develop various techniques to automatically detect and identify tribological related maintenance events (or faults). These techniques may be used in collaboration with the current monitoring system to accurately detect and diagnose abnormal conditions in the course of bearing fatigue life. Technically, this project is to develop an automated condition monitoring scheme, including model establishment, automatic abnormal detection, and fault location finding and fault nature reasoning.

Such a goal requires the development of a series of data analysis and modelling techniques for the observation of data. Consequently, the objectives can be devised as:

- Review the relevant literatures. The research is based on understanding the existing developments in condition monitoring, including modern sensing technology, signal processing, and artificial intelligence and statistical based automated condition monitoring systems. The conduction of these surveys is expected to help the author to find appropriate development strategies and approaches to the identified problems.
- Build diagnostic a training and testing database. Fundamentally, the training data are used to establish the model, while the testing data are tested against the model to evaluate its effectiveness.
- Develop an automated fault detection/diagnostic model with the training data. The model is required to be reliable, consistent and well performed. Thus, a systematic method is necessary to be developed to evaluate the training parameters of the selected learning algorithms and justify the use of the model. Since relevant training parameters have significant influence on the reliability and performance of the model, this development phase is crucial to the whole proposed scheme.

- Develop the methods that optimise the automated detection/diagnostic models, and make the models robust to noise embedded within the training data and changes over time. This requires that the models can be modified and updated throughout the period of application.
- Develop a data fusion technique to fuse applied features and extract or enhance the abnormal trend. This requires use of multivariate approaches which are more powerful, as univariate methods (i.e. single feature) can miss anomalies that do not appear at the extreme tail of a single feature's distribution. The extracted abnormal signatures are also required to connect with the bearing fatigue propagation process and indicate bearing failure mechanisms.
- Develop a threshold defining method, as the proposed scheme is required to automatically identify the significant anomalies and bring these to the operator's attention. The proposed method needs to be sensitive to the tail data in the probability plot, and examined by the fault detection rate and false alarm rate.
- Develop the diagnostic approaches to assist operators in finding the location and causes of the developed faults. It is necessary to investigate the influence of the sensor features on the detected faults, especially for the electrostatic sensors. Since different sensors are related to different detectable fault mechanisms, diagnostic information can be provided in this way.
- Simulate multivariate data with various features. This allows the developed scheme to be examined from a more general perspective, and shows the potential of the approaches which could apply not only to bearing data but to the other areas also.

1.3 Thesis outline

The structure of the remainder of this thesis is summarised below:

Chapter 2 provides a comprehensive review of the latest developments in condition monitoring and fault diagnosis for the condition based maintenance (CBM). Firstly, various sensing technologies used for fault detection or condition monitoring are presented, followed by a review of some state-of-the-art artificial intelligence (AI) based fault detection and diagnosis applications.

In Chapter 3, a summary based on the comprehensive review in Chapter 2 is given, and a proposed methodology for the development of the automated condition monitoring scheme is generated at first. And then, the bearing wear system is introduced, including details of the devices and implemented sensing and signal processing units. Of particular interest is a description of the collected data used to develop the automated condition monitoring scheme.

The task of this project is to develop an automated condition monitoring scheme, including model establishment, automatic abnormal detection, and fault location finding and fault nature reasoning. These are addressed in chapters 4, 5 and 6 respectively.

Specifically, Chapter 4 deals with how to build reliable, consistent and robust automated fault detection/diagnostic models. This chapter involves the topics of model justification by evaluating model training parameters, model adaptation and updating.

Chapter 5 is about automatic anomaly detection, and can be divided into two parts. The first part is the threshold set-up which is used to accurately detect faults by decreasing false alarms and increasing true alarms. The second part is about the data fusion and fault signature extraction. In this chapter, the detected anomalies are specifically analyzed based on the bearing fatigue propagation process.

In Chapter 6, more directed information is provided by reasoning anomaly detection outputs. In other words, the nature of the detected faults is automatically identified with the inputs of the abnormal detection.

Chapter 7 describes the simulation of the multivariate data, and shows the examination results after running through the developed scheme.

Finally, Chapter 8 summarizes key elements of the work presented, and identifies major issues that need to be improved or implemented in the future work.

The thesis is organised into eight chapters, and the general layout of the thesis and topics discussed in each chapter is illustrated in Figure 1.2.

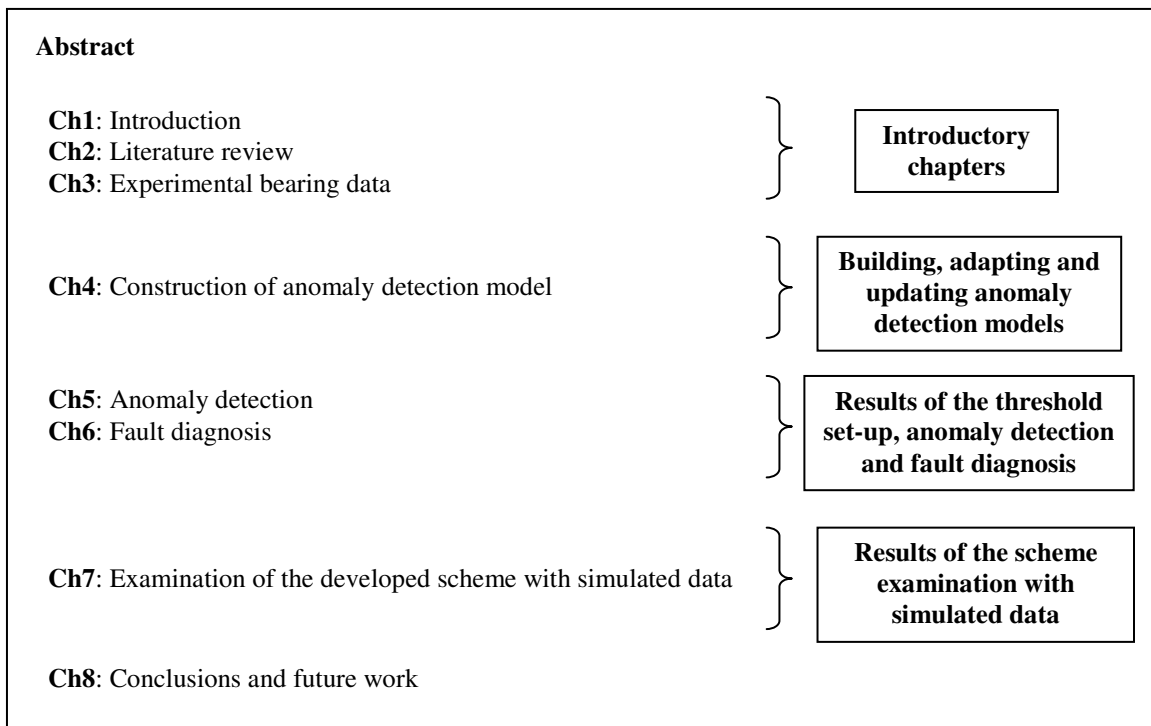


Figure 1.2 Thesis structure and layout

1.4 Contributions from this study

The first contribution of the thesis is the development of a systematic way to construct reliable and consistent anomaly detection model on the basis of the original Gaussian Mixture Model (GMM). Since the GMM-based clustering algorithm involves significant randomness, different initiations with the same training parameters will usually lead to

quite different results. With such a developed method, randomness of clustering can be minimised and consistent solutions can be brought to the anomaly detection in the later stage.

The second contribution is the development of the multivariate-based model adaptation methods using information theory and distance measurements to identify and remove clusters in the Gaussian Mixture Model (GMM) associated with anomalies within the training datasets. This process is considered to be quite important, because not only can the fault masking effects be eliminated, but also the anomaly detection rate can be increased.

The third contribution is a GMM-based model updating method that has been developed for rolling element bearings. The clustering parameters of Occupation Probability of the testing data in the clusters (OP) and the number of trained clusters (NC) are innovatively applied to find the appropriate dataset to be utilised to update the constructed anomaly detection model.

The fourth contribution is a novel threshold set-up strategy that has been developed. The strategy set-up is developed for the anomaly detection indices based on extreme value statistic. The developed strategy can significantly reduce the false alarm rate compared to other conventional methods.

The fifth contribution is the development of a combined statistical anomaly detection system for rolling element bearings. In such a system, the Gaussian Mixture Model (GMM) is used to model the density of the training data to establish the normal model, while additional techniques, including the T-squared statistic and the extreme value statistic, are then calculated based on the normal model as the anomaly detection indices after fusing the applied variables of different sensors, to extract abnormal events from the testing data. With this approach more novel events that cannot be seen clearly in the conventional plots reveal more tribological information that is directly related to bearing failure initiation.

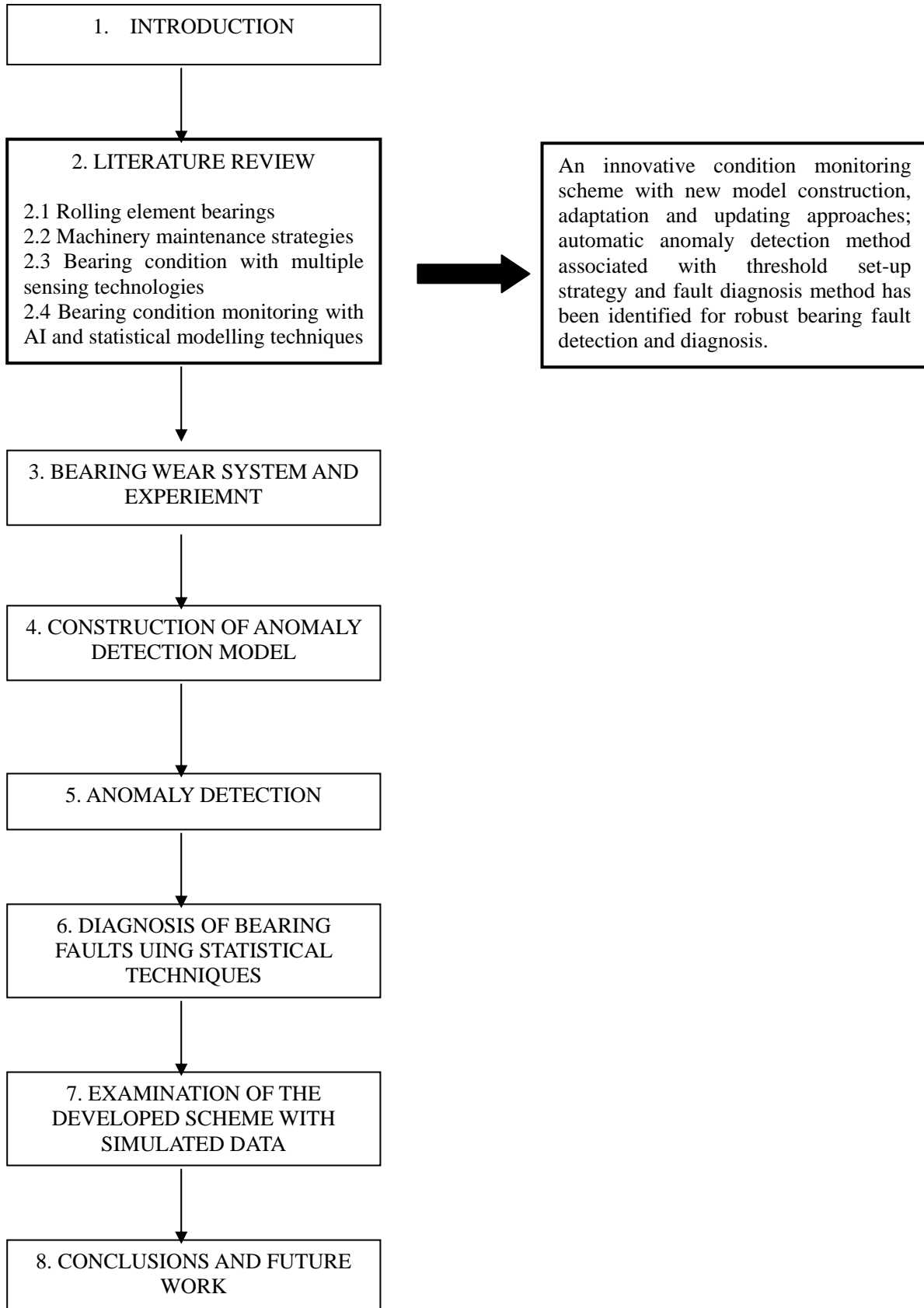
The sixth contribution is the development of a fault localization system for rolling element bearings based on Principal Component Analysis (PCA). The contribution values of the applied variables to the T-squared statistic are calculated, so the truth of which variables are the main causing factors to the anomalies can be revealed.

1.5 References

- [1] M. J. Neale, M. Gee, *Guide to wear problems and testing for industry*, Professional Engineering Published Limited, London, UK, 2000.
- [2] L. E. Alban, *Failure of Gears, ASM Metals Handbook*, 9th edition, Vol. 11, ASM, Metals Park, OH, 1986, pp. 586-601.
- [3] M.J. Neale, *A Guide to the Condition Monitoring of Machinery*, Department of Industry, Council for Terotechnology, HMSO, London, 1979.
- [4] A. Davies, *Handbook of Condition Monitoring*, Chapman and Hall, London, 1997.
- [5] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie. Electrostatic wear monitoring of rolling element bearings. *Wear* 263 (2007) 1492-1501.
- [6] M. Craig, S.L. Chen, T.J. Harvey, R.J.K. Wood, K. Masuda, M. Kwabata, H.E.G. Powrie. Advanced condition monitoring of tapered roller bearings part I, with multiple sensing techniques. Accepted by *Tribology International*, 2008.

THESIS STRUCTURE

KEY OUTCOME



Chapter 2. Literature Review

This chapter aims at giving a comprehensive survey on the subject of applications of various condition monitoring and AI/statistical based techniques for fault diagnosis of the rolling element bearings. A number of literatures are reviewed and classified into four categories. The first category is on publications about the basics of rolling element bearings, while the second category is on maintenance strategies and the need for bearing condition monitoring. The third and fourth parts include literatures about multiple sensing techniques and AI/statistical based decision support systems respectively.

2.1. Rolling element bearings

The term ‘rolling element bearings’ is used to describe a class of bearings in which the main load is transferred through the elements via rolling contact rather than sliding contact [1]. In particular, rolling element bearings exist in a broad range of applications across almost all industries.

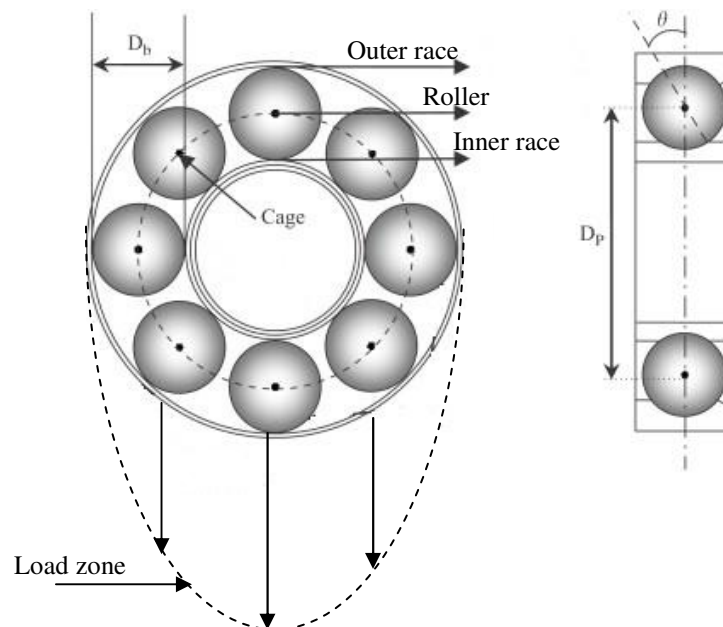


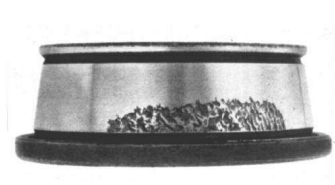
Figure 2.1 Rolling element bearing components and main geometry

There are many types of rolling element bearings, each designed and used for a specific application and load with advantages and disadvantages. The main components of rolling element bearings are the inner race mounted on the rotating shaft, the outer race mounted to the stationary housing, the rolling elements and the cage. The rolling element may be balls and rollers, whereby the balls in the ball bearings transfer the load over a very limited point contact on the raceways, while the rollers transfer the load via line contact with the raceways [2]. Therefore, ball bearings carry lower radial load capacity than roller bearings, but can carry higher axial load. The cage separates rolling elements and prevents contact between them during operation.

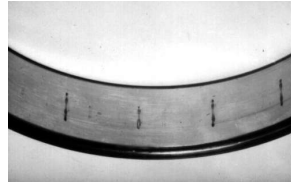
Moreover, the important geometrical quantities of a rolling element bearing include the number of rolling elements Z , the rolling element diameter D_b , the pitch diameter D_p and the contact angle θ . The contact angle θ depends on the type of bearings: it is $\geq 0^\circ$ for ball bearings which carry a radial load, while thrust bearings only carry an axial load with the value of 90° . Figure 2.1 shows these parameters and the components that exist, along with the load zone associated with the fixed outer race.

2.1.1. Bearing failure mechanisms

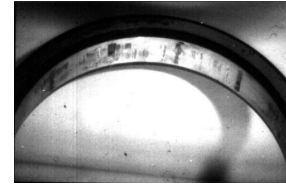
Rolling element bearings may fail due to a number of mechanisms and at different stages through the service life of the bearings. At the early stages, bearings may fail because of inappropriate design, misalignment, overload, false brinelling or contact corrosion. Near the end of bearing service life, bearings failure causes may include: normal rolling contact fatigue, contamination by solids (wear debris) or lubrication film breakdown [3]. Nisbet and Mullet [4] provided a comprehensive overview of the different modes of bearing failure and explained the causes and symptoms of each failure. Figure 2.2 illustrates the images of common bearing failure which are taken with permission from the Neale Consultancy Engineers Ltd website [5].



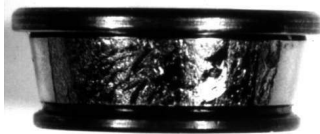
Misalignment of roller bearings causes edge loading and overstressing of the raceway and can result in premature fatigue. This is easily recognized as the damage is clearly restricted to one side



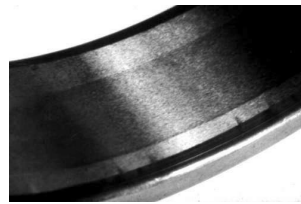
If a bearing is subject to vibration when stationary, fretting damage occurs at the points of contact giving rise to what is known as '**false brinelling**'



Corrosion can occur when the bearing is stationary and is not adequately protected by the lubricant.



Rolling bearings ultimately fail by **fatigue**. This can either be on the rolling elements or on raceways. It starts from a single pit and proceeds from this point.



Abrasive particles in the lubricant will cause **wear**. This is normally most severe with the softer cage material and can lead to cage fracture. Wear damage can be distinguished from lubrication breakdown by the appearance of the tracks.



A characteristic of both types of **lubrication failure** is that the temperature rises through increased friction loss in the bearing. This may allow a breakdown in lubrication to be detected before failure occurs

Figure 2.2 Rolling element bearings failure mechanisms [5]

2.1.2. Rolling element bearing fatigue life

Bearings are one of the most widely used components in rotating machinery, and inevitably suffer the contact surface failure due to rolling contact fatigue (RCF), even they are running in satisfactory environment [6]. RCF occurs due to heavy load applied in the Herzian contact zone, which creates localised defects such as cracks, pits and spalls on the rolling contact surfaces. Operation of rolling element bearings causes fatigue cracks to form within the subsurface material of highly stressed rolling element-raceway contacts. Subsurface cracks propagate and coalesce cause subsequently material removal from the contacting surface. Subsurface cracks are caused by material inhomogeneities, particularly oxide inclusions which are produced in the steel manufacturing process; however, the amount of size of subsurface inclusions has been reduced as a result of the improvements of the steel making process in vacuum procedures [7]. Furthermore, under periodic loading, surface cracks are originated at surface irregularities such as indents or microholes [8]. Surface indentation is the dominant cause of surface originated bearing

failure, and generated debris entrained within raceways and rollers are known as the predominant failure mechanism in rolling element bearings [9].

After initiation, subsurface cracks are stressed at a shallow angle of around 30° by periodic forces, increasing plastic shear deformation which propagates to the fresh metal. During the load cycle, fluid entrapment is another reason for further propagation of the cracks, and the phenomenon found on crack surfaces produces spherical debris with a few microns in diameter [10]. Therefore, impending failure can be detected by monitoring this type of debris. The surface-originated crack extends to the region similar to the layer with maximum shear stress. Ultimately, the subsurface cracks propagate to the surface (see Figure 2.3 (a)), causing removal of surface material. This process is recognised as spallation which generates significant cavities with angular edges, steep sides and flat bottoms [11]; Figure 2.3 (b) illustrates bearing surface pit after spallation.

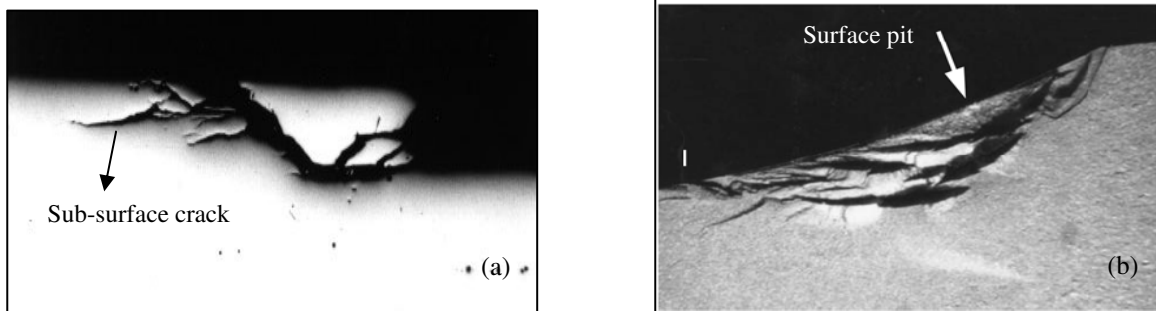


Figure 2.3 (a) A bearing sub-surface crack [14], (b) A bearing surface pit [15]

After the generation of spallation, modifications of surface topography result in an increasing contact stress which then establishes a response that is possibly the dominant mechanism of damage propagation [12]. Spallation is also a reason for the formation of obvious debris types. The major type of debris appears to have a plate-shape with a polished surface and irregular form. Other particle types include spherical ones. Spherical debris generated from the fatigue failure process is usually smaller than $3\ \mu\text{m}$ and several millions of them may be generated in the whole propagation process. However, larger spheres, which are as big as $10\ \mu\text{m}$ can be produced through another mechanism including melting and rapid cooling on the rough surface where flame temperatures exist. Through this process, melted part on the surface is cleared from the main body leading to production of large spherical debris [13].

2.1.3. Bearing defect frequencies

Faults or abnormal conditions induced by the above bearing failure modes give rise to impulses as the components contact the generated defects producing some typical phenomena such as vibration and noise. These impulses are generated periodically and their characteristics are dependent on the location of the defects (i.e. inner race, outer race, rolling elements etc.).

When a unidirectional vertical load is applied to the stationary outer race, the outer race fault would be located in the load zone, and have consistent effects when each rolling element passes through, giving a series of impulses in the signal transmission path. The frequency of generating the pulses is called ‘ball pass frequency in the outer race’.

An inner race fault passes through the load zone at shaft speed, and the series of impulse responses at the ‘ball pass frequency of the inner race’ is modulated by this frequency.

A fault on a rolling element passes through the load zone at cage speed (‘fundamental train frequency’ or FTF) (fixed outer race only), and the series of impulse responses is modulated at this frequency. In this case, the so-called ‘ball spin frequency’ (BSF) is the rate at which the fault strikes the same race.

Equations are used to calculate the defect frequencies (Hz) for the case of the stationary outer race as follows [16]:

- Fundamental Train (or Cage) Frequency (FTF)

$$\omega_c = \frac{\omega_s}{2} \left(1 - \frac{d}{D} \cos \alpha \right) \quad (2.1)$$

- Ball Pass Frequency of the Outer race (BPFO)

$$\omega_{od} = \frac{Z\omega_s}{2d} \left(1 - \frac{d}{D} \cos \alpha \right) \quad (2.2)$$

- Ball Pass Frequency of the Inner race (BPFI)

$$\omega_{id} = z(\omega_s - \omega_c) = \frac{Z\omega_s}{2} \left(1 + \frac{d}{D} \cos \alpha \right) \quad (2.3)$$

- Ball defect Frequency

$$\omega_{re} = 2\omega_b = \frac{D\omega_s}{d} \left(1 - \frac{d^2}{D^2} \cos^2 \alpha \right) \quad (2.4)$$

Where α is the contact angle, ω_s is the shaft rotating speed in rad/s, d is the roller diameter, and D is the pitch diameter, Z is the number of rolling elements.

2.2. Machinery maintenance strategies

2.2.1. The nature of machinery life and operation

To summarise the life of rotating machinery, including fatigue life of rolling element bearings, the classic ‘bath-tub’ curve is usually adopted, and three life stages are generally divided [17], as Figure 2.4 shows. When the machine is new, or after repair, the potential for failure is high, and the time the machine is experiencing is called the running-in period. The machine then goes into a phase of continuous operation with a reduced risk of failure – the steady state and useful life period of the machine. As the rotating machinery begins to wear, the risk of failure increases – the machine goes into its third and then final phase, wear out.

Particularly for bearings, frequent surface contact and interaction between asperities or defect protrusions is usually found in the running-in phase, which leads to the plastic flow of surface peaks, over irregularity of the surface hollows [18]. For this reason, the running condition of the machinery is usually accompanied by impulse shocks which affect the failure rate at a relatively high level. However, during the running-in process, the protuberance on the surface may be polished by the constant passing of rolling

contact, and the machinery then enters into a steady-state stage with a low failure rate. Basically, the steady state running stage should be extended as long as possible and avoiding machinery enters into the wear-out stage which may cause catastrophic failure and breakdown of the machine. To achieve this, improved product design and appropriate maintenance strategy are usually desired to increase the reliability of the machinery.

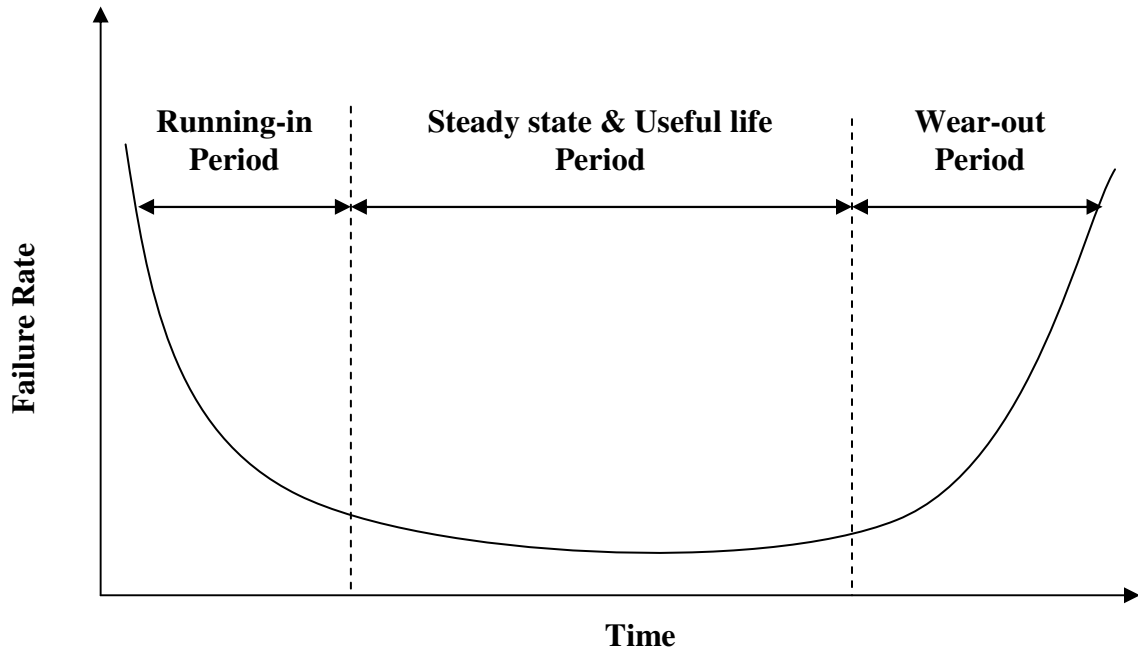


Figure 2.4 Machine life 'bath-tub' curve

2.2.2. Maintenance Strategies

Reliability has been an important and valuable issue in the evaluation of industrial products or machinery for many years. Good product design and manufacturing could obviously benefit higher reliability. However, even the product is perfectly designed and manufactured, occurrence of deterioration is always found over time, due to operating stresses and loads applied in real running environment [19]. Therefore, maintenance is essential to ensure product is operating with an acceptable level of reliability.

The earliest maintenance strategy is called run-to-failure maintenance (see Figure 2.5 (a)), which only takes place when machines break down. This strategy causes unaccepted

downtime, and brings significant loss for the manufacturers and operators. A later maintenance method is time-based preventive maintenance (see Figure 2.5 (b)), which sets periodic intervals to perform maintenance without considering the health of the monitored machinery [20]. However, with the increase in the sophistication level of the modern machinery, unexpected abnormal conditions could not be fully covered by the time-based maintenance. Maintenance actions might be taken too early, e.g. without any evidence of fault or too late after the breakdown. Both situations lead to high cost for the plants [21]. To overcome the disadvantages of these two pre-described maintenance strategies, a more effective maintenance approach, condition-based maintenance (CBM), is implemented, and has been playing a more and more important role in health management of the plant machinery.

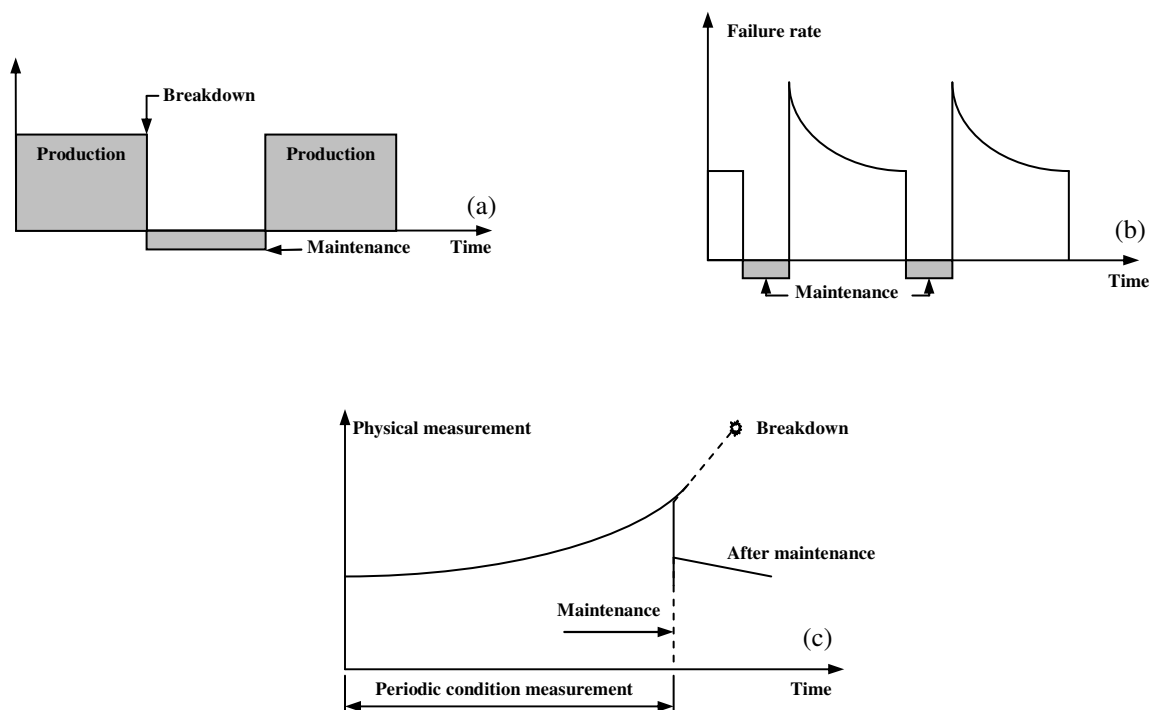


Figure 2.5 Three maintenance strategies: (a) run-to-failure maintenance, (b) scheduled-based maintenance, (c) condition-based maintenance

As Figure 2.5 (c) shows, CBM is a maintenance strategy where the maintenance actions are only permitted when physical measurement shows it to be necessary [22]. CBM tries to collect all the possible physical information and avoid unnecessary maintenance

actions to reduce costs as much as possible. Therefore, CBM is a programme which strongly depends on the collected physical parameters. Various physical parameters could be obtained by condition-monitoring techniques.

2.2.3. The concept of condition monitoring

Since run-to-failure maintenance is almost unacceptable in any industrial plant and time-based maintenance usually causes high cost, as an important part of the condition-based maintenance, condition monitoring was developed in the hope of avoiding unexpected breakdowns and reducing time and financial expenditures.

The definition of Condition Monitoring was given in the British Standard BS3811:1984 [23] as follows: -

'The continuous or periodic measurement and interpretation of an item to determine the need of maintenance. Condition Monitoring is normally carried out with the item in operation, in an operable state or removed but not subjective to major strip down.'

Condition monitoring resolves the shortcomings of the traditional maintenance plans by evaluating the machinery performance criteria, such as temperature, wear rate, pressure, vibration, noise, etc. from the rotating machinery. As a result, service intervals can be extended and scheduled to suit production operations.

Fundamentally, a condition-monitoring programme, as Figure 2.6 shows, includes three main steps [24]:

- Data collection, to obtain data related to the system status through various sensing technologies. Through the 1980s, novel and sophisticated sensing technologies were developed, which employed newly discovered physical properties and high-technology electronics. In section 2.3, various state-of-the-art sensing technologies, particularly used to monitor the health of the bearings, are surveyed.

- Data processing, to analyse data obtained from step one for better understanding and explanation. So far, numerous data processing methods have been specially developed for the specific sensing techniques, and some of them are introduced in section 2.3 followed by review of the sensing techniques.

- Decision support, to provide sufficient information, such as abnormal events, where the faults are in the monitored system, what the causes of the detected faults are and what the remaining lifetime is. These allow the implementation of an appropriate maintenance plan. With the development of modern computation, artificial intelligence and statistical modelling algorithms have been applied to the condition monitoring during the past 20 years to assist humans to facilitate and enhance this process. Section 2.4 reviews the latest developments in this area.

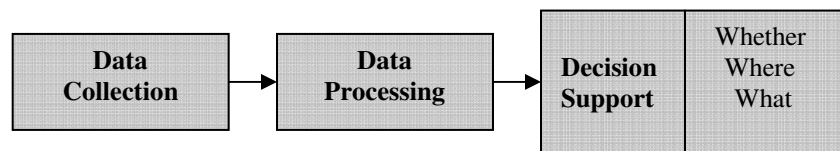


Figure 2.6 Three key steps in the condition monitoring programme

2.3. Bearing condition monitoring using multiple sensing technologies

Nowadays, in order to have an improved and completed understand to the running condition of the rolling element bearings, and avoid catastrophic failure, various sensing techniques have been developed and implemented to monitor bearings in real time. These sensing technologies all have their unique advantages, and play an important role in detecting faults and incipient failure of rotating machinery. Moreover, a large number of parameters are collected and analysed to inspect machine condition, and many studies have been performed to determine which is the most effective one. Unfortunately, not a single parameter has been chosen as being consistently the best. Results have varied

significantly from one application to another. According to [25], the main methods of condition monitoring can be classified into six categories, which are:

- i. *Visual monitoring;*
- ii. *Vibration/noise and shock pulse monitoring;*
- iii. *Wear debris and lubricant monitoring;*
- iv. *Corrosion monitoring;*
- v. *Performance monitoring;*
- vi. *Thermal monitoring.*

Particularly, in this section, the review work is focused on the second and third categories of condition monitoring techniques, since they are not only more relevant to the completed study in this thesis, but are also connected to a new sensing technique - electrostatic monitoring - which requires examination.

2.3.1. Wear debris monitoring

Due to the direct relationship to the cause of the wear, oil analysis is a very useful approach to examine health of the machinery [26]. As a part of the oil analysis, wear debris analysis periodically evaluates oil samples by amounts of measurable attributes: quantity of debris, size distribution, chemical composition and their morphology [27]. These described features are usually applied to establish the relationship to the machinery physical conditions such as wear rate, severity of damage [28], wear mechanism [29] and source of failure [30]. Ferrography is one of the techniques to quantify these features [31], and these features are then used to identify wear conditions of the machinery. An important policy in machinery condition monitoring is to provide early warning of failure; this task can be achieved by comparing the difference of their sizes between normal and abnormal wear debris, because it is assumed that the number and size of the debris collected under abnormal condition increases significantly comparing to that of the normal condition. Although this method has been demonstrated by various applications [32-34], and shown its effectiveness of detecting abnormal wear for a long period of time, debris analysis still receives a great popularity due to its easy-to-operate feature and long-established successful rate in detection.

Nowadays, a series of the latest off-line debris analysis techniques have been developed and implemented to commercial products. HIAC portable oil diagnostic systems (PODS) [35] and Lasernet Fine systems [36] apply optical techniques which is on the basis of the light compensation method, where debris flowing through a light source creates a shadow on a photo-detection unit, which is then used to identify debris size and shape etc. Odahara *et al* [35] used PODS and other off-line approaches for the cooperative purpose to analyse debris generated from bearings and found that during normal conditions, flat trends in the steady state were all with sub 100 μm size indicated by the PODS, which is consistent to the result of ferrographic related indicator, the severity index. However, during abnormal condition, a significant increase in identification of 5 – 50 μm particles was found, correlating with a sharp increase in severity index.

The off-line debris monitoring systems described above allow complicated analysis of oil samples; nevertheless, they require more reasoning support, experienced expert and need to be applied regularly to maximise the usage of this method. Furthermore the wear debris analysis that is dependent on post processing has the restriction in providing real-time information on the dynamic changes of the wear [37]. Therefore, on-line sensing techniques for detecting defects and faults in real time are essential. Nowadays, some on-line wear debris detection devices have been developed and the most commonly used on-line debris analysis is the inductive method. These devices work when oil passing through the transducer coil and the condition of this can be slightly changed as a result of the existence of metallic particles. Ferrous particles increase the inductance because of their high relative permeability, while non-ferrous metallic particles reduce the inductance because of eddy currents occurred in the particles. Thus, the signal phase for ferrous and non-ferrous debris is opposite and enables the device to discriminate between ferrous and non-ferrous debris. Dempsey *et al* [38-39] used an inductive sensor to measure debris size and count particle numbers, and found that the debris mass showed a significant increase when bearing and gear pitting damage began to occur. However, inductive devices are capable of detecting metallic debris typically $>250 \mu\text{m}$ but seem to be insensitive to smaller ones [40]. This shortcoming might result in failure to detect initial stages of material breakdown.

2.3.2. Condition monitoring data collection using vibration analysis

In the early 1950s, vibration monitoring was introduced into various industries, and has been proven to be an effective tool in detecting and diagnosing incipient failures of rotating machinery [41]. As a critical component of the rotating machinery, bearings perform as a cause of vibration because of either modification of compliance or the occurrence of faults within them. Vibrations due to modification of compliance are caused by changes of the loading position when the rolling elements are rotationally operating. This phenomenon gives rise to periodical variation on the raceways. Therefore, the modification of the compliance can produce flexural vibrations no matter what conditions the bearings are in [42], and the signal generated from this type of vibration is usually regarded as the reference or baseline level.

Moreover, the vibration level would be significantly increased if the existence of faults occurs. Two types of bearing faults are usually found and might be classified as 'distributed' and 'local'. Distributed faults generally comprise surface roughness, waviness, misalignment and rolling elements with incorrect size [43-44], which are caused by manufacturing incorrectness, inappropriate installation or abrasive wear with contaminants [45-46]. Localized faults comprise cracks, cavities and spalls on the rolling contact surfaces. Fatigue failure of the rolling elements or raceways is the major failure mechanism of rolling element bearings, and when a fault on an element occurs, sudden changes in the contact forces generate pulses in a very short time. These pulses then derive vibration which can be measured and used to monitor faults in the bearings.

There are generally two approaches adopted to generate localized faults for the observations of vibration signal: the first approach is to test the bearing from good condition to failure and record any changes in their signals [47]. However, this run-to-failure approach takes very long time, as it is not easy for the rolling element bearings or other tribological components to become failure, although some accelerating factors such as over-loading or over-speeding are implemented. The second one is intended to introduce pre-defects on elements of a bearing by pre-indentation, scratching etc, and

then record their parameters to compare with each other or that of good bearings [48-49]. With this method, the pre-introduced defects are varied in size and position to simulate the progress of wear and location of faults, and the vibration parameters are measured to find the correlation with the quantified or qualified defects [49-50].

Whenever an increasing level of vibration level is detected, a further analysis should be carried out to diagnose the possible faults. Traditionally, condition monitoring has been conducted by either a simple overall vibration indication in the time domain or FFT analysis in the frequency domain by using the data processing unit. Furthermore, current signal processing techniques offer a wide variety of methods for analysis of vibration measurements, and aim to extract relevant information from the raw data by transforming it into a recognizable form.

2.3.2.1. Statistical parameters in the time domain

The *time domain* analysis displays how the amplitude of the signal varies over time. Since the amplitude of the transducer's output voltage is proportional to the amplitude of vibration, this analysis shows the vibration amplitude over time [51]. The easiest vibration-based approach in the time domain is the calculation of the root-mean-square (RMS) value. This value can be plotted against time so that any trend can be observed and compared with the pre-defined threshold level. However, as the RMS value represents the overall energy in the signal including noise and other elements that depend on the wear process [52], it may not be the most effective parameter compared with other measuring ones, so it was difficult to find effectiveness for detecting localized faults with this parameter. Some researchers proposed the probability density function for the calculated RMS values [53, 54]. They found that the probability density of RMS in healthy bearing condition follows a Gaussian distribution, while a faulty bearing leads to non-Gaussian distribution (i.e. Weibull or extreme value distribution), which obtains significant tails with lower density. Figure 2.7 shows the typical histograms of RMS values of healthy and inner race faulty bearings. It is found that the data from healthy and faulty bearings are following Gaussian and Weibull distributions respectively.

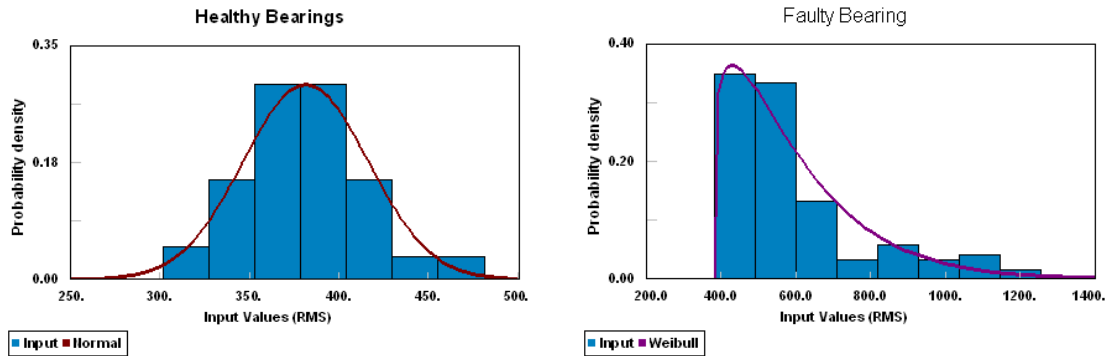


Figure 2.7 Histograms of RMS of healthy (a) and faulty (b) bearings

However, in [54], the phenomenon that data is following Gaussian distribution can also be observed in some bearings with faulty components, and this disadvantage limits the use of identifying bearing faults. Hence, other parameters in the time domain, such as statistical moments, have also been investigated for bearing fault detection. The first and second parameters are the mean value and variance of the examined data respectively. The third moment is normalised and calculated on the basis of standard deviation, named as the skewness. As the fourth moment, Kurtosis is the most frequently used moment parameter for bearing fault detection.

It is because that vibration from a normal bearing is not impulsive and its derived signal generally follows a near Gaussian distribution, this characteristic results in value of the fourth moment-Kurtosis of 3. On the other hand, the vibration from a damaged bearing is impulsive, which affects the distribution, and results in Kurtosis value greater than 3. Dyer and Steward [53] firstly explored the application of Kurtosis for bearing fault diagnosis. They found that the value of Kurtosis was close to 3, if the bearing is in healthy condition. Several other research works [48, 55] also demonstrated the success of Kurtosis in bearing condition monitoring. However, in [53], an obvious disadvantage of the Kurtosis was found that the value usually drops to the level of the healthy bearing even when the defect is further progressed, and unsuccessful to quantify the damaged level. This disadvantage gave rise to a recommendation that bearing component defect frequencies should be taken into account when calculating the Kurtosis; in other words, the Kurtosis should be calculated based on the amplitude at specific defect frequencies

Time domain analysis also contains several other techniques for the bearing fault detection, such as *Crest Factor*, *Spike Energy* [56], *Shock Pulse Method* [57-58], etc. These methods have been applied by the researchers, but have proved not to be 100% reliable, as other factors can also affect the values of the processed parameters.

2.3.2.2. Frequency-domain method

The frequency-domain technique involves frequency analysis of the vibration signal and further processing of the resulting spectrum to obtain clearly defined diagnostic information. In particular, relevant information about bearing faults might be buried within the whole vibration signal, and could not be obtained through the time-domain analysis. Therefore, it is proposed that frequency-domain analysis should be conducted to extract bearing-related information in their particular frequency bands.

Vibration-based frequency domain analysis is the most widely used method for bearing faults detection. Using signal processing techniques to analyze the frequency spectrum, it is possible to determine the defects of the various structural components. Fast Fourier Transform (FFT) has made the intention to obtain the narrowband spectra very effective.

For rolling element bearing defect detection, it is important to note that there is a characteristic rotational frequency for each element within the bearing. If a defect occurs on a bearing element, increasing vibration energy may be found at characteristic frequency of this element. Equations for calculating these frequencies could be found in various publications [59, 60]. However, these characteristic frequencies are normally different from the theoretical values, because of slipping and skidding within the bearings [60]. Therefore, concerns about if these frequencies can be directly applied is raised.

Several researchers [61, 62] have demonstrated effectiveness in bearing fault diagnosis by identifying these frequencies. But in [63], it has been highlighted that it is not easy to extract a significant signature at these frequencies. This is because of the energy involved in cyclic impacts being very weak and usually embedded in much higher levels of vibration generated by macro structural components [64].

To solve the problem discussed above and achieve a more effective frequency analysis, further signal processing is required. Envelope analysis or the high-frequency resonance technique (HFRT) is an important and widely used signal processing technique which identifying bearing faults by detecting abnormal vibration signal at characteristic defect frequencies. Envelope analysis is often explained in terms of amplitude modulation and demodulation [65]. When there are localized defects occurring on the races or the rollers, the vibration caused by impulses becomes amplitude modulated due to periodic changes in the forces. And then the repetitive impacts which have been modulated will be demodulated and will appear in an envelope spectrum.

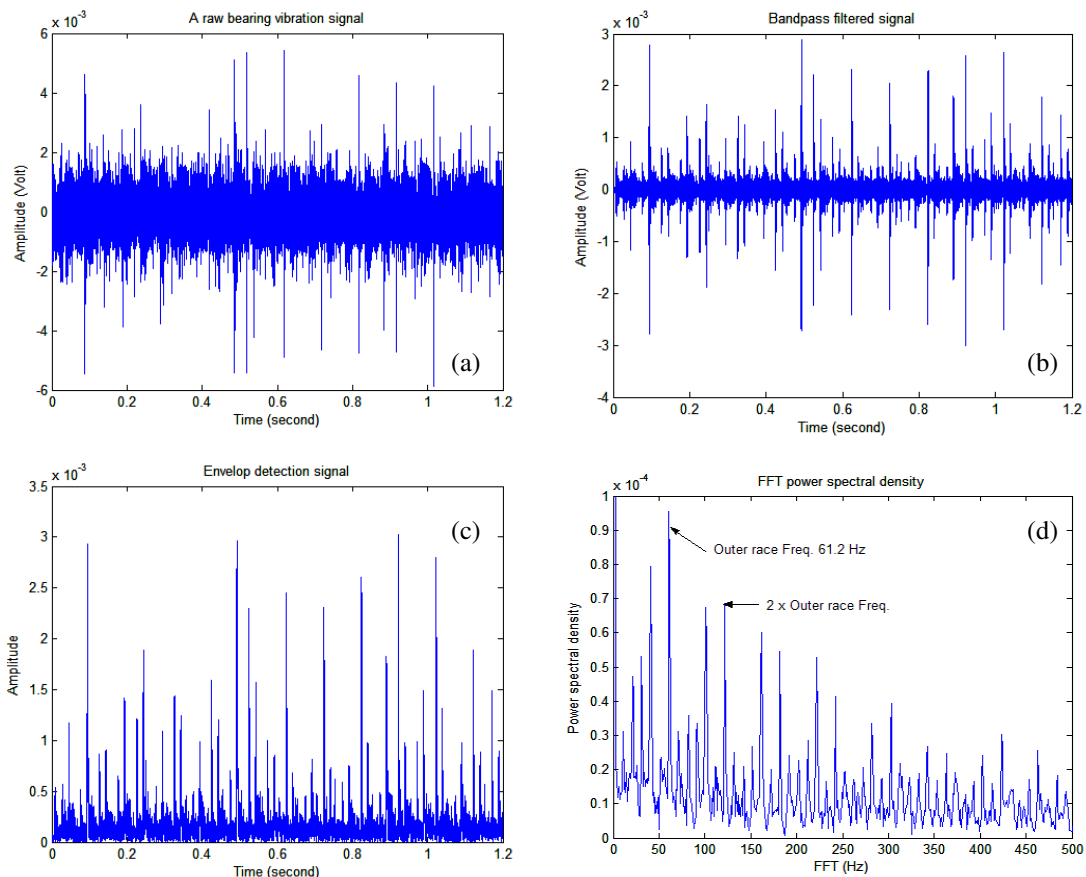


Figure 2.8 The process of envelope analysis

Figure 2.8 shows the four steps of envelope analysis for bearing defect detection. First, the high frequency structural excitation range of a bearing is determined by performing an impact test. Second, the raw vibration signal passes through a band-pass filter centred

around the bearing resonance frequency, so that most of the unwanted signals from other sources can be eliminated to leave bursts of the narrow band bearing signals as Figure 2.8 (b) shows. Next, the band-pass-filtered signal is then rectified by an envelope detector. At this stage, the enveloped time signal should be revealed, as shown in Figure 2.8 (c). Finally, the spectrums of the envelope signal at the bearing characteristic frequencies can be obtained. In Figure 2.8 (d), a bearing outer race defect is identified by the envelope spectrum. The outer race defect frequency and its harmonics (2 times, 3 times...) could be located in the figure. Various applications have demonstrated the effectiveness of envelope analysis in detecting the localized faults of the bearings [66-68].

Generally, the integration of the time and frequency vibration analysis is an effective method to locate the source of the wear problem through frequencies, and to evaluate how serious the damage is by investigating the spectrum. However, as the vibration signal contains too much information, and is hard to classify into different regions through its signal features, this type of condition monitoring technique is therefore inappropriate for the investigation of wear phenomena such as mechanism classification or wear rate evaluation.

2.3.2.3. Time-frequency domain method

Although Fast Fourier Transform (FFT) receives extreme popularity from the researchers and practitioners for applications of vibration monitoring, it is restricted to the stationary signals whose frequency contents do not change over time or exist for the whole duration of the signal. Nevertheless, various kinds of factors, such as the changes of the environment and the faults of the machine of itself, often make the output of signals contain non-stationary components [69]. Therefore, to overcome the disadvantages of the frequency-domain analysis, the time-frequency analysis, such as short time Fourier transform (STFT) with sliding windows and Wavelet transform, has been applied as the most frequently used method to investigate non-stationary signals.

The short-time Fourier transform (STFT) proposed by Gabor in 1946 [70] was the first tool to be used for analyzing the signal in the joint time-frequency domain. The basic idea was to find the spectrum of a signal $x(t)$ at time t by analyzing a small portion of the signal at that time. Wang and McFadden [71] applied STFT to the vibration signal to detect gear failure, and the result shows that the failure of the gear could be detected at an early stage, which is superior to the traditional FFT method. Andrade *et al* [72] compared the performance of STFT with other statistical methods to detect gearbox fault, and found that STFT is an effective approach and performs better than other evaluated methods. However, the crucial drawback inherent in the STFT is that there is a trade-off between time and frequency resolution. Wide windows give good frequency resolution, but poor time resolution. In contrast, narrow windows provide good time resolution, but poor frequency resolution. This problem is the result of choosing a window function, once and for all, and using that window in the entire analysis [23].

Wavelet transform is an alternative method for time-frequency analysis, which is designed to give good time resolution and poor frequency resolution at high frequencies, and good frequency resolution and poor time resolution at low frequencies. As the most frequently used approach, continuous wavelet transform can be defined as [69]:

$$W_x(a, b; \psi) = a^{-1/2} \int x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (2.5)$$

Where $x(t)$ is the input signal, a and b are the scale and time parameters respectively, $\psi(\cdot)$ is a wavelet and ‘*’ represents the complex conjugate. The most commonly used wavelet functions are Morlet, Mexican hat, etc. So far, wavelet transform has shown its effectiveness in condition monitoring and fault diagnosis of the rotating machinery such as gears and bearings. Tse and Peng [73] evaluated the effectiveness of the envelope detection (ED) and wavelet transform for rolling element bearing diagnosis, and the result showed that both methods are effective in detecting bearing faults, but the wavelet transform used less time expenditure. Apart from the standard fault diagnosis, wavelet transform is also applied to fault signature extraction [74], denoising [75], signal

compression [76], etc. and has achieved success at a certain level. It has been indicated by many applications that, due to its adaptive ability, wavelet transform generally performs better than other methods, such as FFT and STFT, etc. However, wavelet transform still suffers from several problems during application. For example, the phase spectrum of the wavelet is not robust to the noise embedded in the signals; occurrence of frequency overlapping is inevitable and usually misleads the analysis of the signals [69]. Moreover, an important issue of how to determine appropriate scale parameters in the wavelet transform is usually ignored, and scales out of ranges may lead to the wavelet transform becoming meaningless [77].

2.3.3. Condition monitoring data collection using the acoustic emission method

Acoustic emission (AE) is a naturally occurring situation caused by external effect, such as mechanical and thermal loading, generates elastic waves because of a rapid liberation of strain energy [78]. The phenomena such as Propagation of cracks, changes of microstructures are usually associated with plastic deformation and are the main causes of AE [79]. Furthermore, the AE signal is usually generated and measured in frequency range that is greater than 100 kHz [78], and this profile has the advantage of detecting the progress of subsurface cracks without significant removal of material. Hence, it is a very effective tool for condition monitoring to indicate the initial stage of the material failure.

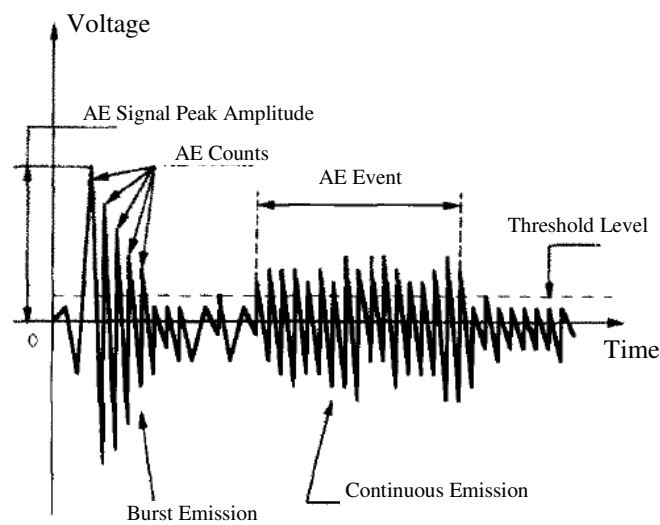


Figure 2.9 Acoustic emission signal: schematic representation [80]

Generally, there are two basic types of signal generated by AE: burst and continuous, as Figure 2.9 shows. These two signals can appear individually or in mixed mode. Furthermore, two types of AE signal analysis are usually recognized: time domain analysis and frequency domain analysis. In real time applications, the use of time domain analysis is common [80], but the amount of research using frequency analysis or even time-frequency analysis has been increasing in recent years. The most frequently used AE quantifiable characteristic in the time domain includes peak amplitude, root mean square (rms), ringdown counts and events [80]. Ringdown counts is the parameter calculating the number of times the AE signal is above the pre-defined threshold line in a specific period of time and provide a simple summary for the observed signal. An AE event not only includes a group of counts but also a transient wave [80]. Furthermore, as mentioned earlier, many signal processing techniques in frequency (e.g. fast Fourier transform (FFT)) and time-frequencies (i.e. wavelet transform) domains have also been applied to the AE signals to extract features for better monitoring.

2.3.3.1. Time domain analysis

The time domain parameters of AE have been widely used for machinery condition monitoring in two aspects, generally. The first one aims to establish links between AE signatures and wear mechanisms or contact conditions for metal-on-metal sliding contacts in real time. The other is intended to use AE responses to detect the local defects of the rotating machinery (i.e. rolling element bearings, gears, etc.).

Various researches have been conducted to establish the correlations between AE parameters and the dominant wear mechanisms or rates of the bearings steel, and certain levels of success have been achieved; for example, McBride *et al* [81] indicated that the material removal from a dry sliding contact was observed with the characteristics of high amplitude, short rise time of AE signals, while in running-in condition, abnormality and plastic flow occurs with lower amplitude signals and much longer rise time. Jiaa and Donfield [82] found that the AE RMS level changed with the change of the running

stages, e.g. running-in, steady state and self-acceleration stage. Sun *et al* [37] stated the investigation into the feasibility of monitoring the single delamination wear mechanism using AE technique, and found that integrated AE RMS could distinguish three wear mechanisms: running-in, delamination/oxidation; oxidation.

Compared with the wear investigation of the sliding contact, AE technique is usually adopted to monitor the local defects of the rotating machinery rather than assigning the specific wear mechanisms. Generally, two approaches have been employed by researchers for simulating localized faults on bearings to study the AE responses and parameters. The first approach is to test the bearing to failure and to observe the changes in their responses. With this approach, Yoshioka and Fujiwara [83, 84] have demonstrated that AE measures could detect the bearing faults earlier than the performance of vibration magnitudes and can also trace causes of AE generation in a ball bearing fatigue life test. They also used AE parameters to detect crack initiation and propagation together with vibration analysis [85]. The other approach is intended to introduce defects with varied sizes on different components of rotating machinery by techniques of mechanical indentation, scratching or other defects, and then record their parameters to compare with each other or that of good bearings. With this method, the pre-indenting defects are varied to simulate the progress of the wear, and the AE parameters are measured to find the correlation with the quantified defects [79, 86]. Choudhury and Tandon [79] employed AE for bearing defect identification on various sized bearings and rotational speeds ranging from 500 to 1500 rev/min. It was found that AE counts increased with accelerating speed for damaged and undamaged bearings while an increasing in load did not result in any significant changes in AE counts for both damaged and undamaged bearings. In addition, they also observed that the parameter of ringdown counts is a very effective parameter for detecting defects in both inner race and roller of the bearings. Al-Ghamd and Mba [87] conducted an experiment in which defects of varying sizes were seeded onto the outer race of the bearings, and both AE and vibration analysis over a range of speed and load conditions were implemented. It was found that AE with the parameters of RMS maximum amplitude and kurtosis offers earlier fault detection and improved identification capabilities than vibration analysis.

Furthermore, the AE technique also provided an indication of the defect size, allowing the users to monitor the rate of degradation on the bearing. Mba *et al* [88, 89] further extended their research by introducing varying seeded defects, a different range of load and speed on the bearing and gear tests. Results showed that parameters of the time domain such as AE counts, RMS and energy performed satisfactory correlation with the pre-defined test conditions.

2.3.3.2. Frequency domain analysis

Apart from the AE parameters of the time domain, the AE power spectrum in the frequency domain has also been used to verify or enhance the analysis. Wang *et al* [90] conducted a test of the bearing steel under the lubricated sliding contact condition, and applied AE and electrostatic techniques to monitor various phases of wear. After the tests, the AE waveform was transformed by FFT and the power spectrum clearly showed an increased power content in a broadband before the onset of severe scuffing; this phenomenon was correlated with the changes of the AE waveform from a continuous signal to a burst type signal. A similar method was also applied by Price *et al* [91], and in their research programme, the AE power spectrum was used to detect a scuffing and pitting fatigue wear regime. The result showed that not only could the two fatigue wear regimes be successfully detected, but a precursor to pitting by the detection of the subsurface cracks could also be possibly measured. However, while dealing with acoustic emission signals, the frequency content is time dependent or non-stationary, and it is usually difficult to process it using these techniques [92]. For this reason, the applications with AE frequency information are found to be limited.

2.3.3.3. Time-frequency domain analysis

Wavelet is one of the most commonly used signal processing techniques in the time-frequency domains to handle non-stationary signals, and has attracted the attention of researchers to apply it to the AE signals for condition monitoring of rotating machinery. Feng and Schlindwein [93] recently developed a Normalized Wavelet Packets Quantifier,

and applied this technique to the AE signal from the faulty bearings. The result shows that, if the appropriate quantifier is selected, both localized defects and advanced contamination faults can be detected and diagnosed, which is superior to the traditional signal processing methods in time or frequency domains. Baydar and Ball [94] applied wavelet transform to both vibration and acoustic signals, and found that acoustic signals processed by wavelet transform are very effective for the prognostic of faults and may provide a powerful tool to indicate the various types of progressing faults in gearbox.

2.3.4. Condition monitoring data collection using electrostatic monitoring

Electrostatic sensing technique is a newly developed on-line sensing approach to detect wear debris. This technique was originally developed by GE Aviation (formerly Smiths Aerospace) for monitoring debris in the gas path of jet engines and turbines [95, 96]. The principle of this is to sense abnormal conditions of electrostatic charge associated with debris by installing a sensor in the jet pipe of a jet engine. These abnormalities may be related to the initialisation of gas path component deterioration, such as blade rubbing, combustion chamber erosion and wear of seal or coatings. This forms the basic concept of the Engine Distress Monitoring System (EDMS). Electrostatic sensors installed in the intake of the engine can be utilised to detect foreign objects entering the engine. This is known as the Ingested Debris Monitoring System (IDMS) [97]. The advantage of the electrostatic system is that it uses a direct measurement of debris generated by engine faults, rather than secondary transmitted effects such as changes of vibration or temperature [40]. The technique can be used to sense various types of debris with different sizes and materials such as ferrous, metallic and non-metallic debris.

The electrostatic monitoring system consists of a passive sensor connected to a signal conditioner (a charge amplifier) from which a voltage signal may be recorded and processed. Figure 2.10 illustrates a diagram of the electrostatic charge sensing system. When an isolated charge particle (E^-) flows through the electrostatic sensor face, an opposing charge (Q^+) will be generated on the probe surface. The internal electrons of the sensor will be redistributed to compensate the additional charge centred around the

sensor face resulting in current flow, which can be quantified and measured. The signal conditioning thus converts the detected charge into a proportional voltage signal [37].

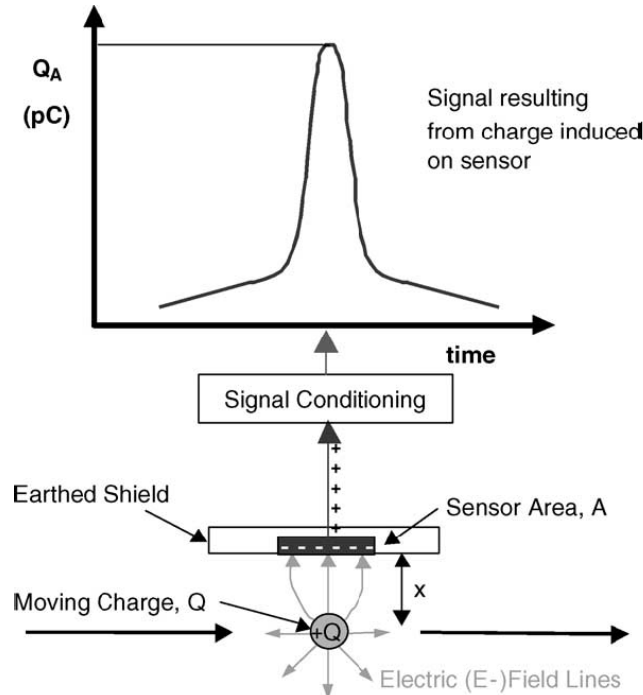


Figure 2.10 Electrostatic charging sensing system [37]

Due to the sensor's finite area, not all of the E -field lines will terminate on the sensor face and hence not all charge $Q+$ will be detected. The dependency of the amount of charge detected by the sensor, Q_A , may be approximated as follows [40]:

$$Q_A \approx \frac{QA}{x^2} \quad (2.6)$$

Where A is the sensor area and x is the distance between the charged particles and sensor face.

2.3.4.1. Electrostatic charging mechanisms

The potential sources of the electrostatic charge in the tribological contacts have been investigated by [98], and found that tribocharging, tribochemistry, contact charging,

triboemission, contact potential difference (CPD) and generated wear debris can all induce additional charges that can be measured. Sun and co-workers [37] summarized possible sources of electrostatic charge and their causes, as Figure 2.11 shows.

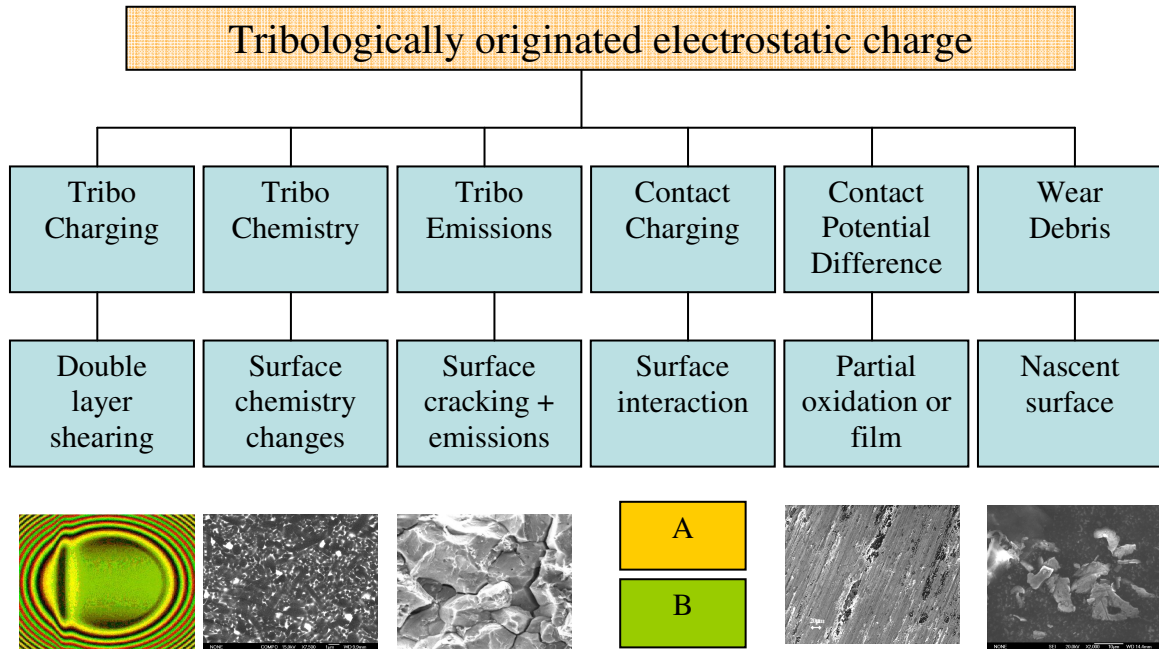


Figure 2.11 Possible electrostatic charge sources in tribology [37].

As a newly developed condition monitoring technique, electrostatic monitoring has its own advantage of directly measuring products of faults produced during component wear processes, especially in the early stages. At the University of Southampton, research has been conducted to evaluate the effectiveness of using electrostatic sensing techniques to monitor the wear of materials. To achieve this goal, two main sensors, electrostatic wear site sensor and oil line sensor, have been developed based on the identification of electrostatic charging mechanisms.

2.3.4.2. Electrostatic wear site sensor (WSS)

WSSs are the passive inductive devices that changes in electrostatic charge in the vicinity of the sensing face [37]. Charge detection is affected by the strength of the charged

source, sensing area and distance from the sensing face. These three factors have been indicated by Equation 2.7. Therefore it is proposed that the sensing face should be located close to critical components, i.e. bearing rolling surfaces, to detect charge that may be correlated to surface distresses or generated debris.

2.3.4.3. Electrostatic oil line sensor (OLS)

In general, electrostatic oil line sensor includes two identical ring sensors located in the oil recirculation system. Ring sensors operate on the same inductive principles as the wear site sensors, with the only difference being the geometry of the sensing face. The sensing face (copper conductor) is wrapped around a glass tube [99], and then, charged sources such as debris or lubricants, or a combination of both, are passed through the bore of the sensor and the charge response is monitored by the usual inductive processes. The benefit of this monitoring tool is being sensitive to sub-100 μm debris and detecting non-metallic particles.

2.3.4.4. Wear detection using a laboratory-based wear rig

Although electrostatic sensing is a relatively new condition monitoring technique, substantial work has been conducted at the University of Southampton for monitoring surface performance of tribo-contacts using laboratory wear rigs. These tests were generally carried out under dry or lubricated conditions, and fundamental wear mechanism identification and early fault indication are the main research objectives.

For oil lubricated tests, electrostatic sensing technique has shown its versatile ability to indicate the onset of material failure [40, 100]. These tests were carried out using laboratory tribometers, and detection of precursor charge activities were achieved prior to the onset of scuffing, which indicates the occurrence of adhesive wear on a surface in the lubricated condition. These precursor charge events were related to one or more of the change sources indicated in Figure 2.12, and could be used as an indication of material failure. Tasbaz *et al* [100] investigated unidirectional and bi-directional lubricated sliding

contacts using pin-on-disc and Plint TE 77 tribometers. The experiment detected the electrostatic precursor charge events up to 1,200 seconds prior to severe scuffing which may eventually lead to catastrophic failure, and found that microstructural changes in contact and regions of increased wear coincided with the occurrence of these precursor charge events, as Figure 2.12 illustrates. Morris and co-workers [40] used two types of lubricants in scuffing tests on the pin-on-disc rig. With both types of oil, they found that additional charge features were developed which are believed to be caused by surface phase charge, associated with first transition scuffing, and hence the onset of failure. They also indicated that surface phase transformed regions will result in a charging mechanism of contact potential difference (CPD) and could be explained as the electrostatic precursor signal prior to scuffing. Furthermore, Wang *et al* [90] investigated acoustic emission and electrostatic responses to the decay of the sliding contact that initially has boundary lubrication and undergoes mild wear but due to oil starvation transforms to a severe wear regime under partial oil entrainment and finally catastrophic scuffing. It is found that electrostatic charge on the wear track shows precursor features at certain locations prior to complete scuffing and failure. This important finding was correlated to the increased power content of the AE spectra before the onset of severe scuffing.

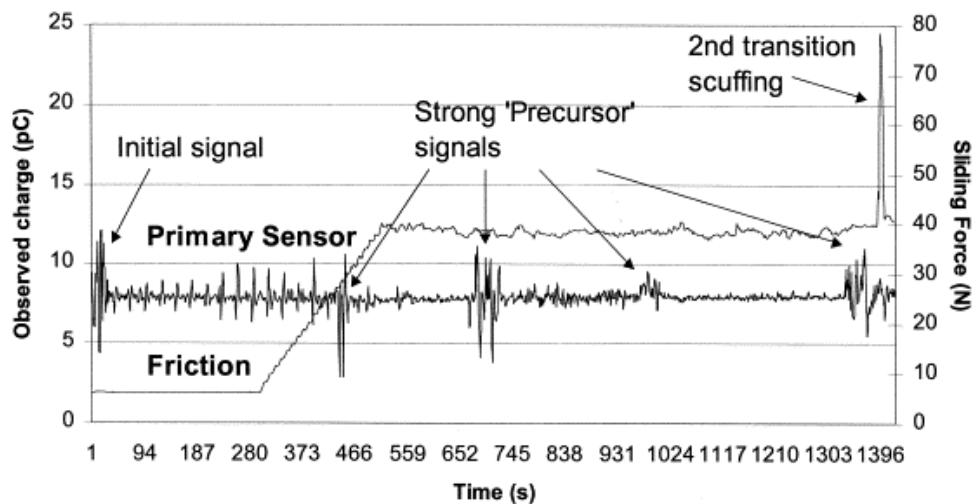


Figure 2.12 Precursors of the lubricated sliding contact [100]

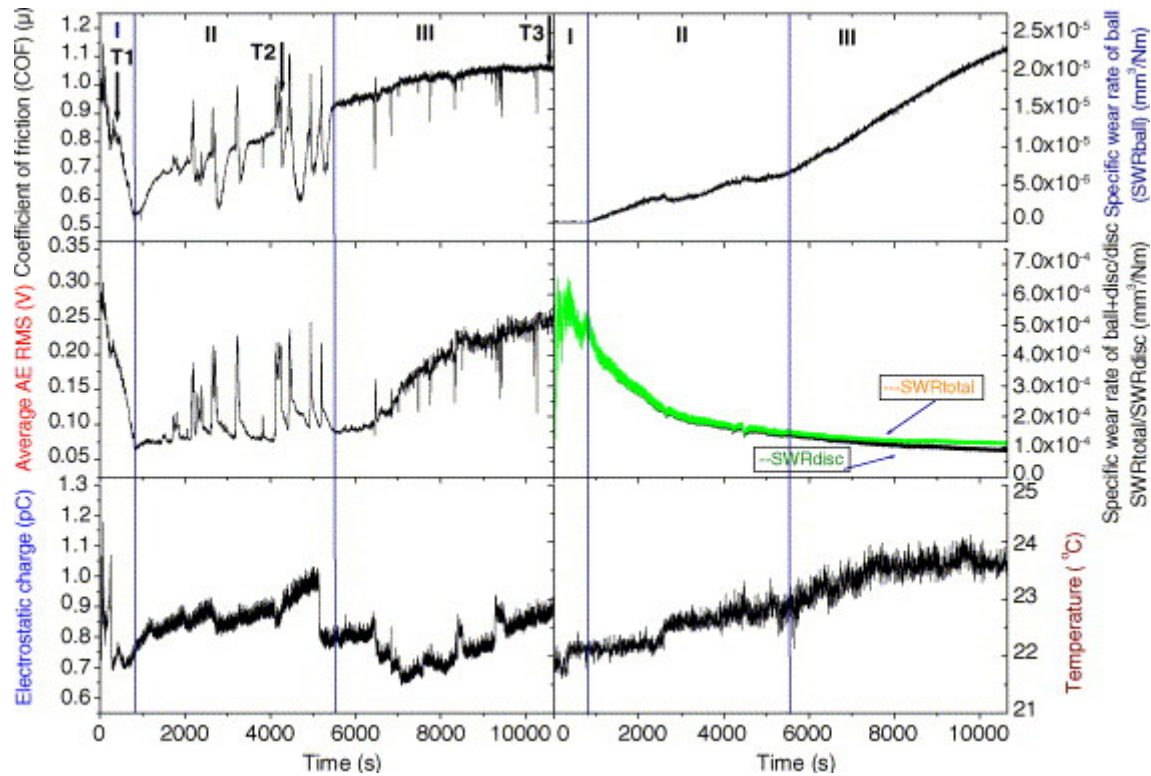


Figure 2.13 Test results from the long-distance dry sliding test [37].

For unlubricated tests, several experiments were conducted using a pin-on-disc wear test rig with conditions designed to generate a delamination/oxidation wear mechanism. Morris *et al* [101] generated a mild-oxidation wear regime, and found that there is a direct correlation between both wear rate and coefficient of friction and the magnitude of electrostatic charge detected. According to the characteristic of the charging signal, four regions of activity were identified. The successful correlation between the electrostatic sensor signal, the delamination wear mechanism and the formation of oxide films proved that electrostatic charge monitoring is a very effective tool to identify the dynamic condition of dry sliding. Sun and co-workers [37] updated the former experiment; they used both acoustic emission and electrostatic sensing technologies to monitor the various phases of generated delamination wear. Tests were conducted using various condition monitoring techniques such as friction, surface temperature, linear wear, acoustic emission and electrostatic signals. The experiment results showed a correlation between measured acoustic emission, electrostatic signals and friction levels and wear rate. Furthermore, three wear regimes were identified through the life of the tests as running-in,

delamination and oxidation. Figure 2.13 illustrates the test result, and distinct acoustic emission RMS and electrostatic charge signal could be found for each regime. It should be noted that during the tests, wear debris samples in each of the region were collected, and microscopy examinations were utilized to examine the worn surface and collected wear debris with the purpose of verifying the wear mechanism in each of the classified regions.

2.3.4.5. Wear detection using realistic wear rig

Powrie and co-workers [120] for the first time applied the electrostatic sensing technique to monitor component degradation in an oil lubricated system. They applied both an electrostatic wear site sensor (WSS) and oil line sensor (OLS) to monitor the wear of the FZG gear rig. During the test, incremental loads were applied to the gear rig every 15 minutes with the charge activity monitored from three WSSs, positioned above, beside and below the gear mesh. Furthermore, OLS was installed in the recirculation pipework to detect generated wear debris. The testing result shows that the charge detected by the WSS is possibly due to phase transformation on the gear tooth surface, rather than being caused by discrete wear debris passing through the sensor. Moreover, the OLS could be used to detect wear debris when large amounts of debris pass through the OLS. According to the testing result, they concluded that increased electrostatic charge signal is detected by the WSS in advance of significant debris being detected by the OLS.

Harvey *et al* [99] extended the wear evaluation work of the electrostatic sensor to a bearing wear rig. Firstly, the bearing was pre-indented on the inner race to simulate the accelerated life test, the WSSs were mounted inside the bearing chamber, and an OLS was added on the oil scavenge line to monitor the wear debris. Besides electrostatic sensors, several other sensing technologies such as vibration, temperature and Macom debris counters using eddy current and ferromagnetism were implemented for rolling element bearing fault detection. From the tests, they found that all monitoring techniques could detect bearing deterioration prior to the function failure. WSSs and vibration could

both detect bearing deterioration 4 hours prior to function failure, while OLS detected increasing production of wear debris 4 hours before the function failure.

As a newly developed condition monitoring technique, electrostatic monitoring has its unique advantage of directly measuring the products of faults, rather than secondary effects such as vibration and temperature exceedance [40]. Furthermore, the collected sensing data could be utilized to investigate the wear mechanisms in the sliding contact wear. Furthermore, electrostatic sensors could detect primary fault mechanisms such as surface phase transformation associated with precursor charge that is undetectable for the other sensing technologies.

2.3.5 Comparison of various condition monitoring techniques

In this chapter, the most frequently used condition monitoring techniques in the field of machinery condition monitoring have been reviewed. Each of the techniques has a unique benefit, and should play an appropriate role on detecting faults during condition monitoring.

Wear debris analysis is the most visual inspection method that directly reflects the condition of the wear components, but this method is of limited capability in providing real-time information on the dynamic wear process, and is not suitable for on-line monitoring. Although on-line debris counters have been developed to accumulate generated debris, they normally could only detect debris with a size of over 100 μm associated with severe fault conditions, and sometimes overlook other fault mechanisms such as fatigue initiation and crack propagation which are free of debris. Despite these concerns, wear debris analysis is still a popular abnormality detection technique, as it could provide solid physical evidence supporting other on-line detection techniques. As discussed in section 2.2.3, due to the lack of obvious identifying features, vibration signals are difficult to use to identify the wear mechanisms of the steel sliding or rolling contacts, but statistical parameters such as RMS and Kurtosis of the time domain and power spectrum calculated through FFT and the envelope technique in the frequency

domain have been widely and successfully applied to detect and locate growing faults. This method is especially useful in sophisticated systems that contain a number of machinery components. Wavelet transform is another popular signal processing technique for vibration analysis, but as [69] indicated, a number of parameters need to be justified, which involves complex calculation; otherwise, the obtained results could be varied without a consistent solution. For this reason, most practitioners still prefer using FFT as the standard diagnostic tool for the vibration analysis. Moreover, vibration could only detect faults with the frequency range up to 50 kHz, and the effectiveness of this technique is that it could only detect surface cracks or other severe fault conditions such as spallation. As a high frequency detection technique, acoustic emission (AE) obtains the advantage of detecting lubrication film breakdown and sub-surface cracks, which induces high frequency signals. To analyse AE signals, data processing techniques in time, frequency as well as time-frequency domains have been utilized. Among them, time domain parameters such as events and counts have been used not only to identify different wear mechanisms of sliding or rolling contacts, but also to quantify defects indicating severity of faults. Nevertheless, very few AE applications with frequency analysis were found, because the AE signals are very non-stationary, and frequency bands of the tracked information vary with the changes of time, which leads to the extraction of frequency-related information difficulty. It has been acknowledged that the detection capability of acoustic emission monitoring is superior to vibration analysis; the reason for this is that the AE can detect the initial failure characteristics such as lubrication film breakdown and progress of sub-surface cracks, while the latter one can only detect defects when they are found on the contacting surfaces. Moreover, as a newly developed approach, electrostatic (ES) monitoring is developed directly relating to the wear of material, and various charging mechanisms have been investigated and are well known [32]. So far, many applications have shown the versatile characteristics of the ES technique; not only can it identify different wear mechanisms and sense generated wear debris, but also some fault mechanisms such as surface phase transformation, sub-surface crack and spallation can be detected by electrostatic sensors. Similar to the vibration analysis, the statistical parameters of the time domain and power spectrum of the frequency domain have also been utilized to analyse ES signals.

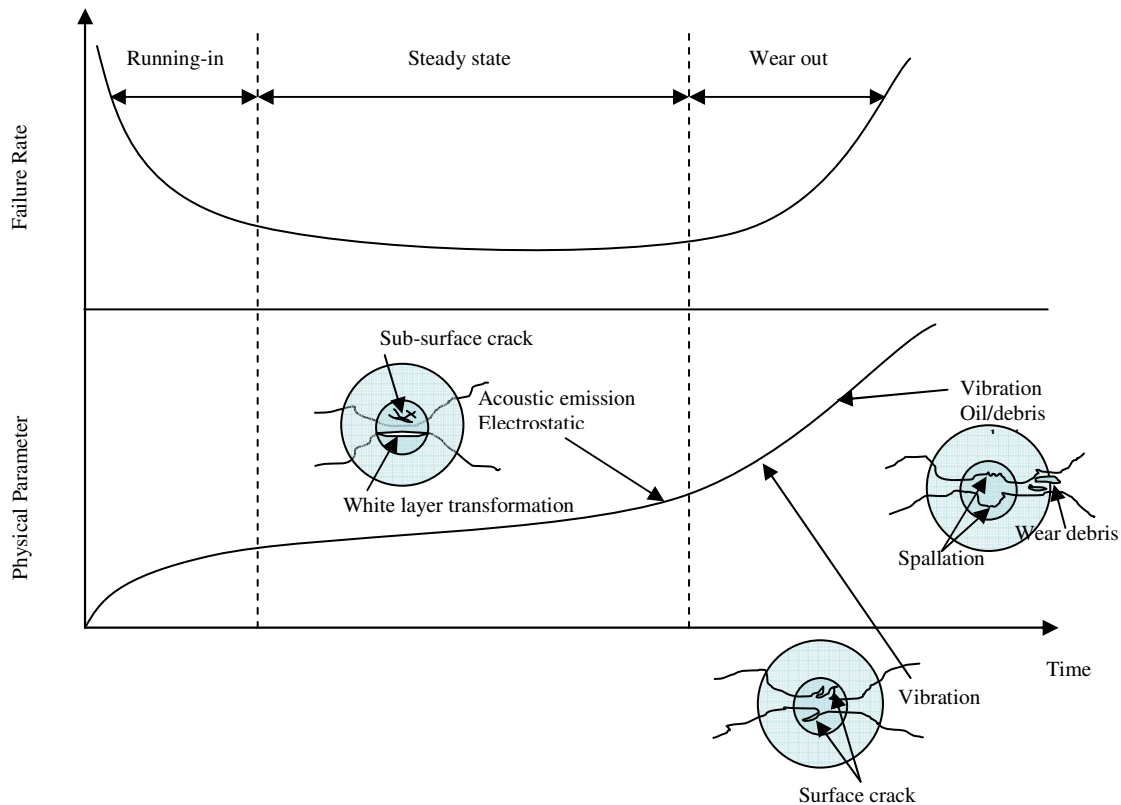


Figure 2.14 Machine operation life time with detection property of the sensors

According to the above analysis, each sensing technique has an advantage with appropriate signal processing techniques and could not be completely replaced by any one of the others. As Figure 2.14 shows, they could be used to detect relevant fault mechanisms in the operation life of the machinery. Although some of their detecting functions overlap, they could still play a complementary role to each other, and enhance the confidence of fault detection. Table 2.1 summarizes the characteristics of these condition monitoring technologies. In modern condition monitoring, multiple sensors are implemented to get the most out of the monitored machinery, and if it is possible, it is recommended that the above described sensing technologies are applied to play complementary as well as cooperative roles.

Therefore, in this thesis, the outputs of the multiple sensors which were equipped to the experimental bearing rig will be examined, and the fault detection and diagnosis analysis are obtained by multiple sensing data fusion and correlation evaluation.

However, it is obvious that multiple sensors bring multivariate data pictures which increase the complexity of the analysis. Moreover, the data plots sometimes could only indicate severe fault conditions; some fatal abnormalities that lead to catastrophic failure are not significant, and might be overlooked. Under these circumstances, it is suggested that an automatic fault diagnosis scheme should be developed to fuse all these sensor features, provide decision support if abnormal events occur, and find out where the fault is and what the causes are.

Sensing Technology	Application Status	Signal Processing	Fault mechanism detection
Wear debris analysis	Off line analysis for wear mechanism identification, but time consuming On line analysis with debris counter to detect debris with different size, but normally detect severe conditions.	Characteristic ferrography (severity index) Optical technique (PODS) Inductive technique (on-line debris counter)	Different type of debris can be linked with different wear mechanism Different size of debris can be linked with wear severity
Acoustic Emission	On-line technique for the wear mechanism analysis On-line technique for the high frequency fault detection	Time domain : RMS, Counts, event Frequency domain: Power spectral density Time-frequency domain: Wavelet transformation	Different features of the signal can be linked with different wear mechanisms. Signal can be used to detect sub-surface crack and the propagation of defects. As a non-stationery type of signal, frequency content of the AE is usually difficult to extract.
Vibration monitoring	Most frequently used method for the machinery on-line condition monitoring. Well-suited for trend monitoring, with considerable scope for performing diagnostic-type analysis.	Time domain: RMS, Skewness, Standard deviation, Kurtosis Frequency domain: Power spectral at different frequency band Time-frequency analysis: wavelet transformation	Collected features can be used for surface crack detection Power spectrum at different frequency bands can be used to locate the developed faults
Electrostatic monitoring	Newly developed technique for wear mechanism analysis and machinery fault detection. Originally developed to detect debris of the gas path engines, now have been applied to monitoring lubricated transmission systems	Time domain: RMS, skewness, standard deviation, kurtosis Frequency domain: Power spectrum at different frequencies	Has the privilege to detect surface phase transformation, sub-surface cracks and the generated ferrous and non-ferrous wear debris of the monitored machinery. Directly relate to wear, and charges could be detected through several well-known mechanisms.

Table2.1 Characteristics of the analysed sensing technologies

2.4. Automatic Machinery Fault Diagnosis Using Artificial Intelligence (AI) and Statistical Modelling Tools

It should be noted that the research method summarized in section 2.3 is a fault diagnosis or wear evaluation process using human knowledge and intelligence. The problem of how to design the automated decision support system assisting people to make appropriate maintenance plans through advanced methods has drawn great attention from researchers in the past 20 years. Artificial intelligence (AI) and statistical-based methods have been the most popular ones, and have been used to develop a great number of automated systems. Techniques for the development of condition monitoring systems can be generally classified into two categories: diagnosis and prognosis. Fault diagnostics is mainly on detection and recognition of faults when they occur. On the other hand, prognostics attempts to predict future status and remaining life time of the machinery. In this section, the author would like to review the most popular and frequently used AI and statistical-based methods for fault diagnostics. The applications in prognostics are not intended to discuss as they are not quite relevant to the research introduced in this thesis.

2.4.1. Multi-class fault classification

As an important development strategy of the diagnostic system, multi-class fault classification has attracted major attention from researchers. The fundamental concept of the multi-class fault classification system is to try to map multiple datasets representing different symptoms of the machinery into multiple categories through various pattern recognition algorithms.

2.4.1.1. Computational expert system

The expert system (ES), also known as knowledge-based systems (KBS), is basically a set of computer programs embodying knowledge about a narrow domain for the solution of problems related to that domain [103]. An ES mainly consists of a knowledge base and an inference mechanism. The knowledge base contains domain knowledge, which may be expressed as any combinations of 'IF-THEN' rules, factual 'statements, frames, objects, procedures and cases, while the inference mechanism manipulates the stored

knowledge to produce solutions. The classical expert system usually requires a large amount of human development time and knowledge for even moderately complex systems.

Su *et al* [104] in 1993 developed a very user-friendly expert system for tribological failure diagnosis of spur gears. It was designed to evaluate a human expert in spur gear failure analysis and can be used by non-experts to determine the gear failure mechanism, to provide preliminary leads into conducting root-cause analysis, and to recommend corrective actions. In this system, IF-THEN rules were heavily used to connect the morphological attributes of the wear debris and classified failure mechanism. Walton [105] developed a diagnostic system to diagnose probable causes for bearing failure, using a large matrix of bearing failure phenomena and related causes, in which each cell contains a 0-100% confidence level that such a phenomenon would suggest a given cause. A knowledge-based expert system is developed which incorporates a magnetic chip detector (MCD), ferrography, filtergram, or filter debris analysis (FDA) [106]. Results are obtained and advice is offered to the user by means of logical reasoning of an experienced analyst in determining the type of metallic material and wear condition of the components based on the morphological attributes of the wear debris.

Expert systems are sometimes developed together with other types of numerical techniques. The combined expert and neural network systems are found in [107-109]. In these applications, the neural network system is designed to extract wear debris morphology information in terms of visual features, while the expert system links the morphological attributes extracted from debris analysis to human expert analysis that is embedded in a knowledge-based system for wear assessment. Figure 2.15 illustrates the flow chart of a combined expert system for wear assessment [109]. Expert systems are usually utilized to evaluate wear, when wear debris analysis is used as the condition monitoring technique. This is because the extracted features of the collected debris samples are relatively discrete and more convenient to be used for classification; furthermore, the wear mechanism classification could be conducted accurately using the IF-THEN rule. On the other hand, fuzzy logic is at the edge of probability theory and the

border of expert systems. The fuzzy logic systems (FLSs) are based on a set of rules. These rules allow the input to be fuzzy, i.e. more like the natural way that humans express knowledge. The use of fuzzy logic can enable ESs to be more practical and knowledge in an ES employing fuzzy logic can be expressed as fuzzy rules (or qualitative statements). A reasoning procedure, the compositional rule of inference, enables conclusions to be drawn by extrapolation or interpolation from the qualitative information stored in the knowledge base. For tribological components [110, 111], fuzzy subsets and corresponding membership functions are used to describe the chosen parameters (i.e. vibration, acoustic emission, etc). Liu and co-workers [110] combined an ES and fuzzy logic classifier to identify bearing faults using the vibration features. In their work, a knowledge base consisting of rules and databases is built to support the fuzzy inference.

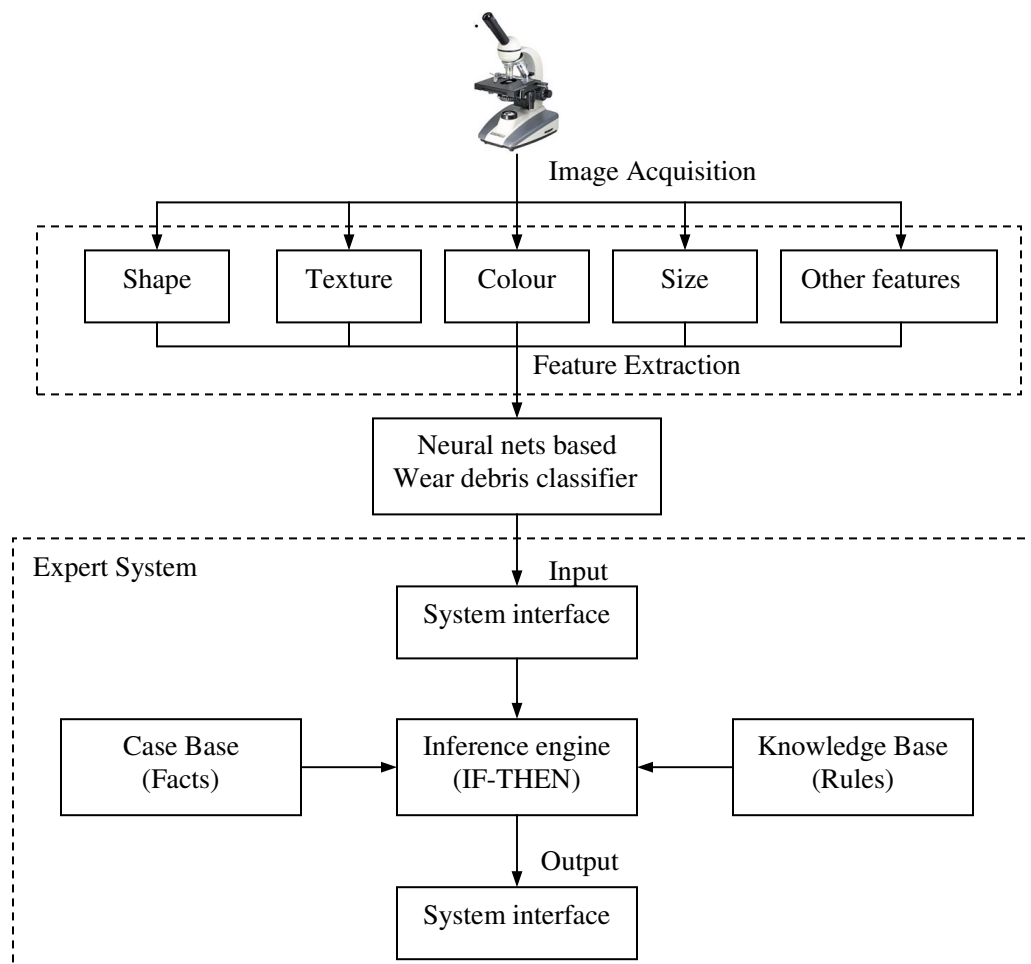


Figure 2.15 Schematic structure of expert system [109]

However, the ES systems need rules to be developed by experts, and this is still extremely time consuming. A small error in these complex rules can produce a significant error in diagnosis; and expert system cannot readily adapt to changes in the operating environment because it has a rigid decision boundary as defined by the human experts.

2.4.1.2. Numerical Artificial Intelligence (AI) approaches

Over the past three decades, various numerical AI techniques have been substantially investigated and applied to machinery condition monitoring. As these intelligent methods are developed with human-like reasoning and decision-making abilities, the applications have shown superior performance over traditional computational approaches. In the literature, the most popular and advanced AI techniques for machine condition monitoring are fuzzy logic systems, artificial neural networks (ANNs), neural-fuzzy systems and evolutionary algorithms (EAs).

(1) Fuzzy logic system

Fuzzy logic is an artificial intelligence technology that emulates the qualitative and inexact nature of human reasoning [112]. The technology applies fuzzy set theory to data, where fuzzy set theory is a theory of classes with rough boundaries and the data belongs in a set based on its degree of membership [39]. The degree of membership can be any value between 0 and 1. Dempsey *et al* [113] used fuzzy logic to develop a feature that defines bearing damage levels based on the accumulated mass measured by the oil debris monitoring system. And then membership values based on the accumulated mass measured by the oil debris sensor were defined for three levels of bearing damage: damage low, damage medium and damage high.

(2) Artificial neural networks

Although the first artificial neuron was produced in 1943 by the neuro-physiologist Warren McCulloch and the logician Walter Pitts [114], only in the past two decades has the real potential of Artificial Neural Networks (ANNs) been realized, and it has attracted more and more attention from the scientists and technologist. Consequently, ANN has

been widely applied to industrial fields such as automatic control, speech processing, and pattern recognition etc.

An ANN is a computational system that is inspired by the structure of the human brain, and it processes information in a similar way to the human brain. It comprises a number of basic processing nodes connected with each other in a complex multi-layer structure which enables a complicated non-linear function to be trained with multi-input and desired multi-output. The ANN then learns the non-linear function by adjusting the weights which are located between the nodes with obtained input and desired output. This procedure is usually named as ANN training. There are various types of neural networks, and the most commonly used ones for pattern recognition and signal processing within the engineering applications are: Multi-Layer Perception networks (MLP), Radial Basis Function networks (RBF) and Self-Organising Maps (SOM).

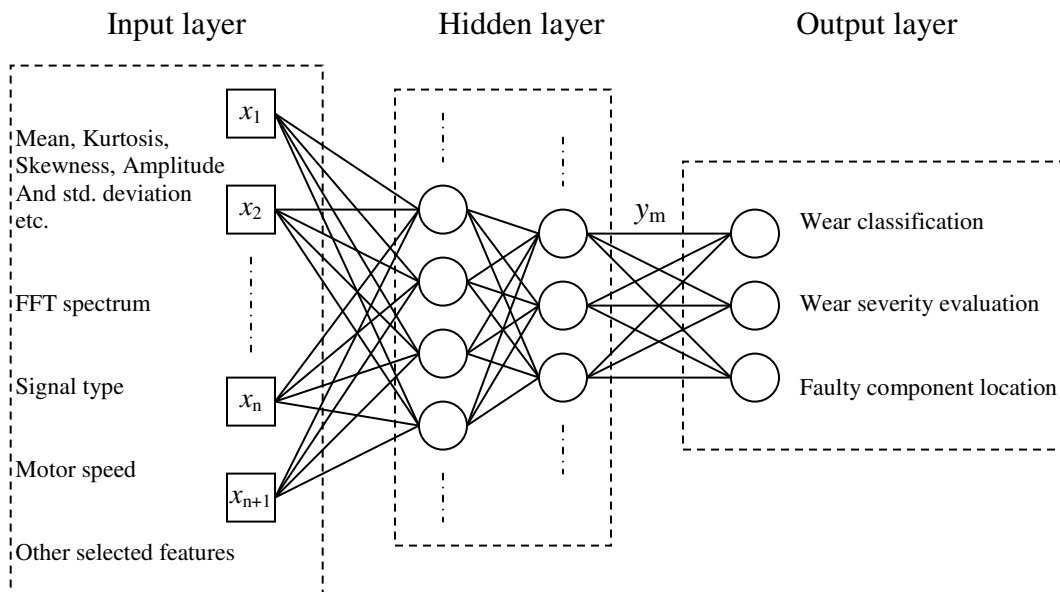


Figure 2.16 Standard supervised training feed forward neural network

There are a number of learning methods to train MLP networks but the feed forward back propagation (FFBP) paradigm has emerged as the most popular training mechanism. Many of the applications [115-119] show the effectiveness of using FFBP for fault

classification and severity evaluation of the machinery components. Figure 2.16 illustrates a typical multilayered feed forward neural network. Neurons in the input layer act as buffers for distributing the input selected feature data $x=(x_1, \dots, x_n)$ to neurons in the hidden layers. Each neuron in the hidden layer sums up its input signals after weighting them and computes its output y_m . Then, outputs of neurons in the output layer—wear classification, wear severity evaluation and faulty components location could be finally achieved. The FFBP networks, however, have a major limitation of difficulty in determining the structure of the networks and number of nodes. Therefore, most of the FFBP networks in pattern recognition to date have intended to use a particular network model without sufficient justification [23]. Wang *et al* [23,119 and 120] developed a systematic approach for training and finding the best performance models from a large number of possible structures of multi-layered FFBP neural networks. They applied this approach to pump fault classification problem, and found that the performance of an MLP network did not necessarily get better when the number of neurons in the hidden layer was increased. In fact, the accuracy of the classification rate is randomly changed with a change in the number of hidden neurons, which made achieving the best performance very difficult.

