



University of HUDDERSFIELD

University of Huddersfield Repository

Smith, Ann

Characterisation of Condition Monitoring Information for Diagnosis and Prognosis using Advanced Statistical Models

Original Citation

Smith, Ann (2017) Characterisation of Condition Monitoring Information for Diagnosis and Prognosis using Advanced Statistical Models. Doctoral thesis, University of Huddersfield.

This version is available at <http://eprints.hud.ac.uk/id/eprint/32609/>

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

<http://eprints.hud.ac.uk/>

CHARACTERISATION OF CONDITION MONITORING INFORMATION
FOR DIAGNOSIS AND PROGNOSIS USING ADVANCED
STATISTICAL MODELS

Ann Smith

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Centre for Efficiency and Performance Engineering
The University Of Huddersfield

February 2017

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and she has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trade marks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Abstract

This research focuses on classification of categorical events using advanced statistical models. Primarily utilised to detect and identify individual component faults and deviations from normal healthy operation of reciprocating compressors. Effective monitoring of condition ensuring optimal efficiency and reliability whilst maintaining the highest possible safety standards and reducing costs and inconvenience due to impaired performance.

Variability of operating conditions being revealed through examination of vibration signals recorded at strategic points of the process. Analysis of these signals informing expectations with respect to tolerable degrees of imperfection in specific components.

Isolating inherent process variability from extraneous variability affords reliable means of ascertaining system health and functionality. Vibration envelope spectra offering highly responsive model parameters for diagnostic purposes.

This thesis examines novel approaches to alleviating the computational burdens of large data analysis through investigation of the potential input variables. Three methods are investigated as follows:

Method one employs multivariate variable clustering to ascertain homogeneity amongst input variables. A series of heterogeneous groups being formed from each of which explanatory input variables are selected.

Data reduction techniques, method two, offer an alternative means of constructing predictive classifiers. A reduced number of reconstructed explanatory variables provide enhanced modelling capabilities ensuring algorithmic convergence.

The final novel approach proposed combines both these methods alongside wavelet data compression techniques. Simplifying number of input parameters and individual signal volume whilst retaining crucial information for deterministic supremacy.

“Looking for structure in a set of numbers, without imposing rigid parametric assumptions, but still within a statistical framework of some kind.” Silverman (1991).

Contents

Abstract	3
Contents	6
Figures.....	11
Tables.....	15
Abbreviations.....	16
Nomenclature	18
Dedications and Acknowledgements.....	21
Chapter 1	22
Introduction	22
1.1 An Overview of Condition Monitoring.....	23
1.1.1 The Importance of Condition Monitoring	23
1.1.2 Information Gathering	23
1.1.3 Data Manipulation	24
1.1.4 Research Application	25
1.2 Motivation	26
1.3 Aim	27
1.4 Objectives.....	28
1.4.1 Time Domain.....	28
1.4.2 Frequency Domain.....	28
1.4.3 Variable Selection	28
1.4.4 Classifier Construction	29
1.4.5 Extended Analysis.....	29
1.5 Organisation of Thesis.....	29
Chapter 2	32
Current Practice in Condition Monitoring.....	32
2.1 The Growth of Condition Monitoring Practice	33
2.1.1 The Condition Monitoring Procedure.....	33
2.1.2 The Condition Monitoring Cycle	34
2.1.3 Data Collection for Informed Decisions	34

2.1.4 Summary.....	35
2.2 Condition Monitoring Recent Research	36
2.2.1 Breadth of Condition Monitoring Methodologies.....	36
2.2.2 Increasing and Hidden Model Complexity.....	38
2.2.3 Selection of Optimum Input Parameter Sets	38
2.3 Envelope Spectra and Fast Fourier Analysis.....	39
2.4 Condition Monitoring of Reciprocating Compressors.....	41
2.4.1 Types of Compressor.....	41
2.4.2 Compression and Expansion Cycle	42
2.4.3 Faults and Implications	43
2.4.4 Applications of Reciprocating Compressors.....	43
2.5 Summary	43
Chapter 3	45
Multivariate Statistical Techniques	45
3.1 Cluster Analysis	46
3.1.1 Review and Current Practice.....	47
3.2 Discriminant Analysis.....	48
3.2.1 Review and Current Practice.....	50
3.3 Naïve Bayes Classification	51
3.3.1 Review and Current Practice.....	53
3.4 Principal Component Analysis	54
3.4.1 Factor Analysis.....	55
3.4.2 Review and Current Practice.....	57
3.5 Support Vector Machines	57
3.5.1 Review and Current Practice.....	59
3.6 Relevance Vector Machines	60
3.7 Genetic Algorithms	62
3.8 Multiscale Principal Component Analysis	63
3.8.1. Review and Current Practice.....	64
3.9 Andrews Plots of Fourier Profiles.....	65
3.9.1 Review	66

3.10 Summary of Multivariate Statistical Techniques	66
Chapter 4	67
Condition Monitoring Data Acquisition from a Compressor Rig.....	67
4.1 Reciprocating Compressors.....	68
4.2 Test Facilities and Data Collection.....	68
4.2.1 Reciprocating Compressor Rig	69
4.2.2 Sensor Details and Specification.....	72
4.3 Experimental Configuration	73
4.4 Data Structure for MATLAB analysis	74
4.5 Summary	74
4.6 Physical Attributes of Output Signal Data and Measurements	75
4.6.1 Low Pressure Measurements.....	75
4.6.2 High Pressure Measurements.....	76
4.6.3 First Stage Vibration Measurements	79
4.6.4 Second Stage Vibration Measurements.....	79
4.6.5 Speed Measurements	82
4.6.6 Angular Speed Measurements.....	82
4.6.7 Current Measurements	83
4.7 Waveform Features	83
4.7.1 First Stage Pressure and Vibration Measurements in Relation to Index Pulse	83
4.7.2 Second Stage Pressure and Vibration Measurements in Relation to Index Pulse	84
4.7.3 Relationship between Cylinder Pressure and Reservoir Pressure	84
4.7.4 Relationship between Speed and Reservoir Pressure	91
Chapter 5	93
Exploratory Data Analysis	93
5.1 Fault Classification Using Statistical Modelling and Graphical Means in the Time Domain	94
5.2 Signal Analysis in the Frequency Domain	101
5.2.1 Frequency Spectra of the First Stage Vibration Signals.....	101
5.2.2 Frequency Spectra of the Second Stage Vibration Signals.....	102