Apart from the FFBP, other neural network models are also usually applied in machinery condition monitoring, such as radial basis function [121], recurrent neural networks [122, 123] and counter propagation neural networks [124]. However, the above ANN models are all required to use supervised learning algorithms which ask for classified input such as *a priori* information and the desired output [118]. For example, the pre-defined fault data were experimentally (with seeded defects) obtained to train the supervised networks. On the other hand, unsupervised learning does not necessarily require any *a priori* knowledge, and the networks learn the information and classify the input vectors according to how they are distributed in the input space. Hoffman and Merwe [125] used self-organizing maps (SOM) which have properties of both vector quantization and vector projection algorithms to evaluate the multiple fault conditions of the rolling element bearings. In this application, the SOM proved to be an easy-to-use tool for initial

data processing and for identifying hidden relationships between features. Furthermore, the SOM obtains the ability to adaptively modify when new input data are implemented.

In practice, knowledge from specific domain expert is usually incorrect, and cannot be directly used as the input of models. Therefore, uncertainty measures of the knowledge are required for ANNs. The widely used uncertainty measures include probability, fuzzy sets and belief networks. A hybrid model of neural network and fuzzy logic (neural-fuzzy network) has been presented [75]. The neural-fuzzy system is an ANN structured upon fuzzy logic principles, which enables this system to provide qualitative description about the fault detection and wear classification process [75]. There are many possible combinations of these two systems, and Wilson Wang and co-workers [126] investigated four types of neural-fuzzy systems, and then compared their performances in gear system monitoring applications.

Although ANN models are probably the most widely used decision support system for machinery diagnostics, there are still several concerns over this popular approach. In the main, neural networks perform associations between events and faults and are excellent for pattern recognition, but little is known about how to structure neural networks into assemblies of co-operative reasoning elements and how to physically explain the network models with *a priori* knowledge [127].

(3) Evolutionary algorithms

An evolutionary algorithm (EA), which simulates a process of how a population naturally evolves, has also shown advantages in applications to machinery condition monitoring. The most frequently used form of EA is the Genetic algorithms (GAs). Zhang and co-workers [128] used the GA to detect faults in rotating machinery. Features from two different components were used to examine the performance of the two-class normal/fault recognition. They compared results with another AI-based method, ANN, and demonstrated that performance is equal to or better than the results obtained by ANN. Moreover, the training times were also found to be shorter than the ANN approach. Genetic algorithms have also been used to make the classification process faster and

more accurate using a minimum number of features which primarily characterize the system conditions with the optimized structure or parameters of ANNs [129, 130]. In recent work [131], the results of ANNs with GAs were presented for fault detection of gears using only time-domain features of vibration signals. In this approach, the features were extracted from the finite segments of two signals: one from a normal condition and the other from faulty gears, and then the GA was applied to reduce the dimension of the selected features before training with ANNs.

2.4.1.3. Statistical Pattern Recognition

Similar to the machinery fault classification with AI approaches, statistical methods have also been used in the pattern recognition way to provide automated condition monitoring techniques for the machinery condition-based maintenance. Within the engineering applications, linear discriminant function, support vector machine (SVM), and clustering analysis are the most widely used systems.

Linear discriminate analysis applies the linear classification rule $q: x \subseteq R^n \rightarrow Y = \{1, 2, \dots, c\}$ which is composed of a set of discriminant functions:

$$f_y(x) = (w_y \cdot x) + b_y, \quad \forall y \in Y, \quad (2.7)$$

Which are linear with respect to the input vector $x \in R^n$ and their parameter vector $w \in R^n$, the scalars $b_y, \forall y \in Y$, introduce the bias to the discriminant functions.

Figure 2.17 (a) shows the binary case $Y = \{1, 2\}$ and its trained linear hyperplane and two classified datasets. C. Li and S. Li [132] applied the pattern classification methodology for detecting defects on the outer race and rollers of the bearings. They established a linear discriminate function-based technique for identifying where the problem is by using two normalized and dimensionless features that are extracted from acoustic emission signals utilizing short-time signal processing techniques. Sun *et al* [133] also applied a piecewise linear discriminator to classify bearing fault conditions.

The basic *SVM* deals with the binary problem, in which the labelled data are classified by a hyperplane defined by a number of support vectors. In the case of a two-dimensional situation, as Figure 2.17 (b) shows, a series of labelled data points for two classes are supposed to be utilized to train a hyperplane that could separate the two defined classes. In mathematical terms, for any point that lies on the boundary line (solid line), it could be written as:

$$[x | (w \cdot x) + b = 0] \quad (2.8)$$

Where w is a vector that defines the boundary line, x is an input vector, and b is a scalar threshold values. At the margins (dotted line), where the support vectors (data points with a circle in Figure 2.17 (b)) are located, the equations for class A and B respectively are:

$$[x | (w \cdot x) + b = +1] \quad (2.9)$$

and

$$[x | (w \cdot x) + b = -1] \quad (2.10)$$

Jack and Nandy [134] first applied the SVM to the rolling element bearing fault detection. They compared the classification performance of both SVM and ANN, and concluded that ANNs generate well and achieve high training success, constant width SVMs achieve a high success rate on the training data, but they do not generalize well, while the averaged width SVMs generalize well, but do not achieve a high training success rate. On the basis of this work, Jack and Nandy [135] also employed a genetic algorithm (GA) for feature selection, and both classifiers discussed above achieve very high training success and generalize well using significantly smaller numbers of features. Samanta and co-workers [129] also compared ANNs and SVMs with GA-based feature selection from time-domain vibration signals. In this work, the roles of different vibration signals and signal processing techniques were investigated, and they found that the performance of SVM is substantially better than ANN with the entire feature set, and the classification with both ANNs and SVMs gave a 100% rate if the GA-based feature selection was applied. Samanta [130, 131] extended work to the gear fault detection using

ANNs and SVMs with GA-based feature selection. The conclusion is similar to the fault diagnosis to the rolling element bearings. However, the research works reviewed above are not optimized and are at a relatively early stage of comparison with ANNs. Alfonso and Nandy [136] investigated SVM at a more advanced level. In their study, the training of SVM is conducted using the Sequential Minimal Optimization (SMO) algorithm, and a mechanism for selection of adequate parameters for SVMs with Radial Basis Function (RBF) kernel is also proposed. This research work found that the combination of the training method and proposed parameters selection mechanism make the fault diagnosis of the rolling element bearings fast and effective.

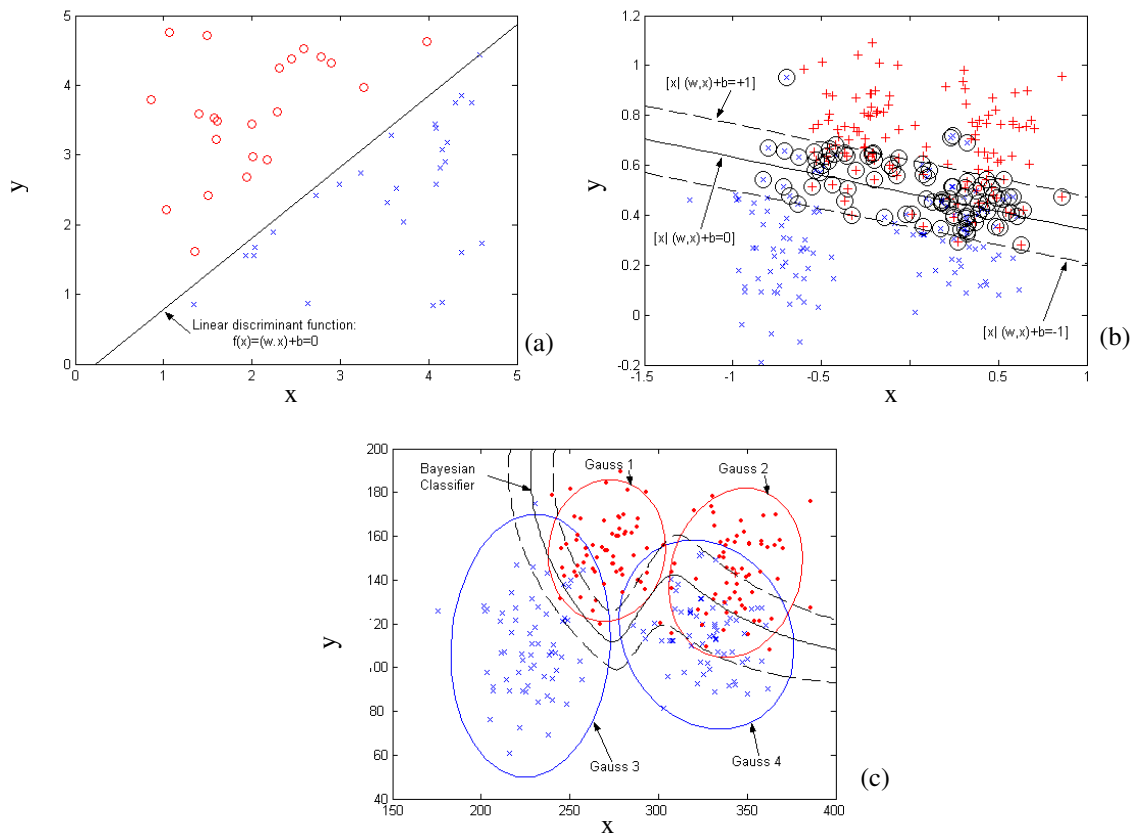


Figure 2.17 Statistical pattern recognition results: (a) linear discriminant classifier
(b) Support vector machine, (c) Bayesian classifier

As Jardine [24] indicates, cluster analysis is a multivariate method that classifies data into different groups such that cases within a group are more similar to one another than to cases in other groups. The intention of this process is to minimise difference within the

same cluster and maximise difference between different clusters. Various researches [137, 138] have been conducted to demonstrate the effectiveness of using clustering techniques for machinery fault diagnosis. Among these applications, distance measures are the main evaluation index to examine similarity between the two cases. These measures are normally developed on the basis of discriminant functions in pattern recognition theory [139]. Most widely used distance indices in machinery condition monitoring include Euclidean distance, Mahalanobis distance and Bayesian distance etc. Examples of the applications for these distance indices can be found in [140-143]. A more advanced method based on symmetric Itakura distance in time–frequency domain has been developed in [144], the method shows the advantage in term of resolution, and more insignificant anomalies can be discovered. Furthermore, similarity measures such as cross-correlation is also used for machinery fault diagnosis [143]; in this method, similarities between the features are evaluated so that the most uncommon one can be recognised as the abnormal feature. Apart from distance and similarity measures, many clustering algorithms are applied directly for determining the number and distribution of signal groups [145], and then the existed signal groups can be used as a classifier for the new detecting signals.

Based on the trained clusters and their support cases, as a clustering classifier, Bayesian classifier is usually developed to construct the hyperplane between the population classes. The general statement of Bayesian theory is:

$$P(w_i | x) = \frac{p(x | w_i)P(w_i)}{\sum_{i=1}^n P(w_i)p(x | w_i)} \quad (2.11)$$

Where $P(w_i)$ is the prior probability of class w_i , $p(x | w_i)$ is the class conditional probability density function (i.e. Gaussian, exponential) and $p(x)$ represents the unconditional probability density function of the observation. Then a decision on a fault is obtained according to the decision rule:

$$P(w_i)p(x|w_i) > P(w_j)p(x|w_j) \quad \forall i \neq j \quad (2.12)$$

Figure 2.17 (c) shows the clustering space consisted of four clusters and its developed classification hyperplane. Although at the moment, the Bayesian classification technique is seldom applied in the field of tribological components fault diagnosis, its effectiveness for classification has been proven in many other application areas, such as plant process monitoring. As the Bayesian approach has the advantage of estimating parameters progressively from sampled data [147], the classification procedural could substantially update the estimation of fault parameters whenever the detected fault is classified.

Hood and Ji [148] proposed an intelligent system using an adaptive learning machine. The system learns the normal behaviour of the detected component. Deviations from the norm are detected and the information is combined in the probabilistic framework of a Bayesian network. The proposed system is thereby able to detect unknown or unseen faults. Lin and Shin [147] also proposed a fault classification scheme based on the Bayesian theory so that fault parameters could be estimated on-line using the most up-to-date information via detection mechanism. Wang and Mcgreavy [149] compared the Bayesian approach with supervised back propagation neural networks, and reported that a major advantage of the Bayesian-based automatic classification approach is that it avoids the need to arbitrarily introduce vigilance factors, which are empirical and have to be estimated by trial and error. Padovese [150] did a similar comparison between the fault classification performance of a statistical Bayesian decision rule and a multilayer perception neural network. He found both methods could obtain excellent performance in fault classification for rolling element bearings by using acoustic noise (higher than a 95% success rate), but discrimination of the results for four bearing fault patterns are better for the neural network than for the Bayesian rule.

2.4.2. Anomaly detection approaches

As Jardine *et al* [24] indicated, one of the main problems inherent in the current supervised or unsupervised multi-class fault diagnosis approaches is that training data

from all classes are required during the training phase. Class labels are essential in order for the decision boundaries to be drawn. However, the problem in the condition monitoring environment is that often only sufficient information from one class may be available, such as ‘normal’ condition; and the number of possible fault conditions is large, and many of them may not be characterized sufficiently to build a conventional multi-class classifier. For those reasons discussed above, the concept of *anomaly detection* is developed to identify abnormal system behaviour, in which a model of normality is constructed, with deviations from the model identified as ‘abnormal’. Figure 2.18 shows the fundamental concept and consisted elements of the anomaly detection model. From the figure, it is seen that, at the beginning, the model of normality is constructed with ‘normal’ data and learning algorithms; next, one of the anomaly detection indices is calculated to indicate and quantify the state of the new arriving data; and finally the so-called ‘contribution values’ or ‘driving factors’ of the applied variables are assessed to provide diagnostic information.

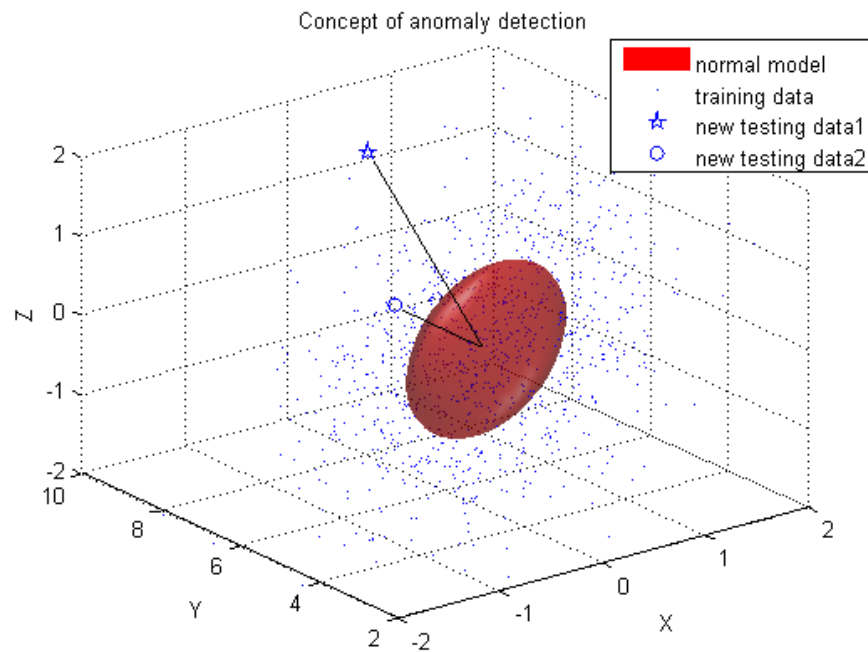


Figure 2.18 Concept of the anomaly detection

The anomaly detection problem introduced above can be regarded as a hypothesis test program with null hypothesis H_0 : Anomaly A is existed, and opposite hypothesis H_1 : Anomaly A is not existed [151]. In a conventional anomaly detection application, hypotheses described above are usually expressed by applying specific underlining distributions, or the distribution parameters, such as location and scale parameters. Anomaly detection indices are then calculated to indicate the condition of the monitored machinery so that the decision can be made to judge whether hypothesis H_0 : Anomaly A is existed or not. Based on such a principle, numerous anomaly detection approaches have been developed and applied in a variety of areas. However, these techniques have not been widely applied and tested in the field of machinery condition monitoring. In this section, the author would like to review the existing anomaly detection methods from the three aspects described above in a more generic perspective, and investigate the possibility of using these approaches in machinery condition monitoring.

2.4.2.1. Construction of a model of normality

In anomaly detection, construction of a model of normality is the first step, and probably also the most important step. Normal data representing the healthy condition of the machinery are usually utilized to be the training data, and then statistical learning approaches are applied to the ‘normal’ training data to summarize their statistical properties.

Fundamentally, it is usually assumed that normal data are generated from a specific underline distribution, which may be estimated from example data. The classical statistical approach is the use of *kernel density estimation* techniques to determine the underling data distribution [152], which may set a threshold between normal and abnormal areas of data space. These approaches are developed on the basis of the assumptions that are made about the type of the underline data distribution, and the parameters of the distribution estimated from observed data. However, these assumptions ignore the real distribution of the ‘normal’ training data, and may usually prove too aggressive, leading to a poor fit between model and observed data [153].

Another conventional approach, *Multivariate Statistical Process Control* (MSPC), which was developed on the basis of statistical process control (SPC), has been well recognised and substantially applied in performance monitoring in chemical plant monitoring [153-158]. The main idea of the MSPC is to measure how deviated of the new data from a baseline dataset representing the healthy condition of the plant and to see if the new data is under the thresholds or not. Traditionally, the normal reference model is built with the technique of principal component analysis (PCA) to find a set of combinations of variables, or major principal components, which describes major variations and trends for the operating data. And then, two statistical parameters, Hotelling's T-squared and Squared Prediction Error (SPE), are calculated as the anomaly detection indices for the current incoming data vectors to see if they fall into the accepted region. Unlike traditional SPC schemes, the extracted principal components (PCs) can reduce the number of features when process performance is monitored. However, as Chen and Liu [159] indicated, it is difficult to detect process faults when the normal data cannot represent a single set of the combination of variables only; the system is more likely to be distributed at several regions.

The approaches discussed above have raised a common problem which is that 'normal' data may not be following a specific distribution and are distributed in a more complicated way. Therefore, both *non-parametric* and *parametric clustering* approaches to summarize or describe 'normal' data more precisely have been developed and demonstrated to be superior to the conventional methods. It should be noted that clustering algorithms are used extensively, not only for data organization and categorization but also for data compression and model construction.

Non-parametric clustering approaches do not make any assumption about the statistical properties of the data [160]; in other words, no specific distribution models will be applied to describe the evaluated data. One of the most well-known examples of this method is the *Parzen window* which involves building a model of normality by estimating the unconditional probability density function $p(x)$ of the 'normal' training data. The density function of this non-parametric kernel-based method is represented by

identical sphere kernel functions around each training pattern in the data which are then summed up for each of the k classes [161]. For most of the applications [162-164], Gaussian kernels are selected for Parzen window methods due to their smooth and convenient statistical properties. However, there are several problems that exist in this approach, including the fact that the size of the assigned window (kernel) function is the same and this property may lead to the capture of local characteristics of the training data distribution difficulty [165]. One simple approach to solve this problem is to assume that the fixed locations of the centres may be chosen randomly from the input data sets. This approach is considered reasonable and provides data distribution in a representative manner [166]. Instead of fixing the centres of the data distribution, other clustering approaches utilize the elliptical basis functions which move the location of their centres in a self-organized learning rule, such as *k-means* clustering and *fuzzy* clustering algorithms. The K-means clustering algorithm [167] is an iterative method of producing a set of prototype patterns μ_j (for $j=1\dots k$), and assigns each case to the specific clusters with certainty. Clifton *et al* [168, 169] applied the k-means clustering algorithm to construct a model of normality, and calculated the Euclidean distance as the novelty score for each case to detect abnormal behaviour of the jet engines. The use of the fuzzy clustering algorithm avoids getting stuck with the local minimum of the objective function in which the classical k-means method suffers [153]. Fuzzy clustering analysis produces in each training data a degree of membership to each cluster. Moreover, fuzzy clustering methods are usually utilized to combine with other statistical approaches to achieve further functions. Frigui and co-workers [170] developed an unsupervised robust c-prototypes algorithm which integrates fuzzy c-means clustering and statistical estimation and the algorithm can remove and merge compatible clusters.

Parametric clustering approaches make an assumption that evaluated data are following specific statistical distributions such as exponential and Gaussian distributions. In practice, Gaussian is the most widely used distribution, and the data can be modelled with means and covariance. Fundamentally, many acquired normal signals are Gaussian like – in fact some signals (e.g. some engine data and transmission vibration signals) give good Gaussians. However, sometimes the data is not truly Gaussian due to instrument

problems or unknown faults, but a mixture of Gaussians can give a good representation of complex distributions. Therefore, a *Gaussian mixture model* (GMM) approach is widely applied to characterize the ‘normal’ training data which can be distributed in the non-Gaussian way. If a multivariate dataset x is modelled by a GMM, its probability density function can be formulated as:

$$p(x; \Theta) = \sum_{m=1}^k w_m p(x; \theta_m) \quad (2.13)$$

Where w_1, w_2, \dots, w_k are the weights of each Gaussian component subject to $w_m > 0$ and $\sum_{m=1}^k w_m = 1$. Each component's density $p(x, \theta_m)$ is a Gaussian probability distribution. And $\theta_m = \{\mu_m, C_m\}$ represents the m^{th} component mean vector and covariance matrix respectively. The complete set of parameters to define the mixture of Gaussians is $\Theta = \{w_1, \dots, w_k, \mu_1, \dots, \mu_k, C_1, \dots, C_k\}$.

To estimate the parameters of the GMM, the log-likelihood of the training data is consistently maximized by using iterative optimization algorithms such as *Expectation-Maximization* (EM) method [171]. The combination of GMM and EM algorithm is a powerful way to describe data and extract their statistical properties. In the last two decades, various applications [172-174] have used this algorithm as a standard approach for anomaly detection. However, like other clustering methods, this parametric clustering approach also suffers a number of problems which have drawn a lot of research interest.

The first problem is that the number of clusters has to be predetermined. An indication of the number of clusters present in the data can be gained by running the EM algorithm a number of times and on each occasion requesting a different number of clusters. The algorithm can return a number of measures that indicate how well the model fits the data. Such measures can be the Akaike's information criterion (AIC) [175] and the Bayesian information criterion (BIC) [176]. McLachlan and Peel [177] have compared the performance of the AIC and BIC, and found that the penalty term of the BIC is more significant than that of the AIC. Moreover, Fraley and Raftery [178] give a systematic

method to select an appropriate number of clusters, incorporating with the hierarchical agglomeration for EM initiation.

The second problem is that EM algorithm involves some randomness and has many local maxima that cause inconsistent solutions from different initial conditions and directly affect anomaly detection. To overcome this random characteristic, Ueda and Nakano [179] proposed the deterministic annealing expectation maximization algorithm (DAEM) by minimizing the thermodynamic free energy instead of maximizing the likelihood function. They introduced a ‘temperature’ parameter and information entropy term in the cost function to smooth out the local maxima in the maximum likelihood function. This ‘temperature’ gradually reduces to achieve a global minimum. Biernacki *et al* [180] chose the best set of initial conditions to obtain maximum likelihood parameters by running a few steps from multiple initial conditions to avoid being trapped in suboptimal solutions. Ding *et al* [181] proposed an adaptive dimension reduction EM algorithm. They projected the dataset to the lower dimension through PCA or any dimension reduction method and ran EM on the subspace. The solutions of the EM were expanded back to original variable space and the possible centres for each cluster were computed.

Apart from the problems of the GMM itself and its associated EM algorithm, two major problems are usually encountered during the applications of models: firstly, the true condition of the ‘normal’ training data may be unknown and data might contain a number of anomalies due to the sensor malfunction and environmental effect, so that the model cannot be expected to respond to the similar anomalies that appear in the new arriving data [182]; secondly, established GMM with the historical training data may have nothing to do with the new arriving data, which leads to incomparable phenomena between the established model and new observed data.

Lauer [183] proposes an approach that solves the problem when the training data contains an unknown number of anomalies. To conduct the algorithm, pre-defined validation dataset or a general estimate of the proportion of anomalies in the training data are necessary. The method started by assuming the data distribution consists of a large

number of normalities and a small number of abnormalities. He defined λ as the percentage of abnormal data, P_N as the distribution of normal data and P_O as the distribution of abnormal data, and then the distribution of the completed training data can be given as: $P(x)=(1-\lambda)P_N(x)+\lambda P_O(x)$. The whole training process can then be summarised by three steps: first of all, estimate $P(x)$ of the evaluated training data by assigning the GMM; and then, divide $P(x)$ into $P_N(x)$ and $P_O(x)$; and finally, determine whether x is an abnormal data by defining the probabilities $P_N(x)$, $P_O(x)$ and the percentage of abnormal data λ . He pointed out that the second step is very difficult to achieve since it usually lacks of information on distribution of abnormal data and their quantities. Therefore, the second step cannot be achieved directly, and alternatively, estimation on the distribution $P_O(x)$ is conducted and the percentage λ and $P_N(x)$ is calculated according to the value of $P_O(x)$. The GMM and its associated EM algorithm are adopted in this approach to estimate the GMM parameters, as described in the first step. After the assignment of the model, the approach can be validated with different values of λ which is used for defining percentage of anomalies. In this method, λ is the key parameter which justifies the relationship between the model or whole training data and abnormal data; and it has been indicated that it usually takes long time to obtain the optimised value, and an automatic searching program used to find satisfactory λ has also been proposed by the author. The approach was evaluated with a simulated datasets of various situations and a real database of medical data with patients, and result showed that the developed algorithm achieved better performance than the traditional approaches. Besides this important method, some other applications also call this anomaly locating problem automated model-free anomaly detection [184], which means the training data is dealing with themselves and detecting faults without any other reference models. Agogino and Tumer [184] incorporated information theory and clustering algorithms to detect faults in the Space Shuttle Main Engine system without too much historical knowledge. Yang and Ma [185] developed a new anomaly detection measurement based on the probability characteristics of intrusion instances and normal instances within the training data, and applied this technique for ‘intrusion detection’ on computer networks.

On the other hand, to overcome the relationship breakdown between the established models and testing data, various methods [186-188] have been developed to enable models to be incrementally updated and learnt over time. Hall and Hicks [186] proposed a method to consistently add Gaussian models based on original GMM space. The new resulting GMM is generally computed in two steps:

- a) Concatenate – suppose a GMM has already been learned up to time t :

$$p^t(x) = \frac{\sum_{m=1}^k w_m^t p(x; \mu_m^t, C_m^t)}{\sum_{m=1}^k w_m^t} \quad (2.14)$$

Where each Gaussian component is represented by its weight w_m^t , mean μ_m^t and covariance C_m^t . Then, if a set of new data is received, the updated model with $k+1$ component could be obtained by trivially combining the original GMM and the new data into a single model:

$$p^{t+1}(x) = \frac{\sum_{m=1}^k w_m^t p(x; \mu_m^t, C_m^t) + w^{t+1} p(x; \mu^{t+1}, C^{t+1})}{\sum_{m=1}^k w_m^t + w^{t+1}} \quad (2.15)$$

- b) Simplify – if possible, merge some of the Gaussians to reduce the complexity of the model while still giving a precise description of the observations. Hall and Hicks [186] proposed to group the Gaussians using the Chernoff bound to detect overlapping Gaussians. Different thresholds on this bound are then tested and the most likely result is kept as the simplified GMM.

However, as Declercq and Piater [187] indicated, this method is too slow for an on-line process, and they used a different criterion to develop an uncertain Gaussian model which provides a quantitative estimate λ of its ability to describe the associated data. With this value, the two given Gaussians could be merged if necessary without drifting from the real data distribution. Furthermore, Arandjelovic and Cipolla [188] addressed the problem of learning the GMM incrementally, which is that previous approaches

universally assumed that new data come in blocks representable by GMMs which are then merged with the current model. They proposed a method of working for the case when new data points arrive one by one, which requires little additional memory.

The approaches discussed above are all statistically based, and besides that, AI methods have also been developed to construct the model of normality. The most commonly used ones include support vector machines (SVMs) which have been reviewed by section 2.4.1.2 as well as self-organizing map (SOM). The SOM, which was initially proposed for clustering and visualization of high dimensional data, provides an unsupervised summary of training data using neural network. Various applications of the SOM to the anomaly detection have been described in [189, 190], while they have also been extended into generic density estimators for statistical anomaly detection [191].

2.4.2.2. Anomaly detection indices

The above described method obtains the advantages of fusing multi-dimensional data, clearly indicating abnormal events and aiding operators to make a decision. However, it is indicated that the traditional approach described above to calculate the control limit is based on the assumption that the normal data follows the single Gaussian distribution, but this assumption is not necessarily satisfied in the more complex processes [192]. The normal data might be distributed in a more complicated way, or it usually consists of a number of examples of ‘normal’ data with a smaller, but unknown, number of ‘abnormal’ data. If the traditional approaches are applied to these two cases, the model runs the risk of fitting non-Gaussian data or ‘abnormal’ data and hence makes their detection more difficult. This might lead to the risk that out-of-control cases are missed or in-control situations are falsely seen as out-of-control. Under this circumstance, Chen and co-workers [192] applied a Gaussian Mixture Model (GMM) to describe the more complex normal training data, and the statistic of log-likelihood to examine how good the established GMM model fitting the current incoming vector was calculated as the fault detection index. Results show that more faulty events could be detected by the GMM-based method than the conventional SPC approach. Zhang [193] also conducted GMM

for the reference model setup. The only difference is that fault detection indices of Hotelling's T-squared and SPE for the current vector are calculated based on each of the clustered components within the normal data. The result also shows the advantages over the traditional approaches.

From the above examples, it is found that GMM is an effective way of overcoming the shortcomings of the traditional SPC method. It has its advantages which can be selected with a combination of the other statistics for the development of novelty detection. Roberts [194, 195] proposed Extreme Value Theory (EVT) for novelty detection that concerns abnormally low or high values in the tails of data distributions. His research is concerned with samples drawn from a distribution whose maxima distribution converges to the Gumbel form. This distribution gives a probability of observing some extreme value. The GMM is used to estimating the distribution of the normal data; hence the EVT probability can be used directly as the EVT distribution is derived from the same range of data that the GMM is fitted on. However, only the component closest to the data point with the minimum Mahalanobis distance is used to calculate the EVT probability as this dominates the EVT statistics.

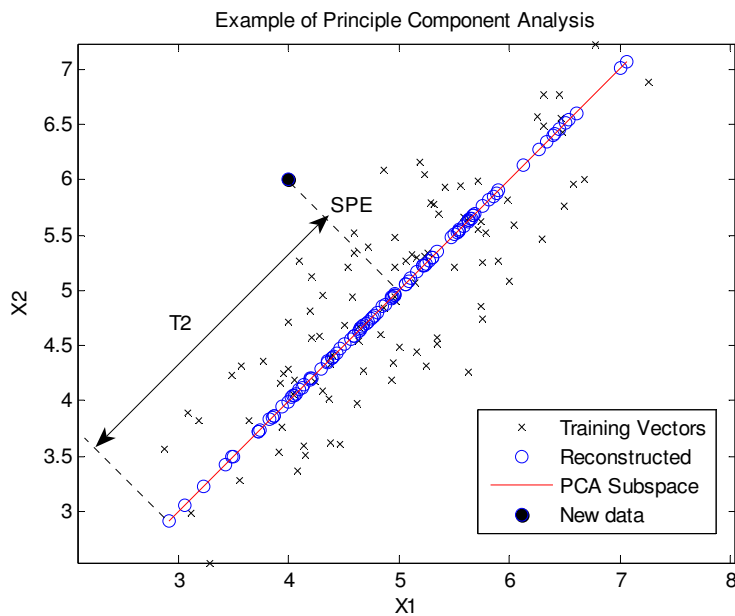


Figure 2.19 Example of the fault detection based on PCA subspace

Novelty detection has always been a hot topic for investigation. However, machinery fault diagnostics is more focused on fault classification rather than detecting the occurrence of the faults. This can significantly reduce the sensitivity and accuracy of the fault diagnosis methods. Therefore, the above anomaly detection approaches have the greatest potential to be considered in the applications of machinery fault detection.

2.4.2.3. Fault diagnosis

It has been shown that fault detection is insufficient for providing adequate decision support. Fault diagnosis is always needed to identify or isolate abnormalities so that the appropriate maintenance plan can be carried out. In the SPC, an unsupervised method based on calculating the weightings of each of the applied variables to the trend of the fault detection indices such as T-squared and SPE has been developed, and variables with higher contribution values to the detected abnormal events can be identified, so that the location and causes of the faults could be determined. This is a complete data-driven approach; no pre-defined knowledge is implemented for the fault diagnosis. Multiple variables are modulated by the PCA model, and then the contribution values of the variables are demodulated in the same scale. This approach has been widely applied in chemical plants to diagnose faults [196-198]. Although faults are not diagnosed automatically with this approach, it could help operators to gain an insight to the abnormal process, so as to make correct judgments.

On the other hand, Callan *et al* [182] developed a so-called ‘influence factor’ trace for each condition indicator in the transmission vibration monitoring system. This ‘influence factor’ trace could be used to locate which condition indicator is driving the detected anomalies. The combination of the anomaly detection index (log-likelihood in this case) and ‘influence factor’ is very valuable, since they could provide diagnostic information.

Other diagnostic approaches also include a diagnostic GMM per-sensor method, in which independent GMMs are trained on individual sensors. Martin [199] applied such a

method and calculated the independent log-likelihood score for each sensor. He reported that useful diagnostic capability was achieved, but required more calculating expenditure.

2.4.3. Feature extraction and data filtering

No matter what scheme the practitioners apply, multi-fault classification or anomaly detection, an appropriate data processing technique and extracted features from the raw data are essential for the model training. This is because these techniques could significantly improve the classification/detection rate and reduce the training time.

Generally, features which are used as the inputs to models are extracted from raw data through various data processing techniques in time, frequency and time-frequency domains. The development of these techniques associated with different sensors has been substantially introduced in section 2.3. But they could be playing different roles in model-based fault classification/detection. Samanta and Al-Balushi [200] investigated the effects of different input signals and features on the success rate of training and testing of neural networks. The features were extracted through signal processing techniques, such as high-pass or band-pass filtration of the vibration signals, with and without demodulation. Furthermore, they also used more elaborate signal processing techniques such as wavelet transform, and found that it does not affect the results significantly. They also found that a substantial reduction in training time was achieved due to processing and the number of extracted features. McCormick [201] examined several feature extraction methods including higher order statistics, velocity of vibrations etc. and their effects on the classification rate. With all the combinations of features, a successful classification rate of over 99% was achieved. Murray *et al* [202] investigated feature extraction information from the wavelet transform, and compared it with the power spectrum of the frequencies. The test results showed that slight improvements were obtained through the use of wavelet transforms. Lei *et al* [203] applied a compensation distance evaluation technique (CDET) to select sensitive features of time and frequency domains before using them as the input of the fuzzy c-means clustering. The results

indicated that the feature selection process is critical to fault classification, and this helped to increase the fault classification rate.

Apart from the feature evaluation study on the model training and testing as discussed above, a very important issue about model training is that learning methods usually suffer from the curse of dimensionality [204] in the sense that if the dimensionality of the data is too high, a very large number of samples are needed to train the model. This requires reducing the dimensionality of the training data. To achieve this, a genetic algorithm (GA) has been applied to select appropriate features for model training. Section 2.4.1.2 has reviewed relevant literatures and showed that this technique could significantly improve the performance of fault diagnosis. On the other hand, PCA and independent component analysis (ICA) of the multivariate analysis are also found to be very valuable in decreasing the dimensionality of the data with a complicated internal structure. In order to better detect lubrication abnormal condition in helicopter gearboxes, Stellman *et al* [205] applied PCA to morphology data for reducing the dimensionality. Allgood and Upadhyaya [206] performed PCA on certain statistics in DC motor diagnosis and prognosis. Sun *et al* [207] incorporated PCA with a C4.5 decision tree to diagnose the faults of the rotating machinery. These approaches have all achieved a better classification rate compared with those without dimensionality reduction techniques. However, as Bishop [204] indicated, the label of the applied features would be lost and replaced by the score vectors (principal components) after the calculation of PCA/ICA. This characteristic may not be serious, if the fault classification rate is the only evaluation index, but it would affect the usage of the approaches which provide diagnostic information through examining the contribution/influence factors of the applied variables, as these variables are related to fault mechanisms and locations.

2.4.4. Comparison and combination of the AI and statistical modelling techniques for machinery condition monitoring

In this chapter, a series of AI and statistical techniques employed in machinery fault classification or anomaly detection have been reviewed. As the descriptions of the

various techniques progressed from items in section 2.3, it should have been noticed that there was a boundary in the development strategy that each system adopted. In fact, this boundary could be drawn based on the type of training data required to construct the model. The first category is the multi-class fault diagnostic systems which require multi-class training data representing different symptoms of the machinery. Based on this strategy, a large amount of automated classification systems have been developed, and are being improved through modification of the algorithms and implementation of the data pre-processing techniques. As a branch of the multi-class fault diagnostic system, expert systems are basically knowledge-based systems, and researchers usually spend more time collecting fault cases than developing the systems themselves. Knowledge-based systems are easy to design, but lack the flexibility property when fresh situations occur. The rest of the multi-class fault diagnostic systems described in this chapter could all be classified as model-based systems. These types of system require more understanding because once the system is learned, the algorithms exploit the information and produce a result as they are designed. It may also be found that there is a need for a discussion of techniques where a reduced level of knowledge is required. Fuzzy logic system needs a knowledge base to support the membership function, while the ANN or GA systems only need knowledge when they are conducting a training process. ANN has been recognized as the most popular and frequently used tool for automated machinery diagnosis and reasoning, because it contains significant advantages. For example, the ANN models are relatively easy to build, and do not require too much knowledge about the models themselves. This benefit especially attracts the attention of engineers, and various engineering applications have been achieved. However, ANN still suffers from several problems. Firstly, at the time of writing, it is not easy to encode these networks with *a priori* knowledge, which means the physical meaning of the models could not be explained. On the other hand, it is not known how the training data is partitioned and their statistical properties could not be extracted; the only result that could be accessed is the classification rate. Similar to the AI classification systems, statistical based approaches have also been adopted for the development. Clustering and support vector machines (SVMs) are the most popular ones, but with these approaches, it can be clearly seen how the training data are partitioned, and these provide physical explanations of the

grouped areas and data. It has been proven by many applications that the statistical based fault classification systems obtain good classification capability, but probably need more system parameters of evaluation and justification.

The second development strategy for machinery condition monitoring is anomaly detection. As an important concept, anomaly detection plays an important role in identifying abnormal events against the models of normality which are built with a single class of 'normal' training data. To construct the normal model, the kernel density estimation method was originally developed with the assumption that the training data follow specific underline distribution. Statistical process control (SPC) was then developed and widely applied in the chemical plants. However, these two methods normally ignored the true condition of training data which may be distributed in the more complex way, and this led to a poor fit between the models and new observed data. Therefore, more sophisticated algorithms such as clustering techniques are investigated to construct anomaly detection models. Generally, the adopted clustering algorithms could be divided into non-parametric and parametric approaches, in which non-parametric approaches such as k-means and fuzzy c-means clustering do not make any assumptions on the distribution of the training data, whereas parametric approaches assume that the training data could be described by a set of known mixture models such as GMM. The discussion about which type of clustering is superior is still on-going, but one thing that is already recognized and well understood is that if the form of the distribution is restricted, a GMM can be made to operate like k-means [208]. This important discovery unveils the relationship between the non-parametric and parametric clustering. Furthermore, Fraley and Raftery [178] also indicated that one advantage of the parametric approach is that it allows the use of approximate Bayes factors to compare the model, and this gives a systematic means of selecting not only the parameterization of the model, but also the number of clusters. As mentioned earlier, the other two important issues for model construction are adaptation and updating which allow models to be more robust over time. It has also been proven that it is more practical to achieve these two requirements with parametric models. After the construction of the model of normality, the anomaly detection index needs to be calculated for each new observed data. Generally, anomaly

detection index could be classified into distance based and probability based. On the other hand, the calculation of the contribution/influence values of each variable proves to be very valuable to aid operators to locate faults and trace causes, although it is not an automatic process. Collected information from the anomaly detection could be used as the inputs to be fed into diagnostic or reasoning inference engines.

Feature selection or data pre-processing techniques are usually conducted before the model construction stage to extract more information from the datasets, as well as overcoming curse of dimensionality problem. Numerical approaches have been applied to form series of feature sets in time, frequency and time-frequency domains, but other AI (e.g. genetic algorithm) or statistical (e.g. PCA and ICA) approaches are utilised to decrease dimensionality and extract variance of the multivariate data. Developing effective pre-processing strategies can be difficult but it is a crucial stage of the process.

Table 2.2 briefly summarizes the most commonly used AI and statistical based approaches for the development of automated condition monitoring systems. It should be noted that each of the described methods has its unique benefits and drawbacks, and has played a role in developing diagnostic systems. The principle of selection of these approaches should be based on the tasks as well as available *a priori* information.

Category	Sub-Category	Application status	Advantages	Disadvantages
Multi-class fault classification	Expert System	Widely used for automatic wear debris classification	does not require too much computation	Confused with novel events
	Fuzzy logic & ANN	Most frequently used methods for machinery automated fault diagnosis	Easy to build the model does not require too much knowledge about model itself	difficult to physically explain the built model Needs multi-class training data
	Genetic algorithm (GA)	Newly developed approach for machinery fault diagnosis Mainly used for feature selection	Increase the fault classification rate.	Time consuming. Needs multi-class training
	Statistical Pattern recognition	Clustering and SVM algorithm are widely used, but seldom used for machinery fault classification	Obtain the good classification Capability, data partitions are Clearly seen.	Needs amount of work for model justification. Multi-class training
Anomaly detection	Kernel Density estimation	Originally developed for anomaly detection, less been used now	Easy to build, if the status of training data is known	Ignore true condition of training data leading to poor fit.
	Statistical Process control (SPC)	Widely utilized in chemical plant monitoring. Started been applied for machinery anomaly detection.	Extract variance of the training data, provides more precise description than kernel density estimation.	Faults could only be detected, if the features Are represented by single set of combination
	Non-parametric Clustering (i.e. k-means)	Newly developed approach A few applications on machinery fault diagnosis	Models are built according to the distribution of training data. Decreasing the level of assumption.	Not easy to extract statistical properties of the training data. Not easy to update models.
	Parametric Clustering (i.e. GMM)	Newly developed approach A few applications on machinery fault diagnosis	Assign cluster membership to the data. Easy to extract statistical properties of the data. Established methods for adapting and updating models. Compare models with varied parameters.	Requires amount of work for model justification. Suffers random initial points problem.
	AI methods	Mostly used method is SOM	Easy to be built.	unknown

Table 2.2 Comparison of the analyzed AI/statistical methods

2.5 Reference List

- [1] J.E. Shigley, C.R. Mischke, R.G. Budynas, *Mechanical Engineering Design*, McGraw Hill, 2004.
- [2] T.A. Harris, *Rolling Bearing Analysis*, New York: Wiley, 1984.
- [3] J.D. Summers-Smith, *An Introduction Guide to Industrial Tribology*, Wiley, 1994.
- [4] T.S. Nisbet, G.W. Mullet, *Rolling Bearings in Service*, Huchinson Benham, London, 1978.
- [5] Neale Consultancy Engineers Ltd,
http://www.tribology.co.uk/services/investigate/r_bearing.htm.
- [6] M.R. Hoeprich, Rolling Element Bearing Fatigue Damage Propagation. *Trans.ASME, J.Tribol.* 114 (1992) 328-333.
- [7] A. Krupp, *Crack Propagation in Metals and Alloys*, Wiley-VCH Verlag GmbH & Co., 2007.
- [8] D. Nelias, F. Ville, Detrimental effects of debris dents on rolling contact fatigue. *Tribology* 122 (2000) 55-64.
- [9] X. Ai, Effect of debris contamination on the fatigue life of roller bearings. *Proc IMechE part J* 215 (2001) 563-575.
- [10] M.C. Smith, The formation of spherical debris in mode 2 fatigue cracks. *Wear* 76 (1982) 105-128.
- [11] P.J.L. Fernandes, Contact fatigue in rolling element bearings. *Engineering Failure Analysis* 4 (1997) 155-160.
- [12] A.V. Olver, The mechanism of rolling contact fatigue: An update. *Proc IMechE part J* 219 (2005) 313-330.
- [13] I. Kleis, U. Muiste, U. Pilvre, H. Uuemois, The physical mechanism of the formation of metal microspheres in the wear process. *Wear* 53 (1979) 79-85.
- [14] I.BRENCO, Spalling and Its Effect on Bearing Performance,
http://www.amstedrail.com/tech_sheets/9401.asp, 1995.
- [15] T.H. Kim, A.V. Olver, P.K. Pearson, Fatigue and fracture mechanisms in large rolling element bearings. *Tribol.Trans.* 44 (2001) 583-590.
- [16] S.A. McInerny, Y. Dai, Basic Vibration Signal Processing for Bearing Fault Detection. *IEEE Transactions on Education* 46 (2003) 149-156.
- [17] A.D. Carter, *Mechanical Reliability*, Macmillan Education, 1986.

- [18] T. Akagaki, K. Kato, Plastic flow process of surface layers in flow wear under boundary lubricated conditions. *Wear* 117 (1987) 179-196.
- [19] J. Tranter, The fundamental of, and the application of computers to, condition monitoring and predictive maintenance, *Proceeding of COMADEM*, Birmingham, UK, 1989, pp. 372-377.
- [20] A. Davies, The intelligent machine, *Manufacturing Engineer*, March 1995.
- [21] I. Watson, The economics of availability, *Manufacturing Engineer*, February 1991.
- [22] Bruel&Kjaer, *Machine Health Monitoring*, Application note, Bruel&Kjaer Ltd., Denmark, 1985.
- [23] L. Wang, PhD Thesis: A Study of Methodologies for Pump Fault Diagnosis in the Waste Water Industry, 2002.
- [24] A.K.S. Jardine, A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20 (2006) 1483-1510.
- [25] A. Dubey, Condition based maintenance-the concept in totality. *Mechanical Engineering Bulletin* 17 (1986) 114-123.
- [26] Z. Peng, N. Kessissoglou, An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis. *Wear* 255 (2003) 1221-1232.
- [27] S. Raadnui, Wear particle analysis – utilization of quantitative computer image analysis: A review. *Tribology International* 38 (2005) 871-878.
- [28] B.J. Roylance, T.M. Hunt, *The wear debris analysis handbook*, Coxmoor Publishing Company, 1999.
- [29] B.J. Roylance, S. Raadnui, The morphological attributes of wear particles – their role in identifying wear mechanisms. *Wear* 175 (1994) 115-121.
- [30] B.J. Roylance, I.A. Albidewi, A.L. Price, A.R. Luxmoore, The development of a computer-aided systematic particle analysis procedure - CASPA. *Lubr.Eng.* 48 (1992) 940-946.
- [31] D. Scott, Predictive maintenance by ferrography. *Wear* 44 (1977) 173-182.
- [32] M.V. Hofman, J.H. Hohson, The development of ferrography as a laboratory wear measurement method for the study of engine operating conditions on diesel engine wear. *Wear* 44 (1977) 183-199.

- [33] P.W. Centers, Laboratory evaluation of the on-line ferrograph. *Wear* 90 (1983) 1-9.
- [34] D. Scott, The application of ferrography to the condition monitoring of gas turbines. *Wear* 90 (1983) 21-29.
- [35] H. Odahara, Y. Sasaki, M. Kawabata, T. Taguchi, Practical Application of Lubricant Oil Analysis for Electric Power Plant. In: World tribology conference, vol. 2, Vienna, 2002.
- [36] J. Reintjes, J.E. Tucker, S.E. Thomas, A. Schultz, L.L. Tankersley, C. Lu, P.L. Howard, T. Sebok, C. Holloway, LaserNet Fines wear debris analysis technology: Application to mechanical fault detection. In: American Institute of Physics, conference proceedings, vol. 657, 2003. p. 1590–7.
- [37] J. Sun, R.J.K. Wood, L. Wang, I. Care, H.E.G. Powrie, Wear monitoring of bearing steel using electrostatic and acoustic emission techniques. *Wear* 259 (2005) 1482-1489.
- [38] P.J. Dempsey, D.G. Lewicki, H.J. Decker, Investigation of Gear and Bearing Fatigue Damage using Debris Particle Distributions, NASA/TM-2004-212883, May 2004.
- [39] P.J. Dempsey, G. Kreider, T. Fitcher, Investigation of Tapered Roller Bearing Damage Detection Using Oil Debris Analysis, *IEEE Aerospace Conference*, Montana, USA. 2005
- [40] S. Morris, R.J.K. Wood, T.J. Harvey, H.E.G. Powrie, Use of electrostatic charge monitoring for early detection of adhesive wear in oil lubricated contacts. *Trans.ASME, J.Tribol.* 124 (2002) 288-296.
- [41] E.D. Palko, Applying portable vibration monitoring instruments. *Plant Engineering* 48 (1994) 62-65.
- [42] A. Choudhury, N. Tandon, A theoretical model to predict vibration response of rolling bearings to distributed defects. *Trans.ASME, J.Vibr.Acoust.* 120 (1998) 214-220.
- [43] T.E. Tallian, O.G. Gustafsson, Progress in rolling bearing vibration research and control. *ASLE Trans* 8 (1965) 195-207.
- [44] F.P. Wardle, S.Y. Poon, Rolling bearing noise-cause and cure. *Chartered Mech.Engr.* July/August (1983) 36-40.
- [45] C.S. Sunnersjo, Rolling bearing vibrations-geometrical imperfections and wear. *J Sound Vibr* 98 (2008) 455-474.

- [46] M.W. Washo, A quick method of determining root causes and corrective actions of failed ball bearings. *Lubric.Eng.* 52 (1996) 206-213.
- [47] T. Williams, X. Ribadeneira, S. Billington, T.Kurfess, Rolling element bearing diagnostics in run-to-failure lifetime testing. *Mechanical Systems and Signal Processing* 15 (2001) 979-993.
- [48] A.F.Stronach, C.J.Cudworth, A.J.Johnson, Condition monitoring of rolling element bearings, *International condition monitoring conference*, Swansea,UK., 2000, pp. 162-177.
- [49] N. Tandon, A comparison of some vibration parameters for the condition monitoring of rolling element bearings. *Measurement* 73 (1994) 285-289.
- [50] N. Tandon, BC. Nakra, Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. *Tribology International* 25 (1992) 205-212.
- [51] H. Amick, S.K. Bui, A review of several methods for processing vibration data. *Proceeding of SPIE* 1619 (1991) 253-264.
- [52] E. Jantunen, A summary of methods applied to tool condition monitoring in drilling. *International of Journal of Machine Tool & Manufacture* 42 (2002) 997-1010.
- [53] D. Dyer, R.M. Stewart, Detection of rolling element bearing damage by statistical vibration analysis. *Trans ASME, J Mech.Design* 100 (1978) 229-235.
- [54] J. Mathew, R.J. Alfredson, Condition monitoring of rolling element bearings using vibration analysis. *Trans.ASME, J.Vibr.Acoust.* 106 (1984) 447-453.
- [55] T. Igarashi, J. Kato, Studies on the vibration and sound of defective rolling bearings. *Bull JSME Transaction* 28 (1985) 492-499.
- [56] J.L. Bleu, M. Xu, Vibration monitoring of seal-less pumps using spike energy. *J Sound Vibr* 29 (1995) 10-16.
- [57] R.J. Alfredson, J. Mathew, Time domain methods for monitoring the condition of rolling element bearings. *Mech Eng Trans,IE Australia* ME10 (1985) 102-107.
- [58] D.E. Butler, The shock pulse method for the detection of damaged rolling bearings. *NDT Int.* (1973) 92-95.
- [59] N. Tandon, B.C. Nakra, Vibration and acoustic monitoring techniques for the detection of the defects in rolling element bearings-a review. *Shock Vibr Digest* 24 (1992) 3-11.

- [60] H. Prasad, The effect of cage and roller slip on the measured defect frequency response of rolling element bearings. *ASLE Trans* 30 (1987) 199-204.
- [61] F.K. Choy, J.Zhou, M.J.Braun, L.Wang, Vibration monitoring and damage quantification of faulty ball bearings. *Trans.ASME, J.Tribol.* 127 (2005) 776-783.
- [62] J.I. Taylor, Identification of bearing defects by spectral analysis. *Trans ASME, J Mech.Design* 102 (1980) 199-204.
- [63] A.B. Johnson, A.F. Stronach, Bearing fault detection in a hostile environment, *International Conference on Condition Monitoring*, Brighton, UK, 21-23 May, 1986. pp. 35-44.
- [64] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International* 32 (1999) 469-480.
- [65] Bruel&Kjaer, *Envelop Detector*, Bruel&Kjaer system development, WB1048.
- [66] P.D. McFadden, J.D. Smith, Vibration monitoring of rolling element bearings by the high frequency resonance technique-a review. *Tribology International* 17 (1984) 3-10.
- [67] P.D. McFadden, J.D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing. *J Sound Vibr* 96 (1984) 69-82.
- [68] P.D. McFadden, J.D. Smith, The vibration produced by multiple point defect in a rolling element bearing. *J Sound Vibr* 98 (1985) 69-82.
- [69] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing* 18 (2004) 199-221.
- [70] Z.Lin, An introduction to time-frequency signal analysis. *Senior Review* 17 (2008) 46-53.
- [71] W.J. Wang, P.D. McFadden, Early detection of gear failure by vibration analysis I. Calculation of the time-frequency distribution. *Mechanical Systems and Signal Processing* 7 (1993) 193-203.
- [72] F.A. Andrade, I.East, M.N.M. Badi, Gearbox fault detection using statistical methods, time-frequency methods and harmonic wavelet-a comparative study, Chipping Norton, pp. 77-85.
- [73] P.W. Tse, Y.H. Peng, R. Yam, Wavelet Analysis and Envelope Detection for Rolling Element Bearing Fault Diagnosis-Their Effectiveness and Flexibilities. *Trans.ASME, J.Vibr.Acoust.* 123 (2001) 303-310.

- [74] B. Liu, S.F. Ling, Q.F. Meng, Machinery diagnostics based on wavelet packets. *Journal of vibration and control* 3 (1997) 5-17.
- [75] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnostics. *Mechanical Systems and Signal Processing* 15 (2001) 963-977.
- [76] M. Tanaka, M. Sakawa, K. Kato, Application of wavelet transform to compression of mechanical vibration data. *Cybernetics and Systems* 28 (1997) 225-244.
- [77] B.A. Paya, I.I. Esat, M.N.M. Badi, Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical Systems and Signal Processing* 11 (1997) 751-765.
- [78] R.V. Williams, *Acoustic Emission*, Adam Hilger, 1980.
- [79] A. Choudhury, N. Tandon, Application of acoustic emission technique for the detection of defects in rolling element bearing. *Tribology International* 33 (2000) 39-45.
- [80] J. Hanchi, B.E.K. Iamecki, Acoustic emission monitoring of the wear process. *Wear* 145 (1991) 1-27.
- [81] S.L. McBride, R.J. Boness, M. Sobczyk, V.E. Viner, Acoustic emission from lubricated and unlubricated rubbing surfaces. *J. Acoustic Emission* 8 (1989) 192-197.
- [82] C.L. Jiaa, D.A. Dornfield, Experimental studies of sliding friction and wear via acoustic emission signal analysis. *Wear* 139 (1990) 403-424.
- [83] T. Yushioka, T. Fujiwara, Application of acoustic emission technique to detection of rolling bearing failure, New York: ASME, pp. 55-75.
- [84] T. Yushioka, T. Fujiwara, A new acoustic emission source locating system for the study of rolling contact fatigue. *Wear* 81 (1982) 183-186.
- [85] T. Yushioka, T. Fujiwara, Measurement of propagation initiation and propagation time of rolling contact fatigue crack by observation of acoustic emission and vibration, Amsterdam: Elsevier (1985), pp. 29-33.
- [86] A. Morhain, D. Mba, Bearing defect diagnosis and acoustic emission. *Proc IMechE part J* 217 (2003) 257-272.
- [87] A.M. Al-Ghamd, D. Mba, A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing* 20 (2006) 1537-1571.

- [88] S. Al-Dossary, R.I.R. Hamzah, D. Mba, Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing. *Applied Acoustics* 70 (2009) 58-81.
- [89] C.K. Tan, D. Mba, Identification of the acoustic emission source during a comparative study on diagnosis of a spur gearbox. *Tribology International* 38 (2005) 469-480.
- [90] L. Wang, R.J.K. Wood, J.Sun, Acoustic emissions from oil lubricated metal on metal sliding contacts. *INSIGHT* 50 (2008) 506-511.
- [91] E.D. Price, A.W. Lees, M.I. Friswell, Detection of severe sliding and pitting fatigue wear regimes through the use of broadband acoustic emission. *Proc IMechE part J* 219 (2005) 85-98.
- [92] E.Y. Kim, A.C.C. Tan, B.S. Yang, V. Kosse, Experimental Study on Condition Monitoring of Low Speed Bearings: Time Domain Analysis. (1999) Brisbane, Australia.
- [93] Y.H. Feng, F.S. Schlindwein, Normalized wavelet packets quantifiers for condition monitoring. *Mechanical Systems and Signal Processing* doi:10.1016/j.ymsp.2008.07.002 (2008).
- [94] N. Baydar, A. Ball, Detection of gear failures via vibration and acoustic signals using wavelet transform. *Mechanical Systems and Signal Processing* 17 (2003) 787-804.
- [95] J. Nurse, C. Petch, C.E. Fisher, Engine gas path integrity monitoring. Aerotech 94, 1994, paper 470/6/52
- [96] H.E.G. Powrie, K. McNicholas, Gas path condition monitoring during accelerated mission testing of demonstrator engine. AIAA paper (1997), 97-2904.
- [97] H.E.G. Powrie, C.E. Fisher, Monitoring of foreign objects ingested into the intake of gas turbine aero-engine, *International condition monitoring conference*, University of Swansea, (1999) pp. 175-190.
- [98] T.J. Harvey, R.J.K. Wood, G. Denuault, H.E.G. Powrie, Investigation of Electrostatic charging mechanisms in oil lubricated tribo-contacts. *Tribology International* 35 (2002) 605-614.
- [99] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Electrostatic wear monitoring of rolling element bearings. *Wear* 263 (2007) 1492-1501.
- [100] O.D. Tasbaz, R.J.K. Wood, M. Browne, H.E.G. Powrie, Electrostatic monitoring of oil lubricated sliding point contact for early detection of scuffing. *Wear* 230 (1999) 86-97.

- [101] S. Morris, R.J.K. Wood, T.J. Harvey, Electrostatic charge monitoring of unlubricated sliding wear of a bearing steel. *Wear* 255 (2003) 430-443.
- [102] H.E.G. Powrie, O.D. Tasbaz, R.J.K. Wood, C.E. Fisher, Performance of an electrostatic oil monitoring system during FZG Scuffing Test, *International condition monitoring conference*, University of Swansea, (1999) pp. 155-174.
- [103] P.J.Graham-Jones, B.G.Mellor, Expert and Knowledge-based systems in failure analysis. *Engineering Failure Analysis* 2 (1995) 137-149.
- [104] Y.L.Su, J.S.Lin, S.K.Hsieh, The tribological failure diagnosis of spur gear by an expert system. *Wear* 166 (1993) 187-196.
- [105] H.W.Walton, Failure diagnostics--application of expert systems, in proceedings of the International Conference and Exhibits on Failure Analysis. *ASM International* (1992) 207-210.
- [106] R.P.Roberge, C.G.Selkirk, G.F.Fisher, Developing an expert system assistant for filter debris analysis. *J Lubr Eng* (1994) 678-683.
- [107] K.Xu, A.R.Luxmoore, An integrated system for automatic wear particle analysis. *Wear* 208 (1997) 184-193.
- [108] K.Xu, A.R.Luxmoore, L.M.Jones, F.Deravi, Integration of neural networks and expert systems for microscopic wear particle analysis. *Knowledge-Based Syst* 11 (1998) 213-227.
- [109] Z.Peng, An integrated system for wear debris analysis. *Wear* 252 (2002) 730-743.
- [110] T.I.Liu, J.H.Singonahalli, N.R.Iyer, Detection of roller bearing defects using expert system and fuzzy logic. *Mechanical Systems & Signal Processing* 10 (1996) 595-614.
- [111] V.Da Silva, A.Silmara, R.Fujimoto, Rolling bearing fault diagnostic system using fuzzy logic, *IEEE International Conference on Fuzzy Systems*, The University of Melbourne, Australia, 2001, pp. 816-819.
- [112] B.Geropp, Artificial neural networks and fuzzy-logic used for reliable machine diagnosis, *COMADEM 97*, Espoo, Finland, pp. 564-573.
- [113] P.J.Dempsey, W.Morales, A.A.Afjeh, Investigation of spur gear fatigue damage using wear debris. *Lubrication Engineering* 58 (2002) 18-22.
- [114] D.Siganos, Why neural networks?, *web page: www-dse.doc.ic.ac.uk/*.
- [115] D.T.Pham, P.T.N.Pham, Artificial Intelligence in engineering. *Intl.J.of Machine Tools & Manufacture* 39 (1999) 937-949.

- [116] B.Samanta, K.R.Al-Balushi, S.A.Al-Araimi, Gear fault detection using wavelets and artificial neural network. *International Journal of COMADEM* 6 (2003) 22-30.
- [117] D.Chen, W.J.Wang, Classification of wavelet map patterns using multi-layer neural networks for gear fault detection. *Mechanical Systems and Signal Processing* 16 (2008) 695-704.
- [118] I.E.Alguindigue, A.Loskiewicz-Buzak, R.E.Uhrig, Monitoring and diagnosis of rolling element bearings using artificial neural networks. *IEEE Transactions on Industrial Electronics* 40 (1993) 209-217.
- [119] L.Wang, A.D.Hope, Bearing fault diagnosis using multi-layer neural networks. *Insight: Non-Destructive Testing and Condition Monitoring* 46 (2004) 451-455.
- [120] L.Wang, A.D.Hope, H.Sadek, Vibration-based condition monitoring of pumps in the waste water industry. *INSIGHT* 42 (2000) 500-503.
- [121] D.C.Baillie, J.Mathew, A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing* 10 (1996) 1-17.
- [122] D.W.Dong, J.J.Hopfield, K.P.Unnikrishnan, Neural networks for engine fault diagnostics, New York, pp. 636-644.
- [123] C.J.Li, T.Y.Huang, Automatic structure and parameter training methods for modeling of mechanical systems by recurrent neural networks. *Applied Mathematical Modelling* 23 (1999) 933-944.
- [124] P.Deuszkiewicz, S.Radkowski, On-line condition monitoring of a power transmission unit of a rail vehicle. *Mechanical Systems and Signal Processing* 17 (2003) 1321-1334.
- [125] A.J.Hoffman, M.T.Merwe, The application of neural networks to vibrational diagnostics for multiple fault conditions. *Computer standards & Interfaces* 24 (2002) 139-149.
- [126] W.Wang, F.Ismail, F.Golnaraghi, A neuro-fuzzy approach to gear system monitoring. *IEEE Transactions on Fuzzy Systems* 12 (2004) 710-723.
- [127] R.Callan, B.Larder, The implementation of advanced diagnostic and prognostic concepts-practicle tools for effective diagnostic and prognostic health management, *IEEE Aerospace Conference* 2003, Vol. 7-3165
- [128] L.Zhang, L.B.Jack, A.K.Nandi, Fault detection using genetic programming. *Mechanical Systems and Signal Processing* 19 (2005) 271-289.

- [129] B.Samanta, K.R.Al-Balushi, S.A.Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence* 16 (2003) 657-665.
- [130] B.Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing* 18 (2004) 1273-1282.
- [131] B.Samanta, Artificial neural networks and genetic algorithms for gear fault detection. *Mechanical Systems and Signal Processing* 18 (2004) 1273-1282.
- [132] C.J.Li, S.Y.Li, Acoustic emission analysis for bearing condition monitoring. *Wear* 185 (1995) 67-74.
- [133] Q.Sun, P.Chen, F.X.Zhang, Pattern recognition for automatic machinery fault diagnosis. *Trans.ASME, J.Vibr.Acoust.* 126 (2004) 307-316.
- [134] L.B.Jack, A.K.Nandi, Support vector machines for detection and characterization of rolling element bearing faults. *Proc IMechE part C* 215 (2000) 1065-1074.
- [135] L.B.Jack, A.K.Nandi, Fault detection using support vector machines and artificial neural networks augmented by genetic algorithms. *Mechanical Systems and Signal Processing* 16 (2002) 373-390.
- [136] A.Rojas, A.K.Nandi, Detection and classification of rolling element bearing fault using support vector machine. *IEEE Transaction on Computer Science* 25 (2005) 153-158.
- [137] V.A.Skormin, L.J.Popvack, V.I.Gorodetski, M.L.Araiza, J.D.Michel, Applications of cluster analysis in diagnostics-related problems, *1999 IEEE Aerospace Conference*, USA.
- [138] M.Artes, L.Del Castillo, J.Perez, Failure prevention and diagnosis in machine elements using cluster, *Tenth International Congress on Sound and Vibration*, Stockholm, Sweden, pp. 1197-1203.
- [139] J.Schurmann, *Pattern Recognition: A Unified View of Statistical and Neural Approaches*, Wiley, New York, 1996.
- [140] H.Ding, X.Gui, S.Yang, An approach to state recognition and knowledge-based diagnosis for engines. *Mechanical Systems and Signal Processing* 5 (1991) 257-266.
- [141] W.J.Staszewski, K.Worden, G.R.Tomlinson, Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition. *Mechanical Systems and Signal Processing* 11 (1997) 673-692.

- [142] S.K.Goumas, M.E.Zervakis, G.S.Stavarakakis, Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction. *IEEE Transactions on Instrumentation and Measurement* (2002) 497-508.
- [143] X.Lou, K.A.Loraro, Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing* 18 (2004) 1077-1095.
- [144] M.C.Pan, P.Sas, H.Van Brussel, Machine condition monitoring using signal classification techniques. *Journal of vibration and control* 9 (2003) 1103-1120.
- [145] A.R.Webb, *Statistical Pattern Recognition*, Wiley, West Sussex, England, 2002.
- [146] C.K.Mechefske, J.Mathew, Fault detection and diagnosis in low speed rolling element bearing. Part II: The use of nearest neighbour classification. *Mechanical Systems and Signal Processing* 6 (1992) 309-316.
- [147] T.H.Lin, K.G.Shin, A Bayesian approach to fault classification. *IEEE Transaction on Computer Sciences* 13 (1990) 58-66.
- [148] C.S.Hood, G.C.Ji, Probabilistic network fault detection. *IEEE Transaction on Computer Sciences* 18 (1995) 1872-1876.
- [149] X.Z.Wang, C.McGreavy, Automatic Classification for mining process operational data. *Ind.Eng.Chem.Res.* 37 (1998) 2215-2222.
- [150] L.R.Padovese, Comparison between probabilistic and multilayer perceptron neural networks for rolling element bearing fault classification. *International Journal of Modelling and Simulation* 22 (2002) 97-103.
- [151] J.Ma, C.J.Li, Detection of localized defects in rolling element bearings via composite hypothesis test. *Mechanical Systems & Signal Processing* 9 (1995) 63-75.
- [152] B.W.Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: London, 1986.
- [153] M.Markou, S.Singh, Novelty detection: A review. *Signal Processing* 83 (2003) 2481-2497.
- [154] T.Kourti, J.F.MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics Intell.Lab.Syst.* 28 (1995) 3-21.
- [155] J.V.Kresta, J.F.MacGregor, T.E.Marlin, Multivariate statistical monitoring of process operating performance. *Can.J.Chem.Eng.* 69 (1991) 35-47.
- [156] P.Nomokos, J.F.MacGregor, Multivariate SPC charts for monitoring batch processes. *Technometrics* 37 (1995) 41-59.

- [157] J.Liu, K.W.Lim, R.Srinivasan, X.T.Doan, On-line process monitoring and fault isolation using PCA, *2005 IEEE International Symposium on Intelligent Control*, Limassol,Cyprus.
- [158] S.J.Qin, Statistical process monitoring: basics and beyond. *Journal of Chemometrics* 17 (2003) 480-502.
- [159] J.H.Chen, J.L.Liu, Mixture principal component analysis models for process monitoring. *Ind.Eng.Chem.Res.* 38 (1999) 1478-1488.
- [160] M.J.Desforges, P.J.Jacob, J.E.Cooper, Application of probability density estimation to the detection of abnormal conditions in engineering. *Proc IMechE part C* 212 (1998) 687-703.
- [161] E.Parzen, On estimation of probability density function and mode. *Ann.Math.Stat.* 33 (1962) 1065.
- [162] C.Bishop, Novelty detection and neural network validation, *IEE Conference on vision and image signal processing*, 1994, pp. 217-222.
- [163] D.Y.Yeung, C.Chow, Parzen window network intrusion detectors, in: *Proceeding of 16th International Conference on Pattern Recognition*, 11–15 August, 2002, IEEE Computer Society, vol. 4, pp. 385–388.
- [164] L.Tarassenko, Novelty detection for the identification of masses in mammograms, *International Conference on Artificial Neural Networks* (1995), pp. 442-447
- [165] L.P.M.Johnson, M.A.Kramer, Probability density estimation using elliptical basis function. *AIChE J* 40 (1994) 1639.
- [166] D.Lowe, Adaptive radial basis function nonlinearities, and the problem of generalisation, *First IEE international conference on artificial neural networks*, London,UK, pp. 171-175.
- [167] L.Tarassenko, *A guide to neural computing applications*, Arnold, UK, 1998.
- [168] D.A.Clifton, P.R.Bannister, L.Tarassenko, Application of an intuitive novelty metric for jet engine condition monitoring, *Proceedings of the International Conference on Condition Monitoring*, Harrogate UK, 2007, pp. 1149-1158.
- [169] D.A.Clifton, P.R.Bannister, L.Tarassenko, A framework for novelty detection in jet engine vibration data, *Proceedings of the International Conference on Condition Monitoring*, Harrogate UK, 2007, pp. 305-312.
- [170] H.Frigui, R.Krishnapuram, A robust algorithm for automatic extraction for an unknown number of clusters from noisy data. *Pattern Recognition Letter* 17 (1996) 1223.

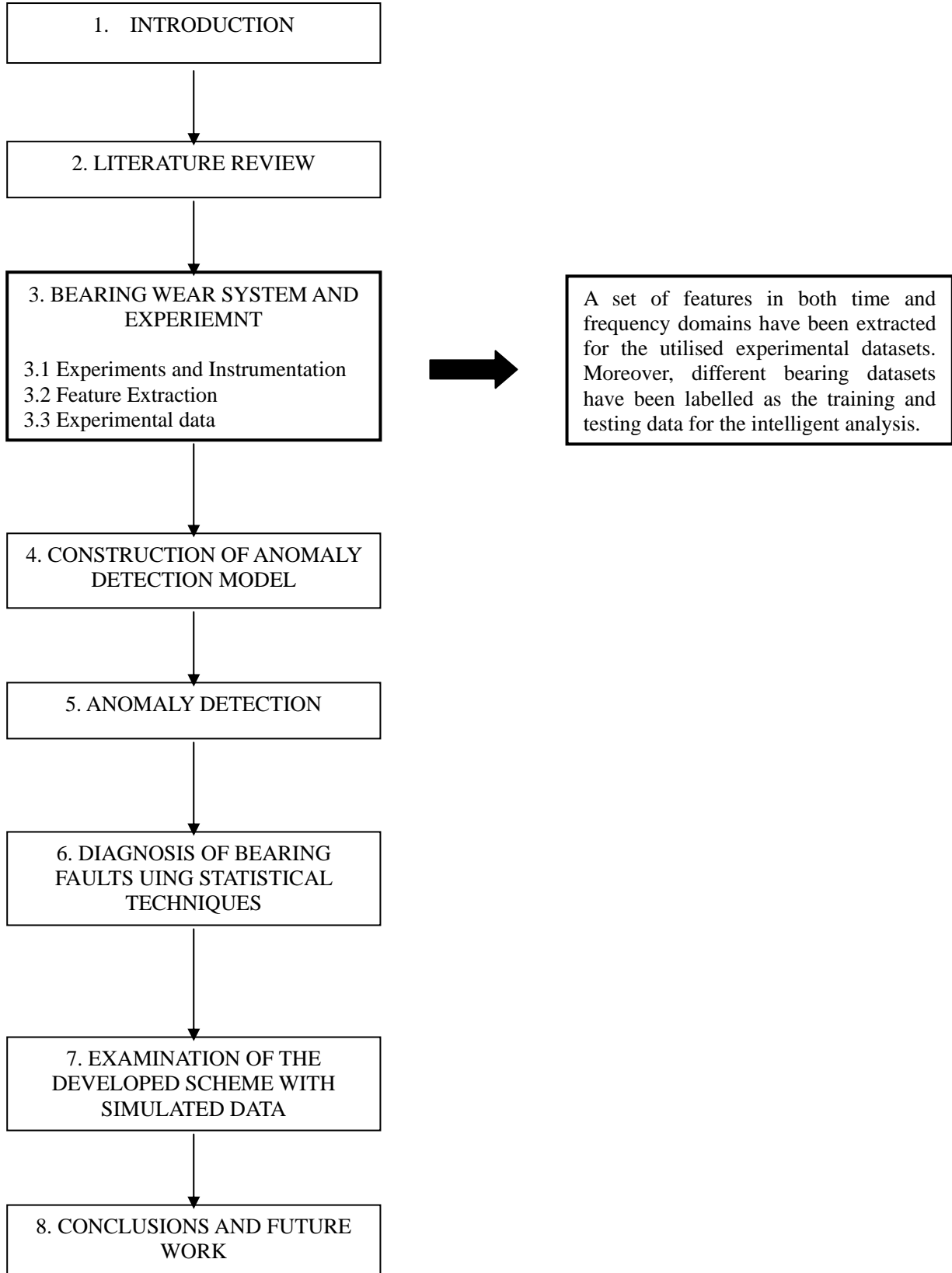
- [171] A.P.Dempster, N.M.Laird, D.B.Rubin, Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B* 39 (1977) 1-38.
- [172] S.Roberts, L.Tarassenko, A probabilistic resource allocating network for novelty detection. *Neural Computation* 6 (1994) 270-284.
- [173] D.Y.Yeung, Y.Ding, Host-based intrusion detection using dynamic and static behavioral models. *Pattern Recognition* 36 (2002) 229-243.
- [174] L.Tarassenko, A.Narirac, N.Townsend, P.Cowley, Novelty detection in jet engines, *IEE Colloquium on Condition Monitoring, Imagery, External Structures and Health*, pp. 41-45.
- [175] H.Akaike, A new look at the statistical model identification. *IEEE Tansction on Automatic Control* 19 (1974) 716.
- [176] G.Schwarz, Estimating the dimension of a model. *Ann.Stat.* 6 (1978) 461.
- [177] G.J.McLachlan, D.Peel, *Finite Mixture Models*, Wiley: New York, 2000.
- [178] C.Fraley, A.E.Raftery, How Many Cluster? Which Clustering Method? Answer via Model-Based Cluster Analysis. *The Computer Journal* 41 (1998) 578-588.
- [179] N.Ueda, R.Nakano, Dterministic Annealing EM algorithm. *Neural Networks*, 11 (1998) 287.
- [180] C.Biernacki, G.Celeux, G.Govaet, Choosing Starting Value for the EM for getting the highest likelihood in Multivariate Gaussian Mixture Models. *Comput.Stat.Data Anal.* 41 (2003) 561.
- [181] C.Ding, X.He, H.Zha, H.D.Simon, Adaptive dimension reduction for clustering high dimensional data, *International Conference on Data Mining (ICDM'02)*, Piscataway, NJ.
- [182] R.Callan, B.Larder, J.Standiford, An integrated approach to the development of an intelligent prognostic health management system, *IEEE Aerospace Conference Montana, USA, 2006*.
- [183] M.Lauer, A mixture approach to novelty detection using training data with outliers, *12th european conference on machine learning* (2001), pp. 300-311
- [184] A.Agogino, K.Tumer, Entropy based anomaly detection applied to space shuttle main engines, *IEEE Aerospace Conference, Montana, USA, 2006*.
- [185] Y.J.Yang, F.Y.Ma, An entropy-based unsupervised anomaly detection pattern learning algorithm. *Journal of Harbin Institute of Technology* 12 (2005) 81-85.

- [186] P.Hall, Y.A.Hicks, A method to add gaussian mixture models, *Technical report, University of Bath*, 2005.
- [187] A.Declercq, J.H.Piater, On-line simultaneous learning and tracking of visual feature graphs, *Online Learning for Classification Workshop, CVPR ' 07*.
- [188] O.Arandjelovic, R.Cipolla, Incremental learning of temporally-coherent gaussian mixture models, *BMVC ' 05*.
- [189] M.L.D.Wong, A.K.Nandi, Modified self-organizing map for automated novelty detection applied to vibration signal monitoring. *Mechanical Systems & Signal Processing* 20 (2006) 271-289.
- [190] A.Ypma, R.P.W.Duin, Novelty detection using self-organizing maps. *Progress of connectionist based information systems 2* (2008) 1322-1325.
- [191] H.Yin, N.M.Allinson, Self-organizing mixture networks for probability density estimation. *IEEE Transactions on Neural Networks* 12 (2008).
- [192] T.Chen, J.Morris, E.Martin, Bayesian control limits for statistical process monitoring, *International Conference on Control and Automation (ICCA2005)*, pp. 409-414.
- [193] F.Zhang, A Mixture Probabilistic PCA Model for Multivariate Processes Monitoring, *2004 American Control Conference*, Boston, Massachusetts, pp. 3111-3115.
- [194] S.Roberts, Novelty detection using extreme value statistics. *IEE Proc.on Vision, Image and Signal processing* 146 (1999) 124-129.
- [195] S.Roberts, Extreme value statistics for novelty detection in biomedical signal processing, *1st International Conference on Advances in Medical Signal and Information Processing*, pp. 166-172.
- [196] P.Miler, R.E.Swanson, Contribution Plots: A Missing Link in Multivariate Quality Control. *Applied Mathematics and Computer Science* 8 (1998) 775-792.
- [197] E.L.Russel, L.H.Chiang, R.D.Braatz, *Data-Driven Techniques for Fault Detection and Diagnosis in Chemical Processes*, Springer, London, 2000.
- [198] R.Treasure, U.Kruger, J.E.Cooper, Dynamic Multivariate Statistical Process Control using Subspace Identification. *Journal of Process Control* 14 (2004) 279-292.
- [199] R.A.Martin, Unsupervised Anomaly Detection and Diagnosis for Liquid Rocket Engine Propulsion, *IEEE Aerospace Conference*, Montana, USA, 2005.

- [200] B.Samanta, K.R.Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems & Signal Processing* 17 (2003) 317-328.
- [201] A.C.McCormick, A.K.Nandi, Classification of the rotating machine condition using artificial neural networks. *Proc IMechE part C* 211 (1997) 439-450.
- [202] A.Murray, J.Penman, Wavelets as an alternative to the FFT in condition monitoring schemes using ANN's, *COMADEM 1996*, Sheffield, UK, pp. 177-186.
- [203] Y.Lei, Z.He, Y.Zi, X.Chen, New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. *Mechanical Systems & Signal Processing* 22 (2008) 419-435.
- [204] C.Bishop, *Neural Networks for Pattern Recognition*, Oxford, UK, 1995.
- [205] C.M.Stellman, K.J.Ewing, F.Bucholtz, Monitoring the degradation of a synthetic lubricant oil using infrared absorption, fluorescence emission and multivariate analysis: a feasibility study. *Lubrication Engineering* 55 (1999) 42-52.
- [206] G.O.Allgood, B.R.Upadhayaya, A model-based high frequency matched filter arcing diagnostic system based on principal component analysis (PCA) clustering, *Applications and Science of Computational intelligence III*, Bellingham, pp. 430-440.
- [207] W.X.Sun, J.Chen, J.Q.Li, Decision tree and PCA-based fault diagnosis of rotating machinery. *Mechanical Systems & Signal Processing* 21 (2007) 1300-1317.
- [208] C.Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

THESIS STRUCTURE

KEY OUTCOME



Chapter 3. Methodology and experiment of bearing wear system

This chapter generally consists of four parts. Section 3.1 proposes the development scheme for this study on the basis of summary of the survey in Chapter 2. In section 3.2, experiment of the bearing wear system, including instrumentation and experimental procedure will be introduced. Sections 3.3 and 3.4 will provide the details about the feature extraction and summary of the obtained datasets for the intelligent analysis respectively.

3.1. Methodology of the condition monitoring scheme

Chapter 2 attempts to summarise recent research and development in condition monitoring techniques and automated decision support systems implementing the condition based maintenance (CBM) programme. Various techniques, models and algorithms have been reviewed following the three main steps of a condition monitoring process, namely data collection, data processing and maintenance decision-making.

In order to collect data, various sensors have been developed to reflect the health of the monitored components. It has been discussed that each of these sensing technologies has their unique benefit, and should not be easily neglected. Valuable features from multiple sensors are recommended for implementation for the multivariate analysis. In this study, one of the aims is to identify various bearing operation modes such as running-in, sub-surface cracks, cracks and spallation in the operation life, and according to the reviewed literatures, some mechanisms can be detected by multiple sensors whereas several others could only be detected by specific ones. Therefore, in this study, multiple sensors have been implemented and are expected to extract multiple bearing failure mechanisms in a combined or unique way. Moreover, signal processing techniques in time, frequency and time-frequency domains are essential to extract reliable sensor features. However, as Peng and Chu [1] indicated, there is no standard rule to select wavelet functions for different tasks, and the problem of how to determine the scale ranges of wavelet transforms is usually ignored. In order not to involve too much algorithm justification

work for wavelet transformation, it is preferred that the statistical parameters in the time domain and power spectrum of the frequency domains are utilized to form multiple feature sets. Moreover the extracted features are not always clean, and always contain missing or sparse data (i.e. single spikes), and these unusual data will sometimes confuse the fault detection results (i.e. increasing false alarm rate), so the problem of how to handle these unusual data is becoming important.

Having obtained the processed multivariate features, decision support needs to be carried out to provide information on whether the machinery is abnormal or not, where the anomalies are and what the causes are for these abnormalities. Among these tasks, the latter two have drawn great attention from researchers, and various automated fault classification systems have been developed. However, irrespective of the approaches mentioned above, the general procedure is almost unified for an automated scheme: a set of experimental data representing various conditions of the bearing needs to be generated by simulating or introducing faults on different components (e.g. inner race, outer race or rollers of bearings) or with varied severity (e.g. different sizes of defects), and then, the selected system can ‘learn’ from such training datasets. Finally, the testing dataset is applied to validate how well developed the system is by calculating the classification rate. Although the performance of these developed systems is consistently being improved with the adaptation of algorithms or implementation of data pre-processing and feature selection techniques, there are still several concerns over this development strategy: 1) it is not easy to obtain training data covering all aspects of bearing symptoms, for example, data of bearing sub-surface cracks; 2) simulated faults represent realistic bearing fault conditions, as bearings usually experience much more complicated propagation processes, even if the seeded faults were introduced beforehand; 3) multi-class fault classification strategy is focused on accurate fault classification, but this seems insufficient if the task is to detect novel events (i.e. different bearing wear mechanisms such as crack and spallation) from the datasets. It should be noted that these concerns are not intended to doubt the classification capability of these developed systems but to point out that new strategies might be needed if the task is to extract and explain more valuable information from the time history of the datasets.

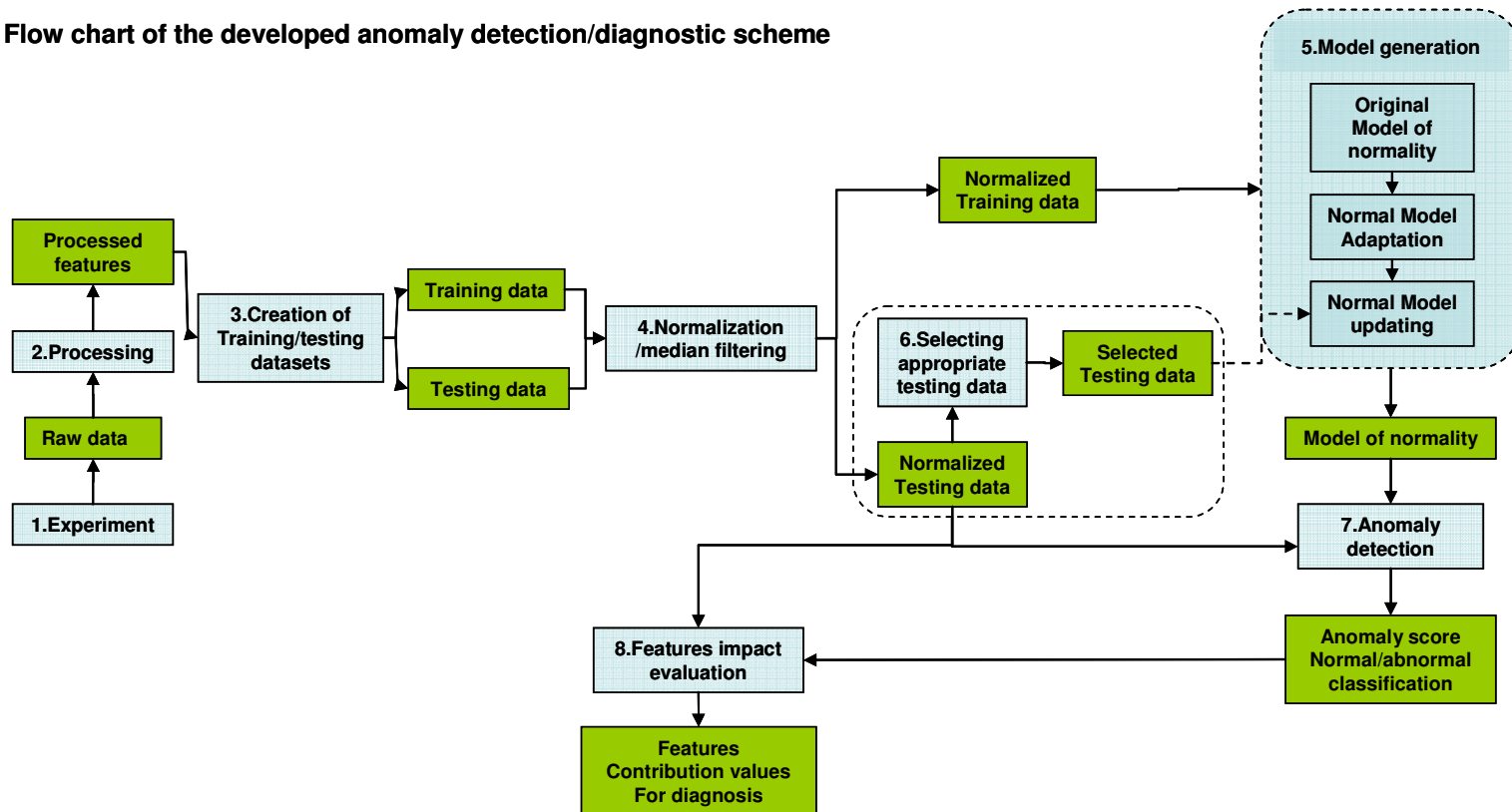
In order to provide early warning of the potentially ill-defined possible failures, and to cope with the limitation of the multi-class training data which are required during the training phase of the machine learning methods, anomaly detection is conducted to identify whether the machine condition deviates from a model of normality. Under this circumstance, several unsupervised machine learning techniques such as Gaussian mixture model (GMM), k-means clustering and self-organizing map (SOM) were applied to summarize the distribution of the training data. Although it is reported in many studies that the detection performance could be significantly improved by characterization of the training data, a very practical problem is also proposed which is that the training data might include a certain amount of unknown anomalies, so that the model could not be expected to respond to these anomalies that have been used to construct the model of normality. Therefore, the established model needs adaptation to make sure it is sensitive to these anomalies. Apart from fault detection, in the applications of SPC, the so-called contribution values are also calculated for each applied feature to examine which features are driving the detected anomalies so that diagnostic information can be obtained. The described anomaly detection strategy has three major advantages: 1) the model is built with a single class of normal data which is the most easy-to-obtain data; 2) the operating conditions can be monitored through time histories of anomaly detection indices to extract abnormal information; 3) the diagnostic information is given by an unsupervised method without prior knowledge, and the feature impact can be examined at any period of interest to extract the ‘where’ and ‘what’ information. These characteristics of anomaly detection satisfy the current requirement for information extraction, hence anomaly detection is chosen as the fundamental development strategy for bearing fault detection and diagnosis.

Figure 3.1 shows a flow chart of the methodology of the proposed scheme. At the beginning, multiple features are extracted from the raw data by signal processing techniques, and both training and testing datasets are created for further analysis. Next, the Gaussian Mixture Model (GMM) is utilized to characterize the training data, and a preliminary anomaly detection model is established. Nevertheless, as described above,

the constructed model needs adaptation to make sure it is sensitive to anomalies, so innovative methods are adopted to locate and trim Gaussian components that are associated with anomalies. After the model adaptation, cleaned training data and an adapted anomaly detection model are obtained, and these two elements are then used to generate a threshold line for anomaly detection. The scheme then calculates the anomaly detection index (Hotelling's T-squared statistic in this study) of the testing data against the adapted model to discover abnormal events which are above the threshold line. Finally, Principal Component Analysis (PCA) is applied to the adapted model and cleaned training data, and the testing data is fed into the PCA model, so that the features contribution values can be estimated to obtain diagnostic information.

More details and the theoretical background of the methodology used to develop the proposed scheme have been introduced in Chapter 2 of this thesis, and information on feature extraction and utilized experimental bearing datasets are provided in Chapter 3. Chapters 4, 5 and 6 present the results of the anomaly detection model construction and anomaly detection as well as diagnosis, respectively. In order to further investigate the robustness of the developed approaches, multivariate data are also simulated in Chapter 7. Finally, conclusions and future work are given in Chapter 8. Table 3.1 summarizes the key elements of the developed scheme, and could be utilized as the reference throughout the whole thesis.

Flow chart of the developed anomaly detection/diagnostic scheme



Please see the page 94 for more detailed explanation:

1. The bearing tests were carried out to obtain the multiple sensing data. On the other hand, synthetic multivariate data are also implemented to validate Developed scheme

2. The Signal processing techniques were Applied to get various features in both time and frequency domains.

3. Normally, baseline tests are used as training data to build the model of normality; while the failure tests are used as the testing data to detect anomalies.

4. Normalization was applied to both training & testing data, and the rogue data are also filtered by median filtering techniques.

5. Main work of the project: various training parameters and their effect on the model building anomaly detection were evaluated. And two key problems In the anomaly detection were addressed by the developed approaches of model adaptation and updating.

6. The method is developed to select appropriate Testing data to update the model of normality.

7. Anomaly detection is conducted by calculating the Hotelling's T-squared distance between the testing samples and origin of the model of normality

8. Contribution values of the applied features to the T-squared distance are compared and the diagnostic Information could be provided.

Figure 3.1 Flow chart of the developed scheme

The explanation and keywords in this table can be used as the reference throughout the thesis

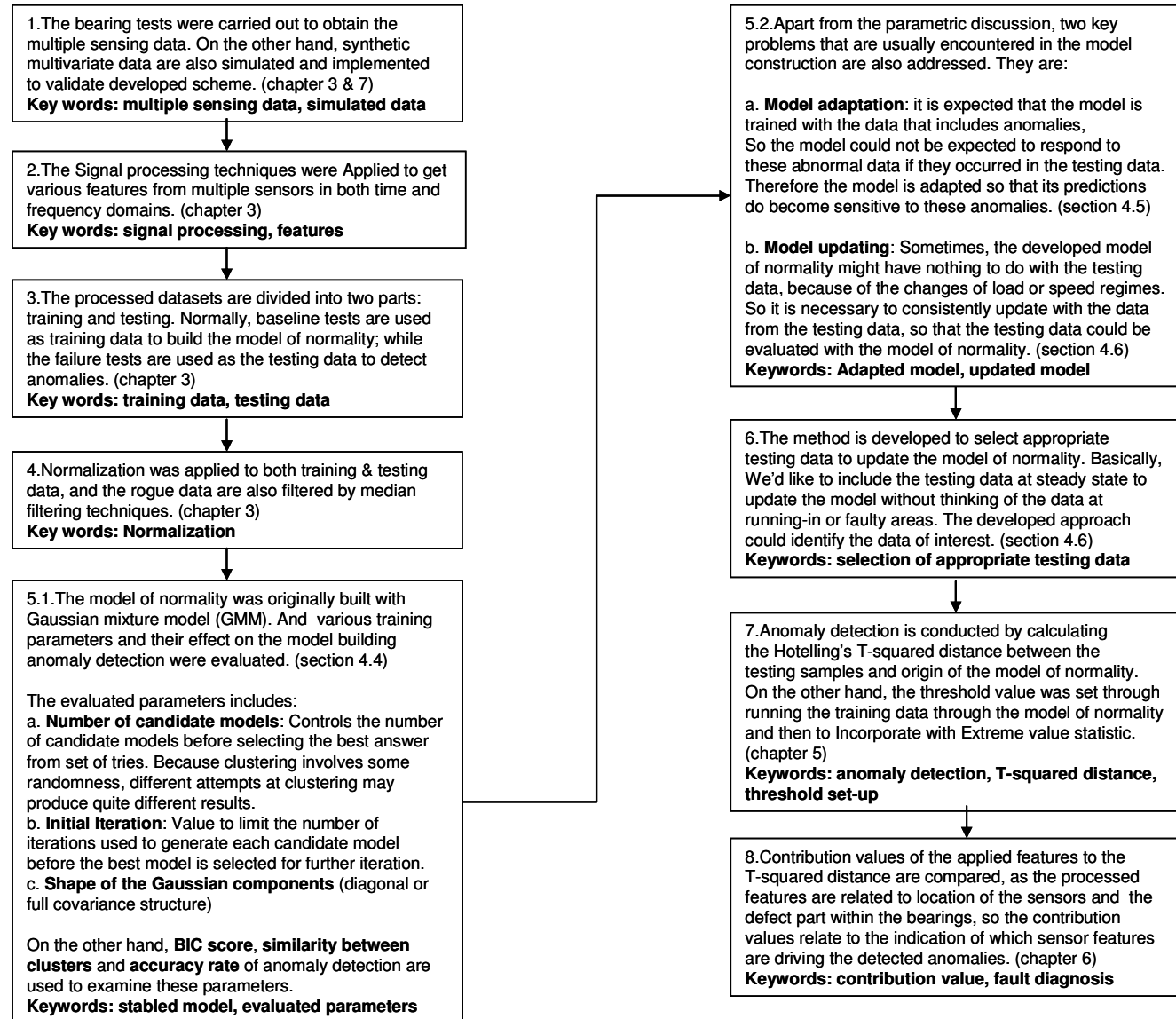


Table 3.1 Key elements in the developed scheme

3.2. Experiment and instrumentation

The original proposed bearing condition monitoring strategy at the beginning of this study has been developed and verified through both experimental and simulated data. This chapter presents how the experimental data were obtained with the details of the bearing tests. It should be noted that the experimental bearing data utilised in this thesis was originally collected by Dr. Terry Harvey and Mr. Mark Craig [2, 3]. The following sections will briefly describe the experimental set-up and test conditions used to run accelerated testing of taper roller bearings, as well as the signal processing techniques deployed for feature extraction.

3.2.1. General

The bearing rig is positioned on a concrete block with a steel-mounting frame upon an oil-drip tray. The experimental parameters used for the bearing accelerated tests are shown in Table 3.2 [2].

General	
Speed:	2500 rpm
Maximum load:	20 kN (200% max. dynamic load)
Lubricant:	TRB 115-032
Lubricant flow rate:	4 litres min ⁻¹
Bearings	
Outer race:	LM67010
Cone:	LM67048
Contact angle:	13.5°
Pitch diameter:	1.78" (45.226 mm)
Inner diameter:	1.25" (31.75mm)
Outer diameter:	2.328" (59.13mm)
Bearing width:	0.6467" (14.43 mm)
Rollers per row:	19
Mean roller diameter:	0.2436 in. (6.186 mm)

Table 3.2 Experimental conditions

The bearings used in the experiments are taper roller bearings (Timken LM67010 and LM67048) see Table 3.2 for specification. The rig utilises four bearings: two support bearings are located in the two end housings and two test bearings are installed in the centre housing. Figure 3.2 shows a section of the bearing test chamber.

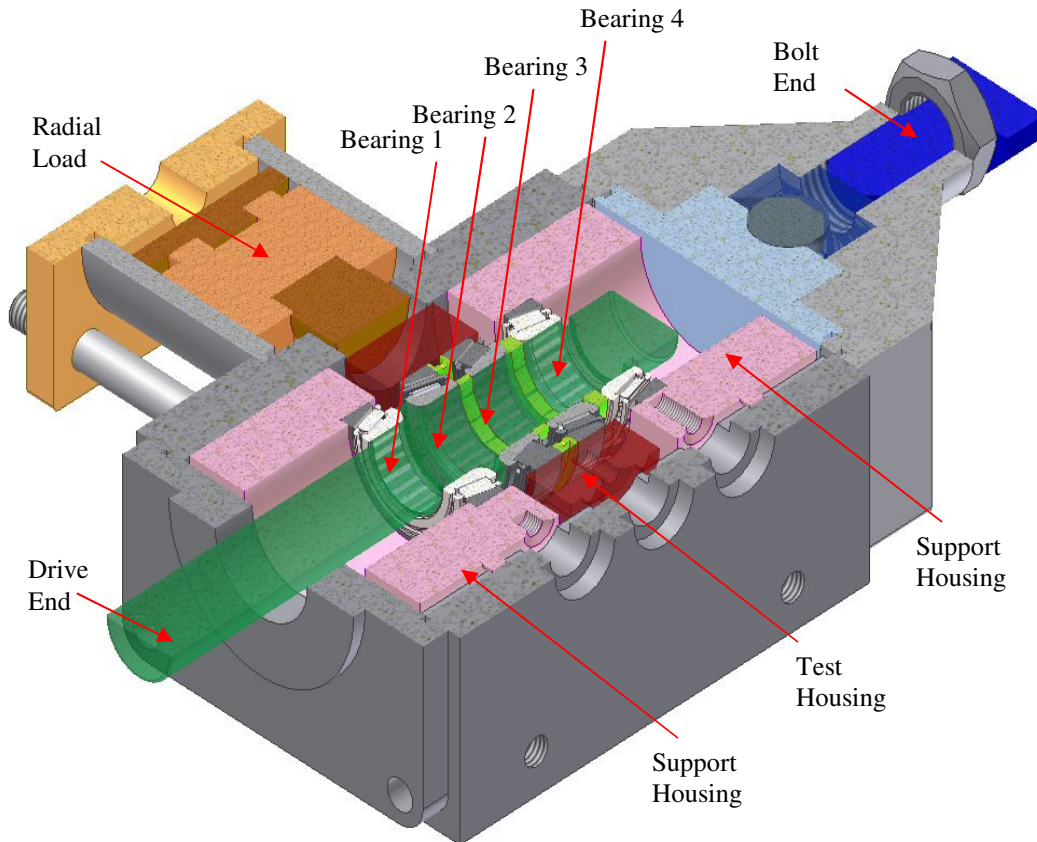


Figure 3.2 Bearing test chamber

There were two types of tests carried out, and their generated data were used in this thesis: one is the base-line test, in which healthy bearings were tested to obtain the base-line data; the other one is the run-to-failure tests that were accelerated by having an indent pre-seeded on the inner race of the test bearing No.2 prior to each test. These tests were conducted by Dr. Terry Harvey and Mr. Mark Craig, and more details about the tests can be found in [2, 3]. As Figure 3.3 (a) shows, the pre-indenting conducted by Timken company was on the inner race and achieved by applying a Rockwell Hardness testing machine with a 'C' diamond and a Rockwell C scale. This process applies a 150-kgf load with an included angle of 120° . The profile of pre-indentation on the modified surface is

shown in Figure 3.3 (b). The loading of the test bearings is also included as another driving factor of accelerated test; a hydraulic system was used to apply pressure (radial load) to a piston, which pushes against the centre housing accommodating the two test bearings (one with pre-indentation), as shown in Figure 3.2. Furthermore, drive end and bolt end in the figure respectively imply two ends of the shaft which connects the driving motor and is tightened by a bolt.

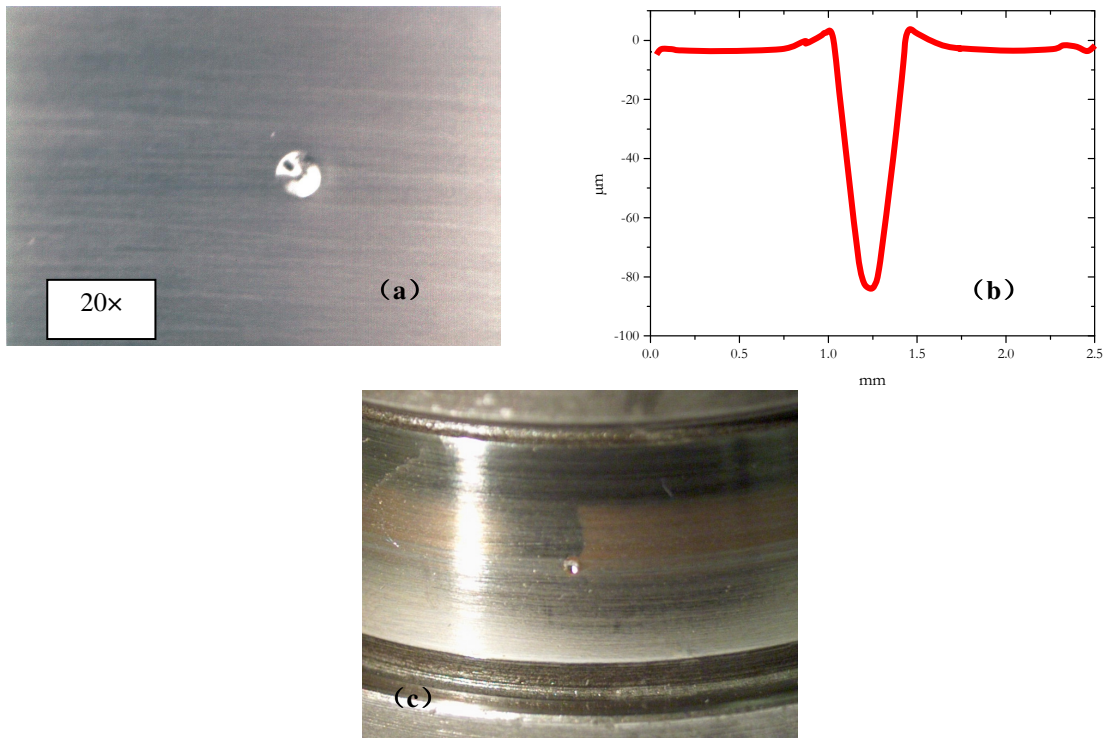


Figure 3.3 (a) optical image of indent; (b) profile from 2-D Talysurf profilometer with 3 μm diamond tip stylus; (c) optical image of inner raceway and indent location

The oil recirculation system (see Figure 3.4) consists of mainly copper and rarely nylon pipework with brass connectors [2]. The flow rate is controlled and monitored by the bypass system and flow meter respectively. The oil circulation system contains two types of filter: conventional paper filters which can capture debris up to 25 μm, positioned before the test housing, and the one for the collection of 6 μm debris, after the test housing but before the oil pump. The second type of filter is magnetic, to enable collection of ferrous debris generated during testing, and this filter is located just before the bearing chamber.

A hydraulic system applied a radial load to a piston which induced pressure to the test housing. The bearing load was 20 kN, which produces an equivalent Hertzian contact pressure between the roller and raceway of 2.5 GPa, with a line contact width of 200 μm . The end load was applied axially by the screw thread. It should be noted that the test was accelerated by overloading the test bearings, approximately 200% dynamic rating. The reason for applying the dynamic loading rating rather than static load is that the static load of a bearing block for a given diameter ballscrew is a measure of the capacity of the bearing block to support the load under non-operating conditions. The dynamic load rating is a measure of the applied operating load the bearing blocks can withstand. In a radial load application, the load is usually supported by raceways or rollers. The screw is primarily used to move the load. For this reason, the dynamic load rating of the bearing is more critical than the static load. As the Table 3.5 shows, all the bearing tests utilised in this thesis were accelerated with the dynamic load rating, and more introduction about dynamic load rating calculation can be found in [2,3].

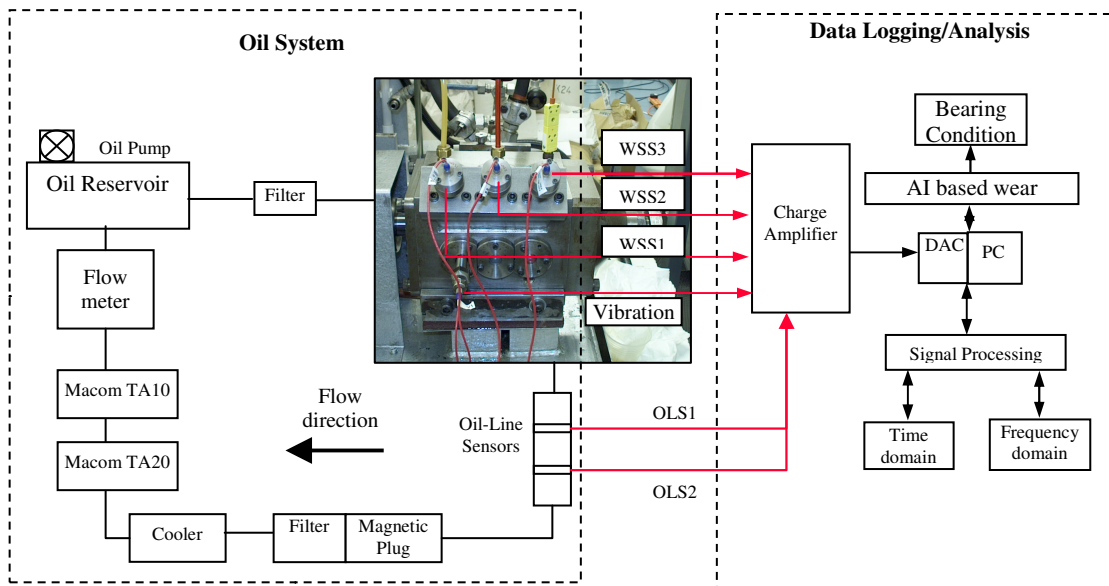


Figure 3.4 Schematics of the whole bearing monitoring system

3.2.2. Sensing technologies

There are three sets of monitoring devices fitted to the rig and these include: vibration sensing, electrostatic sensing and two Macom Technologies devices [4] measuring debris production. The installation positions of all the sensors can be found in Figure 3.5.

Among these sensors, a single vibration sensor is mounted on a vibration stock installed on the front housing. The vibration sensor is a piezoelectric type, with a sensitivity of 28.9 pCg^{-1} . The set-up of this parameter can be found in [2, 3].

The electrostatic sensors employed to monitor bearing performance can be divided into two types. The first type is a wear-site sensor (WSS), Which detects electrostatic charge in the vicinity of the bearings. There are three electrostatic wear-site sensors installed in the bearing rig, WSS1 and WSS3 monitor two support bearings (#1 and #4 respectively); WSS2 monitors two test bearings (#2 and #3), as illustrated in Figure 3.5. The second type of electrostatic sensor is the oil-line sensor (OLS). Generally, it consists of two ring sensors located in the oil recirculation system, which detect charge associated with the passing debris generated from the bearings.

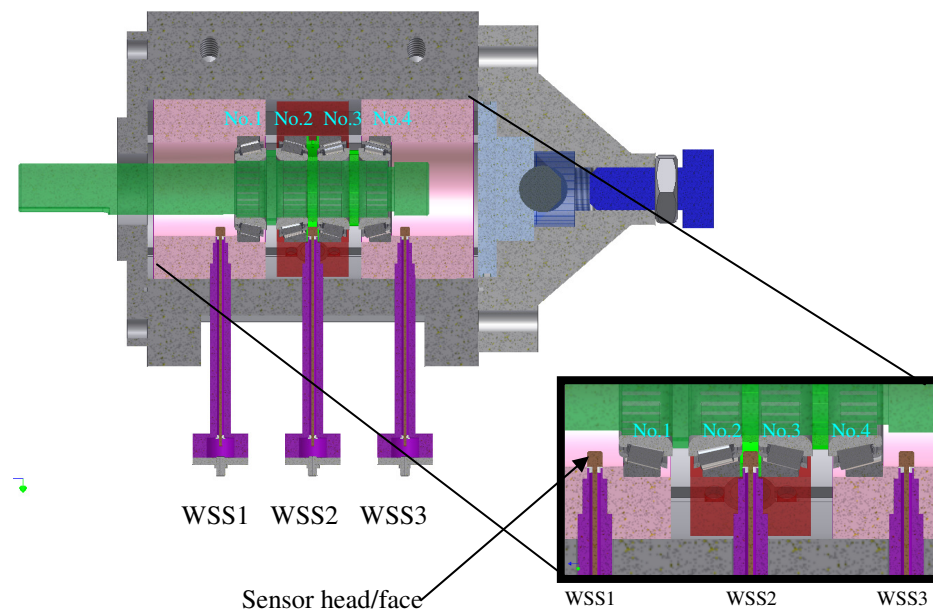


Figure 3.5 Installation positions of the WSS sensors

The Macom TA10 works when oil passing through the transducer coil and the condition of this can be slightly changed as a result of the existence of metallic particles. Ferrous particles increase the inductance because of their high relative permeability, while non-ferrous metallic particles reduce the inductance because of eddy currents occurred in the particles. Thus, the signal phase for ferrous and non-ferrous debris is opposite and enables the device to discriminate between ferrous and non-ferrous debris. Furthermore, the signal magnitude provides a measure of debris size and concentration. The measurement sensitivity of the TA10 is down to 50 μm [2, 4].

The Macom TA20 is triggered at the time of oil samples flowing through the device. The accumulated ferrous debris collected by sensor probe can modify magnetic flux in the internal circuit. This modification of flux is then converted into an electrical signal, which correctly reflects the total amount of collected debris. The TA20 detects magnetic saturation while the flow cleans the accumulated particles as a result of a flush cycle where the magnet is drawn back from the oil line. A non-magnetic boundary in the middle of the oil and the magnet helps to keep integrity of oil line. Finally, the magnet precisely returned to the original position in the oil line after the completion of the flush cycle. Furthermore, factor of time is taken into account while measuring bulk particulate load, and rate of debris production can therefore be easily estimated [2, 4].

Apart from the on-line sensing techniques, off-line debris analysis was conducted with the oil samples collected at a drain point during testing. Due to the large volume oil stored in the reservoir, the reduction of oil is assumed to have a limited effect on the lubrication of the bearings. These off-line techniques are generally used as the assistant physical evidence to the on-line data analysis.

As a type of off-line analysis, Ferrographic analysis indicated the severity of wear index $SI = (AL^2 - AS^2)$ where AL and AS represent the fractal area filled by large and small debris respectively. In the experiments conducted by Mr. Mark Craig [3], the large abnormal debris is greater than 15 μm and small normal debris is sub 5 μm [5, 6].

The HIAC portable oil diagnostic system (PODS) type 2 was also used in the Mark Craig's experiments as another off-line test. An average count from 3 test cycles is applied to ensure consistent debris diffusion throughout the whole sample, which is accumulated according to bin size between 4-64 μm and defined using National Aerospace Standard (NAS) 1638 [3].

3.2.3. Data logging & analysis system

Both electrostatic and vibration sensors are connected to a charge amplifier via low-noise microdot cables. The amplifier converts the charge signals to measurable voltage outputs that are fed into an A/D acquisition system. The raw voltage signal is then conditioned with a signal processing unit and processed in both time and frequency domains. Finally, AI-based data mining techniques are carried out to obtain the bearing condition monitoring and fault diagnosis. The sampling rate for vibration and WSSs raw data was set to 8190 Hz, and for the OLS was set to 2048 Hz in the experiments [2, 3]. The band-pass frequency requirement of the charge amplifier was fulfilled by a high pass filter of 1Hz and low pass filter of 10 kHz, which excludes effect of dc drift and high frequency noise. This band-pass frequency set-up was experimentally tested and provided by the GE Aviation. The on-line signal processing system, developed by GE Aviation, includes processed parameters in both time (i.e. RMS value) and frequency (i.e. Power spectral density) domains recorded every 2 seconds.

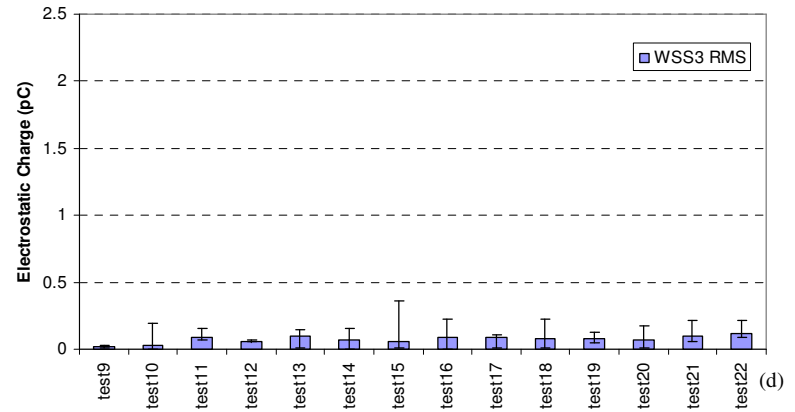
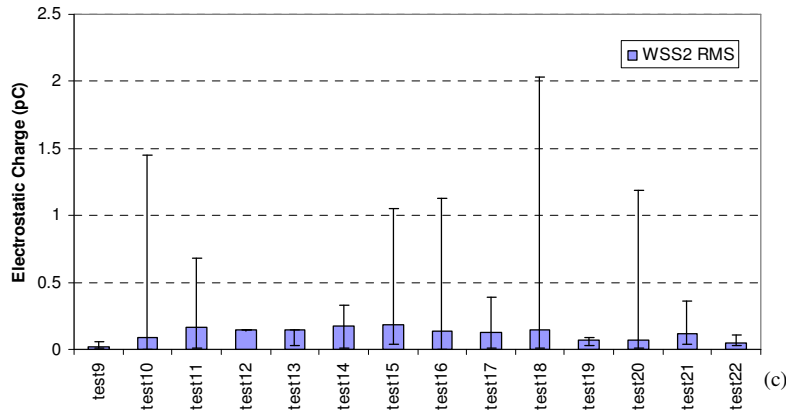
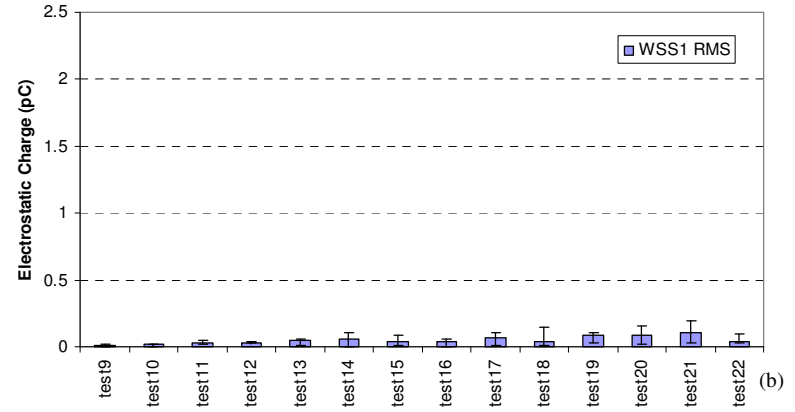
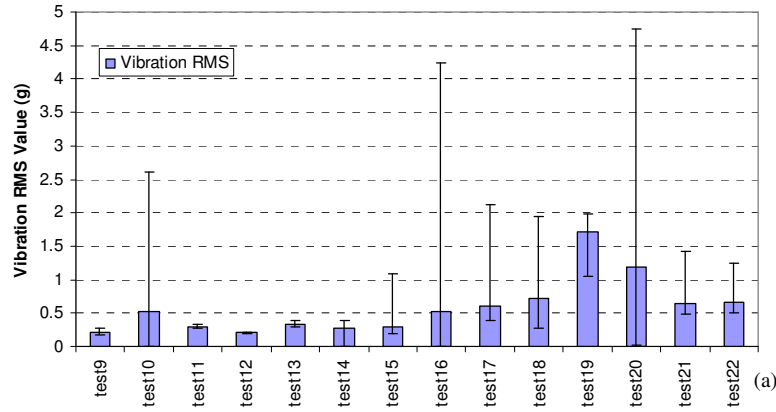
3.3. Feature extraction

3.3.1. Time domain features

The post-processed signal features from the three types sensors (vibration, WSSs and OLS) were extracted to construct the training and testing data sets. For vibration and WSS sensors, the RMS value is selected as the time domain feature. For the OLS analysis, three time domain parameters including RMS values of the two oil line sensors and the cross-correlated value that couples the two OLS sensors to identify charge moving at the oil flow rate were utilized to generate the training data set. This training and testing set was produced one point every two seconds [2, 3].

Figure 3.6 illustrates the RMS values of the three types of sensors. It is found that, with the same processed feature, such as vibration RMS value in Figure 3.6 (a), mean value, minimum value and maximum value (each error bar implies the maximum, mean and minimum RMS values during each test) between different tests deviate significantly. This phenomenon appears obviously between the baseline tests (i.e. tests 9, 11 and 12) with the healthy bearing condition and the run-to-failure tests (i.e. tests 16, 17, 18 and 20) with the fatigue damaged condition at the end of the tests. In baseline tests, it is seen that the mean, minimum and maximum values are closed to each other; this is due to the flat trend and low value of the data throughout the tests. On the other hand, maximum values dramatically increase in the run-to-failure tests because of the increasing trend and higher values associated with substantial damages. Furthermore, the RMS values between five sensors also show great difference. For example, WSS1 and WSS3 in Figures 3.6 (b) and (d) are found with lower RMS values with relatively flat trend compared to the ones in WSS2. Since these wear site sensors monitor different situations and bearings with varied fault conditions, which provide the ability to extract diagnostic information in this study. From Figure 3.6, two discovered characteristics (differences between the tests and differences between the sensors for each individual test) can be found and maybe utilised in the later analysis. For example, the differences between the Max/Mean/Min values in baseline tests (e.g. test 9, 11, 12 and 19) are much smaller than those in run-to-failure tests (all except baseline tests); the difference between the sensors can help to locate the faults and this information is expected to be automatically extracted by the developed approaches.

For the time domain features, other statistics such as standard deviation, skewness and kurtosis can also be calculated and implemented. However, having considered the number of data points, and in order not to trigger the problem of ‘curse of dimensionality’ (introduced in section 2.4.3), these time domain features have not been included in the datasets for the intelligent analysis, but may be considered in the future study.



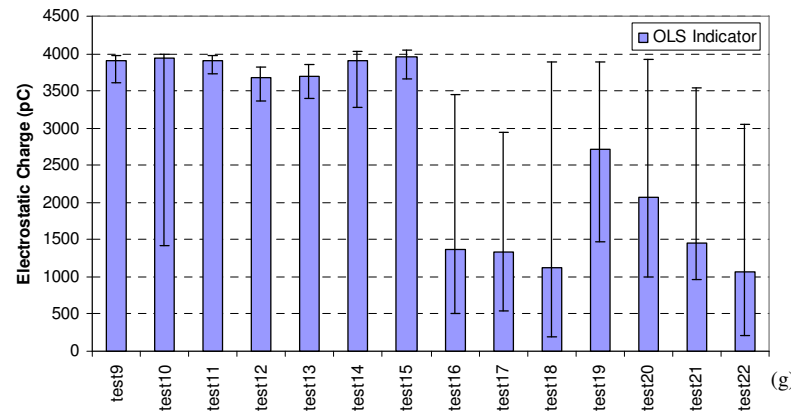
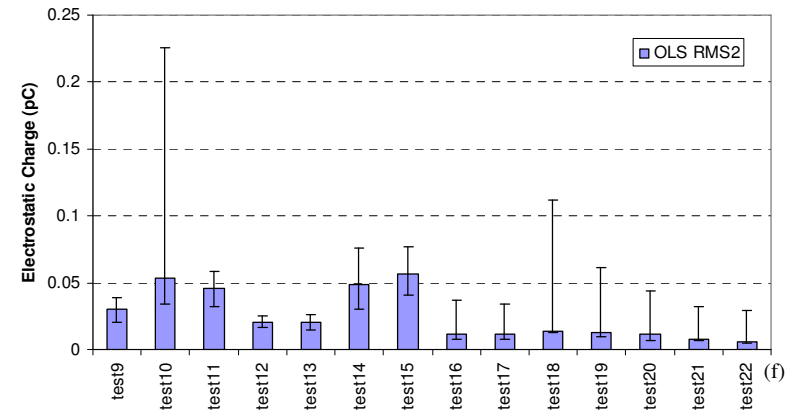
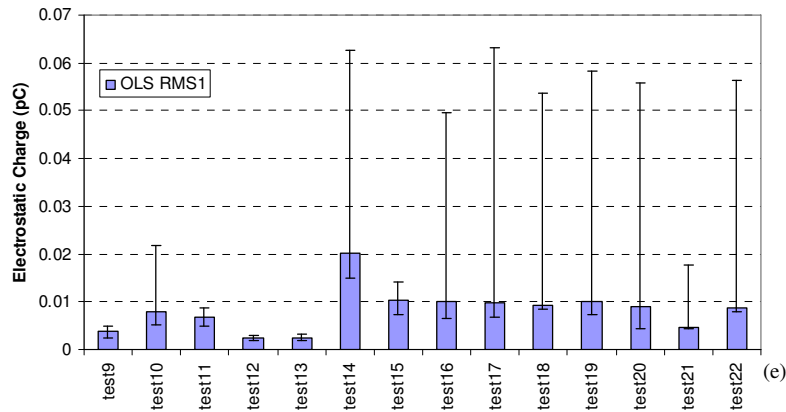


Figure 3.6 Max, Mean and Min RMS values of the 5 applied sensors for all the 14 tests
(Max/min RMS values are top/bottom of the error bar)

3.3.2. Frequency domain features

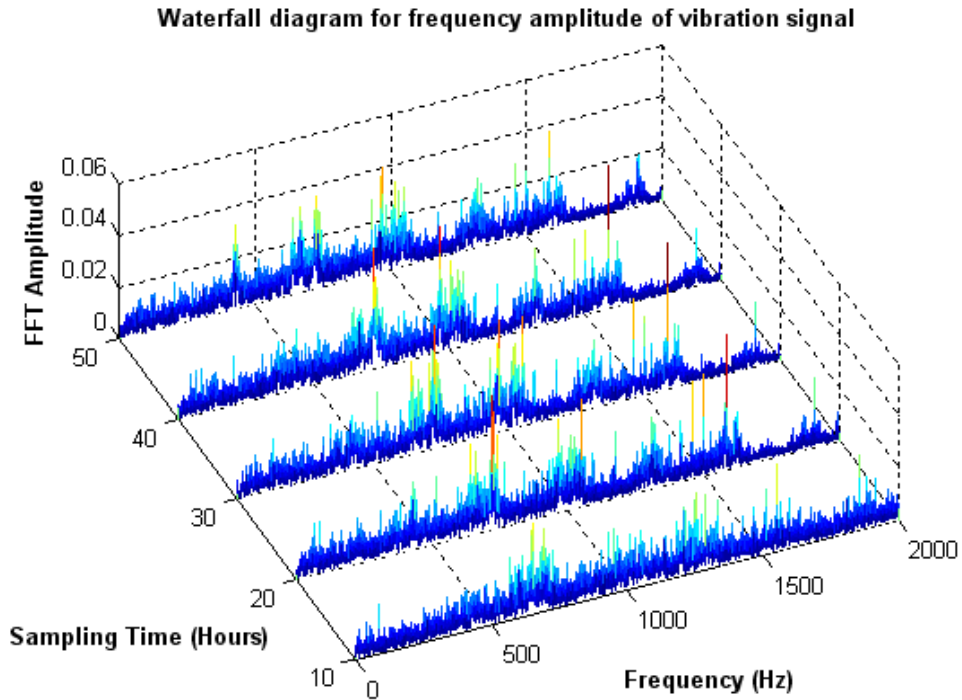
To form the vibration and WSSs training set, six dimensional features were selected for each of the two sensing technologies. In these features, the first and second dimensions were chosen as the RMS value of the time domain and the energy at the rotational frequency respectively. The third to sixth features are based on the energies at the bearing defect frequencies, i.e. races, rollers and cage. These features are used to reveal the information on the bearing components. Each energy was externally clocked using the once-per-revolution signal from the shaft tachometer. Datasets of RMS value and energies at bearing defect frequencies were produced at two points per minute. The bearing defect frequencies are listed as the equations (2.1)-(2.4) show and Table 3.3 summarises calculated frequencies for each frequency-domain feature. Figure 3.7 illustrates typical FFT spectra of vibration and WSS2 electrostatic signals at different sampling times obtained from test 16 data which shows a typical steady increasing trend of the amplitude at particular frequency bands.

Shaft Rotating speed (rpm)	2500
Tapered Roller Bearing element frequency	Value (Hz)
Shaft rotating (Tacho) frequency	42
Fundamental train (or Cage) frequency	18.063
Roller spin frequency	149.617
Inner Race defect frequency	448.473
Outer race defect frequency	343.194
Rolling element defect frequency	299.233

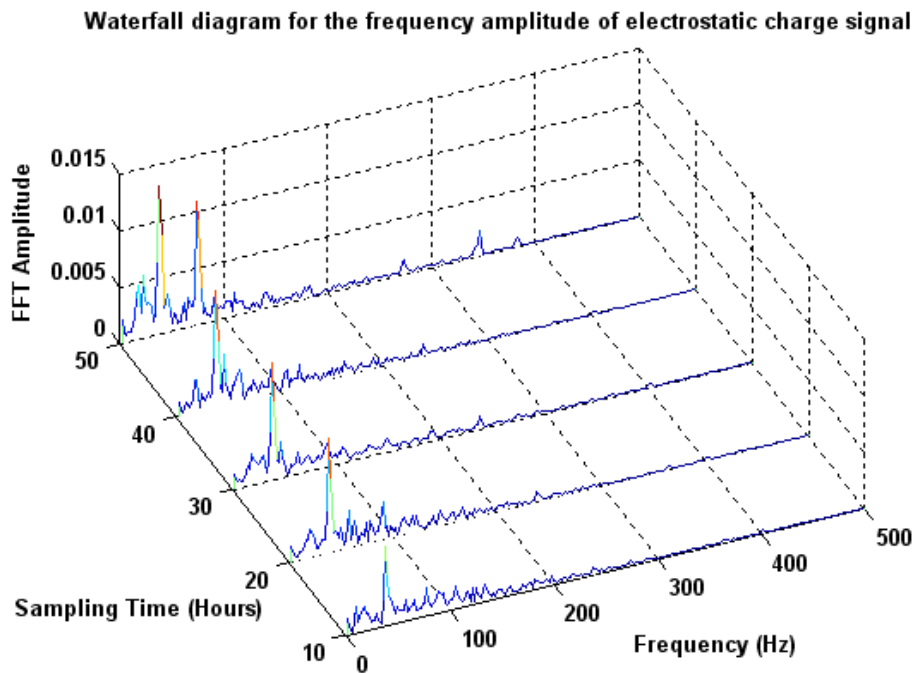
Table 3.3 Calculated frequencies for bearing elements

From Figure 3.7, it is found that, with the time progressing, amplitude at specific frequency bands steadily increases, i.e. vibration amplitude at around 1000Hz and WSS2 amplitude at 500 Hz. These amplitude values at specific frequency bands can be collected continuously as the time passes and then used to form the processed feature sets as the function of time for the intelligent analysis at later stages. Take WSS2 amplitude for example, it is seen that at particular frequency bands, such as 40-100 Hz and 300-400 Hz, the amplitude of the feature steadily increased, which means the elements of the bearing

associated with these frequencies may be experiencing a wear propagation process, and the proposed approaches are expected to extract or enhance this trend.



(a)



(b)

Figure 3.7 Waterfall plots for the frequency spectrum, (a) shows the plot of vibration signal, (b) shows the plot of WSS2 signals

Table 3.4 summarises the 27 features discussed above in both time and frequency domains. These features include the RMS values for vibration, WSSs and OLS; and the five spectral energies at bearing defect frequencies for vibration and WSSs. These features are extracted for both training and testing data and used not only in a combined way for the multivariate anomaly detection but also in a separate way to reveal the diagnostic information

Vibration	WSS1	WSS2	WSS3	OLS
1.RMS	7.RMS	13.RMS	19.RMS	25.RMS1
2.Tacho Energy	8.Tacho Energy	14.Tacho Energy	20.Tacho Energy	26.RMS2
3.Cage Energy	9.Cage Energy	15.Cage Energy	21.Cage Energy	27.Cross-correlated values
4.Roller Energy	10.Roller Energy	16.Roller Energy	22.Roller Energy	
5.Outer Race Energy	11.Outer Race Energy	17.Outer Race Energy	23.Outer Race Energy	
6.Inner Race Energy	12.Inner Race Energy	18.Inner Race Energy	24.Inner Race Energy	
Front housing	Vicinity of Bearing1	Vicinity of Bearings2&3	Vicinity of Bearing4	Oil-line

Table 3.4 Extracted features for the data analysis

It should be noted that the bearing fault frequencies used as features in the thesis are the theoretical values calculated for each of the bearing rolling elements. In practice, roller bearings experience slippage, which causes random variations in the fault frequencies. In addition, bearing fault frequencies usually have extremely low energy. Many research work have reported that the vibration based bearing diagnostics can be based on more advanced signal processing techniques such as envelope analysis, which shows up bearing fault frequencies indirectly via modulation of excited structural resonance. It is, therefore, likely that this choice of features could lead to the difficulties in the different stages of the data fusion process, and evaluation of more advanced signal processing techniques are essential for the future study.

3.4. Summary of the datasets

In this thesis, 14 sets of data are utilized to validate the proposed anomaly detection and diagnostic scheme. Table 3.5 summarises the acquired datasets and their test conditions and functions. From the test condition used, there are two modes of test, the first one is

the baseline test, in which bearings are free from defects; the second one is the defect test, in which the inner race of one of the test bearings was pre-indented to accelerate the fatigue failure process (For details of the test conditions, such as bearing pre-indentation and dynamic load setup can be found in section 3.2.1). According to the investigation of the development strategy of the condition monitoring scheme, the central idea of utilization of these data is: the baseline datasets are collected to build the anomaly detection model representing the normal operating condition, while the defect test datasets are evaluated against the established model to detect anomalies as well as diagnosing detected anomalies. As Table 3.5 indicates, baseline tests 9, 11, 12 and 19 are used to construct the anomaly detection model, while the remainder of the data are evaluated against the model as the testing data. It should be noted that although all the defined testing data have been evaluated against the model, five cases are chosen to present in this thesis based on their different damaged conditions. For example, test 20 data [3] contains the insignificant precursor information; the anomaly detection and diagnostic results in this thesis will show how the prognostic window is significantly improved relative to the original feature plots. Another test (test 16) [2] will also be implemented to demonstrate the capability and benefits of the proposed scheme that the prognostic information can be extracted and enhanced.

Figure 3.8 illustrates the input pattern matrix for each of the 14 bearing datasets, and it should be noted that the values of the matrices have been normalised and scaled, and suitable for being the input of modelling. In each matrix, extracted 27 features at selected sampling periods are presented. Take test 20 for example, feature values in the running-in period (5 hours) are relatively high, while the values in the steady state (10-50 hours) become steady and low. However, when the bearing was entering the wear-out stage (60 hours), values of the features increases dramatically. These discovered characteristics in the input pattern matrix are expected to be analysed and found in intelligent processing stage.

3.5 References

- [1] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing* 18 (2004) 199-221.
- [2] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Electrostatic wear monitoring of rolling element bearings. *Wear* 263 (2007) 1492-1501.
- [3] M. Craig, S.L. Chen, T.J. Harvey, R.J.K. Wood, K. Masuda, M. Kwabata, H.E.G. Powrie. Advanced condition monitoring of tapered roller bearings part I, with multiple sensing techniques. Accepted by Tribology International, 2008.
- [4] <http://www.macomtech.net>
- [5] B.J. Roylance, T.M. Hunt, *Wear debris analysis*. Coxmoore Publishing Company First Edition 14. 1999.
- [6] T.M. Hunt, *Handbook of wear debris analysis and particle detection in liquids*. London and New York: Elsevier Applied Science. 51. 1993.

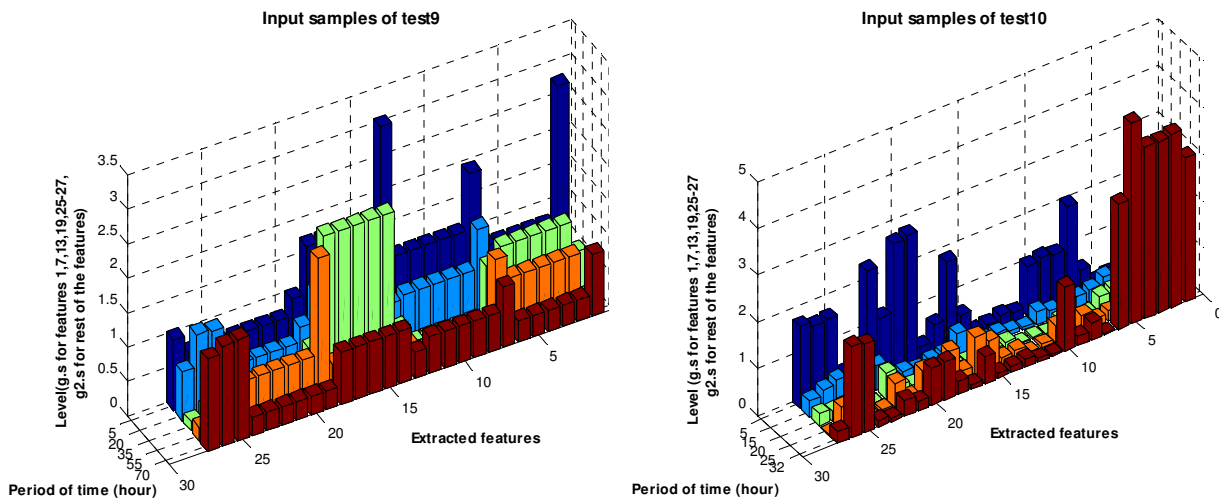


Figure 3.8 Data matrix of the 14 sets of bearing data

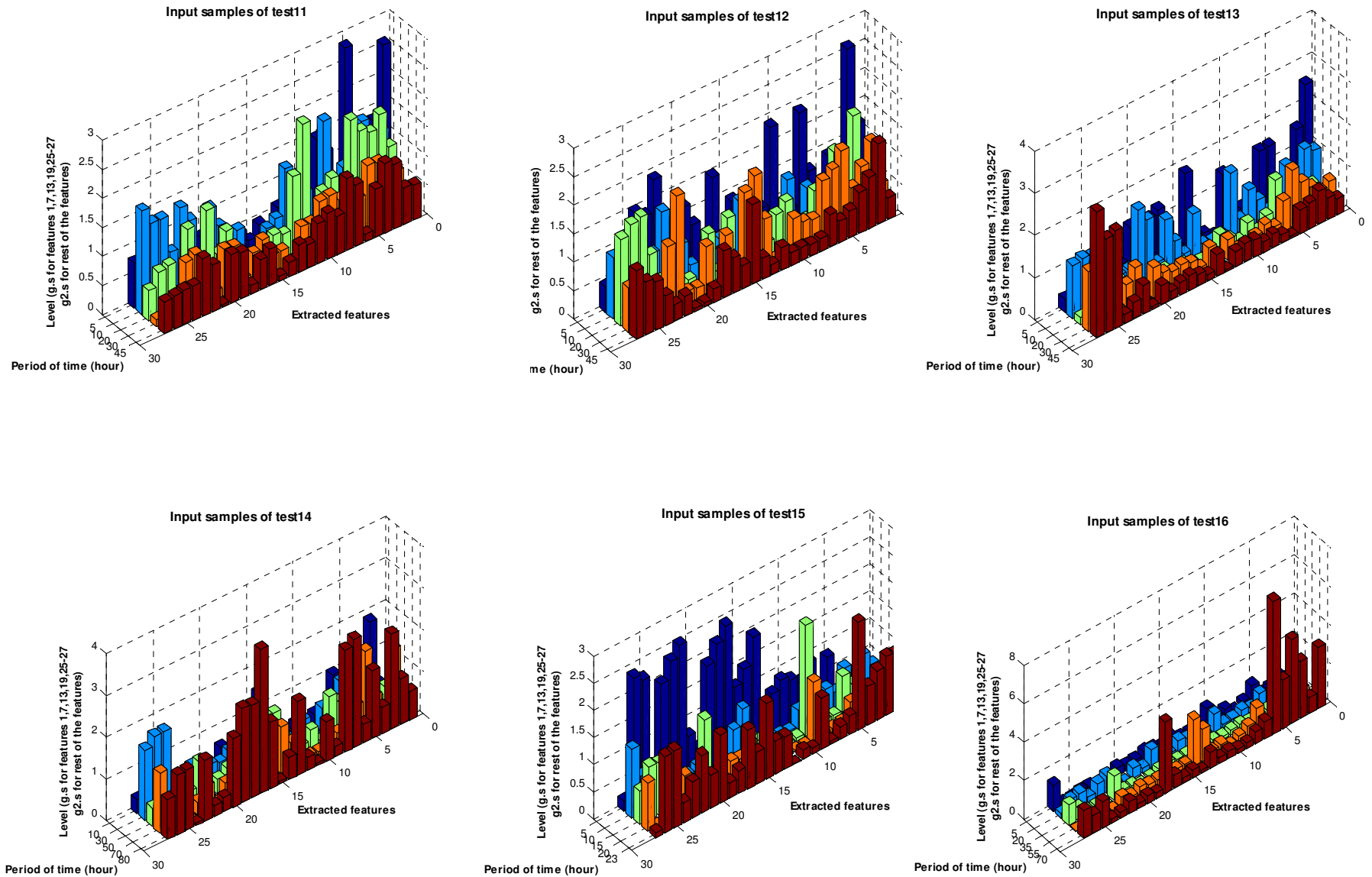


Figure 3.8 Data matrix of the 14 sets of bearing data (cont.)

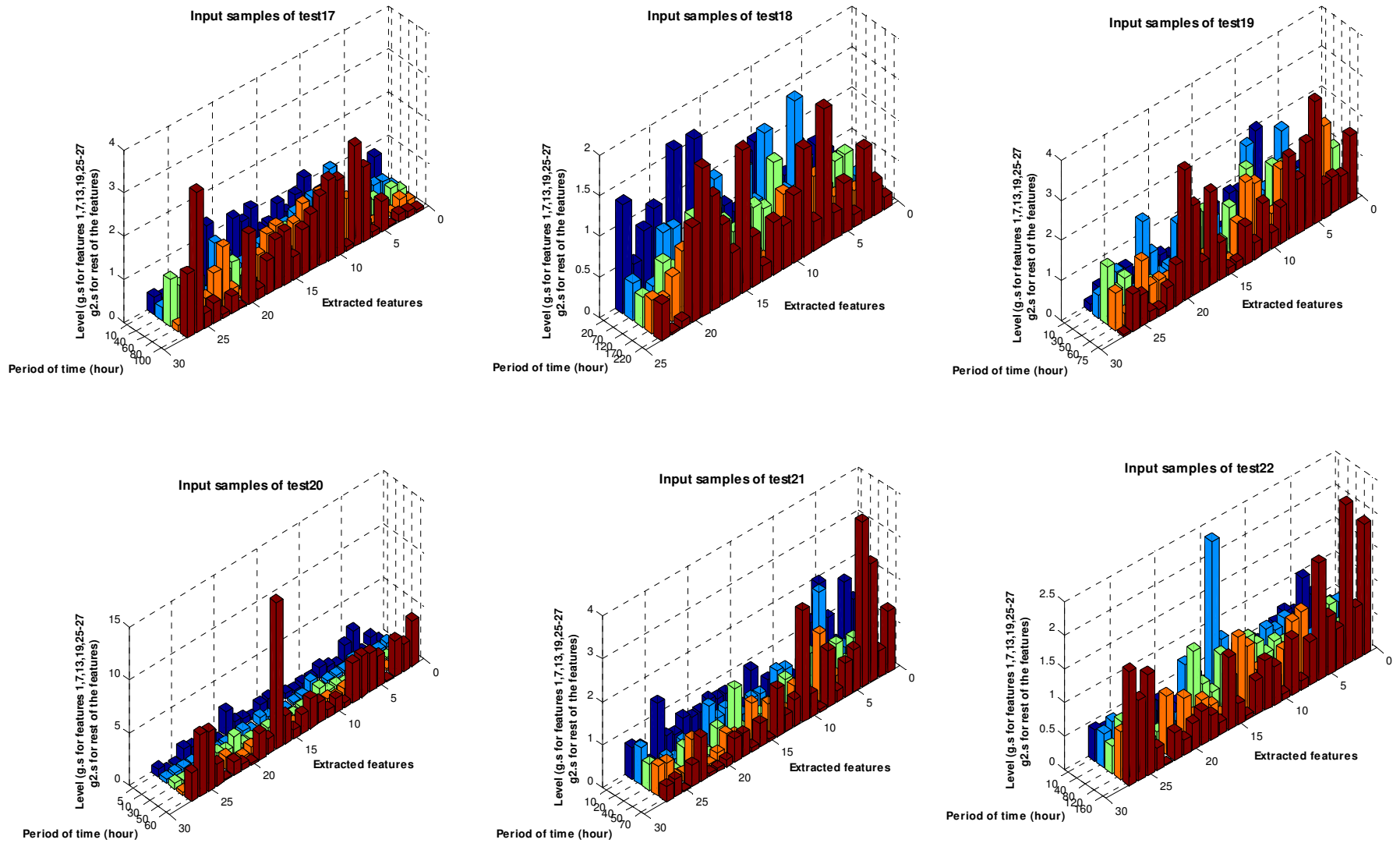


Figure 3.8 Data matrix of the 14 sets of bearing data (cont.)

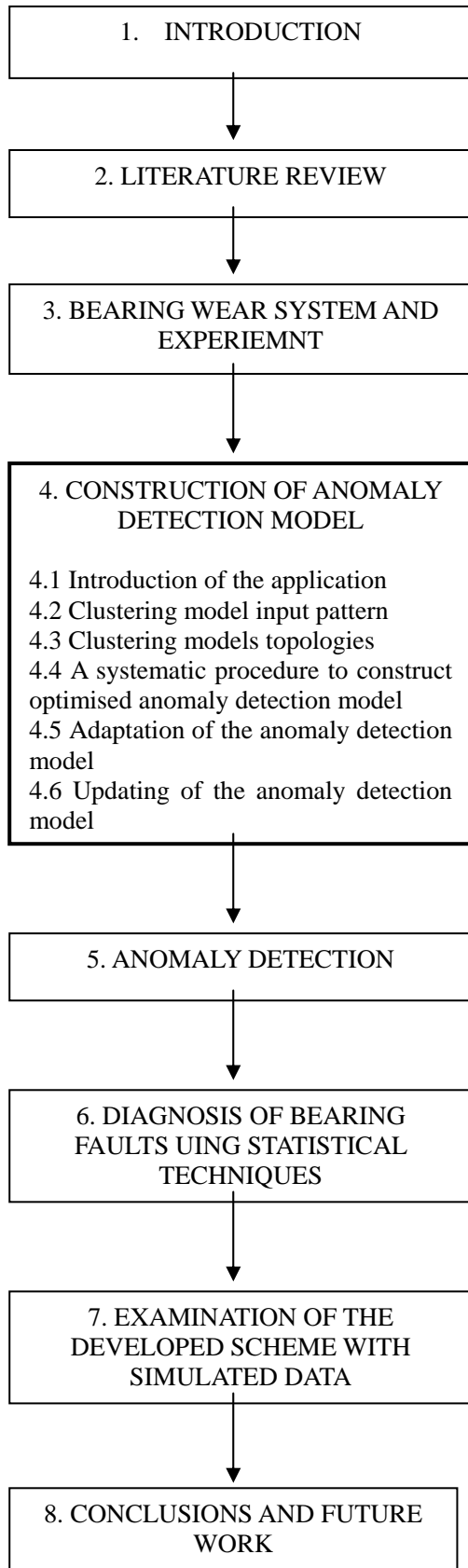
Test Number	Test Load / kN(max.)	Load changes	Test Duration / hours	Defect Bearing #	Speed/rpm
9	15	Increased to 15 kN at 4 hour 25 min.	72	None (baseline)	2500
10	15	Increased to 15 kN at 4 hour 38 min.	34	2	2500
11	15	Increased to 15 kN at 4 hour 15 min.	48	None (baseline)	2500
12	15	Increased to 15 kN at 4 hour 25 min.	48	None (baseline)	2500
13	15	Increased to 15 kN at 4 hour 00 min.	48	2	2500
14	20	Increased to 15 kN at 4 hour 00 min. Increased to 20 kN at 71 hour 8 min.	85	3	2500
15	20	Increased to 15 kN at 2 hour 49 min. Increased to 17.5 kN at 3 hour 36 min. Increased to 20 kN at 4 hour 5 min.	24	2	2500
16	20	Increased to 15 kN at 4 hour 34 min. Increased to 17.5 kN at 24 hour 5 min. Increased to 20 kN at 48 hour 15 min.	72	2	2500
17	17.5	Increased to 15 kN at 4 hour 8 min. Increased to 17.5 kN at 30 hour 00 min.	105	2	2500
18	17.5	Increased to 15 kN at 4 hour 10 min. Increased to 17.5kN at 27 hour 5 min.	233	2	2500
19	20	Increased to 20 kN at 4 hour	80	None (baseline)	2500
20	20	Increased to 20 kN at 4 hour	62	2	2500
21	20	Increased to 20 kN at 4 hour	76	2	2500
22	20	Increased to 20 kN at 4 hour	165	2	2500

Test Number	Available features	Damage (Comments)	Weight loss (races)	Application	Date
9	6 Vib,18 WSS, 3 OLS	None	2.1 mg	Training data	Jun-02
10	6 Vib,18 WSS, 3 OLS	Major damage to inner and outer races of bearing 2	304.9 mg	Anomaly detection & diagnosis	Jul-02
11	6 Vib,18 WSS, 3 OLS	None	0.95 mg	Training data	Aug-02
12	6 Vib,18 WSS, 3 OLS	None	0.93 mg	Training data	Sep-02
13	6 Vib,18 WSS, 3 OLS	None (Debris recovered 13 mg)	13 mg	Anomaly detection & diagnosis	Sep-02
14	6 Vib,18 WSS, 3 OLS	Damage near defect on the inner race of bearing 3, damage to shaft from bearing rotation	11.3 mg	Anomaly detection & diagnosis	Oct-02
15	6 Vib,18 WSS, 3 OLS	None, but the motor was found damaged at later period of running, causing increasing level of vibration	3 mg	Anomaly detection & diagnosis	Oct-02
16	6 Vib,18 WSS, 3 OLS	Major damage to inner race of bearing 2	1033.8 mg	Anomaly detection & diagnosis	Dec-02
17	6 Vib,18 WSS, 3 OLS	Major damage to inner race of bearing 2	61.19 mg	Anomaly detection & diagnosis	Feb-03
18	6 Vib,18 WSS, 3 OLS	damage to inner race of bearing 2, but lots of damage to one roller of bearing 2	149.03 mg	Anomaly detection & diagnosis	May-03
19	6 Vib,18 WSS, 3 OLS	None	0.3716 mg	Training data	Mar-07
20	6 Vib,18 WSS, 3 OLS	Major damage to outer race and rollers of bearing 4	255.79 mg	Anomaly detection & diagnosis	Jan-07
21	6 Vib,18 WSS, 3 OLS	Small spalls to the inner race of bearing 2	23 mg	Anomaly detection & diagnosis	July-08
22	6 Vib,18 WSS, 3 OLS	Small spalls to the inner race of bearing 2	19 mg	Anomaly detection & diagnosis	Oct-08

Table 3.5 Datasets utilized within the thesis

THESIS STRUCTURE

KEY OUTCOME



A series of new approaches have been developed for model establishment to achieve robust fault detection and diagnosis, particularly:

1. A systematic procedure has been developed to build optimised models by evaluating number of training parameters, such as Number of Candidate models (4.4.3), Initial Iterations (4.4.4) and shape of the Covariance structure (4.4.5). The designed evaluation indices have shown that the quality of the generated models can be significantly improved by justifying these parameters.
2. The innovative methods for model adaptation have been developed. The fundamental methods are developed based on the information theory and distance measure (4.5.1), and these approaches have also been optimised to achieve more reasonable results (4.5.3).
3. A new procedure has been designed to select appropriate datasets to update anomaly detection model by investigating two originally developed parameters.

Chapter 4. Construction of Anomaly Detection Model

In order to carry out automatic and data-driven anomaly detection, before examination of any bearing failure tests, anomaly detection models need to be established with the baseline data which represents the healthy condition of the bearings. The following sections will describe related topics about model construction, including data status (normal or abnormal) identification, topology, construction of optimised models, adaptation of anomaly detection model and model updating and merging.

4.1. Introduction to the application of mixture model based clustering in tapered roller bearing condition monitoring

Generally, the mixture model based clustering approach is used to construct an anomaly detection model in the proposed taper roller bearing condition monitoring scheme. In practice, using clustering models to build a normal model for bearing fault detection and diagnosis will require the establishment of the baseline database representing the healthy condition of the bearings. To fulfil this requirement, the laboratory bearing rig was running with a series of perfect bearings to obtain baseline data, and then this was followed with the further processing steps to get the optimised model ready for fault detection and diagnosis. After model construction, testing data can be tested against the model to see whether they are healthy or not. Concept and example of the anomaly detection can be found in section 2.4.2 as a reference.

The general steps of the clustering approach to construct anomaly detection models consist of:

- *Signal processing to extract features, and identify training and testing datasets.*
- *Use training data to construct optimised models by evaluating a set of cluster training parameters.*
- *Adopt novel approaches to identify anomalies in the training data and their associated clustering sub-spaces. This is step is very important, since the fault masking effect in the anomaly detection can be eliminated through this process.*

- *To ensure the established models are consistently robust, the appropriate testing data points also need to be selected to update the existing models.*
- *The adapted and updated models are ready for the applications.*

This procedure is illustrated in Figure 4.1. It is seen that normalised training data is used in step 1 to build stabled original normal model with a unique systematic approach (approach will be introduced in section 4.4), and then the original model will be checked in step 2 with the novel methods (Series of new methods will be explained in section 4.5) to ensure that the anomalies within the normal model can be substantially removed. Thirdly, appropriate testing data will be constantly selected for updating model; this procedure is conducted in steps 3 and 4, and will be explained with details in section 4.6.

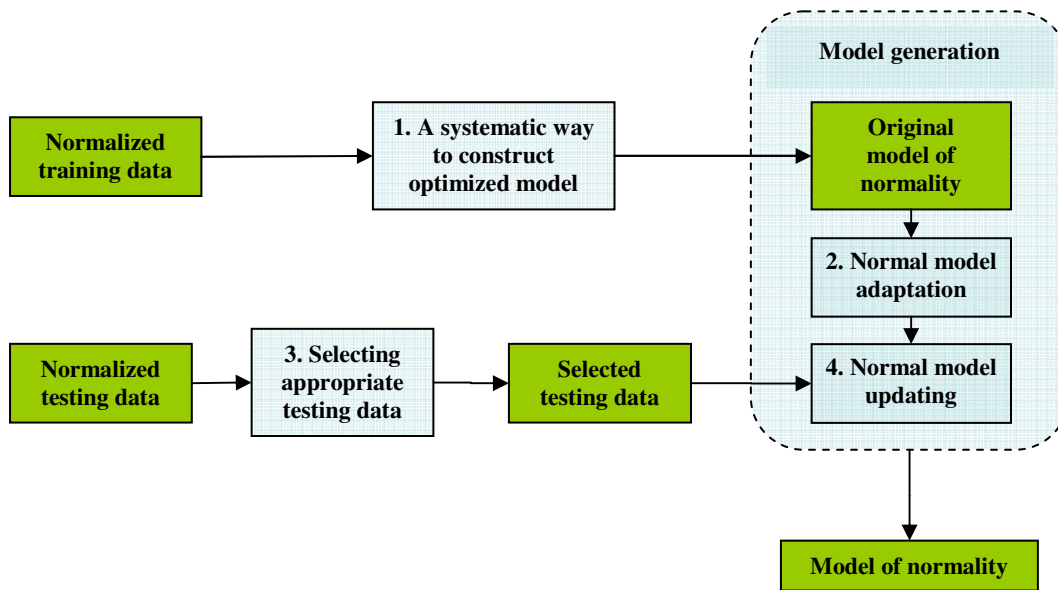


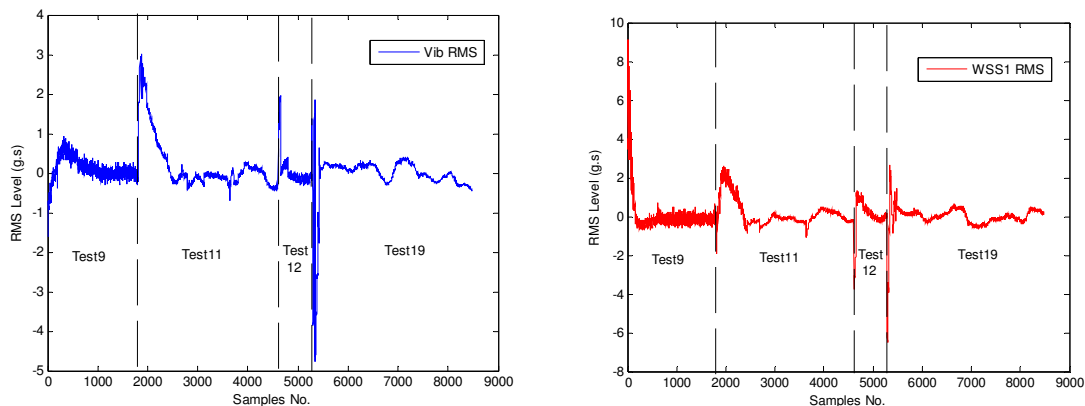
Figure 4.1 The procedure of using the clustering approach to construct the anomaly detection model

For the model optimisation, the quality of the obtained models with different training parameters is not only assessed by several evaluated indices including BIC score and similarity measure, but also by the model adaptation and anomaly detection rates. For the model adaptation, the developed novel approaches are firstly demonstrated with the two selected dimensional bearing data which contain typical patterns of anomalies and are then followed by the application of all the bearing baseline data. For the model updating,

a set of testing data is employed to demonstrate the method of how to select appropriate testing points to compare with the normal data and update existed models. The applications of the approaches for the anomaly detection will be presented in Chapter 5.

4.2. Clustering model input patterns

Figure 4.2 shows example traces of processed RMS values from 6 sensors from the 4 baseline tests. From these figures, it is seen that amount of abnormal trend data are found in the features, and these abnormal data are expected to have an impact on the anomaly detection causing fault masking phenomenon, and crucial to be eliminated from the training data. There are 8,476 data points for each of the 27 extracted features, hence a total of 228,852 data points are applied to train the anomaly detection models. At the current stage, a histogram is used to help define the state (i.e. normal or abnormal) of the baseline data in order to conduct model adaptation rate (Anomalies correctly identified in the training data/Total anomalies in the training data) calculations in the later stages. Figure 4.3 illustrates the histogram example of the vibration RMS value. From this plot, the data which are distributed with very low densities are defined as the anomalies, and the two red lines are the threshold values which are set by the mean value of the normal fit plus one standard deviation. The same selection criterion is used for the other 21 features. The reason for choosing the one standard deviation as the reference value (not the 1.5σ , 2σ or 3σ) is that the other values are too conservative so that more obvious anomalies cannot be identified, and the effect of this set is worthwhile to be evaluated in the future study.



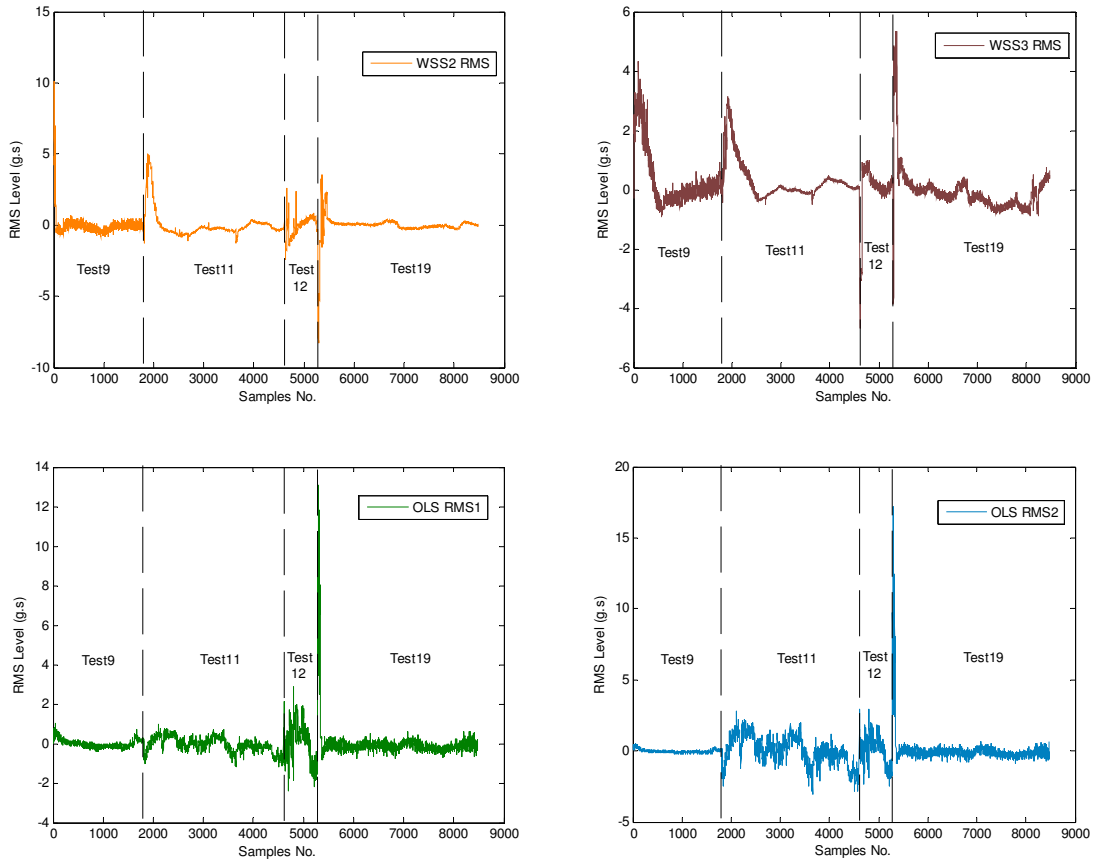


Figure 4.2 RMS values of the 6 sensors for the baseline data

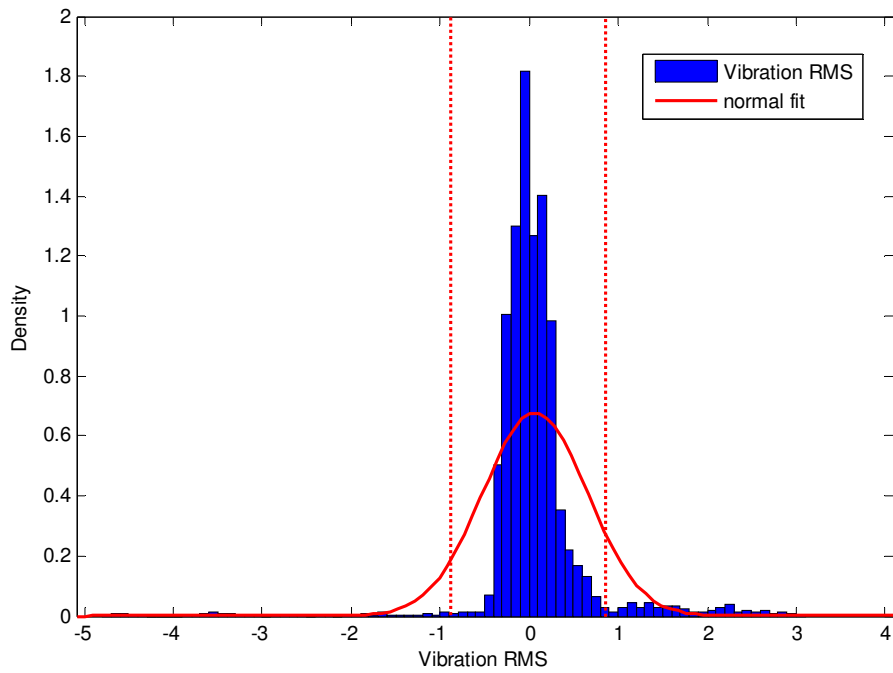


Figure 4.3 Histogram of the vibration RMS values of the baseline data

Pre data identification	Total number	Normal data	Abnormal data
Total data	8476	7923	553
Test 9	1817	1782	35
Test 11	2786	2438	348
Test 12	681	647	34
Test 19	3192	3056	36

Table 4.1 Pre identification for the baseline data

Table 4.1 shows the ‘normal/abnormal’ data distribution as judged by the histogram technique. The model adaptation rate will be calculated based on anomalies identified from the above.

4.3. Clustering Model Topologies

Clustering is a method that intends to group observations into clusters so that observations within the same cluster are more closed to one another compared with observations belonging to other clusters [1]. As a type of clustering method, a mixture-model based clustering technique has attracted numerous attentions. This technique is developed based on probabilistic models where all observations can have a degree of membership to all the assigned clusters. And these observations assume that a mixture of underlining probability distributions has generated them. In a mixture model, a cluster is regarded as a component that is represented by a type of probability distribution (e.g. Gaussian and exponential). The association of an observation with a cluster is evaluated as a probability of membership.

As the Gaussian distribution was selected as the underling probabilistic model in the current study, the parameters of mixture model consists of the mean, covariance and mixing coefficients for each cluster. To estimate these parameters, Dempster *et al.* [2] introduced the Expectation-Maximization (EM) algorithm and demonstrated its application to a range of problems. EM can be applied to clustering, and the mixture

model starts with a random partitioning of the training data and usually assumes an equal mixing coefficient (prior probability) for each cluster. The parameters for each cluster are estimated. The points are then re-assigned following which the parameters are re-estimated. This process continues until convergence. Details of the EM algorithm is presented below.

Each Gaussian density, k , has a mean vector μ_k , and covariance matrix Σ_k . The rotation of each component can be varied by placing restrictions on the form of Σ_k . For example, if Σ_k is restricted to a diagonal form (all off-diagonal entries are zero) the component principal axes will not be rotated; otherwise, the axes will be rotated according to off-diagonal entries. The form of $f_k(x_i | \mu_k, \Sigma_k)$ is used to denote the density of observation i from component k :

$$f_k(x_i | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)\right\} \quad (4.1)$$

Where p is the number of variables (the dimension of each observation's description). Each cluster contributes to the density of an observation. A cluster's contribution is weighted by a mixing coefficient $P(k)$, and the unconditional probability function is:

$$f(x_i) = \sum_{k=1}^K P(k) f_k(x_i | \mu_k, \Sigma_k) \quad (4.2)$$

Note that each mixing coefficient must have a value between 0 and 1 and the coefficient must be equal to 1.

$P(k | i)$ is regarded as the probability of clustering component k . Given an observation i , the posterior probability for this component is defined with the Bayesian rule, and indicated as:

$$P(k | i) = \frac{f(x_i | \mu_k, \Sigma_k) P(k)}{f(x_i)} \quad (4.3)$$

The EM algorithm is applied to the problem of parameter estimation. In particular, the parameters, μ_k , Σ_k , and $P(k)$ of the mixture Gaussian density function need to be adjusted by the EM algorithm. To achieve this, two steps are usually iterated. The E-step (Expectation) calculates $\hat{P}(k|i)$ which acts as the conditional expectation of posterior component probability $P(k|i)$, with the given observation i and the latest estimates for μ_k and Σ_k . The M (Maximization) step updates the parameters μ_k and Σ_k by replacing $P(k|i)$ with its updated estimate $\hat{P}(k|i)$. With N training data points, the equations for computing the next estimated parameters are:

E-step:

$$\hat{P}(k|i) = \frac{f(x_i | \hat{\mu}_k, \hat{\Sigma}_k) \hat{P}(k)}{f(x_i)} \quad (4.4)$$

M-step:

$$\hat{\mu}_k = \frac{\sum_{i=1}^N \hat{P}(k|x_i) x_i}{\sum_{i=1}^N \hat{P}(k|x_i)} \quad (4.5)$$

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^N \hat{P}(k|x_i) [x_i - \hat{\mu}_k]^T [x_i - \hat{\mu}_k]}{\sum_{i=1}^N \hat{P}(k|x_i)} \quad (4.6)$$

$$\hat{P}(k|x) = \frac{1}{N} \sum_{i=1}^N \hat{P}(k|x_i) \quad (4.7)$$

The solution of maximum log-likelihood function, L , is obtained by iterating the E and M steps until each parameter has converged to within a tolerance criterion.

$$L(\theta, P) = \sum_{i=1}^N \sum_{k=1}^K \log P(k) f(x_i | k) \quad (4.8)$$

Where

$$\theta = (\mu_1, \dots, \mu_k, \Sigma_1, \dots, \Sigma_k, P_1, \dots, P_k)$$

The method of determining the number of clusters typically requires multiple runs of the EM with various number of clusters. An information index for each run is evaluated to determine the most suitable number of clusters. One such indication is the Bayesian

Information Criterion (BIC). In this index, Bayesian approach is utilized to select the best model (with each model having a different number of clusters). The posterior probability for each model conditioned on the training data is estimated and the model with the highest probability is considered to be the best model. So the BIC is a rule that adds a penalty factor to the log-likelihood. This penalty looks at the number of parameters used in the cluster. The BIC is usually written in a form whereby the lower the BIC the better the model. The rule is [3]:

$$\text{BIC} = -2 L(\theta, P) + V \quad (4.9)$$

Where L is log likelihood and $V = \ln(\text{number of cases}) * \text{Num of parameters}$.

4.4. A systematic procedure to construct an optimised anomaly detection model

The training mixture model based clustering algorithm is a sophisticated and relatively random process, especially when the patterns to be trained are distributed in a complicated manner. Therefore, finding the clustering model with the best performance is an important but difficult task. So far, most of the current research and applications ignore this fact, and choose the clustering training parameters randomly because the selected training parameters could lead to best performance of the model or it might be within the range of the best ones. However, this is not good enough, since the effect of randomness will not only lead to uncertainty in the training results but also the anomaly detection results that follow. The study in this section would like to demonstrate that a systematic way of finding the best performance models from a large number of possible structures with different combinations of training parameters is needed, because the performance of a model depends on several parameters in the mixture model based clustering topology with the EM algorithm, such as:

- *Number of clusters in a clustering model*
- *Covariance structure*
- *Maximum iteration*
- *Stopping tolerance*

However, it is widely known that the EM algorithm involves some randomness, and can cause inconsistent solutions from different initial conditions; these varied solutions at clustering may produce quite different results and directly affect anomaly detection. Therefore, the challenge becomes how can the practitioners reduce such randomness, and obtain a consistent optimized model indicated by the BIC. Therefore, the goal in this study is to identify a method that would give the highest likelihood in a fixed number of iterations, and provide consistent solutions. The key issue is finding the trade-off between adequately searching the parameter space with initial positions versus iterating the EM algorithm sufficiently to get close to the maximum [4]. Given a total iteration constraint, in [4], a three step search/run/select (S/R/S) strategy for maximizing the likelihood considered:

1. Building a search method for generating p initial positions. This could be based on random starts or short runs of the standard EM algorithm.
2. Running the EM algorithm a set number of times at each initial position with a fixed number of initial iterations.
3. Selecting the solution that provides the highest likelihood among the p trials for further iteration.

Based on the above principle, in our approach, a set of candidate models $p(x; \Theta)_{can}$ at different initiation points are generated for each clustering attempt (with different number of clusters $k= 1, \dots, m$) before the best model (within the candidate models) with the highest likelihood L_{max} is selected for further EM iteration. And then, the same procedure with unchanged training parameters (i.e. number of candidate models, initial iteration for candidate models) needs to be repeated several times to examine whether the optimum models (indicated by BIC) show more stable pictures. Otherwise, the number of the candidate model n_c could be increased until a more stable optimized model is obtained. Principally, increasing the number of the candidate model can reduce the opportunity of missing the best clustering solution, but this also requires more time expenditure. The flowchart in Appendix A1 shows the general procedure of the approach.

Based on the above issues that should be considered during the design of a clustering training programme for the bearing condition monitoring problem, a systematic way of finding the best performance clustering model from a large number of possible clustering models is proposed in this study. Since one of the purposes of this study is to achieve stable models, the evaluated training parameters in this study include:

- *Number of clusters in a clustering model*
- *Number of candidate models for each clustering attempt*
- *Initial iteration for each candidate model*
- *Structure of the covariance matrix*

It should be noted that due to limited time, the evaluations of other two common training parameters, *maximum iterations* and *stopping tolerance*, are ignored in this training programme and set with the fixed values of 250 and 0.0001 respectively.

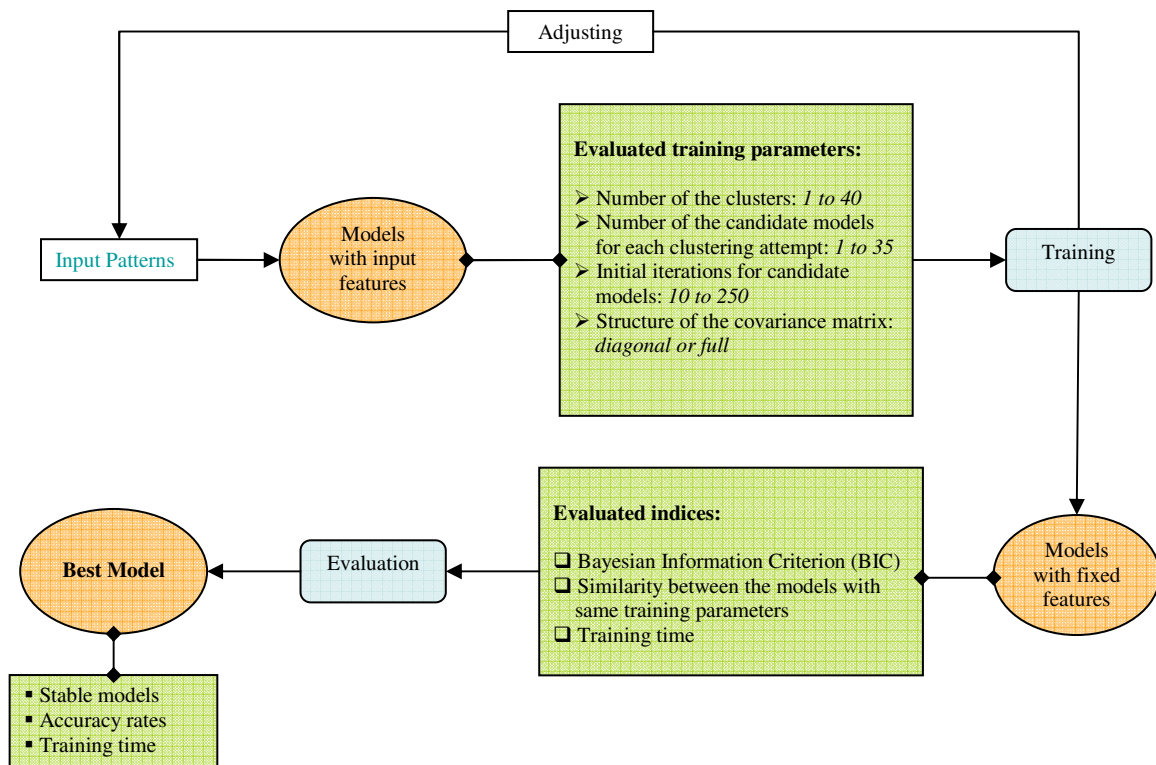


Figure 4.4 The programme for finding the best performance of the clustering model

The whole training procedure is illustrated in Figure 4.4. The loop starts with the definition of the input for the clustering models, then the variables of the clustering topology are required. For the proposed clustering models, the parameters include the number of clusters which runs from 1 up to 40, the number of candidate models for each clustering attempt which is between 1 and 35, the initial iterations for candidate models which is greater than 10 and smaller than 250, and the covariance matrices with diagonal and full structures. It should be noted that selected ranges for these parameters were set based on the initial examination and modelling; in these selected ranges, changes of the modelling results with different parameters are found significant, and therefore is more reasonable to demonstrate the importance of justifying the training parameters. Once these parameters are determined, the clustering training can be initialised, and after the model is trained, the programme should go back to the starting point in the loop. In the meantime, the model being trained is examined by the evaluated indices. The BIC score is used to select the optimum model from the series of models with varied 1 to 40 clusters. The similarity is then used to compare the differences between clustering attempts with the same training parameters. The accuracy rates of the model adaptation and anomaly detection are also important indices to inspect the quality of the models. Furthermore, the training time is also recorded for reference purposes.

4.4.1. Model training programme and evaluation

4.4.1.1. Model training scheme and matrix

Features from the baseline tests 9,11,12 and 19 were used to establish the initial model of normality, and in this case, a clustering algorithm is running with different numbers of clusters ($k= 1\dots40$), while the BIC score for each clustering attempt is plotted to identify the optimized model (with the lowest BIC value). This procedure is called multiple clustering in this study for convenience. On the other hand, the number of candidate models n_c for each clustering attempt within the multiple clustering process is set between 1 and 35, and the initial iteration for each candidate model is between 10 and 250. Figure 4.5 illustrates the concept of the model training programme. With each

number of candidate models and its associated initial iteration, the multiple clustering repeated 3 times, and tables 4.2 and 4.3 show the model training programme with the diagonal and full covariance structures respectively. Take Table 4.2 for example, with diagonal covariance structure, for each fixed initial iteration, the parameter of number of candidate models is modified between 1 and 35, and with each parameters combination (fixed initial iteration and one of the changed number of candidate models), the multiple clustering repeats 3 times. Similarly, for the next evaluated initial iteration value, the procedure follows the instruction described above. Having finished the diagonal-structured models, evaluation of the full-structured models as the Table 4.3 shows can be completed in the same way. Once the model training is finished, the result of different training can be plotted against the training parameter, and the changes can be conveniently observed.

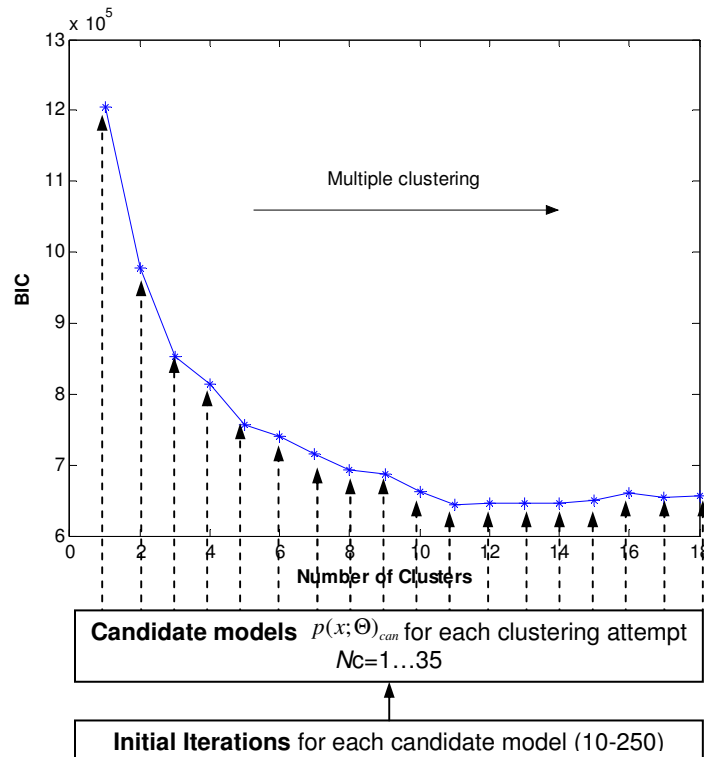


Figure 4.5 Concept of the training scheme

4.4.1.2. Evaluation indices

The quality of the trained clustering models is initially evaluated by the BIC score, which shows how good the trained model fits the data; the clustering with the lowest BIC score should be selected within each multiple clustering process. Moreover, since each multiple clustering process with the same training parameters is repeated three times, to see whether the obtained optimum models look consistent and stable with the increase of the number of candidate models and their associated initial iterations, another evaluation index of similarity between the optimum models with the same training parameters is also implemented.

The similarity measure should depend on the mean value and covariance matrix of each cluster that determines the degree of sharing of each point among all clusters. The similarity measure between clusters i and j can be defined as [5]:

$$\text{SIM}_{ij} = 1 - \frac{d_c(G_i, G_j)}{\left[\sum_{k=1}^n (\max_{l=1}^m \{x_{lk}\} - \min_{l=1}^m \{x_{lk}\})^2 \right]^{1/2}} \quad (4.10)$$

Where the denominator term in the equation on the right hand side denotes the Euclidean distance of the minimum and maximum values of each input vector on all the data for the operating region, and the nominator term refers to a metric d_c for clusters G_i and G_j , which is defined as:

$$d_c^2(G_i, G_j) = \|\mu_i - \mu_j\|^2 + \text{tr}((C_i - C_j)^T (C_i - C_j)) \quad (4.11)$$

Where the first term of the equation on the right hand side is the Euclidean distance as the dissimilarity measure between two cluster centres μ_i and μ_j . Since similarity cannot be determined only based on the distances between cluster centres, the second term is used to explain the size difference between the two clusters (C represents the covariance matrix of the cluster). The notation ‘tr’ represents the trace.

Having confirmed the similarity calculation between the two clusters, the similarity between the two clustering models can be defined using equations 4.10 and 4.11. As the proposed similarity is examined among the 3 clustering attempts which are with the same training parameters, at the beginning, each cluster in one of the clustering attempts n is compared with clusters in the other two attempts, and then the maximum similarity value is selected for each evaluated cluster in attempt n , followed by a summing up of these maximum similarity values, and finally the number of clusters in the n^{th} clustering attempt is divided to ensure the values fall into $0 < \text{SIM} \leq 1$. Equation 4.12 shows the normalized similarity between the clustering models which is repeated for the other two clustering attempts to complete the whole comparison.

$$\begin{aligned} \text{Normalised SIM (between } n^{\text{th}} \text{ clustering attempt and other two)} = & \\ \frac{\text{SIM}(1,n)_{\text{max}} + \dots + \text{SIM}(k,n)_{\text{max}}}{\text{Number of clusters in } n^{\text{th}} \text{ clustering attempt}} & \quad (4.12) \end{aligned}$$

And $\text{SIM}(k, n)$ represents the similarity between cluster k in n^{th} clustering attempt and the clusters of the other two clustering models.

Apart from the quality (BIC score) and similarity assessment, the performance of the clustering models is also inspected by calculating their model adaptation and anomaly detection rates. The successful detection rate of clustering models is very important for all fault detection problems. The higher the accuracy rate that is achieved, the better the model performs. 100 percent accuracy rate is desired at all occasions. However, it is not always achievable in practice. The accuracy rate of the clustering model depends very much on the nature of the problem, the available inputs for the model training and the quality of the trained and adapted models. As the complexity of the problem increases, it becomes more difficult for the models to train their input patterns.

The accuracy rate of anomaly detection (true positive) [6] is defined as the percentage of correctly detected anomalies in the total defined anomalies, which follows the equation 4.13. On the other hand, the false alarm (false positive) rate is the percentage of the normal patterns incorrectly detected as the anomalies in the total defined normal data. The false alarm rate is calculated as equation 4.14 below.

$$tp(\%) = \frac{\text{Anomalies correctly detected}}{\text{Total anomalies}} \quad (4.13)$$

$$fa(\%) = \frac{\text{Normalities incorrectly detected as anomalies}}{\text{Total normalities}} \quad (4.14)$$

4.4.2. Examples of the training results

Figure 4.6 (a) shows the BIC plot of the 3 repeated multiple clustering (varied number of clusters from 1 to 40) with 10 candidate models. From this plot, the optimum models for each repeated multiple clustering could be selected with the number of clusters 30, 29 and 29 respectively, and illustrated as the figures 4.6 (b), (c) and (d) show. From these figures, it is seen that the produced solutions are quite different from each other, especially for the generated clusters which are found to be widespread with low support cases. The similar unstable solutions are also found in the clustering results with 15 candidate models, shown in figures 4.6 (f), (g) and (h), which show that the effect of randomness is still strong, and leads to difficulty in choosing an appropriate model for anomaly detection. However, as the number of candidate models increases to 20, the optimum solutions to the repeated 3 clustering processes become more consistent and stable. As the figures 4.6 (j), (k) and (l) show, data which are distributed very widely are almost grouped by the clusters with a similar shape in the 3 repeated clustering processes, although the clusters' mean and covariance values are slightly different from each other. Therefore, the anomaly detection model is chosen from these 3 consistent solutions with the lowest BIC score.

Figures 4.7 (a)-(l) illustrate the diagonal clustering models with 10 (4.6(a)-(d)), 15 (4.6(e)-(h)) and 20 (4.6(i)-(l)) candidate models. It is seen that the diagonal models have a similar trend with the full covariance models: the repeated models are not stable when the low number of candidate models are applied, while the models become more reliable and consistent when the number of candidate models is increased to 20.

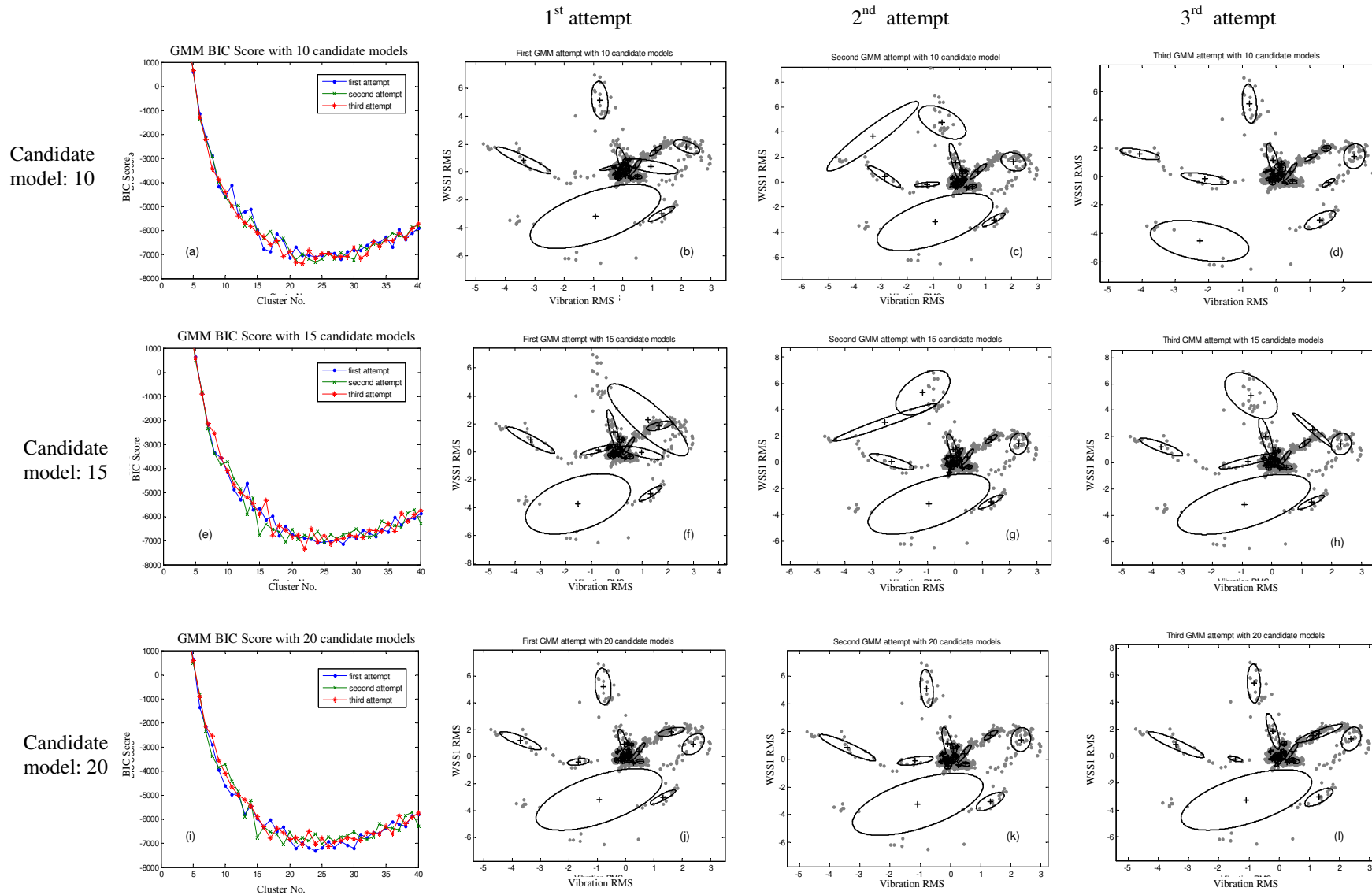


Figure 4.6 Gaussian mixture models with different candidate models under the full structure

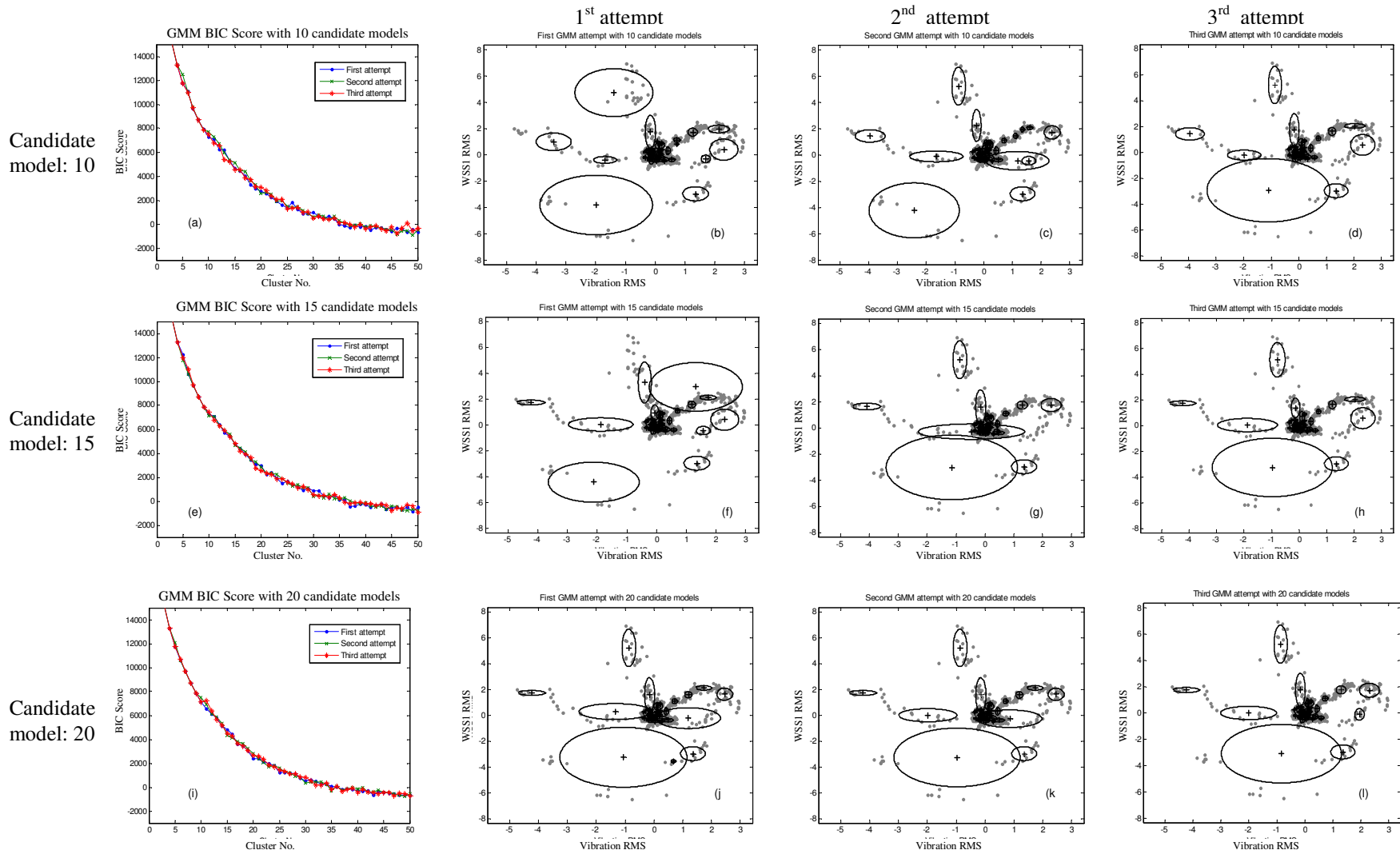


Figure 4.7 Gaussian mixture models with different candidate models under the diagonal structure

4.4.3. Evaluation of the number of candidate models

Training time and BIC score are the two important outputs from the model training. Initially, the models with good quality were chosen based on these two results. The lower the BIC value reaches and the shorter the training time, the better the model is. In this section, these two parameters are plotted against one of the evaluated parameters, the number of the candidate models, to see whether the quality of the obtained models can be improved with the increase of the candidate models for each clustering attempt. However, this seems insufficient, if the stability and randomness of the repeated multiple clustering also needs to be evaluated through calculating the index of similarity between the optimum models, which are selected from the repeated 3 multiple clustering processes with the same training parameters. Hence, the similarity is also plotted and assessed against the varied number of candidate models. Table 4.4 summarises the evaluated parameter- number of candidate models and its evaluation indices in section 4.4.3.

Evaluated parameter	Evaluation index
Number of candidate models (NCM)	BIC score (Section 4.4.3.1)
	Training time (Section 4.4.3.1)
	Similarity between models (Section 4.4.3.2)

Table 4.4 Evaluated parameter and its evaluation index

4.4.3.1. Lowest BIC score and training time

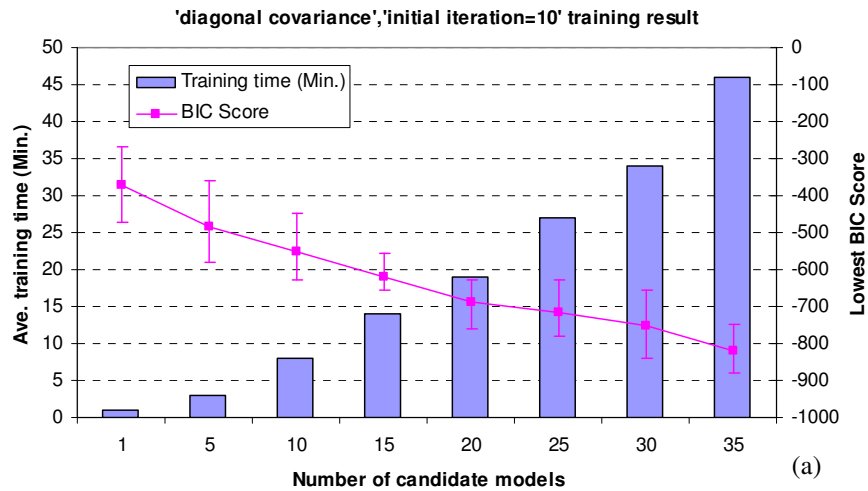
Figures 4.8 and 4.9 illustrate the lowest BIC scores of the optimum clustering models and their training time under the diagonal and full covariance structure respectively. In each figure, there are two lines. One shows the change of the lowest BIC score versus the number of candidate models (pink), and the other one shows the training time change with the number of candidate models (blue).

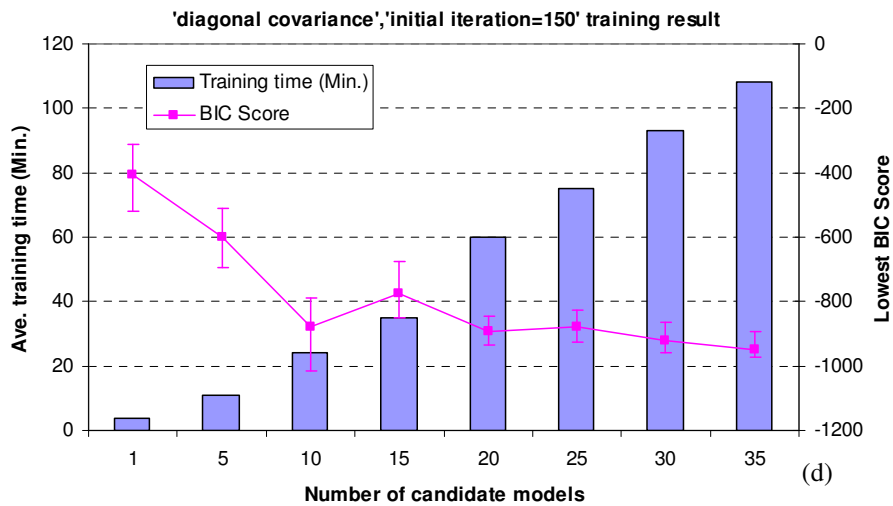
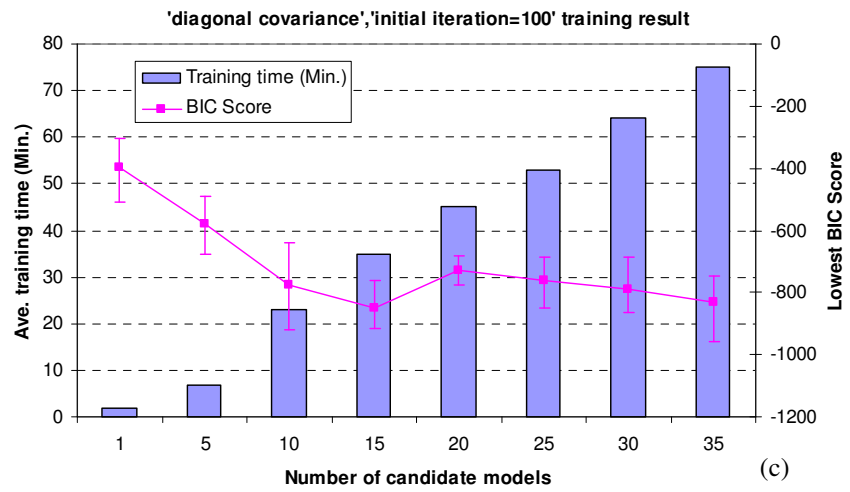
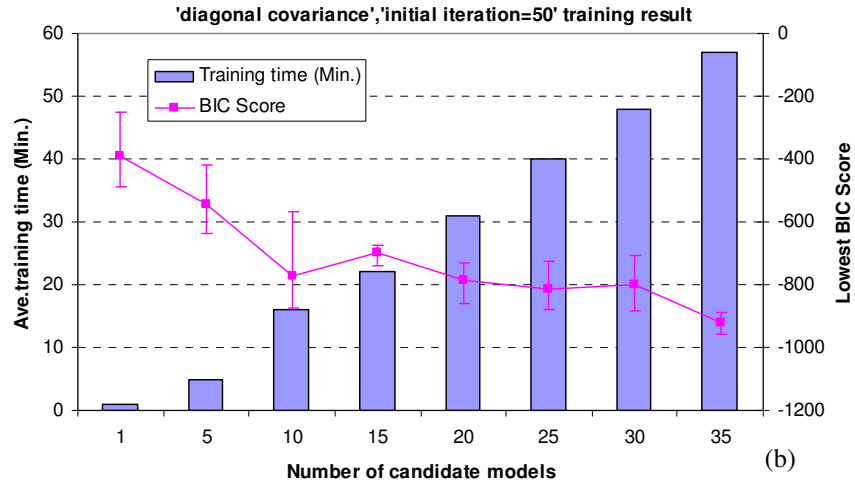
(1) Diagonal covariance structure

From the relationships shown in Figure 4.8, for almost all the cases with the same diagonal covariance structure and different initial iterations (10-250), the clustering training time tends to linearly increase when the number of candidate models grows. It is

seen from these figures that the training time for the models with a single candidate model are reasonably low, and all are below 10 minutes, but they dramatically increase to dozens of minutes when the number of candidate models reaches 35, and especially the models with 250 initial iterations and 35 candidate models (Figure 4.8 (f)), the time cost is greater than 2 hours.

However, the lowest BIC scores which are used to evaluate the quality of the trained models are not really following the linear trend as the increase of candidate models, except the models that are with 10 initial iterations. It is seen from Figure 4.8 (a) (with 10 initial iterations) that the lowest BIC scores show a decreasing trend and the value reaches the lowest point of -820. However, the other cases with different initial iterations illustrate another characteristic trend. It is found that the BIC scores of the models with the first 3 numbers of candidate models (1, 5 and 10) have an obvious decreasing trend and as the number of candidate models increases, the BIC scores become relatively flat and close to each other, which does not significantly improve the quality of the trained models.