5.2.3 Envelope Spectrum Analysis of Vibration Signals.....	102
5.2.4 Envelope Spectra for the Motor Current Signals.....	107
5.3 Preliminary Investigation into Fault Diagnosis in the Frequency Domain	111
5.4 Summary	118
Chapter 6	120
Multivariate Classifiers Using Reduced Number of Input Parameters.....	120
6.1 Identification of Separable Classes.....	121
6.1.1 Andrews Plots of the Second Stage Vibration Signal.....	121
6.1.2 Andrews Plots of the Motor Current Signal	121
6.2 Variable Clustering	124
6.2.1 Clustering of Second Stage Vibration Signal Envelope Harmonics.....	124
6.2.2 Cluster Analysis of the Motor Current Signal Envelope Harmonics.....	127
6.2.3 Cluster Analysis of the Second Stage Pressure Signal Envelope Harmonics	129
6.2.4 Investigative cluster scatter plots	130
6.3 Discriminant Analysis.....	133
6.4 Naïve Bayes Classification	136
6.5 Summary	141
Chapter 7	142
Multivariate Classifiers Using Variable Reduction Methods	142
7.1 Principal Component Analysis of Envelope Harmonics	143
7.2 Confirmatory Factor Analysis.....	146
7.3 Support Vector Machine Classification	147
7.4 Summary	148
Chapter 8	151
A Combined Approach to Variable Reduction Incorporating Signal Compression .	151
8.1 Combined Variable Reduction Technique	152
8.2 Multiscale PCA	153
8.3 Data Compression	156
8.4 Summary	166
Chapter 9	167

Review of Thesis Achievements	167
9.1 Review of Thesis Objectives.....	168
9.1.1 Time Domain.....	168
9.1.2 Frequency Domain.....	169
9.1.3 Variable Selection	169
9.1.4 Cluster Construction.....	170
9.1.5 Extended Analysis.....	171
9.2 Contributions to Knowledge	172
9.3 Conclusions.....	174
9.3.1 Initial Exploratory Analysis Findings.....	174
9.3.2 Class Profiling and Variable Selection	175
9.3.3 Classifier Construction	177
9.3.4 Extended Analysis.....	177
9.4 Novelties.....	178
References.....	180
Publications.....	189
Appendices	190
A1. Experimental Compressor Rig.....	190
A2 MATLAB Code	194
A2.1 MATLAB Code for Cluster Analysis	194
A2.2 MATLAB Code for Discriminant Analysis.....	194
A2.3 MATLAB Code for Naïve Bayes Classification	196
A2.4 MATLAB Code for PCA	198
A2.5 MATLAB Code for SVM.....	199
A2.6 MATLAB Code for MSPCA.....	201
A2.7 MATLAB Code for Andrews plots	203

Figures

Figure 3.1 SVM Illustration of Margins and Hyperplanes.	59
Figure 4.1 Broom WadeTS9 Reciprocating Compressor Rig.	71
Figure 4.2 Comparison of the Low Pressure Measurements per Section. (a) Section 1, (b) Section 2 and (c) Section 3.	76
Figure 4.3 Annotated Plot of First Stage (Low) Pressure Cycle.	77
Figure 4.4 Comparison of First and Second Stage Pressure Measurements.	78
Figure 4.5 Comparison of Compression and Discharge Phases for First and Second Cylinder Pressure Stages.	78
Figure 4.6 First Stage Vibration Measurements of Each Segment.	80
Figure 4.7 Comparison of the Second Stage Vibrations In Each Of the Three Number Segments.	80
Figure 4.8 First and Second Stage Vibrations Highlighting Offset Actions of Suction and Discharge.	81
Figure 4.9 Comparison of Speed Measurements by Section Displaying Sensor Pulses. (a) Section 1, (b) Section 2 and (c) Section 3.	81
Figure 4.10 From Top: Inspection Of Pulses In First Hundredth Of A Second; The Differences Between Successive Pulses; Number Of Samples Between Successive Pulses; The Instantaneous Angular Speed Which Can Be Seen To Fluctuate Periodically.	85
Figure 4.11 Motor Current During First Cycle.	85
Figure 4.12 Comparison of First Stage Pressure and Vibration Signals in Comparison to Index Marker (First Section).	86
Figure 4.13 Comparison Of First Stage Pressure And Vibration Over First Cycle Relative To Pulse.	86
Figure 4.14 Second Stage Pressure And Vibration Cycles With Index Reference. ...	87
Figure 4.15 Comparison Of Second Stage Pressure And Vibration Within First And Second Pulses In Relation To Index Pulse.	87
Figure 4.16 Relative Values of IAS, First Stage Pressure and Vibrations In Comparison to the Index.	88
Figure 4.17 Standardised Values for Direct Comparison [$IAS-430$; $10 * pL / \text{Mean } (pL)$; $10 * pH / \text{Mean } (pH)$].	88
Figure 4.18 Comparison Of First Stage Cylinder Pressure By Reservoir Load. Note: 80psi=5.52bar, 90psi=6.21bar, 100psi=6.90bar, 110psi=7.58bar, 120psi=8.27bar. ...	89
Figure 4.19 Comparison Of Second Stage Cylinder Pressure By Reservoir Load. Note: 80psi=5.52bar, 90psi=6.21bar, 100psi=6.90bar, 110psi=7.58bar, 120psi=8.27bar. ...	90
Figure 4.20 Speed of Motor under Varying Loads (psi). Note: 80psi=5.52 bar, 90psi=6.21 bar, 100psi=6.90 bar, 110psi=7.58 bar, 120psi=8.27 bar.	92