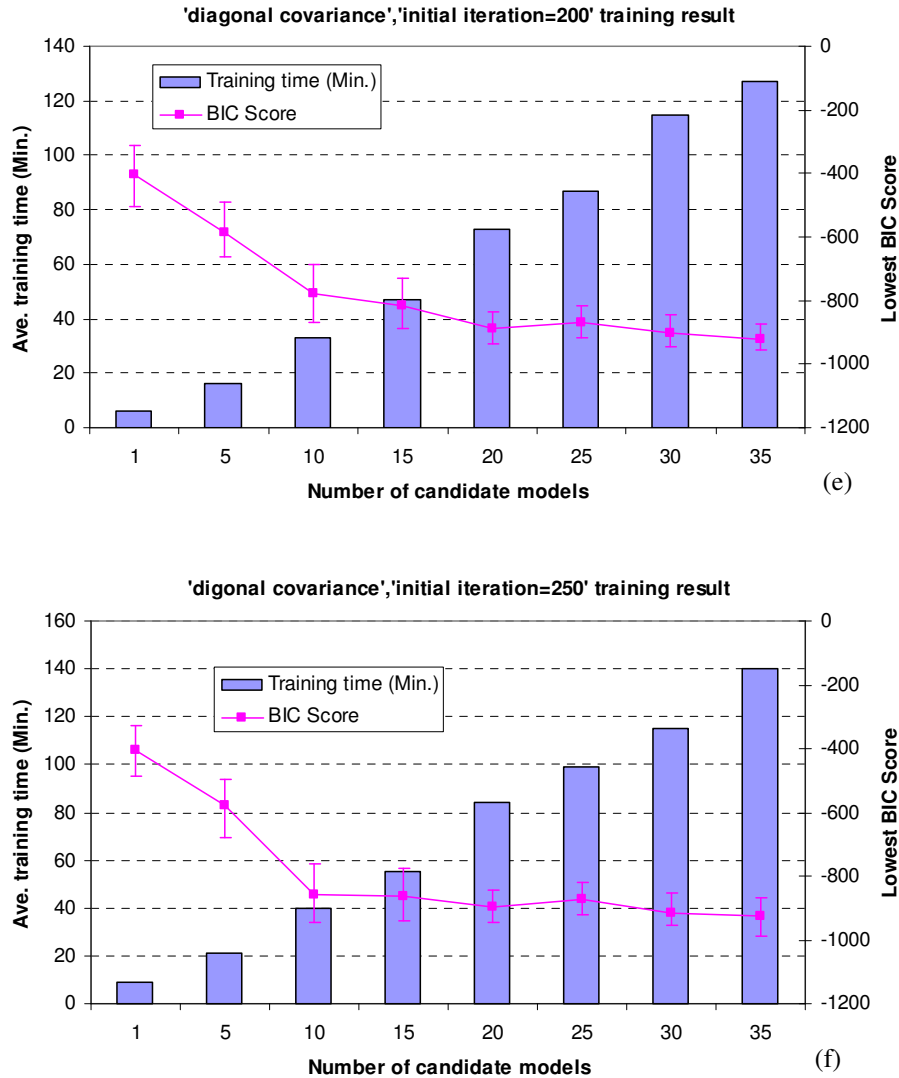


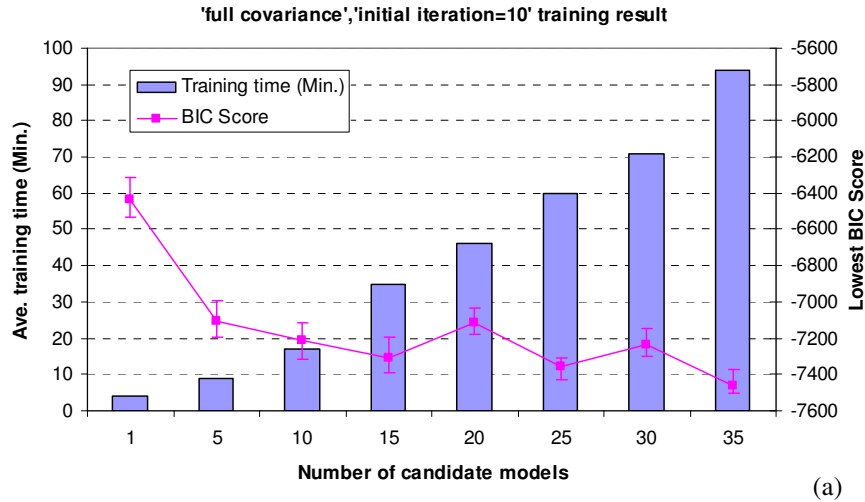
Figure 4.8 Training result of the clustering models under the diagonal covariance and different initial iterations (10-250) (a-f)

(2) Full covariance structure

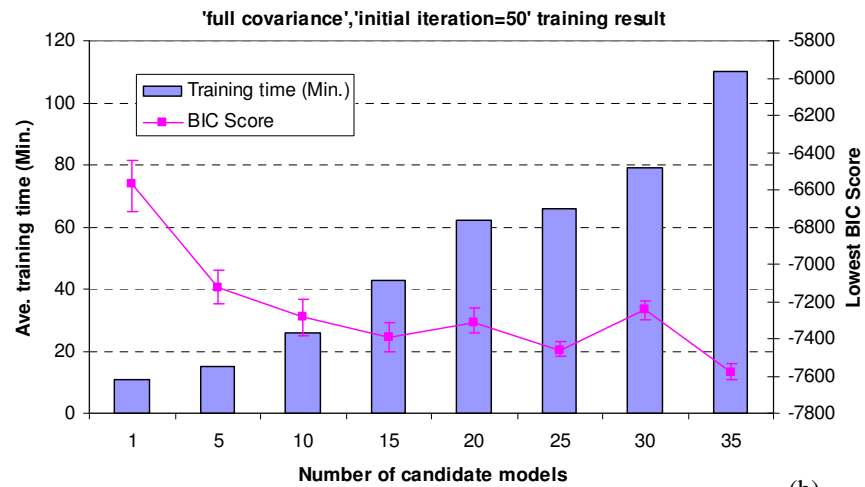
Figures 4.9 (a)-(f) show all the cases with the same full covariance structure and different initial iterations (10-250). Similar to the models with the diagonal covariance structure, the clustering training time of the full covariance models are also with the linear increasing trend as the number of candidate models grows. It is found from these figures that the training time for the models with a single candidate model are no longer low, and most of them have reached double figures, and not surprisingly, significantly increase to the high level which are all above 1 hour when the number of candidate models reaches

35, and the training time for the models with 250 initial iterations (Figure 4.9 (f)) climb to the highest point with 278 minutes.

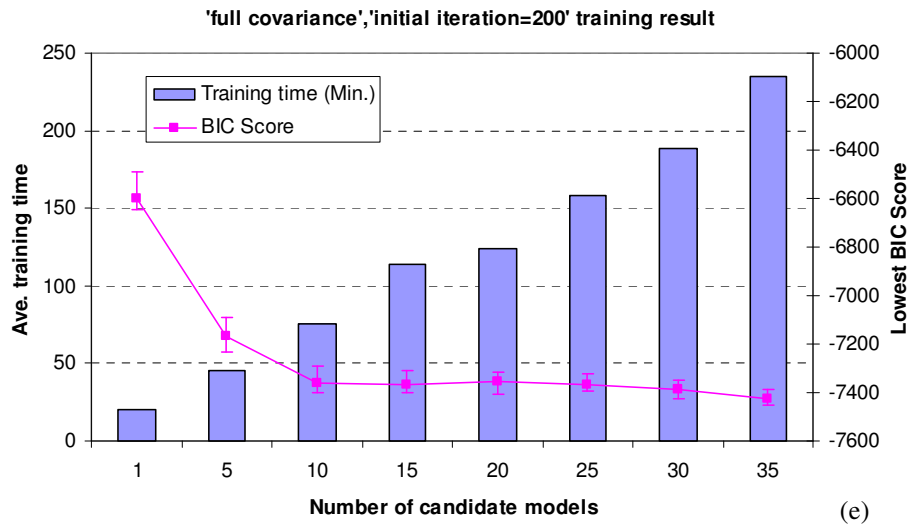
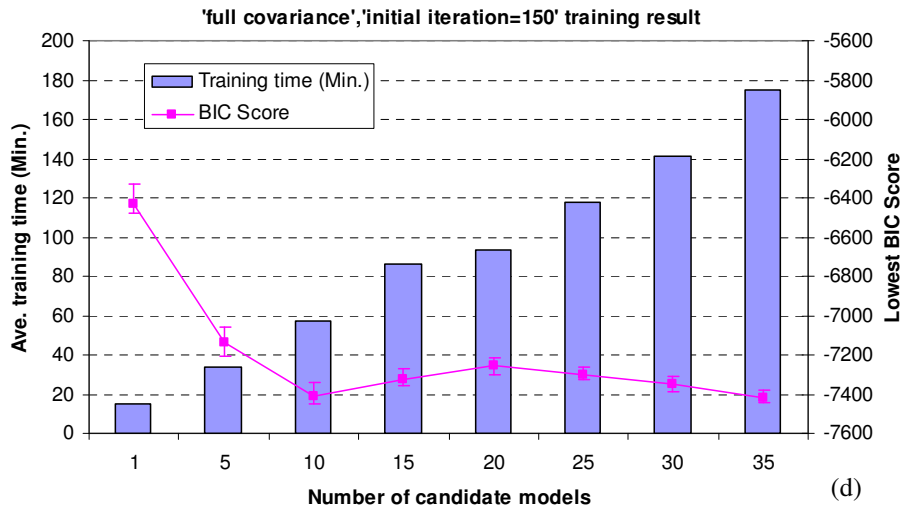
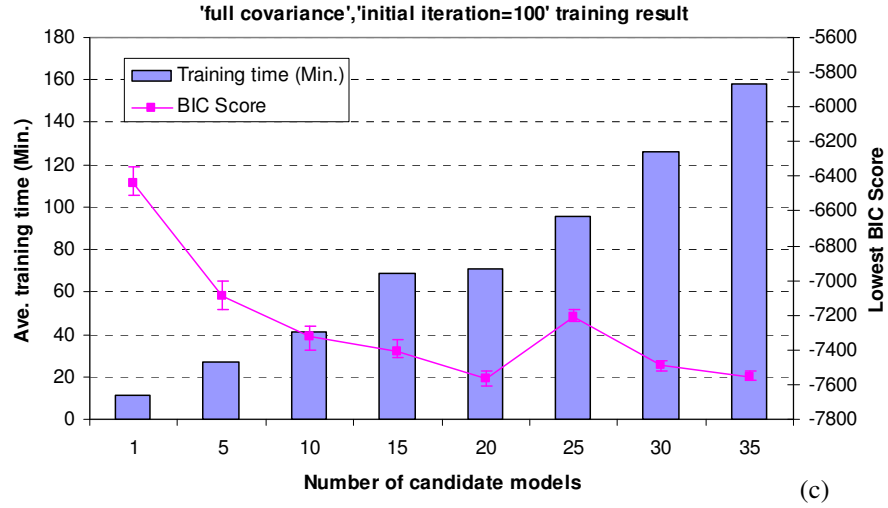
However, similar to the diagonal models, the lowest BIC scores of the full models do not follow the linear decreasing trend as the increase of candidate models. Nearly all the cases with different initial iterations illustrate the same trend. It is found that the BIC scores of the models with the first 4 numbers of candidate models (1, 5, 10 and 15) illustrate an obvious decreasing trend and as the number of candidate models increases, the BIC scores become relatively flat and fluctuated, which does not significantly improve the quality of the trained models. Therefore, it is not always true to assume that choosing a ‘large’ number of candidate models can achieve better quality models.



(a)



(b)



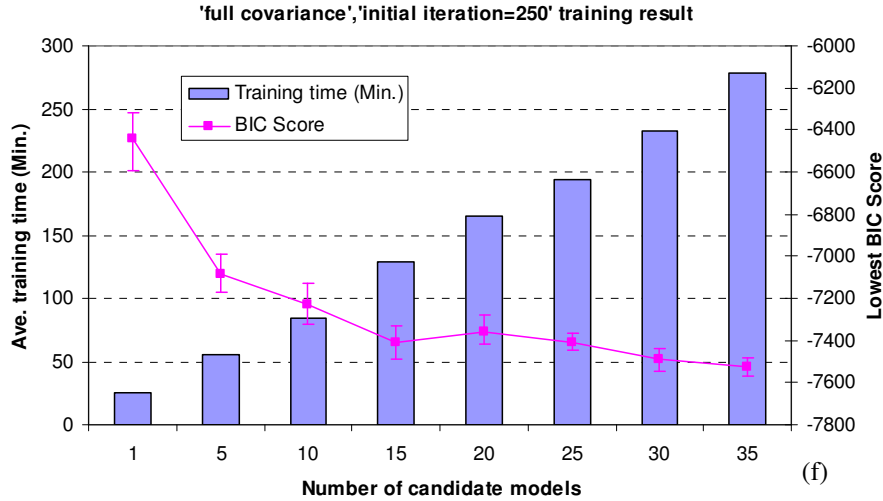
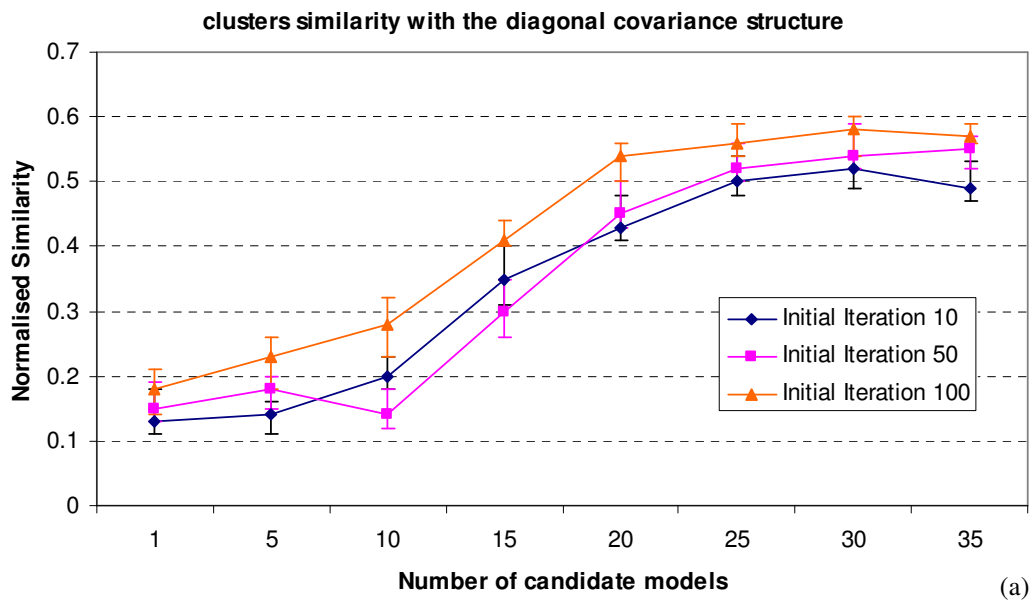


Figure 4.9 Training result of the clustering models under the full covariance and different initial iterations (10-250 (a-f))

4.4.3.2. Similarity between the clustering models

The repeatability of the clustering training has always been difficult, because the model is initialized randomly at the beginning of the training. Therefore, in this study, the measure of similarity between the models with the same training parameters is calculated and used to examine whether such an index could be improved by increasing the number of candidate models.



(a)

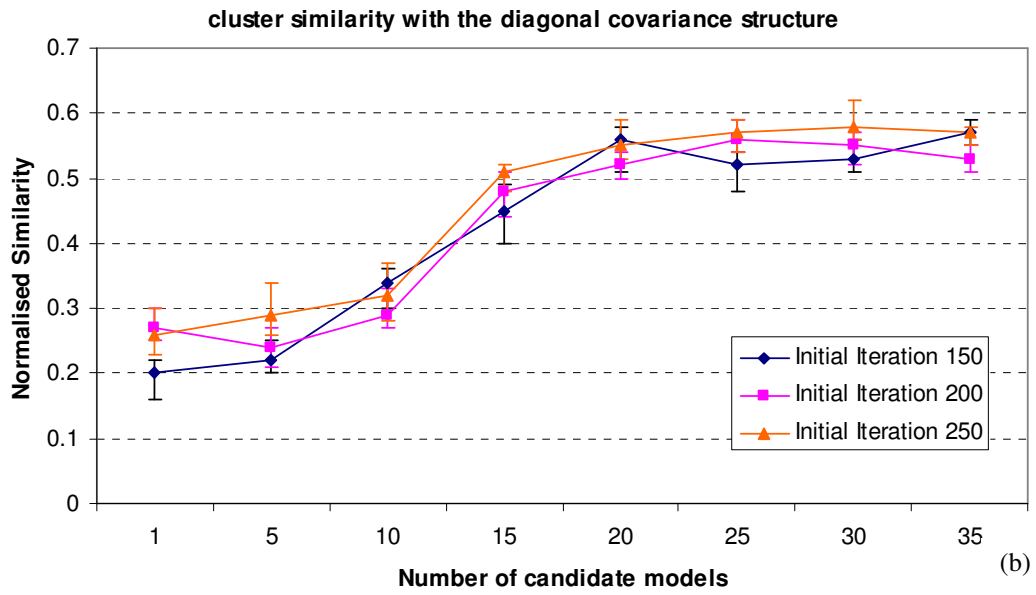
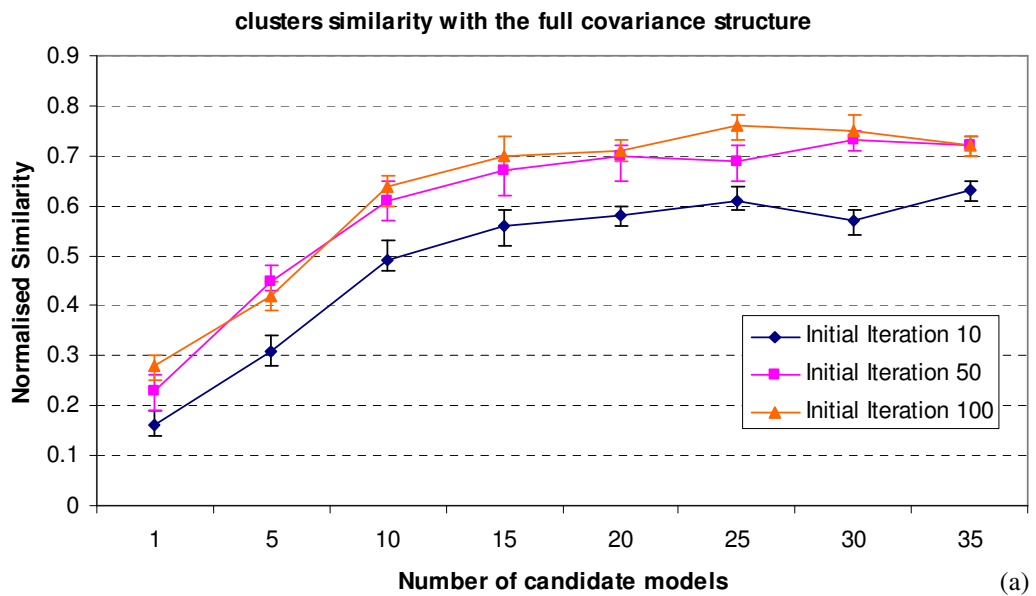


Figure 4.10 Similarity measures between clustering models under the diagonal covariance and different initial iterations (10-250)



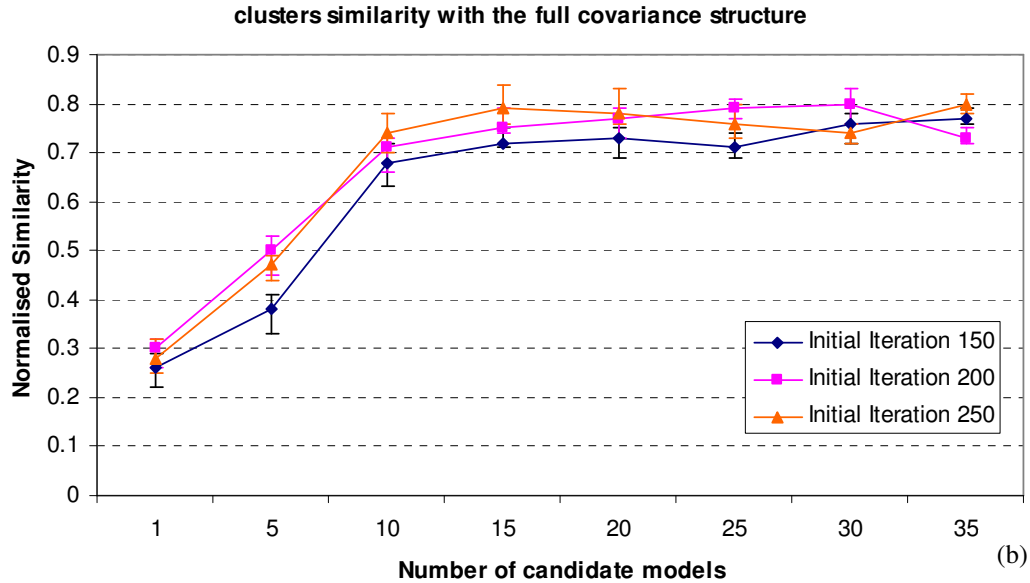


Figure 4.11 Similarity measures between clustering models under the full covariance and different initial iterations (10-250)

Figures 4.10 (a) and (b) plot the calculated model similarity against the increased number of candidate models under the same diagonal structure and different initial iterations (10-250). It is seen from the figures that the calculated similarity associated with varied numbers of candidate models almost have the same characteristic trend under the 6 different initial iterations. For the first 4 numbers of candidate models (1-15), an obvious increasing trend can be found under the 6 initial iterations, which verifies the assumption that the modelling stability and repeatability could be improved by increasing the number of candidate models. But the values of similarity measure for the next 4 numbers of candidate models (20-35) seem to follow a relatively flat trend, which does not have a significant impact on the stability of the modelling. This feature is examined and compared with the lowest BIC trend in the last section, and these two types of trends were found to have a consistent characteristic in that the values of the first 4 numbers of candidate models show the linear trend and then become flat and close to each other for the next 4 numbers of candidate models.

For the models under the full covariance structure, figures 4.11 (a) and (b) show similar calculation results. It is found that the trends of values under the different initial iterations

also have the ‘initially increasing and then flat’ trend, and seem to have a similar characteristic to the lowest BIC trend under the full covariance structure, in which the first 3 numbers of candidate models have significant influence on the calculated values while the rest of the number of candidate models are relatively weakly connected.

In this section, the parameter of number of candidate models (NCM) has been evaluated by three evaluation indices, and found that the quality and consistency of the models can be significantly improved when values of the NCM are changing between 1 and 15. This finding can be used to find appropriate and optimised models for the anomaly detection. In the next section, similar procedure would be applied to another evaluated parameter initial iteration (II) to see its impact on the model construction.

4.4.4. Evaluation of initial iteration

Similar to the evaluation approach to the number of the candidate models, in this section, the lowest BIC score, training time and similarity are also evaluated against the variation of the initial iteration for the candidate models. Table 4.5 summarises the evaluated parameter- initial iteration and its evaluation indices in section 4.4.4.

Evaluated parameter	Evaluation index
Initial Iterations (II)	BIC score (Section 4.4.4.1)
	Training time (Section 4.4.4.1)
	Similarity between models (Section 4.4.4.2)

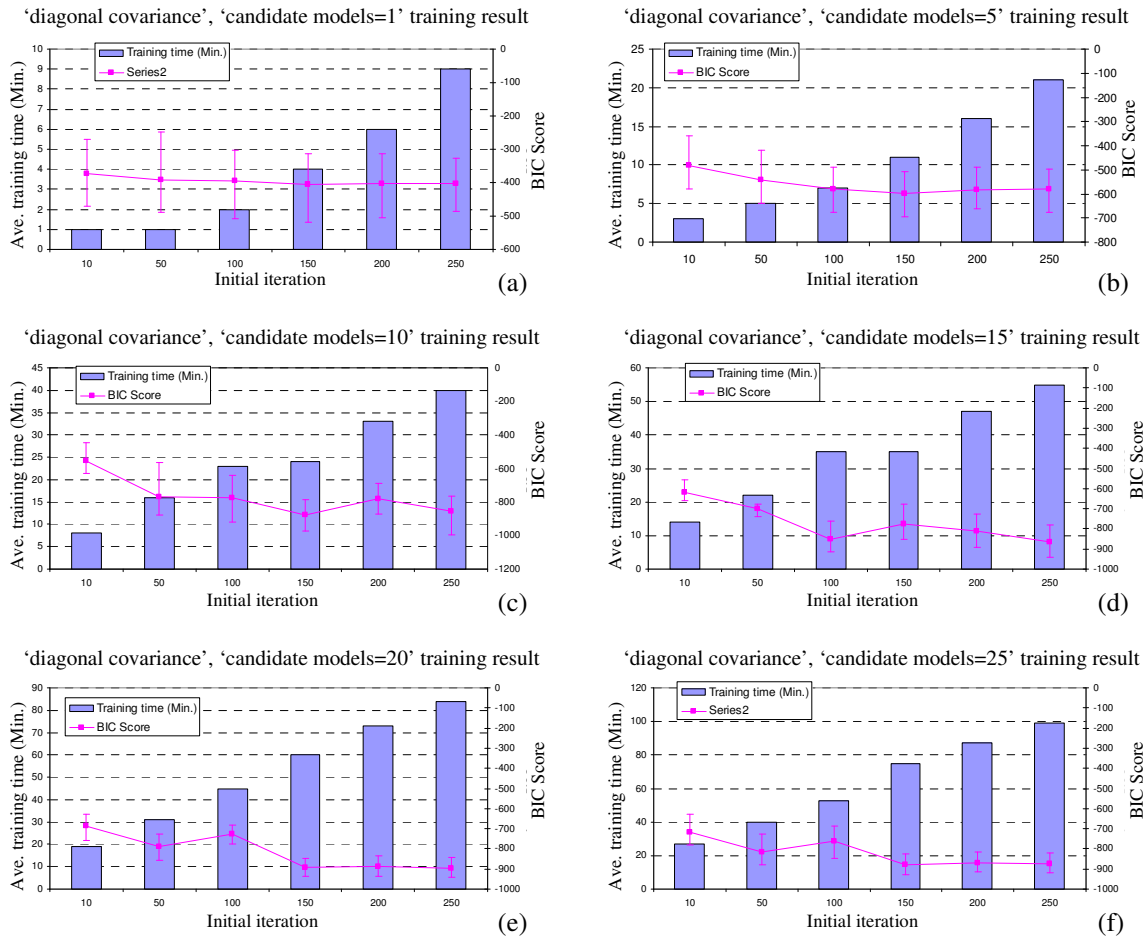
Table 4.5 Evaluated parameter and its evaluation index

4.4.4.1. Lowest BIC score and training time

From the relationships shown in figures 4.12 and 4.13, for almost all the cases with diagonal or full covariance structures and different numbers of candidate models (1-35), the clustering training time tends to linearly increase when the value of the initial iteration grows. These trends are also found to be similar to the ones that are with the increase of the number of candidate models in Section 4.4.3.1. The only difference is that

the increasing trend of the time expenditure for the initial iteration is not as significant as the number of candidate models.

However, the lowest BIC scores that are used to evaluate the quality of the trained models are not really following the linear trend as the increase of initial iterations. As Figure 4.12 shows, the lowest BIC scores of the diagonal covariance models generally show a decreasing trend, which is consistent with the trend of the number of candidate models. However, other cases of the full covariance models (see Figure 4.13) illustrate another characteristic. It is found that the BIC scores of the models do not have an obvious trend and as the number of candidate models increases, the BIC scores are relatively fluctuated and close to each other, which does not significantly improve the quality of the trained models.



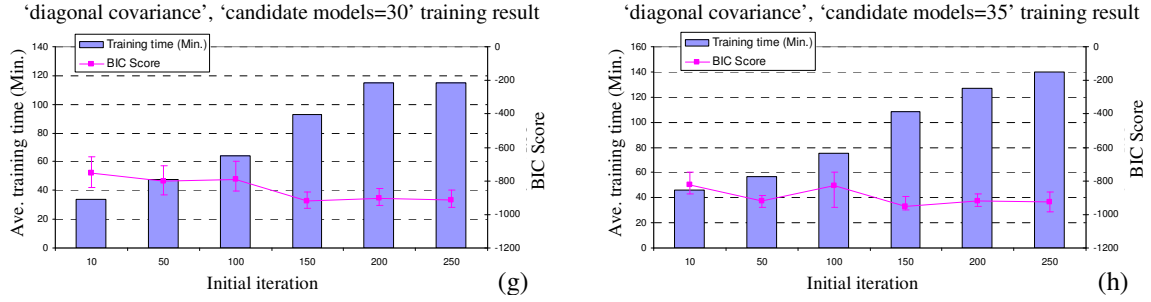


Figure 4.12 Training result of the clustering models under the diagonal covariance and different candidate models (1-35 (a-h))

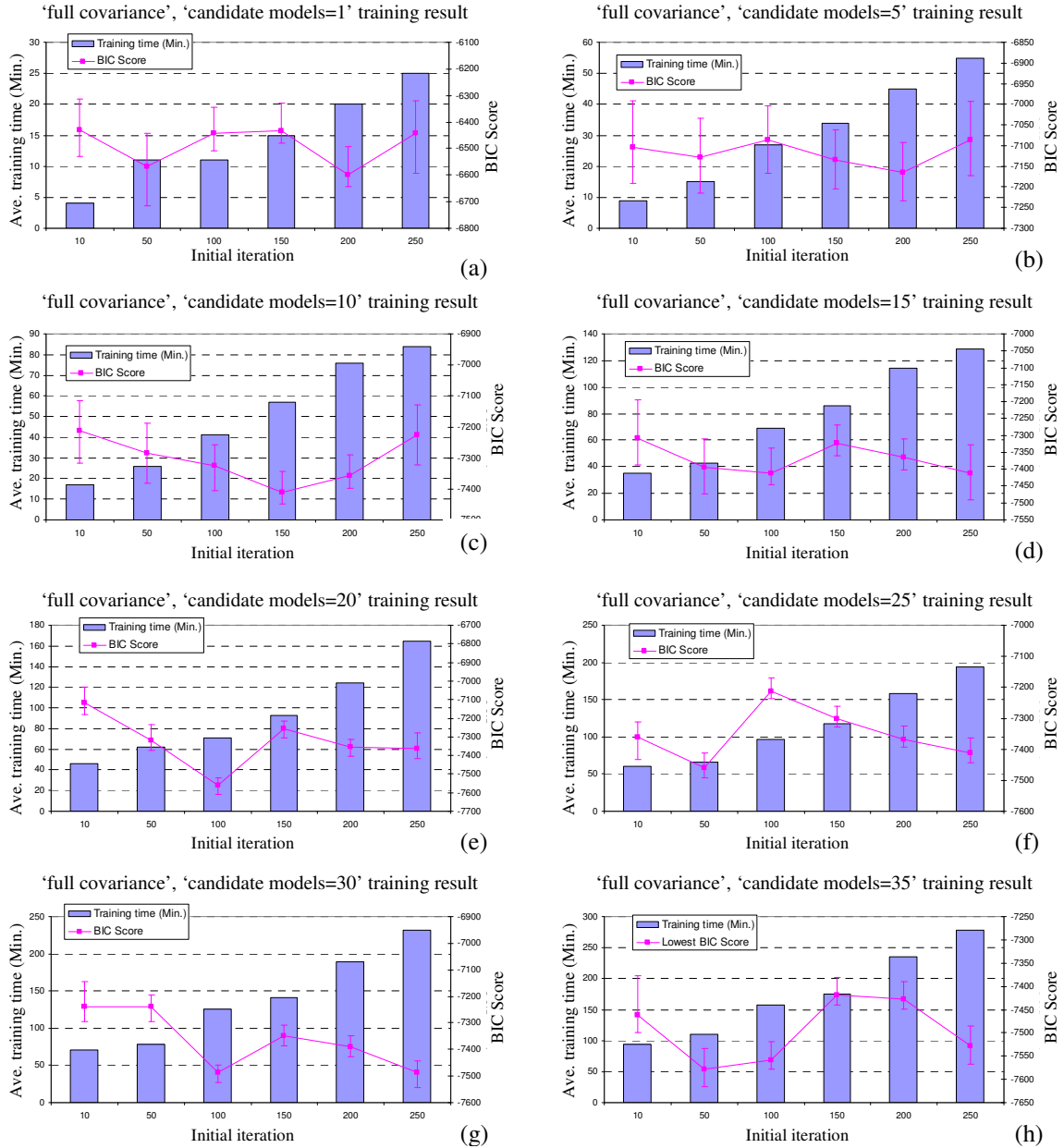


Figure 4.13 Training result of the clustering models under the full covariance and different candidate models (1-35 (a-h))

4.4.4.2. Similarity between the clustering models

For the models under the diagonal and full covariance structures, figures 4.14 and 4.15 show the similarity calculation results respectively. It is found that the trends of initial iteration values under the different number of candidate models also have the ‘increasing firstly and then flat’ trend, and seem to have a similar characteristic to the similarity trend of the number of candidate models in Section 4.4.3.2, in which the first 3 numbers of candidate models have a significant influence on the calculated values while the rest of the number of candidate models are relatively weakly connected.

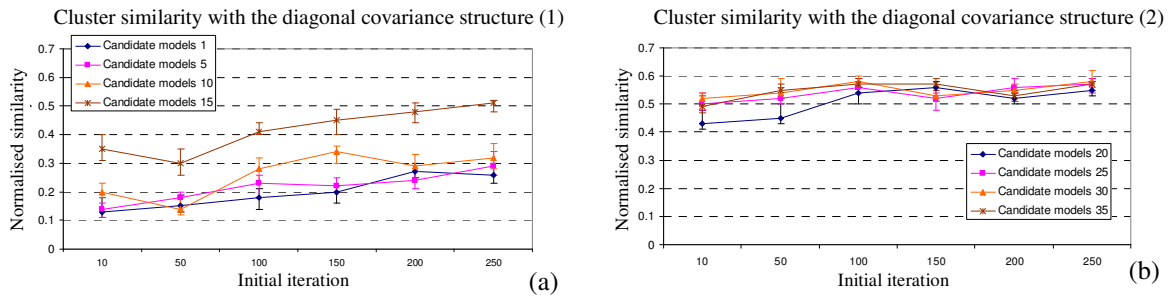


Figure 4.14 Similarity measures between clustering models under the diagonal covariance and different candidate models (1-35)

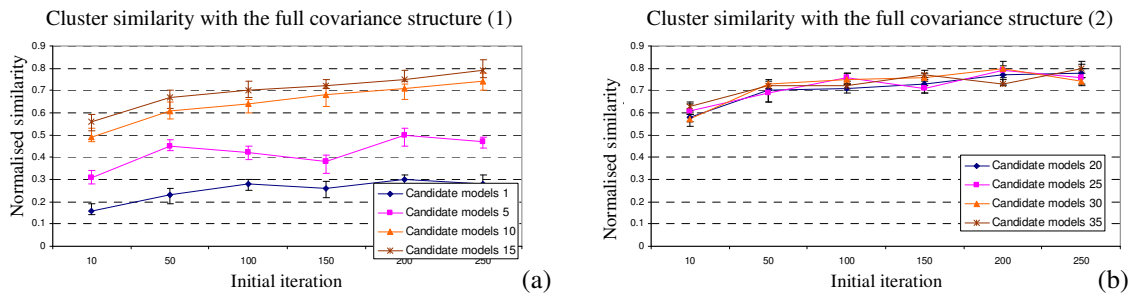


Figure 4.15 Similarity measures between clustering models under the full covariance and different candidate models (1-35)

Section 4.4.4 has evaluated another training parameter-initial iteration (II) with evaluation indices, and found that model performance can still be improved by changing the value of II, but seems not as significant as the NCM.

4.4.5. Evaluation of covariance structure

The above evaluated parameters, the number of the candidate models and its associated initial iterations are both examined under the diagonal and full covariance structures, and in this section, these two types of covariance structures are compared to see whether they could lead to any differences for the evaluated indices, training time, lowest BIC score and similarity.

Figures 4.16 (a) and (b) illustrate the training time of the diagonal and full covariance structures respectively with the variation of the number of candidate models and initial iterations. It is seen that the training time of the two types of models are both increasing with the growth of the two training parameters, but the time cost of the full models seems to be much higher than that of the diagonal models.

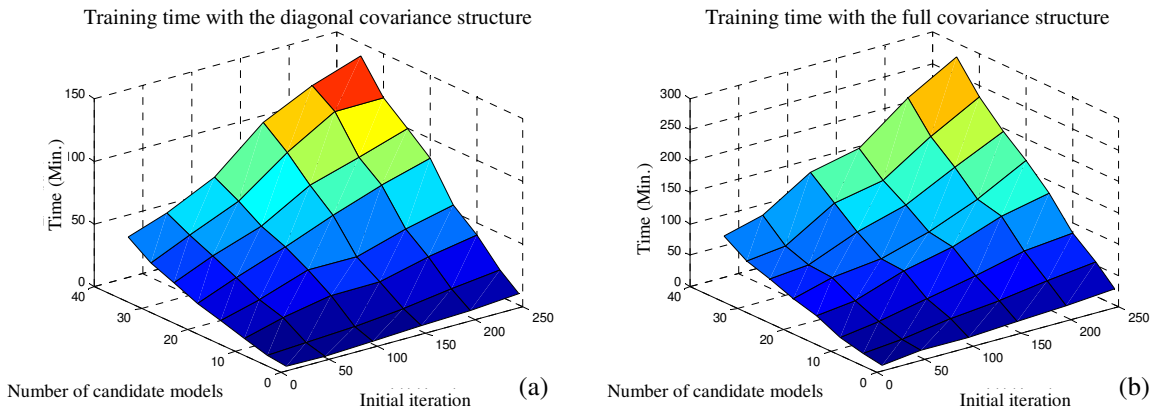


Figure 4.16 Training time of clustering models with different parameters

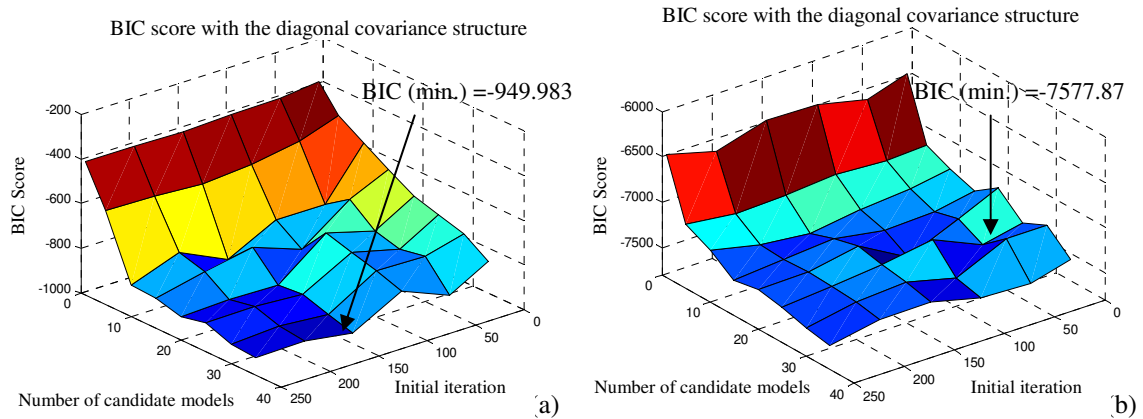


Figure 4.17 BIC score of clustering models with different parameters

The BIC scores of the diagonal and full covariance models are presented in figures 4.17 (a) and (b). It is found that the BIC values of the full models are much lower than those of the diagonal models. However, the diagonal and full models have different structures and characteristic parameters and have an influence on the BIC calculation; therefore, the BIC scores between these two covariance structures are not comparable.

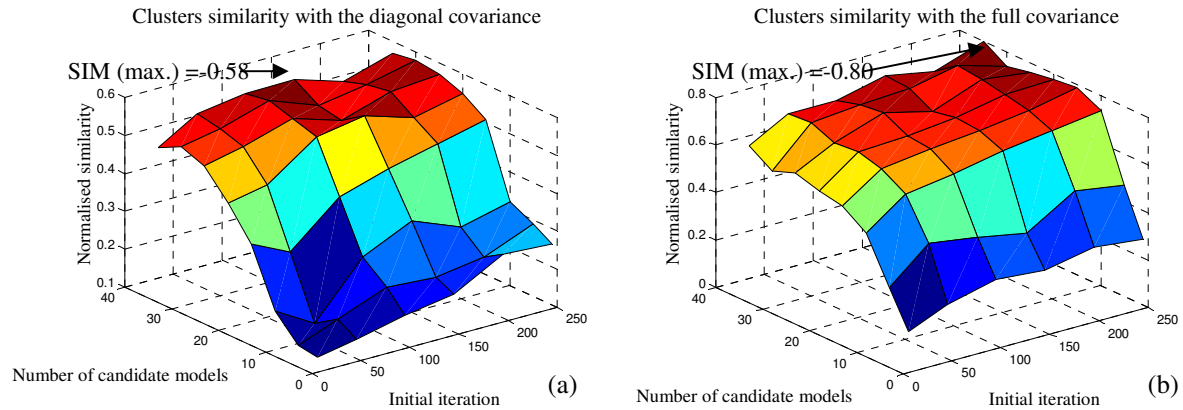


Figure 4.18 Similarity measures between clustering models with different parameters

Figures 4.18 (a) and (b) show the similarity values of the diagonal and full covariance models. It is seen that the shapes of the two models are quite similar, but the highest value of the full covariance models could reach around 0.8 while the highest value of the diagonal covariance is only as high as around 0.6.

4.4.6. Summary

Section 4.4 has introduced a systematic procedure for using clustering models for constructing the anomaly detection model. In this procedure, a systematic way of approaching the consistent and stable models is proposed in this study.

- Training time of the clustering models is evaluated with the changes of the number of candidate models, initial iterations and covariance structure. It is found that the training time consumption is linearly increased with the growth in the

- number of candidate models and initial iterations, but full models generally cost much more time than diagonal ones under the same training parameters.
- 3 training parameters are also varied to calculate corresponding BIC scores, and found that for both diagonal and full structures, BIC values are obviously influenced and have the expected decreasing trend when candidate models are with small numbers, such as 1, 5, 10 and 15; but values become stable or slightly changed if the number of candidate models keep increasing. On the other hand, initial iterations also contribute to changes of BIC scores, but seem not as significant as the number of candidate models. This is probably due to the fact that the opportunity of missing the more appropriate solution with a low number of candidate models is increased, causing unstable BIC indication; on the other hand, when the number of candidate models reaches certain values, it is more possible to select qualified models with similar BIC values.
 - The similarity index is an important parameter to measure the stability and consistency between the clustering models. The calculated results indicate that the similarity index responds to both of the number of candidate models and initial iterations, but seems to only be effective when these two parameters are with small numbers, such as 1, 5, 10 and 15 for the number of candidate models and 10, 50 and 100 for initial iterations. Furthermore, the values of similarity for full covariance models are generally higher than those of diagonal models.
 - From the results, since the quality of the clustering model (evaluated by BIC score) and stability between the models (evaluated by similarity measure) are not significantly affected by the training parameters after they have reached a certain level, it is unnecessary to consistently increase the value of training parameters, because it involves much more training time. Hence, it should be carefully evaluated when selecting appropriate models for anomaly detection.
 - The BIC and similarity are only the indices which are used to assess the models themselves, and it is also very necessary to examine their anomaly detection capabilities, and to see the effect of the training parameters on the rate of the anomaly detection. This will be discussed in Chapter 5.

4.5. Adaptation of the anomaly detection model

It was explained in Chapter 2 that the amount of unknown anomalies might be included in the training data which are modelled by the GMM. So the challenge is how to identify areas of cluster space that are associated with the anomalies. In this section, the developed novel approaches that are used to identify and eliminate anomalies in the training data will be presented. Initially, the original methodologies are introduced and demonstrated with the selected bearing data, and then the improved optimized approaches which are based on the fundamental ones will be presented. The flowchart in Appendix A2 shows the general procedure of the model adaptation. Furthermore, Table 4.6 summarises the structure of the context which will be introduced in section 4.5.

Fundamental method (4.5.1)		Optimised method (4.5.3)	
Information theory (4.5.1.1)	Distance method (4.5.1.2)	Multivariate segmentation (4.5.3.1)	Density changes (4.5.3.2)
Demonstrated by two dimensional bearing data (4.5.2.1)	Demonstrated by two dimensional bearing data (4.5.2.2)	Application to the integrated baseline data (4.5.4.1)	Application to the integrated baseline data (4.5.4.2)

Table 4.6 Structure of the context in section 4.5

4.5.1. Fundamental methodology

To define the proposed approach, the data from the bearing tests were reviewed beforehand, and it was found that there are basically two types of anomalies: 1) an abnormal trend (unknown fault) is occurring at occasional time intervals of the time series; 2) discrete noise (instrumental or environmental effect) is spreading throughout the whole test, and is distant from the normal data. Therefore, the developed method needs to satisfy the requirements of: 1) locating regions of space associated with low and occasional support; 2) locating regions with points that are distant from other regions.

4.5.1.1. Method based on information theory

To meet the first requirement, it is desirable to develop an uncertain measure of anomalies occurring in the time series. And in the developed approach, the statistic of

entropy in information theory is applied to cope with the existence of occasional occurring anomalies. For n partitioned categories (equal time intervals) in the time series, and for a training vector x with all samples having a cluster ID v , the entropy E is defined as [7]:

$$E(x = v) = \sum_{i=1}^n -p_i \log_2 p_i \tag{4.15}$$

and

$$\sum_{i=1}^n p_i = 1$$

Where p_i is the probability of training vector x coming from cluster v of category i occurs.

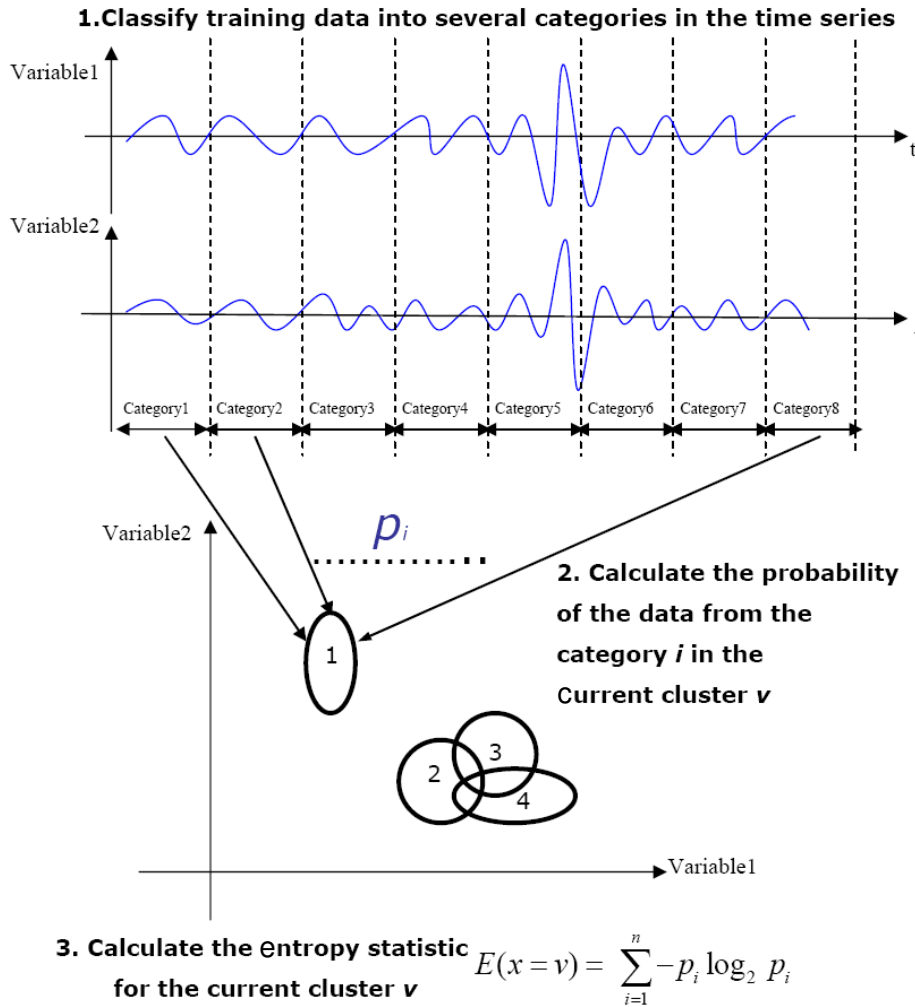


Figure 4.19 Steps of the entropy-based approach to remove anomalies in the training data

Figure 4.19 shows the concept and steps of using the entropy statistic for finding the anomalies embedded in the GMM subspace. In the first step, the time series is discretized into several categories $(1, \dots, n)$. Here, each bin width is set equally according to experience, but the issue of time series segmentation [8] needs to be considered to enable the categorization to be more reasonable and intelligent (i.e. to automatically segment abnormal trend into separate category). And then, the probability of the data from category i needs to be calculated in each trained cluster. Finally, the entropy statistic is applied to examine which elements each cluster consists of. Hence, the main idea of the entropy-based anomaly detection is to explore whether the support cases in a particular cluster are from the occasional time intervals (i.e. an abnormal trend with different distribution) or multiple time intervals (i.e. normal data with similar distribution). If data in a cluster is just from one time interval, the probability of the data points from this time interval in the current cluster will be high, i.e. approx 100%, and the probabilities of the data points from other time intervals are 0%, so that the entropy score for the current cluster will be low or close to zero. Otherwise, the entropy score of other clusters supported by multiple categories' data will be much higher. Based on this method, the clusters with different data distribution can be distinguished.

4.5.1.2. Method based on the distance

Although anomalies are found to occur occasionally in the time series of the training data in most situations, anomalies are also appearing frequently throughout the test with strong certainties, and the entropy-based method can be insensitive to these anomalies. Hence, to fill the second requirement of locating the distant region, a distance-based method is developed to measure the data dispersion clusters with distant points from other regions being eliminated from the calculation. In this study, Hotelling's T-squared distance is calculated between two clusters [9], using equations (4.16) and (4.17):

$$T^2 = \frac{n_A n_B}{n_A + n_B} (\mu_A - \mu_B)^T \hat{\Sigma}_{AB}^{-1} (\mu_A - \mu_B) \quad (4.16)$$

$$\hat{\Sigma}_{AB} = \frac{1}{n_A + n_B - 2} ((n_A - 1) \hat{\Sigma}_A + (n_B - 1) \hat{\Sigma}_B) \quad (4.17)$$

Where n_A and n_B are the number of support cases in clusters A and B ; μ_A and μ_B are sample means of clusters A and B , respectively; $\hat{\Sigma}_A$ and $\hat{\Sigma}_B$ are estimates of covariance matrices of the clusters A and B , respectively.

4.5.2. Demonstration with two-dimensional bearing data

In order to demonstrate the proposed method, two-dimensional electrostatic data with various characteristic features are selected from the test 19 as Figure 4.20 (a) shows. Obvious characteristic features of the selected bearing data such as the running-in feature (at the start of the test for each variable), abnormal increasing trend (at the end of the test for variable ‘a’) and discrete noise throughout the test (indicated in variable ‘b’) appeared. Moreover, data are trained with the EM algorithm to obtain the model, and 11 clusters are obtained to form the GMM space, as illustrated in Figure 4.20 (b).

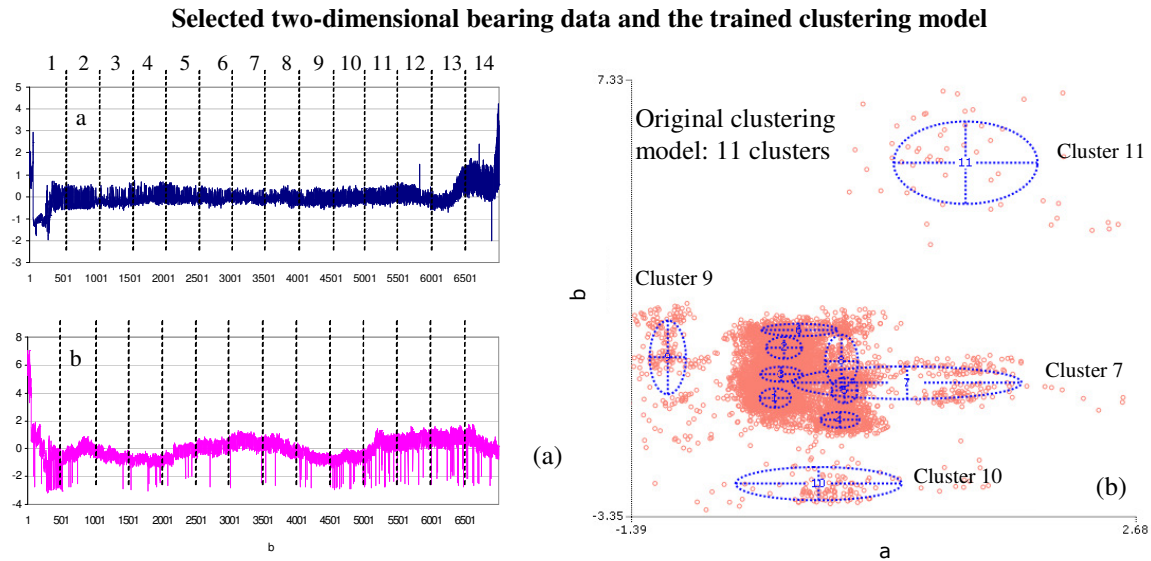


Figure 4.20 Selected two-dimensional electrostatic sensing data (a) and the trained clustering model (b)

4.5.2.1. Result of the information theory method

Following the procedure of the approach introduced in Section 4.5.1.1, a two-dimensional electrostatic bearing dataset with 7,000 data points were divided into 14 equal categories,

as shown in Figure 4.20 (a); next, the content of each cluster is explored, and the probabilities of the divided categories in each cluster are calculated, as Figure 4.21 (a) shows. It is seen that most of the clusters are supported by the cases from multiple categories with closed probabilities between 10% and 20%, except clusters 9 and 11. In these two clusters, nearly all the support cases are from category 1 with probability values of 98% and 99% respectively. After the calculation of the probabilities, the entropy scores of each cluster are estimated, and from Figure 4.21 (b), an entropy score of clusters 9 and 11 are seen as significantly lower than the other scores of clusters with values of 0.02 and 0.0017 respectively, because of the concentrated support cases. Thus, these two clusters are highly suspected to be the regions associated with occasional occurring anomalies, and it is suggested that they should be removed from the original normal space.

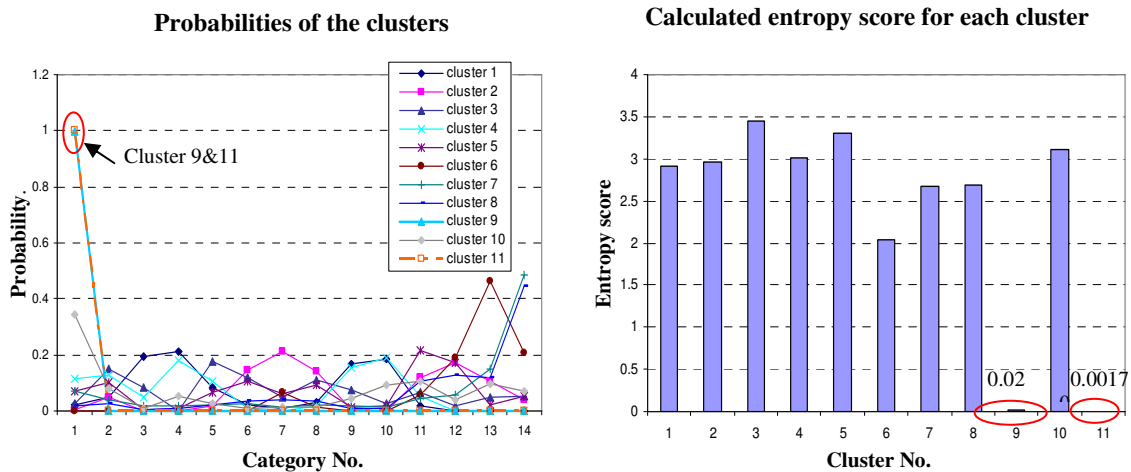


Fig 4.21 (a) Consisted probabilities of the clusters

Figure 4.21 (b) Calculated entropy score for each cluster

Figure 4.22 (a) shows the dataset after the removal of clusters 9 and 11 and their associated support cases. It is clearly seen that the abnormal trend at the running-in area has been successfully filtered out. However, the abnormal trend at the end of the test of variable 'a' and continuous noise in variable 'b' are still in the dataset. If the adapted clustering model is examined as Figure 4.22 (b) shows, it can be found that the associated regions of these two remaining features are clusters 7 and 10 which are supported by the cases from multiple categories in the time series and that is why the entropy score of these two clusters failed to be used, indicating abnormal behaviour.

It has been indicated at the beginning of Section 4.5.1 that another important feature of the encountered anomalies is their distant characteristic. From the adapted dataset of Figure 4.22 (a), the remaining anomalies are distant from normal data, and their values cover more of a magnitude of bins in the ‘y’ axis than the normal data which are just distributing within a certain level. Therefore, it is suspected that the entropy scores of the clusters associated with these anomalies will be higher than the ones with the normal data, if the classified categories are set in the magnitude direction instead of the time direction.

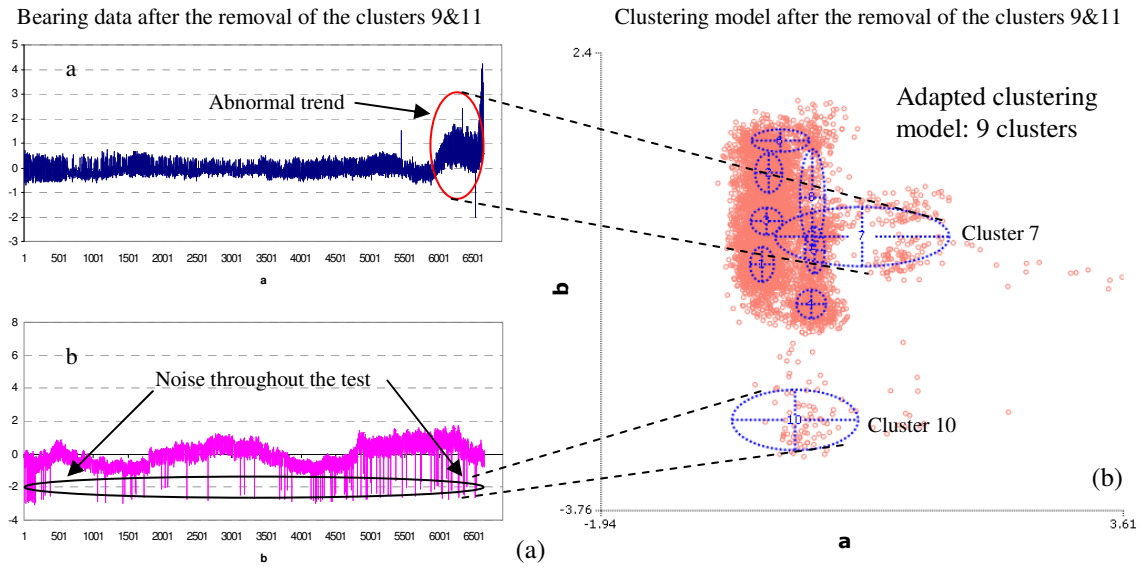


Figure 4.22 Selected two-dimensional electrostatic sensing data after the removal of clusters 9 and 11 (a) and the adapted clustering model (b)

Nevertheless, the values of variables are different from each other, and it is impossible to set the categories separately for each variable, otherwise the multivariate samples would be confusing about which categories of the magnitude direction they belong to. To accommodate this, a virtual space of $\mu_v = [0, \dots, 0]$; $\Sigma_v = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}$ is created, and the Hotelling's T-squared distance between the multivariate samples and the virtual space can be estimated, so that the separated variables are fused, and the categories could be set in the direction of distance.

Figure 4.23 (a) illustrates the Hotelling's distance between the simulated data and the virtual space. In this figure, 12 categories are assigned in the distance direction, and

following the entropy-based anomaly detection procedure, the entropy score for each cluster in the GMM space can be obtained as Figure 4.23 (b) illustrates. It is seen that the entropy scores of clusters 7, 9, 10 and 11 are significantly higher than the other ones which validates the support cases in these 4 clusters are distant from the normal data and are widely distributed in the distance direction. Figure 4.24 shows the two dimensional bearing data and the adapted clustering model after the removal of clusters 7, 9, 10 and 11 and their associated cases. It is found that most of the anomalies have been successfully removed, and both variables present a perfectly flat trend.

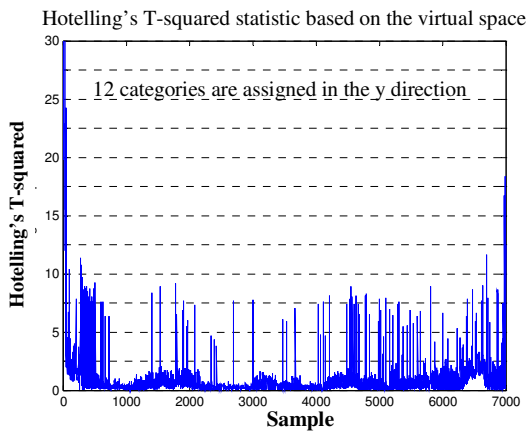


Fig 4.23 (a) Hotelling's T-squared statistic of the data based on the virtual space

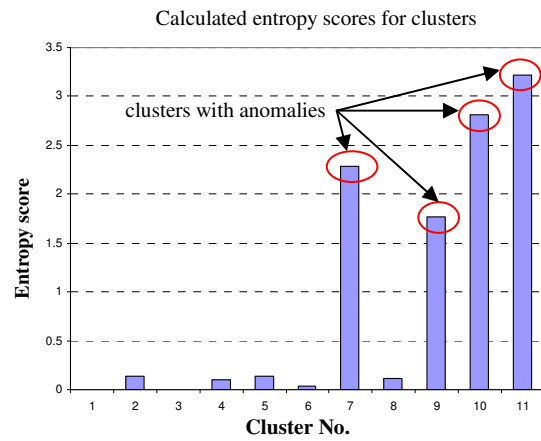


Fig 4.23 (b) calculated entropy for each cluster with the segmentation in the 'y' direction

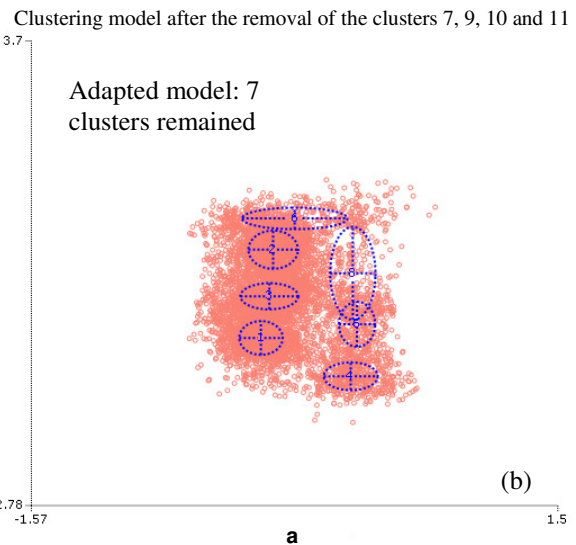
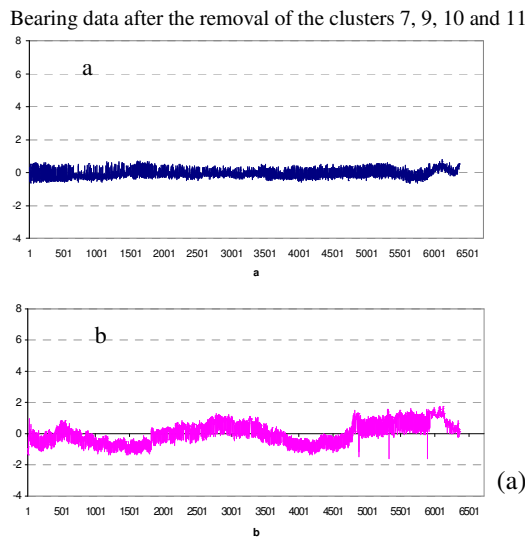


Figure 4.24 Bearing data after the removal of clusters 7, 9, 10 and 11(a) and their adapted clustering model (b)

From the above experimental study, the proposed entropy and distance-based approaches show the capability of identifying anomalies within the training data. The combination of the categorization methods in both time and magnitude directions is powerful, as the occasional and wide-spread characteristics of the anomalies could not be caught by the derived entropy scores.

4.5.2.2. Results of the distance method

Apart from the approach with the entropy statistic, the second strategy to remove anomalies in the training data is the measure of cluster distance. With the second anomaly detection strategy, Hotelling's T-squared distance is calculated between each of the Gaussian components (clusters). Figure 4.25 (a) illustrates the calculated result. It can be seen that clusters 9 and 11 are recognized as being distant from the other clusters, so the support cases associated with these two clusters are identified as anomalies. Figure 4.25 (b) shows the calculated cluster distance after the removal of clusters 9 and 11. It is seen that the distance of clusters 7 and 10 to the other ones are higher. Therefore, clusters 7, 9, 10 and 11 are defined as the abnormal regions. This discovery is found to be consistent with the findings of the entropy-based method.

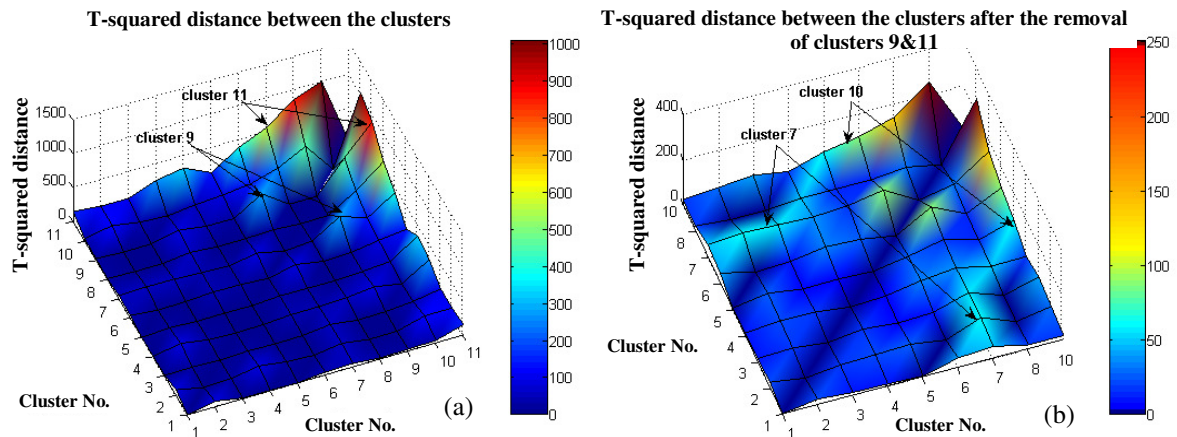


Figure 4.25 Hotelling's T-squared distance between the clusters (a) and the distance after the removal of clusters 9 and 11 (b)

4.5.3. Optimised approaches

4.5.3.1. Entropy statistic associated with automatic multivariate segmentation approach

For the entropy-based method, however, a difficult task is to determine how many categories are appropriate to be assigned to the training data. Different categorizations may directly affect detection. Principally, abnormal data in the time series are desired to be segmented from other types of data, and the clustering algorithm seems to obtain the capability of grouping similar data points. But they cannot be directly applied to a multivariate time-series, because apart from similarity, data points in the clusters must also be contiguous in the data ID, and so the ID related data have to be considered during clustering. In such a case, the categorization problem can be defined as conditional clustering: data points should be grouped based on their similarity, but conditioned to the data ID, which means all data points in a cluster must come from a successive data point.

Since the data ID related variable t_n is included and independent from other features x_n , the EM based clustering algorithm is implemented by the membership function of data ID points:

$$A_i(t_n) = \exp\left(-\frac{1}{2} \frac{(t_n - \mu_m^t)^2}{\sigma_{m,t}^2}\right) \quad (4.18) \quad \beta_i(t_n) = \frac{A_i(t_n)}{\sum_{j=1}^c A_j(t_n)} \quad (4.19)$$

Where μ_m^t and $\sigma_{m,t}^2$ are the means and variances of the membership function.

Hence, the joint unconditional probability function could be adapted from equation (4.1) and (4.2) to:

$$p(x; \Theta_m) = w_m \frac{1}{(2\pi)^{\frac{d}{2}} |C_m|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} (x - \mu_m)^T C_m^{-1} (x - \mu_m)\right\} \times \frac{1}{(2\pi)^{\frac{1}{2}} |\sigma_{m,t}^2|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \frac{(t_n - \mu_m^t)^2}{\sigma_{m,t}^2}\right\} \quad (4.20)$$

And then, the algorithm could follow the classical EM iteration to estimate the maximum likelihood of Θ . It should be noted that this data ID conditional clustering is independent of the process of anomaly detection model construction, and the entropy score of the clusters in the anomaly detection model can be calculated based on the automatic time-series categorization results obtained in this section.

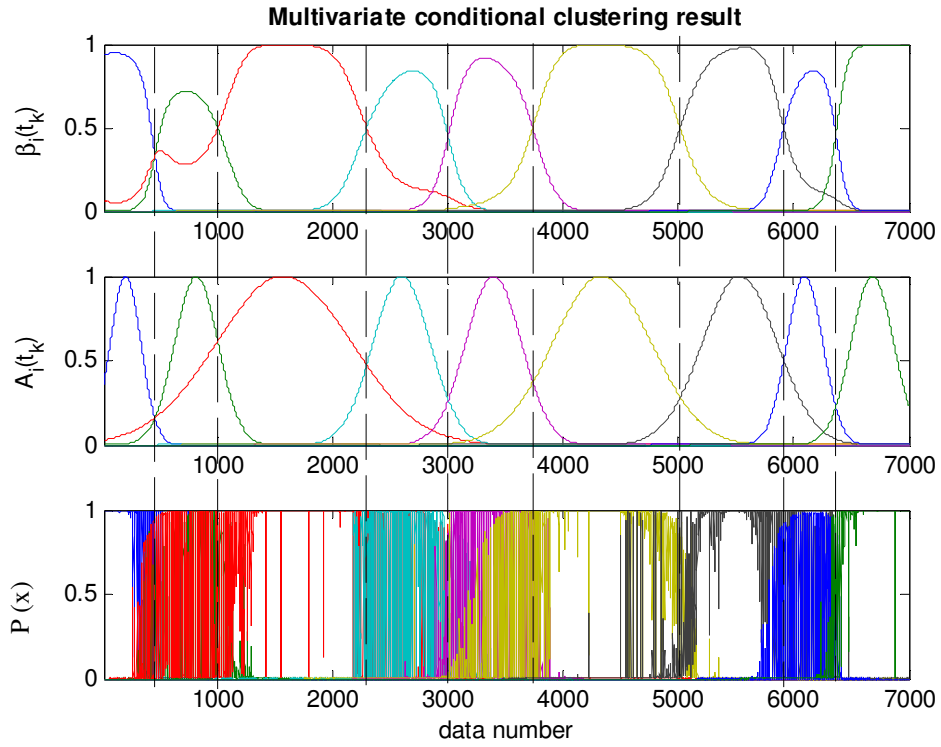


Figure 4.26 Result of the multivariate conditional clustering for the bearing data

two dimensional bearing data with 9 assigned categories in the time series

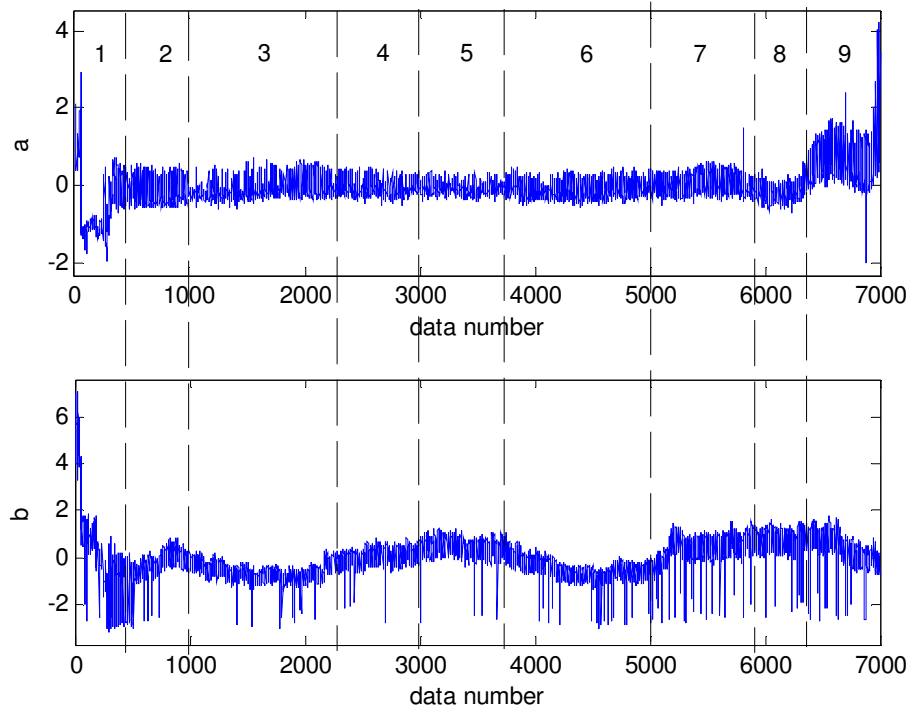


Figure 4.27 Two-dimensional bearing data with 9 assigned categories

Figure 4.26 shows the conditional clustering result and it is found that 9 conditional clusters are obtained and this result was assigned to the two-dimensional data as Figure 4.27 illustrates. With this categorization result, the constructed model as Figure 4.20 (a) shows can be re-analyzed to examine which one of the 9 categories the support cases of each cluster (Figure 4.20 (a)) is from. And then, the entropy score for each cluster can be re-calculated based on new categorization results in the time series.

Figure 4.28 is the calculated entropy score for each cluster in the trained model (Figure 4.20 (b)). Compared to the entropy scores based on manual categorization results (Figure 4.21 (b)), clusters 9 and 11 were found to be lower by both method, but cluster 7 was only found to have the low entropy score in Figure 4.28. This is because the abnormal trend at the end of variable ‘a’ (figure 4.20 (a)) was found in the support cases of cluster 7 and has been intelligently separated from other types of data by the conditional clustering approach (introduced in this section and shown in Figure 4.27), while the abnormal trend of variable ‘a’ was found occupying two more categories in the original manual categorization approach (see figure 4.20 (a)).

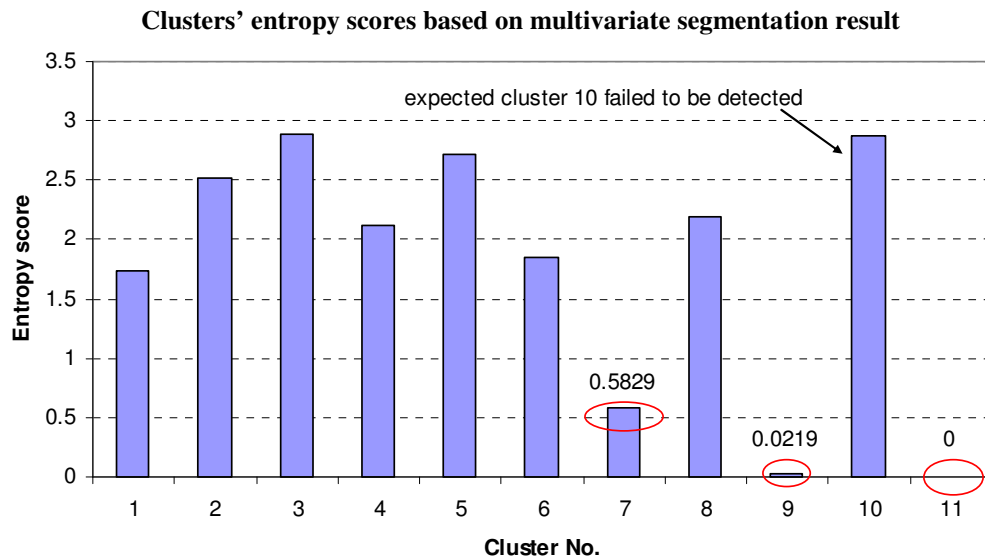


Figure 4.28 Clusters' entropy scores based on the multivariate segmentation result

On the other hand, cluster 10 still failed to be detected, as its support cases are the noise throughout the test and it occupies multiple categories, even though it was assigned by the intelligent approach.

4.5.3.2. Entropy statistic associated with the inspection of density changes

In Section 4.5.2.1, while using the entropy score to measure the uncertainty, the categories were also assigned in the magnitude direction to find anomalies which are widely distributed in the ‘y’ direction. However, similar to the original time series category assignment, the categories in the magnitude direction were also set manually and the adaptation of the anomaly detection model was strongly affected by this randomness. It has also been introduced in Section 4.5.2.1 that the categories in the magnitude direction was set in the fused T-squared statistic (see Figure 4.23 (a)), and it is desired to develop a method to automatically set categories in the ‘y’ direction of the T-squared statistic.

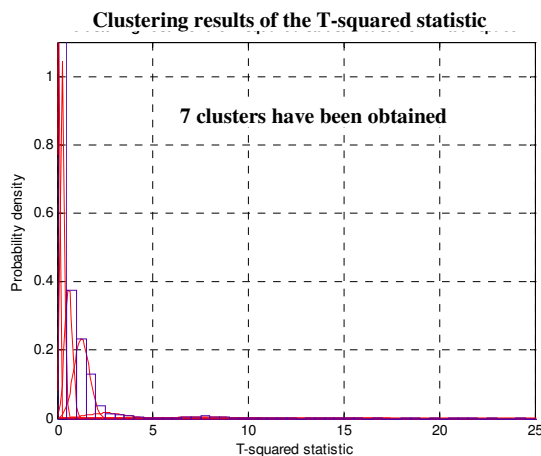


Fig 4.29 (a) clustering result of the T-squared Statistic based on virtual space

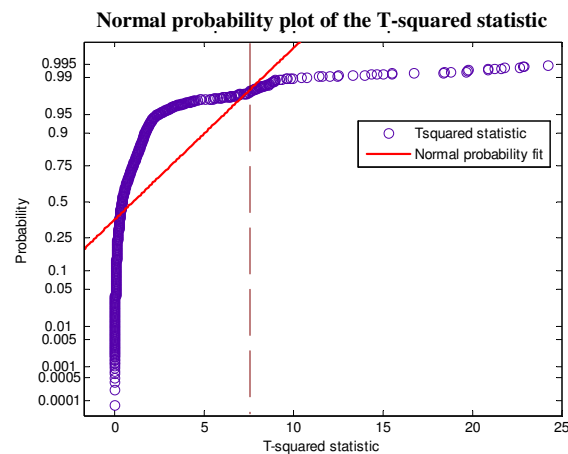


Fig 4.29 (b) normal probability plot of the T-squared statistic

To examine the T-squared statistic, the data points were plotted on the normal probability plot (see Figure 4.29 (b)), and a long tail can be seen in the figure. Hence, the method should be developed to separate these tail points from other types of data. Moreover, the histogram of the T-squared statistic was also plotted, and it was found that the density of the normal data is changing frequently and with high values, but the density of the tail or

abnormal region is almost unified with a very low value. Based on the requirement and observation, the T-squared statistic was modelled by the GMM with the standard EM process, and 7 Gaussian components (clusters) were obtained. It is seen in Figure 4.29 (a) that data in the normal regions are modelled by multiple Gaussian components with higher densities while the anomalies in the tail region are modelled by fewer components with lower densities. Based on this training result, the categories can be set while the changes of the densities, and then the entropy values of the clusters (see Figure 4.20 (b)), can be calculated.

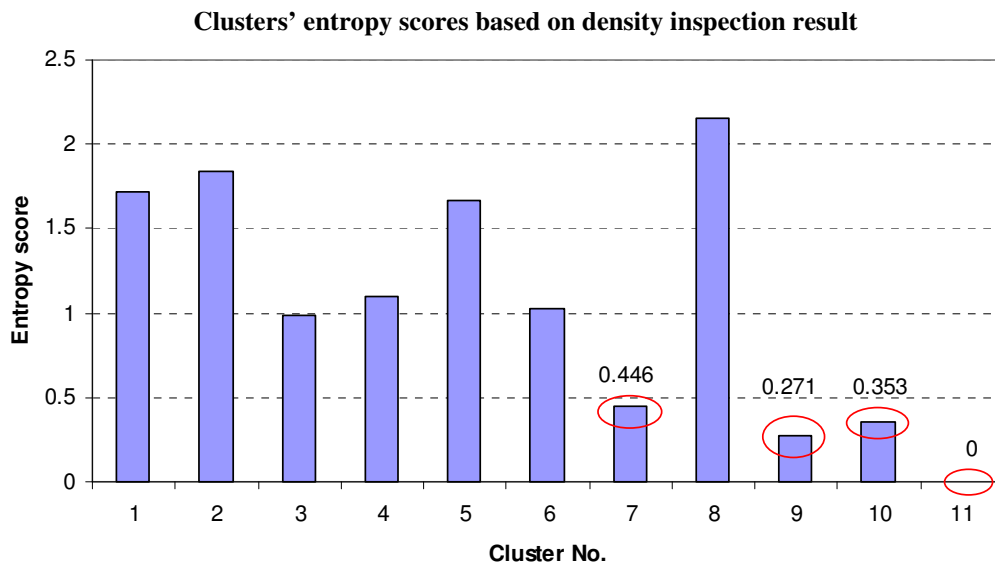


Figure 4.30 Clusters' entropy scores based on the density inspection result

Figure 4.30 illustrates the calculated entropy scores for the clustering space (Figure 4.20 (a)). It is seen that clusters 7, 9, 10 and 11 were found with significantly lower entropy scores than others. Furthermore, in previous analyses, these 4 clusters were found to be associated with anomalies; therefore, the results obtained in this section are successful and consistent with the previous results.

So far, approaches of using the entropy statistics to measure uncertainties have been introduced. Table 4.7 summarizes the entropy scores obtained by both original and optimized methods. The scores of the clusters that are significantly lower or higher than other ones have been highlighted with a red colour. It is found that the clusters 9 and 11

can be identified as the regions associated with anomalies by all the original and optimised methods, but only the methods which assign the categories in the magnitude direction can locate all the 4 clusters (7, 9, 10 and 11) associated with anomalies. Furthermore, the posterior probabilities of all the 11 clusters in the trained GMM (see Figure 4.20 (b)) are also listed in the table and the cluster number is sorted in the ascending order. It is interesting to find that the clusters which are identified as the abnormal regions also have lower posterior probabilities; in other words, these clusters are all supported by fewer cases. The reason for this finding is that in the time series, the anomalies are occurring occasionally, and in most situations, are distributed widely in the magnitude direction. Therefore, the number of the support cases, occasional occurrences and the wide distribution can be investigated cooperatively and can have an impact on the calculation of entropy scores.

Cluster No.	Posterior probability	Pure entropy method	Entropy method in the 'y' direction	Entropy with segmentation	Entropy with density inspection
1	0.257	2.920	0.001	1.734	1.713
2	0.242	2.968	0.129	2.512	1.835
3	0.182	3.444	2.38E-11	2.894	0.990
4	0.076	3.014	0.104	2.121	1.096
5	0.054	3.299	0.135	2.717	1.669
6	0.050	2.046	0.035	1.855	1.024
7	0.048	2.677	2.284	0.582	0.446
8	0.032	2.685	0.112	2.189	2.156
9	0.028	0.022	1.769	0.022	0.270
10	0.019	3.105	2.815	2.869	0.352
11	0.007	0.001	3.215	0.002	0.001

Table 4.7 Posterior probability of the trained clusters and their entropy score with different methods

4.5.4. Application of the proposed approaches to the baseline data

In this section, the demonstrated optimized entropy-based methods as well as the distance based method will be adopted to the combined baseline data that have been introduced in Section 4.2. The examined anomaly detection model is the optimized model that has been generated in Section 4.4.2, as Figure 4.6 (j) illustrates.

4.5.4.1. Entropy statistic associated with multivariate segmentation

Figures 4.31 (b) and (c) show the $A_i(t) = p(t_k|\eta_i)$ Gaussian membership function (see Equation 4.18) and $p(z_k|\eta_i)$ conditional probabilities (see Equation 4.20). It can be seen that with these two parameters, the proposed method described in 4.5.3.1 has segmented the multivariate data into 10 categories. The segmentations were then applied to the in the baseline time series and the results for the vibration RMS data are shown in Figure 4.31 (a). From this figure, it can be seen that some abnormal data are separated into individual categories from the normal data, e.g. data points in the third, sixth and seventh categories.

Each of the cases in the anomaly detection model (shown in Figure 4.6 (j) with 30 clusters) are categorised based on the above multivariate segmentation and the category identification of all cases in each of the 30 clusters are summarised in Figure 4.32 (b). It can be seen that most of the clusters are supported by the cases from more than two categories (shown as more than two colours in the figure). Clusters 13, 14, 17, 18, 20, 21, 22, 24, 27 and 30 have supporting cases from only one category (shown as single colour). In these clusters, nearly all support cases are from a single category (dominant probability value $> 99\%$), which means that the support cases in these clusters are occurring occasionally in the time series and are suspected as anomalies. The entropy scores of each cluster are calculated and shown in Figure 4.32 (a). The entropy scores of the clusters suspected to include anomalies are seen to be significantly lower than that of the others. This is because their support cases are from single category. This also confirms that these clusters are associated with anomalies, and should be removed from the original clustering space.

Figures 4.33 and 4.34 illustrate the anomaly detection models before and after the adaptation, i.e. after the removal of clusters 13, 14, 17, 18, 20-22, 24, 27 and 30. It can be seen that the clusters associated with anomalies that distributed widely and discretely have been removed after the adaptation. Most of the remaining clusters have relatively low variance values and high probability densities. However, two of the clusters (nos. 23 and 28 in Figure 4.34) are not in this category. They are widely spread compared to other

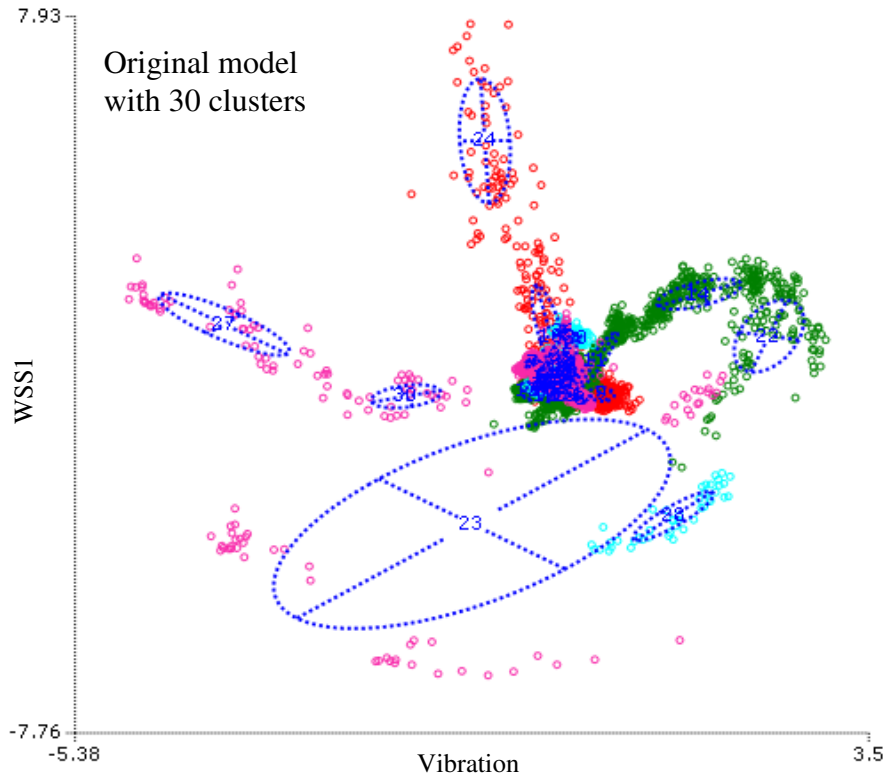


Figure 4.33 Original model with 30 clusters

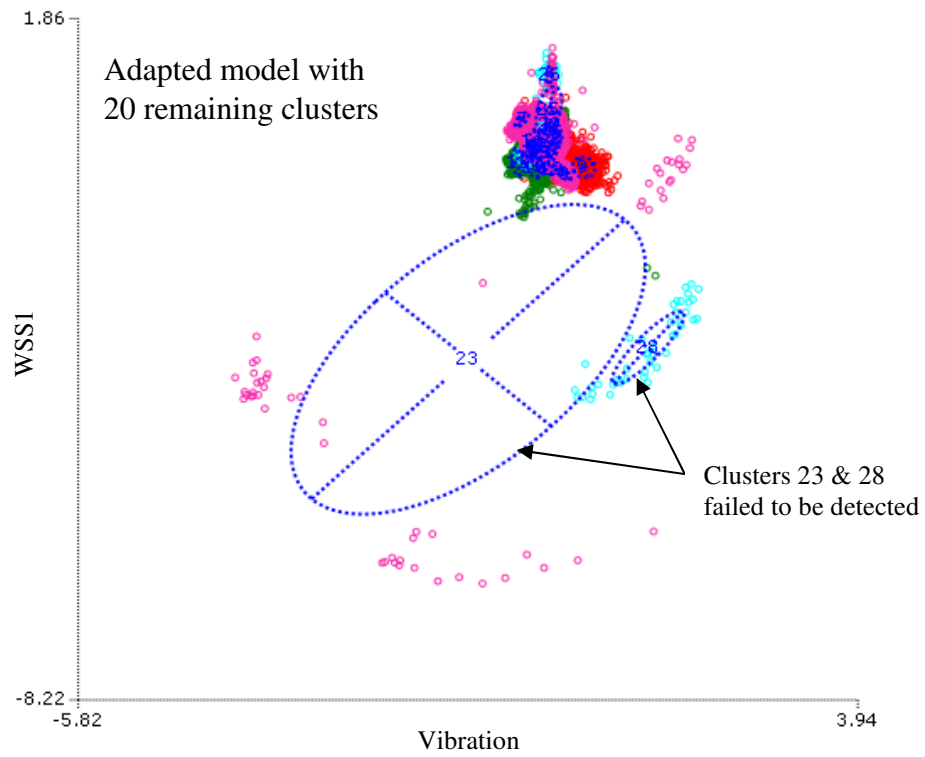


Figure 4.34 Adapted model with 20 remaining clusters

4.5.4.2. Entropy statistic associated with inspection of density changes

It has been found that using the multivariate segmentation approach to adapt the anomaly detection model can detect most of the anomalies and their support regions, but entropy values of clusters 23 and 28 are calculated high and failed to be detected as the abnormal region. Moreover, as mentioned in Section 4.5.3.2, categories can also be set in the magnitude direction, and with this approach, the T-squared statistic of the baseline data (see figure 4.35 (a)) was evaluated by GMM associated with the EM algorithm, and 11 Gaussian components (clusters) were obtained. Hence, 11 categories can be set as the density of the data changes. Figure 4.35 (b) illustrates the categorization result. It is seen that the outliers between 5000 and 6000 data points only occupy small number of categories while the normal data throughout the test spreads into more frequently changed categories.

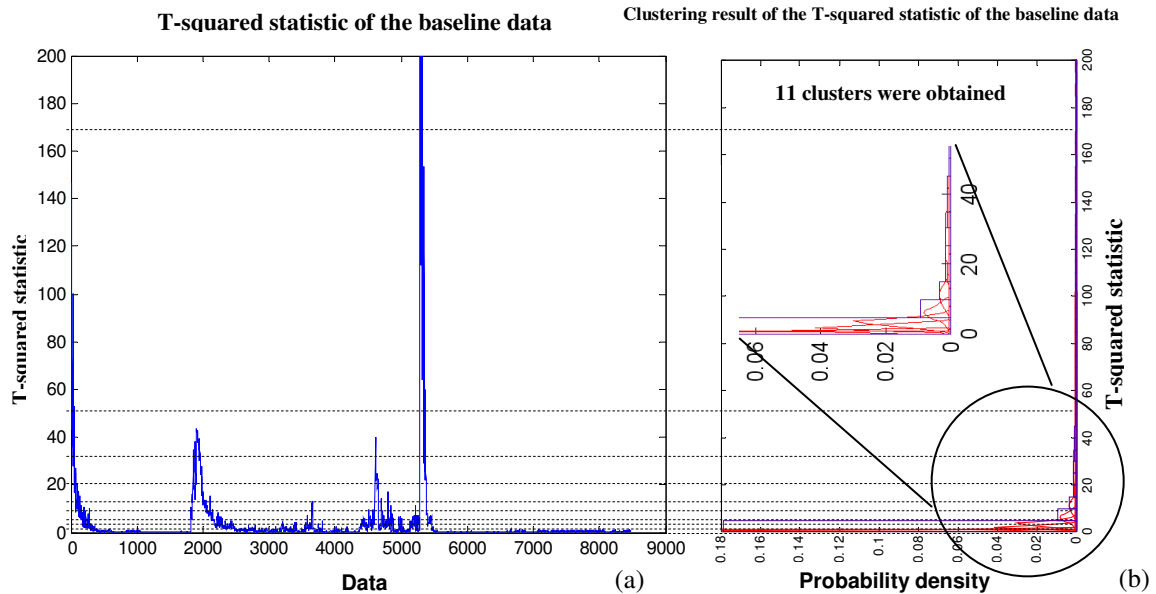


Figure 4.35 T-squared statistic of the baseline data (a) and the trained 11 clusters (b)

Figure 4.36 shows the calculated entropy scores for the original clustering space (Figure 4.33). It is found that clusters 23 and 28 have lower entropy values, and these two clusters are the ones that have not been detected in the results which were obtained from the optimized time series segmentation method. The success of detecting clusters 23 and 28 proves that with the density changing method, the anomalies that are widely distributed in

the magnitude direction can be detected, and it is a necessary implementation to the time series segmentation approach.

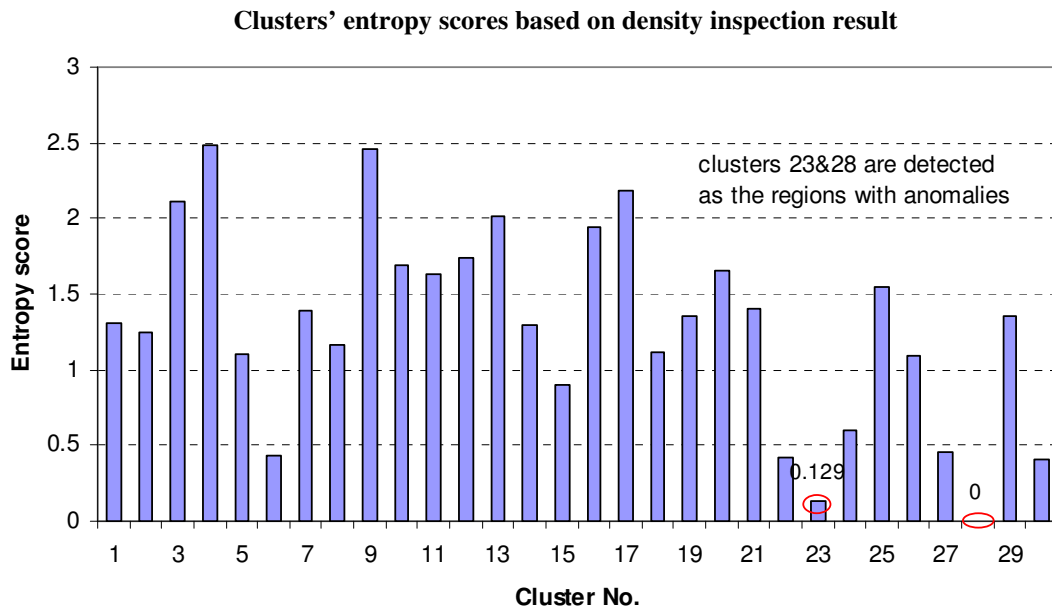


Figure 4.36 Entropy scores of the clusters based on the density inspection result

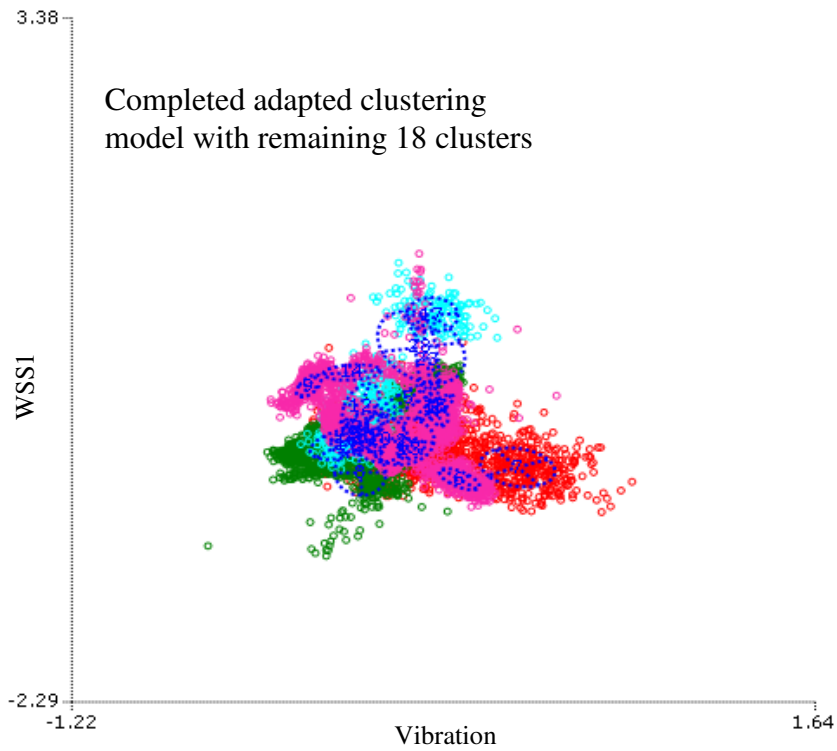


Figure 4.37 Adapted model with 18 remaining clusters

Figure 4.37 shows the adapted anomaly detection model after the removal of clusters that have been detected by both the segmentation and density changing methods. It is found that the remaining clusters have relatively low variance values as well as high probability densities.

4.5.4.3. T-squared distance between the clusters

Figure 4.38 (a) illustrates the calculated result. It can be seen that cluster 23 is recognized as being distant from the other clusters, so the support cases associated with these two clusters are identified as anomalies. Figure 4.38 (b) shows the calculated cluster distance after the removal of cluster 23. It is seen that the distances of clusters 14, 24 and 27 to the other ones are higher. Therefore, clusters 14, 23, 24 and 27 are defined as abnormal regions. This discovery is found to be consistent with the findings of the entropy-based method.

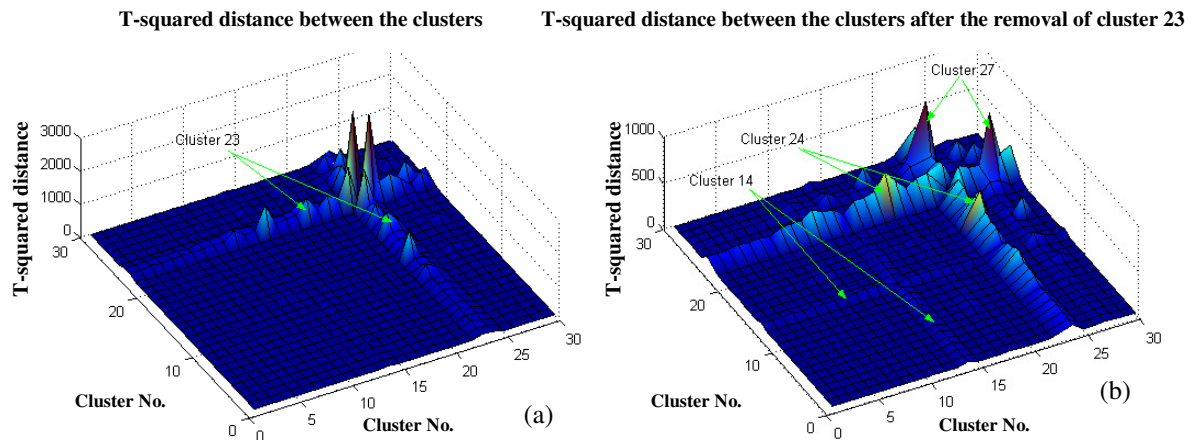


Figure 4.38 Hotelling's T-squared distance between the clusters (a) and the distance after the removal of cluster 23 (b)

4.5.5. Effect of training parameters on the adaptation of the anomaly detection models

It has been discussed in Section 4.4 that the GMM training parameters have an impact on the distribution of the clustering space as well as the model adaptation. Therefore, in this section, the influence of training parameters on the model adaptation will be discussed, and the anomaly detection rate (true positive rate) and false alarm rate are the evaluation indices.