Figure 5.1 Healthy and Faulty Signals Comparison: First Stage Pressure Measurements (load 100psi).....	95
Figure 5.2 First Stage Pressure for Each of the Five Single Fault Cases.....	95
Figure 5.3 Boxplots displaying first stage pressure characteristics by fault (load 100).	96
Figure 5.4 Scatter Plot to Show the Mean Vs Kurtosis of First Stage Pressure Per Number Segment.	99
Figure 5.5 Plot of Mean and Kurtosis for the Vibration Measurements at First Stage with Load 100psi (6.90 bar).....	99
Figure 5.6 Class Distributions in Healthy Compressor Rig.....	100
Figure 5.7 Spectrum Plot: Fundamental Frequency For Each Machine State, First Stage Vibration Signal.....	103
Figure 5.8 Frequency Spectra for First Stage Vibration Measurements to 32nd Harmonic (Fundamental Frequency 7.3Hz).	103
Figure 5.9 Second Stage Frequency Spectra for Healthy System and Four Machine Faults.	108
Figure 5.10 Envelope Spectrum of First Stage Vibration Signal.....	108
Figure 5.11 Envelope Spectra Of Second Stage Vibration Signal.....	109
Figure 5.12 Envelope Spectra Displaying Amplitudes At Decisive Harmonics In Second Stage Vibration Signal.....	109
Figure 5.13 Envelope Spectrum for Motor Current Showing Peaks at Fundamental Frequency Multiples.	110
Figure 5.14 Envelope Spectrum for Motor Current Against Harmonic Order.	110
Figure 5.15 Comparison of Model Complexity 24 Sample SVM.....	112
Figure 5.17 Comparison of Envelope Spectra of Faults to Healthy (baseline).	113
Figure 5.18 Features Selected by Class.	115
Figure 6.1 Andrews Plot Showing the Feature Profiles by Class Against time(s)...	122
Figure 6.2 Andrews Class Profiling using Second Stage Vibration Signal Against time (s).....	122
Figure 6.3 Andrews Plots for each of the Five Cases using Motor Current Signal Against time(s).	123
Figure 6.4 Andrews Plots using Motor Current: Medians and Quartiles Against time(s).	123
Figure 6.5: Dendrogram Displaying Feature Similarity (colour threshold T=6.1) Second Stage Vibration Envelope Feature. Clustering Based on Euclidean Distance between Pairwise Observations.....	125
Figure 6.6 Dendrogram of Motor Current Feature Clustering Based on Euclidean Distance Between Pairwise Observations.....	127
Figure 6.7 Dendrogram for Second Stage Pressure Envelope Features.	129

Figure 6.8 Scatter Plots Showing Classification Similarities for Homogeneous Harmonics 3 and 5 from the Second Stage Vibration Signal..... 131

Figure 6.9 Scatter Plot: All Cases Using Envelope Spectrum Features 4 and 7 from the Second Stage Vibration Signal..... 131

Figure 6.10 Comparison of Homogeneous and Heterogeneous Feature Pairings. 132

Figure 6.11 Two-Dimensional Cluster Plot Displaying the Distinct Groups of Healthy and ICL Using Envelope Features 4 and 6. (Second Stage Vibration Signal)..... 132

Figure 6.12 Discriminant analysis Using Features 4 and 7 To Separate the Healthy and Inter-Cooler Leak Classes..... 134

Figure 6.13 Successful Discrimination between the Healthy and ICL using Features 4 and 6. 134

Figure 6.14 Scatter Plot of All Five Classes using Features 4 and 6 Alone..... 135

Figure 6.15 Naïve Bayes Classification Tree Using Two Input Parameters (Features 4 and 6). 137

Figure 6.16 Second Stage Vibration, NB 10 Parameter Model With 82% Successful Classification Across All 5 Groups..... 140

Figure 6.17 Second Stage Vibration, NB 32 Parameter Model With 75% Successful Classification Across All 5 Groups..... 140

Figure 7.1 Fault Clustering Using the First Two Principal Components. 145

Figure 7.2 Fault Clustering Using the First Three Principal Components Accounting for 81% of the Total Variation..... 145

Figure 7.3 SVM Classification Linear Kernel Function ('Healthy' and 'Not healthy') Using Harmonics 4 and 7. 149

Figure 7.4 SVM Including Hard-Margin, Using Harmonics 4 and 7 ('Healthy' and 'Not Healthy'). 149

Figure 7.5 SVM Linear Kernel Using Harmonics 4 and 6 (SVL Identified). 150

Figure 7.6 SVM Including Hard-Margin Using Harmonics 4 and 6 (SVL Identified) 150

Figure 8.1 PCA Using Reduced Number of 10 Input Parameters Indicated by CA 153

Figure 8.2 First Stage Multiscale PCA Results..... 155

Figure 8.3 Improved Multiscale PCA Signal Comparisons..... 155

Figure 8.4 Distributional Comparison of 1. 4th Harmonic Amplitudes and 2. Compressed 4th Harmonic Amplitudes..... 157

Figure 8.5 Original Signal Envelope Harmonic 4..... 158

Figure 8.6 Compressed Signal Envelope Harmonic 4..... 158

Figure 8.7 Distributional Comparison of 1. 6th Harmonic Amplitudes and 2. Compressed 6th Harmonic Amplitudes..... 162

Figure 8.8 Comparison of Distributions of Compressed Signals 1. 4th Harmonic 2. 6th harmonic. 162

Figure 8.9 Scatter Plot by Class Using Compressed Signals..... 163

Figure 8.10 NB Classification for all 5 Classes using Two Compressed Harmonics (4 and 6).....	163
Figure 8.11 Andrews Plot using Compressed 4th and 6th Harmonics Only.....	164
Figure 8.12 Surface Plot of Original Envelope Harmonics.....	165
Figure 8.13 Wavelet Decomposition of Fault Profiles.....	165