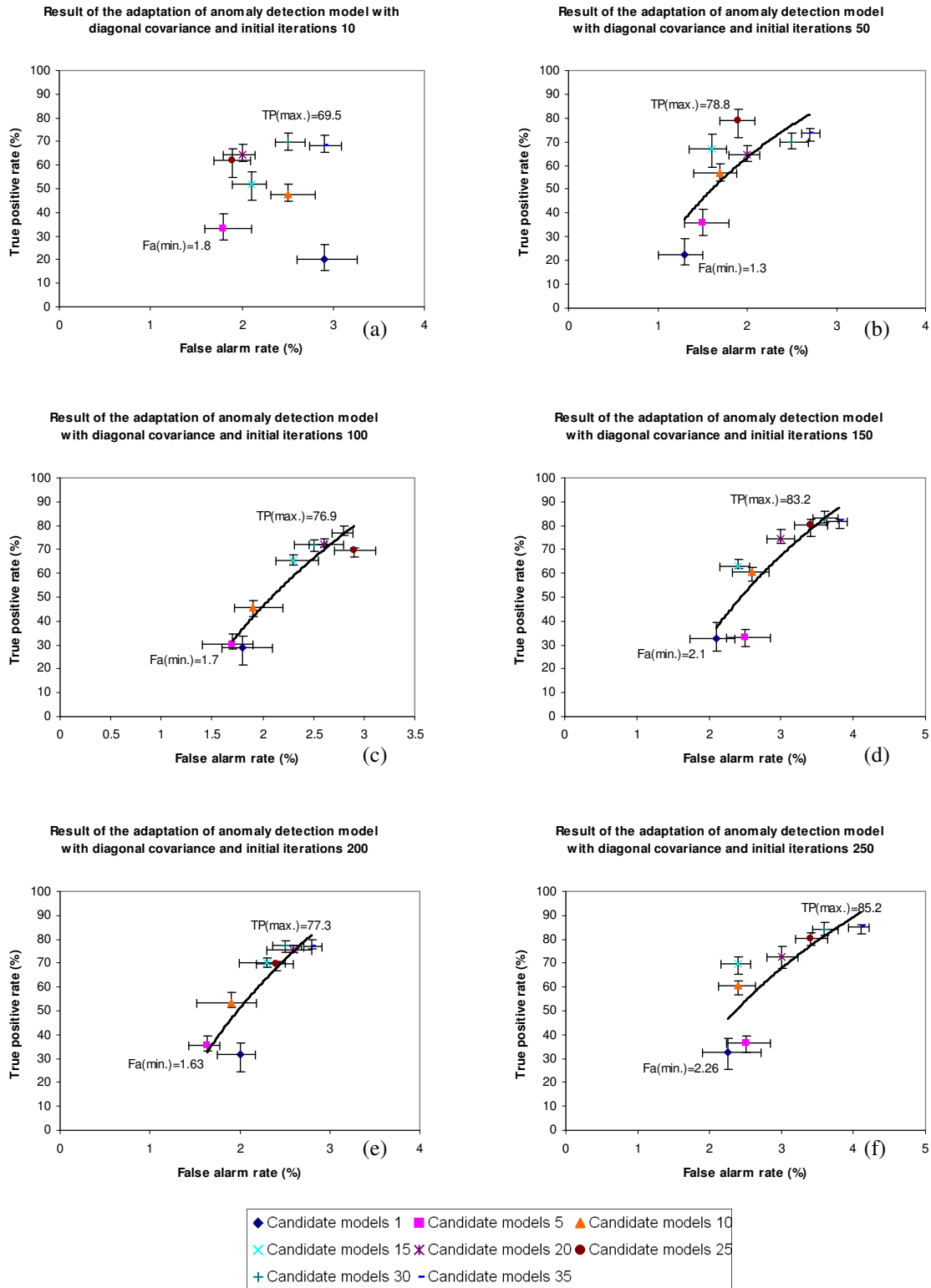


Figure 4.39 Anomaly detection rate versus false alarm rate with diagonal covariance structure

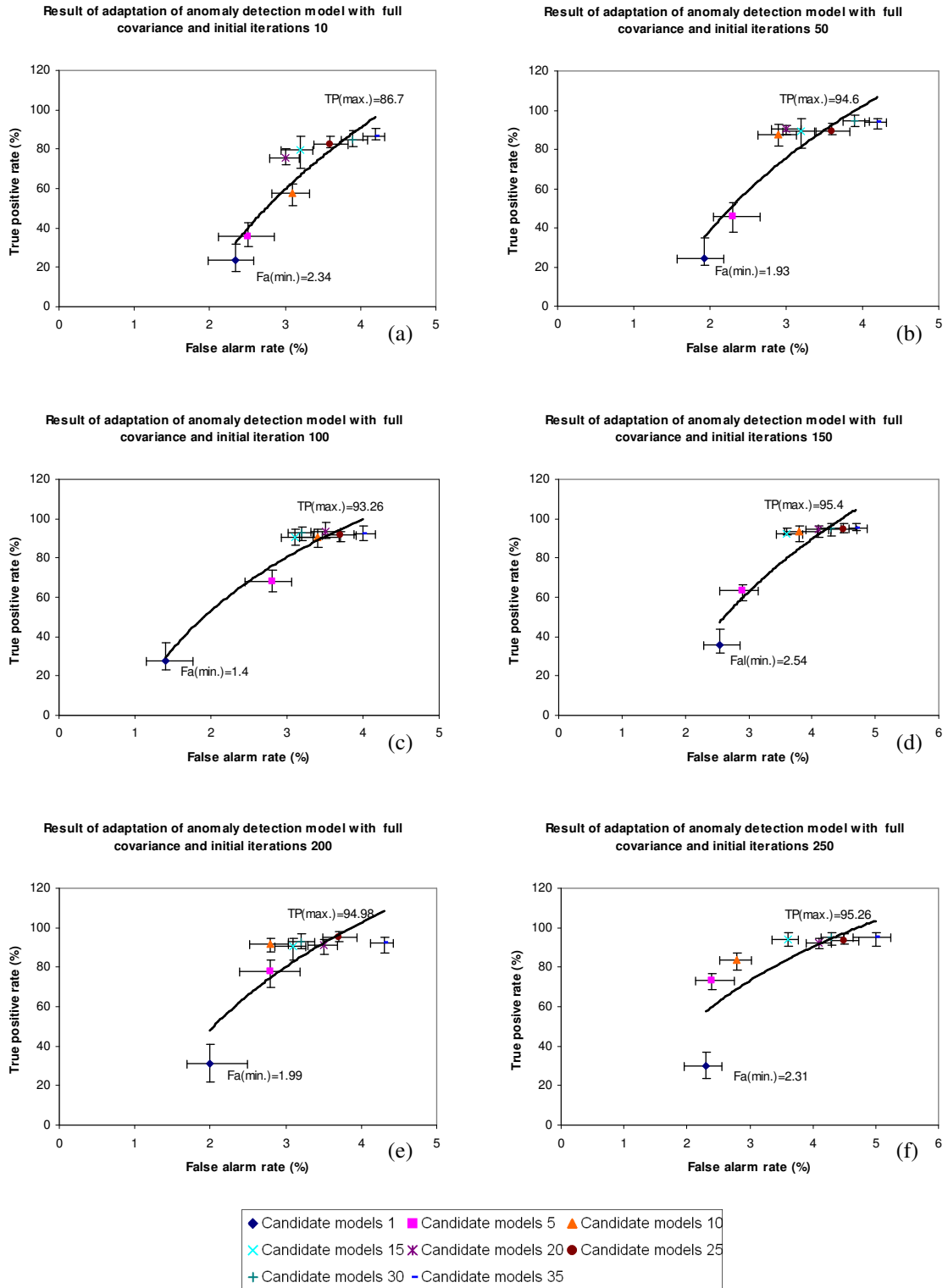


Figure 4.40 Anomaly detection rate versus false alarm rate with full covariance structure

Figures 4.39 and 4.40 illustrate the figures of anomaly detection (true positive) rate versus false alarm rate with the changes of the number of candidate models and initial iterations, under the diagonal and full covariance structure. The anomaly detection rates generally show a trend where the parameter of the number of candidate models increases for almost all the cases with different initial iterations and covariance structure, which proves that increasing the number of candidate models is crucial for the model adaptation. However, false alarm rates which are desired to be reduced are also increased with a low and accepted level. Furthermore, the parameter of initial iteration seems to have insignificant impact on the evaluation indices, but the full covariance models are found to lead the higher detection and false alarm rate compared to that of diagonal models.

4.5.6. Summary

In this chapter, the approaches to adapt the anomaly detection model have been introduced. Generally, the approaches can be divided into entropy-based method and distance-based method.

Method	description	Advantages	Disadvantages
Pure entropy method in the time series	categories are assigned to the time series to see the consisted elements in the clusters are from single category or not	The anomalies which are occasionally appeared in the time series could be detected	The categories are assigned manually , and the results are dependant on the categories
Pure entropy method in the magnitude direction	Multivariate are fused by the T-squared distance, and the categories are assigned in the magnitude direction.	The anomalies which are wide spread could be detected, as they normally occupy more categories.	The categories are assigned manually.
Entropy method associated with multivariate segmentation in the time series	Clustering conditioned to the data point ID is used to automatically assign the categories in the time series.	Categories in the time series are assigned in a more intelligent way.	Conditional clustering involves more calculations and justification.
Entropy method associated with density inspection in the magnitude direction	Density inspection is conducted to the fused T-squared statistic, and the categories in the magnitude direction could be set when the density changes.	Categories in the magnitude direction are assigned in a more intelligent way.	Clustering is involved again when the density inspection is conducted.
Distance between the clusters	Distance between the clusters are calculated, and the clusters which are distant from others are identified as abnormal.	Simple method, less computation	Difference between the normal and abnormal clusters are not always obvious.

Table 4.8 Summary of the proposed approach to adapt the anomaly detection model

For the entropy-based method, the main idea is to explore the components contained in each trained cluster, and examine whether they are from occasional or multiple pre-defined equal categories in both time and magnitude directions; this could be achieved by calculating the probabilities of data from each category in a particular cluster, and this will ultimately be indicated by the entropy statistic. Based on this original idea, the optimized methods were also developed to automatically assign the categories in both time and magnitude directions, and this showed more satisfied results. Furthermore, the distance-based method is an important implementation to examine the distant clusters, but this method needs to be further developed, because the distance threshold for identifying distant clusters is unclear at the time of writing. These methods have been summarized in Table 4.8. It should be noted that the methods have their unique characteristics and should be applied cooperatively to achieve the most reasonable results.

4.6. Updating of Anomaly Detection Model

Although the adaptation of the anomaly detection model could minimize the risk of missing the wanted anomalies, the modified model is still not robust enough, because of the unexpected rotating speed or load changes which could affect the background level of the testing data. Therefore, the constructed anomaly detection model would become incomparable to the new observed data with a different range. To solve this problem, several researches have been conducted to incrementally add new Gaussian components to update the original GMM space [10, 11]. In the anomaly detection task, the new Gaussian components could be associated with some new observed data. The difficulty is that the selection of appropriate new observed data for mixture model updating is blind and unwanted data such as running-in or faulty data may be included. Therefore, it is necessary to develop an intelligent multivariate method to choose appropriate data for model updating. Appendix A3 shows the general procedure of model updating.

4.6.1. Approach to discover the difference between the two groups of data

It has been recognized that the data used to update the anomaly detection model is desired to be the ones representing the steady state of the machinery. Thus, the developed method

should be capable of identifying the normal data from the other type of points. In this section, the approach to discover the difference between the two groups of data will be introduced first, before the demonstration of how to update the model is described in Section 4.6.2.

4.6.1.1. Static clustering

To demonstrate the proposed approach, the data from test 16 was selected, and in this section, the extracted features from the three types of sensors (Vibration, WSS and OLS) are used as inputs into the clustering model to examine the behaviour of the features during the different running stages. From the conventional signal plots, three distinctive wear stages - running-in, steady state and wear-out - could be identified; samples were collected from these stages to see if they could be grouped separately.

It is assumed that the bearing wear conditions are varied at different running stages, and are independent of each other. 300 samples were randomly collected between 30~32.5, 57.5~60 and 60~62.5 hours from each sensor response. This preliminary test was not intended to detect any specific type of wear phenomena, but to be used to validate the ability of the clustering model to classify the different running stages.

Figures 4.41 and 4.42 illustrate the clustering results and their projections with the six vibration and WSS2 features. From the figures, it could be found that both vibration and WSS2 features have three trained clusters, and these clusters have their corresponding support cases. To define these clusters, three groups of parameters - Mean Vector, Covariance Matrix and Mixing Coefficients - could be used. However, these parameters are not appropriate if the study purpose is to explore whether the samples collected at different running stages could be grouped separately by trained clusters. Therefore, it is better to further investigate the purity in each of the trained clusters, or in other words, to find the sources of the data points in each cluster.

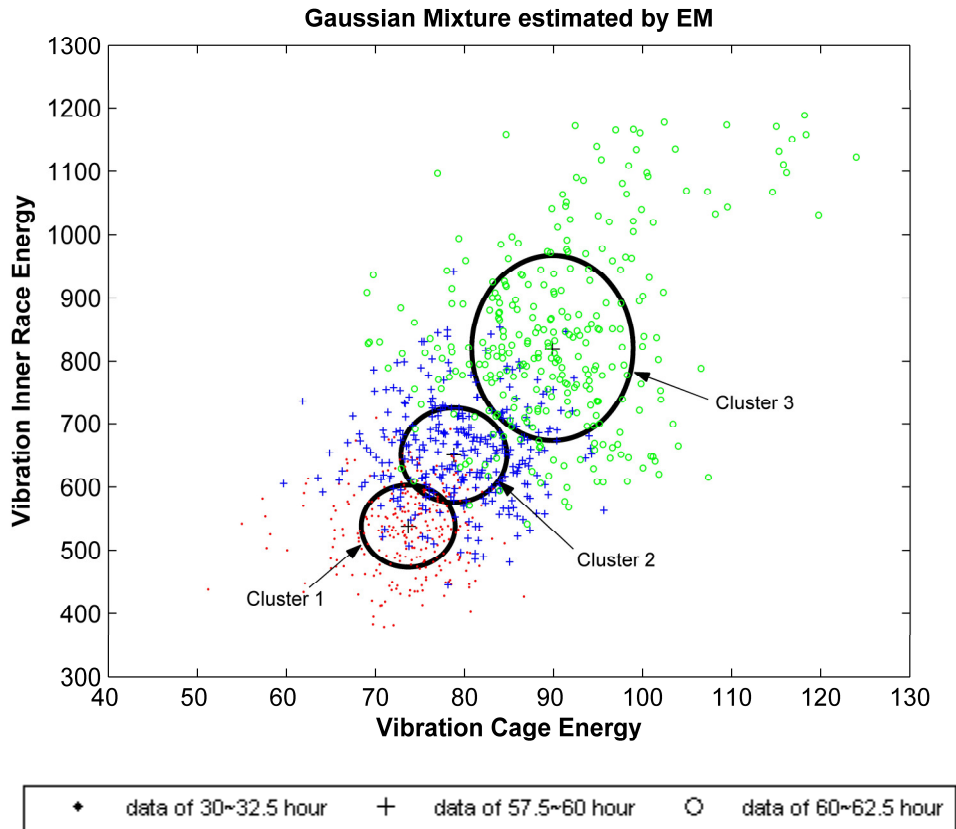
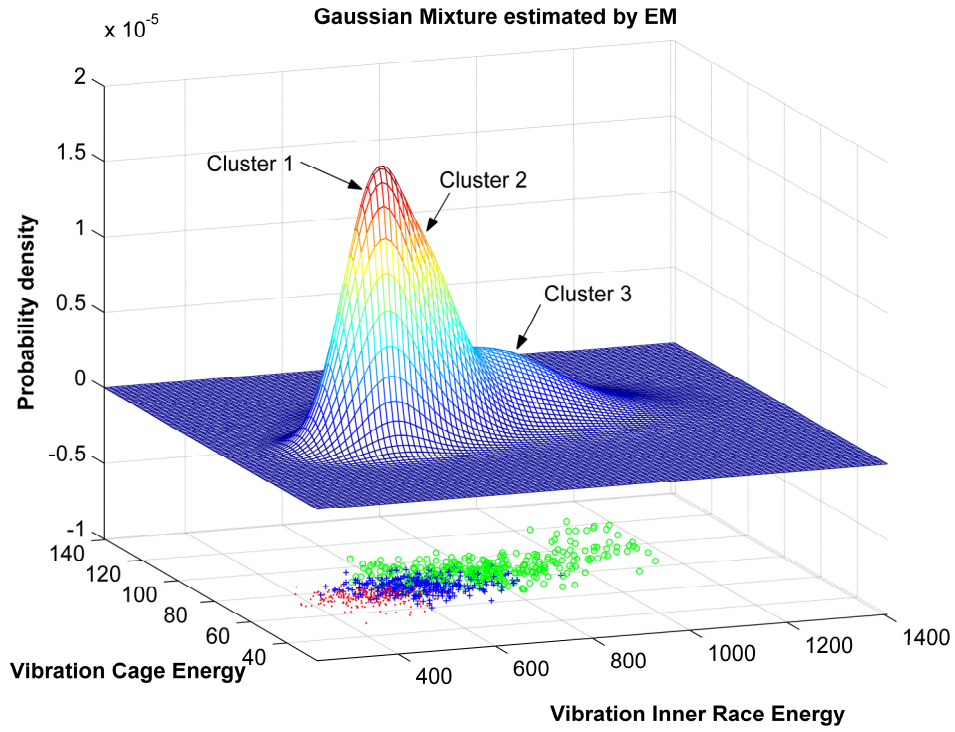


Figure 4.41 Result of vibration clustering analysis

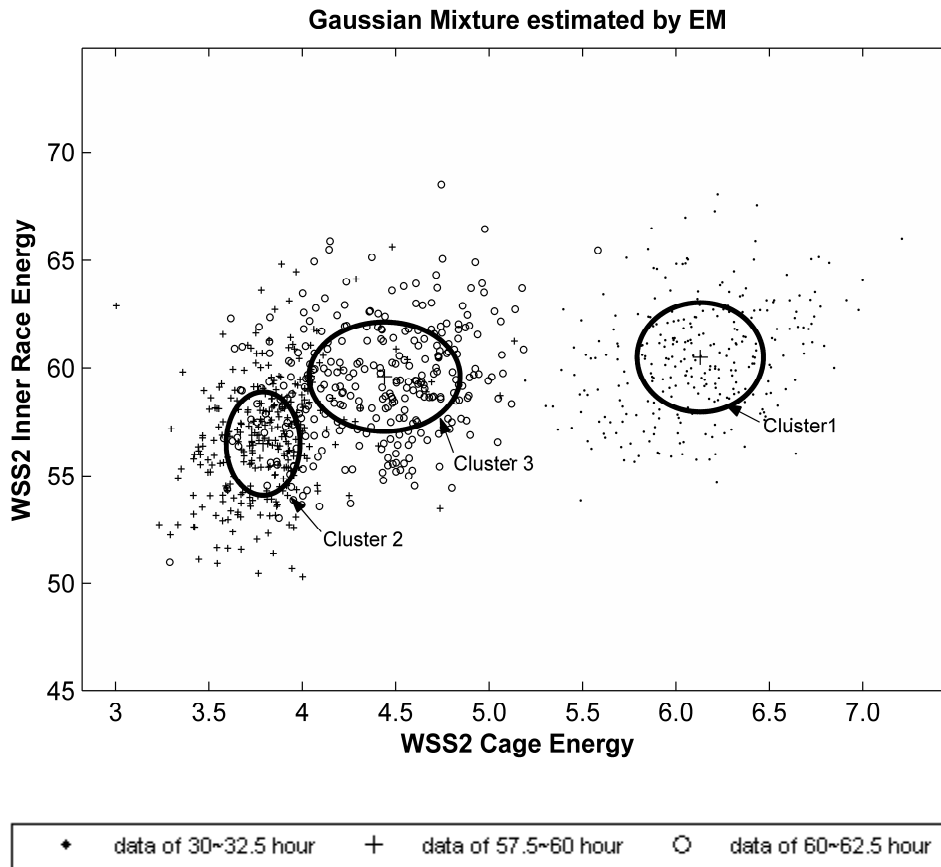
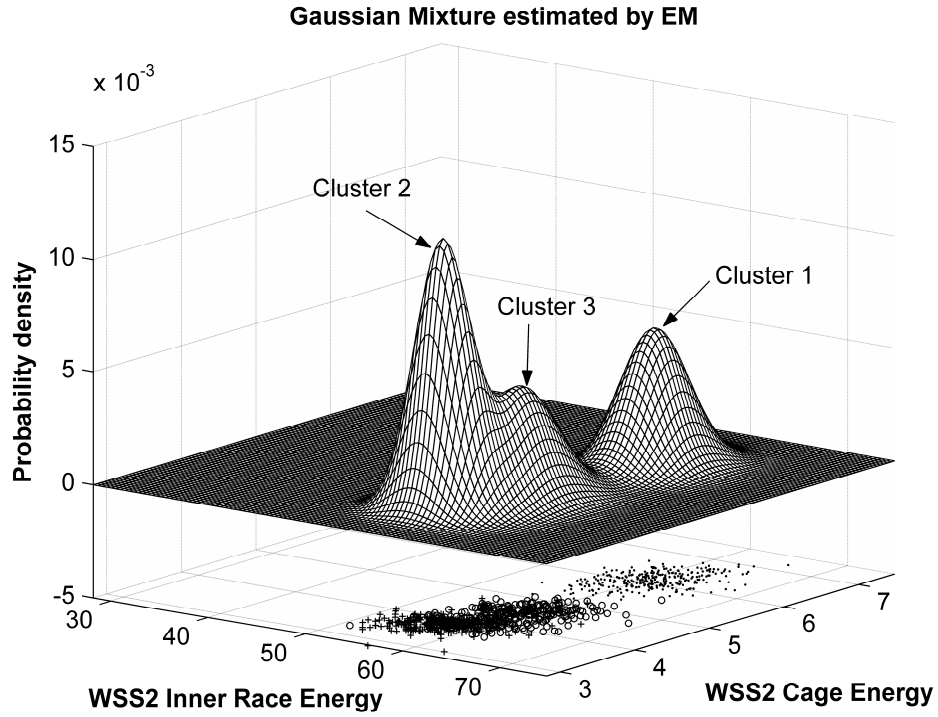


Figure 4.42 Result of WSS2 clustering analysis

Table 4.9 shows the results of cluster purity exploration for the three types of applied sensors. The numbers outside the brackets represent the support cases coming from each of the sampling periods, while the numbers in the brackets are the occupation percentages for the support cases. For the vibration sensor, which is discovered in each of the clusters, there is a major occupation component. In cluster 1, all the samples from 30~32.5 hours are grouped together, and the occupation percentages of 57.5~60 hours and 60~62.5 hours are very low at 0.16% and 0.04% respectively. This could be explained by the fact that the bearing condition at 30~32.5 hours is different from the ones at the other two running stages. In cluster 2, the purity approaches 100%, supported by most of the samples of 57.5~60 hours. In cluster 3, samples of 60~62.5 hours are dominant, reaching 87.4%, but there are also 43 support cases from 57.5~60 hour; this is because the two sampling times are together, and some samples are not easily distinguished.

For the WSS2, the exploration result is similar to that of the vibration sensor. In each of the trained clusters, there is a significant major occupation class. Nevertheless, the purity investigation of the WSS1, WSS3 and OLS is different from that of the vibration and WSS2. From Table 4.5, it is seen that there are 2 trained clusters for both WSS1 and WSS3, and in each of the clusters, there is no dominant occupation class and the number of constituent components are very close to each other. A similar result of inseparability is also found from the purity search of the OLS, although 4 clusters are obtained after the training.

From the above analysis, three running stages can be identified from using the clustering method based on the vibration and WSS2 features. Thus, it is essential to analyze the signals from both vibration and WSS2 sensors with clustering analysis. However, the clustering result seems not to be satisfied if the samples are from WSS1, WSS3 and OLS. This may be due to the samples being collected randomly and some characteristic features might have been overlooked. Under these circumstances, it is necessary to investigate the data from these sensors at continuous (from the beginning till the end) running stages to detect any wear condition by clustering.

Purity Search	30-32.5 Hours	57.5-60 Hour	60-62.5 Hour
Cluster1(Vib.)	300 (99.8%)	0.365(0.16%)	0.003(0.04%)
Cluster2(Vib.)	0.04(0.06%)	257(99.9%)	0.004(0.04%)
Cluster3(Vib.)	0.22(0.1%)	42.9(12.5%)	300(87.4%)
Cluster1(WSS1)	142.7 (31.6%)	161.2(35.7%)	147.5 (32.7%)
Cluster2(WSS1)	157.3 (35.1%)	138.8 (30.9%)	152.5 (34.0%)
Cluster1(WSS3)	160.3 (35.3%)	153.6 (33.9%)	139.8 (30.8%)
Cluster2(WSS3)	139.7 (31.3%)	146.4 (32.8%)	160.2 (35.9%)
Cluster1(WSS2)	293(99.3%)	0.108(0.05%)	1.81(0.65%)
Cluster2(WSS2)	0(0%)	235(81%)	59.6(19%)
Cluster3(WSS2)	7.43(2.4%)	65.3(21%)	239(76.6%)
Cluster1(OLS)	81(35.8%)	68(30%)	77(34.2%)
Cluster2(OLS)	74(34.7%)	70(32.9%)	69(32.4%)
Cluster3(OLS)	76(32.3%)	79(33.6%)	80(34.1%)
Cluster4(OLS)	69(30.5%)	83(36.7%)	74(32.8%)

Table 4.9 Result of cluster purity exploration

4.6.1.2. Dynamic difference detection process

In this section, a dynamic difference detection process is simulated to detect any new conditions in data during the test. Clustering is applied to base-line data from a known healthy bearing test (test 9) and data from continuous running stages (test 16) to see if the newly sampled data could be grouped separately from the known healthy data during the training. In this approach, although the training data are from two groups, i.e. base-line and newly sampled, the Bayesian Information Criterion (BIC) controls the number of training clusters. Furthermore, the number of base-line data and newly sampled data at different running stages are both set to 300.

This method first randomly samples a specified number of points with a fixed length T from the time series. Next, each sampled point, originally in the time domain, is transformed into the feature space by calculating a set of feature values. At the same time, the prepared base-line feature sets representing the healthy condition of the bearing are mapped into the feature space as well. Thirdly, the feature space is clustered by a mixture-model method to examine whether the sampled features are abnormal and could

be grouped separately from the base-line features by wear. Finally, the results of clustering are explained according to the property of the particular sensing technologies. Figure 4.43 illustrates the working procedure of the clustering-based bearing wear detection technique.

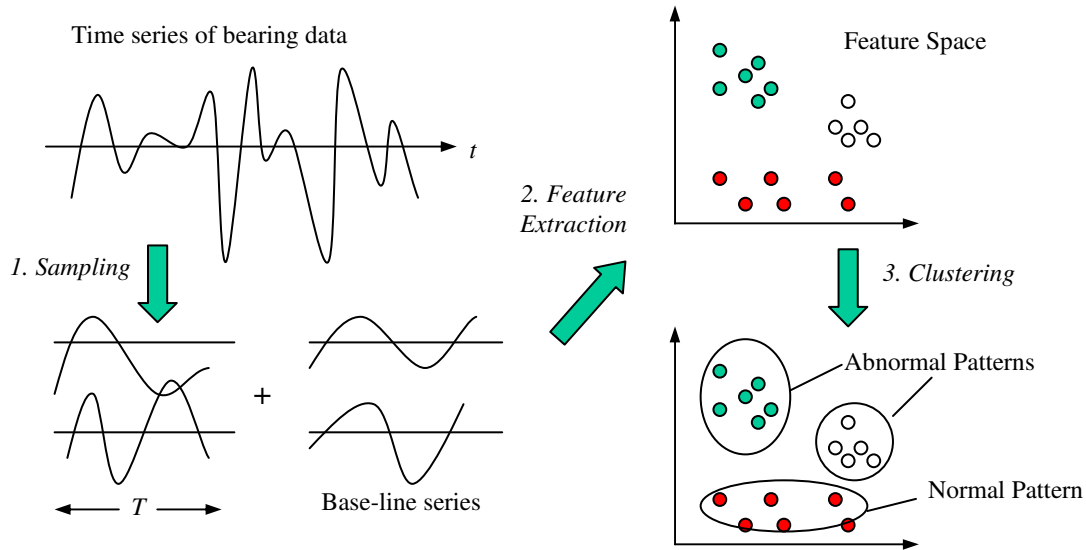


Figure 4.43 Working procedure of clustering-based wear detection

(1) Results

Figure 4.44 illustrates a few examples of the dynamic clustering process with the six vibration features. From the figure, it can be seen that the new data are inseparable from the base-line data between 40~55 hours (a-f). From hour 57.5, most of the newly inputted data starts to form a separate cluster. Furthermore, when the new data are collected at 60~62.5 hours and 62.5~65 hours, very clear new clusters are created, and new data are further grouped into two more clusters. Another interesting phenomenon is that the number of the trained clusters at each of the sampling stages is constant at 2, until 60~65 hours when the number of the trained cluster increases to 3 and 7.

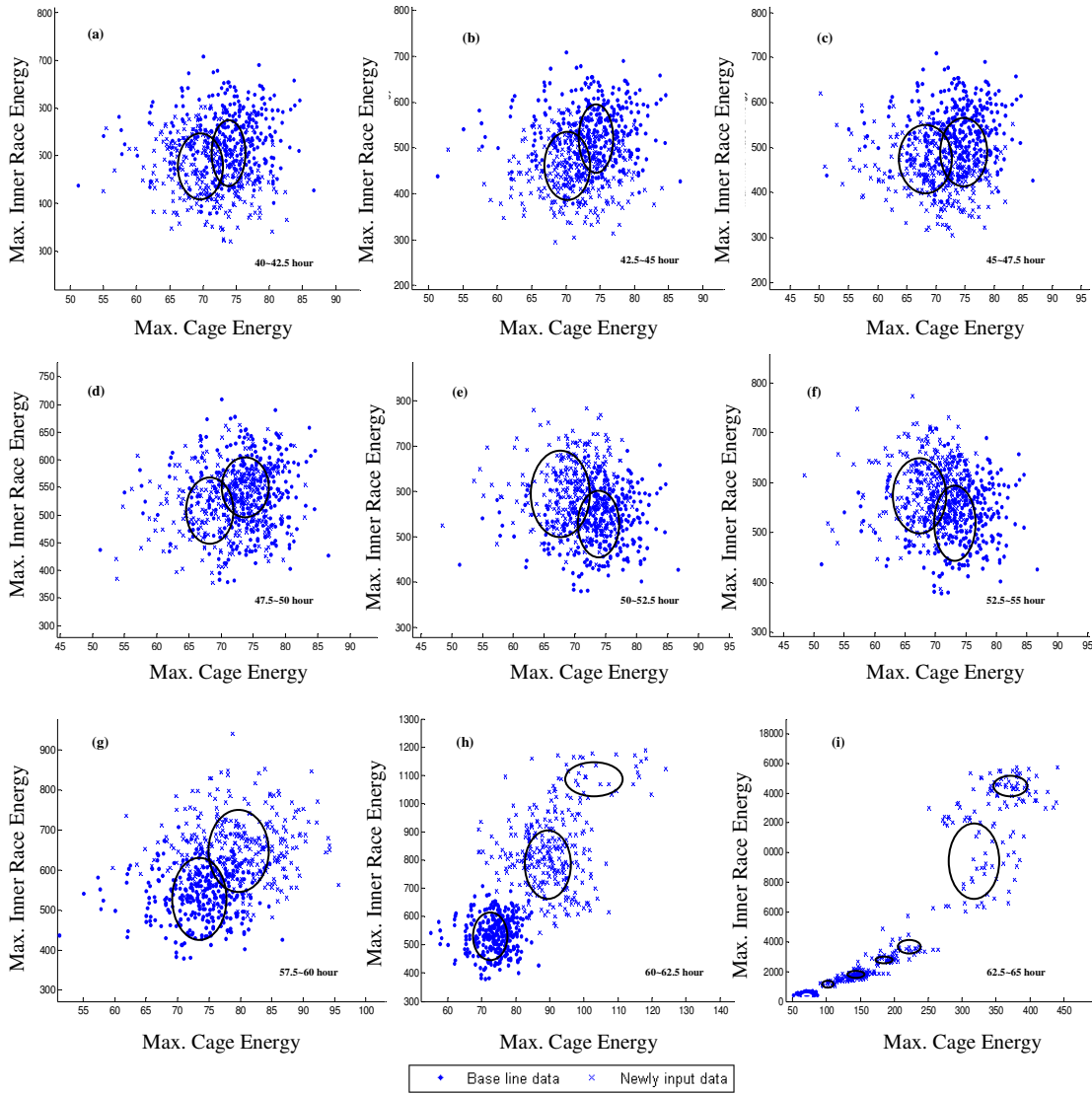


Figure 4.44 Dynamic clustering process of vibration features (Figures 5(a-i) illustrate the clustering examples of vibration features at a particular sampling period)

A similar approach was applied to the WSS2 features (see Figure 4.45). In the first three examples, the new data could not be independently grouped from the base-line data. But in the rest of the trained samples, the new data are nearly separated from the base-line data, and the occupation probabilities of the new data in the clusters reach above 95%. Besides that, the distributions of the three clustering samples are found to be significantly different from the other trained results when they are recorded at 47.5~50 (d), 50~52.5 (e) and 57.5~60 (g) hours. These three trained examples all show a very discrete characteristic, and the number of trained clusters is higher than other samples.

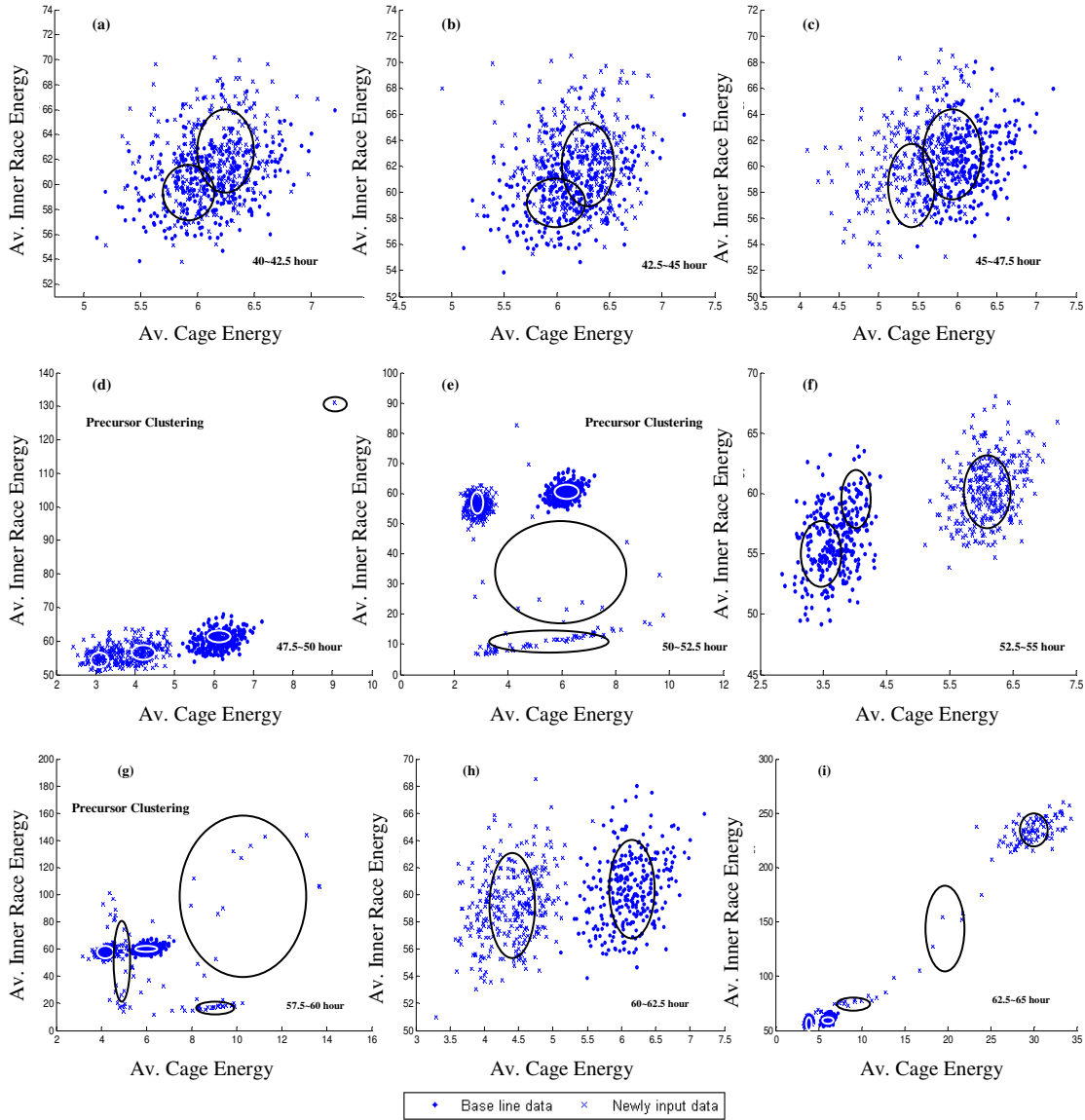


Figure 4.45 Dynamic clustering process of WSS2 features (Figures 6 (a-i) illustrate the clustering examples of WSS2 features at a particular sampling period)

In order to make it clear, the original plots of WSS2 features were traced, and the discrete charge events were frequently found between 48~60 hours. Figure 4.46 shows the zoomed-in picture of the WSS2 inner race energy, and the discrete charge could be clearly identified. To explain this interesting phenomenon, the link between wear in the oil lubricated contact and electrostatic charge mechanisms needs to be understood. This link has been indicated elsewhere [12-14] and the discrete signals which have also been found in figures 4.45(d), (e) and (g) are caused by the surface phase transformation associated with the first transition scuffing, which hence leads to the onset of failure. The

presence of surface-transformed regions will result in contact potential differences (CPD) and generate discrete electrostatic precursor signals prior to the severe adhesive wear (scuffing) which can lead to catastrophic failure. However, rolling element bearings do not suffer the adhesive wear, and the causes of these discrete signals could not be explained by surface phase transformation with full confidence; they might be the result of the generated wear debris or some other sources. Therefore, the reasons for the uncommon clustering result with the samples at 47.5~50 (d), 50~52.5 (e) and 57.5~60 (g) hours is not very clear, but the occurrence of these discrete charge events which was identified with clustering is still invaluable, as they could be used to detect some abnormal conditions, and provide a precursor to severe wear. Here, we call them the precursor cluster.

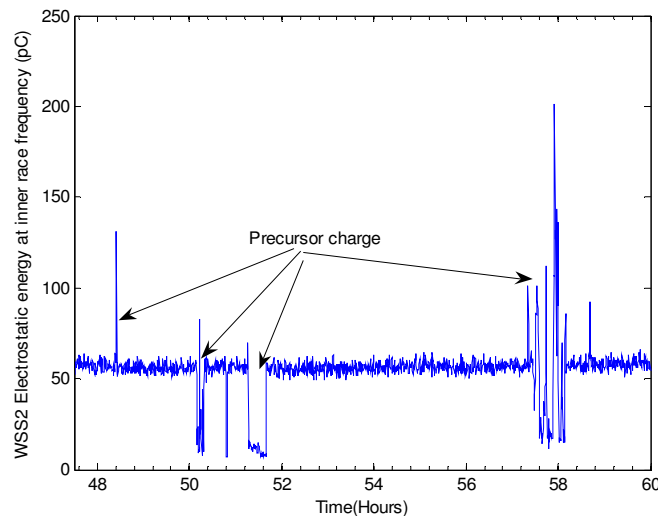


Figure 4.46 WSS2 precursor signal of the inner race energy

The dynamic clustering for the electrostatic oil line sensor features showed no separation from the base-line throughout the test until 62.5~65 hours. The occupation percentage of new data in the clusters suddenly increased from around 50% to 83%~100%. It has also been found that the number of clusters varies between 4 and 5 in the whole test. Figure 4.47 shows two clustering examples of the OLS features before and after hour 62.5 respectively.

For the three sensors analyzed above, the dynamic clustering process has shown that two parameters are important, the occupation percentage of the newly sampled data in the clusters (OP) and the number of trained clusters (NC), to vary with the sampling time. Thus, these two parameters are valuable tools with the potential for bearing fault detection.

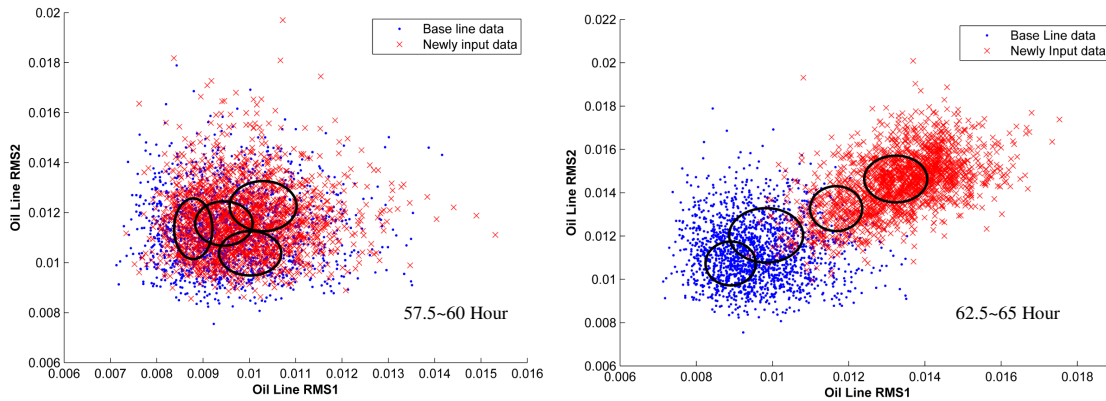


Figure 4.47 Samples of OLS dynamic clustering process

(2) Discussion

Figure 4.48 shows the occupation probability of new data in the clusters that new data occupy, most starting from 35 hours for all sensors. The vibration features show a steady increasing trend: all the OP values are below 60%, until it is calculated at 55~57.5 hours, when it reaches 69%, and then, the OP values keep increasing until it is 100% at 62.5~65 hours. The new vibration data started to become grouped independently at 55~57.5 hours, which means the wear condition is firstly detected by vibration features at 55~57.5 hours. As the sampling time passes, the more severe wear conditions which have the corresponding OP values (e.g. 57.5~60 hours with 78%, 60~62.5 hours with 95% and 62.5~65 hours 100%) could also be detected.

For the OP trend of WSS2 features, the value suddenly increased from 68% to nearly 100% at 47.5~50 hours. And then, the OP keeps steady with a value of around 100%. It is found that the new WSS2 data could be grouped independently and earlier than the vibration data, which could be described as the precursor processes, and detected by WSS2 sensor, while the vibration could only detect the severe wear condition when the cracks occur on the contact surface.

The trend of OP for the oil line sensor is easier to explain, as the oil line sensor was installed in the recirculation pipework, and therefore it is not sensitive to the contact surface changes. It is designed for sensing discrete wear debris passing through the sensor, hence the OP values keep low for most of the time, although the contact surface has been greatly modified by wear. At the sampling time of 62.5~65 hours, the OP value suddenly reaches 83%~100%, as spalling occurs causing massive amounts of wear debris to pass through the oil line sensor.

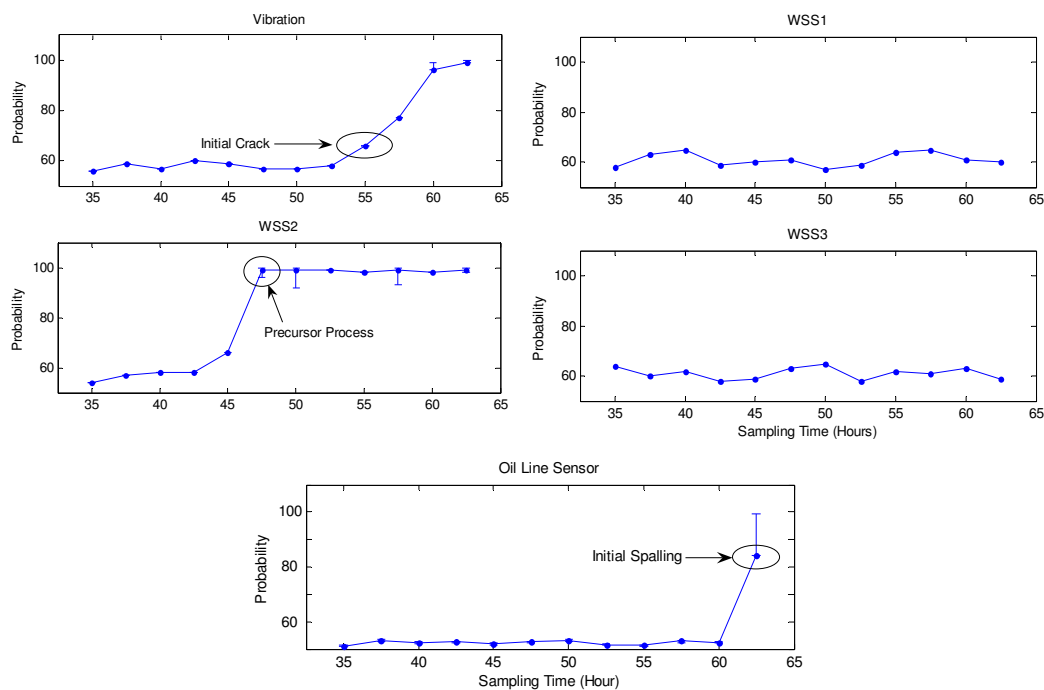


Figure 4.48 Occupation percentage of the new data in the cluster that newly input data occupies most

The OP trend of WSS1 and WSS3 are also illustrated, and the flat trends were found throughout the test. As the WSS1 and WSS3 were utilized to monitor the two support bearings (bearing 1&4), the reason for this phenomenon might be explained by the fact that two support bearings were not experiencing the severe wear in the whole testing procedure, and the OP parameter did not detect any abnormal conditions.

Another interesting clustering parameter is the number of the trained clusters (NC), which might be used for wear detection. Figure 4.49 shows the trends of NC for the 3 types of

sensors. For vibration, differences in NC only occur when the bearing contact surface enters into the severe wear condition. However, for the WSS2 NC trend, four significant different values were found at 47.5~50, 50~52.5, 57.5~60 and 62.5~65 hours. The clustering results with the samples at the first three sampling periods are called precursor clusters. Hence, the WSS2 NC values could be used to find out if the precursor signals exist, so that the severe wear could be pre-warned by the precursor NC values. The NC trend for the oil line sensor does not show any distinguishing features. So the examination of the NC value is of little value for the oil line sensor. The NC values of WSS1 and WSS3 showed a consistent flat trend with the OP trend, which enhanced the confidence that the two support bearings were not severely damaged.

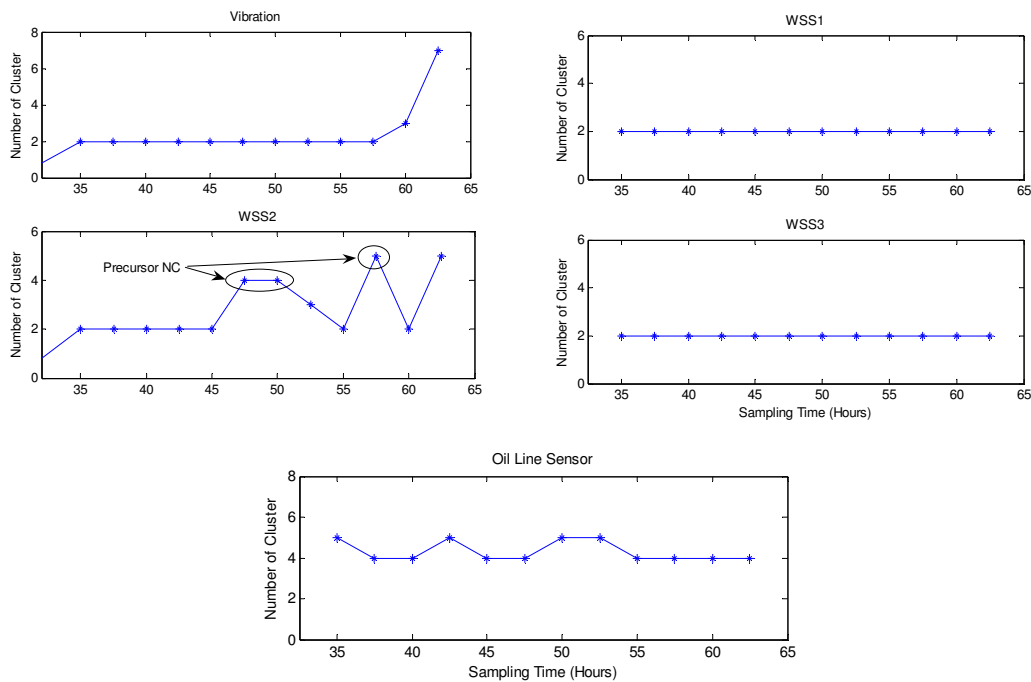
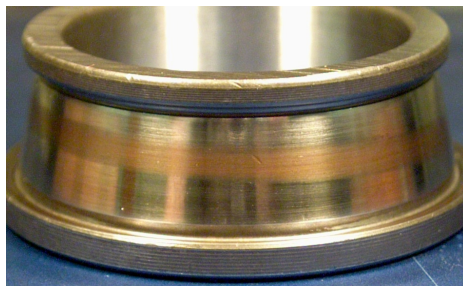


Figure 4.49 Number of the trained cluster

In order to validate the above analysis, the pictorial evidence of the inner race damage is shown for all four bearings. In Figure 4.50, the photographs clearly show that the failure of the bearings during the test originated from the inner race of bearing 2. That is the reason that the precursor process and precursor NC values could be found from the OP and NC trend of WSS2. The figure also indicates that the fatigue damage (spalling) occurred to the inner race of bearing 2, associated with high OP values of vibration and

OLS at the end of the test. However, no perceivable damages were found on the inner races of bearings 1 and 4, which proves the assumption that there is no significant damage in these two bearings with a flat trend of OP and NC values.

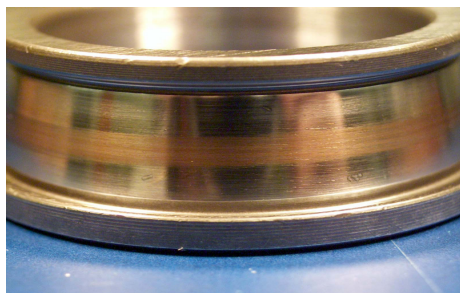
From the above analysis, the clustering parameters of OP and NC are both valuable for wear detection with particular sensing technologies. OP is useful for all the three sensors, and WSSs shows the advantage with clustering OP value to detect precursor processes that indicate contact distresses and wear, and shows that OP values with vibration features are sensitive to relatively severe conditions, and for the electrostatic oil line sensor, OP values could only detect the occurrence of severe wear (spalling) associated with massive amounts of wear debris. However, the parameter of NC seems only meaningful for WSS features, as it could be used to find and describe the electrostatic precursor signals that prove the advantage of the electrostatic wear-site sensor.



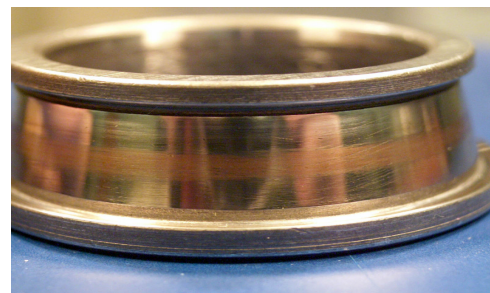
Bearing 1 (WSS1)



Bearing 2 (WSS2)



Bearing 3 (WSS2)



Bearing 4 (WSS3)

Figure 4.50 Inner race examination of the four bearings

4.6.2. Approaches to select appropriate dataset to update the anomaly detection model

Having confirmed that occupation probability and the number of the trained model are useful to identify the difference between two groups of data, this section will demonstrate how to use these two parameters to select appropriate data which are used to update the anomaly detection model. Figure 4.51 (a) shows the simulated two-dimensional data which consist of typical features observed in the bearing run-to-failure test. In the dataset, the features of running-in, step changes, discrete fault signal and wear-out are added. On the other hand, as Figure 4.51 (b) shows, a 4-components GMM is also simulated and is to be used as the normal model. It is noted that this GMM is created with the intention of deviating from the simulated run-to-failure test in order to emphasize the importance of updating the anomaly detection model. If the run-to-failure data are plotted into the GMM space, as Figure 4.51 (c) illustrates, most of the run-to-failure data are deviated from the Gaussian kernels, and this results in most of the Hotelling's T-squared statistic from the run-to-failure data being above the threshold line, which will thus trigger a very high false alarm rate.

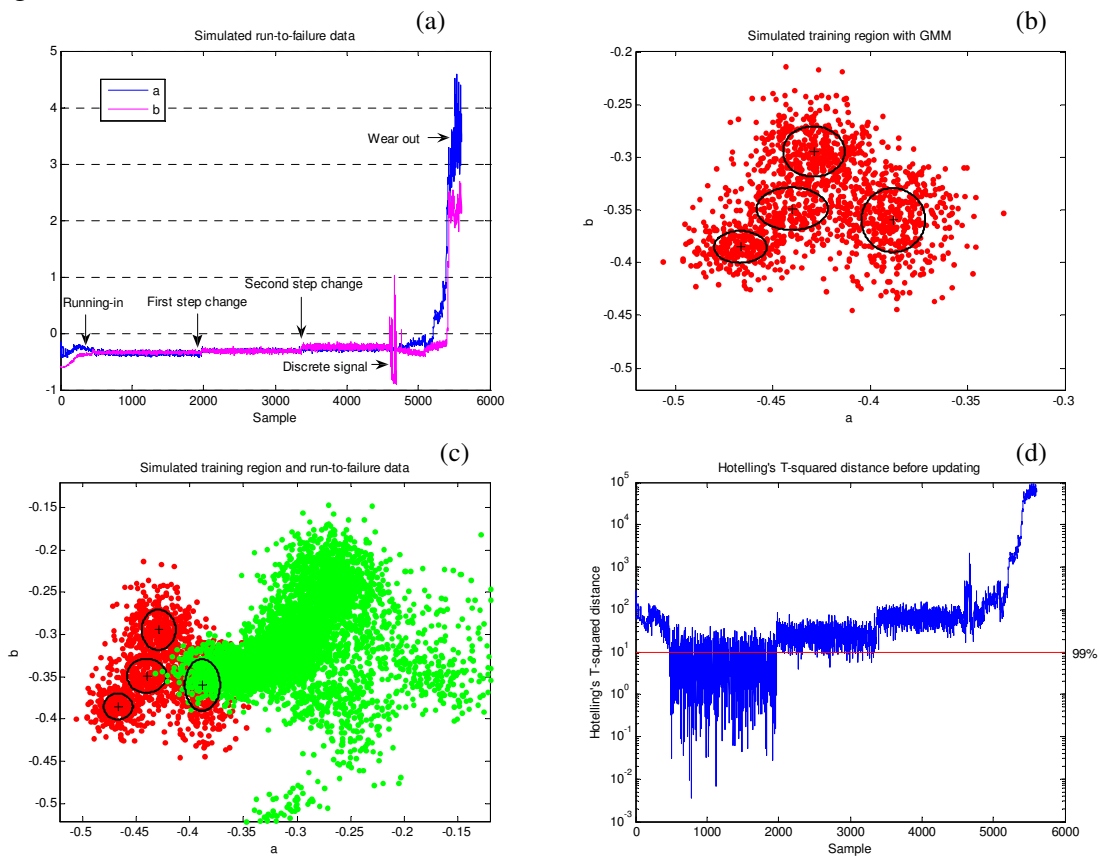


Figure 4.51 Simulated dataset: (a) Simulated testing data (b) Simulated training data with 4 Gaussians (c) Scatter plot of the GMM space and testing data.(d) Hotelling's T-squared statistic of the testing data based on 4 components of GMM.

Figure 4.51 (d) is the Hotelling's T-squared based anomaly detection result which indicates that a large amount of data are exceeding the threshold line, even though some of them belong to a steady normal state. This is due to the difference of background level between the training and testing data. Therefore, appropriate testing data need to be selected to update the original constructed GMM space. In this case, as the testing data contains various kinds of data, and background normal data are preferred to be included, the developed method should be able to identify the characteristic difference between the normal and other types of data. After the observation of the testing data, as illustrated in Figure 4.51 (a), it is found that data at running-in or faulty areas generally have high magnitude fluctuations, and if a set of categorized windows are set in these areas, two neighbouring datasets have a high probability of being clustered independently. On the other hand, data at a steady normal state are relatively normally distributed, and two neighbouring datasets are not easily separated, and may almost equally appear in the assigned clusters. Based on this important characteristic, the proposed method is summarized as:

- 1) Assigning a number of categorized windows in the time series, and number of points in each group is fixed and equalled.
- 2) Modelling every two neighbouring groups (previous and current) by EM algorithm.
- 3) Setting the 'number of clusters' as two. The method is not supposed to accurately describe the two neighbouring datasets, but to examine how separated the two datasets are, and see whether they could form independent clusters.
- 4) Checking the consisted elements in each cluster. As the number of data points of two neighbouring groups (previous and current) are equalled, and each group would occupy most in one of the two clusters. So the clusters are named based on which group of data occupies the most.
- 5) Plotting the occupation probability of each of two neighbouring groups (previous and current) for two clusters.

Figure 4.52 shows the clustering samples of the three neighbouring groups. The first sample is groups 2 and 3 at the running-in area, and it is seen that most of the data from

each group could form independent clusters, and the dominant data group could be found in each cluster; while the second sample from a steady state indicates a different feature: the occupation probabilities of two neighbouring groups are close to each other in the clusters, which means the dataset of groups 4 and 5 are mixed together, and could not be clustered independently. In the third sample, the data of two neighbouring groups are experiencing a step change, and the clustering result is similar to the first sample.

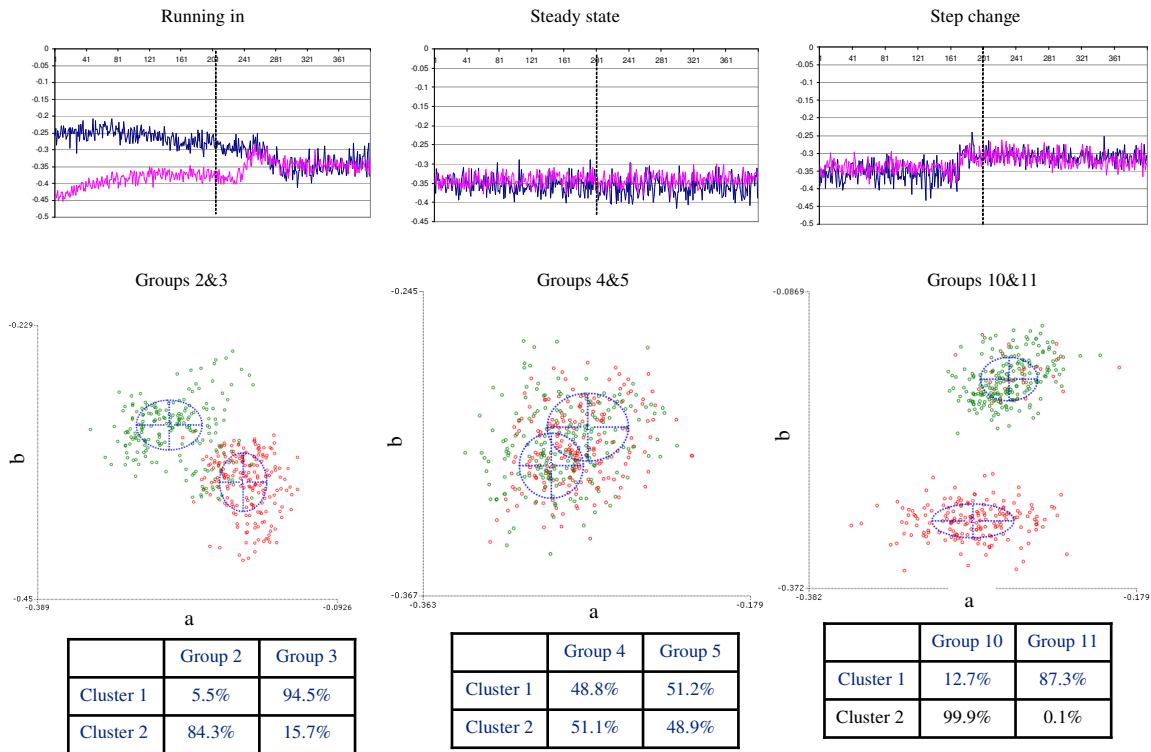


Figure 4.52 Samples of clustering results with two neighbouring datasets

If the probabilities of the two clusters are plotted according to which group of datasets (previous and current) occupies most, the occupation probability charts could be obtained, as Figure 4.53 illustrates. From these two charts, it is clearly found that the occupation probabilities of two groups of data have a large deviation in both clusters when they are at the running-in, step change, faulty and wear-out areas, while the occupation probabilities of the two groups of data are much closer to each other when they are at the steady normal states. Therefore, the occupation probability chart could be used to recognize data at steady normal states from other types of data, and in this case, it is advisable to utilize groups of 5, 11 and 19 datasets to update the anomaly detection model.

In addition to the simulation data, experimental data are also examined in this way to select appropriate points to update the model; examples of these tests can be found in Appendix C.

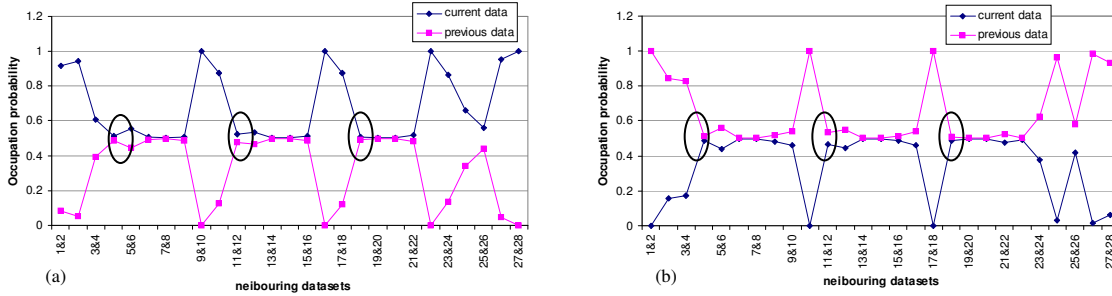


Figure 4.53 Occupation probabilities (a) cluster that current data occupies most; (b) cluster that previous data occupies most.

4.6.3. Anomaly detection model updating and merging processes

4.6.3.1. Concatenate process

Having determined which group of testing data should be appropriately selected for updating the anomaly detection model, the updating procedure could be followed with the method suggested by Hall and Hicks [10]:

Concatenate – suppose a GMM has already been learnt up to time t :

$$p^t(x) = \frac{\sum_{m=1}^k w_m^t p(x; \mu_m^t, C_m^t)}{\sum_{m=1}^k w_m^t} \quad (4.21)$$

Where each Gaussian component is represented by its weight w_m^t , mean μ_m^t and covariance C_m^t . Then, if a set of new data is received, the updated model with $k+1$ components could be obtained by trivially combining the original GMM and the new data into a single model:

$$p^{t+1}(x) = \frac{\sum_{m=1}^k w_m^t p(x; \mu_m^t, C_m^t) + w^{t+1} p(x; \mu^{t+1}, C^{t+1})}{\sum_{m=1}^k w_m^t + w^{t+1}} \quad (4.22)$$

Figure 4.54 (a) illustrates the anomaly detection results after the GMM space is updated with the group of 5 dataset. It is found that most of the steady state data before the first step change have been classified as normal, but the other steady state data are still above the threshold line. Therefore, the group of 11 datasets is applied for updating, and in Figure 4.54 (b), nearly all the steady state data before the second step change are below the threshold line. The group of 19 datasets is finally extracted to update the model, and as Figure 4.54 (c) indicates, the steady state data are successfully treated as normal, and the running-in, discrete fault signal and wear out behaviours are also accurately detected.

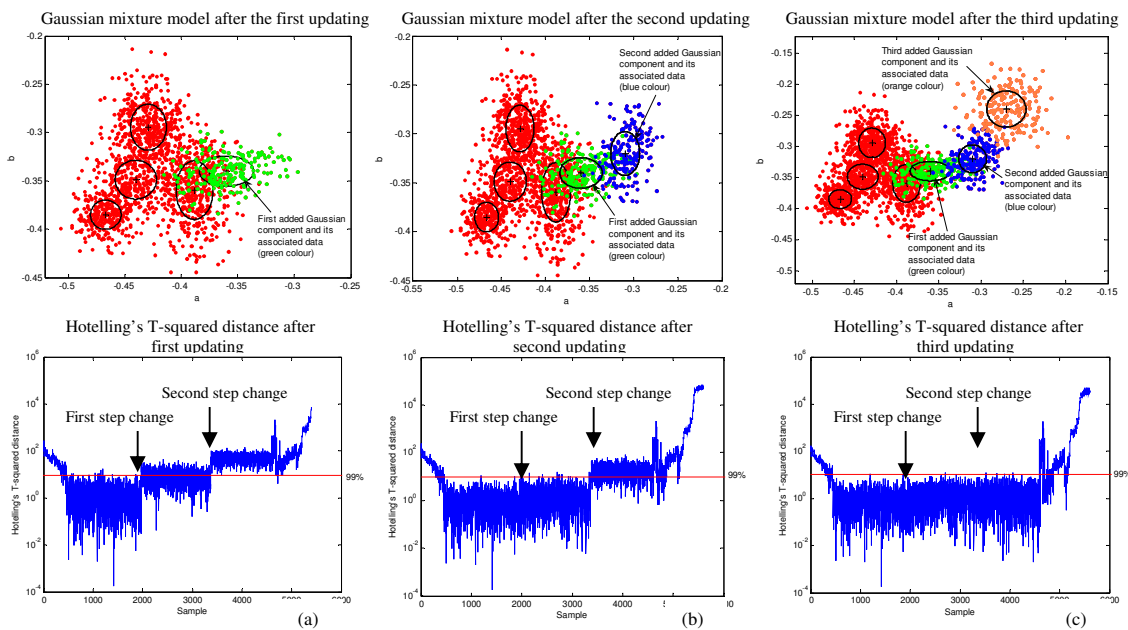


Figure 4.54 Anomaly detection results after the model updating with the group of 5 dataset (a), groups of 5 and 11 datasets (b) and groups of 5, 11 and 19 datasets (c).

4.6.3.2. Merging and simplifying process

Simplify – if possible, merge some of the Gaussians by measuring the similarity which could be determined based on the distances between the Gaussian centres and difference of their covariance size. It is noted that the BIC index should be consistently used after each merging process, to ensure the optimized standard of the updated model, as it directly affects the accuracy of anomaly detection.

Cluster No.	1	2	3	4	5	6	7
1	1	0.81516	0.80172	0.87255	0.91731	0.78924	0.59573
2	0.81516	1	0.76417	0.86681	0.80137	0.70753	0.59607
3	0.80172	0.76417	1	0.89185	0.72162	0.59222	0.4126
4	0.87255	0.86681	0.89185	1	0.80635	0.67978	0.51475
5	0.91731	0.80137	0.72162	0.80635	1	0.87058	0.67668
6	0.78924	0.70753	0.59222	0.67978	0.87058	1	0.78505
7	0.59573	0.59607	0.4126	0.51475	0.67668	0.78505	1

Table 4.10 Similarity values between the clusters

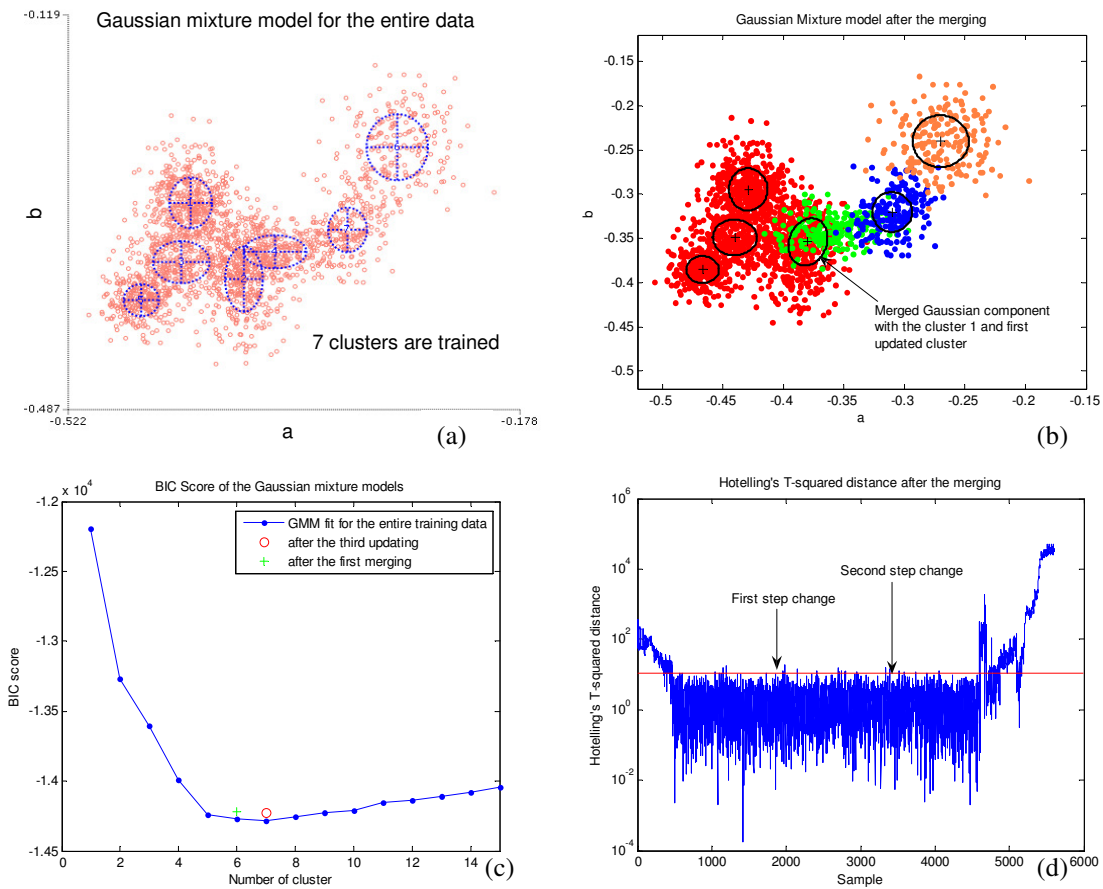


Figure 4.55 Results of the clustering merging: (a) Clustering space for the entire training data; (b) Clustering space after the merging; (c) BIC scores for different clustering spaces; (d) T-squared statistic based on the merging space

To decide whether two Gaussian components (clusters) G_i and G_j can be simplified into one, and to check if the similarity between the clusters has a value close to one, say, exceeding a given threshold $SIM_{min}=0.9$, the similarity between the clusters can be

calculated. Following the equation of 4.10, the similarity values between the 7 clusters (see Figure 4.54 (c)) is obtained as Table 4.10 indicates. It is found that the similarity between clusters 1 and 5 is above the threshold and reached about 0.92. Therefore, these two resulting Gaussians are merged using the usual equations (4.23-4.26) supplemented by the combination of the cumulative density functions [10]:

$$\omega = \omega_i + \omega_j \quad (4.23) \quad \mu = \frac{1}{\omega} [\omega_i \mu_i + \omega_j \mu_j] \quad (4.24)$$

$$C = \frac{\omega_i}{\omega} [C_i + (\mu_i - \mu)^T (\mu_i - \mu)] + \frac{\omega_j}{\omega} [C_j + (\mu_j - \mu)^T (\mu_j - \mu)] \quad (4.25)$$

$$F(x) = \frac{1}{\omega} [\omega_i F_i(x) + \omega_j F_j(x)] \quad (4.26)$$

Figure 4.55 (b) shows the merged model, and Figure 4.55 (c) illustrates the BIC scores of the model before (red circle) and after (green cross) the merging; these two values are seen to be close to each other, and if compared to the BIC values measured for the entire training data as Figure 4.55 (a) illustrates, the difference between the BIC score of the optimized model (7 clusters) and those of the updated and merged models is limited. Thus, the quality of the model is not changed significantly after the merging process, and the testing data can be calculated based on this merged model. Figure 4.55 (d) is the T-squared score of the testing data, and the testing result is quite similar to the one that is obtained based on the third updated model (see figure 4.54 (c)). Although the result does not indicate any improve through merging, it costs less time because of the reduction of model complexity.

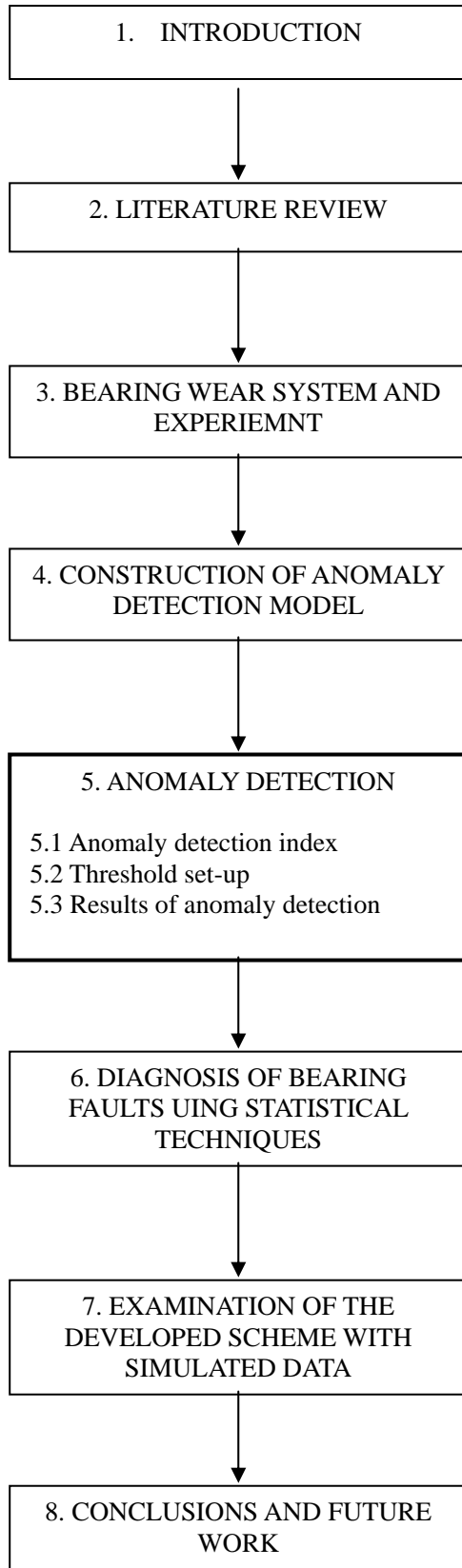
4.7 Reference

- [1] C.M. Bishop, *Neural Networks for Pattern Recognition*. 1995, Oxford, UK, ISBN 0 19 853864 2 (pbk).
- [2] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, B 39 (1977) 1-38.
- [3] C. Fraley, A.E. Raftery, How Many Clusters? Which Clustering Method? Answer via Model-Based Cluster Analysis. *The Computer Journal* 41 (1998) 578-588.
- [4] C. Biernacki, G. Celeux, G. Govaet, Choosing Starting Value for the EM for getting the highest likelihood in Multivariate Gaussian Mixture Models. *Comput.Stat.Data Anal.* 41 (2003) 561.
- [5] L.K. Hyung, Y.S. Song, K.M. Lee. Similarity Measure between Fuzzy sets and between elements. *Fuzzy Sets Syst.* 62 (1994) 291.
- [6] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters* 27 (2006) 861-874.
- [7] R. Callan. *Artificial Intelligence*, 2003, Palgrave, UK, ISBN 0-333-80136-9.
- [8] E. Keogh, S. Chu, D. Hart, M. Pazzani. Segmenting Time Series: A Survey and Novel Approach. *Data Mining in Time Series Databases*. World Scientific Publishing Company. (2003).
- [9] T. Anderson, *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley and Sons, 2nd edn., 1984.
- [10] P. Hall, Y.A. Hicks. A method to add Gaussian mixture models. *Tech. Report, University of Bath*. (2005).
- [11] M. Song, H. Wang. Highly efficient incremental estimation of Gaussian mixture models for online data stream clustering. *Intelligent Computing: Theory and Applications*. (2005).
- [12] O.D. Tasbaz, R.J.K. Wood, M. Brown, H.E.G Powrie, G. Denuault, Electrostatic monitoring of oil lubricated sliding points contacts for early detection of scuffing. *Wear* 230 (1999) 86-97.

- [13] S. Morris, R.J.K. Wood, T.J. Harvey, H.E.G. Powrie, Use of electrostatic charge monitoring for early detection of adhesive wear in oil lubricated contacts, *ASME Journal of Tribology* 124 (2002) 288-296.
- [14] H.E.G. Powrie, O.D. Tasbaz, R.J.K. Wood, C.E. Fisher, Performance of an electrostatic oil monitoring system during FZG Scuffing Test, *Proceedings of the International Conference on Condition Monitoring, University of Wales Swansea, UK*, 1999, pp. 155–174.

THESIS STRUCTURE

KEY OUTCOME



An anomaly detection index: Hotelling's T-squared statistic (5.1) has been adopted to extract abnormal information from the testing data. A new threshold strategy for the T-squared statistic has also been developed based on GMM and EVT (5.2). Four sets of testing data haven been presented in this thesis to show that not only the anomaly detection rate can be significantly improved but more prognostic information can also be extracted.

Chapter 5. Anomaly detection

5.1. Anomaly detection index

In machine condition monitoring, univariate (i.e. single variable) methods can sometimes miss anomalies that do not appear at the extreme tail of a single variable's distribution. Multivariate approaches are more powerful to enhance and extract these anomalies. In this thesis, Hotelling's T-squared statistic is applied to fuse the multivariate (i.e. multiple features) and to calculate the distance between the centre of the anomaly detection model and the new testing samples.

Based on the assumption that random features can be approximated by a Gaussian distributed random variable, Hotelling's T-squared statistic can be defined as follows:

$$T^2 = \frac{n_x n_y}{n_x + n_y} (\mu_y - \mu_x)^T \hat{C}_{xy}^{-1} (\mu_y - \mu_x) \quad (5.1)$$

$$\hat{C}_{xy} = \frac{1}{n_x + n_y - 2} ((n_x - 1) \hat{C}_x + (n_y - 1) \hat{C}_y) \quad (5.2)$$

Where \hat{C}_x and \hat{C}_y are estimates of covariance matrices of the samples X and Y , respectively; and μ_x and μ_y are sample means of X and Y , respectively.

As for each new test vector y_i , the normal model is a Gaussian mixture with k components. A decision is required to identify which of the components should be used to calculate the Hotelling's T-squared statistic for the new data sample. It is proposed that the Mahalanobis distance [1, 2] is calculated for every new data sample based on each of the Gaussian mixtures. The Gaussian component with the minimum Mahalanobis distance is then chosen for the current observed sample:

$$k_{\min} = \arg \min_k \sqrt{(y_i - \mu_k)^T C_K^{-1} (y_i - \mu_k)} \quad (5.3)$$

5.2. Setting-up thresholds for bearing failure detection

This section introduces the proposed approach to set the threshold and probabilistic boundary for anomaly detection. At the beginning, the principle of the Extreme Value Theory (EVT) is presented as a statistical method to set probabilistic thresholds which are more sensitive to the anomalies. Next, the proposed approach that uses the EVT to set the threshold for the T-squared statistic is demonstrated using baseline data. The next two parts are about the evaluation of the effect of model adaptation on the threshold set-up and comparison with other conventional methods respectively. Appendix A4 illustrates the procedure of setting up the threshold level.

5.2.1. Approach of using Extreme Value Theory (EVT) to set a threshold level for anomaly detection (threshold for distance measure and probabilistic boundary)

Extreme value statistics is a branch of statistics that effectively models the tails of the distributions, concerning where the extreme large or small values extracted from the distribution of ‘normal’ data are expected to lie.

Consider a set of data $X = \{x_1 \dots x_m\}$ of m is drawn from a distribution $D(x)$. In this thesis, the X is assumed to be generated from a Gaussian distribution $P(x) = N(\mu, \sigma^2)$ and defines $x_m = \max(X)$. Then, according to one of the fundamental principles in ‘classical’ extreme value statistics, the Fisher-Tippett theorem [3], the probability distribution $P_E(x)$ describing where the expected most extreme of those m samples is lying towards the Gumbel distribution (Gumbel distribution is a limited form of the generalized extreme value distribution which describes maxima):

$$P(x_m \leq x | \mu_m, \sigma_m) = \exp\{-\exp(-y_m)\} \quad (5.4)$$

Which the Extreme Value Distribution (EVD) is termed, where y_m is the reduced variable:

$$y_m = \sigma_m^{-1}(x_m - \mu_m) \quad (5.5)$$

Where μ_m and σ_m are the location and scale parameters of the EVD as equations 5.6 and 5.7 show. It is seen that the classical EVT assumes that location and scale parameters are dependent only on m .

$$\mu_m = (2 \ln m)^{1/2} - \frac{\ln \ln m + \ln 2\pi}{2(2 \ln m)^{1/2}} \quad (5.6)$$

$$\sigma_m = (2 \ln m)^{-1/2} \quad (5.7)$$

However, the theory introduced above is only effective when x_m is univariate. On the other hand, the assumption that the $X = \{x_1 \dots x_m\}$ is generated from Gaussian distribution $P(x) = N(\mu, \sigma^2)$ seems insufficient if the X is distributed in a more complicated way. Therefore, it is reassumed that in [1, 2], multivariate X is generated from the Gaussian Mixture Model (GMM) $P(x) = N(\mu_k, C_k)$, and can be easily evaluated with the Mahalanobis distance in its standard form:

$$h(x)_{k^*} = \sqrt{(x - \mu_{k^*})^T C_{k^*}^{-1} (x - \mu_{k^*})} \quad (5.8)$$

Hence, the information of multivariate vector x_m can be transmitted by univariate value $h(x)_k$, and in equations 5.4-5.7, the parameters x_m and m can be replaced by $x_m = h(x)_k$, $m_k = mp(k)$ (the selection of component k of GMM for specific vector is based on the minimum $h(x)$), therefore, the cumulative EVT probability can be modified from equation 5.4:

$$P(x|m) = P(h(x)_{k^*} | m_{k^*}) = \exp\left\{-\exp\left(-\left(\frac{h(x)_{k^*} - \mu_m}{\sigma_m}\right)\right)\right\} \quad (5.9)$$

5.2.2. Threshold set-up for distance measure (T-squared distance)

Based on the above-described theory, the probabilistic boundary ($P(x) \geq 95\%$ or 99%) could be drawn for univariate or multivariate physical features. However, as introduced in the previous section, the Hotelling's T-squared distance is selected as the anomaly

detection index, and the EVT probability can not be calculated directly for this statistic. Therefore, the conventional EVT approach needs to be modified and the EVT probability of the T-squared statistic needs to be evaluated.

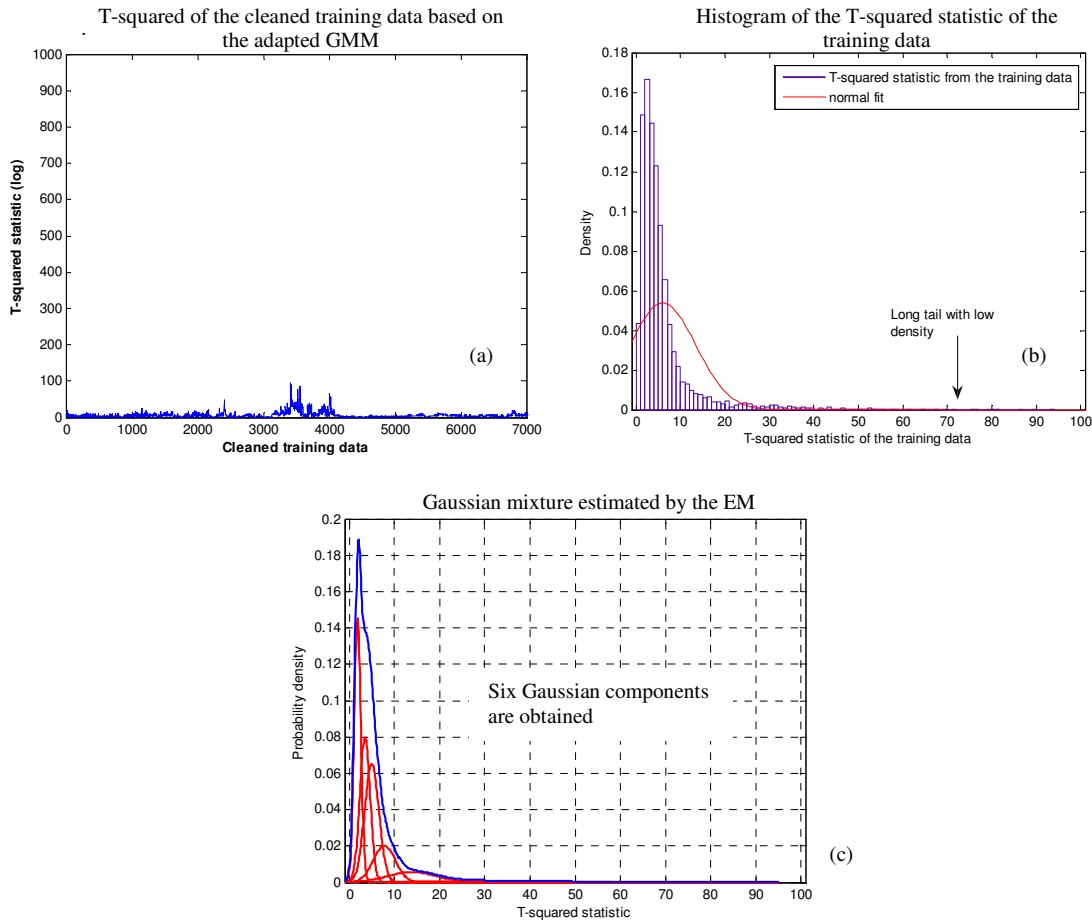


Figure 5.1 Statistical analysis of the T-squared distance from the training data

As discussed in Section 5.1, the anomaly detection index is calculated based on the constructed anomaly detection model. In this study of setting up the threshold, the predicted T-squared statistic of the training data (TSOT for short) based on the model of normality (the optimized model is chosen as Figure 4.34 shows) is used to generate the alarm threshold. Figure 5.1 (a) shows the calculated TSOT. The density of TSOT is shown in Figure 5.1 (b), and shows that the density does not follow a single Gaussian distribution which provides a poor fit to the data and contains a very long tail; hence, it would not be correct to base a threshold on a mean value plus several standard deviations. In order to describe TSOT more precisely, the GMM is applied to identify different

regions of density of TSOT instead of a single Gaussian model. It is seen in Figure 5.1 (c) that there are 6 Gaussian components (clusters) being created. Finally, the threshold line should also lie towards the tail of the cumulative distribution function (CDF) $P(x)$, so that the anomalies could be detected when $P(x)$ is $> 95\%$ or 99% (determined beforehand). In this particular study, the TSOT at different regions of density are modelled with the Extreme Value Theory (EVT) [4] so that the threshold line could be set at the tail of the cumulative distribution.

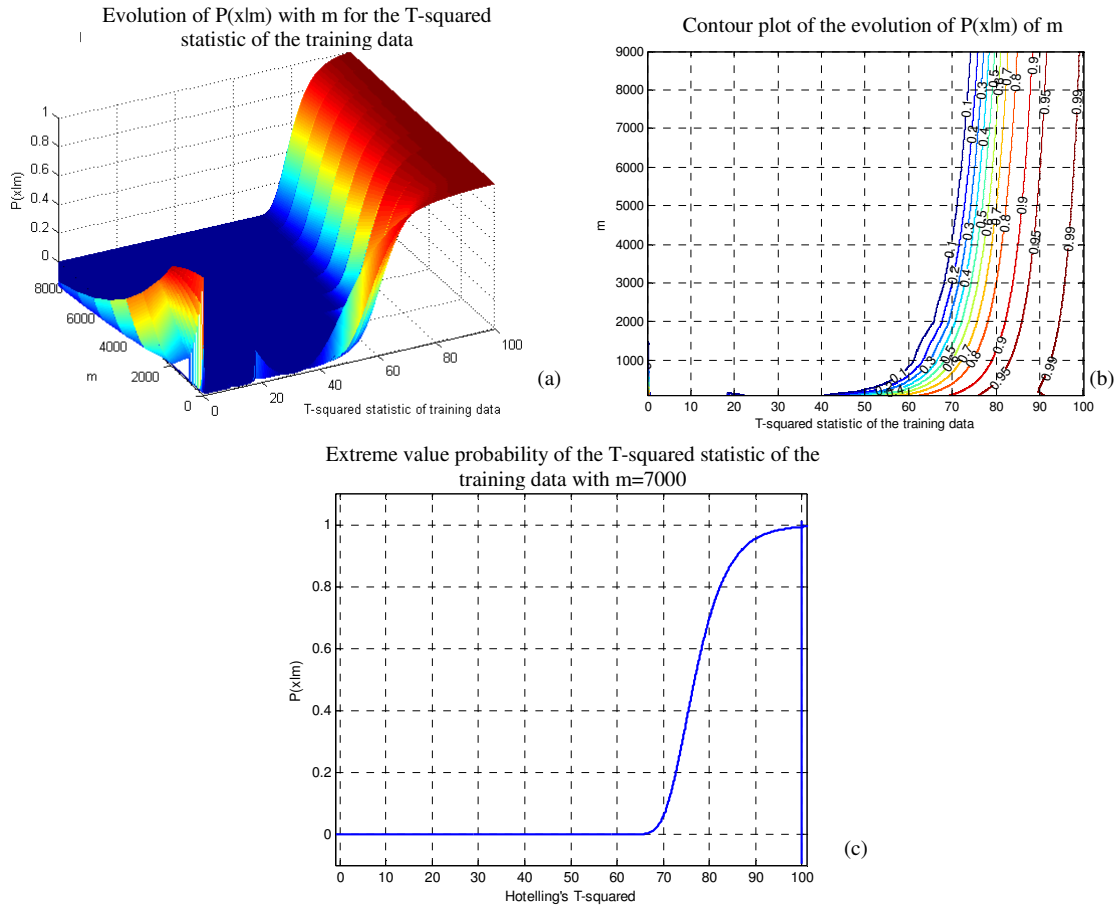


Figure 5.2 Threshold set-up using extreme value statistic

In section 5.2.1, the idea is introduced that the location and scale parameters of the EVT probability are only dependent on the number of observed data points (m) which are used to fit the model. Therefore, variation of value m has an influence on the calculation of probability. Figure 5.2 shows the evolution of $P(x|m)$ with m for the TSOT. The number of observed data points increases and so the resultant extremal probabilities become less conservative. The value of m increases linearly from 2 to 8,000. If the threshold is set

when the probability reaches 99%, as the figure illustrates, the threshold alerting values have a non-linear increasing trend with different m . It is also found that the difference between the 99% related values becomes smaller when the m is larger than 7,000. Therefore, in this study, the $m=7000$ is selected to set the threshold value. Figure 5.2 presents the EVT cumulative probability of the TSOT, and it is found that the threshold line is set to 99.2 when the probability reaches 99%.

According to the analysis discussed above, the approach to set the threshold level for univariate anomaly detection index can be summarized as follows:

- 1) Calculate anomaly detection index of the training data based on the constructed model of normality.
- 2) Inspect the density changes of the anomaly detection index of the training data using GMM.
- 3) Select appropriate parameter m .
- 4) Calculate EVT probability of the anomaly detection index, and set a threshold level when the probability reaches a certain level (i.e. $P(x) > 95%$ or $99%$).

5.2.3. Probabilistic boundary set-up

Different from the threshold set-up of the anomaly detection index, the probabilistic boundary can be calculated directly with multivariate features based on the constructed model of normality.

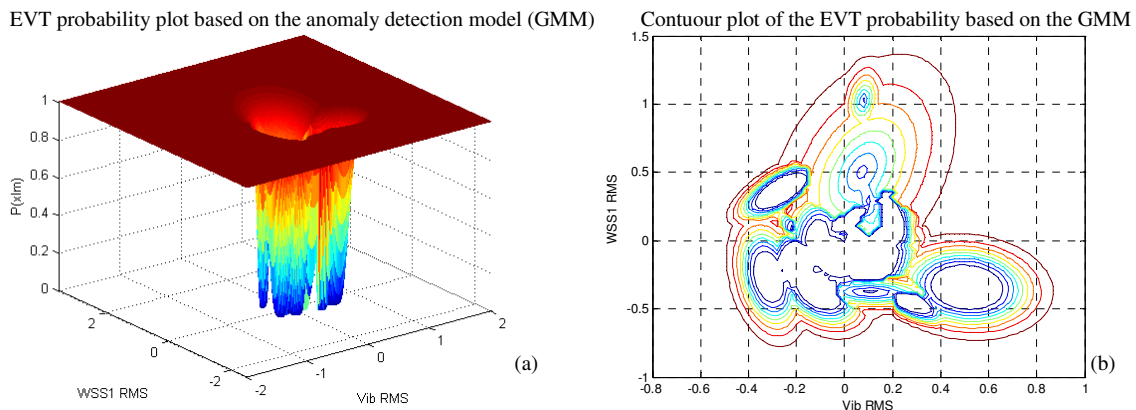


Figure 5.3 Probabilistic boundary after the model adaptation

Figure 5.3 shows the surface mesh plot of the EVT probability for the first two features (vibration RMS and WSS1 RMS) space. Figure 5.3 (b) shows the boundaries with different probability values, and any related data can be plotted in this two-dimensional space to check their status with different probability boundaries.

5.2.4. Effect of model adaptation on the threshold set-up

In this section, the method is repeated, but applied to the original training data and constructed model of normality (see Figure 4.33) before adaptation. The purpose of this study is to examine the effect of model adaptation on the threshold set-up.

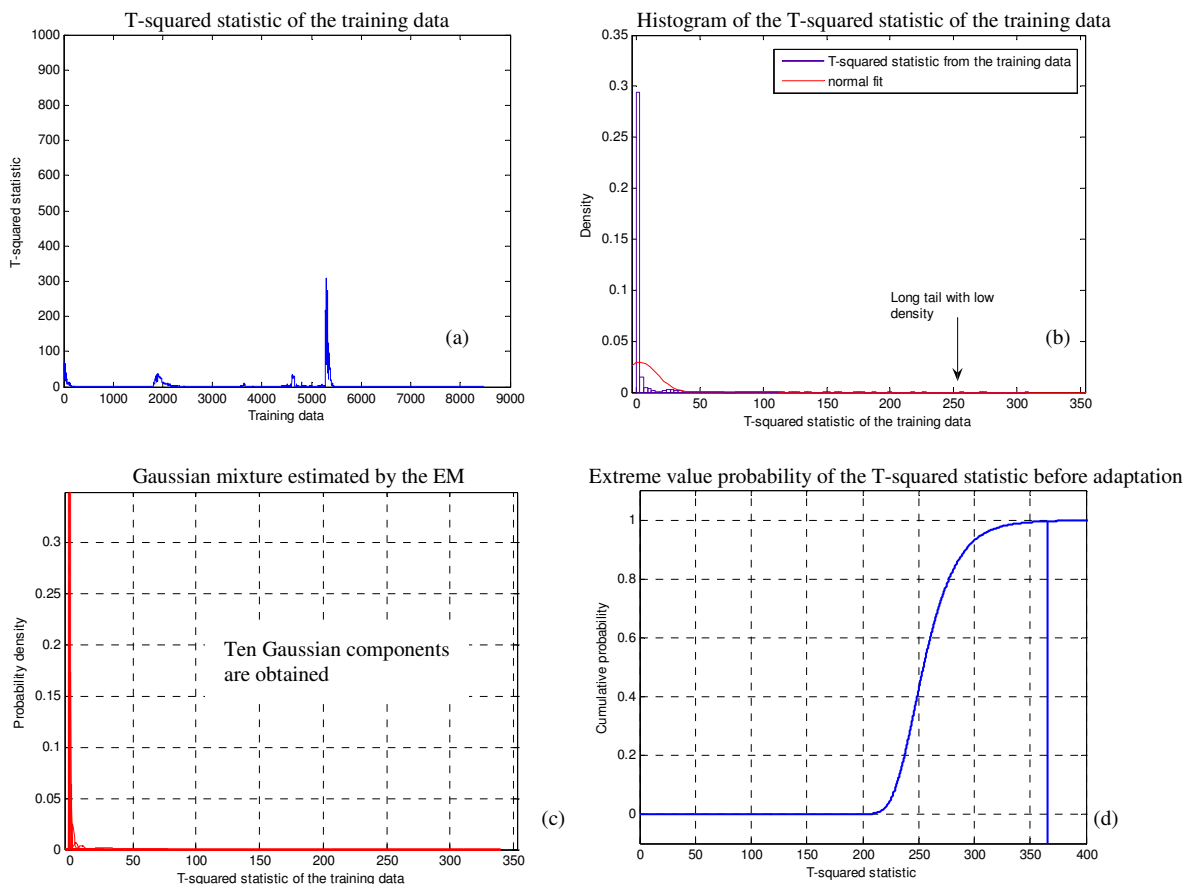


Figure 5.4 Threshold set-up before the adaptation of the model

Figure 5.4 illustrates the procedure of setting the threshold for T-squared statistic. It is seen that the calculated T-squared statistic of the original training data (see Figure 5.4 (a)) has a long tail in the statistical plot (see Figure 5.4 (b)), and the GMM with 10 Gaussian

components were provided to precisely describe the density changes of the data, as Figure 5.4 (c) illustrates. Finally, the EVT probability is calculated with selected $m=8000$ for the data, and the threshold value is set to 364.1 when $P(x)=99\%$.

On the other hand, the probabilistic boundaries with the original training data and model of normality are also drawn. Figure 5.5 shows the surface mesh and contour of the EVT probability for the original model.

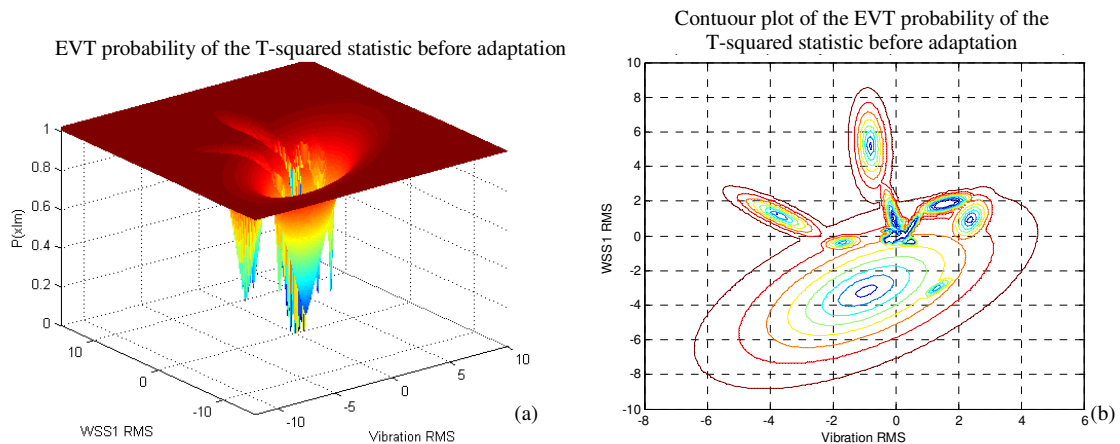


Figure 5.5 Probabilistic boundary before the model adaptation

From the threshold level and probabilistic boundary for the original training data and model of normality, it is found that the value and occupied area are both much larger than the ones created with cleaned training data and an adapted model. This is because a number of abnormal data which are easily sensed by the EVD are distributed very widely in the GMM space, and the threshold value and probabilistic boundaries are also affected by their wide spread distribution. Once these abnormal data and their associated clusters are removed, the distribution of the training data becomes more concentrated, and results in a low threshold value and smaller RMS occupied area.

5.2.5. Comparison of the developed method with the conventional methods

It is confirmed that the threshold and probabilistic boundary can be set by calculating the cumulative EVT probability. However, conventionally, a cumulative Gaussian distribution is used to set the threshold and boundary at the tail of its distribution. The

difference between the Gaussian and EVT distribution can be clearly demonstrated in the TSOT PDF function plot, as Figure 5.6 shows. It is seen that for both TSOT before and after the model adaptation, the PDF of EVT is more sensitive to the data in the tail position, while the Gaussian fit is more centred on the data with higher density. This obvious difference forms the motivation of the comparison between the Gaussian and EVT CDF.

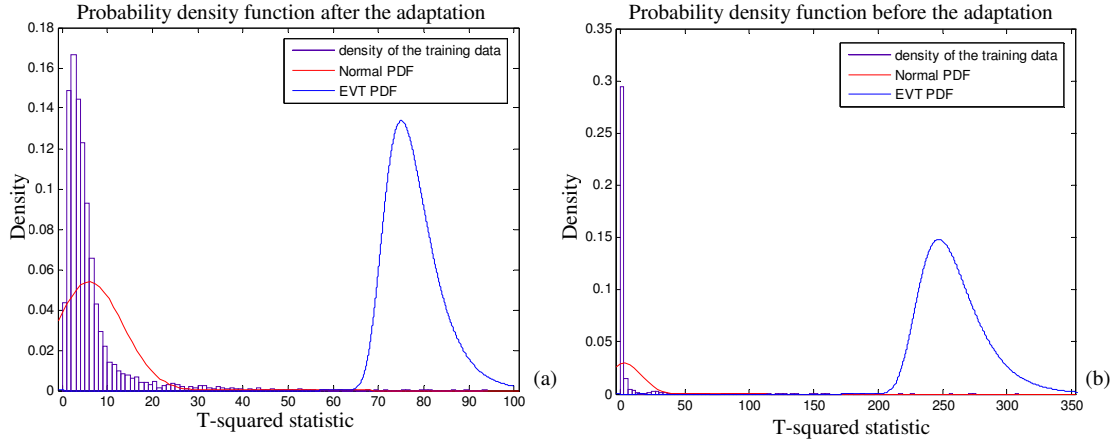


Figure 5.6 Comparison between the EVT and Gaussian PDF

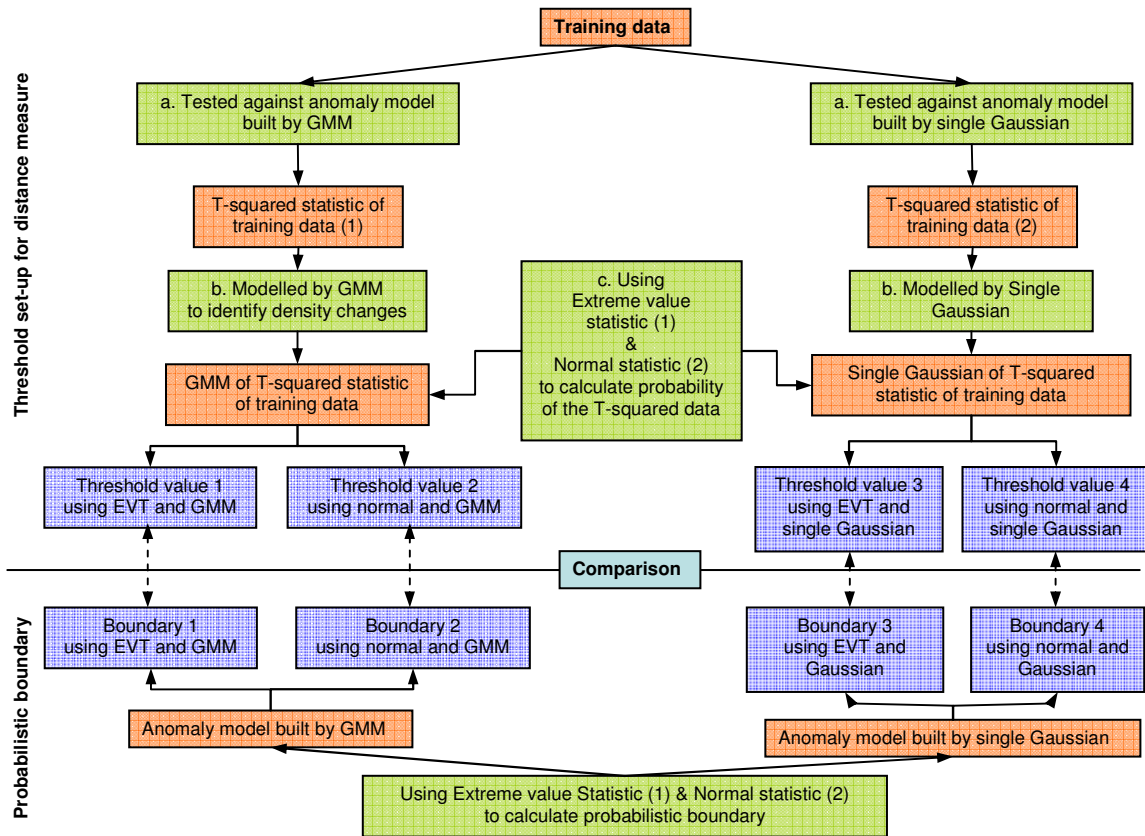


Figure 5.7 Flowchart of the threshold and probabilistic boundary set-up

On the other hand, the threshold line should also lie towards the tail of the cumulative distribution function (CDF) $P(x)$, so that the anomalies could be detected when the $P(x)$ is $>95\%$ or even 99% (determined beforehand). Conventionally, normal CDF could be used to achieve this function based on a single pair of Gaussian parameters μ and σ or multiple sets of μ_k and σ_k generated by GMM. Figure 5.8 (with the red and orange curve) shows the normal CDF based on single and multiple sets of Gaussian parameters respectively. It is seen that the threshold line has been set when the probabilities reach 99% with the corresponding values of 28 and 37 respectively. However, it is well known that normal distribution is not sensitive to the data at the extreme tail shown in Figure 5.1(b), and may treat these data as anomalies; this will lead to a high false alarm rate when conducting anomaly detection. Figure 5.7 summarizes the procedure of the threshold set-up using different statistics and models.

As a branch of Extreme Value Distribution, the Gumbel distribution is more sensitive to the tail region on the probability plot, and the majority of the data will not be in alert. The green and blue curves in Figure 5.8 illustrate the cumulative extreme value probability of the training data based on single Gaussian and GMM respectively. It is seen that the value of GMM-EVT method increases to around 100 and higher than the value of Gaussian-EVT, if the threshold line is set at a 99% position. Therefore, it is concluded that the GMM is used to correctly describe density of the training data's T-squared statistic. The EVT is then applied to indicate the probability of the data point belonging to the extreme value distribution. The proposed method shows a more tolerated capability that more normal data will not be triggered as the abnormal. Table 5.1 shows the calculated threshold values with different methods and model status.

Method	Before adaptation	After adaptation
Gaussian-normal	49.2	28.8
Gaussian-EVT	77.4	47.9
GMM-normal	57.7	36.2
GMM-EVT	364.1	99.2

Table 5.1 Summary of the threshold-set results

Moreover, Figures 5.9 illustrate the probabilistic boundary with the 4 compared methods before and after the model adaptation respectively.

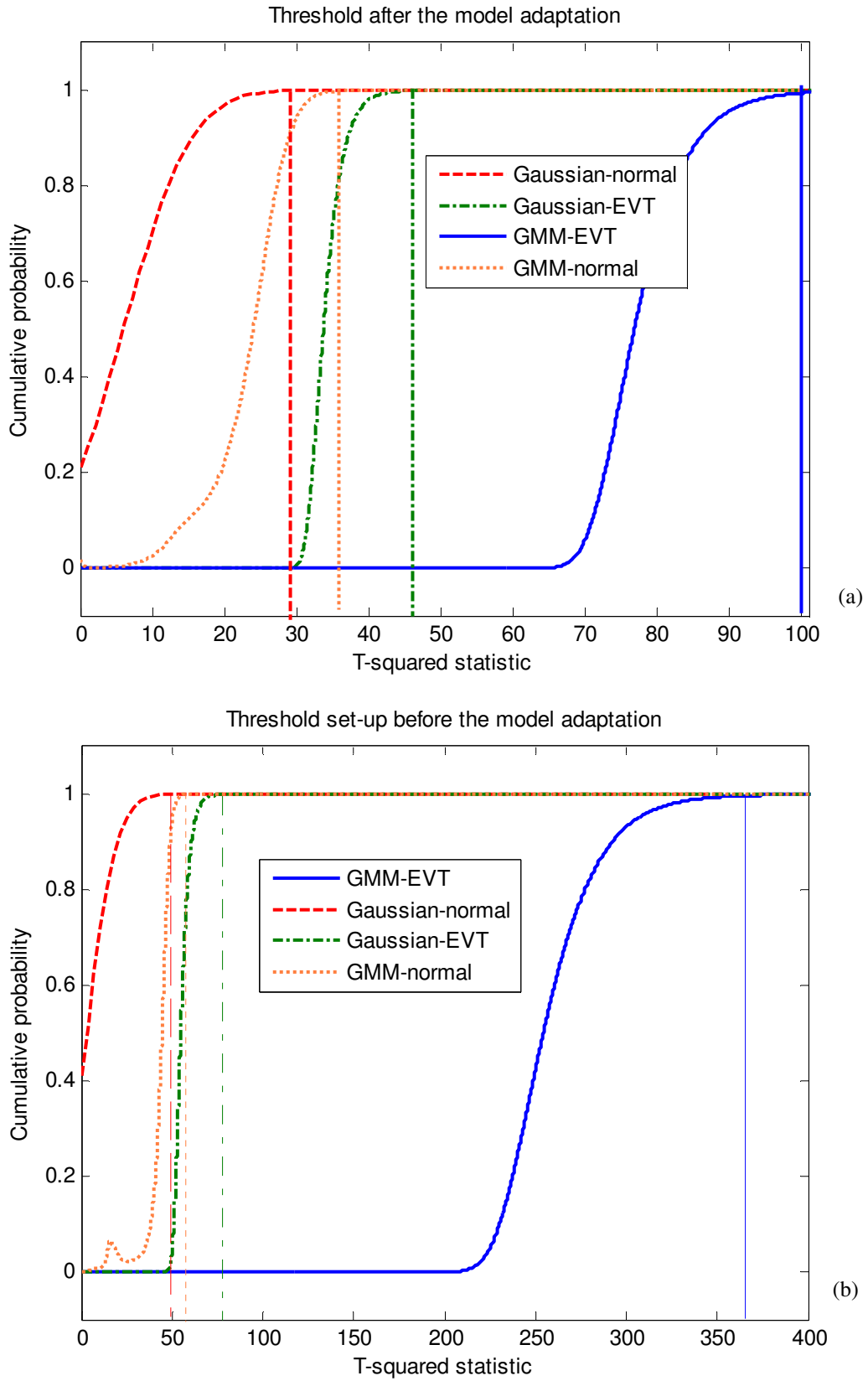


Figure 5.8 Threshold set-up before and after the adaptation of the model

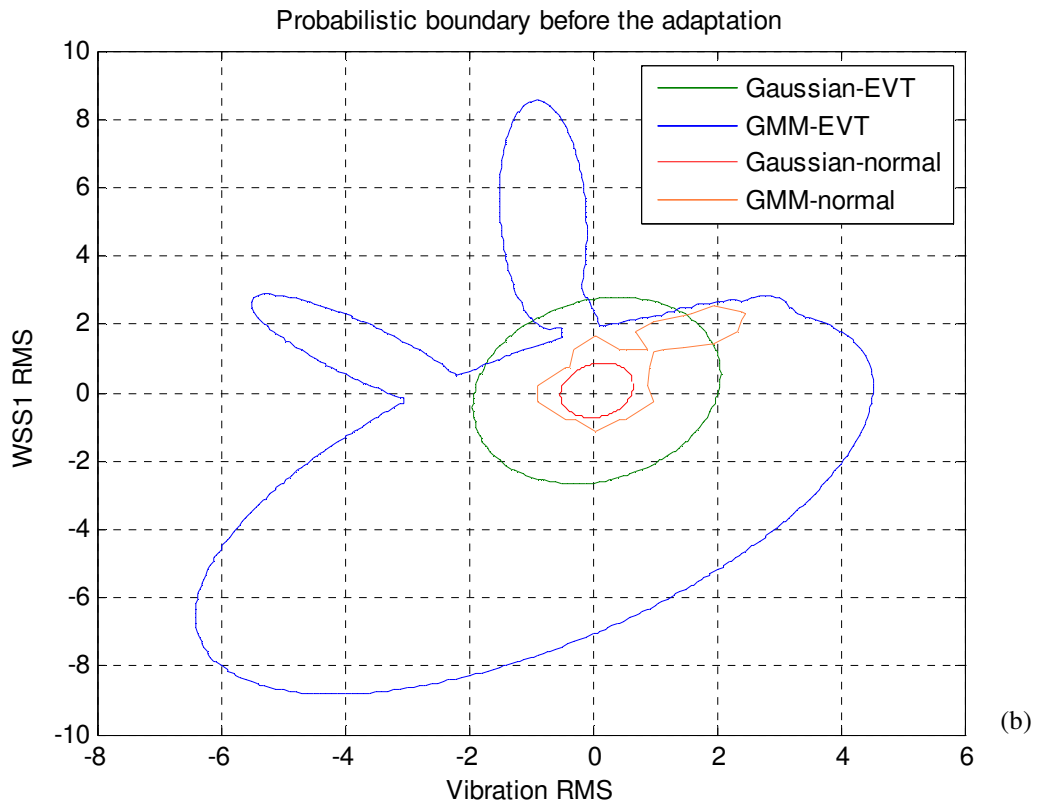
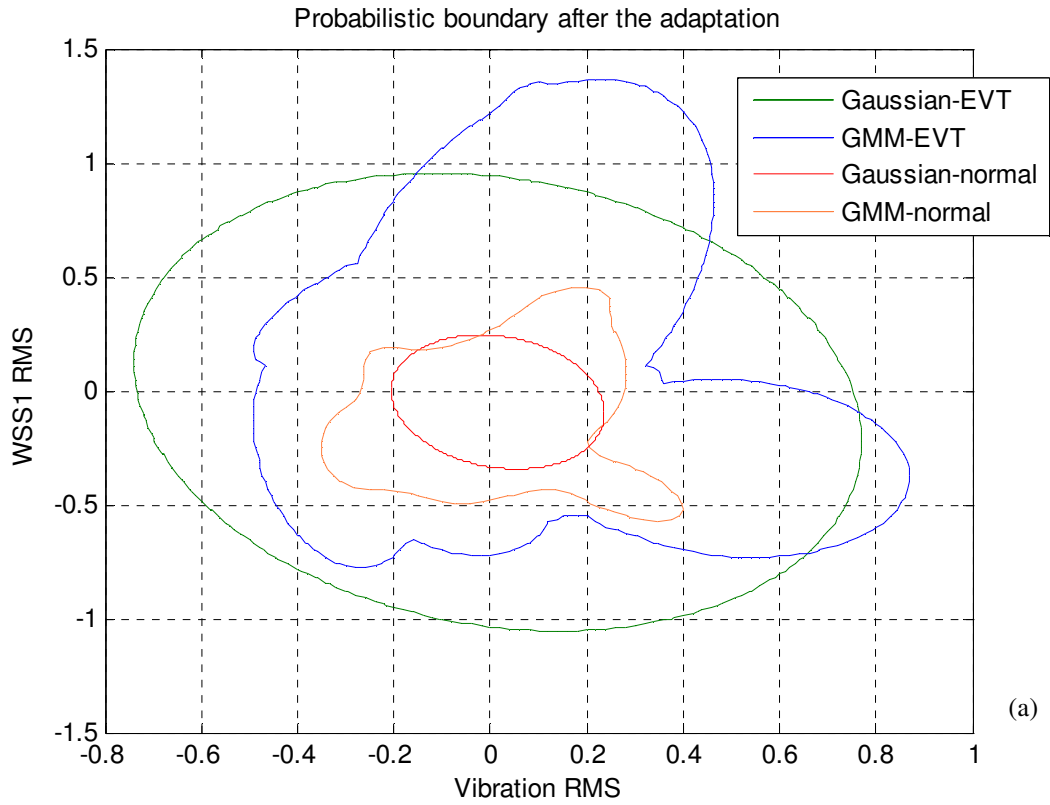


Figure 5.9 Probabilistic boundary set-up before and after the adaptation of the model

5.3. Results of the anomaly detection

Anomaly detection is a problem that arises in a number of applications in condition monitoring. Primarily, it is used to indicate whether the monitored events are working properly or not. There is no precise definition of what anomalies are but, intuitively, they can be defined as follows: “an anomaly would be an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [5]. There are two main approaches to anomaly detection. The first approach is based on estimation of unconditional probability density. In this case a decision on novelty is made using a probability density function (pdf) $p(x|normal)$ and a pre-defined threshold, the test point x is considered as an outlier while the $p(x|normal)$ is less than the threshold [6]. The second approach to anomaly detection is based on the description of the data region R that corresponds to the normal operating regime. The test event is considered as an anomaly when it is outside the region R . This approach has been widely applied in the statistical process control (SPC) domain [7-9]. Appendix A5 shows the general procedure for anomaly detection.

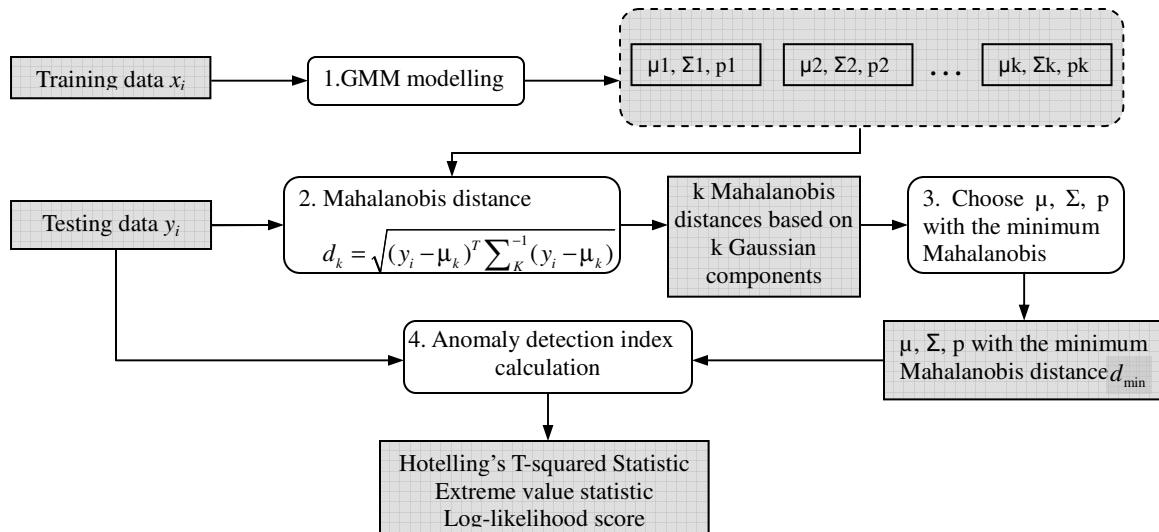


Figure 5.10 Approach to select appropriate Gaussian component for anomaly calculation

However, as discussed in Chapter 4, the Gaussian mixture model is used as the anomaly detection model which consists of k Gaussian components. Therefore, to judge the abnormality of a test point y_i , a decision has to be made on which of the k component

should be used to calculate the proposed anomaly detection indices. Figure 5.10 illustrates the approach to select appropriate Gaussian component for each test point to compute anomaly detection indices. After obtaining multiple Gaussian components in step 1, the proposed problem is solved in step 2 by calculating the Mahalanobis distance for each test point based on each Gaussian component. The Gaussian component with the minimum Mahalanobis distance [1, 2] will then be chosen for the current test sample (it is achieved in step 3) to calculate the fault detection indices, including extreme value statistic, Hotelling's T-squared statistic or log-likelihood.

In the following sections, analysis procedure of the anomaly detection will be introduced at first; an example will be given to explain how the anomaly detection results are interpreted. And then, an overview of the anomaly detection results of all 11 evaluated tests will be summarised and compared in section 5.3.2. Finally, four typical tests with distinctive features will be analysed with details in sections 5.3.3 and 5.3.4.

5.3.1. Analysis procedure of the anomaly detection

Generally, the aims of conducting anomaly detection can be divided into two parts. The first one is to extract or detect any abnormal trends which are related to rolling element bearing failure process; this requirement is fulfilled by the Hotelling's T-squared statistic. The second one is incorporating developed threshold set-up strategy to automatically isolate the extracted anomalies from normal data.

To achieve the first aim, time history of the T-squared statistic is plotted at first, and detected abnormal trends are then correlated to the phase of the machinery life cycle in rolling element bearings and the physical evidence that was observed in the course of testing. Figure 5.11 illustrates an example of the T-squared statistic calculated from the test 20 (details of this test can be found in section 5.3.3). It is seen that the T-squared statistic follows the well-known 'bath-tub-curve': in the running-in period (between 0 and 15 hours), bearing was experiencing a highly dynamic process that involves surface interaction between asperities and defect protrusions, this behaviour is usually associated with the generation of plate-like particles which can be verified by debris counter [11];

and then bearing was entering the steady state which lasted for about 18 hours, and in this period, there was also a significant reduction in debris detection; finally, activities of the T-squared statistic between 52 and 63 hours suggest that the bearing may be spalled generating discrete areas of nascent surface and large-sized particles. Apart from the classic three running phases of the machinery life cycle, activities between 42 and 43 hours are also worthwhile to be investigated, since they may be the invaluable precursors to the wear-out period. It should be noted that in the plotted T-squared statistic, a number of single spikes are also found (single spike at 12 and 48 hours), and have the obvious abnormal characteristic. However, at the time of writing, the cause for these anomalies is difficult to trace, because they may be the indication of bearing distress, or from the malfunction of instruments, or the contaminants in the lubricants. Therefore, there are generally three types of interesting activities (trends relating to the bath-tub-curve, early interesting consistent signs and single spike) are found, but the current study only focuses on the consistent abnormal trends in the T-squared statistics and the first two types of interesting activities will be analysed for each testing datasets in the following sections.

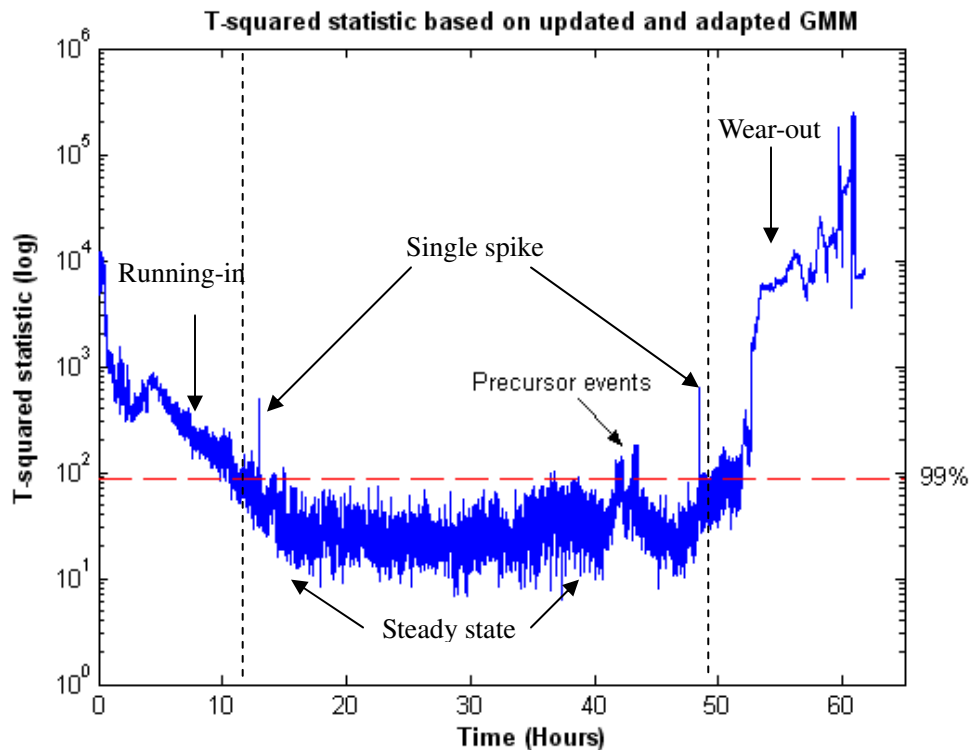


Figure 5.11 Example of the T-squared statistic and its detected abnormal features (test 20)

To achieve the second aim, threshold set-up strategies are investigated and developed. The quality of this work is evaluated by two parameters-anomaly detection rate (or true positive rate; the higher the value is, the better the strategy is) and false alarm rate (the lower the value is, the better the strategy is). Furthermore, in order to demonstrate that the Gaussian-EVT threshold strategy (see section 5.2 for details) is superior to other evaluated methods as well as the importance of effect of model adaptation on threshold set-up, a evaluation strategy is designed as Figure 5.12 shows. In the figure, models are divided as the adapted and non-adapted at the beginning; and then in each category, single Gaussian and GMM models are evaluated with their associated threshold set-up strategies. Ultimately, the evaluation of this study is intended to demonstrate that the adapted GMM model associated with the GMM-EVT threshold method can achieve the best anomaly detection result.

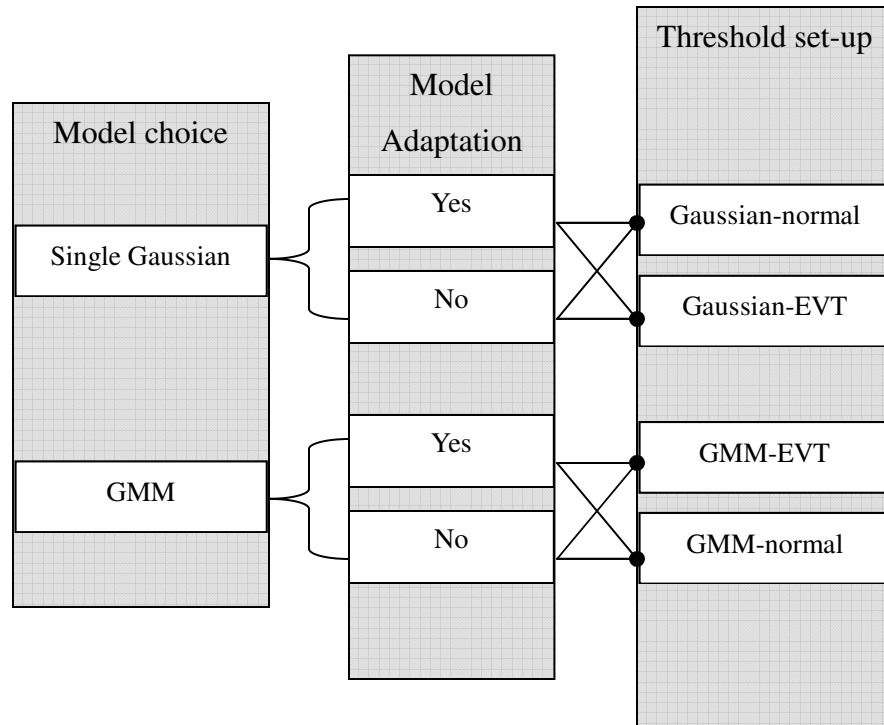


Figure 5.12 Evaluation strategy of different models and thresholds

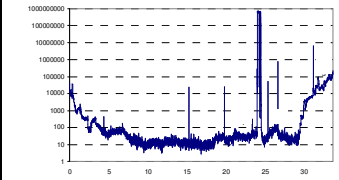

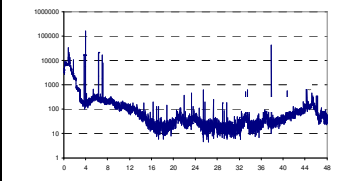

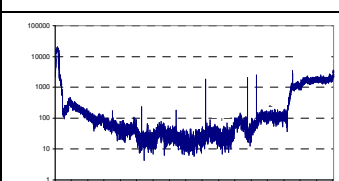
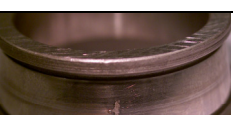
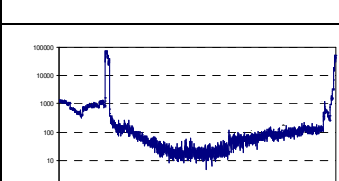

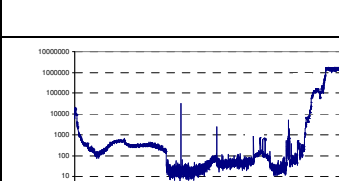

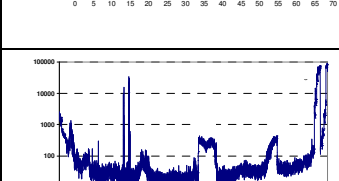

5.3.2. Overview of the anomaly detection results

Table 5.2 summarises the anomaly detection results with Hotelling's T-squared statistic of

all the 11 evaluated run-to-failure tests (in the second column of the table). It is noted that these results were predicted based on the adapted anomaly detection model with 18 clusters obtained in Section 4.5.4; and according to the well known bath-tub-curve and discovered early interesting signs, the series of the testing results can be described as 4 stages: running-in, steady state, early interesting signs and wear out. The discussion on the test results will thus be focused on these stages. In the third, fourth and fifth columns of the table, the descriptions of the calculated statistics were presented, and the photographic evidence of the mostly damaged bearing components is provided in the sixth column. The purpose of presenting this table is that the links between the detected abnormal trends or features and the observed elements of the bearings can be generally established or assumed, and then more analyses in details can be conducted through finding the correlations between the variables (i.e. energy at inner race frequency) and implementing more physical evidences (i.e. information of the collected debris).

From this table, it is found that the results of the T-squared statistic vary significantly from one another and their associated bearing element inspection with different damaged size (level of damage) verifies that the T-squared testing results can reflect the spallation size of the bearings. Figure 5.13 (a) illustrates the weight loss of the damaged component and for all the 12 run-to-failure tests. It is seen that the weight loss of some tests is very limited, i.e. tests 13-15, while others have obvious weight loss, such as tests 16 and 20. Correspondingly, the maximum T-squared value of the testing results were also plotted in Figure 5.13 (b), and it was found that the values obtain a similar trend with the weight loss of the element. Therefore, it is interesting to find that, if the run-to-failure data is tested against the same anomaly detection model, the obtained T-squared results have a correlation with the weight loss of the damaged bearing failure component.

Based on the T-squared statistics found in the detection results, the tests are set into two categories: 1) anomaly detection agrees with conventional analysis, e.g. tests 10, 14, 15, 16, 17, 18, 22 and 23; 2) model detects more and earlier signs than conventional analysis, e.g. tests 13, 20 and 21. Among these tests, tests 13, 16, 18 and 20 are more typical than others and will be presented separately in sections 5.3.2 to 5.3.4.

Test No.	T-squared statistic	Running stage	Links between the T^2 and bearing inspection	Additional features	Physical evidence
10		Runing-in:0-8 hours Steady state:9-27 hours Wear-out:28 hour till the end	Increasing trend starting from 28 h in the wear out region can be related to the visualized inner race spallation of bearing 2. Small abnormal trend extracted at 27 h might be the early failure indication of the bearing, and needs to be confirmed with more evidences.	Significant deviation at 24 h was caused by malfunction of the WSS sensors.	 304.9 mg loss to the inner race of bearing
13		Runing-in:0-18 hours Steady state:9-37 hours Insignificant abnormality: 38 hour till the end	Insignificant increasing trend starting from 38 h is extracted (not seen clearly in the original features) and can be related to the tiny area of spallation closed to the pre-indentation.	Additional multiple and single peaks can be found throughout the test.	 13 mg loss to the inner race of bearing 2
14		Runing-in:0-25 hours Steady state:25-72 hours Abnormal trend: 73 h till the end	Increasing trend starting from 73 h may be the reason for the detection of the defect on the inner race of bearing 3.	None	 11.3 mg loss to the inner race of bearing 3
15		Runing-in:0-10 hours Steady state:11-23 hours Abnormal trend: 24 h till the end	Increasing trend at the end of the test seems not consistent to the final inspection, in which bearing was found not damaged. The failure of the motor can be explained as the reason for the detection of the significant abnormal trend in the T^2 .	Motor was found slowed dramatically at the end of the test, and vibration of the whole system was also found severe.	 3 mg loss to the inner race of bearing 2
16		Runing-in:0-24 hours Steady state:25-62 hours Abnormal trend: 63 h till the end	Inner race of bearing 2 was found substantially damaged, and this may be related to the increasing trend starting from 62 h and two sets of early signs at 50 h and 57.5 h.	The increasing trend and discrete spikes at around 50 h and 57.5 h were detected	 1033.8 mg loss to the inner race of bearing 2
17		Runing-in:0-18 hours Steady state:19-100 hours Abnormal trend: 100 h till the end	Inner race of bearing 2 was found spalled, and this needs to be confirmed relating to the detected abnormal trends at the end of the test	Two deviation trends at 60 h and 80 h were detected, these two features might be the early indication of inner race spallation.	 61.19 mg loss to the inner race of bearing 2

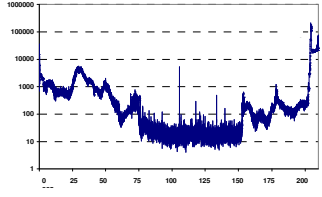

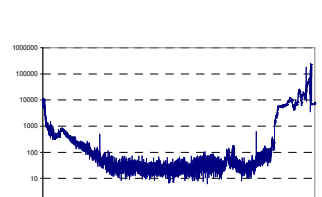

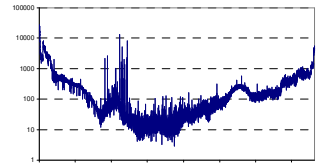
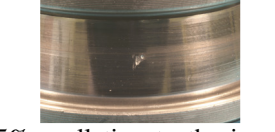
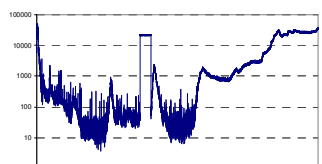

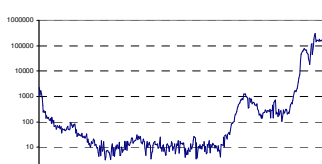
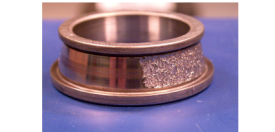
18		<p>Runing-in:0-75 hours Steady state:76-190hours Abnormal trend: 190 h till the end</p>	<p>One of the rolling elements was found significantly damaged, while a small area of spallation was found closed to pre-indentation on the inner race of bearing 2. These two features may be linked to the increasing trend starting from 200 h appeared in the statistic.</p>	<p>The spikes at about 150 h and 180 h were detected. These two features might be the early indications of the rolling element damage or inner race small defect.</p>	 <p>149 mg loss to the rolling elements of bearing 2</p>
20 (Mark)		<p>Runing-in:0-13 hours Steady state:14-47 hours Abnormal trend: 48 h till the end</p>	<p>Bearing inspection showed that one of the rolling element and the outer race of bearing 4 were found damaged, but it was also found that inner race of bearing 2 has a small spallation closed to pre-indentation. These can be linked to the increasing trend in the wear-out area appeared in the statistic.</p>	<p>A small amount of anomalies at 43 hours was detected (not seen in the original features). This invaluable feature might be the precursor of the outer race damage.</p>	 <p>255.79 mg loss to the outer race of bearing 4</p>
21 (Mark) (with AE)		<p>Test lasted for about 76 hours</p>	<p>Increasing trend starting from 60 hours. Need to identify the cause of this increasing trend in Chapter 6.</p>	<p>Early signs at 24 hours and second early signs at 55 hours can be correlated with AE amplitude.</p>	 <p>5% spallation to the inner race of bearing 2</p>
22 (Mark) (with AE)		<p>Test lasted for about 163 hours</p>	<p>Increasing trend starting from 120 hours. Need to identify the cause of this increasing trend in Chapter 6. Similar damage on the inner race of bearing 2 was found, but the dynamic responses in the T^2 may be caused by slightly longer spall.</p>	<p>Early signs at 70 hours and second early signs at 117 hours can be correlated with AE amplitude to find cause of it.</p>	 <p>5% spallation to the inner race of bearing 2</p>
23 (Mark)		<p>Test lasted for about 68 hours</p>	<p>Increasing trend starting from 50 hour can be correlated to the extensive spallation which was found on the inner race of bearing 2.</p>	<p>Early signs at 50 hours and needs to be examined to see whether it is the early warning of the final significant damage which started at 70 hours.</p>	 <p>50% spallation to the inner race of bearing 2</p>

Table 5.2 Summary of the anomaly detection results

Weight loss of failed components in different tests

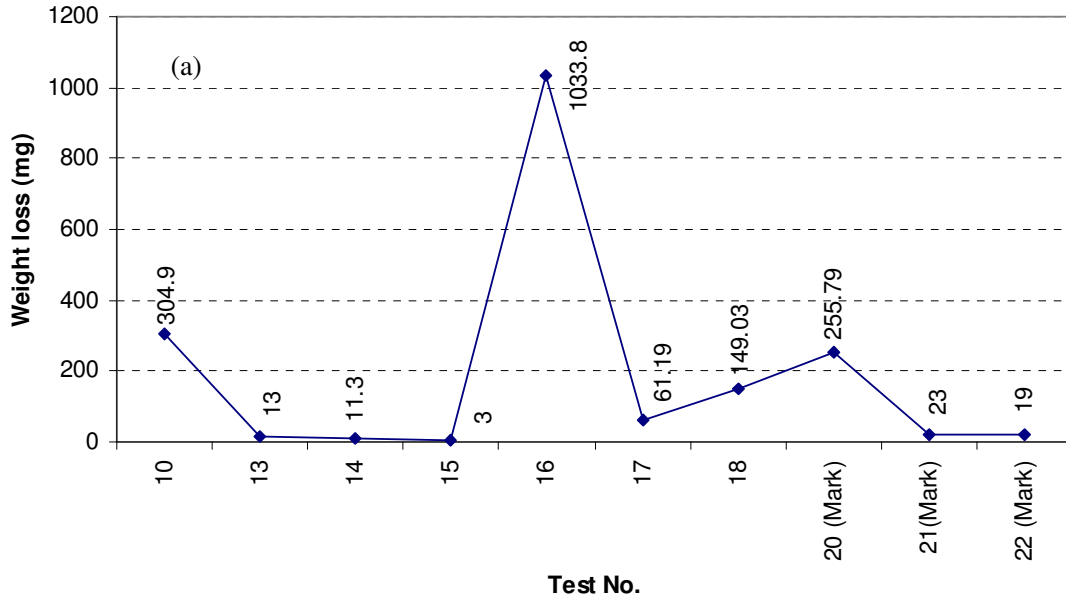


Figure 5.13 (a) Weight loss of the failed components

Maximum T-squared value in different tests

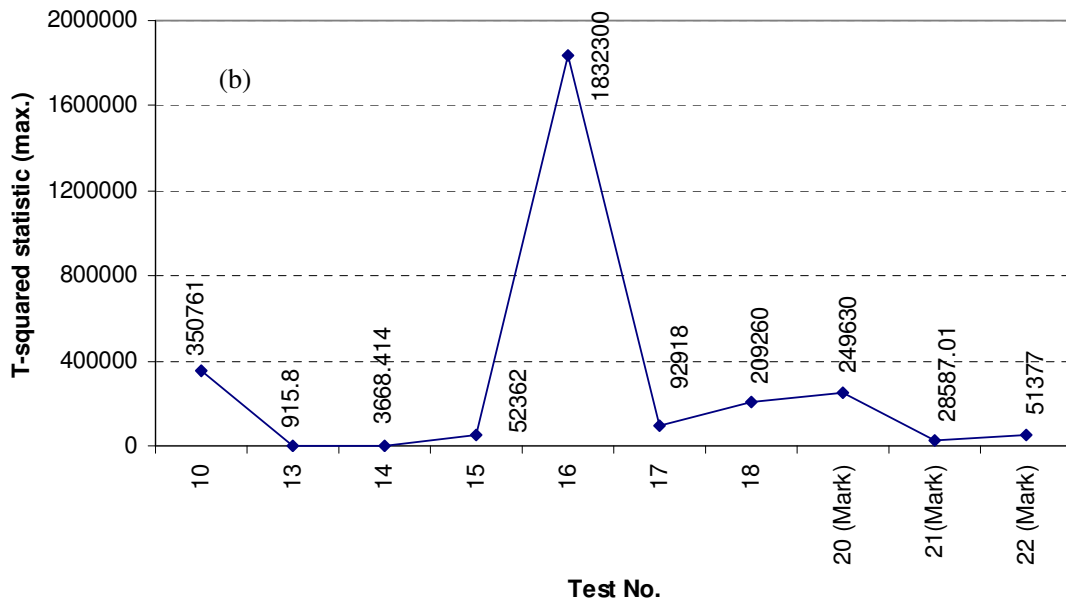


Figure 5.13 (b) Maximum T-squared value

In the following sections (5.3.3-5.3.4), detailed analyses of four typical tests which not only detect obvious abnormal trends but also extract additional signs are presented. The tests studied have different results, i.e. no significant damage at the end; expected

spallation to the inner race of the bearing, etc. The sections will not only focus on the anomaly detection rates with the combinations of different types of anomaly detection models (e.g. single Gaussian or GMM and before or after the model adaptation) and varied threshold level set-up strategies (introduced in sections 5.2 and 5.3.1) but will also explain the detected anomalies. The main aim of Section 5.3 is to demonstrate various anomaly detection results to validate that the:

- Developed adapted anomaly detection model based on GMM is more robust than other evaluated models, such as single Gaussian model, GMM before adaptation.
- Developed threshold level (GMM-EVT strategy) works well providing improved anomaly detection rate and false alarm rate. It is noted that, to conduct automated anomaly detection, every single point in the tests needs to be pre-defined as normal or abnormal. The method is to utilize a histogram to identify the data of every feature with high or low density with respect to normal or abnormal data. The detail of the method can be found in Section 4.2.
- Model can detect faults; faults can be correlated to the physical observations.

5.3.3. Tests 16 and 18: model can detect anomalies appeared in conventional analysis

(1) Test 16

Table 5.3 shows the anomaly detection results (T-squared statistic plus anomaly detection and false alarm rates) with the test 16 data based on different types of anomaly detection models (single Gaussian and GMM) and threshold level set-up strategies (Gaussian-normal, Gaussian-EVT, GMM-normal and GMM-EVT). It is seen in Table 5.4 that although the value and trend of the T-squared statistics calculated with different anomaly detection models show some differences, apart from the first index (T-squared statistic calculated on the basis of single Gaussian model before adaptation, three distinctive areas of interest can be found in other three indices; they are: running-in region (0-24 hours), early signs (50 and 57.5 hours) and wear out regions (starting from 60 hours). Hence, in this test, variation of the models seems to have little impact on the trend of the T-squared statistic.

However, apart from the trend examination, the anomaly detection rate (or true positive rate (TP)) and false alarm (FA) rates using different models associated with different threshold levels are also investigated. It is seen from Table 5.3 that the anomaly detection rate for the models after the adaptation are higher than the ones before adaptation, but the false alarm rate is also found to increase after model adaptation. Moreover, for both models before and after the adaptation, the threshold level set-up with the GMM-EVT strategy obviously decreases the false alarm rate compared to Gaussian-normal, Gaussian-EVT and GMM-normal strategies, while maintaining a relatively high anomaly detection rate.

Anomaly model	Threshold set-up		Detected trend
Before adaptation	Gaussian-normal	Gaussian-EVT	Testing result based on single Gaussian model
Single Gaussian	TP: 0.23	TP: 0.16	
	FA: 0.01	FA: 0.01	
Before adaptation	GMM-normal	GMM-EVT	Testing result based on GMM
GMM	TP: 0.79	TP: 0.41	
	FA: 0.37	FA: 0.05	
After adaptation	Gaussian-normal	Gaussian-EVT	Testing result based on single Gaussian model
Single Gaussian	TP: 0.94	TP: 0.92	
	FA: 0.64	FA: 0.52	
After adaptation	GMM-normal	GMM-EVT	Testing result based on GMM
GMM	TP: 0.97	TP: 0.95	
	FA: 0.23	FA: 0.14	

Table 5.3 Comparison of the anomaly detection of test 16 with different anomaly detection models and threshold set-up strategies

It is clearly found that the anomaly detection result with the adapted GMM and GMM-EVT threshold set-up strategy can not only detect significant abnormal trends but also provides the second highest anomaly detection rate (95%) and the third lowest false alarm rate (14%) out of eight evaluated strategies, which is the most satisfactory result. Therefore, this result is selected for further analysis.

Figure 5.14 (a) illustrates the four probabilistic boundaries with the test 16 data in a two-dimensional space (Vibration RMS and WSS1 RMS). Details of the boundary setting up can be found in section 5.2. From Figure 5.14 (a), test data are identified as normal or abnormal by seeing if they are projected within or outside the boundaries. It is seen that most of the data are falling outside the Gaussian-normal and GMM-normal boundaries, leading to the situation that more normal data might be wrongly identified as abnormal compared to that with the Gaussian-EVT and GMM-EVT boundaries. This result agrees to the anomaly detection results using the T-squared statistic and threshold level shown in Table 5.4. Figure 5.14 (b) shows the trend of conventional analysis-vibration RMS and WSS2 RMS traces throughout the test. These two parameters are plotted as a reference in comparison with the calculated T-squared statistic to examine whether the significant abnormal trends in conventional analysis can be detected by the T-squared statistic.

Figures 5.14 (c) illustrates the Hotelling's T-squared statistic of test 16 data based on adapted GMM and GMM-EVT threshold strategy. It is noted that this test has been previously published in [10], and result shows that WSS2 and vibration RMS can indicate bearing deterioration at the end of test. From Figure 5.14 (c), it is clearly seen that the running-in region (0-24 hours) and the wear out region (after 60 hours) have been successfully detected as anomalies by the predefined threshold. Also, obvious anomalies at around 50 hours and 57.5 hours as well as mild anomalies at about 38 hours are detected. The abnormal features at wear-out region and activities between 50 and 57.5 hours were also seen in the conventional time domain plots (see Figure 5.14 (b)). Hence, in this test, the abnormal trends clearly seen in the original plots can be successfully detected by the T-squared statistic and automatically isolated by the threshold level.

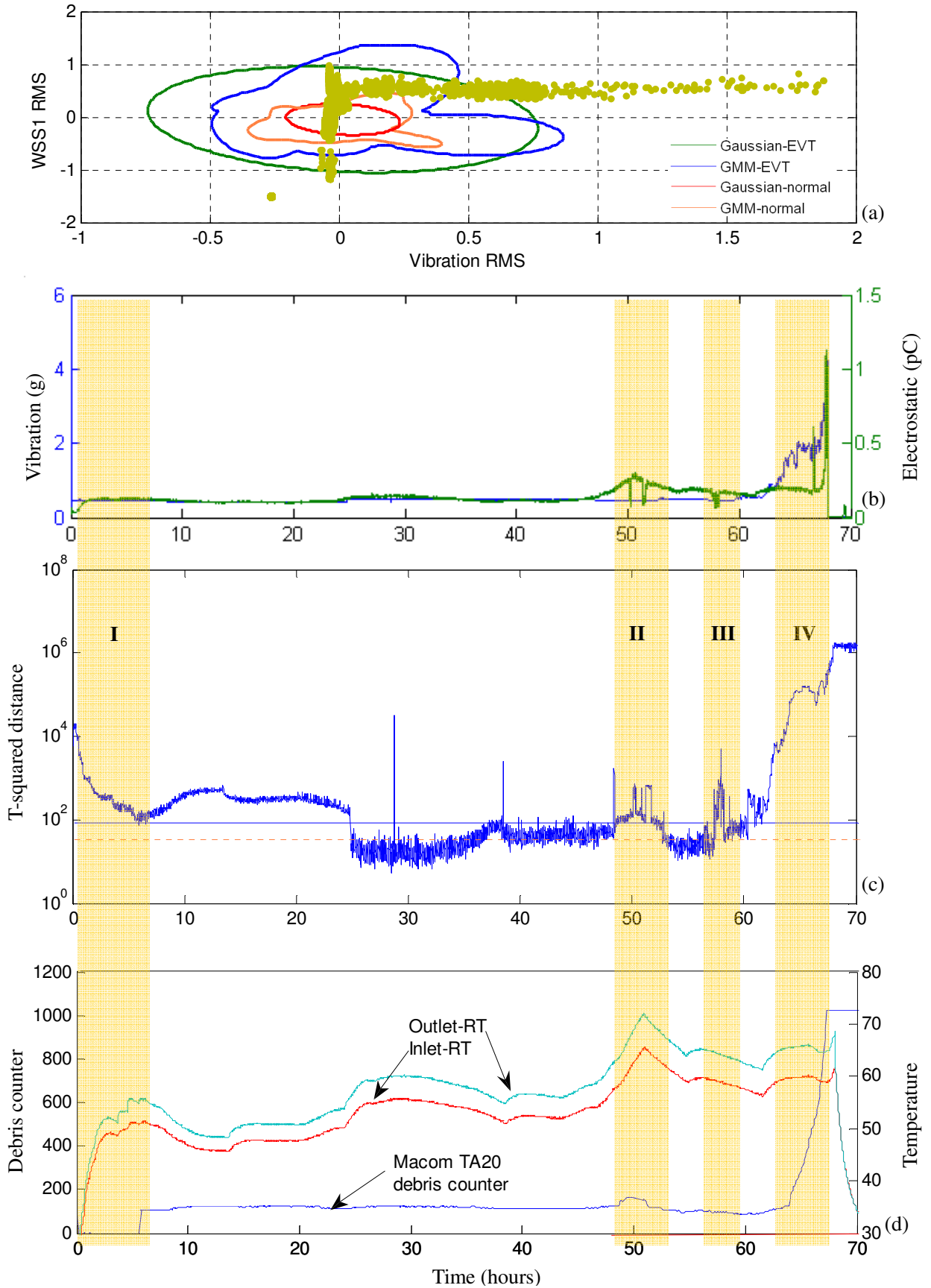


Figure 5.14 (a) Probabilistic boundaries, (b) vibration and WSS2 RMS traces, (c) anomaly detection results of the test 16 and (d) trends of the debris counter and temperature

Figure 5.14 (d) illustrates the thermocouple and TA20 ferromagnetic debris counter responses. It is found that abnormal activities can be detected at running-in and wear-out areas in these two parameters and it is understood that a relatively large amount of debris associated with a strong heating process usually occurs in these two regions [11]. Moreover, the debris monitoring results by the TA20 debris counter as well as thermocouples also confirm an obvious abnormal trend at around 50 hours. This trend at 50 hours provides the correlation and physical evidence for the detection of the early signs by the T-squared statistic. This early sign is debris related and might be the indication of the fatigue failure initiation generating a small amount of debris and higher temperature.

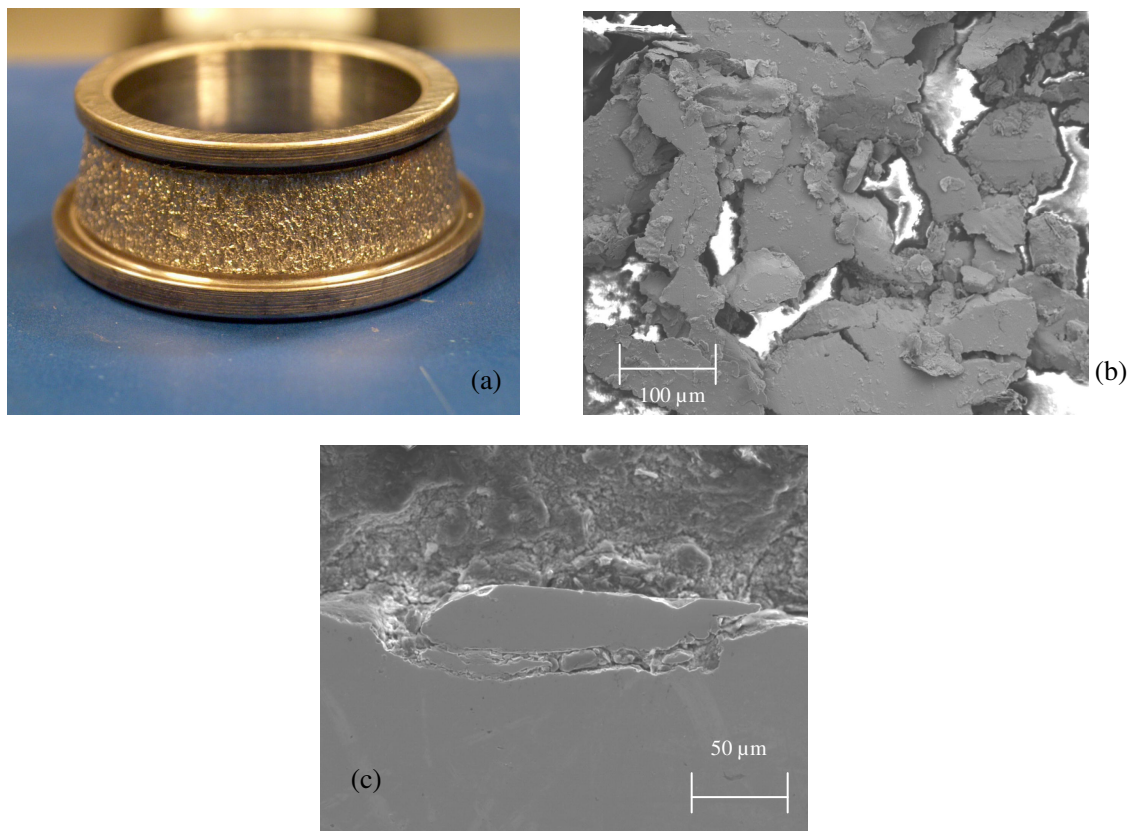


Figure 5.15 (a) Condition of inner race of bearing 2 after the testing 16, (b) SEM image of debris particles collected at the end of the test and (c) SEM image cross-section of the inner race from bearing 2 [10]

Furthermore, the photographic evidence clearly shows that failure of the bearings during the accelerated test originated from the inner race of bearing 2, as shown in Figure 5.15 (a). The inner race exhibited extensive spallation of 75% of the surface, and this provides evidence that the T-squared value at the end of test 16 is quite high and the T-squared

statistic is above the threshold value and consistently lasted for a longer time (i.e. 15 hours). Furthermore, the SEM image of the debris obtained from the magnetic plug (see Figure 5.15 (b)) shows a variety of debris sizes from tens to hundreds of microns. The larger particles generally show some cracking, probably due to crushing between the rollers and races. Figures 5.15 (c) shows SEM images of the damaged cross-section on the inner race of bearing 2. It can be seen that sub-surface cracking was the main failure mechanism that could produce both large and small fatigue debris particles with 5-50 μm .

From the above analysis, the significant abnormal events in the T-squared statistic (see Figure 5.14 (c)) starting from around 60 hours correspond to the fatigue failure of the inner race of bearing 2; this process also leads to a rise in running temperature and number or size of the particles (see Figure 5.14 (d)). Nevertheless, this diagnostic information can only be provided at the end of the test by examining the bearing surfaces, and diagnostic information of the early signs (abnormal events at 50 and 57.5 hours) cannot be given. So far, it is unclear whether the abnormal events detected between 50 and 57.5 hours are the initiation of the fatigue failure of bearing 2. Hence, further evidence and diagnostic information will be discussed in Chapter 6.

(2) Test 18

Test 18 is another run-to-failure test with a significant abnormal trend in the conventional time/frequency analysis, but having more complicated failure conditions (failure occurred to the unexpected bearing component); this might lead to difficulties in anomaly detection. Therefore, this test is worthwhile to be introduced to validate that the developed approach is robust to different situations.

Test 18 data will also be evaluated against different models and their associated threshold set-up strategies to give further evidence of optimisation of the anomaly detection models and threshold set-up strategies. Table 5.4 lists the anomaly detection results using the combinations of four types of models and four different threshold set-up strategies. It is found that the calculated T-squared trend based on different models varies significantly, especially between the models before and after the adaptation. Compared to the trends

based on models before adaptation, T-squared trends based on adapted models have a more obvious running-in region before the 80 hours, and clear early signs between 170-200 hours. This phenomenon demonstrates the benefit of conducting model adaptation, because more buried original abnormal trends (i.e. obvious running-in trend appeared in vibration and WSS2 RMS values, as Figure 5.16 (b) shows) have been recovered through model adaptation process. This advantage is also reflected by the significantly improved anomaly detection rate and false alarm rate, and found that the adapted GMM plus GMM-EVT threshold provides the best result, so their generated T-squared statistic will be further investigated.

Anomaly model	Threshold set-up		Detected trend
	Gaussian-normal	Gaussian-EVT	
Before adaptation			Testing result based on single Gaussian model
Single Gaussian	TP: 0.17	TP: 0.12	
	FA: 0.08	FA: 0.05	
Before adaptation			Testing result based on GMM
GMM	TP: 0.19	TP: 0.13	
	FA: 0.18	FA: 0.04	
After adaptation			Testing result based on single Gaussian model
Single Gaussian	TP: 0.99	TP: 0.98	
	FA: 0.21	FA: 0.10	
After adaptation			Testing result based on GMM
GMM	TP: 0.98	TP: 0.97	
	FA: 0.18	FA: 0.05	

Table 5.4 Comparison of the anomaly detection of test 18 with different anomaly detection model and threshold set-up strategies

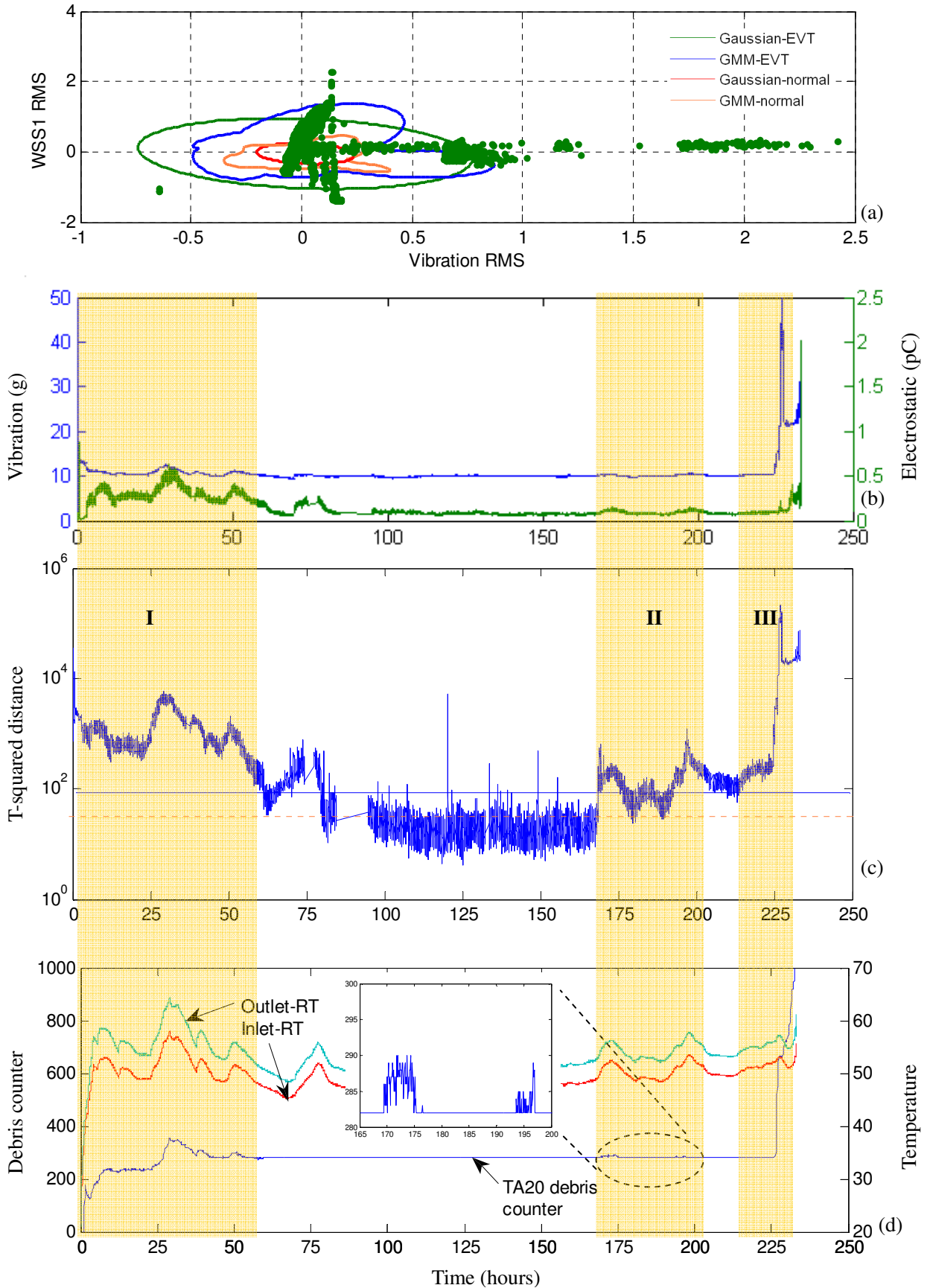


Figure 5.16 (a) Probabilistic boundaries, (b) vibration and WSS2 RMS traces, (c) anomaly detection results of the test 18 and (d) trends of the debris counter and temperature - 220 -

Figures 5.16 (a) and (b) illustrate the two dimensional features of test 18 on four probabilistic boundaries and time history of conventional RMS analysis respectively. Figure 5.16 (c) illustrates the calculated Hotelling's T-squared statistic based on adapted GMM model and GMM-EVT threshold level. The T-squared statistic shows that it agrees with the conventional analysis in Figure 5.16 (b) and has a similar trend with significant consistent abnormal events over the first 80 hours and at the end of the test. Moreover, a number of events between 170 and 200 hours (early signs) are also found to exceed the threshold line in the statistic. In Figure 5.16 (d), the parameters of thermocouple and Macom debris counter were also found to have activities in the above-mentioned periods of time. The events of the first 80 hours and the ones at the end of the test can be explained as the running-in and wear-out areas respectively, and the generated debris and the heating process were usually accompanied with the running process. However, for abnormal events between 170 and 200 hours, the debris and the rise in the temperature are also clearly seen, and these activities might be fatigue initiation of the component, but this needs to be corroborated by further evidence.

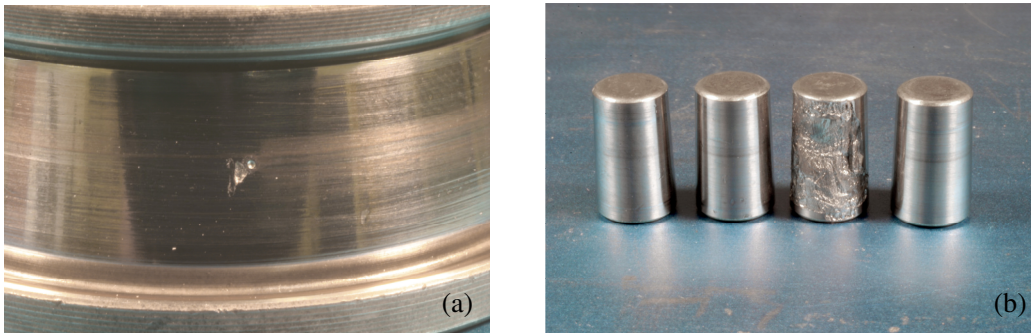


Figure 5.17 (a) Condition of inner race and (b) rolling elements of bearing 2 after the testing 18 [10]

Figure 5.17 illustrates pictures of the inner race and rollers from bearing 2 after the test. It was found that one of the rolling elements was unexpectedly fatigue damaged; therefore, the activities detected in the T-squared statistic at the end of the test could be due to the failure of the rolling element in bearing 2. However, a small damage near the artificial indentation was also found on the inner race of bearing 2; thus, the cause of the abnormal activities between 170 and 200 hours can not be determined, since they could be from either the initial spallation of rolling element or inner race damage. In Chapter 6, more analyses will be conducted to see whether the precursors detected between 170 and 200 hours are caused by inner race defect or initial spallation of the rolling element.

5.3.4. Tests 20 and 13: model can extract information not seen in conventional analysis

The two examples introduced above have shown that the approach developed in this study can not only detect abnormal situations that shown in the conventional Vib/ES RMS analysis, but also provide satisfactory anomaly detection and false alarm rates with the developed model adaptation approach and GMM-EVT threshold. However, some abnormal features are supposed to appear insignificantly and cannot be seen clearly in the conventional analysis, and it is hoped that the approach can discover or enhance these invaluable anomalies.

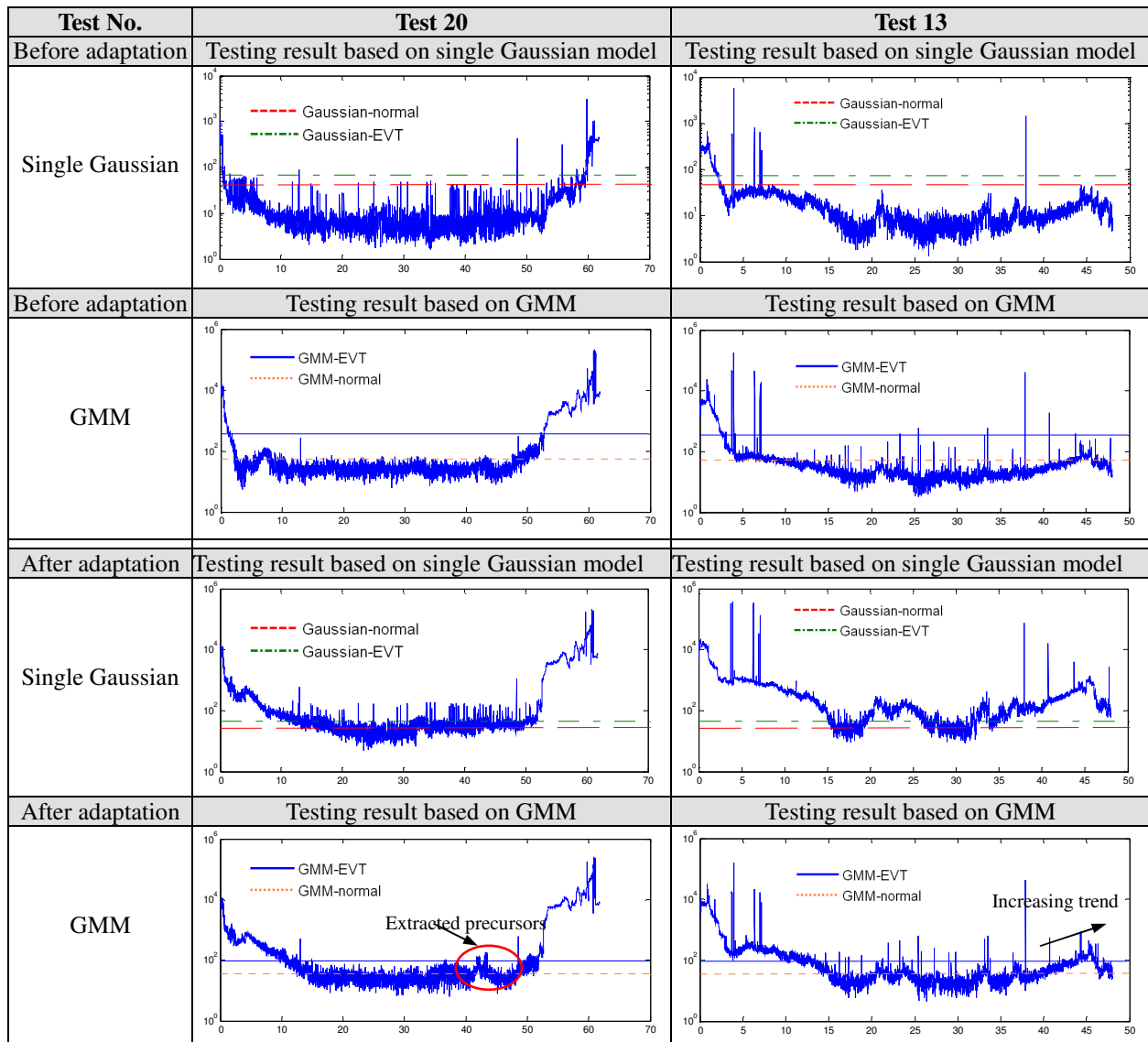


Table 5.5 Comparison of the anomaly detection of tests 13 and 20 with different anomaly detection models and threshold set-up strategies

(1) Test 20

Test 20 is a type of data that no early signs of failure can be seen clearly in the conventional time/frequency domain analysis. Figure 5.18 (a) shows the RMS values of the vibration and WSS3 sensors, and found that before the wear-out region starting from 54 hours, no obvious additional abnormal activities can be found. Therefore, data was run through the developed models to see whether any additional knowledge can be revealed. Table 5.5 gathers the anomaly detection results based on different models and their associated thresholds. It is found that all four T-squared statistics appeared in Table 5.5 have obvious running-in and wear-out regions, but results based on single Gaussian models (both before and after model adaptation) appear to show much more noise in the steady state than those based on GMM. Furthermore, from the results based on GMMs, the adapted GMM model can produce more information that a clear early interesting sign can be discovered at around 43 hours, and this interesting sign cannot be found in the statistic based on non-adapted GMM model. This is because the original model (prior to adaptation) included unexpected anomalies in the training data (tests 9-12). Thus, the non-adapted anomaly detection model could not be sensitive to the anomalies which are similar to the ones in the training data. This information is invaluable, since it not only provides a possible early indication of bearing failure, but also validates the benefit of model adaptation again (was demonstrated in the test 18 for the first time), although it needs more physical evidence to verify the meaning of its existence.

Figure 5.18 (b) shows the Hotelling's T-squared statistic of test 20, which was obtained using the adapted GMM model. This figure shows that there is evidence of this behaviour at the start of the test, which is due to the bearing running in [11]. The wear out to failure (i.e. high wear rate) from about 54 hours dominates the signals in this period. However, as discussed above, an interesting evidence of changes in advance of 54 hours can be found: an overall increase at around 36 hours, with obvious peaks between 43 and 45 hours. It is also noted that these detected prognostic activities could not be seen clearly in the conventional time/frequency analysis (shown in Figure 5.18 (a)).

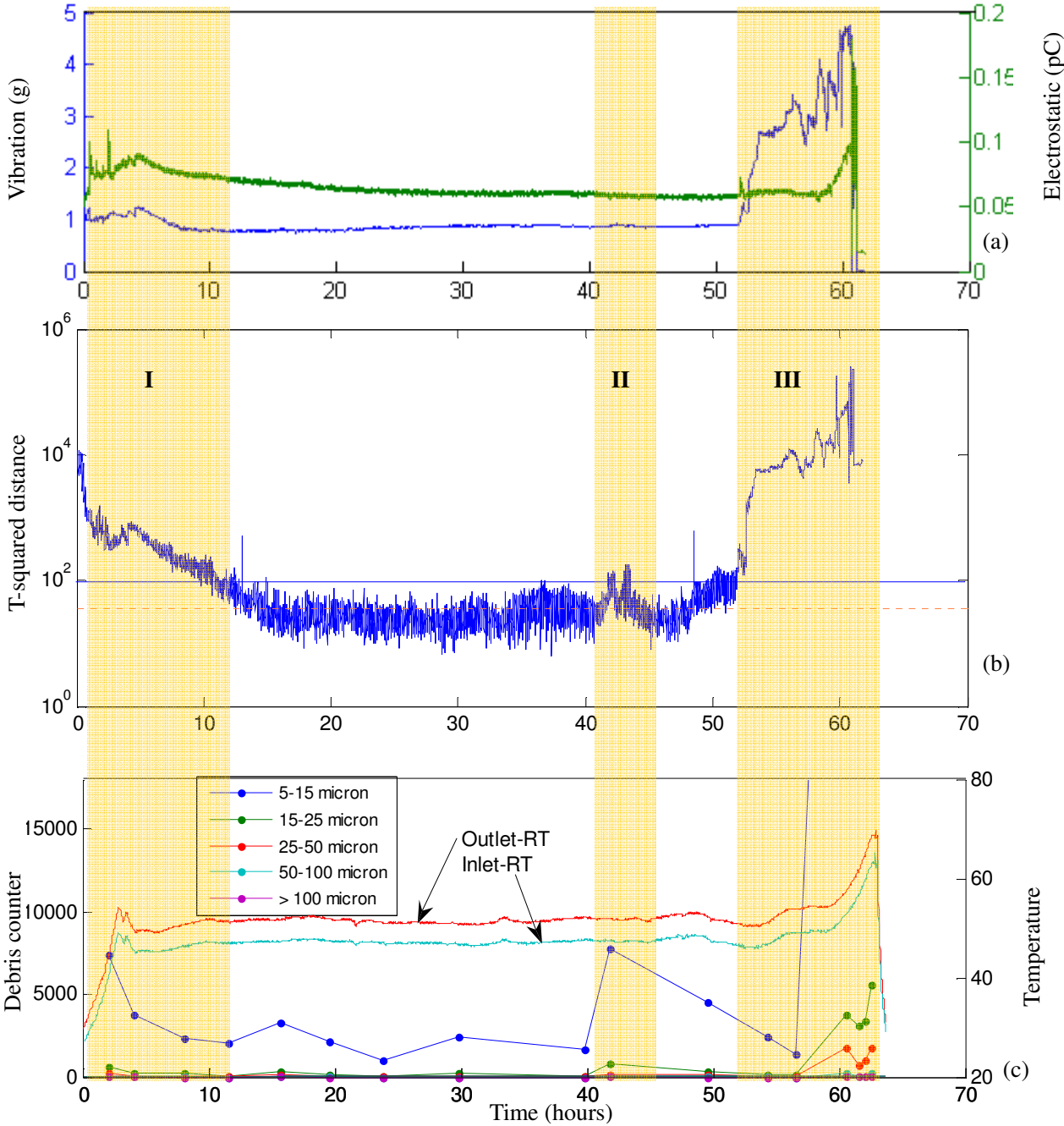


Figure 5.18 (a) vibration and WSS2 RMS traces, (b) anomaly detection results of the test 20 and (c) trends of the debris counter and temperature

According to the above analysis, the following three periods of interest are detected as abnormal by the developed approach: the first period of interest at the beginning of the test due to the running-in process; the second period of interest between 43 and 45 hours is the early sign before the wear-out region; the third period of interest starting from 54 hours is the wear-out region and associated with significant damage to the bearing

surfaces, this physical phenomenon has been validated by the final bearing surface inspection. Figures 5.19 (b) and (c) show the outer race and rolling elements of the bearing 4 respectively. It is found that significant fatigue damage has occurred on these two components, and it is clear that the increasing trend of the wear-out region is caused by the spallation of material. So far, the cause of the third period of interest between 43 and 45 hours visualized in Figure 5.18 (b) is unclear; they may result from wear debris generated by delamination from the pre-indent on the inner race of bearing 2 or surface debonding by subsurface cracks. Figure 5.19 (a) illustrates the SEM image of the pre-indent area on the inner race of bearing 2; it is found that the initial stage of spallation with minor material removal is found and close to the pre-indentation; and second period of interest that appear between 43 and 45 hours might be the reflection of this physical process. However, the early signs are possibly the prognostic indication of the fatigue initiation of the failing rolling elements or outer race of bearing 4, as Figures 5.19 (b) and (c) show. Therefore, the cause and location of this abnormal situation needs to be confirmed by further analysis.

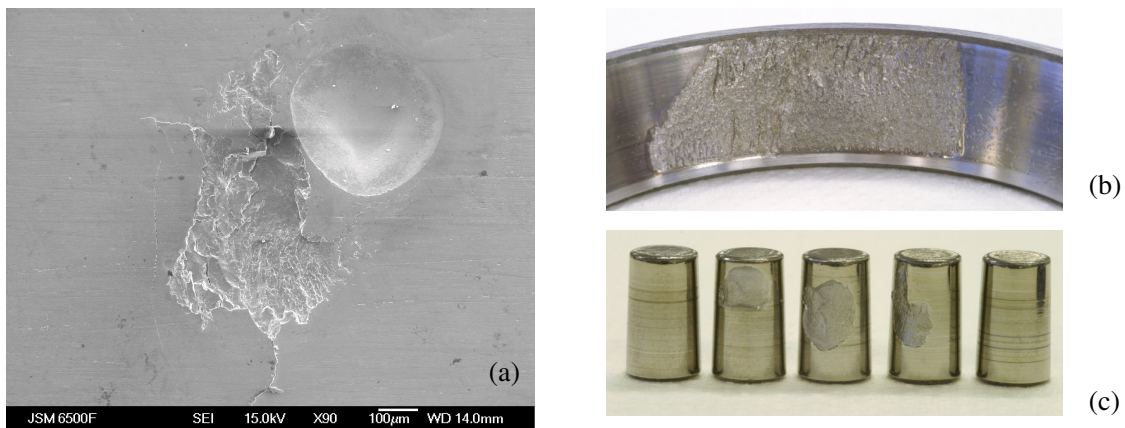


Figure 5.19 (a) Condition of inner race, (b) outer race and (c) rolling elements of bearing 4 after the testing 20[11]

Apart from the photographic images, it is also shown in Figure 5.18 (c) that the off-line debris counter PODS detected increases in debris production with a size between 5-15 μm at the time of 43 hours. This correlates well with the detected activities by the T-squared statistic over the similar period of time. Hence, these abnormal events are invaluable, as they could be used to detect debris-related abnormal conditions that result from distress or fatigue failure initiation of the bearings.

(2) Test 13

Test 13 is a test which did not show extensive damage apart from a moderate spall close to the pre-indentation on the inner race of bearing 2. Figure 5.20 shows the inner race of bearing 2 after the running of the test. It is seen that a tiny spallation occurred close to the pre-indentation. The data from this test is evaluated to see whether the models can make an appropriate response to a relatively moderate fault conditions within the bearing and extract any indicating signals of the moderate abnormal conditions.

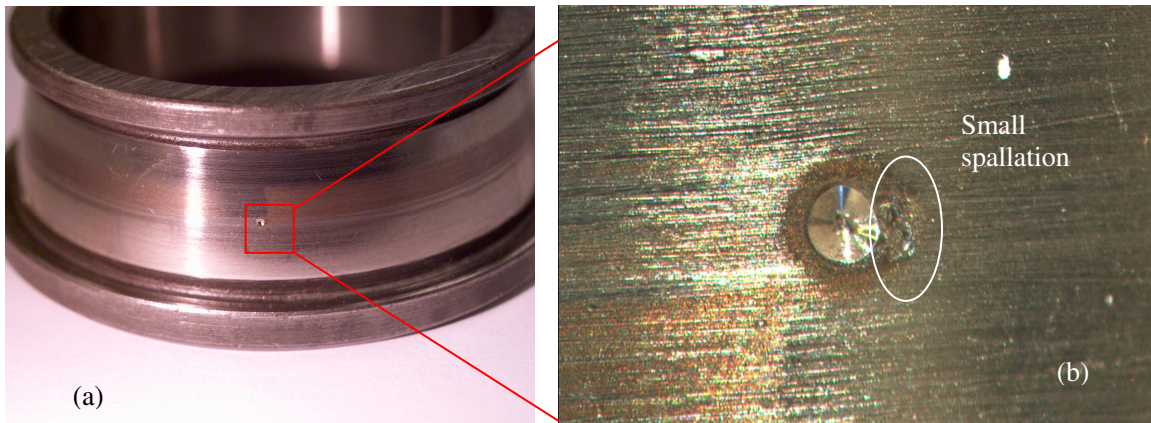


Figure 5.20 Condition of inner race of bearing 2 after the testing 13 [10]

Table 5.5 also shows the anomaly detection results from the test 13 based on different types of anomaly detection models and threshold level set-up strategies. It is found that the trends of the T-squared statistic based on different models have a very similar shape: an obvious running-in condition before 15 hours, a relatively flat trend of the steady state and a slight increasing trend starting from 40 hours. Therefore, the type of models seems to have little impact on the trends of T-squared statistic. However, the anomaly detection and false alarm rates based on different thresholds are varied seriously (the result is not shown in Table 5.5) and GMM model associated with GMM-EVT threshold achieved the most satisfactory result. Therefore, their generated T-squared statistic, as Figure 5.21 (b) shows, was selected for further analysis.

Apart from the running-in region between 0 and 15 hours, in region III (Figure 5.21 (b)) of the calculated T-squared statistic, only a very mildly increasing trend is found and is consistent with the temperature trend during this period (see Figure 5.21 (c)). This

indicates that the T-squared statistic could be used to detect small abnormal events which might be related to initial spallation on the inner race of bearing 2 accompanied with the heating process. Although this assumption needs more evidence to validate, the small damage shown in Figure 5.20 is probably the cause of the increased trends of T-squared statistic and temperature at the end of the test.

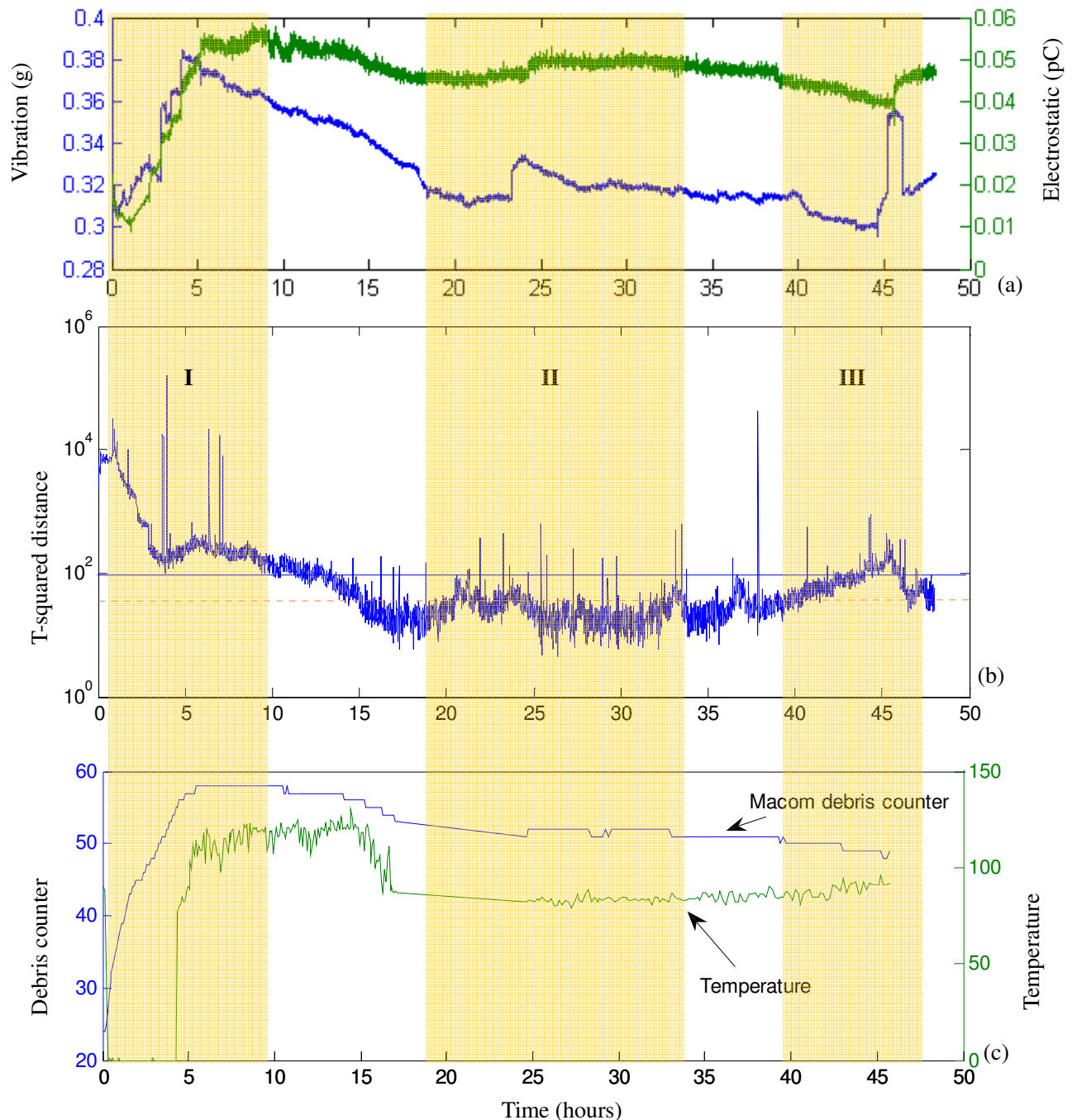


Figure 5.21 (a) vibration and WSS2 RMS traces, (b) anomaly detection results of the test 13 and (c) trends of the debris counter and temperature

Furthermore, two original features are also utilized as an implementation tool to the T-squared statistic for the anomaly detection. It is seen in Figure 5.21 (a) that in region III, an insignificant trend can also be seen in both vibration and WSS2 RMS values, which supports that in region III, abnormal events were highly suspected and found to deviate from the established model of normality.

From this case, it is found that the T-squared statistic has the potential to be used to detect small sized damage, and further diagnostic information such as whether the insignificant trend at the end of the test in the T-squared statistic can be sensed by the inner race related features etc. will be provided in Chapter 6.

5.3.5. Summary of the anomaly detection

There are three aims for conducting anomaly detection in this chapter. The first aim is to demonstrate that the developed anomaly detection model was robust and can make appropriate responses to anomalies. The second is to automatically isolate abnormal events from the normal ones using the proposed threshold set-up. Finally, the developed scheme is expected to extract more anomalies which can not be seen clearly in the original plots.

According to the obtained results and discussions, different anomaly detection strategies will often lead to quite different results which can be reflected by the trend of the anomaly detection index (T-squared statistic in this study), anomaly detection rate and false alarm rate. In Section 5.3, different anomaly detection strategies (with different anomaly detection models and their associated thresholds) have been applied to the four series of test, and found that the evaluated models (GMM or Single Gaussian and before or after adaptation) can all extract significant abnormal trends, such as running-in and wear-out regions, and changes of models seem to have a slight impact on the detection of obvious trends; this might be because the anomalies in these areas are large enough and deviated significantly from the centre of the models, so that the anomalies are outside the confidence regions of all varied models. However, the process of model adaptation seems

to have a significant influence on the detection of small abnormal trends, i.e. abnormal trends between 170-200 hours in test 18. It is found that the trends in this period of time were not seen on using non-adapted models. This is probably because a small amount of anomalies appeared in the original training data and the model adaptation process eliminates data which is similar to the small abnormal trends.

Moreover, in all four presented tests, it can be seen that the application of adapted GMM can increase the anomaly detection rate while the adoption of the proposed GMM-EVT threshold set-up method can decrease the false alarm rate. Although the combination of these two techniques cannot provide the best values of anomaly detection and false alarm rates individually, the most satisfactory can be obtained from this proposed approach. Furthermore, it is also found that the application of the GMM to build the anomaly detection model can optimise the trend of the anomaly detection index and extract hidden anomalies (test 20). This is invaluable, since this newly extracted information might be the early indication of bearing failure, and could provide warning in time. However, the detected anomalies should be explained associated with more physical information such as surface inspection of the bearings, debris counter, and thermocouple etc. The evaluated datasets have all been cross-correlated with this physical evidence, and have been found to have reasonable correlations.

Therefore, the testing results have shown that:

- 1) The anomaly detection approach can detect significant trends detected by the conventional time/frequency domain analysis (see tests 16 and 18)
- 2) The adapted GMM associated GMM-EVT threshold can provide most satisfactory anomaly detection and false alarm rate.
- 3) The anomaly detection approach can extract or enhance insignificant abnormal trends that may be relating to the prognostic information (see tests 13 and 20).

Although the proposed anomaly detection strategy obtains the advantages of simplifying pictures of the multivariate data, automatically dividing normality and abnormality, enhancing the detected abnormal trend and extracting more interesting information etc,

the fused anomaly detection index buried the labels of the physical variables which may contain invaluable ‘where’ or ‘what’ diagnostic information, and extra development methods are necessary to reveal more information about which variables are driving the detected or extracted anomalies. In this section, the causes of various detected anomalies are still unclear, and questions have been proposed in each presented case. Table 5.6 summarises the detected anomalies of the four presented tests in section 5.3, and proposed faults to be confirmed or diagnosed in the next chapter.

Test No.	Anomalies detected by the T-squared statistic	Faults to be confirmed or diagnosed
16	Running-in region (0-24 hours); early signs at 38, 50 and 57.5 hours; wear-out region (starting from 62 hours)	Is the early sign occurred relating to the inner race of bearing 2? Can the spallation occurred on the inner race of bearing 2 be sensed by the inner race related features?
18	Running-in region (0-80 hours); early signs between 170 and 200 hours; wear-out region (starting from 215 hours)	Is the early sign relating to the small damage on the inner race of bearing 2 or the initial spallation on the rollers of bearing 2?
20	Running-in region (0-12 hours); early signs between 43 and 45 hours; wear-out region (starting from 52 hours)	Is the early sign relating to the small damage on the inner race of bearing 2 or the initial spallation on the rollers of bearing 4?
13	Running-in region (0-17 hours); insignificant abnormal trend at the end of test (starting from 38 hours)	Can the small spallation close to the pre-indentation be sensed by the inner race related features?

Table 5.6 Summary of the detected anomalies and the issues need to be confirmed or diagnosed

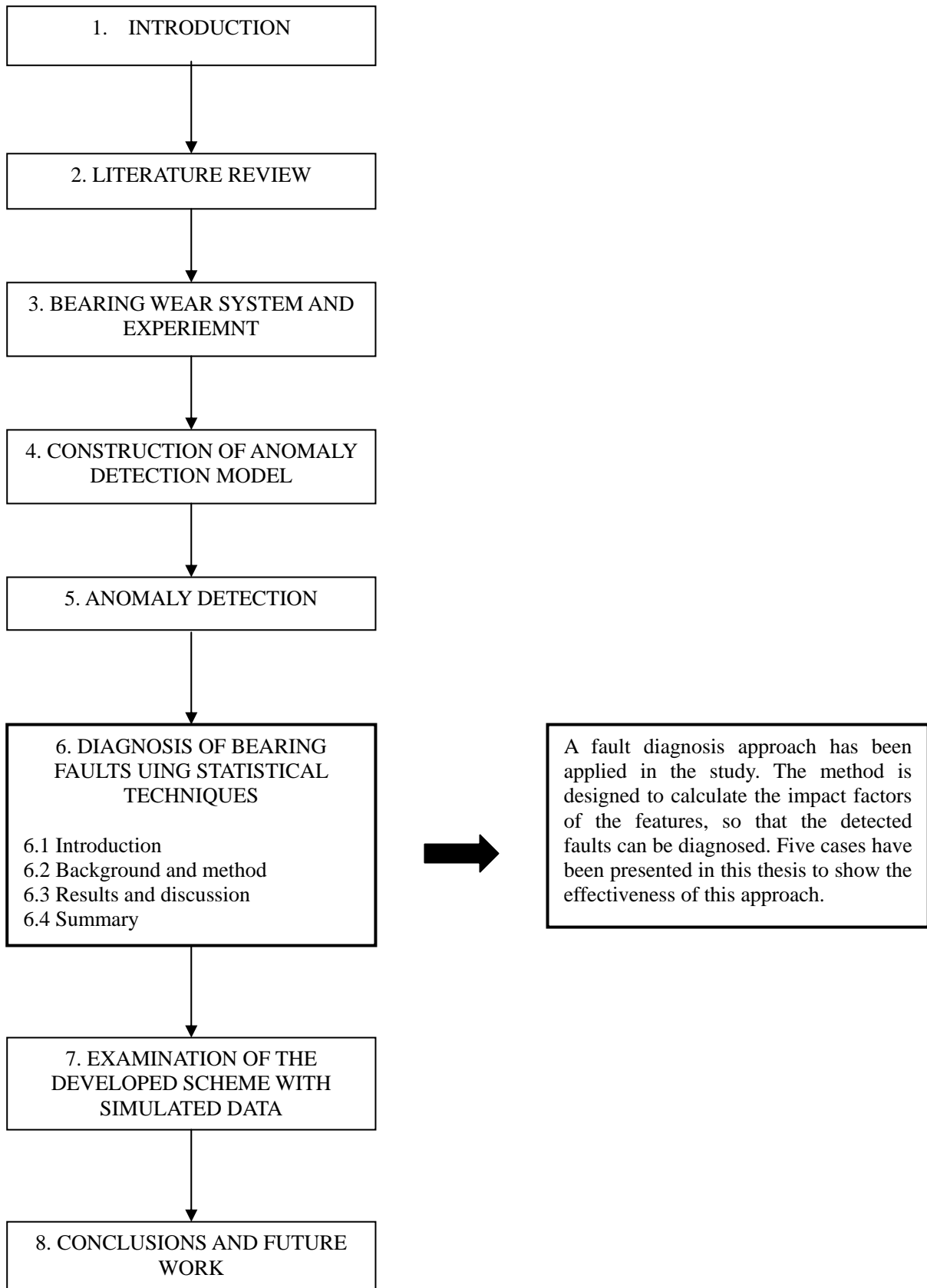
5.4 References

- [1] S. Roberts, Novelty detection using extreme value statistics. *IEE Proc.on Vision, Image and Signal processing* 146 (1999) 124-129.
- [2] S. Roberts, Extreme value statistics for novelty detection in biomedical signal processing, *1st International Conference on Advances in Medical Signal and Information Processing*, pp. 166-172.
- [3] R.A. Fisher, L.H.C. Tippett. Limiting forms of the frequency distributions of the largest and smallest members of a sample. *Proc. Camb. Philos. Soc.* 24, 1928.
- [4] E.J. Gumbel. *Statistics of Extremes*. Columbia University Press, New York, 1958.
- [5] D. Hawkins, *Identification of Outliers*, London, Chapman and Hall, 1980.

- [6] C. Bishop, Novelty detection and neural network validation, *Proc. IEE Conference on Vision and Image Signal Processing*, pp. 217-222, 1994.
- [7] J.V. Kresta, J.F. MacGregor, T.E. Marlin, Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng* 69 (1991) 35–47.
- [8] T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics Intell. Lab. Syst.* 28 (1995) 3–21.
- [9] P. Nomikos., J.F. MacGregor. Multivariate SPC charts for monitoring batch processes. *Technometrics.* 37 (1995) 41–59.
- [10] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie. Electrostatic wear monitoring of rolling element bearings. *Wear* 263 (2007) 1492-1501.
- [11] M. Craig, S.L. Chen, T.J. Harvey, R.J.K. Wood, K. Masuda, M. Kwabata, H.E.G. Powrie. Advanced condition monitoring of tapered roller bearings part I, with multiple sensing techniques. Accepted by Tribology International, 2008.

THESIS STRUCTURE

KEY OUTCOME



Chapter 6. Diagnosis of bearing faults using statistical techniques

6.1. Introduction

In Chapter 5, the detailed process of detecting abnormal events was demonstrated, but it seems insufficient to carry out an appropriate maintenance plan if the obtained information is just based on the unanalysed anomalies. Furthermore, the possible causes of the bearing faults were also proposed in the last chapter, and need further evidence to verify the assumptions. Therefore, in this chapter, it is necessary to develop a diagnostic method to recognize and isolate the detected faults. Moreover, all the features are fused by one of the univariate fault detection indices, and this process simplifies the complicated data plots making it is easy to establish a threshold level for the entire system; however, the trend of each feature, which could be correlated to the fault location or mechanism, is also masked. Hence, the motivation of revealing these features to analyse their influence on the detected anomalies is clear.

The research carried out in this section proposes a new diagnostic scheme based on the statistical data-driven approach. It explores the use of Principal Component Analysis (PCA) to detect developing faults in the case of a bearing wear rig. It is acknowledged that the early detection of faults is a challenge; the approach requires small fault conditions are tested first as well as severe fault conditions. In order to verify the capability of PCA to detect bearing abnormal conditions, the PCA-based model was initially established for normal operating conditions. Any unexpected event such as a fault condition of the bearing would cause a significant deviation from the established PCA model. In this work, the Hotelling's T-squared and Squared Prediction Error (SPE) statistics were calculated to detect faults. Apart from the fault detection, the contribution values of each monitored variable to the statistics of Hotelling's T-squared and SPE were also calculated to isolate the dominating variables with the highest contribution values, so that the causes for the detected abnormal events could be analysed.

6.2. Background and method

6.2.1. PCA algorithm

The PCA is a technique of linear statistical predictors that has been applied in various fields of science. The primary objective of the approach is data compression. PCA was originally developed in the 1900s, and has now re-emerged as an important technique in data analysis [1]. It seeks to find a series of new variables which extracts the maximal amount of variability in the data and linearly maps multi-dimensional data onto lower dimensions with minimum loss of information (see Appendix D for PCA algorithm).

6.2.2. Fault detection indices

The PCA-based fault detection is done via statistical hypothesis tests on two indices, Hotelling's T-squared and squared prediction error (SPE) statistics, respectively in the principal component subspace and residual space. The T-squared statistic measures distance to the origin in the principal component subspace, and the SPE statistic, also known as squared prediction error (SPE), is used to measure variability that breaks the normal process correlation and hence that usually indicates an abnormal condition (see Appendix D for the definition of fault detection indices).

6.2.3. Contribution plots

Contribution plots are known as the diagnostic tools for fault identification [2-4]. The indices used for fault diagnosis with contribution plots could be T-squared and SPE introduced in the last section. A contribution plot on PCA scores indicates the significance of the effect of each variable on the T-squared statistic, while the contribution plots on SPE indicate the significance of the effect of each variable on the SPE statistic. Therefore, the variables with the largest contribution value are considered major contributors to the fault. However, it is argued that although contribution plots are easy to calculate, they are often used and required to interpret the plots.

The individual contribution of the j^{th} physical variable to the i^{th} principal component score is:

$$c_j^{t_i} = p_{ij} x_j \frac{t_i}{\lambda_i} \quad (6.1)$$

Where t_i is the value of the i^{th} principal component score, λ_i is the variance of the i^{th} principal component scores. p_{ij} is the element of i^{th} row and j^{th} column of the loading matrix that represents the normal condition. x_j is the value of j^{th} physical variable.

The individual contribution of the j^{th} physical variable to the Hotelling's T-squared distance is actually the j^{th} physical variable to the overall principle component scores which have the first largest eigenvalues. Thus, the contribution value to the Hotelling's T-squared distance could be calculated as:

$$c_j^{T^2} = \sum_{i=1}^r c_j^{t_i} \quad (6.2)$$

6.2.4. Steps for bearing fault detection and diagnosis with PCA

- a) Acquire original training data matrix X which represents the normal conditions.
- b) Establish the PCA subspace and obtain the loading matrix P with normal attributes.
- c) Select k principal components in the established PCA model, based on the cumulative percentage (capture more than 80% variance in most of the applications). The first 3 steps could be achieved through the calculations described in Section 6.2.1.
- d) Obtain the new sampling data in the fixed time interval, scale it to Y .
- e) Calculate the PCA statistics, T-squared distance and SPE.