Tables

Table 4.1 Two-stage Broom-Wade Reciprocating Compressor Specification.....	71
Table 4.2 Displays the Mean Revolutions Per Minute and Corresponding Standard Deviations for Each of the Three Sections.	82
Table 4.3 Peak Values of First and Second Stage Pressures.....	90
Table 5.1: Summary Statistics for First Stage Pressure Measurements (Load 100Psi=6.90Bar).	97
Table 5.2 Summary Statistics for First Stage Vibration (Load 100).....	100
Table 5.3 Summary of Frequency Spikes and Amplitudes by Fault Type,	104
Table 5.4 95% Confidence Intervals of the First Harmonic Amplitudes.....	105
Table 5.5 95% Confidence Intervals of Amplitude Spectra at Given Harmonics....	105
Table 5.6 Spectrum Amplitudes for the Second Stage Vibration Signal.....	107
Table 5.7 Faults Induced in Simulations.....	114
Table 5.8 Average Classification Rates.....	114
Table 5.9 Feature Usage by Class.	115
Table 5.10 Dominant Features Used in Classification.....	116
Table 5.11 Classification Rates for mRVM.....	117
Table 5.12 Classification Results using mRVM with Varying Numbers of Input Parameters and Classes.	118
Table 6.1 Second Stage Vibration Envelope Feature Cluster Groups (Group threshold set at 6.0 with sub-sets shown in brackets forming prior to this distance).....	126
Table 6.2 Motor Current Harmonic Feature Cluster Groupings.....	128
table 6.3 Classification steps using measurements on features 4 and 6.....	136
Table 6.4 Effect on Classification Rates of Inclusion of Harmonic 14	138
Table 6.5: Classification Rates for All 5 Classes Using the Second Stage Vibration Envelope Harmonics.	139
Table 6.6 Comparison of Classification by Class	139
Table 6.7 Classification Success Rates per Number of Groups and per Model.....	139
Table 7.1 Summary of Eigenvalues and PC Variance for the first three PCs.....	144
Table 7.2 Cumulative Sums for The First 14 Principal Components	144
Table 7.3 Class Scores on the First Two Principal Components.....	144
Table 7.4 Factor Loadings and Specific Variance for Key Harmonics.....	146
Table 8.1 One-way ANOVA 4 th Harmonic.	156
Table 8.2 One-way ANOVA 6 th Harmonic.	156
Table 8.3 Calculation of Test Sensitivity and Specificity.....	160

Abbreviations

ANN(s)	Artificial Neural Network(s)
ANOVA	Analysis of Variance
BDC	Bottom Dead Centre
CA	Cluster Analysis
CI	Confidence Interval
CM	Condition Monitoring
DA	Discriminant Analysis
DE	Differential Evolution
DFT	Discrete Fourier Transform
DPLS	Discriminant Partial Least Squares
DVL	Discharge Valve Leak
DWT	Discrete Wavelet Transform
FA	Factor Analysis
FDA	Fishers Discriminant Analysis
FFT(s)	Fast Fourier Transform(s)
FT	Fourier Transform
GA	Genetic Algorithms
H	Healthy
IAS	Instantaneous Angular Speed
ICL	Inter Cooler Leak
IDWT	Inverse Discrete Wavelet Transform
KPCA	Kernel Principal Component Analysis
LB	Loose (drive) Belt
LDA	Linear Discriminant Analysis
MANOVA	Multivariate Analysis of Variance
MEMS	Micro Electromechanical Systems

mRVM	Multiclass Relevance Vector Machines
MSPCA	Multi Scale Principal Component Analysis
NB	Naïve Bayes
OAA	One Against All
OAO	One Against One
PC(s)	Principal Component(s)
PCA	Principal Component Analysis
psi	Pounds per Square Inch
PSO	Particle Swarm Optimisation
RC(s)	Reciprocating Compressor(s)
RMS	Root Mean Square
rpm	Revolutions Per Minute
RVM	Relevance Vector Machines
St. Dev.	Standard Deviation
SVL	Suction Valve Leak
SVM(s)	Support Vector Machine(s)
TDC	Top Dead Centre
TLBO	Teaching-Learning Based Optimisation
WT	Wavelet Transform

Nomenclature

$X(f)$	FFT
f	Frequency
t	Time
N	Number of Samples
$x(t)$	Signal value at time t
X	Matrix of Observations (n observations by p variables)
$d(p, q)$	Euclidean Distance between points p and q
$D_M(\underline{x})$	Mahalanobis Distance
$\underline{x} = (x_1, x_2, \dots, x_n)^T$	Observation
$\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$	Population Mean Values
S	Data Covariance Matrix
Σ_0, Σ_1	Population Covariance Matrices
T	Threshold Value
c	Decision Criteria Threshold Value (homoscedacity)
d_{ij}	Pairwise Euclidean Difference between the i^{th} and j^{th} observations
$d_{(r,s)}$	Average linkage
$w(p) = (w_1, \dots, w_p)_{(p)}$	Weight or Loading Vectors
$z_{(i)} = (z_1, \dots, z_p)_{(i)}$	Vector of PC Scores
λ_i	The Variance of Z_i
C	The Covariance Matrix
X_i	The i^{th} standardised score,
a_i	The factor loading
F	The factor value

e_i	The portion of X_i specific to the i^{th} test only.
$\sum a_i^2$	The Communality of X_i
$k(x, y)$	Kernel Function
α_i	Hyperplane Parameters
D	Set of Training Data
φ	Kernel Function
$w \sim N(0, \alpha^{-1}I)$	The Weight Vector
$w = (w_1, w_2, \dots, w_m)^T$	The Prior Weights
α_j	The Variances of the Prior
x_1, \dots, x_n	The Training Set Input Vectors.
t_1, \dots, t_n	The Target Values
$\varphi(x) = (\varphi_1(x), \dots, \varphi_m(x))^T$	The Basis Functions
L	Wavelet Decomposition Level
A_L	Wavelet Approximation Models (m by $\frac{n}{2^L}$)
(D_1, \dots, D_L)	Wavelet Details
pL	Low Pressure (first stage of compression operation)
pH	High Pressure (second stage of compression)
vL	Vibration (first stage of operation)
vH	Vibration (second stage of operation)
ang	Angular speed of the crank shaft
ind	Index angular speed,.
i	Motor current,.
Pd	Load pressure in holding tank
Fs	Sampling Frequency,
θ	Crank Shaft Angle (degrees)
ω	Angular Velocity
fn	The n^{th} Envelope Harmonic Feature
R^2	Goodness of Fit Statistic

α	Significance Level (Type 1 Error)
β	Type 11 Error
$1-\alpha$	Specificity
$1-\beta$	Sensitivity or Power

Dedications and Acknowledgements

I wish to acknowledge the University of Huddersfield for facilitating my doctoral study. Especially my advisory team, Dr Fengshou Gu and Professor Andrew Ball, for stoic refusal to display the slightest wonder at my gargantuan achievements and perfect prose thus forcing measureable academic improvements.

Dedicated to Robert, Eddie and Daisy, my true inspiration. I trust my focus on doctoral research throughout your teenage years has equipped you with the fortitude to forge meaningful priorities in life.

Chapter 1

Introduction

This chapter outlines the research area including motivation, aims and objectives plus details of the thesis organisation.

An overview of Condition Monitoring is presented, its importance measured and the process of information collection and data manipulation summarised.

1.1 An Overview of Condition Monitoring

1.1.1 The Importance of Condition Monitoring

Condition monitoring (CM) is concerned with preventing, or at the very least predicting, impending component failure. The past fifty years have seen exponential growth in the use of CM in process control for an increasing number of applications. CM is largely similar to medical diagnosis in that symptoms are detected and tested with the aim of diagnosing the problem. Alongside more traditional scheduled maintenance, based on set periods of time or production outputs, modern process monitoring is complemented with maintenance on demand. Maintenance based on real time observations of operation with respect to expected normal running principles [28, 79 and 80]

The ultimate aim of CM is to prevent failure and unplanned or emergency intervention. Ideally interceding at the earliest possible onset or deviation from optimal operation. Thus efficiency of systems and components is enhanced through preventive and corrective maintenance supplemented with routine overviews and services [6 and 27].

With increasing process performance and product quality CM becomes ever more vital. Gathering up to the minute information on operating condition and process efficiency. Fault management through continuous supervision ensures deviations from normal healthy operation are detected practically at onset. Immediate corrective measures enhance the efficiency and lifetime of both systems and components [29].

Supervision of processes utilising fault detection methods and diagnostic fault classification or inference methods is designed to highlight anomalies in the current behaviour of a process compared to expectations [10 and 55]. Thus aiding timely intervention should non-normal output measures be recorded. Ideally through appropriate corrective actions undesirable outcomes such as component malfunctions and failures should be minimised if not entirely eradicated.

1.1.2 Information Gathering

As the complexity of machinery increases, both component and system wise, so do the complexity and costs of maintenance programmes. A reliable method of assessing health and monitoring performance is required. Incorporating sensors into complex systems to collect output signals provides up to the minute condition information. Thus

continual assessment of different components at all stages of operation is possible. Sensors may be temporarily affixed for scheduled assessment of processes or be permanently attached and continuously monitored [107]. Accelerometers, for example, externally attached to mechanical components are non-intrusive and provide continuous localised vibration measurements. Response times to identification of deviations from normality are timely thus immediate steps can be taken should a process be operating at less than optimal efficiency [16].

Classical methods of process monitoring, suitable for overall process management include checking measurable variables against pre-set tolerances. Alarms being triggered should the boundaries on a number of process control rules be violated [30, 78 and 100]. Alarms signifying appropriate intervention from operators unless a suspected dangerous state of the process is indicated when the necessary counter action or possible shutdown would be automated.

1.1.3 Data Manipulation

To locate and identify a specific fault it is necessary to convert collected signals into physical attributes and remove extraneous noise due to unrelated factors. Sophisticated modelling of signal patterns requires advanced data processing and multivariate statistical techniques. Analysis of data collected under known process conditions informs statistical models for future diagnostic use. Thus the major characteristics of faults can be determined and catalogued for comparison of current condition [18, 19 and 20].

CM covers a vast number of diverse applications, techniques and monitoring methods. One area of study being detection and identification of faults in mechanical industrial processes. Enhanced multivariate statistical techniques are applied to signals retrieved from various stages of process operation. Vibration signals are particularly useful in describing mechanical process state of health [55 and 87]. Advances in model sophistication are prevalent using pattern detection of frequencies or combinations of frequencies from spectra. Surges in computational capabilities have facilitated massive increases in data processing potential. A search of current literature, however, reveals little or no specific research into the underlying characteristics of the input variables.

1.1.4 Research Application

Reciprocating compressors (RC) are an intrinsic part of many industrial processes. For example, oil refineries, gas pipelines, chemical plants and refrigeration plants [1, 2 and 11]. Component failure is therefore potentially life threatening as well as costly and time consuming. Efficiency and continued performance of these processes rely on early detection of RC component deterioration. Component failure in RCs is broadly due either to mechanical failure or loss of elasticity in sealing components. The latter being the most prevalent. Monitoring RCs is difficult mainly due to inaccessibility of component parts hence the importance of performance monitoring through suitably positioned sensors for signal capture [54, 72 and 82]. Although informative internally inserted pressure sensors are intrusive whilst accelerometers can be externally attached without direct interference with operation. Vibration signals collected via accelerometers provide high levels of information facilitating detailed and accurate assessment of system condition. RCs are susceptible to a multitude of faults, both due to mechanical and elastic deterioration, which can occur in isolation or combination. Extracting useful information is often hindered by the large amount of noise captured along with measured signals. For meaningful analysis this extraneous variation needs to be filtered out [12, 16, 28 and 87]. In reciprocating machines the problem is all the more apparent than in rotating mechanical systems due to the greater vibration amplitudes and increased complexity of interacting component parts. Thus the potential complexity of the RC faults lends itself perfectly to investigation through complex multivariate statistical models and their testing. Techniques which are then easily transferable to other mechanical processes with less intricacy.