- f) Plot variable contribution chart for violation of SPE or T-squared distance to find out the dominating process variable.
- g) Scale the T-squared distance statistic contribution of each physical variable to the unit variance.

$$\hat{c}_j^{T^2} = \frac{c_j^{T^2}}{\sqrt{\gamma_j^{T^2}}} \quad (6.3)$$

Where $\gamma_j^{T^2}$ is the covariance matrix of the contribution value calculated with recorded normal data.

However, in the case of a large number of features, it can be difficult to interpret the contribution plots, the contribution from one variable may be propagated to other variables in calculation [5]. This ‘transferring’ effect can reduce the difference between contributing and non-contributing variables. An effective approach is to use multiple-block fault diagnosis [6]. Qin et al. [7] report the use of hierarchical contribution plots for this process. The process is usually partitioned into several blocks based on the knowledge of the process in terms of sensing technologies. It is suggested in [6] to divide the process into sections that describe a unit or a specific physical operation (i.e. variables which are used to monitor one specific bearing). After grouping into blocks, the block T-squared statistic and SPE are calculated and examined against their control limit; the variable contributions in that block can then be further examined.

6.3. Results and Discussion

6.3.1. Feature selection

As discussed above, the bearing wear rig is implemented in this project as the case study to validate the developed approaches. Five groups of features, as Table 6.1 shows, have been selected to evaluate the conditions of the monitored bearings. The details of these features can be found in Chapter 3.

All bearings		Bearings 1	Bearings 2 & 3	Bearing 4
OLS	Vibration	WSS1	WSS2	WSS3
RMS1	RMS	RMS	RMS	RMS
RMS2	Tacho Energy	Tacho Energy	Tacho Energy	Tacho Energy
Indicator	Cage Energy	Cage Energy	Cage Energy	Cage Energy
	Roller Energy	Roller Energy	Roller Energy	Roller Energy
	Outer Race Energy	Outer Race Energy	Outer Race Energy	Outer Race Energy
	Inner Race Energy	Inner Race Energy	Inner Race Energy	Inner Race Energy

Table 6.1 Selected features and their relevant functions

Within these features, vibration and electrostatic oil-line features are used to monitor the shocks and generated wear debris of the whole bearing system respectively, according to their characteristic; but the vibration features have the additional functions of revealing bearing element information (faults that occur on rollers, outer race or inner race). Moreover, since there are three electrostatic wear-site sensors on the system that monitor the four bearings respectively, WSS features cannot only indicate bearing element information but also the corresponding bearing numbers.

6.3.2 Establishment of the normal subspace

In order to carry out the fault detection, normal healthy data are collected to establish the reference model; this process is similar to the fault detection approaches developed in Chapter 5. However, in this work, two methods were chosen to extract the principal components of the normal healthy data to establish a PCA subspace. The first method is based on the percentage of captured variance. Here, the first 11 principal components captured more than 80% percent of the original data matrix (see Table 6.2 for details), as it can capture most of viability and eliminate most of noise in the data [5-7]. The second method is a graphical representation, where the last principal component to be chosen should show a sharp change on the graph. From Figure 6.1, 11 principal components should be selected. Therefore, the loading matrices of the first 11 principal components were extracted to establish the normal subspace. After the establishment of the PCA subspace representing the normal condition of the bearings, datasets from several experiments which have different wear conditions at the end of the test were examined to

verify the diagnostic capability of the proposed approach. The evaluated cases are the ones which have been introduced in the anomaly detection chapter, and the detected anomalies and their cause assumptions in Chapter 5 are expected to be traced and confirmed in the following sections.

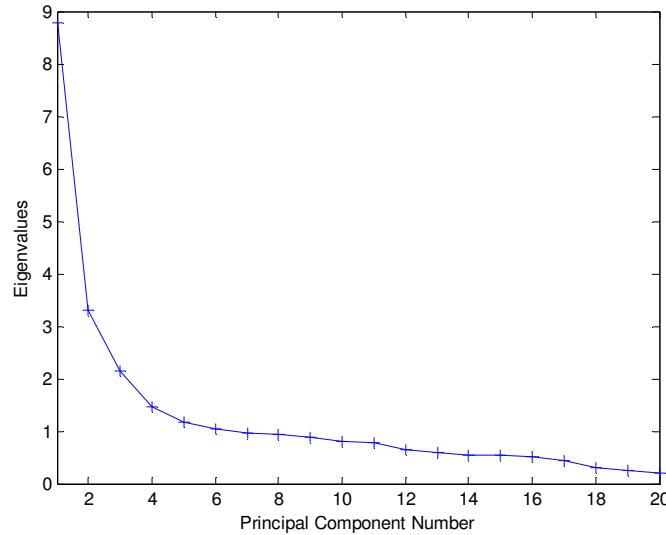


Figure 6.1 Eigenvalues of the principal components

Principal Component Number	Eigenvalue of Cov(X)	% Variance captured this PC	% Variance captured total
1	8.77E+00	32.48	32.48
2	3.31E+00	12.27	44.74
3	2.14E+00	7.92	52.66
4	1.46E+00	5.42	58.09
5	1.18E+00	4.38	62.46
6	1.03E+00	3.82	66.29
7	9.73E-01	3.6	69.89
8	9.38E-01	3.47	73.36
9	8.74E-01	3.24	76.6
10	8.14E-01	3.01	79.61
11	7.83E-01	2.9	82.51
12	6.49E-01	2.41	84.92
13	5.99E-01	2.22	87.13
14	5.42E-01	2.01	89.14
15	5.27E-01	1.95	91.09
16	5.03E-01	1.86	92.95
17	4.40E-01	1.63	94.58
18	3.06E-01	1.14	95.72
19	2.60E-01	0.96	96.68
20	2.06E-01	0.76	97.44

Table 6.2 Obtained principal components and their parameters

6.3.3. Case 1 (expected failure to the inner race of bearing 2, test 16)

As introduced in Section 5.3.3, the experimental result of this case has shown significant fatigue damage on the inner race of the test bearing 2. It is also seen from Figure 5.14 (c) that the Hotelling’s T-squared distance has a corresponding abnormal trend with significant and consistent novel events starting from 60 hours. These two discoveries reveal the correlation between the physical damage and detection of abnormal events in the monitored signals. However, it seems it is impossible to obtain diagnostic information from the anomaly index, indicating that the spallation occurred to the inner race of bearing 2. Furthermore, some precursors observed between 50 and 57.5 hours are also found to exceed the confidence limits; the reasons for these events have been assumed in Section 5.3.3 to be the fatigue initiation accompanied with a small number of debris. In the following figure, the contribution values to the T-squared statistic are calculated to see whether they can support presumptions and provide more information.

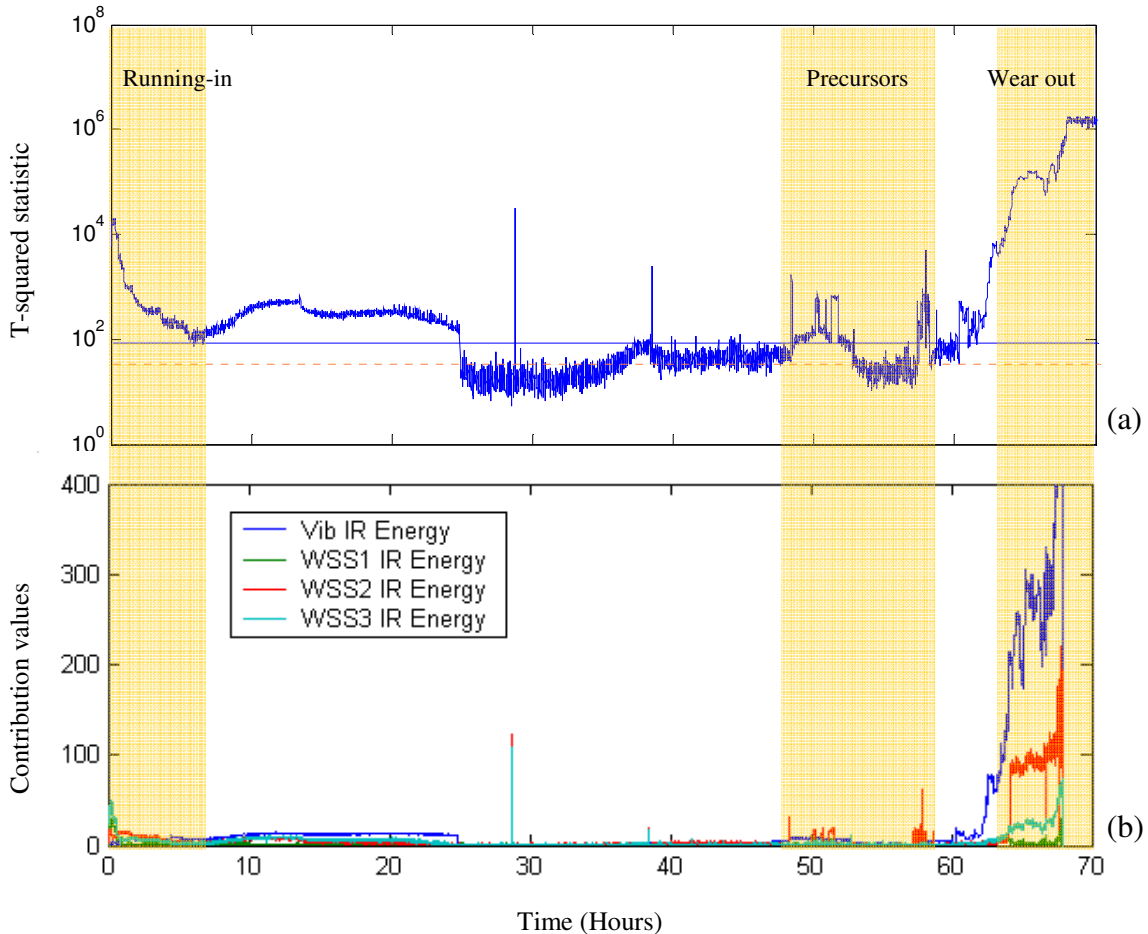


Figure 6.2 Contribution values of inner race energy of four sensors

Figure 6.2 shows the contribution values of the inner race energies of four sensors. From this figure, the rank of impact level of evaluated features to the periods of interest can be clearly seen, i.e. from 60 hours, the Vib. IR energy shows the most dominant impact on the trend, while the WSS1 IR energy is the least. Moreover, as the figure illustrates, there are three periods of interest which are worth examining in terms of the contribution values of the different bearing elements (rollers, outer race and inner race) of the sensors.

In order to have a greater insight of these values, the contribution figures of the three series of features in the fixed time intervals (running-in, precursor and wear-out) are plotted with individual plots for each bearing element, as Figure 6.3 shows.

Figures 6.3 (a)–(c) show the features' contribution values in the running-in region. It is clearly seen that the inner race elements dominate the activities, and Vibration and WSS2 inner race energies are the two strongest impact factors during the running-in region. Therefore, it is assumed that running-in activities are probably due to the modifications to the bearing 2 inner race surface giving contacting difference and impulse shocks.

For the prognostic information (see Figures 6.3 (d)–(f)), vibration inner race energy seems to have the strongest impact on the increasing level of the precursors between 49 and 52 hours, but WSS2 inner race energy contributes to the additional peaks. Moreover, WSS2 inner race and Rlg El energies are dominant for the precursors at 57.5 hours. Since WSS2 is used to monitor bearings 2 and 3 in which bearing 2 was pre-indented in the inner race and expected to fail, this diagnostic information could be obtained: 1) It occurred within bearing 2. 2) Activities are probably caused by an initial spallation at the artificial defect in the inner race of bearing 2. This correlates with the increasing activity seen on the TA20 particle counter. Although the cause of the precursors still cannot be confirmed, the diagnostic information that faults occurred on the inner race of bearing 2 enhances the confidence of the assumption.

It is clearly seen that vibration and WSS2 inner race energies dominate the wear-out region (see figures 6.3 (g)–(i)), which enhances the confidence that the extended

spallation occurred to the inner race of bearing 2. This is because large amounts of debris and hugely modified contact surfaces between the rollers and inner race of bearing 2 occurred due to the rolling contact fatigue generating high charge and strong impact energy, which are sensitive to WSS and vibration sensors, respectively. So far, the diagnostic information for the increasing trend of the wear-out region has been extracted from the features, and it is found to be consistent to the bearing surface inspection (see figures 6.3 (j)-(l)).

From the above analysis, it is found that the abnormal events detected between 0-2.5 hours were caused by the running-in process particularly seen within bearing 2. The WSS2 features could detect the abnormal events at 50 and 57.5 hours, which are presumed to be precursor events with a small amount of released debris from contact surfaces of the inner race of bearing 2. Finally, the contribution values of vibration and WSS show that the severe symptom occurred to the inner race of the bearing 2. Therefore, nearly all the abnormal activities are found on the inner race of bearing 2, which demonstrates a complete picture of component operating mechanism (i.e. from running-in, steady state, precursors and wear-out). Moreover, the contribution values of OLS features (see Figure 6.4) are also plotted and it is found that a significant impact is observed in the wear-out area, but seldom seen in the other regions, such as precursors and running-in; this might be due to the fact that the size of the debris generated in the wear-out region is big enough, and can be easily sensed by the OLS features.

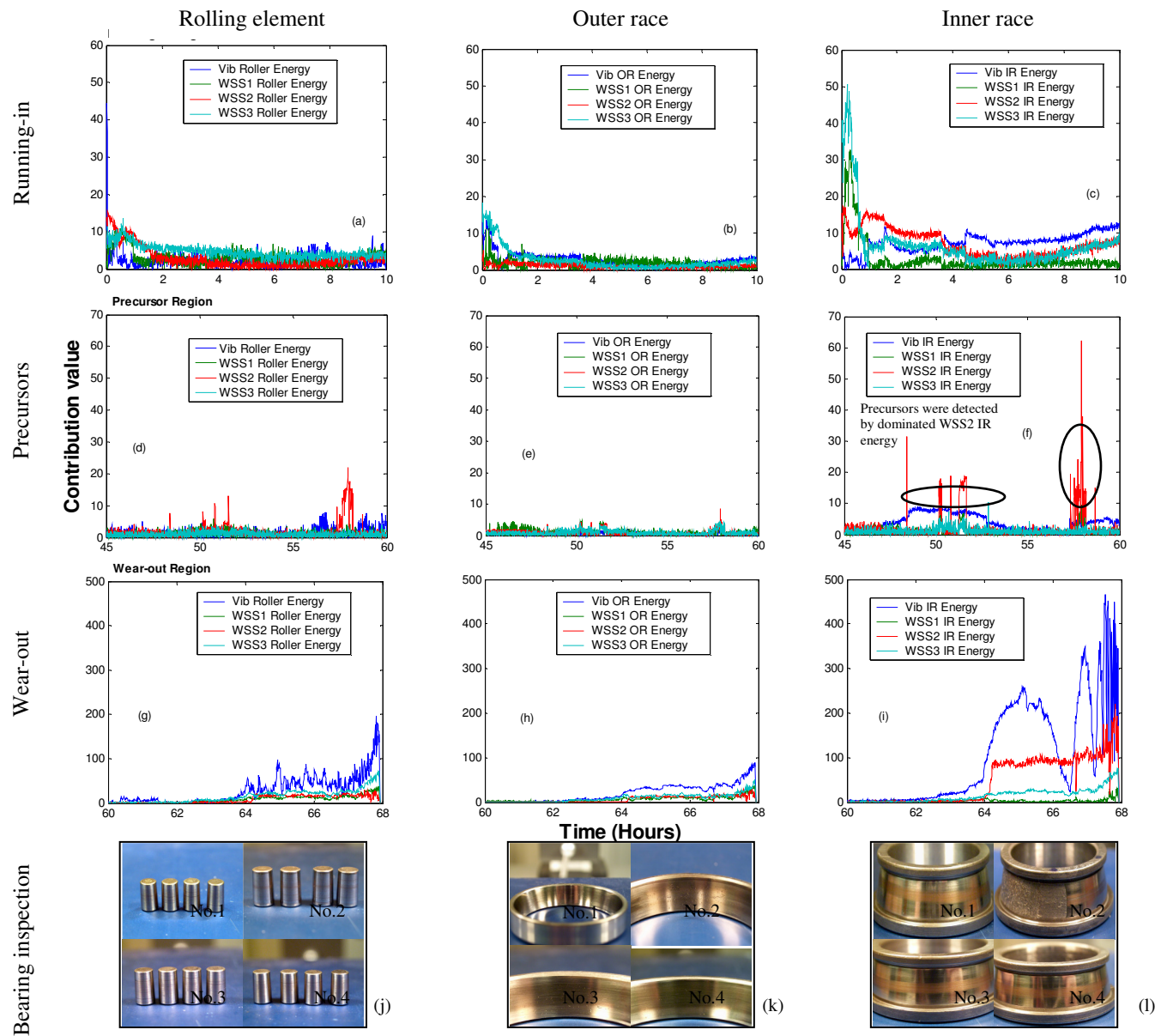


Figure 6.3 Contribution values of the bearing element energies at 3 periods of interest (test 16)

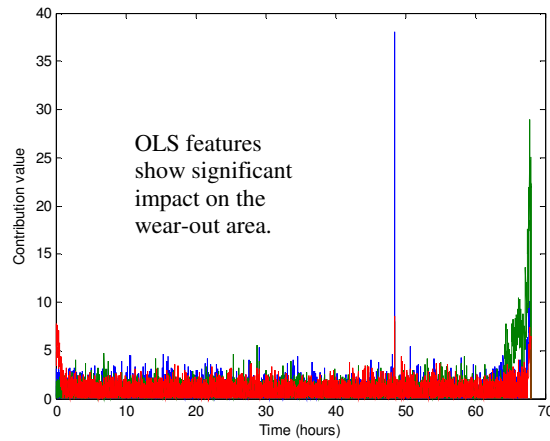


Figure 6.4 Contribution values of the OLS features (test 16)

6.3.4. Case 2 (unexpected failure to the rolling element of bearing 2, test 18)

Test 18 was another test with detected precursors and a significant abnormal trend at the end of running, but it had an unexpected failure condition-fatigue failure that occurred to one of the rollers of bearing 2. The anomaly detection index extracted a number of events of interest, but all this information raised the questions of where they occurred and what caused them. Are the precursors detected between 170-200 hours the early indication of the failed rollers of bearing 2, or the reflection of small spallation observed beside the pre-indent on the inner race of bearing 2? And can the contribution values indicate that the failure occurred to the rollers of bearing 2? In the following section, the calculated contribution values of bearing elements in the three periods of interest are analysed.

In the running-in region (see figures 6.5 (a)-(c)), the variable of WSS2 rolling elements occupies the leading positions with the highest contribution values. This phenomenon is consistently found in the contribution plots up to 80 hours in Figure 6.5 (a). This interesting phenomenon might be explained by the rolling elements of bearing 2 experiencing an installation damage or having surface defect, which was generating an amount of debris; but this assumption needs to be further confirmed by the later analysis.

Between 165 and 170 hours (see figures 6.5 (d)-(f)), it is found that the variable of the WSS2 rolling element feature has the greatest contribution to detected abnormal precursor events. Furthermore, the TA20 was also affected by this abnormal event (see Figure 5.16 (d)). As the WSS sensor might have responded to the generation of debris, and the TA20 detected occurrence of the debris, it is likely that wear debris generation occurred at this point in the test. For the period of 190 to 200 hours, the WSS2 Rolling element energy still shows the highest contribution value, while vibration features have low impact. Moreover, the variable of TA20 detected particles again, so the causes to the abnormal event at the time of 190 to 195 hours could also be explained by the generated wear debris. This illustrates the advantage of the WSS sensors in detecting wear debris or surface charge from the monitored components. Furthermore, as the WSS2 rolling element variable indicates the dominating role, it is reasonable to assume that the precursor events happened to the rollers of the monitored bearing 2 due to entrainment of particles into the rolling element-raceways contact. Therefore, at this point, it is almost certain that the precursors detected between 170 and 200 hours are the early failure indication of rolling elements of bearing 2, not the true reflection of the small spallation next to the pre-indent on the inner race of bearing 2.

In the last example of 220 to 225 hours (see figures 6.5 (g)-(i)), it is found that all the vibration variables are found to provide significant contributions to the abnormal T-squared value, while only the WSS2 energies at bearing rolling element defect frequency show obvious high contribution values compared to the energies at other defect frequencies. This suggests severe wear occurred to the rolling elements of bearing 2. Furthermore, contribution values of OLSs are insignificant (see Figure 6.6), which might be because the amount of debris is small and additional signals were buried in the background level.

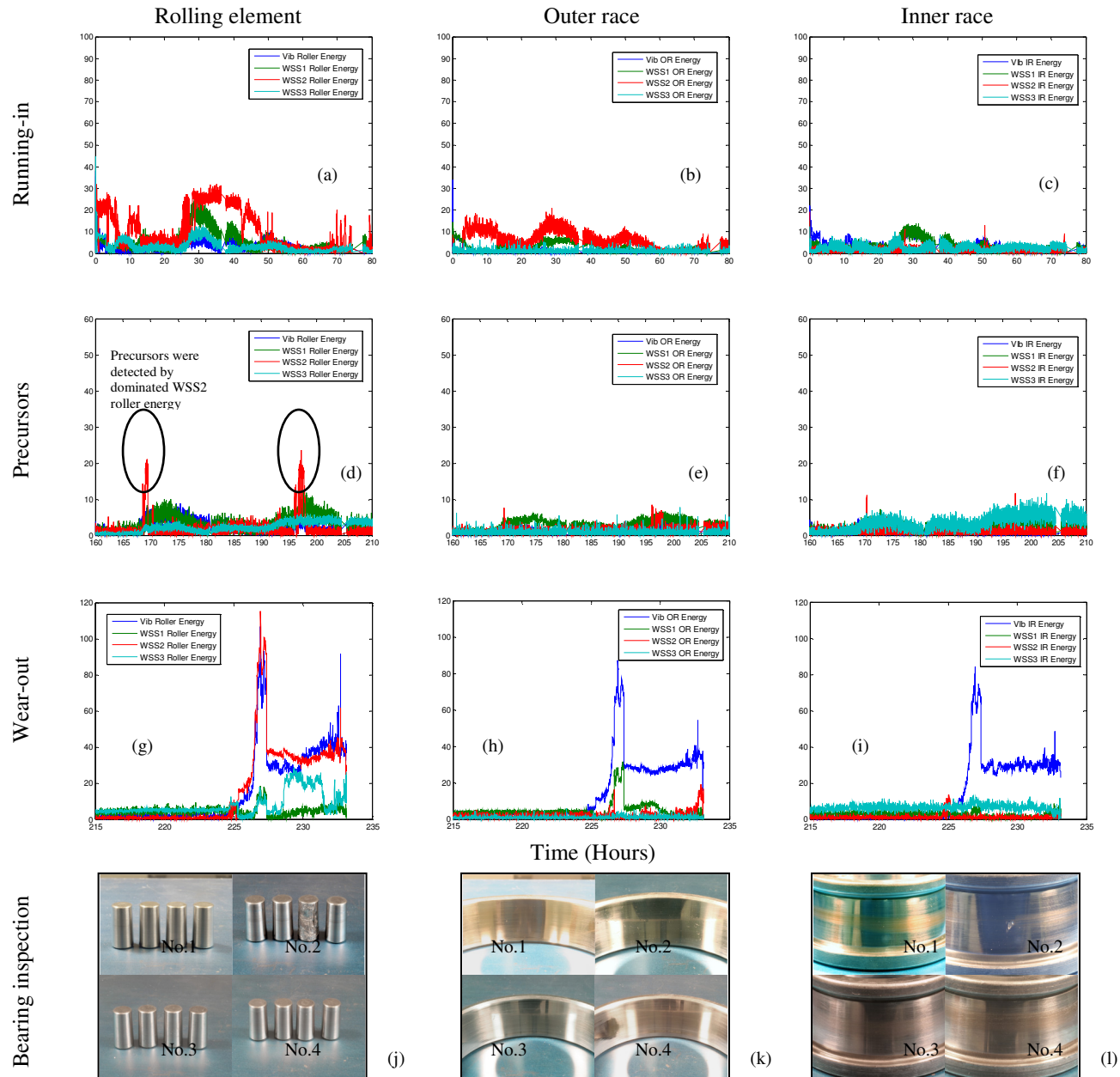


Figure 6.5 Contribution values of the bearing element energies at 3 periods of interest (test 18)

Figure 6.5 (j-1) shows different parts of the monitored bearings at the end of the test. It can be seen that one of the rolling elements of the monitored bearing 2 was significantly damaged, while the races are relatively healthy. Hence, diagnostic information from the contribution values is consistent to the failure component-rolling element of bearing 2, and can obtain a full picture of rolling contact fatigue of this failed unexpected failure component.

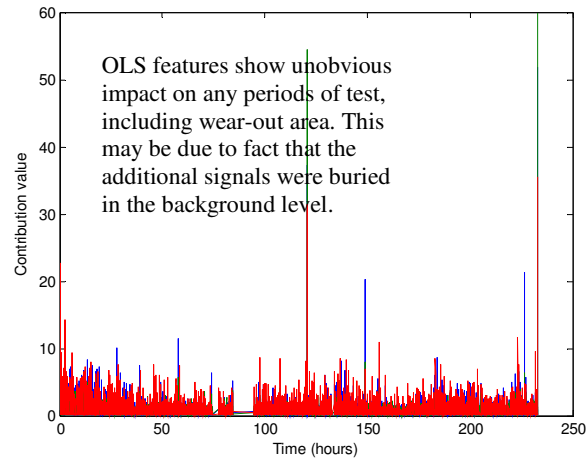


Figure 6.6 Contribution values of the OLS features (test 18)

6.3.5. Case 3 (unexpected failure to the rolling elements of bearing 4, test 20)

Although the two examples introduced above have different failing components, they still occur in bearing 2 which is expected to fail. In test 20, however, fatigue failure occurred to an unexpected bearing which was without pre-indent. Therefore, contribution values are expected to be examined to see whether they are robust enough and can respond to different situations. Furthermore, precursors at 43 hours which cannot be visualized clearly in the original features are discovered in anomaly detection, and off-line debris has confirmed that it is the debris related anomalies, but the location and cause of it is still unknown, and fatigue initiation of the failure component of bearing 4 and abnormal signal of the small spallation on the inner race of bearing 2 (similar to the test 18) are the two possible reasons, and require confirmation in the following section.

Figure 6.7 (a)-(c) illustrate the absolute contribution values of the bearing component energy for the four sensors during the running-in region. It is seen that the contribution value of Vib. Rlg El energy is predominant between 0–6 hours, and WSS3 Rlg El energy also shows relatively high impact between 2–6 hours. As for the other examined features, especially the energy at outer and inner race defect frequencies, they appear to be independent of the running-in behaviour. Combining the explanation of the running-in process in [8], the dominant vibration phenomenon is probably due to the rolling element surface protuberance, defect or debris giving impulse shocks.

At the second period of interest (prognostic activities between 36 and 43 hours), it can be seen that the WSS3 Rolling element energy dominates, while the other Rlg El energies show insignificant contributions to the detected abnormal trend. As the contribution values of outer race energy shows, Figure 6.7 (e), vibration and WSS3 elements have a relatively high impact, but seem to be less significant than WSS3 Rlg El energy. Figure 6.7 (f) shows the contribution value of IR energies, the difference between the previous two bearing elements, Rlg El and OR, and it is not easy to distinguish which variables are the major contribution resource to the prognostic activities, and which have a relatively weak impact. According to the above comparison, it is found that the WSS 3 Rlg El and outer race energies obtain the highest impact on the detected prognostic events (between 37 and 43 hours), from which two diagnostic facts can be discovered: 1) abnormal conditions occurred within the bearing 4, since it was monitored by WSS3; 2) as the WSSs are designed to detect charge generated between the contacting surfaces and wear debris, it can be inferred that the delamination or initial spallation occurred between the contacting surfaces of rolling elements and outer race of bearing 4, and the small amount of debris generated is also detected by the PODS off-line debris counter (see Figure 5.18 (c)).

Furthermore, nearly all the features contribute to the abnormal condition at the third period of interest (wear-out region between 54-63 hours), but the Rlg El and OR energies of vibration and WSS3 are the strongest (see figures 6.7 (g)-(i)). This could be explained by spallation within bearing 4 due to rolling contact fatigue, hence large amounts of

debris and modified contact surfaces cause high charge levels and high vibration energies which are reflected by WSS3 and the vibration sensor respectively. Moreover, the OLS variables also show a significant trend during this period (see Figure 6.9 for details). This is because the OLSs are installed in the oil line, and the generated fatigue debris suspended in the oil flow passes through the OLSs, causing significant variations.

The final bearing inspection revealed that the rolling elements and outer race of bearing 4 were fatigue damaged (see Figure 5.19 (b) and (c)), which shows that not only is the diagnostic information obtained from the above analysis correct, but also the assumption that the dominant vibration energy at the running-in region is due to the defects on the rolling element of bearing 4 is reasonable, and these defects might be the reason causing the unexpected failure of bearing 4.

However, the individual rolling element contribution value between the WSS2 and WSS3 cannot be obviously distinguished from the figure, and to reveal which variables are the main factors to drive the detected precursor events, combined contribution values were calculated for the different groups of variables based on the established PCA model. Figure 6.8 illustrates the calculated result of the combined contribution values with the different variables. It can be seen from Figure 6.8 (d) that the contribution value with the WSS3 feature set obtains the abnormal discrete trend at around 43 hours, while the score with other feature sets show a relatively flat trend. As the WSS3 sensor monitors bearing 4, it is presumed that the precursor events happened within bearing 4 which had the potential to fail within a short time. Furthermore, the abnormal event trend at the wear-out area is mainly driven by the vibration and OLS variable sets, as the spalling occurred generating large amounts of wear debris. At the end of the test, four bearings were disassembled to examine fault conditions, and it was found that the outer race and rolling elements of bearing 4 were fatigue damaged, which supports the assumption that precursor events occurred within the bearing 4 and will lead to the failure of the system.

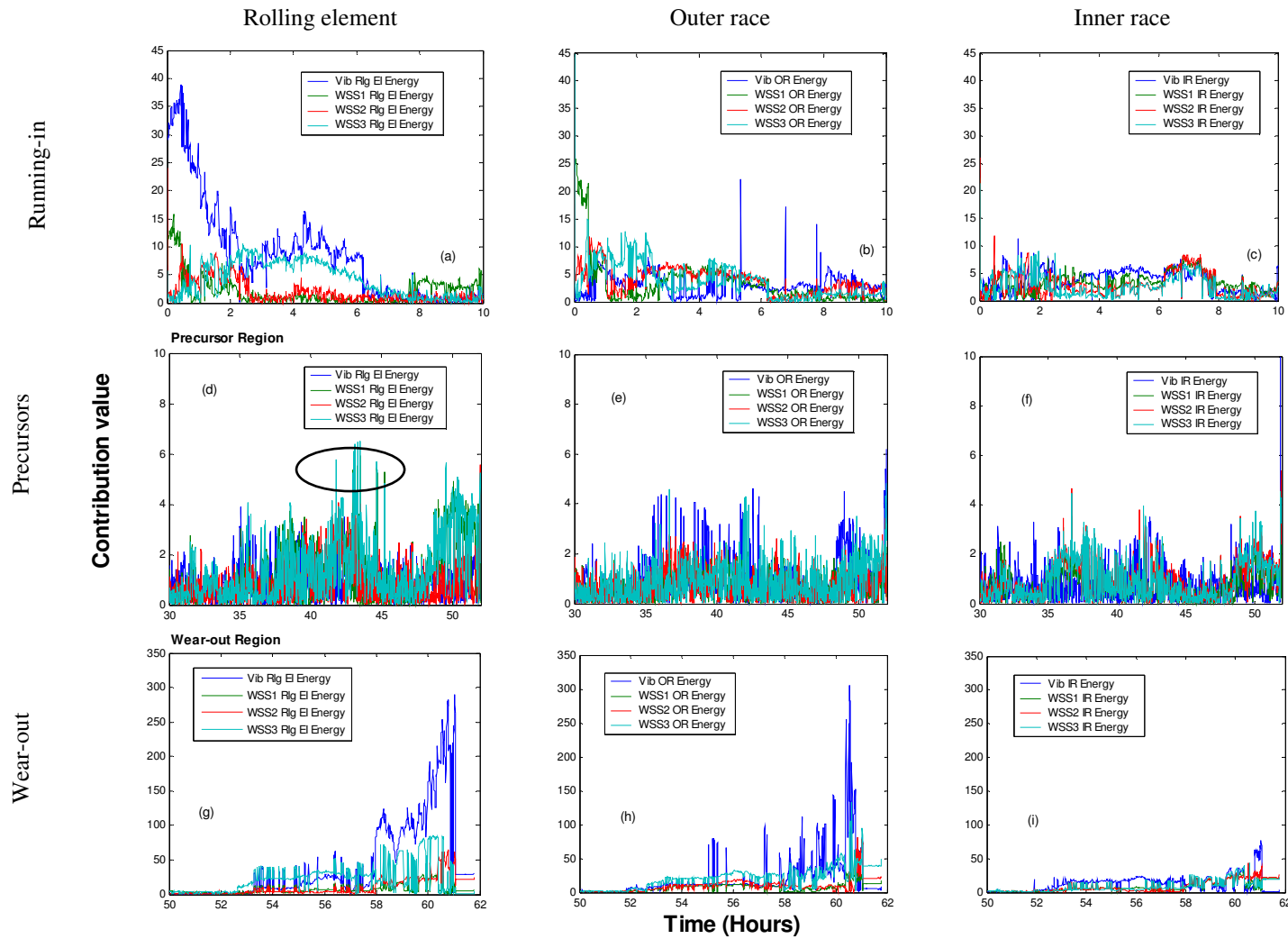


Figure 6.7 Contribution values of the bearing element energies at 3 periods of interest (test 20)

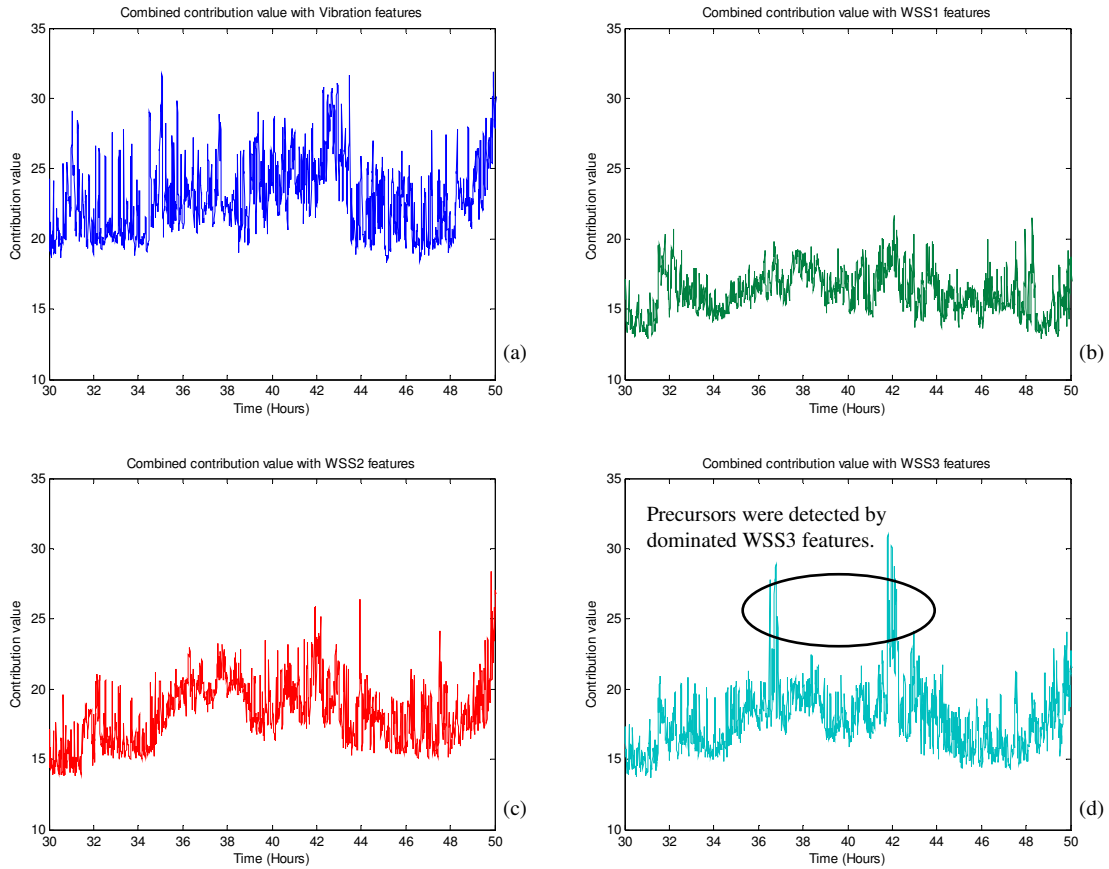


Figure 6.8 Combined contribution values of the different sensors in the precursor area

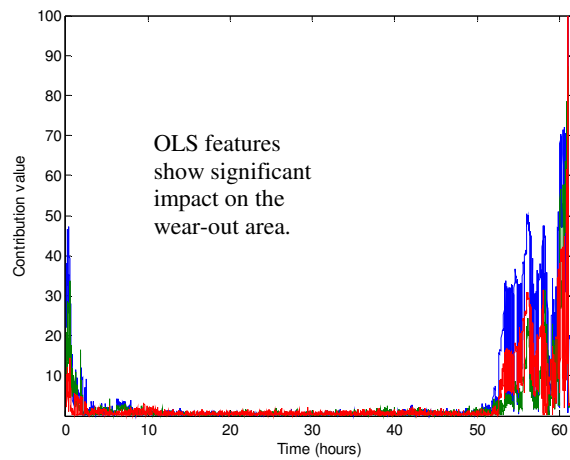


Figure 6.9 Contribution values of the OLS features (test 20)

6.3.6. Case 4 (almost normal at the end of the test, test 13)

Apart from the tests with significant failure conditions, in Chapter 5, test 13 that did not show significant defects on the parts of the bearings was also collected. The whole test lasted for around 48 hours, and the calculated fault detection index has been illustrated by Figure 5.20 in Chapter 5. From the calculated result, Hotelling's T-squared statistic shows that a relatively constant increasing trend occurred starting from 39 hours, and the values between 44 and 46 hours are above the 99% confidence limit which is established through the threshold set-up strategy introduced in Chapter 5. Although this increasing trend was insignificant, it is still valuable to find out which bearing or what part of the bearing was generating the defect.

Therefore, to diagnose the detected abnormal condition, the contribution values of each selected variables are calculated throughout the test. Figure 6.10 shows the calculated contribution values to the statistic of Hotelling's T-squared statistic at three fixed time intervals – running-in, steady state and wear-out areas. It is found that the calculated contribution values for each of the variables are close to each other in the fixed time interval. However in each of the divided time intervals, the variable with the highest contribution value is not consistent, which supports the assumption that there are no dominating features when the components are in a healthy condition. Finally, in the wear-out area, the rolling element and outer race energies of the sensors have low and closed contribution values, but the inner race energies of the vibration and WSS2 sensors have the highest impact on the insignificant trend at the end of the test. From the above analysis, the defects which are the cause of the increasing trend are suspected of occurring on the inner race of bearing 2. Moreover, the final bearing inspection in Chapter 5 showed that a small defect did occur to the inner race of bearing 2 accompanied with the increasing of the temperature.

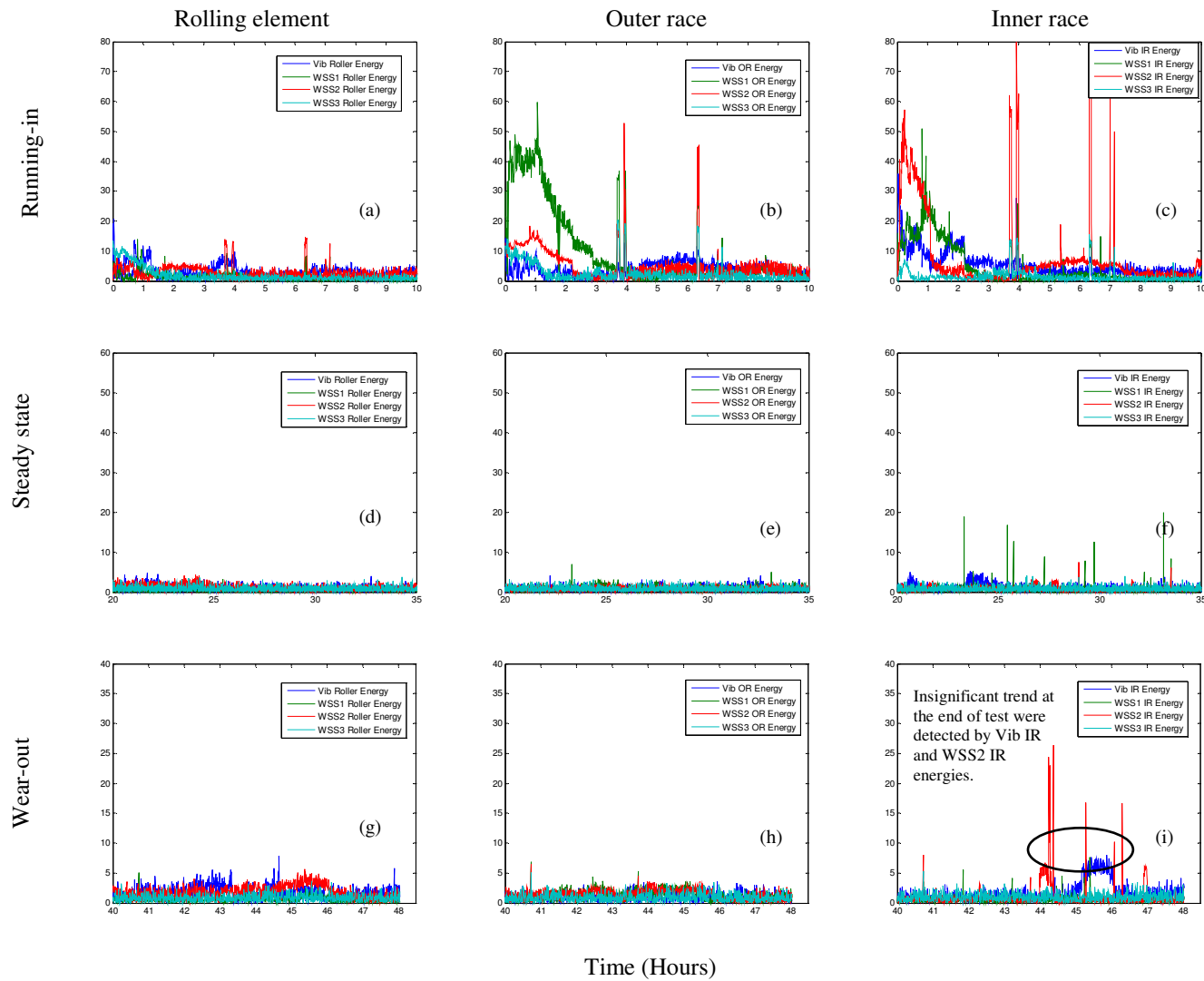


Figure 6.10 Contribution values of the bearing element energies at 3 periods of interest (test 13)

Figure 6.11 illustrates the contribution values of the OLS features. As expected, a very low impact is found at the end of running, which means no large-sized debris was generated, resulting in relatively unmodified contact surfaces.

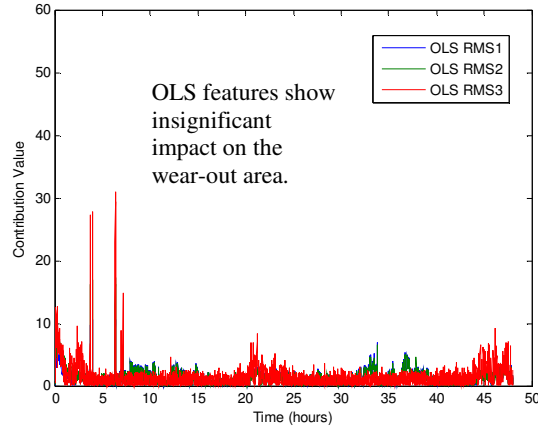


Figure 6.11 Contribution values of the OLS features

6.3.7. Case 5 (correlation to the AE amplitude, test 21)

At the later stage of the tests [8], the sensor of acoustic emission (AE) sensor was equipped to the experimental rig, and is supposed to detect some abnormal behaviour in terms of the changes of the sub-surface condition. In this section, AE amplitude is used as an assistant tool to the contribution values to see whether any additional diagnostic information can be obtained.

Figure 6.12 (a) shows the T-squared statistic calculated with 27 features and the collected AE amplitude respectively. It is seen from the figure that in region I, running-in behaviour was experienced for around 25 hours and within this area, multiple peaks were found between 20 and 25 hours. In contrast, AE amplitude plotted in the figure is also found to have additional activities in the running-in region, but seems to have no corresponding peaks to the multiple peaks in the T-squared statistic between 20 and 25 hours. After a constant steady state for both evaluated parameters, a single peak is found in AE amplitude at 48 hours, but related activities cannot be visualized in T-squared statistic. And then, consistent and additional activities which show an increasing trend till the end is found in AE amplitude, and this increasing trend is also found to have the

corresponded activities in T-squared statistic, such as 'C' shape trend between 56 and 60 hours and a constant increasing trend till the end of running.

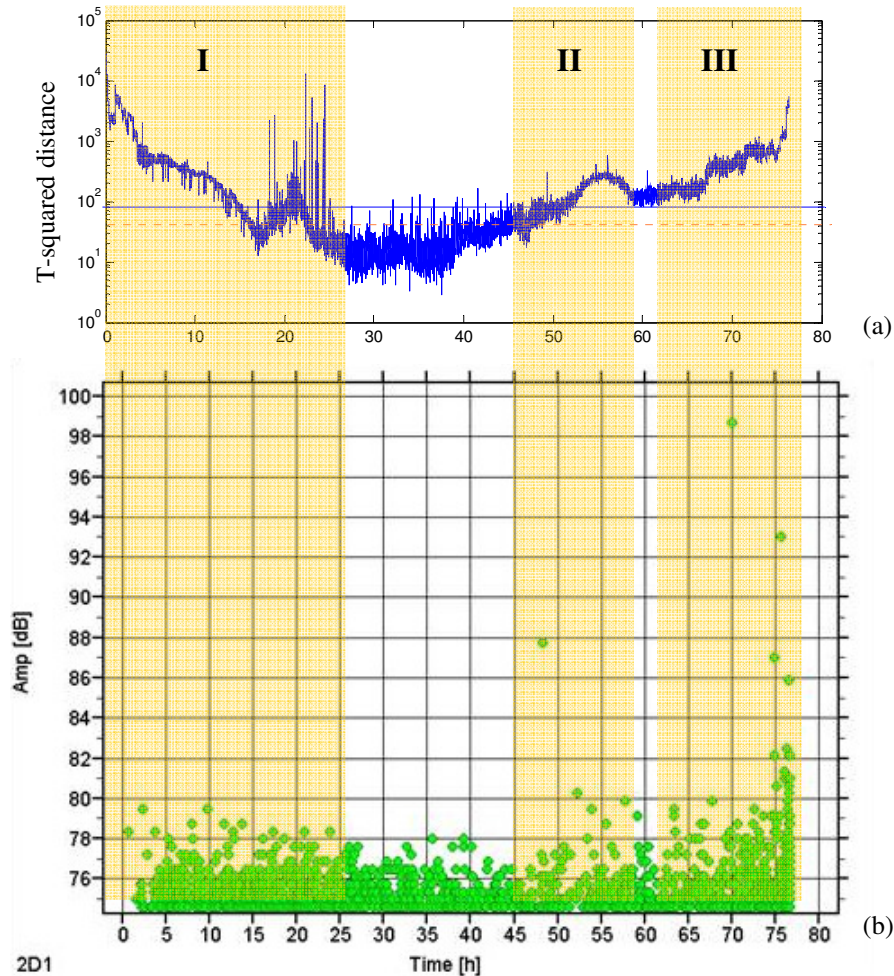


Figure 6.12 T-squared statistic of the test 21 data (a) and AE amplitude (b)

Obviously, three interesting areas are found with additional activities above the threshold and increasing trends. These activities are expected to be further analysed with the combination of AE amplitude and calculated contribution values. Figures 6.13 (a)-(c) illustrate the bearing element contribution values in the running-in area. It is found that for the multiple peaks detected between 20-25 hours, vibration inner race energy shows the dominant impact to the additional increasing-decreasing trend, but the multiple peaks seem to be contributed equally by other WSS features. Therefore, the reason for this

running-in activity is unknown; it might be because of the reflection of the inner race running-in condition, but there is no extra evidence to validate it.

At 48 hours, it is found that a single high peak appears in AE amplitude, but is invisible in the T-squared statistic. Figures 6.13 (d)-(f) show the contribution values in the period of precursors. It is clearly seen that WSS2 outer race energy (the red colour in Figure 6.13 (e)) has some impact factor at 48 hours; this new discovery can establish the correlation with the single high peak in AE amplitude at the same time. Furthermore, vibration inner race energy (the blue colour in Figure 6.13 (f)) shows the strongest effect on the 'C'-shaped trend between 56 and 60 hours. Although the bearing number of this activity cannot be confirmed because other impact factors of WSS features are close to each other, the pre-indent was on the inner race bearing 2, which enhances the possibility that the prognostic abnormal condition occurred to the inner race of bearing 2.

Figures 6.13 (g)-(i) show the contribution values of the features in the wear-out region. It is found that the WSS2 inner race energy (the red colour in Figure 6.13 (i)) shows the relatively dominant impact on the increasing trend at the end of running. This diagnostic information is also found to be consistent with the final inspection of the bearings which indicates that a small spallation occurred to the inner race of bearing 2. Furthermore, from the contribution values of the OLS features (see Figure 6.14), an increased level is also seen from around 60 hours, which supports that a number of debris was generated associated with an increase in acoustic emission and vibration.

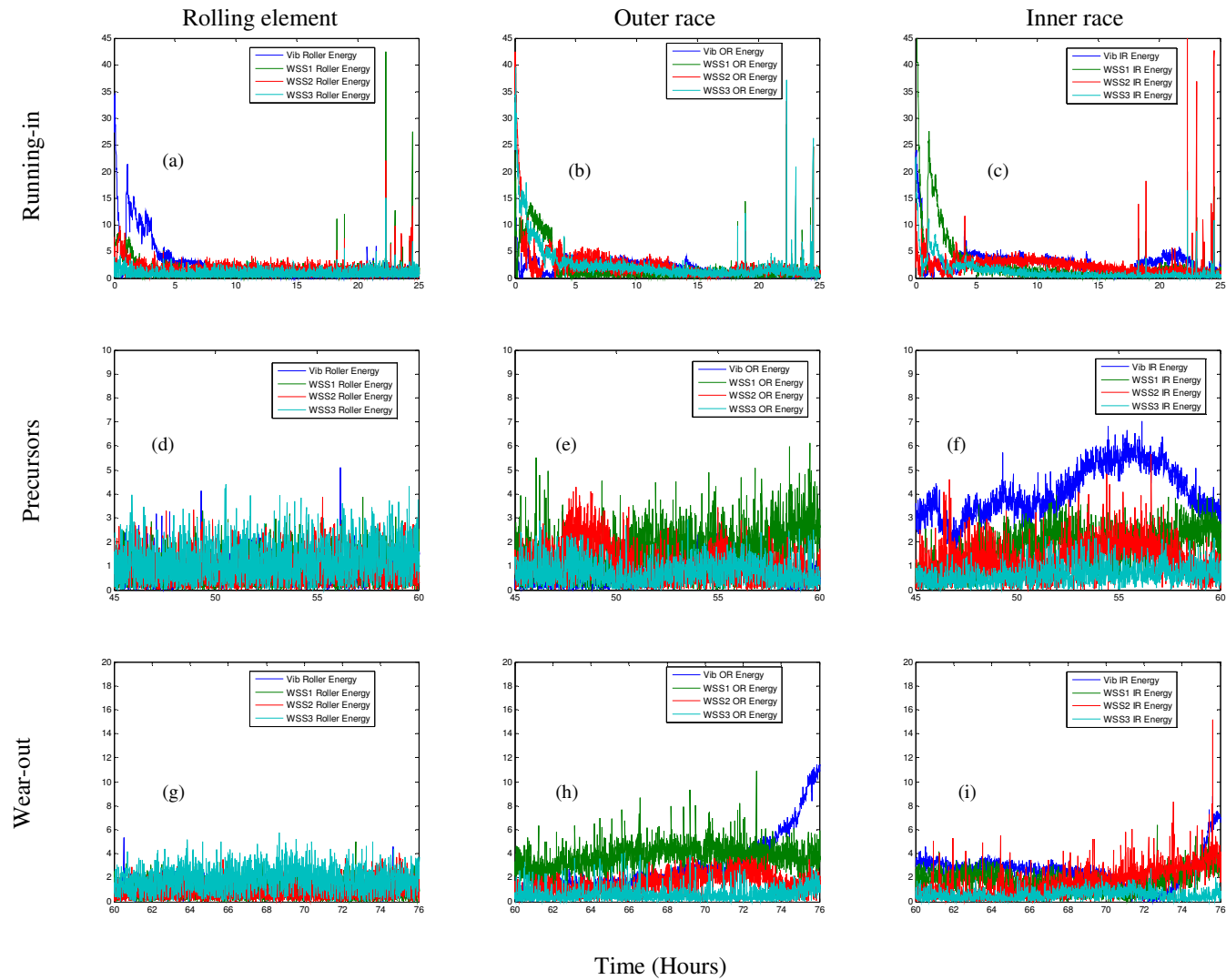


Figure 6.13 Contribution values of the bearing element energies at 3 periods of interest (test 21)

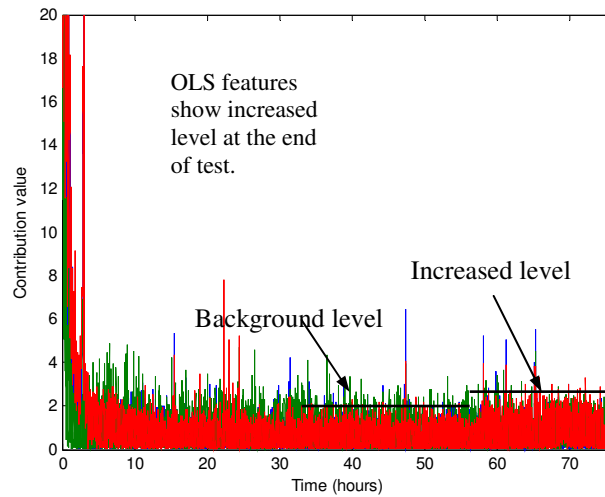


Figure 6.14 Contribution values of the OLS features (test 21)

In this case, AE amplitude was firstly utilized to correlate to the T-squared statistic and its contribution values. From the analysed results, it seems that additional AE activities can find related correlations in the T-squared statistic, apart from the single peak that solely appeared in AE amplitude at 48 hours (recovered in contribution values by WSS2 outer race energy), but come across difficulties when using contribution values to explain these detected anomalies, especially for the confirmation of faulty bearing numbers.

6.4. Summary

This study suggests that multivariate statistics based on principal component analysis can be used as a reasonable tool for fault detection in rotating machinery. Abnormal events, such as the occurrence of faults, can change the correlation structures of the measured variables. The nominal PCA model developed for normal operating conditions is no longer valid for the faulty conditions. Moreover, the contribution charts of the variables to the calculated statistics, such as T-squared statistic, is also applied to isolate the dominant variables, so that the causes for the faults could be found. In this work, the analysis of the contribution values to the abnormal T-squared statistic showed a consistent agreement with the bearing examined faults observed at the end of each test.

	Cases identified by both developed scheme and conventional plots	Cases that developed scheme identified but not seen by conventional plots
Test10	1. Activities in the running-in regions 2. Spallation activities starting from 28 hours	1. Detected precursors at 27 hours 2. Diagnostic information on detected precursors
Test13	1. Activities in the running-in region 2. Insignificant activities at the end of the test	1. Diagnostic information on the activities at the end of the test
Test14	1. Activities in the running-in region 2. Activities at the end of the test 3. Diagnostic information on the activities at the end of the test	None
Test15	1. Activities in the running-in region 2. Activities at the end of the test	1. Diagnostic information on the activities at the end of the test
Test16	1. Activities in the running-in regions 2. Precursors at 47.5 and 57.5 hours 3. Spallation activities starting from 60 hours 4. Diagnostic information about bearing spallation	1. Diagnostic information about the detected precursors
Test17	1. Activities in the running-in region 2. Precursors at 55 and 85 hours 3. Spallation activities starting from 95 hours	1. Diagnostic information on the detected precursors 2. Diagnostic information on bearing spallation
Test18	1. Activities in the running-in region 2. Precursors at 55 and 85 hours 3. Spallation activities starting from 95 hours	1. Diagnostic information on the detected precursors
Test20	1. Activities in the running-in regions 2. Spallation activities starting from 54 hours	1. Debris related prognostic activities detected at 43 hours 2. Diagnostic information about the detected precursors 3. Diagnostic information about bearing spallation
Test21	1. Activities in the running-in regions 2. Spallation activities starting from 50 hours 3. AE activity at 48 hours	1. Spallation occurred in inner race of bearing 2. AE activity occurred in bearing 2
Test22	1. Running-in activities 2. Spallation activities	1. Precursors at around 100 and 117 hours 2. Spallation occurred to the inner race of bearing 2

Table 6.3 Significant findings of the two cases relative to original plots

12 sets of experimental data have been tested against the developed model and 5 of them have been presented in the thesis. Significant new information has been successfully mined by calculating the Hotelling's T-squared statistic and features contribution values. For test 16 data, diagnostic information was extracted showing that precursors and spallation occurred to the inner race of bearing 2. These two sets of information were difficult to see in conventional plots. The diagnostic information obtained from the test

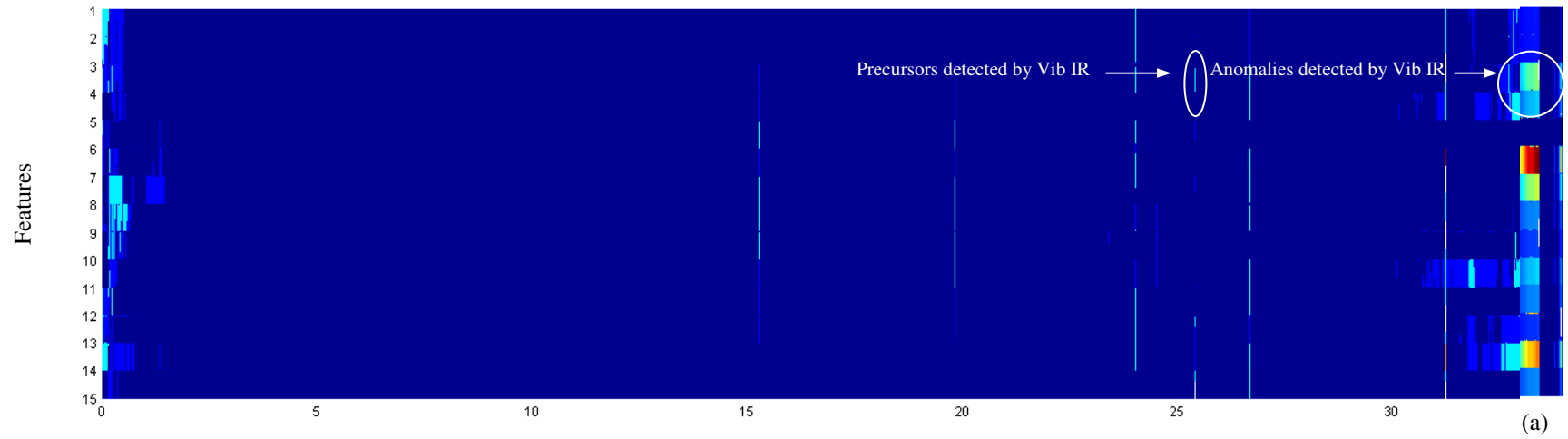
18 data clarifies that precursors between 170 and 200 hours are due to the fatigue failure initiation of the rolling element of bearing 2, not the indication of small spallation of bearing 2. For test 20 data, prognostic activities missed by conventional plots were observed approximately 10 hours before spallation; diagnostic information were also provided indicating that precursors and spallation occurred to the rolling elements and the outer race of bearing 4. Contribution values from test 13 data identify the reason for the small defect of the inner race of bearing 2. Other testing results have also shown that the application of anomaly detection and diagnosis methods can reveal more information. Table 6.3 summarizes these findings compared to the original data plots.

In order to clearly present contribution values and reveal dominant features throughout the tests, colour maps of the contribution values have been drawn as Figure 6.15 (a-j) shows. The magnitude of the colour map represents the contribution values; the y axis is the number of the evaluated features which have been indexed, as Table 6.4 shows. It should be noted that only 15 features were used to plot the colour map, because these features were found to reveal most of the information. Therefore, readers could identify which features are dominant at the specific periods of the tests. The general explanations of these maps can be related to the description of the contribution value in Table 5.3 of Chapter 5.

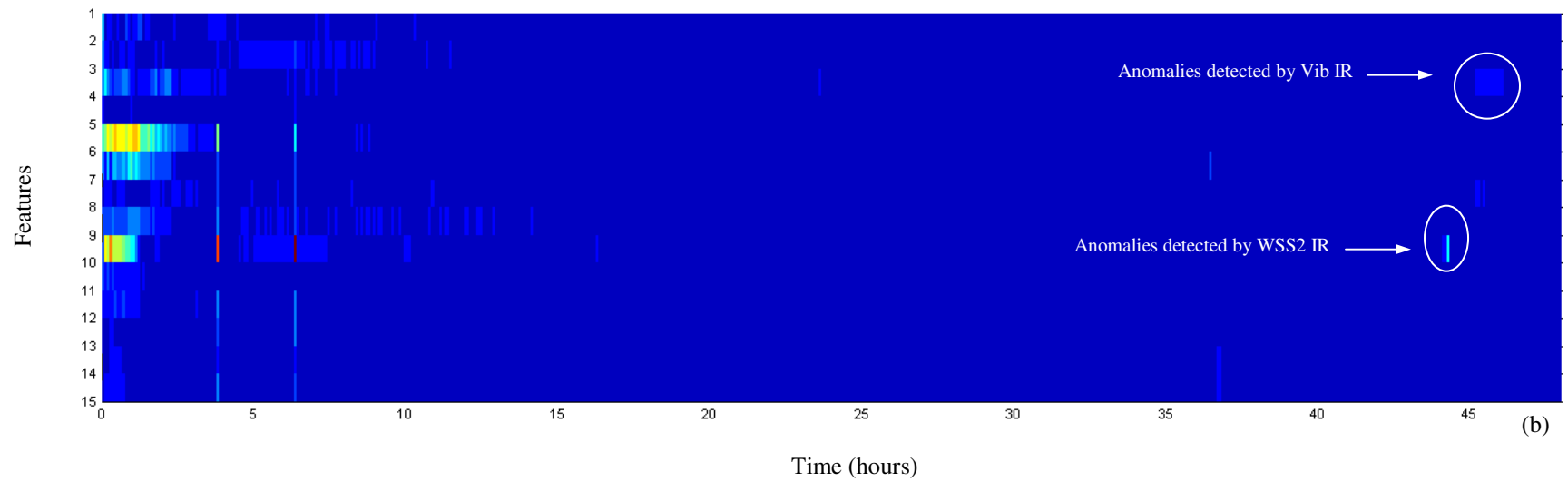
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Features	Vib. Rlg.	Vib. OR	Vib. IR	WSS1 Rlg.	WSS1 OR	WSS1 IR	WSS2 Rlg.	WSS2 OR	WSS2 IR	WSS3 Rlg.	WSS3 OR	WSS3 IR	OLS 1	OLS 2	OLS Ind.

Table 6.4 Feature number indication for the colour plot

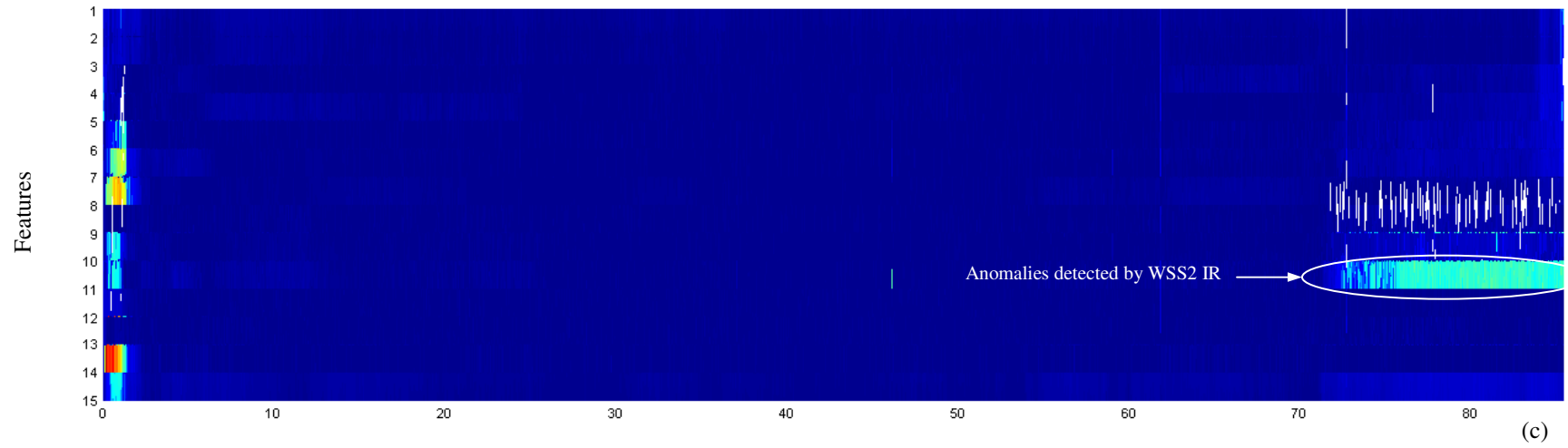
Test 10



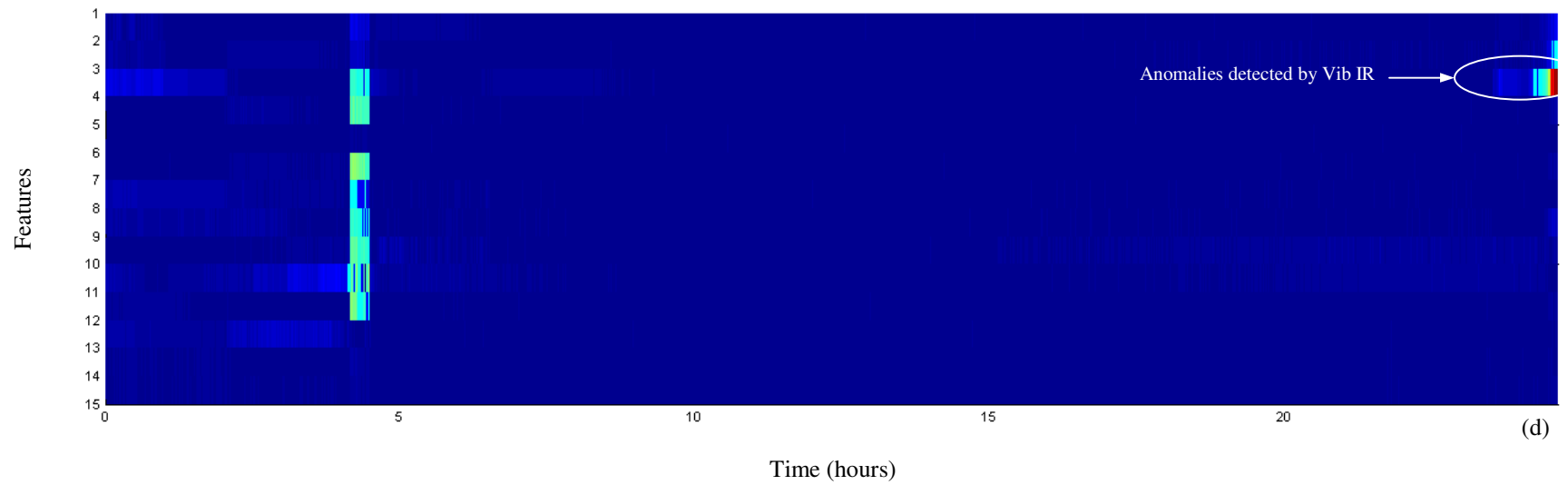
Test 13

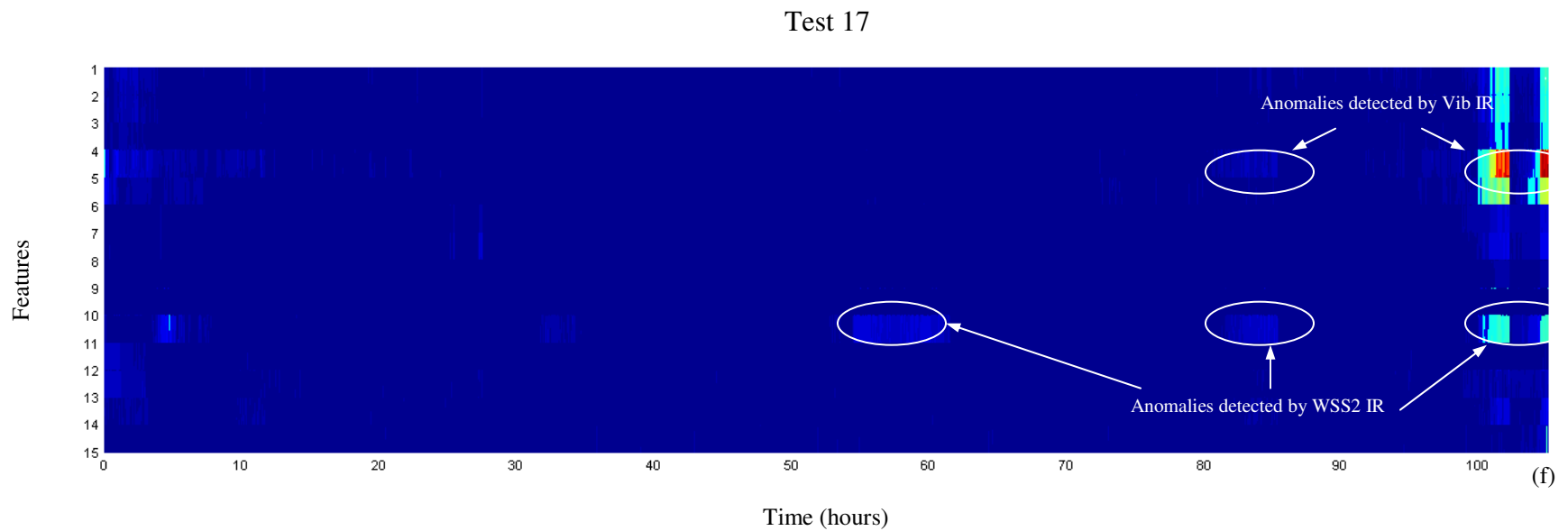
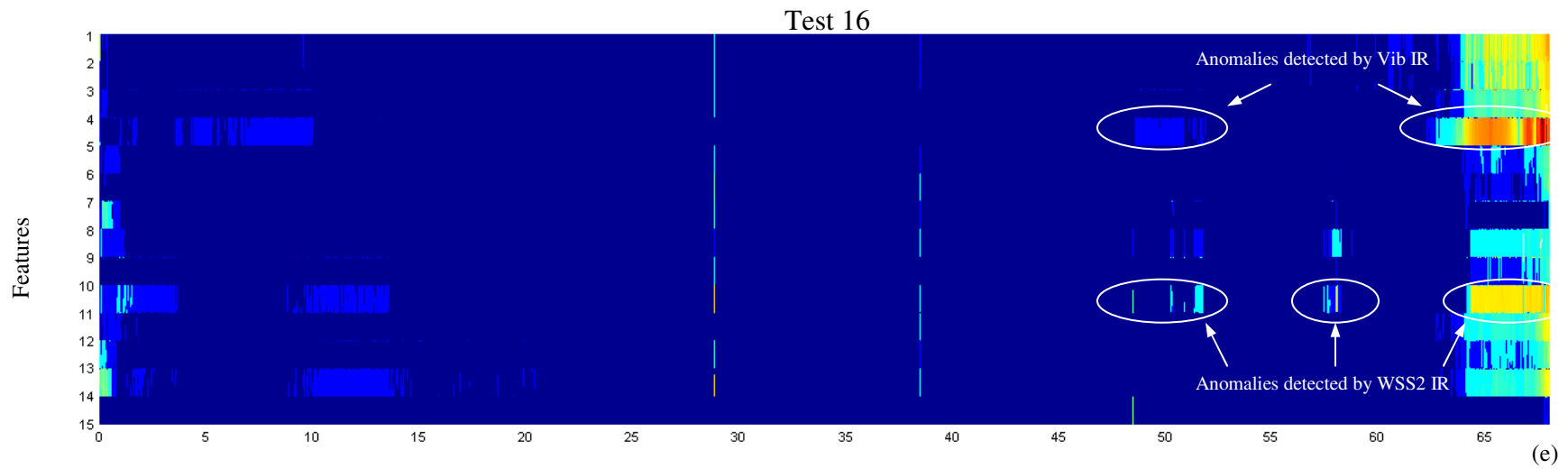


Test 14

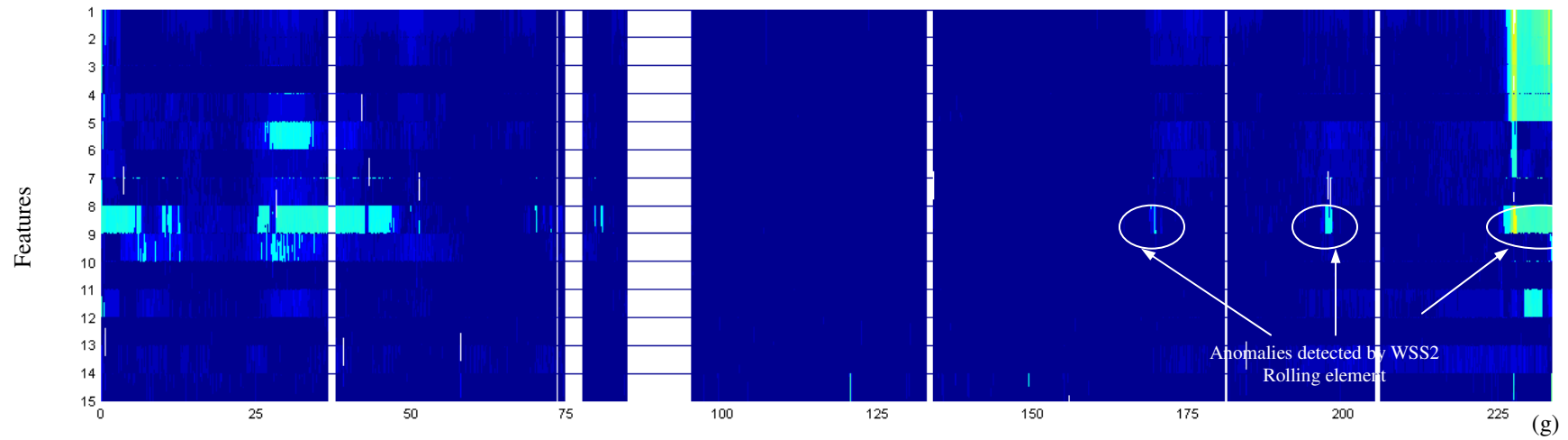


Test 15

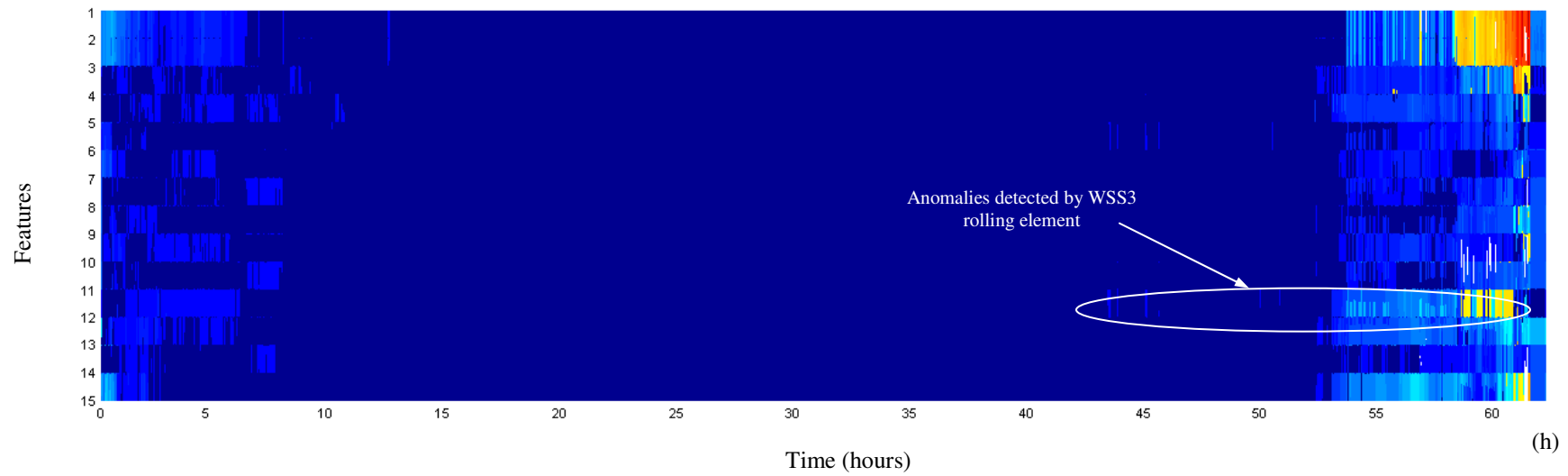




Test 18



Test 20



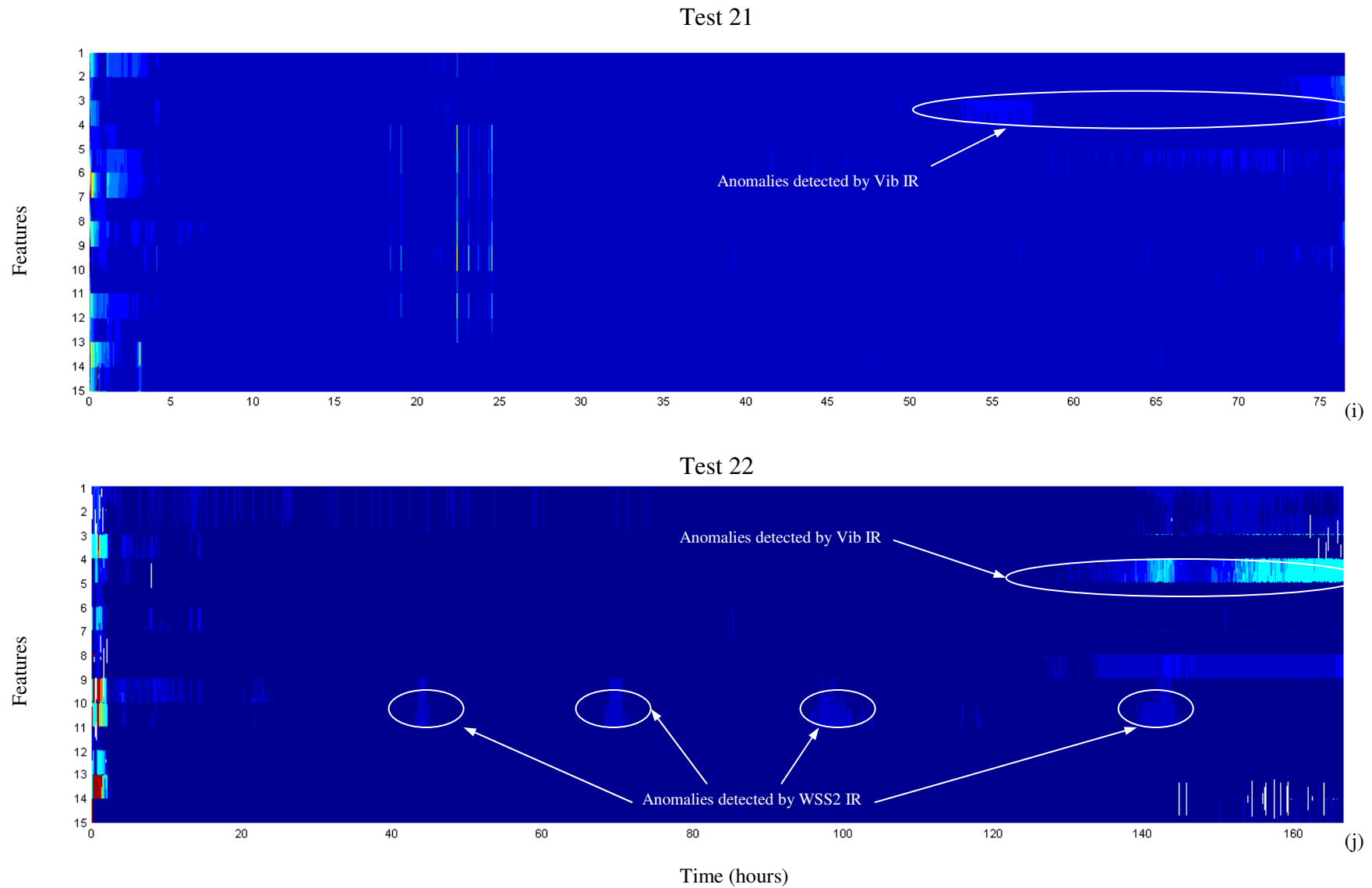


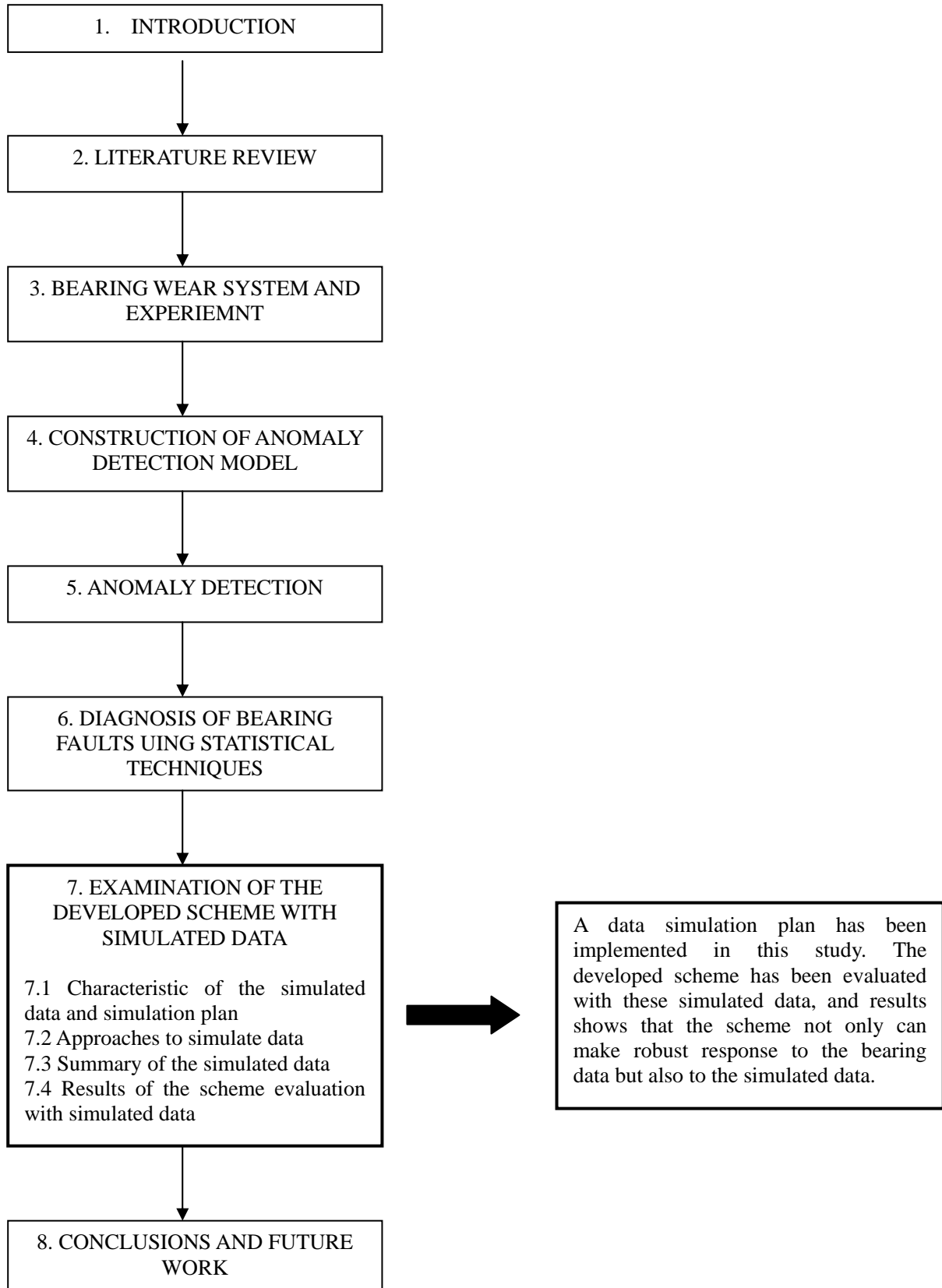
Figure 6.15 Colour maps of the contribution values

6.5 References

- [1] Y.M. Sebzalli, X.Z. Wang. Knowledge discovery from process operational data using PCA and fuzzy clustering. *Engineering Application of Artificial Intelligence*. 14 (2001) 607-616.
- [2] P. Miller, R.E. Swanson, C.F. Heckler. Contribution Plots: A Missing Link in Multivariate Quality Control, *Applied Mathematics and Computer Science* 8 (4) (1998) 775 – 792.
- [3] E.L. Russel, L.H. Chiang, R.D. Braatz, *Data-Driven Techniques for Fault Detection and Diagnosis in Chemical Processes*, Springer, London, (2000)
- [4] R. Treasure, U. Kruger, J.E. Cooper. Dynamic Multivariate Statistical Process Control using Subspace Identification, *Journal of Process Control*, 14 (3) (2004) 279 – 292.
- [5] S.J. Qin, S. Statistical process monitoring: basics and beyond. *J. Chemometrics*. 17 (2003) 480-502.
- [6] J.F. MacGregor, C. Jaeckle, C. Kiparissides, M. Koutoudi. Process monitoring and diagnosis by multiblock PLC methods. *AIChE J.* 40 (1994) 826-828.
- [7] S.J. Qin, S. Valle-Cervantes, M. Piovoso. On unifying multiblock analysis with applications to decentralised process monitoring. *J. Chemometrics*. 15 (2001) 715-742.
- [8] M. Craig, S.L. Chen, T.J. Harvey, R.J.K. Wood, K. Masuda, M. Kwabata, H.E.G. Powrie. Advanced condition monitoring of tapered roller bearings part I, with multiple sensing techniques. Accepted by *Tribology International*, 2008.

THESIS STRUCTURE

KEY OUTCOME



Chapter 7. Evaluation of the developed scheme with simulated data

From the previous chapters, although the developed anomaly detection and diagnosis system was demonstrated with the rolling element bearing tests, the number of available experimental data was still relatively low. Therefore, multivariate simulated data are generated using the Gaussian mixture approach to further evaluate the bearing fault detection and diagnosis scheme.

7.1. Characteristics of the simulated data and simulation plan

To simulate multivariate data, the characteristics of the bearing data need to be reviewed and extracted as the reference. Figure 7.1 shows two typical features of the experimental bearing data. In these two features, typical characteristic trends of the bearing data can be seen, such as a discrete trend which usually occurs in running-in and precursor regions; a flat normal trend in the steady state period; an increasing trend at the end of the test representing significant abnormal conditions; and an oscillation trend with noise which are also usually seen throughout the tests. Therefore, these observed characteristics are expected to be used as examples for the proposed simulated data. It should be noted that the reason for including noise in the simulated data is to validate that the developed approaches are robust to noise and able to identify them in the training data.

Before conducting simulation, a data simulation strategy needs to be confirmed. Since the developed scheme involves anomaly detection model construction/adaptation, anomaly detection and fault diagnosis, the number of the simulated data, the usage of the data and the characteristics of the data are determined by these three main tasks. To build the anomaly detection model, baseline data representing the healthy condition need to be prepared, and a series of abnormal features such as discrete anomalies, abnormal trends and noise are added to the flat baseline level to demonstrate the capability of the model adaptation. To conduct anomaly detection and fault diagnosis, run-to-failure tests need to be simulated, and this includes usually encountered anomalies such as step change,

significant increasing trend, discrete trend, etc. Within these simulation data, seven baseline tests and four run-to-failure tests are expected to be prepared:

- **Baseline test**
 - To simulate 7 sets of baseline data.
 - To build the model of normality.
 - To generate various abnormal features.
 - To examine the capability of the developed approach to remove anomalies within the baseline (training) data, and justify the procedure of building the stable and reliable model.

- **Run-to-failure test**
 - To simulate 4 sets of run-to-failure data.
 - To test against the normal model.
 - To generate various abnormal features.
 - To examine the capability of the developed approach to detect and diagnose the anomalies in the run-to-failure (testing) data.
 - To compare the anomaly detection results before and after the removal of anomalies within the baseline (training) data.

Table 7.1 summarizes the proposed simulated data with specific characteristics which are appearing in the typical experimental bearing data [1, 2], such as noise with the form of spikes, abnormal trend, oscillation trend etc. To prove the capability of the model adaptation, these abnormal features are implemented in a more complicated way. Apart from the single abnormal feature in Baseline 1, 2 and 3, other Baseline data are all simulated with multiple abnormal features: i.e. multiple discrete trend in Baseline 4; abnormal trend plus noise in Baseline 5; and oscillation trend as well as noise in Baseline 6. The purpose of mixing these abnormal trends in one baseline test is to see whether the developed model adaptation method is robust enough to handle more complicated situations, such as the situation that anomalies appear throughout the test.

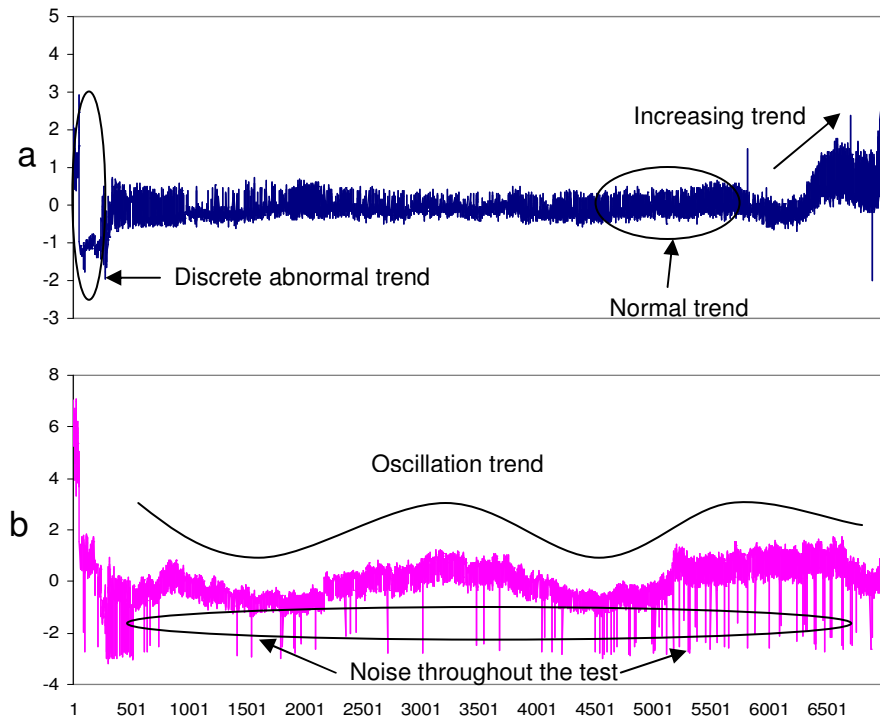


Figure 7.1 Typical experimental bearing data with characteristic features

Moreover, the run-to-failure data which are used to test against the model to extract fault detection and diagnostic information consists of four datasets. In each test, single or double abnormal features with different levels are added to see whether the model can respond to either the significant trend or small anomalies. In the following section, the approach to simulate related abnormal characteristics is introduced.

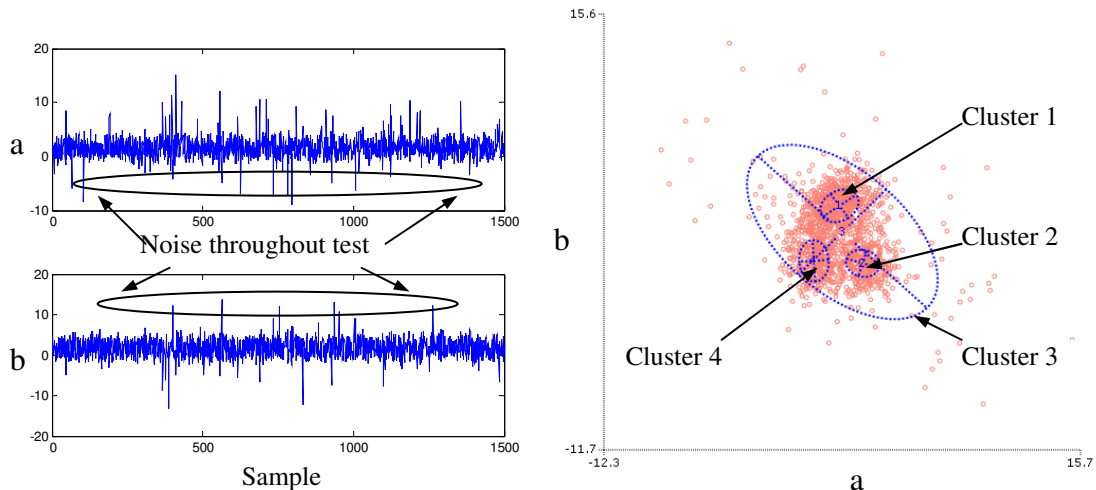
Simulating dataset	Dimensions	Characteristic features	Example
Baseline 1	27	Noise with the form of spikes throughout the test (for single or multiple variables)	
Baseline 2	27	Single abnormal trend	
Baseline 3	27	Discrete abnormal trend	
Baseline 4	27	Multiple abnormal trend	
Baseline 5	27	Abnormal trend + noise	
Baseline 6	27	Oscillation + noise	
Baseline 7	27	Oscillation + abnormal trend + noise	
Run-to-failure 1	27	Step changes	
Run-to-failure 2	27	Discrete trend + step changes	
Run-to-failure 3	27	Discrete trend + increasing trend	
Run-to-failure 4	27	Discrete trend + step change + increasing trend	

Table 7.1 Data simulation plan and the characteristic features

7.2. Approaches to simulate multivariate data with a mixture of Gaussians

A multivariate Gaussian (or normal) distribution is an extension of a univariate Gaussian. In a single dimension, a normal distribution has the familiar bell-shaped curve. In two dimensions, each feature is itself a normal distribution. If the two dimensions are independent then they tend to cluster as a circular cloud of points if they are correlated, and then they form an ellipse. This can be extended to any number of multiple dimensions. If two or more independent sources are generating data and are then pooled together, a multivariate mixture can be obtained. Often the data are Gaussian or are nearly Gaussian and can be modelled by a mixture of Gaussians.

It is well known that the means and covariance structure are the main parameters to describe the Gaussian components. What if we wanted to estimate the parameters instead? Assume for now that we know the number of data sources. With the data that has clear distinct clusters, the model fit can be plotted using 1 standard deviation ellipse. But the example is more challenging if the clusters overlap each other. In fact, in this chapter, the clusters have been drawn without colouring by source to emphasize this fact. The ellipses are again the model fit and the red '+' markers indicate the true centres.

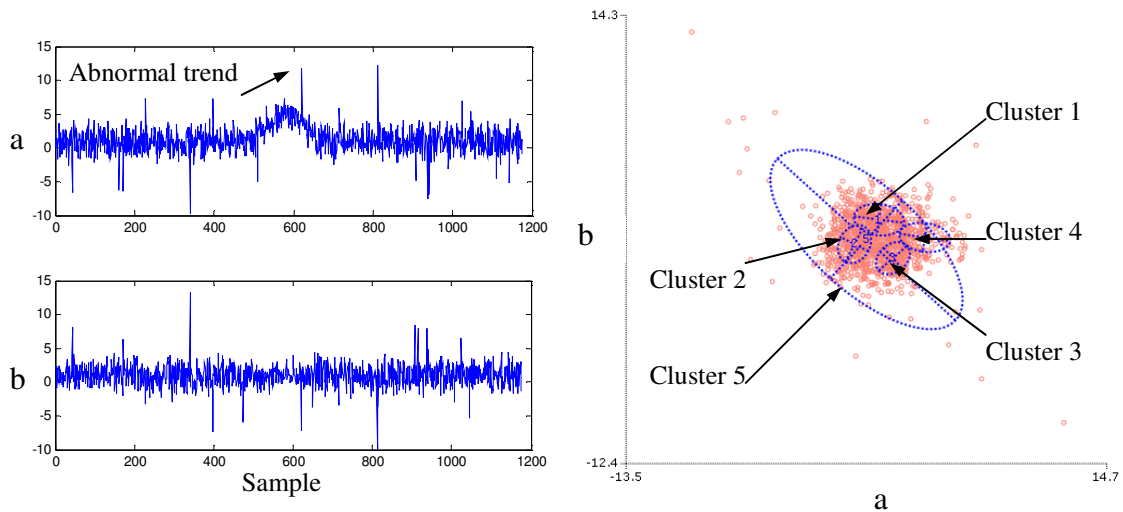


Probability		Mean		Covariance	
w1	0.375544	m1	1.534858 3.591193	V1	1.256424 0.285107
w2	0.227285				0.285107 1.05224
w3	0.21884	m2	2.899584 -0.021647	V2	0.983157 -0.089108
w4	0.178332				-0.089108 0.723122
		m3	1.706034 1.946734	V3	31.20429 -16.72109
					-16.72109 29.96687
		m4	0.041119 0.135753	V4	0.754669 -0.062199
					-0.062199 1.702356

Figure 7.2 Approach to simulate noise throughout the dataset

a) Generating noisy data: method of simulating noise throughout the dataset, as the Figure 7.2 shows. The general procedure is: 1) The normal patterns are generated by a mixture of Gaussians (clusters 1, 2 and 4) with small variance and high probability density; 2) The noise (outliers) are randomly generated according to a Gaussian distribution (cluster 3) with large variance and low probability density.