1.2 Motivation

The motivation arose during previous research detecting and diagnosing faults in RCs wherein classifiers were established using sophisticated modelling techniques [2]. For algorithmic convergence it was necessary to restrict the number of input parameters to as few as 15 hence decisions had to be made as to which parameters to include and which to omit [101 and 102]. Whilst it was apparent that certain parameters had particular relevance to a given class being repeatedly used in its presence, feature selection was more intuitive than rigorously quantitative. To optimise explanatory power of the model a rigorous selection procedure was necessary. Random selection of 15 variables from the possible 32 is clearly not a sensible proposition. Analysis of harmonic feature characteristics for homogeneity should allow a reduction in the number of input variables (harmonic features) through selection of a smaller number of heterogeneous harmonics and elimination of those with contributions to variability already accounted for by others. Thus by applying the underlying model building principles of sparse multivariate regression analysis, reduced feature sets with optimal explanatory powers should be realised [57, 61 and 73].

This research is primarily concerned with reducing the volume of input parameters prior to model construction whilst maintaining classification accuracy and avoiding bias. Methods of reducing input parameters are considered in three broad groups a) inputting fewer original variables identified through variable clustering to offer the highest levels of non-repeated information b) by using variable reduction techniques transforming all original variables into a reduced number of linear combinations of those original variables and c) a combination of the procedures detailed in a) and b). In addition the possibility of volume reduction through data compression is considered [77]. Throughout effectiveness of the selection criteria are illustrated by construction and evaluation of multivariate classifiers using both classical statistical and machine learning model building techniques.

1.3 Aim

To explore the impact on multivariate classifier efficacy of differing input parameter selection techniques. Introduce robust selection criteria through identifying variable properties and correlations. Maintain the highest possible levels of model efficiency, avoiding bias and reducing input parameter volume with application to reciprocating compressor fault identification.

1.4 Objectives

1.4.1 Time Domain

To examine the characteristics of output signals from a reciprocating compressor rig to identify their physical attributes under normal, healthy, operating conditions.

To explore inter-relationships between signal measurements during healthy operations in the time domain.

To identify the most richly informative, non-intrusive output measurements with respect to their potential for explaining operating condition.

1.4.2 Frequency Domain

To convert output signals from the time to the frequency domain and examine characteristics of their frequency spectra.

To de-noise output signals using fast Fourier transforms and identify salient features with respect to size and position of harmonic amplitudes.

To utilise envelope spectra harmonic amplitudes to monitor condition of a process, detect and identify any faulty components present.

1.4.3 Variable Selection

To identify plausible class separation through class profiling Fourier transforms of output signals.

To identify the output signal with optimal powers of explanation and minimal intrusive effect on the operating system.

To identify homogeneous groups of envelope harmonics through variable clustering in order to select reduced numbers of input parameters with optimal explanatory power.

1.4.4 Classifier Construction

To select representative group members, from each of the heterogeneous cluster groups, for inclusion as input parameters in multivariate statistical modelling.

To establish models, or classifiers, to predict state of health of reciprocating compressors using reduced numbers of input parameters whilst maintaining classification accuracy. Also to investigate effect on model accuracy of varying numbers of input parameters in detecting and identifying increased numbers of faults.

To establish multivariate classifiers through variable reduction techniques from classic multivariate statistical and machine learning methodologies.

1.4.5 Extended Analysis

To consider the impact of data compression on input parameter distributions, and possible effect on richness of signal information and ability to identify and classify faults.

1.5 Organisation of Thesis

This thesis is organised into nine chapters with the data analysis covering four main sections of investigation:

- Chapter 5: Exploratory data analysis in both the time and frequency domains.
- Chapter 6: Investigations into correlations and inter-dependencies, of the envelope spectra harmonics, to establish variable groupings.
- Chapter 7: Classifier construction using the two proposed methodologies, reducing the number of input parameters and variable reduction techniques.
- Chapter 8: Combined methodologies incorporating data compression and identification of further work.

Chapter 1 introduces the thesis and outlines the research area including the motivation for the work studied, the overall research aim and research objectives. This is extended in chapter 2 where the background to existing condition monitoring methodology from a variety of disciplines is explored giving a brief overview of each area. Also further investigates use of vibration signals in detection of fault occurrence in mechanical processes.

Chapter 3 gives a detailed account of the theory and applications of the multivariate statistical techniques employed in the analysis in subsequent chapters of the thesis. Their implementation in Matlab and possible adaptation to the field of CM.

The engineering application of reciprocating compressors is reviewed in Chapter 4, including differing types of compressor, their basic functions and purpose, common faults and their implications. The test facilities and data collection process are also described with respect to the test rig employed, the output signals captured and the seeded faults to be identified.

The first data analysis section begins with Chapter 5 which contains the exploratory data analysis, examination of individual signal outputs and characteristics in both the time and the frequency domain. Inter-relationships between the output signals measured and some preliminary investigations into fault diagnosis in the frequency domain.

Chapter 6 heralds the start of the novel investigations and the second stage of the data analysis. Initially inspecting envelope harmonics in the frequency domain for key fault characteristics through class profiling using Fourier transforms with results displayed in the form of Andrews' plots to determine potential for class separation. There follows a pre assessment of the variable characteristics through variable clustering from which heterogeneous representatives are selected for incorporating in the modelling process.

Illustrative models are examined from both classical multivariate statistical methods and machine learning technologies in the third analysis stage. Classifiers are constructed and assessed using discriminant analysis and Naïve Bayes methods. The focus of the analysis in this chapter being the construction of multivariate classifiers for fault detection using reduced numbers of input parameters, identified through prior variable clustering, whilst maintaining fault classification accuracy. Modelling continues in chapter 7 which investigates the efficacy of multivariate classifiers constructed using variable reduction methods. Again examples are taken from both multivariate statistics and machine learning methodologies, namely principal component analysis and support vector machines. Once again the focus of the chapter being the ability of the input parameter selection technique to facilitate construction of

models with high classification capabilities and low computational output requirements.