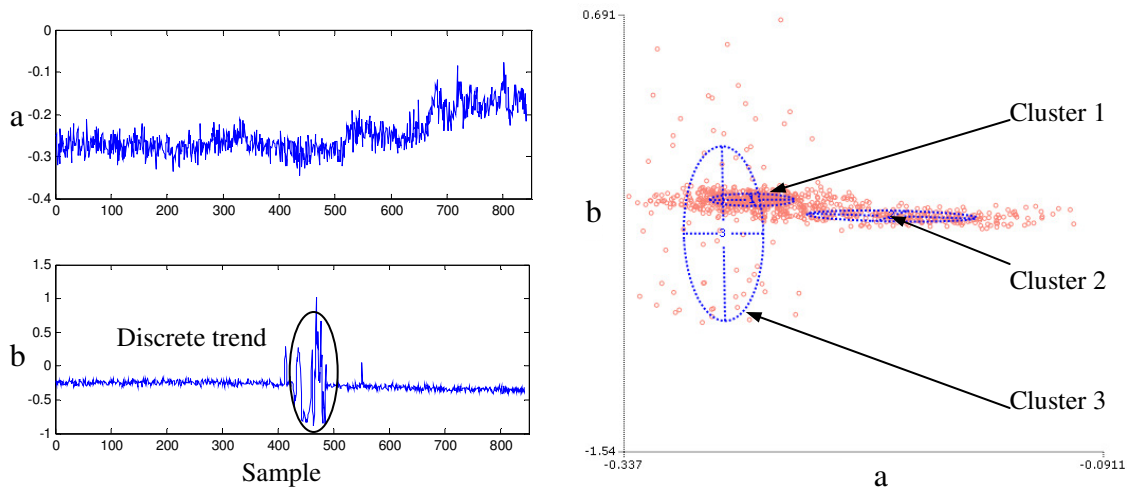
b) Generating an abnormal trend in a data set: method of simulating an abnormal trend, as the Figure 7.3 shows. The general procedure is: 1) The normal patterns are generated by a mixture of Gaussians (clusters 1~3) with small variance and high probability; 2) The abnormal trend is added by steadily changing the mean of the Gaussians in the short sampling periods, and the generated trend forms cluster 4. It is note that, the trend could not be directly generated by cluster 4; it has to be generated by steadily changing the means and variance in the time series. 3) The noise (outliers) are randomly generated according to a Gaussian distribution (cluster 5) with large variance and low probability density.



Probability		Mean		Covariance	
w1	0.33581	m1	1.33601 2.112464	V1	2.171521 -0.171643
w2	0.307569	m2	-0.107725 0.649884		-0.171643 0.891544
w3	0.244067	m3	2.153288 -0.183959	V2	0.956281 0.237258
w4	0.085091	m4	4.031542 1.053826		0.237258 1.293784
w5	0.027462	m5	0.596573 0.927701	V3	0.985821 0.274709
					0.274709 0.918612
				V4	2.249902 -0.220694
					-0.220694 0.763534
				V5	32.01176 -20.30512
					-20.30512 28.52138

Figure 7.3 Approach to simulate abnormal trend

c) Generating a discrete trend in a dataset: method of simulating discrete trend, as the Figure 7.4 shows. The general procedure is: 1) The normal patterns are generated by a mixture of Gaussians (clusters 1, 2) with small variance and high probability; 2) the discrete trend are generated according to a Gaussian distribution (cluster 3) with large variance. Please note: the difference between this and noise generation is that the discrete trend is generated in a particular sampling period, not throughout the test.



Probability		Mean		Covariance	
w1	0.553899	m1	-0.271387 -0.253149	V1	0.000469 -2.61E-05
w2	0.327838	m2	-0.199762 -0.33594		-2.61E-05 0.000994
w3	0.118263	m3	-0.286072 -0.4257	V2	0.001888 -0.000356
					-0.000356 0.000811
				V3	0.000417 -0.00036
					-0.00036 0.199692

Figure 7.4 Approach to simulate discrete trend

d) Generating an increasing trend in a ‘noisy’ dataset: method of simulating increasing trend and noise as the Figure 7.5 shows. The general procedure is: 1) The normal patterns are generated by a mixture of Gaussians with small variance and high probability (cluster 1); 2) The step changes could be added by increasing the mean of the Gaussians (cluster 2). To make the step change more fluctuated, the variance of the Gaussians could also be changed; 3) The noise is simulated by cluster 3.

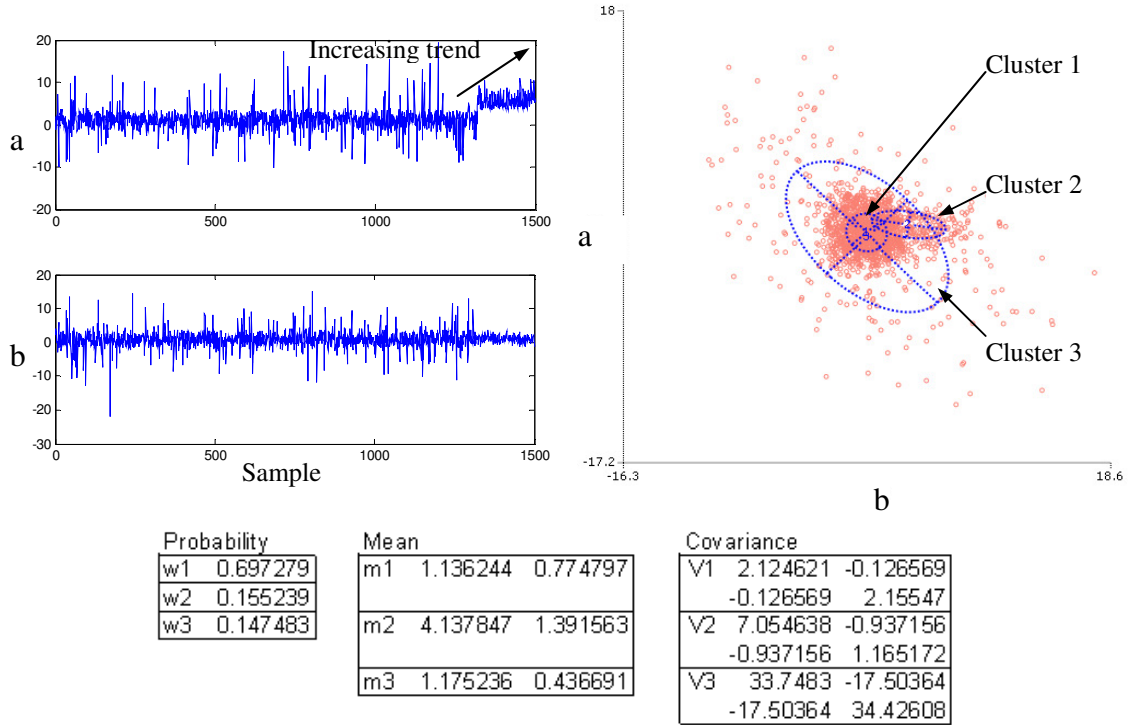


Figure 7.5 Approach to simulate step changes

7.3. Summary of the simulated data

Having confirmed the simulation plan and approaches to the desired abnormal features, 12 simulation datasets have been generated as the Table 7.2 summarises. For the baseline data, particular abnormal features were implemented to the first two features, and were expected to be detected with the model adaptation methods; and the rest of the 25 features are with the entire flat healthy trend. For the run-to-failure tests, abnormal trends were added with different aims. For example, in RTF2, different anomalies were generated in different features, and the approach is expected to locate each feature to identify their associated anomalies. Details of the aims and methods for each test can be found in Table 7.2.

Baseline data			
Simulating dataset	Dimensions	Method	Example (first two features)
Baseline 1	27	<p>1) The normal patterns are generated by a mixture of Gaussians with small variance and high Probability density (clusters 1-4)</p> <p>2) The noise (outliers) are randomly generated according to a Gaussian distribution with large variance and low probability density (cluster 5).</p>	
Baseline 2	27	<p>1) The normal patterns are generated by a mixture of Gaussians with small variance and high Probability</p> <p>2) The abnormal trend is added by steadily changing the mean of the Gaussians in the short sampling periods, and the generated trend forms cluster 4.</p>	
Baseline 3	27	<p>1) The normal patterns are generated by a mixture of Gaussians (clusters 1, 2 and 3) with small variance and high Probability</p> <p>2) The discrete trends are generated according to a Gaussian distribution (cluster 4) with large variance. Its difference from the noise generation is that the discrete trend is generated in a particular sampling period.</p>	

<p>Baseline 4</p>	<p>27</p>	<p>1) The normal patterns are generated by a mixture of Gaussians (clusters 1-8) with small variance and high Probability;</p> <p>2) The discrete trends are generated according to a Gaussian Distribution (cluster 9-10) with large variance.</p>	
<p>Baseline 5</p>	<p>27</p>	<p>1) The abnormal trend is added by steadily changing the mean of the Gaussians in the short sampling periods, and the generated trend forms cluster 5 and 6.</p> <p>2) The noise (outliers) are randomly generated according to a Gaussian distribution with large variance and low probability density.</p>	
<p>Baseline 6</p>	<p>27</p>	<p>The oscillation trend is generated by changing Gaussian means and variance throughout the test (the method is similar to the increasing trend simulation) but limiting the means of each simulated period within a certain level.</p>	

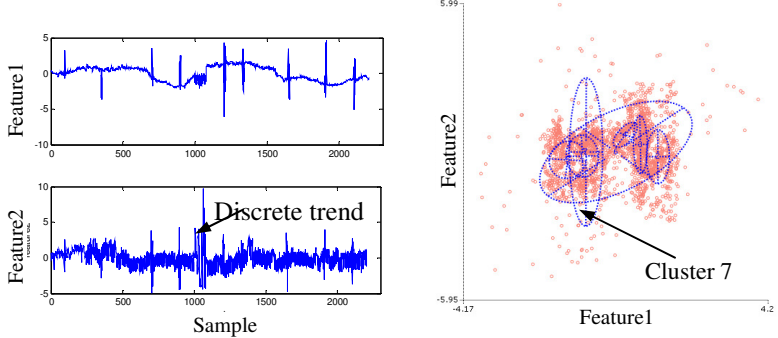
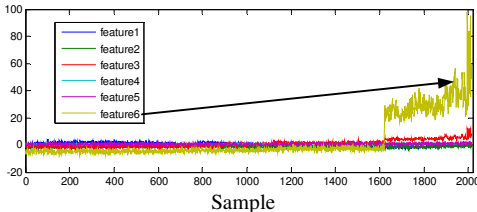
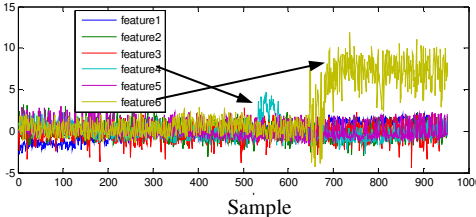
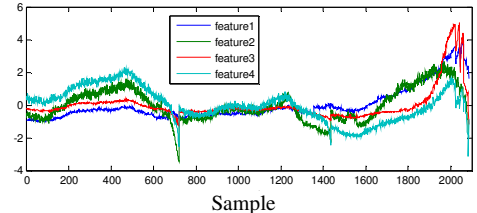
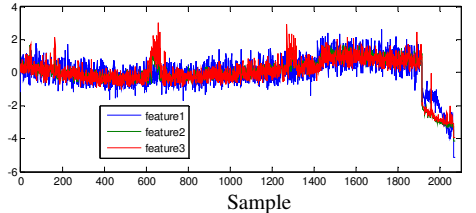
<p>Baseline 7</p>	<p>27</p>	<p>Based on Baseline 6, the abnormal discrete trend is generated by cluster 7.</p>	
<p>Run-to-failure (RTF) data</p>			
<p>RTF1</p>		<p>Features are simulated with different levels of abnormality; this is achieved by changes of means and variance of every feature.</p> <p>The aim of this simulation is to prove the diagnostic tool (contribution factor) could identify different abnormal levels and identify the dominant factor-feature 6.</p>	 <p>Different features are added into the different variables.</p> <p>The aim of this is to use the contribution factor to locate variables with different abnormal features, to indicate the small anomaly and step changes that occurred in feature 4 and 6 respectively.</p>
<p>RTF3</p>		<p>Oscillation trend is simulated throughout the test. Discrete trend and increasing trend with different levels are added into different variables, and the diagnostic tool is used to identify different abnormal features in different variables.</p>	 <p>Three encountered abnormal features are simulated, and different features are simulated with the similar levels in different variables. It is expected that the contribution factors of these variables will be close to each other.</p>

Table 7.2 Summary of the simulation data

7.4. Results of the scheme evaluation with simulated data

In this section, the developed integrated scheme will be examined with the simulated data following such a procedure: 1) Construction of an anomaly detection model with 7 baseline data; 2) Adaptation of the constructed anomaly detection model; 3) Anomaly detection with 4 RTF data; 4) Fault diagnosis with 4 RTF data.

(1) Construction of a stable and consistent anomaly detection model

It was introduced in Section 7.3 that 7 sets of baseline data in total have been simulated, and the anomaly detection model is built with these 7 sets of data. Following the procedure of finding a consistent and best modelling solution, which was introduced in Section 4.4; the optimized model is constructed as shown in Figure 7.6. It is seen that 20 clusters have been trained. Then colours represent that the cases are from the 7 tests. It is found that a couple of clusters such as clusters 14 and 19 are mainly supported by purple colour points which are from one single test. Thus, these two clusters are suspected to be associated with anomalies, because they are occurring occasionally in just one single dataset. For other clusters, the support cases are with multiple colours suggesting they are from multiple tests; this indicates the cases in these clusters are from the normal condition, as they frequently occur in the different datasets.

To confirm whether these two clusters are supported by the anomalies and to conduct the appropriate model adaptation, the support cases in these clusters will be examined with the developed approaches introduced in Chapter 4, to examine whether they occur occasionally in the time series or are widely distributed in the magnitude direction. Only these two requirements are met, the support cases can only be identified as the anomalies.

In the following, 7 simulated baseline datasets are combined together in a sequent order to form a multivariate time series, and the multivariate segmentation approaches is conducted to separate any abnormal features appeared in the time series from normal ones.

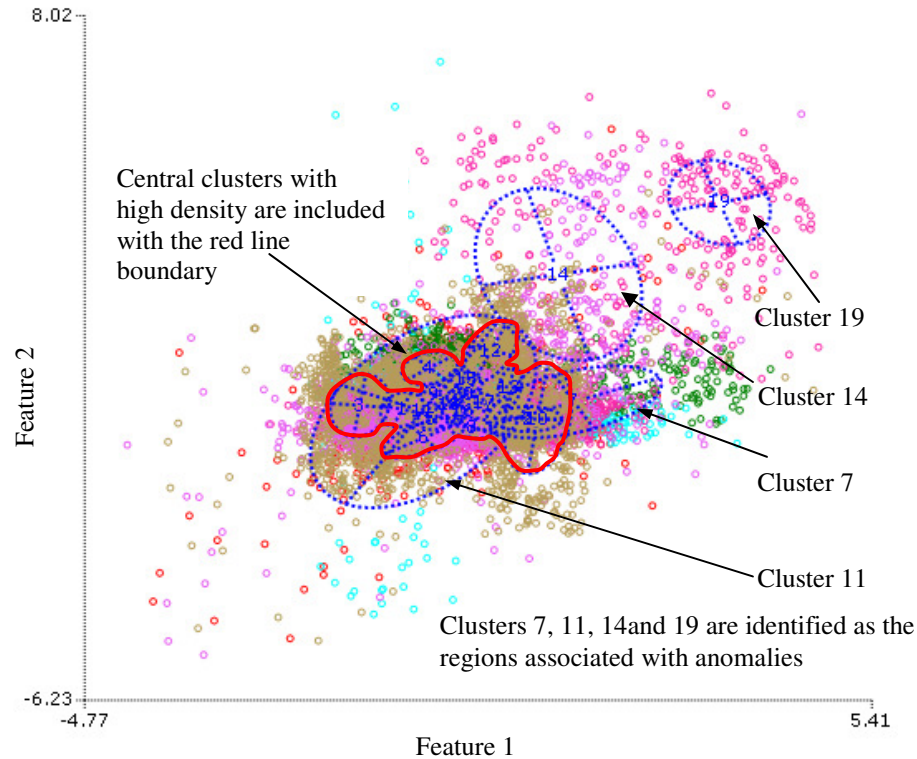


Figure 7.6 The anomaly detection model built by baseline data using GMM (with 20 clusters)

(2) Identification of the anomaly related clusters and adaptation of anomaly detection model

To adapt the anomaly detection model, the approaches of multivariate time-series segmentation (introduced in Section 4.5.3.1) and magnitude density inspection (introduced in Section 4.5.3.2) are adopted, and the cluster entropy scores are calculated based on these automatic segmentations.

Figure 7.7 shows the multivariate segmentation result and it is found that 9 conditional clusters are obtained and this result was assigned to the combined baseline data shown with two features for convenience (27 features in total) as Figure 7.7 illustrates. With this categorization result, the constructed model shown in Figure 7.6 can be analyzed to examine which one of the 9 categories the support cases of each cluster (Figure 7.6) is from. Then the entropy score for each cluster can be calculated based on the new categorization result in the time series.

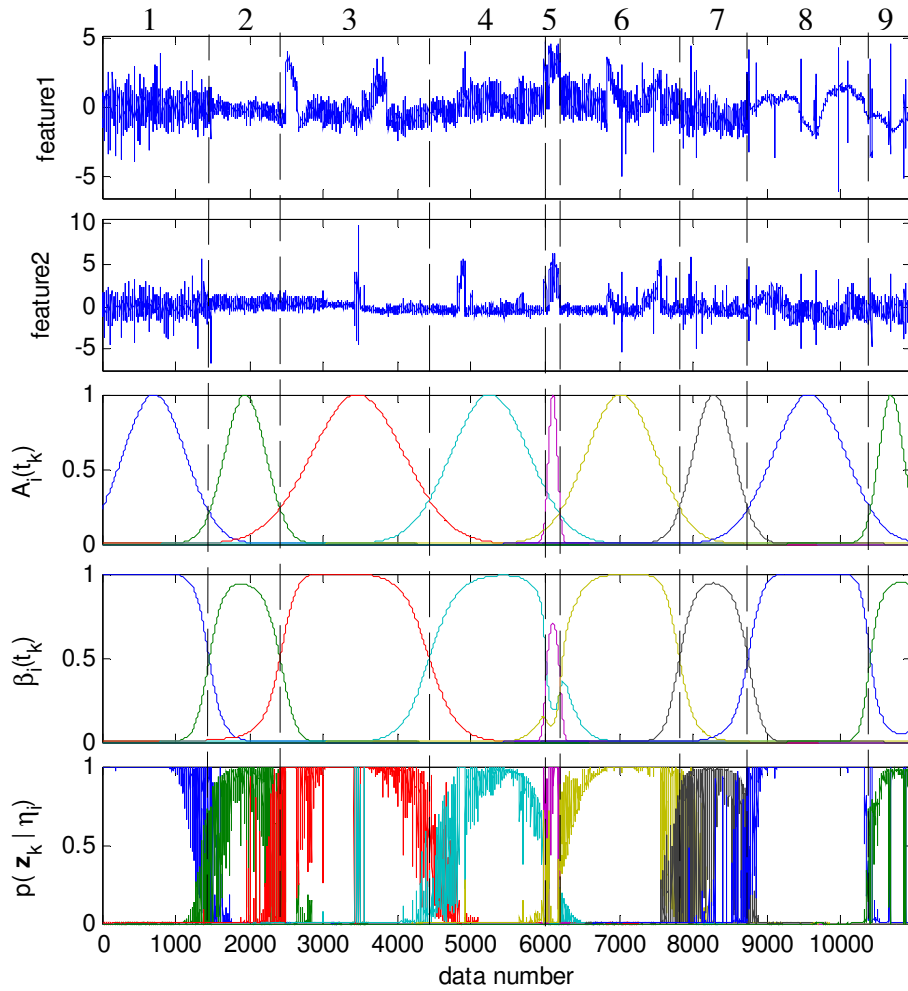


Figure 7.7 Multivariate segmentation for the baseline data

Figure 7.8 (a) shows the entropy score of clusters and it is found that the entropy scores of clusters 14 and 19 are seen significantly lower than the other scores of clusters, because of the unified support cases. Thus, these two clusters are highly suspected to be the regions associated with occasional occurring anomalies, and it is suggested that these should be removed from the original normal space. Figures 7.8 (b) and 7.8 (c) illustrate the adapted model and the first two features of the dataset respectively after the removal of clusters 14 and 19 and their associated support cases. It is clearly seen that the several abnormal trends in the second features have been successfully filtered out. However, the abnormal trend in the first feature and continuous noise in both features are still in the dataset. And these abnormal characteristics are expected to affect the final anomaly detection, and it is recommended that they are removed from the dataset.

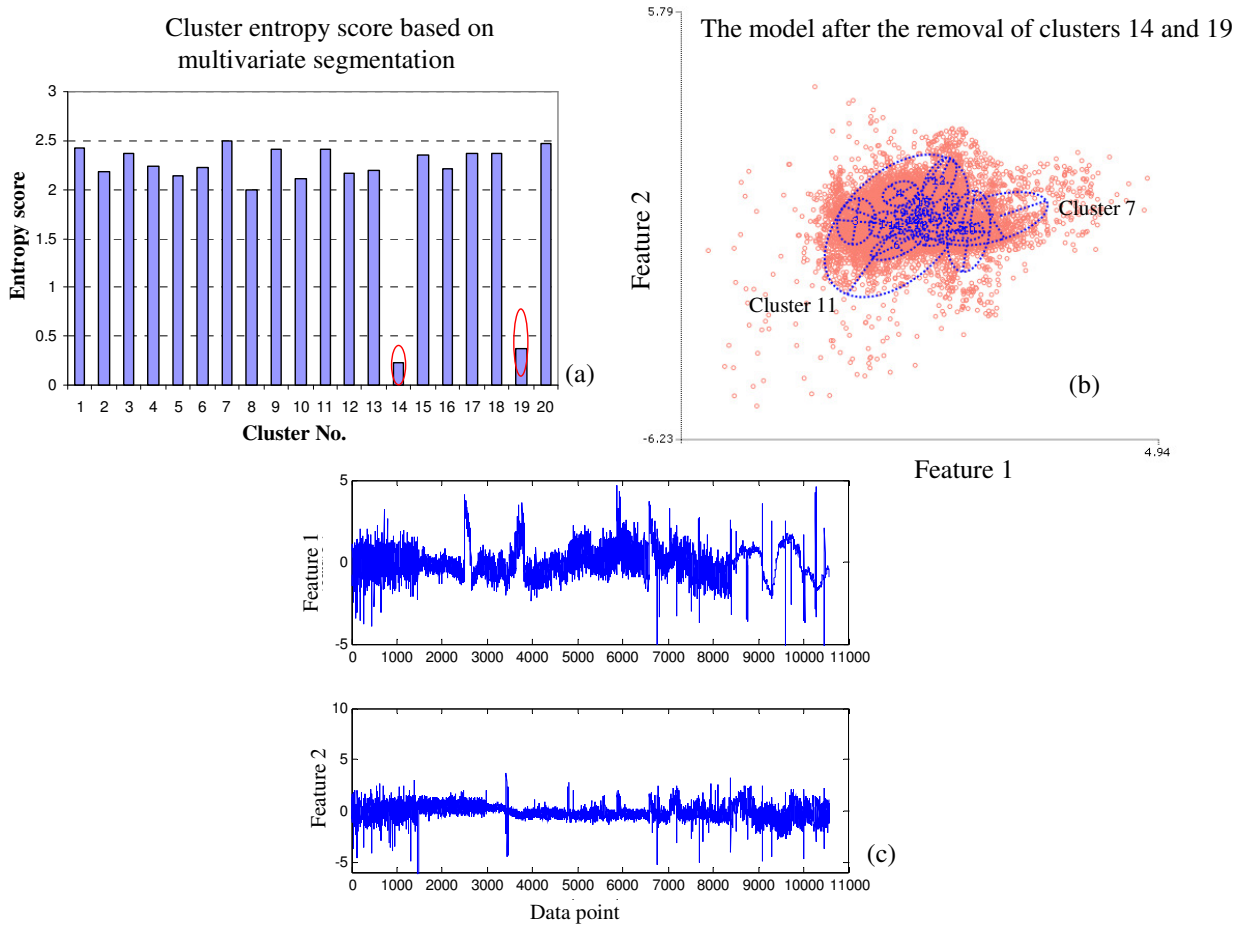


Figure 7.8 Results of the entropy examination

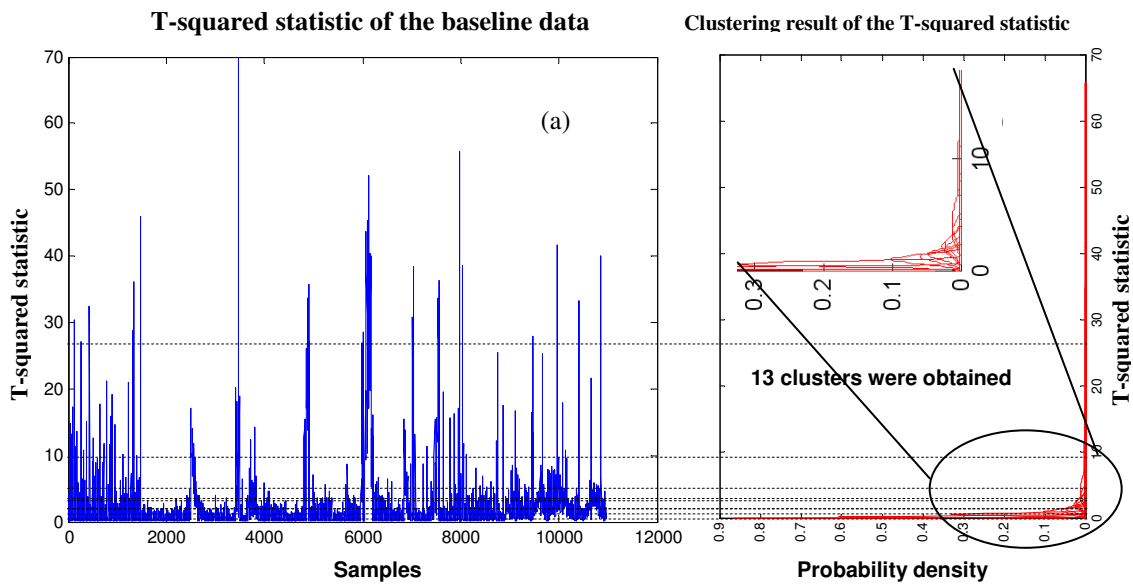


Figure 7.9 Density changes inspection for the T-squared statistic of the baseline data

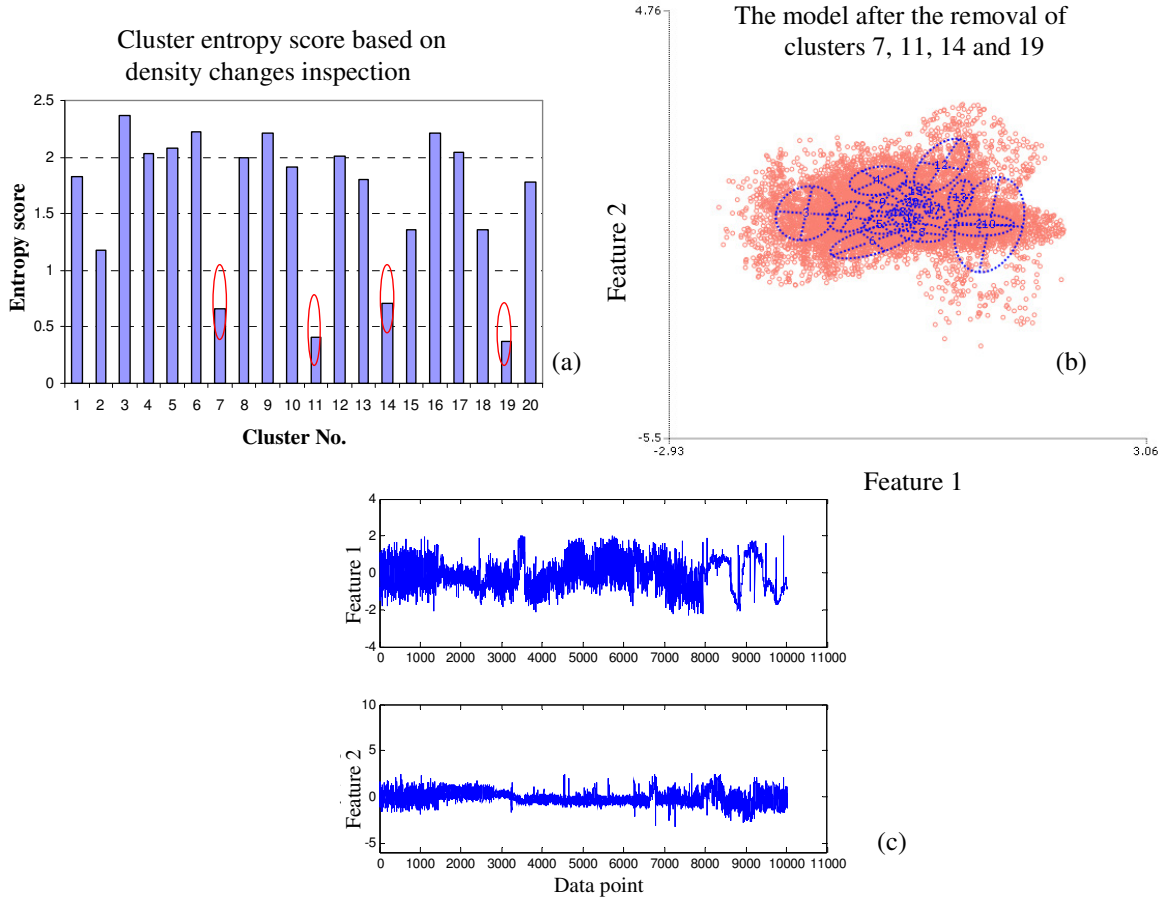


Figure 7.10 Results of the density changes inspection

Furthermore, as Section 4.5.3.2 highlights, categories can also be set in the magnitude direction, and with this approach, the T-squared statistic of the baseline data (see Figure 7.9 (a)) was evaluated by GMM associated with EM algorithm, and 13 Gaussian components (clusters) were obtained. Hence, 13 categories can be set as the density of the data changes. Figure 7.9 (b) illustrates the categorization result in the magnitude direction, and the entropy scores of the clusters can be calculated based on this categorisation result.

Figure 7.10 (a) shows the calculated entropy scores for the original clustering space (Figure 7.6). It is found that apart from clusters 14 and 19 which have also been detected by the time series segmentation method, clusters 7 and 11 are also with the lower entropy values, this might be due to the fact that the support cases in these clusters are distributed widely in the magnitude direction. Figures 7.10 (b) and (c) show the adapted model and dataset after the removal of clusters 7, 11, 14 and 19, and it is found that the remaining

clusters are more concentrated, which is consistent to the red line boundary drawn in Figure 7.6 and most of the abnormal characteristics have been filtered out from the combined baseline data which currently are within a certain low level. Therefore, the adapted model, as Figure 7.10 (c) shows, is used for the following anomaly detection.

(3) Anomaly detection results

After the construction of the anomaly detection model, 4 RTF tests are proposed to be examined against the model. Figure 7.11 shows the anomaly detection results of the 4 RTF tests based on the adapted model (see Figure 7.10 (b)). However, more anomaly detection tests based on different combinations of models and threshold set-up strategies were also conducted to compare the performances. Table 7.3 shows the calculations of the true positive and false alarm rates of the different anomaly detections for 4 RTF tests. It is found that for 4 RTF tests, the true positive rate of the tests based on adapted models can be significantly increased compared to the ones before the adaptation. Moreover, for both situations of before and after the model adaptations, the proposed GMM-EVT threshold set-up strategy can decrease the false alarm rate compared to the other strategies.

Therefore, with the simulated data, the developed scheme has again validated to be superior to the other anomaly detection methods. As Figure 7.11 shows, the proposed anomalies within the 4 RTF tests can be successfully detected. For RTF1, a step change at the end of the test was recognized above the threshold level, a small abnormal trend and significant increasing trend was detected in RTF2, an abnormal oscillation trend can be seen throughout the RTF3 test, and the two small abnormal trends and an increasing trend can also be identified at the end of RTF4.

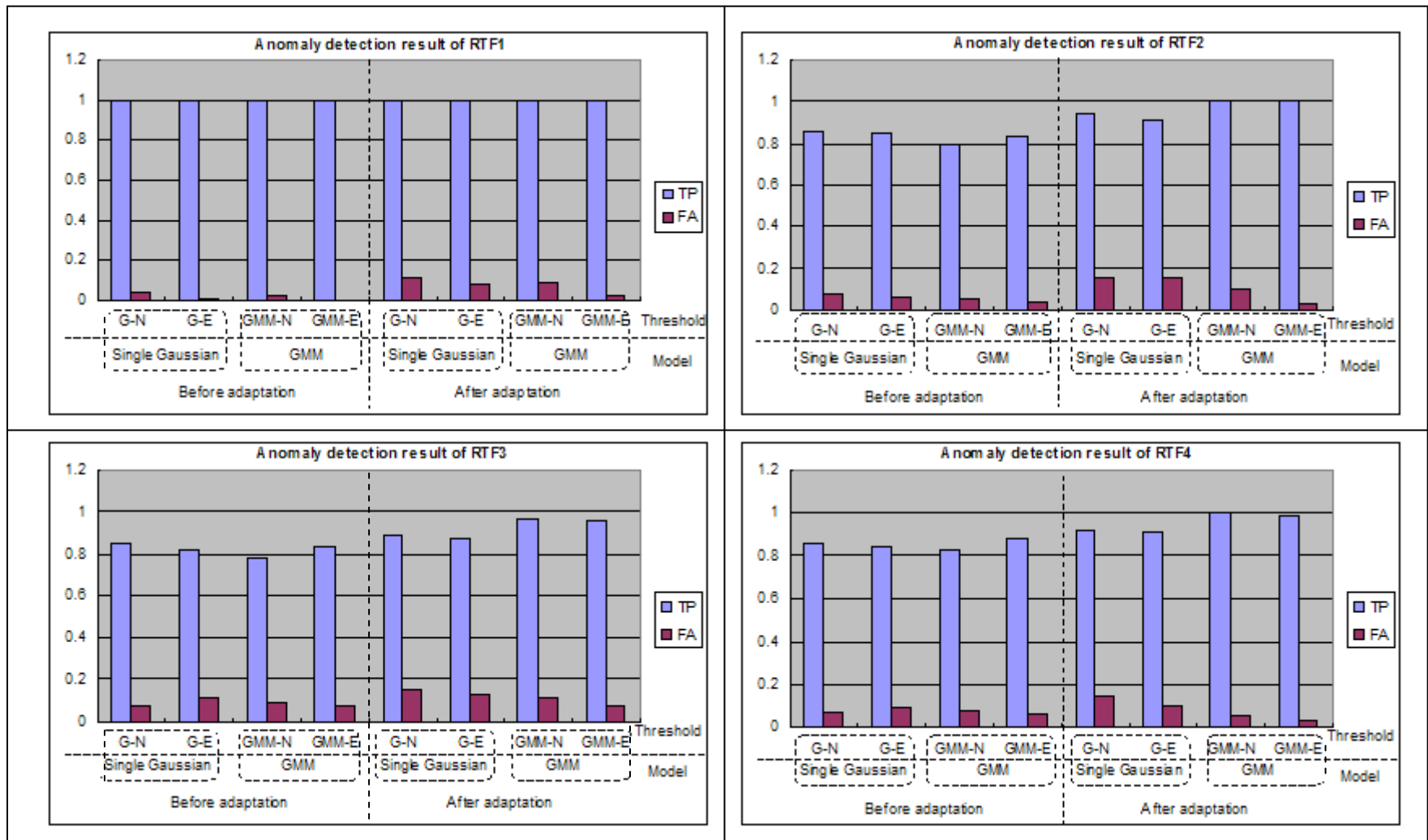


Table 7.3 Summary of the anomaly detection rates based on different models and threshold set-up strategies

G-N: Gaussian-normal
 G-E: Gaussian-EVT
 GMM-N: GMM-normal
 GMM-E: GMM-EVT

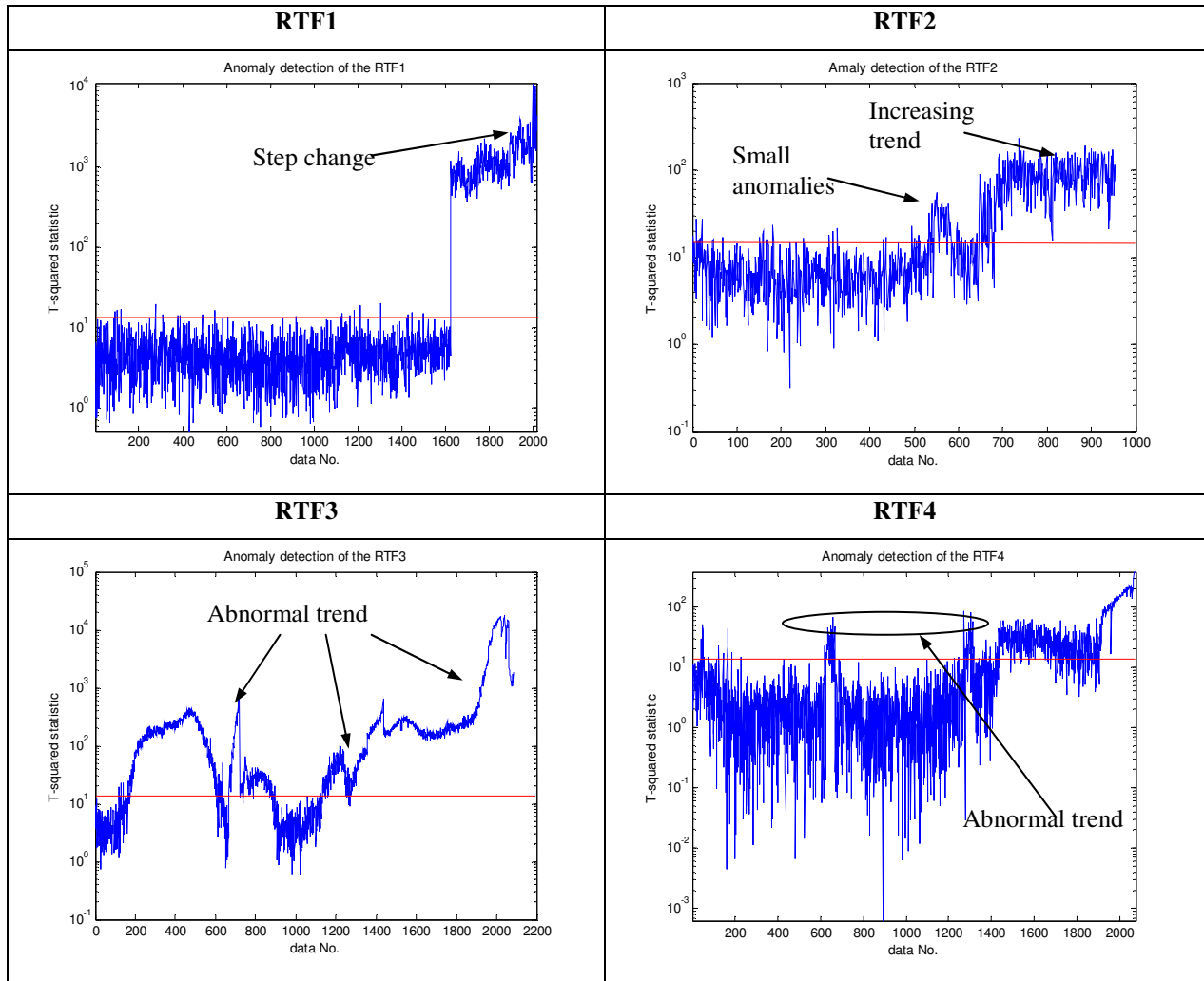


Figure 7.11 Results of the anomaly detection based on the adapted model and GMM-EVT threshold

(4) Fault diagnosis

For the detected anomalies, contribution values are proposed to calculate for examining whether the provided diagnostic information is consistent to the pre-arrangement (i.e. ranking of the features and their influences on the anomalies, as the Table 7.2 summarises). Figure 7.12 illustrates the contribution values of the applied features for 4 RTF tests. It is seen in RTF1 that feature 6 shows the strongest influence on the step change, while feature 3 provides the second strongest impact; this diagnostic information is also found to be consistent to the pre-arrangement of the original features. In RTF2, feature 4 is found to have a significant contribution to the small abnormal trend while the increasing trend at

the end is found to be dominated by feature 6; these two findings are found to be the same as the original variables as well. For the other two RTFs, the provided information also demonstrates that the proposed scheme is an effective diagnostic approach.

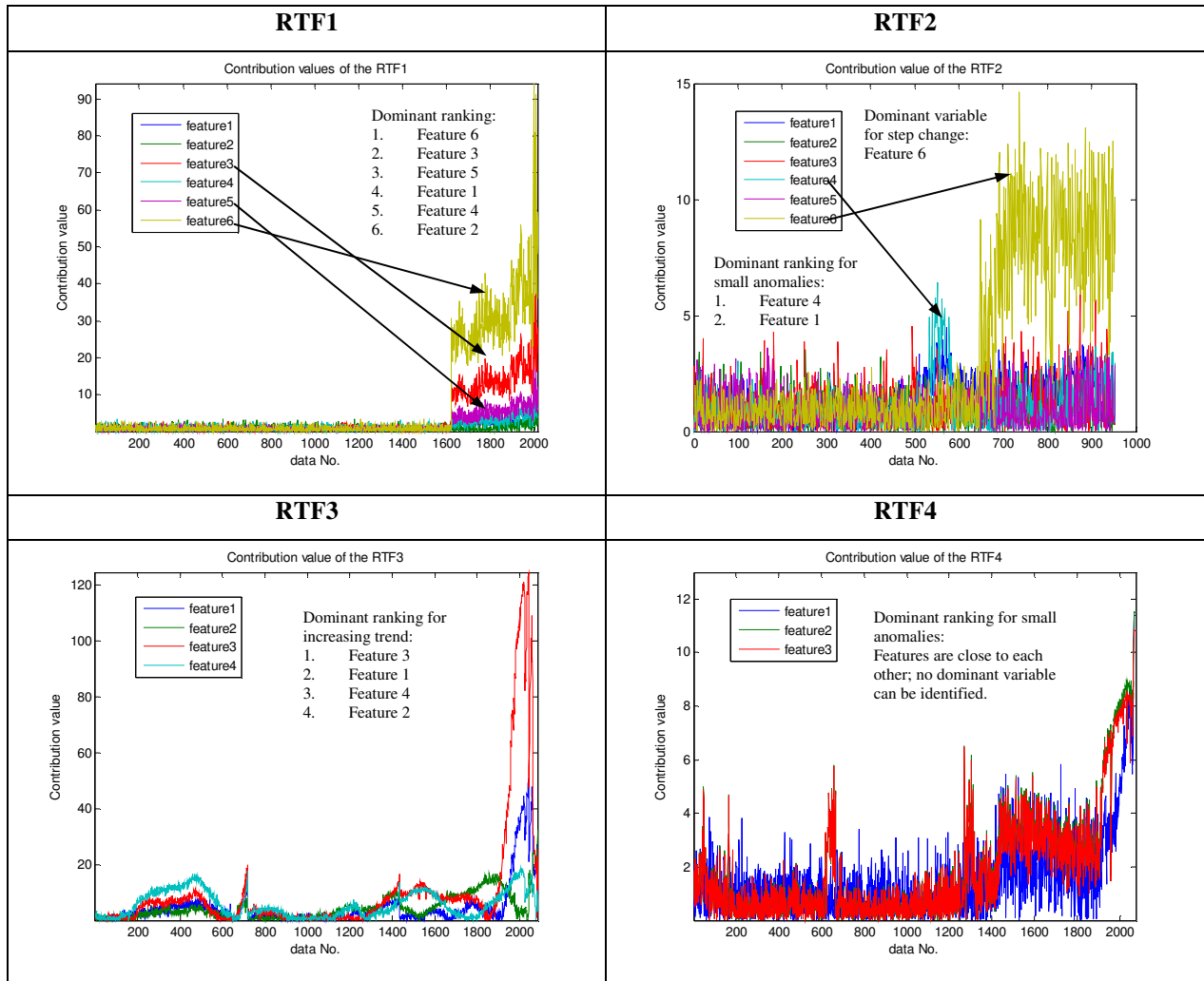


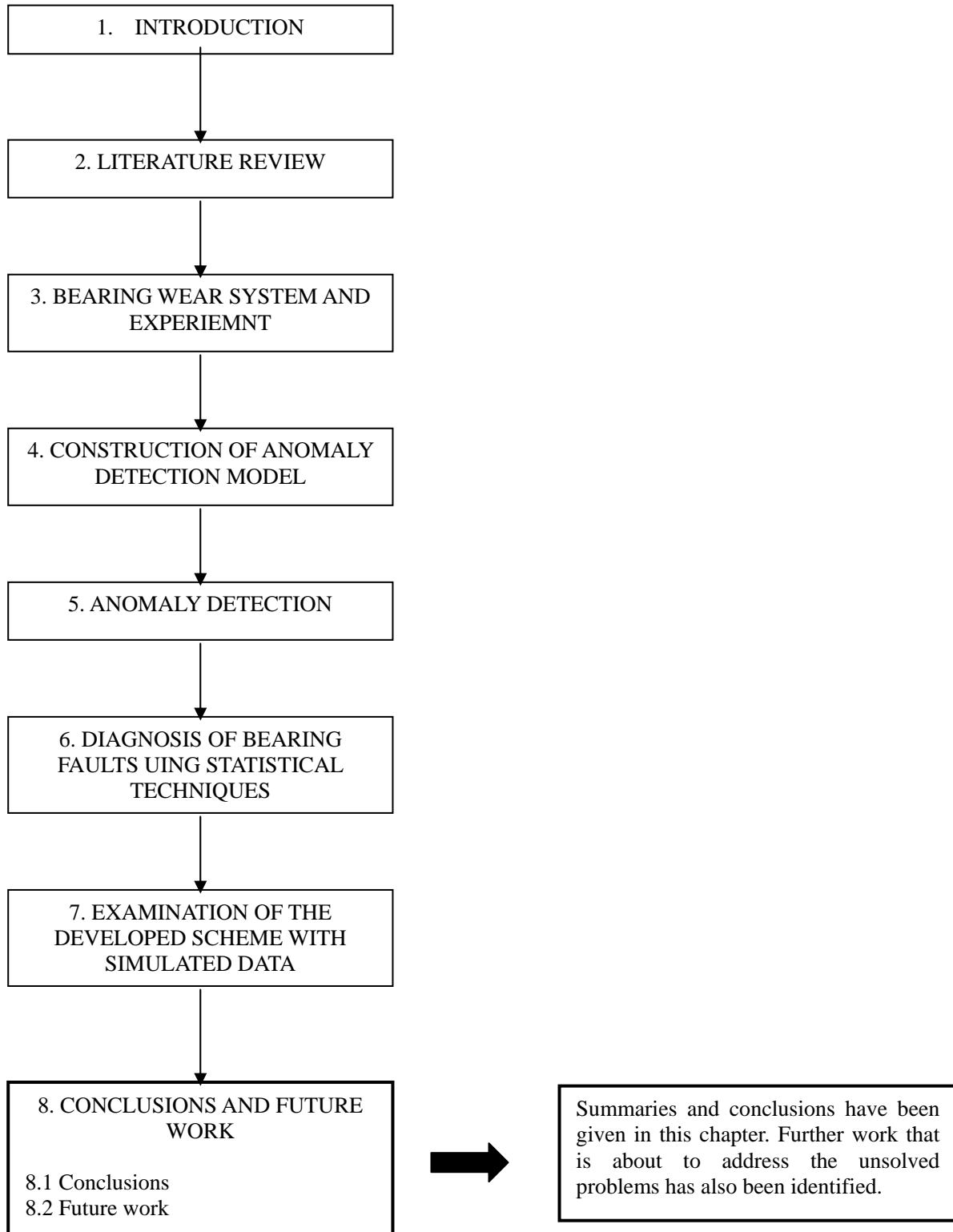
Figure 7.12 Results of the contribution value for the fault diagnosis

7.5. References

- [1] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie. Electrostatic wear monitoring of rolling element bearings. *Wear* 263 (2007) 1492-1501.
- [2] M. Craig, S.L. Chen, T.J. Harvey, R.J.K. Wood, K. Masuda, M. Kwabata, H.E.G. Powrie. Advanced condition monitoring of tapered roller bearings part I, with multiple sensing techniques. Accepted by *Tribology International*, 2008.

THESIS STRUCTURE

KEY OUTCOME



Chapter 8. Conclusions and Future work

This thesis presents a combination of data analysis, theoretical investigation and model development aimed at robust detection of early of bearing distress associated with longer term severe failure mechanisms. A wide range of literature has been critically reviewed throughout the whole period of the study. Relevant technologies are covered from multiple condition monitoring technologies to modern artificial intelligence and statistical based fault diagnosis systems. Based on the existing techniques and previous research, an innovative condition monitoring scheme has been designed and developed not only for tapered roller bearings but for more generic machinery maintenance applications. To verify the capability of the new scheme, a series of tapered roller bearing experimental data and simulated multivariate data was utilized. In the scheme, a systematic method to find the best performing anomaly detection model has been developed. A GMM-based clustering technique was selected as the fundamental development tool, and the developed systematic approach was found to be effective. To make the model more robust, innovative methods have been developed to adapt and update the anomaly detection models, and are shown to be crucial to improve the performance of the anomaly detection. Furthermore, multiple features were fused by Hotelling's T-squared statistic to extract novel events in the history of the time series. The results have shown detection of early signs for bearing failure as well as the location of such distress.

8.1. Conclusions

The aim of this study was to develop an automated condition monitoring scheme for machinery fault detection and diagnosis. This scheme is aimed at providing decision support to determine if machinery is faulty or not, as well as where the fault is and what causes these abnormal events.

8.1.1 Summary of the completed work

At the beginning of the thesis, the relevant literature has been reviewed. This review included various sensing technologies for data collection, advanced data processing

approaches to extract useful features from the original dataset, and the most frequently applied AI based fault diagnostic methods. The aim of carrying out this comprehensive survey was to isolate the most appropriate approaches to use within the current project. On the basis of established knowledge, a proposed methodology flowchart was developed, and has been used as a guideline throughout this project.

Following on from the literature review, diagnostic training and testing database was built, in which the training dataset was constructed with baseline tests (with healthy bearings) while the testing data was grouped with the run-to-failure tests (with defect bearings).

After the construction of the database, the anomaly detection model was established. In this thesis, the Gaussian mixture model (GMM) is chosen as the fundamental tool for establishing the model. However, the GMM involves a relatively complicated training process and various training parameters, and it is almost impossible to generate a consistent and well-performing model by just randomly selecting the values of training parameters. Therefore, the training parameters, such as the number of the clusters, the number of the candidate models, the initial iterations and covariance structures were assessed to evaluate their influence on the model performance.

Moreover, the problem that healthy training data might contain an amount of unknown anomalies was considered, and innovative multivariate based model adaptation methods using information theory and distance measurements were developed. And these approaches were used to identify and remove clusters in the Gaussian mixture model associated with anomalies within the training datasets. This process is considered quite important, because not only the fault masking effects can be eliminated, but also the anomaly detection rate can also be increased. Apart from model adaptation and in order to make the model more robust, a novel approach to select the appropriate data to update the model is also implemented.

Before conducting anomaly detection, the threshold level was set to achieve an automated process. In this thesis, a novel approach based on extreme value theory was developed. One of the important issues identified in this project was to detect abnormal events in real time accurately. Under this consideration, several novel fault detection techniques were developed. Firstly, a mixture model based clustering algorithm was applied by investigating the trained parameters of purity and the number of trained clusters; these two trained parameters were then used as the fault detection indices to find any abnormal events. Next, three statistical parameters including Hotelling's T-squared statistic, extreme value probability and log-likelihood were calculated based on an established normal reference model developed with a Gaussian mixture model and healthy training data.

After completing the fault detection approaches, a complete data-driven method was designed for fault diagnosis based on Principle Components Analysis (PCA). As the candidate variables were fused by the fault detection indices, the trend of each variable was also masked by the Gaussian Mixture Model (GMM); the PCA was then utilized to reveal these variables again by calculating contribution values of the detected faults, so that the location and causes of the faults could be determined.

It is obvious that the above fault detection and diagnosis are all numeric statistical tools and may not be fully demonstrated, if only bearing experimental data is applied to verify their performances. Under this circumstance, numerous multivariate data were simulated to evaluate the new condition monitoring scheme and its performance.

Generally, proposed automated condition monitoring for the machinery fault diagnosis scheme has been constructed and developed. Several cases with the data of bearing wear rig were also tested to examine the capability of the developed scheme.

8.1.2. Conclusions

The conclusions drawn from the results of the completed work are as follows:

1. In order to find a systematic way to build a reliable and consistent anomaly detection model, several GMM training parameters were evaluated and it was found that both the parameters of number of candidate models (NCM) and initial iteration (II) have an impact on the performance (indicated by BIC score) and stability of the models (indicated by similarity measure), when their values are low, such as 1, 5 and 10 for NCM and 10, 50 and 100 for II. Generally, the value of the BIC decreases from -6400 to -7200 (lower the better) and similarity between the models can be improved from 0.2 to 0.8 (larger the better, 1 is maximum). However, model performance and stability only slightly change, when the values of these two parameters keep increasing. Furthermore, the models with a full covariance structure generally perform better than the ones with a diagonal structure.
2. One of the important innovations in this thesis is the development of an approach to identify and remove anomalies within the training data. To achieve this, the entropy statistic of the information theory can be used to locate clustering regions associated with anomalies that are occurring occasionally in the time series. To identify if the particular feature occurs occasionally in the time series, a new multivariate time series segmentation method has been designed. With this method, about 60% of the anomalies can be identified, but anomalies of different segmented categories are still grouped into one cluster. This leads to difficulty in recognition. Therefore, another new method in which the categories are assigned in the magnitude direction has been developed; this important implementation increased the detection rate to about 90%. Moreover, a so-called distance based method is used to calculate distance between clusters, and to locate and remove distant ones. The result shows that the combination of these two methods can locate most of the abnormal patterns in the training data, leading to a significant increase in the anomaly detection rate in the later stages.

3. To ensure that the developed model was consistently robust to new events, the approach to select appropriate data to update the anomaly detection model has also been generated. The approach utilises two new parameters namely the; occupation probability (OP) and number of clusters (NC) to identify steady state data from other data types (i.e. running-in data, noise and abnormal trend) in the run-to-failure tests, so that the model can be consistently updated to ensure its robustness. The results of the model updating can solve the problem that new events have nothing to do with the historical training data due to the changes of speed range or load regime.
4. To set the threshold level for anomaly detection, extreme value theory and GMM are used for the first time to set the alarm level for the anomaly detection index. The result shows that new threshold defining method (GMM-EVT approach) can significantly decrease the false alarm rate with approximately 20% lower than that of the other evaluated approaches (i.e. Gaussian-normal, Gaussian-EVT) in average.
5. The developed clustering based fault detection method was found to be more sensitive to ‘small’ abnormal events/anomalies compared to conventional plots, and a number of precursors up to 10 hours prior to the fatigue failure have been extracted which can be correlated to valuable tribological findings. Furthermore, the developed anomaly detection model with the GMM can decrease the false alarm rates and minimise noise up to 20%, by using three anomaly detection statistics: Hotelling’s T-squared distance, extreme value probability and log-likelihood which were calculated for anomaly detection.
6. The contribution values of each variable to the detected abnormal events provide useful information and aid operators to locate and trace detected faults such as spallation that occurred on the inner race of bearing 2 (one of the test bearings), as well as to indicate the impact of the applied features. Further more, the amount of information that has been discovered cannot be seen clearly in the original

processed features. Generally, about 70% of the detected anomalies have been successfully diagnosed by this approach, and verified by the physical analysis, such as bearing inspection, SEM and debris analysis.

7. A series of multivariate data using the Gaussian mixture approach have been simulated to repeat the automatic data fusion process. Moreover, the unique advantages of this approach have been demonstrated during the construction and validation stages. Several tests were carried out to verify its capability and the results showed a satisfactory anomaly detection rate of 90% in average.
8. Most importantly, this study has developed an automated condition monitoring system including model optimisation, model adaptation, anomaly detection and fault diagnosis. These system elements are not only developed separately, but also incorporated into the system, so that the data can be processed step by step within the system.

8.2. Future work

Although this study has made significant advances for automated bearing fault detection and diagnosis, there are still places that need further development and optimization.

As proposed in the methodology flow chart, data pre-processing should be carried out before the fault detection and diagnosis, since it directly affects the test results. In the future work, the data pre-processing strategies should be designed separately for both training and testing data. As discussed in the fault detection chapter, the training data usually contains a number of outliers which have a significant influence on the testing result, if the training data contained outliers are similar to the novelties within the test data. So far, the entropy based method has been developed to identify outliers within the training data, as it is a relatively new topic, the potential of further investigation is suggested. On the other hand, testing data pre-processing should focus on handling missing value, sparse value and trend analysis. It should be mentioned that the PCA based fault diagnosis method introduced in Chapter 6 was developed with the assumption

that the normal reference data could represent a single set of combined variables. This problem has been solved with the application of GMM. In future work, the contribution values should be calculated based on the mixture principle components to see if the contribution values could reveal more valuable information. Finally, it is necessary to conduct bearing tests with the AI analysis on-line or real-time, so the tests can be stopped at the time when the analysis shows the precursor events and investigate the bearing components for the location and cause to see if small sub-surface cracks (early delamination) could be detected. In general, the developed scheme needs to be improved by implementing more meaningful techniques such as data filtering, and carrying out more data tests to update knowledge in the future study, in order to solve more realistic problems like reasoning the nature of the faults, isolation of the sensor faults and providing decision support for the machinery health management. In the following, more specific points are given to show which aspects of the scheme needs to be further developed:

- 1) Data pre-processing: There are many other types of anomalies in the real bearing data (e.g. due to instrumentation malfunctions and maintenance actions) that could mask important abnormal trends due to bearing problems. One example of this had been found in the data review. The current data pre-processing performs a kind of differencing or filtering and does not obtain the ability to separate an abnormal trend and step change or existence of noise from each other. It is recommended that other trend pre-processing options be explored.
- 2) Advanced signal processing: More advanced signal processing techniques such as envelop detection and wavelet processing are worthwhile to be conducted for the raw data, to generate new features for the intelligent analysis. The purpose of doing this is to examine whether there will be more information discovered through the combination of the advanced signal processing and intelligent analysis.
- 3) Model adaptation: The application of this technology is relatively new and more validation process is needed to refine the models. The preliminary result of the

anomaly detection showed that there are a large number of T-squared statistics with high values being generated by the models. In some cases this is because of the situation that the model adaptation was maybe too severe, hence the approach to re-adapt the model is needed. However, in most cases the modelling is identifying real anomalous data. A great percentage of this is due to the instrumentation malfunctions. Although many problems related to bearings had been rectified, these instrumentation problems are more difficult to trace with the current anomaly detection capability. Furthermore, the process of model adaptation needs to be optimised. Current method suggests removing clusters with low entropy scores, but seems without sophisticated criteria to justify, and where to stop the finding procedure is also not defined. Hence, further work can be carried out based on the above concerns.

- 4) Probabilistic threshold set-up: Setting a solid threshold on the anomaly detection index to separate abnormal trends from normal data is not a particularly perfect method and restricts system capability such as on-line monitoring. Current threshold set-ups were established based on the evaluation of the historical training data, and suitable for off-line analysis. However, real-time information such as the status of the testing data has not been taken into account. There could be a difficult to compare the established model and the real-time data if there is a weak link between them. Therefore, a probabilistic measure of an anomaly detection index of the testing data is more reasonable, and extremely helpful if this measure could be utilised to update the anomaly detection model and the threshold.
- 5) Anomaly detection: The established theoretical models are required to be further validated by the process of data mining of features with abnormal trends. Furthermore, as a part of the anomaly detection, model diagnostics can identify the features that are driving the trends and it would be very valuable to include these diagnostics in the system. Hence the established diagnostic information can be tested by mining the identified driving features. The future work on the anomaly detection is desired to be conducted around the issue of discovering more diagnostic knowledge through anomaly detection capability.

- 6) Fault diagnosis: The current fault diagnostic capability is achieved by data-driven approaches; no more directed or automated information could be provided. Therefore, automated reasoning is required to be developed by learning the information across bearings. Reasoning could be performed not only on the form of the anomaly such as abnormal trends, step changes and single spikes etc. but also on the features driving a trend to provide more insight details on the cause of anomalies. Finally automated reasoning could be achieved by searching for any similar stored cases in the established database.

- 7) Feature evaluation: In the current study, the features of vibration and electrostatic sensors have been assessed. Other new-implemented features such as acoustic emission and debris counter are required to be evaluated to examine their effect on the anomaly detection and diagnosis.

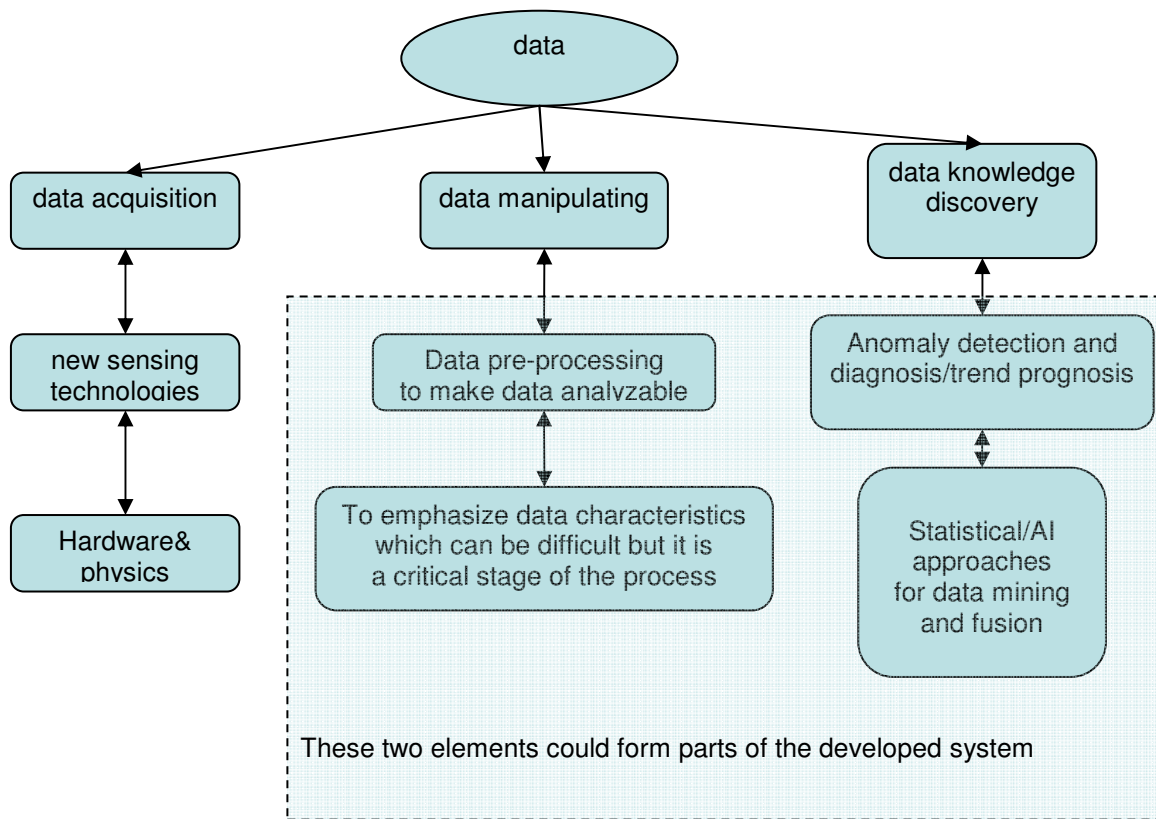
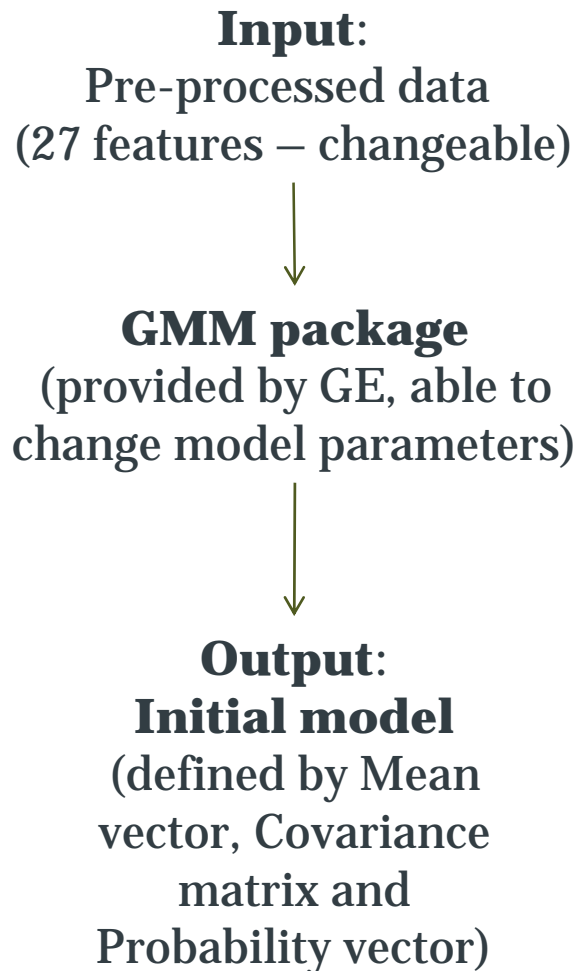


Figure 8.1 Knowledge structure of the proposed future study

8) Data interpretation: The developed system seems to be capable of performing anomaly detection and diagnostic behaviour with the demonstrated experimental bearing data and simulated multivariate data. However, there are still some defects throughout the whole procedure. a) In the data pre-processing stage, unexplained noise and anomalies existed in the pre-processed features; this might be due to the sensing malfunction, data logging problem or even the environmental effect. This is necessary to be verified in the future study. b) In the anomaly detection, some expected abnormal trends were not significantly detected. This is probably because the model is over trimmed or still contains unwanted anomalies, so that the model cannot respond to similar abnormal trend in the run-to-failure test data; or abnormal trends do not actually exist in the original processed features. Therefore, the future development is required not only to confirm the reason but to quantify the effect of model adaptation on the anomaly detection results. c) In the fault diagnosis, the situation of failure components related features are not dominant and close to other health components related features are sometimes occurred. Hence, the correct diagnostic information cannot be provided. This phenomenon is required to be further explored to confirm whether it is the algorithm related problem or the true reflection of the original features. d) To conclude the points described above, Figure 8.1 shows the knowledge structure that have the potential to be further analysed. It is seen that three processing elements (data acquisition, data manipulating and data knowledge discovery) consists of the whole developed system, and work are conducted surrounding the core working objective—data. As far as the author concerns at the time of writing, current developed systems mostly put their focus on the systems themselves, but lack of the interpretation on the inputs and outputs of the systems. Therefore, behaviour of the data distribution is sometimes not fully explained with the physical understandings, and leaves the gap between the processed data and real engineering problems. Some efforts have been tried to solve this problem in the current study, but improvements are required in the future study.

Appendix A: Key elements and flow chart for each main development

Appendix A1: The training process



Training is to:

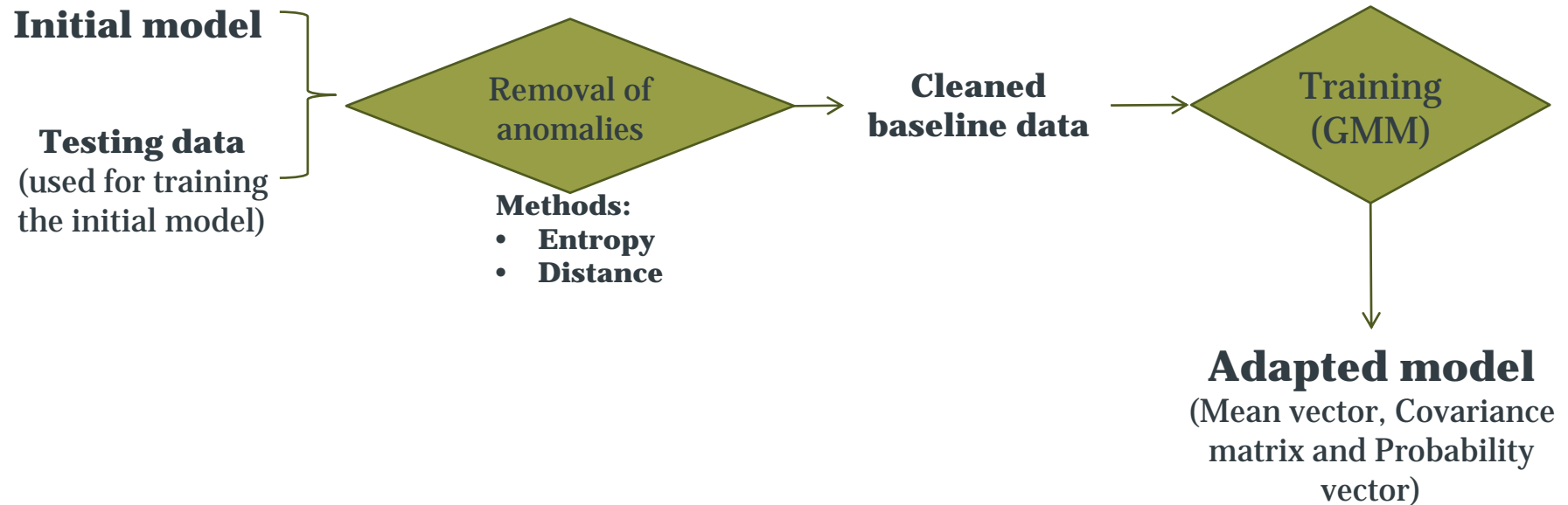
- Optimise model parameters (training parameters)
- Find number of clusters (x) in the initial model (BIC)
- Find the best candidate model (out of y no. of candidates that have x clusters)
(this is justified by maximizing log-likelihood)

Notes:

- The GMM package is provided by GE;
- Nick has developed methods for finding the best models in terms of the number of clusters and the most stable model (due to the randomness from model initiation).

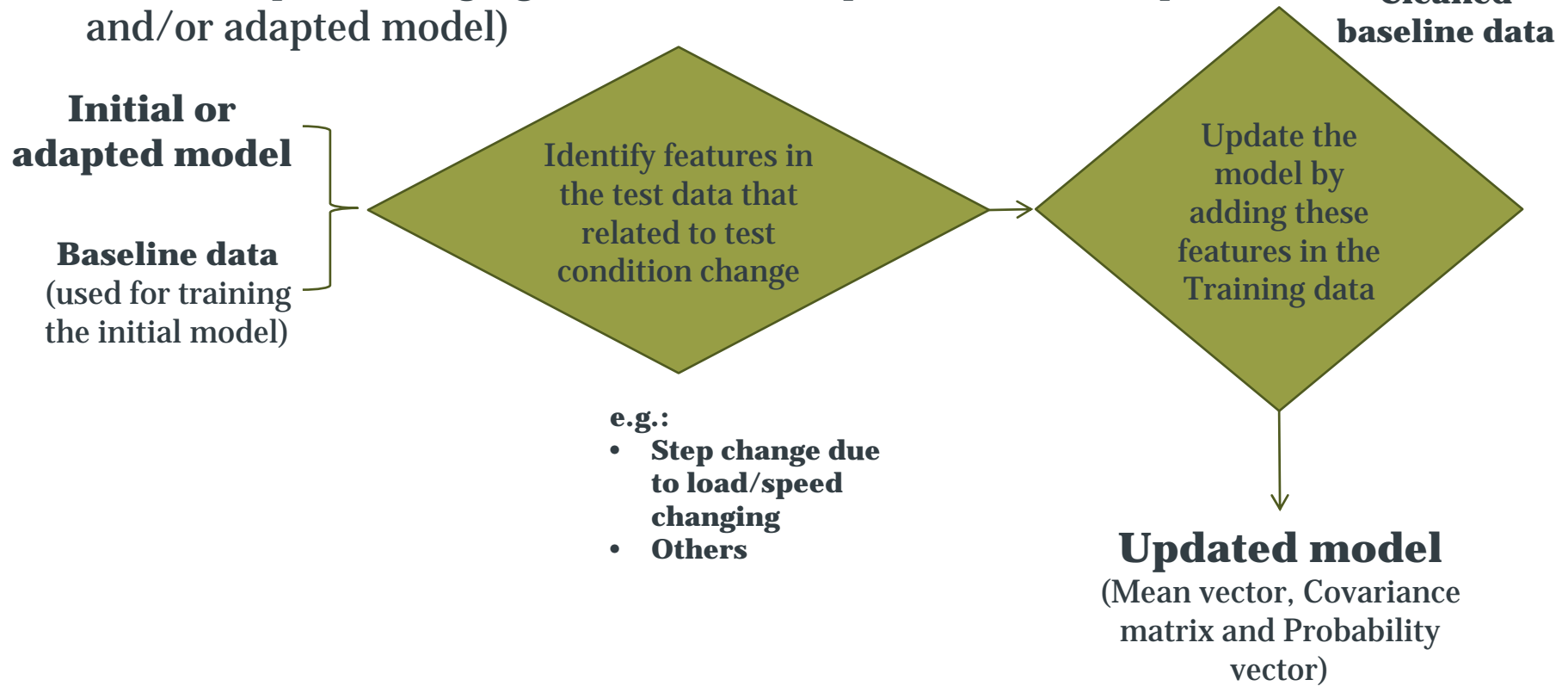
Appendix A2: Model Adaption

(Apart from model training and fault detection/diagnosis, Nick has also developed methods to adapt the initial model)



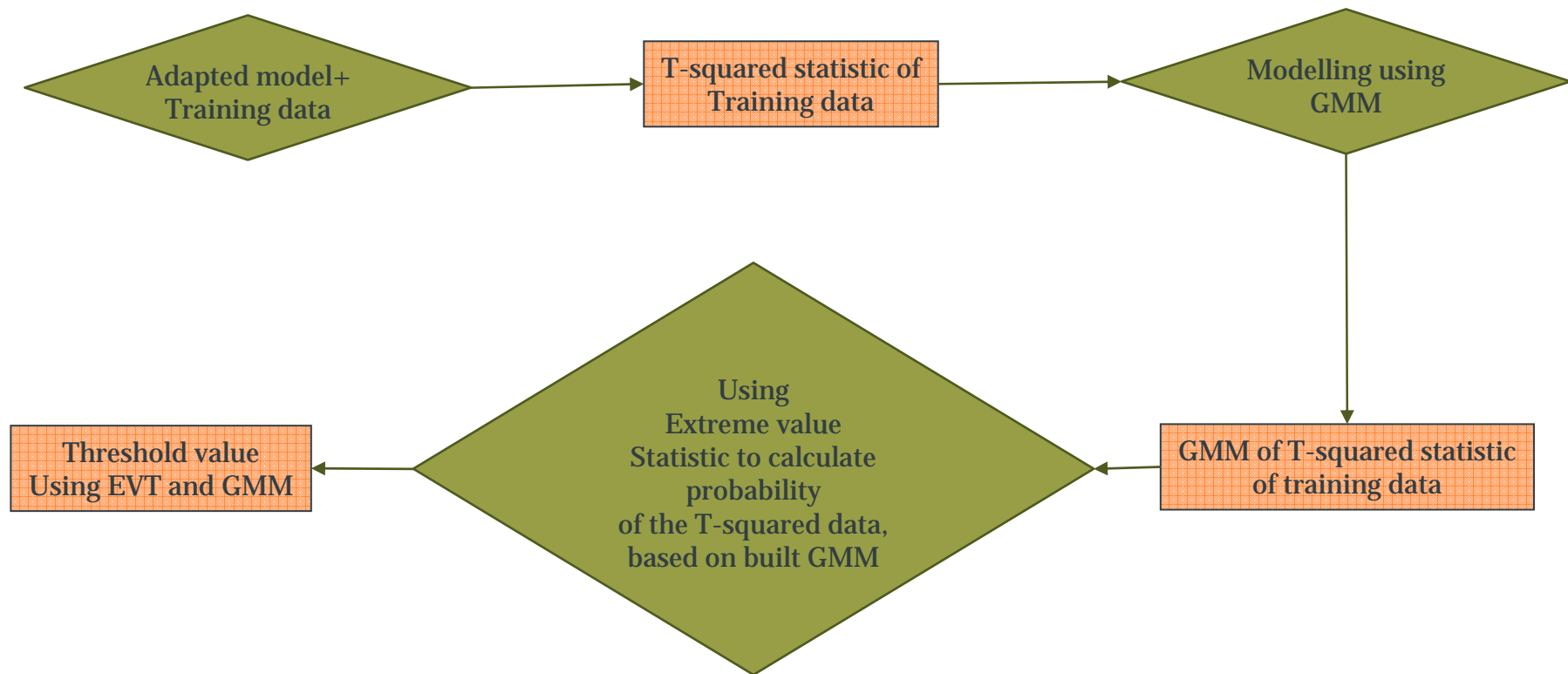
Appendix A3: Model Updating

(The adapted model can be updated by test data based on known conditions, e.g. , load and speed changing. Nick has developed methods to update the initial and/or adapted model)



Appendix A4: Threshold set-up

Nick has developed methods to set threshold value for the calculated anomaly detection index, based on extreme value statistic and GMM.



Appendix A5: The model adaptation and fault detection

The initial model
(Mean vector, Covariance matrix and Probability vector)
(Output from model training)



Exported to a Matlab program (by Nick)



Test data (pre-processed)



Results:

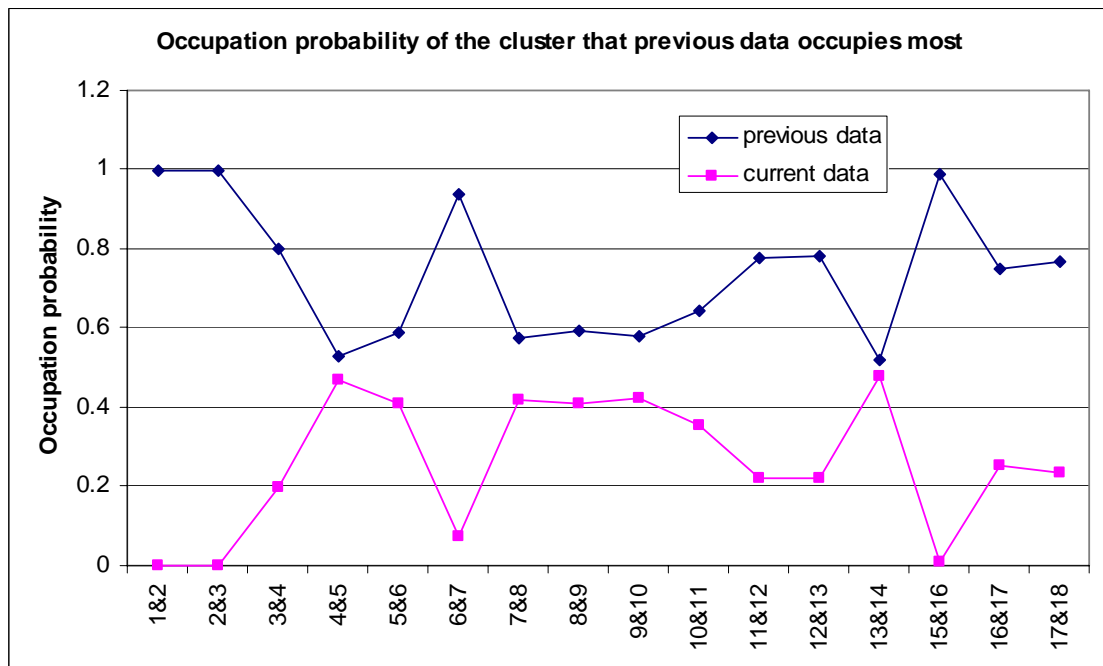
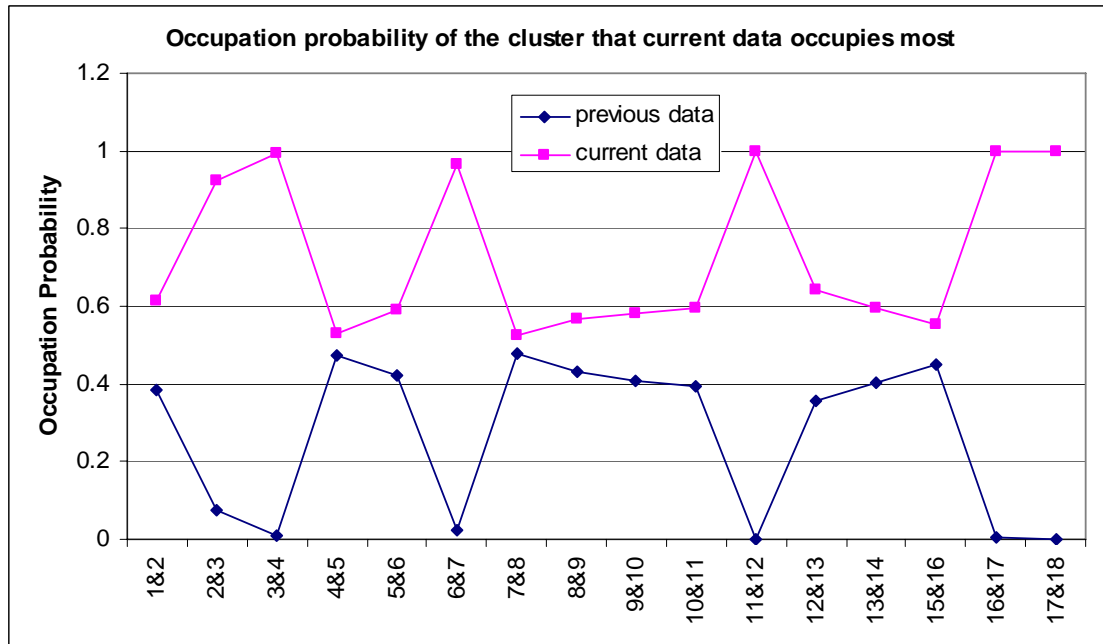
- T^2 statistics and M-distance (anomaly detection)
- Contribution values based on PCA (fault diagnosis)

Appendix B: Developed approaches and their corresponding developing programs

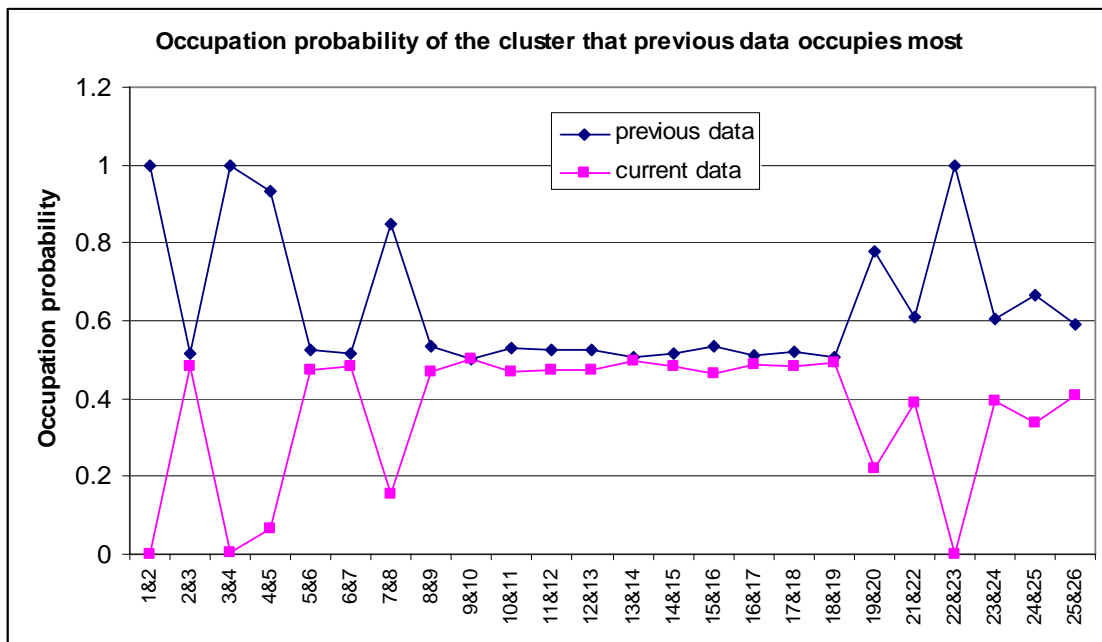
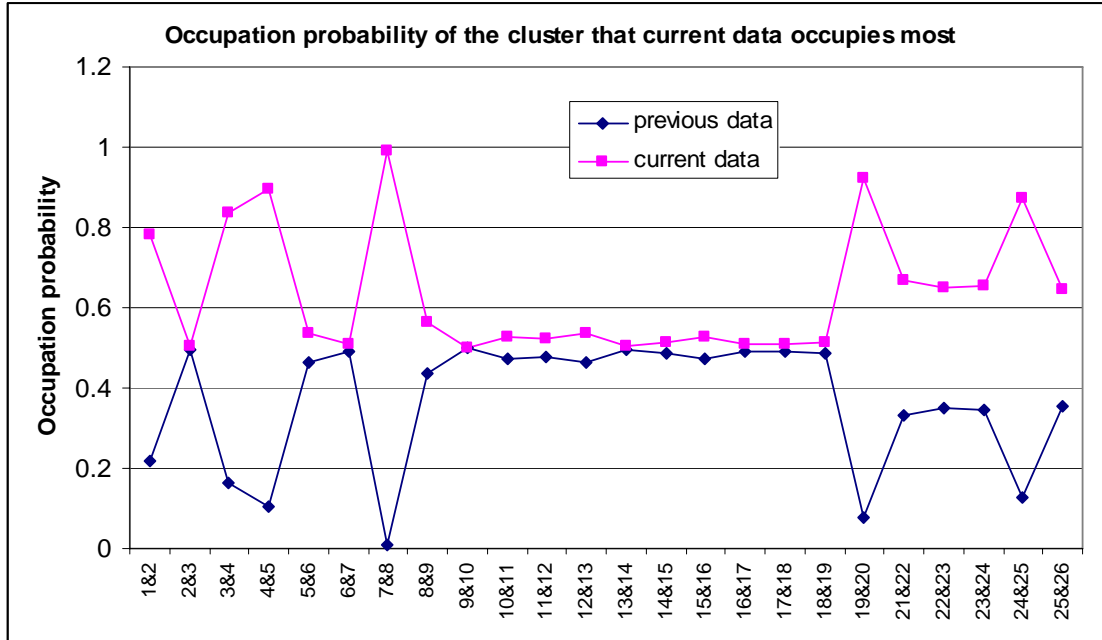
Developed approach	The tool (program) used
Applying GMM to build the anomaly detection model	GE data mining software
Method of adjusting the training process of the GMM to obtain the optimized model	GE data mining software and the program developed with MATLAB (B1)
Approaches to adapt anomaly detection model	Program developed with MATLAB (B2)
Approach to select appropriate data to update anomaly detection model	Program developed with MATLAB (B3)
Approach to set threshold level	Program developed with MATLAB (B4)
Approach to detect anomalies using T-squared statistic	Program developed with MATLAB (B5)
Approach to diagnose faults using contribution values	Program developed with MATLAB (B6)

Appendix C: Method of selecting appropriate data to update model (bearing data)

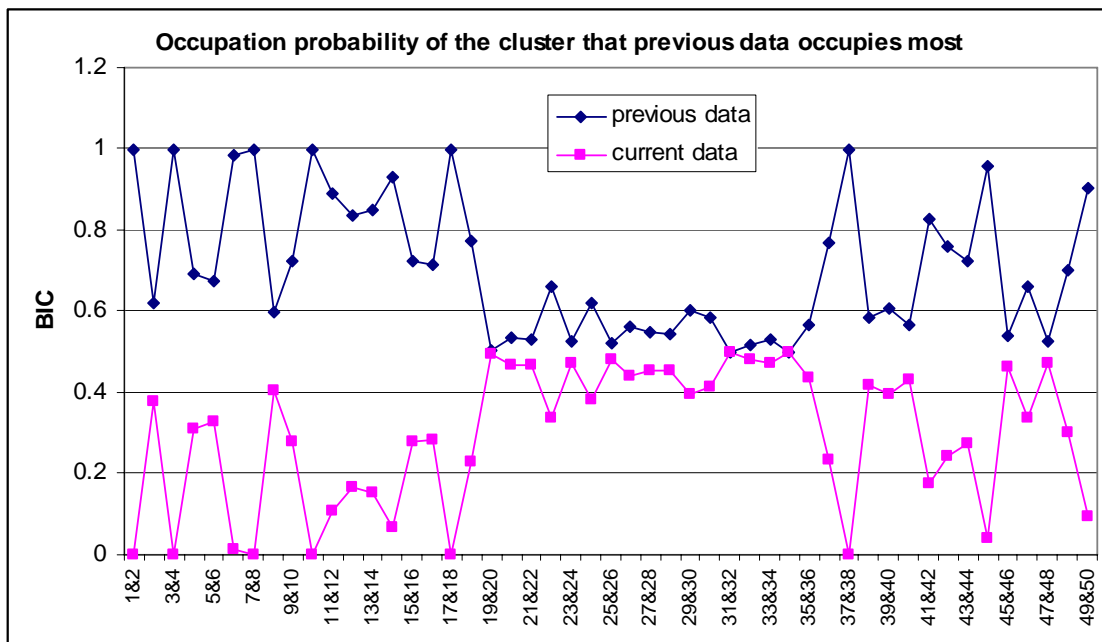
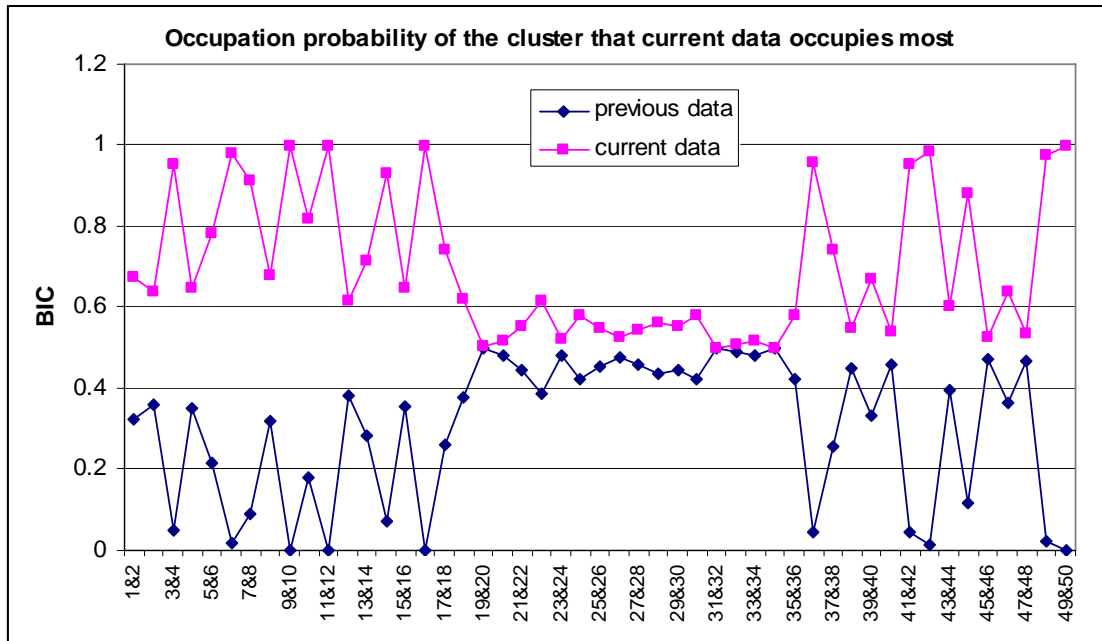
Test16 (500 points in each category)



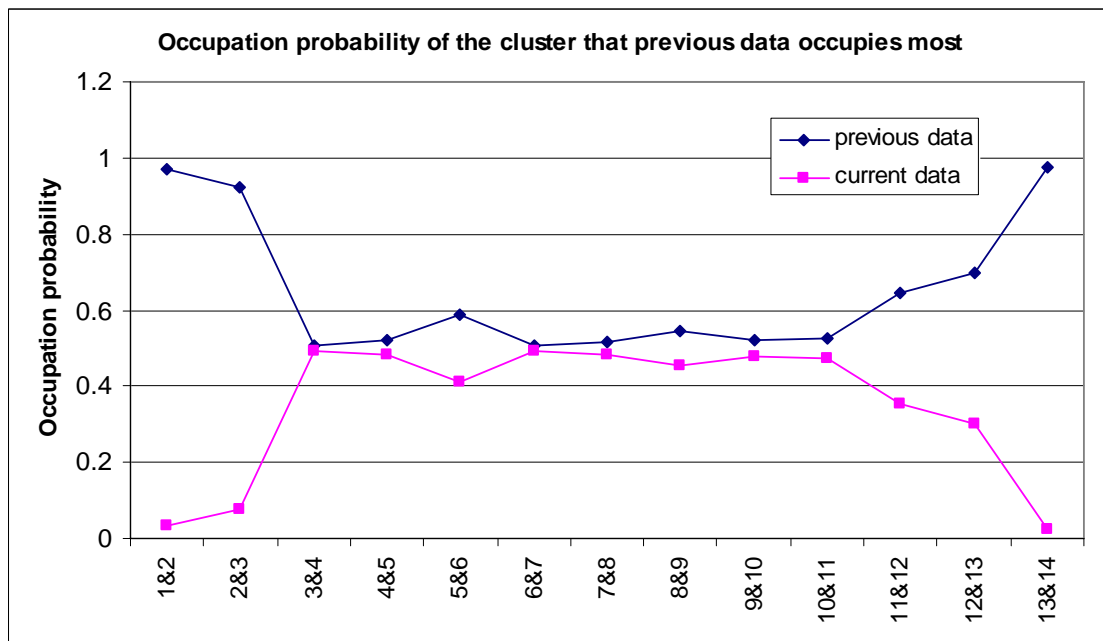
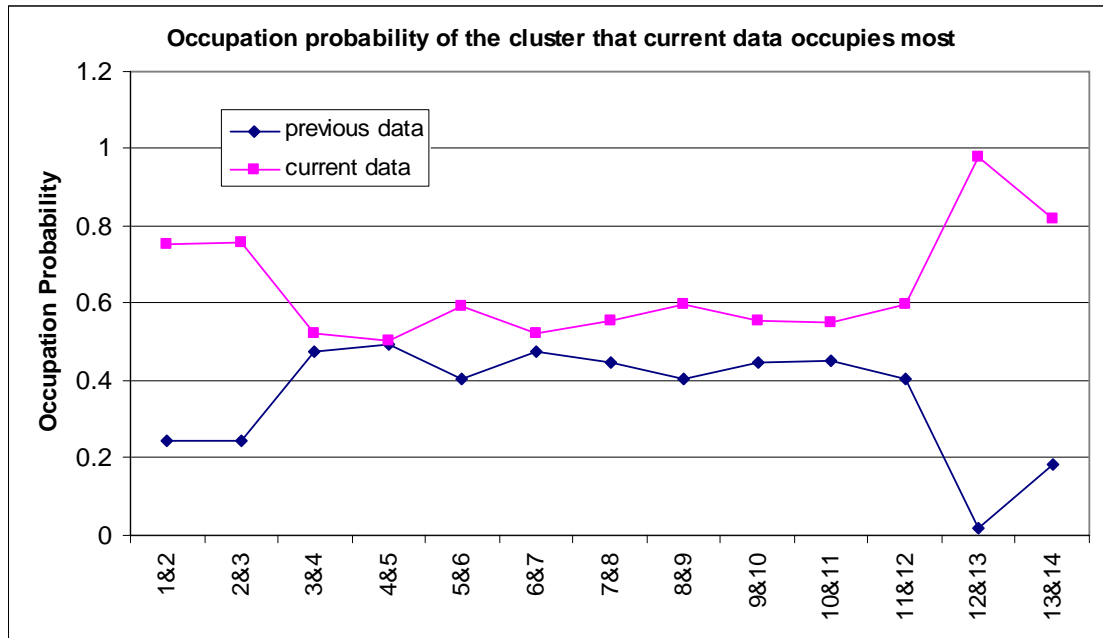
Test17 (500 points in each category)



Test 18 (500 points in each category)



Test 20 (500 points in each category)



Appendix D: Algorithm of PCA and definition of fault detection index.

Consider a data input matrix X ($n \times m$) representing n observations of m variables. PCA is conducted to transform X as follows:

$$X = TP^T = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T = \sum_{i=1}^m t_i p_i^T \quad (\text{D-1})$$

Where p_i is called principal component loading vector and:

$$P^T P = I \quad (\text{D-2})$$

t_i is called principal component score vector and:

$$T^T T = \Delta = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m) \quad (\text{D-3})$$

Where λ_i ($i=1, 2, \dots, m$) are eigenvalues of $X^T X$.

If the singular value decomposition (SVD) factors for X are given by: $X = UDV^T$, where U is a $m \times m$ matrix, D is a $m \times n$ matrix and V^T is a $n \times n$ matrix. The matrix U and V are orthonormal, i.e. $U^T U = I$ and $V^T V = I$. The n columns of V contain the eigenvectors of matrix $X^T X$ and the diagonal elements of the D matrix contain the square roots of the corresponding eigenvalues of matrix $X^T X$. Therefore, the loading matrix P and the score matrix T can be obtained as $P = V$ and $T = UD = XV$.

Normally, the first k ($k < m$) principal components capture most of the variance, while less important components, which mostly describe noise information in the data, can be abandoned without loss of significant information. By doing so, input matrix X can be reconstructed as:

$$X = \hat{X} + E = \sum_{i=1}^k t_i p_i^T + E = T_k P_k^T + E \quad (\text{D-4})$$

Where T is the principal component score matrix, P is the principal component loading matrix and E is the modelling residual. The score vectors span a lower dimensional subspace used for further analysis.

The use of T-squared distance and SPE for fault detection can be explained with a simple example shown in Figure D-1. The training data (open circle) are synthetically generated from the three-dimensional Gaussian distribution. The PCA is applied to obtain two principal components (lines). The two indices of a new input sample (full circle) could be calculated as shown in Figure D-1.

It should be noted that the version of Hotelling's T-squared distance which is based on PCA subspace is different from the one described in Chapter 5, and could be expressed in terms of principal components:

$$T^2 = zz^T = (XP_k \Delta_k^{-1/2})(XP_k \Delta_k^{-1/2})^T = t_k \Delta_k^{-1} t_k^T \quad (D-5)$$

For the calculation of SPE, the equation can be written as:

$$SPE = rr^T = (X - XP_k P^T)(X - XP_k P^T)^T \quad (D-6)$$

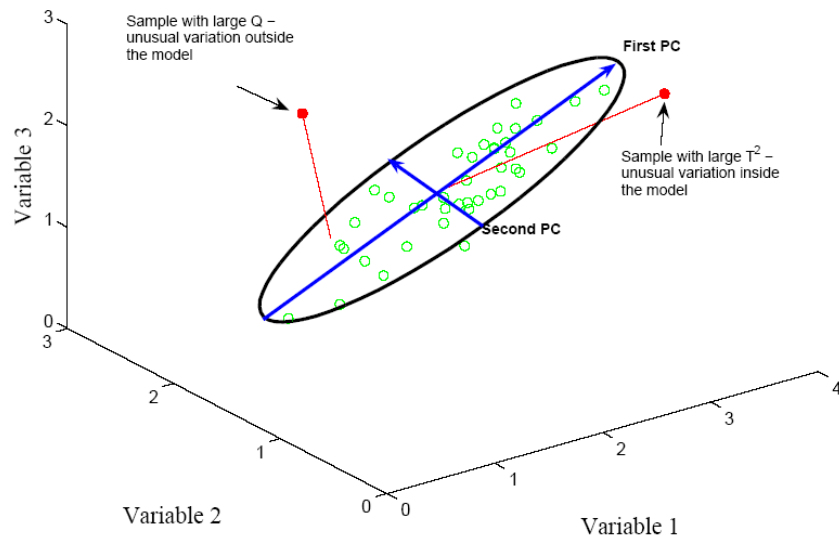


Figure D-1 Example of the principal component analysis