Combined methodology is considered in chapter 8 where the concept of data compression is also introduced. Further discussion is included of the potential for data compression and a combination of the techniques explored in chapters 6 and 7 to produce optimally sparse unbiased multivariate classifiers with reduced input parameter volume. This chapter forms the basis for identification of further research the primary focus being the construction of classifiers using the conjoined parameter selection techniques.

Chapter 9 addresses the overall research conclusions, reviews the thesis achievements and objectives and details the thesis contributions to knowledge and novelties.

Chapter 2

Current Practice in Condition Monitoring

This chapter reviews CM from its onset in the 1960s to current day and examines the CM spectrum. In particular the CM of mechanical processes and the use of vibration signals in fault detection analysis. A review of the various types of compressor systems and their typical applications is presented. Also a brief overview of maintenance protocols, historical and emergent. Potential operational problems, major sources of poor performance or breakdown and impact on all aspects with respect to cost and efficiency. Prospect for further faults to develop due to worn or broken parts, increases in running costs and consequences of downtime due to major incidents.

2.1 The Growth of Condition Monitoring Practice

With computer technology becoming readily available throughout the 1960s and being cheap in comparison to salaries, computer monitoring of systems was introduced and greatly increased during the 1970s. This was further enhanced and developed in the 1990s to include diagnostic testing of systems. From then condition monitoring of systems and components has become the norm [80 and 107]. As process automation or condition monitoring expanded so did the demand for increasing process performance and product quality. Human operators were thus released from monotonous and arduous tasks [55]. This automation expansion is also evident in areas of technical production and precision mechanical devices which operate partially automatically or with integrated automatic functions. Thus an industrial process can be continually assessed with relevant and timely intervention should pre-determined limits or thresholds be violated. Maintaining quality of outputs and machine performance. However, whilst these methods provide a protective environment for operation, in particular for safety related issues, greater damage due to major faults may be detected too late. Enabling early diagnosis of the root cause before extensive damage and deterioration in performance is evident is essential [80, 87, 96 and 107].

2.1.1 The Condition Monitoring Procedure

Monitoring of processes is conducted on three sub-levels of information processing; control level (feedback control), supervision level (observation and supervision of process) and higher level (management). Thus scheduled maintenance, based on set periods of time or production outputs, is complemented with maintenance on demand, based on real time observations of operation with respect to expected normal running principles [96 and 107].

Fault management through continuous supervision and diagnosis with performance and lifetime prognosis falls into three major categories maintenance, repair and reconfiguration with the first two being either planned or unplanned emergency maintenance [100]. The ultimate aim of condition monitoring is to prevent failure and unplanned or emergency intervention. To intercede at the earliest possible onset or deviation from optimal operation thus enhancing efficiency and lifetimes of systems and components. This is achieved through preventive and corrective maintenance supplemented with routine overviews and services [6].

Naturally as the complexity of a machine increases so its expected maintenance costs and monitoring programme increase in size and complexity. A reliable method of assessing health and monitoring performance is thus sought to enable the machine lifetime to be prolonged to a maximum whilst maintaining an optimal operative state. De Botton et al (2000) describe the four-fold benefit of improving productivity, product quality, profitability and overall effectiveness of a system.

2.1.2 The Condition Monitoring Cycle

The condition monitoring cycle of detection of abnormalities, diagnosis or classification of faults and prognosis is continual and ideally an intrinsic part of the maintenance cycle. Timely intervention through perpetual signal measurement and data analysis complementing time based maintenance schedules [30, 56 and 85].

The major focus of condition monitoring being its prognostic and fault prediction abilities, the power of which is determined by selecting the appropriate analytical techniques. Making correct distributional and theoretical assumptions is key to model robustness and strength of inference. Likewise a condition based rather than time interval based maintenance regime ensures near optimal performance for the duration of process operation. An ever increasing drive for environmental supremacy and the highest safety standards alongside increasing global competition forces companies to strive for the greatest possible achievements in both performance and quality [63 and 76].

2.1.3 Data Collection for Informed Decisions

Decisions are made daily about manufacturing processes in the presence of variability. Process specific knowledge allows quantification of the risks associated with various courses of action while statistics provides a common language to communicate information. A process is considered to be in a state of statistical control if any variations in observed measurements can be attributed to chance variation only not assignable variability which it is feasible to detect and identify, for example a loose belt, leaky valve etc. [78]. Thus a process under statistical control operating under the influence of common causes of variation only should be operating within the upper and lower control limits. Consequently in optimising process performance in all

aspects it is essential that deviations from the norm and or mechanical faults are detected, quantified and corrected in a timely manner [29].

Data extracted via appropriately positioned sensors which capture output signals in a non-intrusive manner give relevant and timely information on the state of the observed process. Hence the current condition of a system can be assessed, failures predicted and remaining operating times estimated where applicable. Through suitably established statistical models condition predictors and reliability prognosis in systems is possible and of vital importance in maximising all aspects of efficiency [6, 55, 85 and 107].

Supervision of processes utilising fault detection methods and diagnostic fault classification or inference methods is designed to highlight anomalies in the current behaviour of a process compared to expectations. Thus aiding timely intervention should non-normal output measures be recorded. Ideally through appropriate corrective actions undesirable outcomes such as component malfunctions and failures should be minimised if not entirely eradicated [100].

2.1.4 Summary

Optimisation goals can only be achieved if the right data is collected and appropriately analysed in sufficient time to enable corrective action to be taken. Data relevant to the analytic means and pertinent to the point of performance deviation. Increasingly sophisticated methods of analysis are being developed to cope both with growth in data collection and need for the upmost precision of detection and classification techniques. Generally these methodologies originate from either multivariate statistical techniques or from the machine learning environment. Technologies are more often than not application specific. Particular signals also lend themselves to certain areas of investigation more than others. Vibration signals, for example, tend to be poor indicators when applied to electrical systems as the system wiring interferes with output data; whereas they form the basis of much reliable diagnostic testing of mechanical processes.

2.2 Condition Monitoring Recent Research

2.2.1 Breadth of Condition Monitoring Methodologies

In [10] a review of 177 CM articles is presented classified by techniques and monitoring methods. Over the decade studied it was found approximately 20% were employing neural network techniques with around 12% of all studies analysing vibration data. These proportions are roughly similar today although 'wavelet transform techniques' might be a more appropriate collective noun than neural networks.

Vibration signals can be collected via accelerometers in a non-intrusive manner, the signal amplitudes of the fast Fourier transforms are widely recognised as efficient pattern recognition features [12, 16 and 77]. Main areas of application being mechanical driven machinery as smaller electrical applications with additional wiring and transducers interfere with signal outputs. Major strengths of vibration analysis being its use for permanent as well as intermittent monitoring along with the high likelihoods of identifying specific faulty components [99]. Technological signal capture is further advanced through the use of micro electromechanical systems (MEMS) accelerometers directly attached to reciprocating compressor (RC) rotors [39]. Similarly motor current signature analysis which combines voltage and current outputs is a quick response and relatively cheap technique although is limited by equipment damage at low loads.

Off and on-line tests are relatively popular although off-line tests are very time consuming and subject to untypical stresses components being detached from their usual supply. In addition to the long lead time and the actual fault and detection tests occurring almost at shutdown. On-line testing on the other hand whilst maintaining attachment to relevant components suffers performance decline throughout the testing period [12].

Temperature monitoring via internal motor sensors is particularly productive from an information point of view, offering immediate results without the need for complex analysis. Applications include bearing related mechanical faults and defects of systems deprived of air or water cooling, however, the extremely high cost of thermal imaging cameras and related software is prohibitive.

Mathematical techniques offer faster monitoring and greater precision hence the continual growth in available techniques. Principal component analysis (PCA) was found to be most frequently used in application to chemical processes [7] with wavelet transforms rapidly increasing in importance as a signal processing tool. Wavelet

transform coefficients being utilised in the form of decomposed acoustic signals to monitor tool wear [111]. It is claimed in [45] that wavelet transform features offer faster, more reliable and sensitive means of monitoring than Fourier analysis. However, much research conducted utilising FFTs would contradict this claim giving equally precise and accurate diagnoses [1, 8, 21, 32, 35 and 46].

Neural networks remain the most oft used approach to system monitoring employed increasingly often alongside fuzzy logic or genetic algorithms (GA) for more complex fault diagnosis. Incorporating neural networks with 2 and 3-D graphs it is claimed in [59] leads to greater visualisation although essentially the characteristics and fundamental properties of the input variables remain hidden. GA are self-learning tools applied to training data to assign each case to one of a number of pre-determined sets related to normal operation and a number of faults. They may be incorporated into knowledge based or expert systems which do not require prior knowledge of the process but can be prone to overfitting. GA and artificial neural networks (ANNs) were combined in the analysis of vibration data, specifically employing GA to automate the fault classification process and minimise human interaction [70]. It was felt by the authors in [67] many models were not addressing specific problems being based on broad assumptions so failing to accurately identify dynamic characteristics.

Models derived from multivariate data alone are prone to limitations if the faults are not known a priori and as the quantities of data and complexities of analysis continue to grow it is argued [22]. Solutions in the form of multi-dimensional path visualisations are offered in [24 and 43]. Acoustic signals from RCs are analysed using wavelet techniques [103 and 104] whilst wave matching differential evolution (DE) algorithms, which are akin to GA for feature extraction, are used prior to classification with support vector machines (SVM) in [83]. The authors in [83] arguing SVM have better generalisation than ANN especially for small sample problems. Limitations of these studies being a priori knowledge of the working principle to extract waveform features and the potential for hidden extraneous vibrations due to external effects. Findings in [104] concur, the authors claiming faster training and greater classification accuracy. However, in [63] it is noted that whilst data mining, ANN, SVM and other wavelet CM approaches might produce highly laudable results in a controlled experimental environment they show great weakness when applied in the real environment especially in monitoring wind turbulence effects. Perhaps confirmation that analysis via 'black box' techniques with no direct relation to the original signals and extraneous factors produces less than optimal solutions.

2.2.2 Increasing and Hidden Model Complexity

Enhanced CM procedures monitor increasingly large numbers of faults for ever increasing volumes of data. Specific details of analysis though remain largely hidden within computerised learning technology. In addition there is a lack of evidence describing the nature and characteristics of the original input parameters. Ensemble clustering entails merging input cluster groups from a given data set to generate an improved final clustering fit [117]. However, the focus is squarely on gaining enhanced cluster separation rather than identifying pre-determined clusters. Where clusters are fixed such as RC fault types this methodology offers no solution. There is no apparent research analysis on associations between original variables with clustering very much a post classification procedure especially popular in process control. SVM and NB analyses are conducted employing random forests post hoc in [31]. Random forests provide grouping akin to clustering. Also correlation proximity is frequently employed in spectral clustering of samples post classification, the clustering parameters determined by trial and error with similar outcome to SVM. Another study, [71], highlights the inherent difficulties of on-line control systems violating computational capabilities in time constrained systems with large numbers of input parameters. Whilst there is much research into the value of particular output data measurements, vibration analysis, current signature analysis, temperature analysis etc. investigation of individual parameters within particular measurements is not apparent.

2.2.3 Selection of Optimum Input Parameter Sets

The problem in machine parameter setting for optimal outputs is considered in [86]. Present day machine process monitoring involves data sets with large numbers of input parameters. Identifying which factors will most affect the cost and quality of the products is key. Selection of optimum parameter sets for analysis is of paramount importance to satisfy all the conflicting objectives of the process. Due to the necessity to restrict numbers of input parameters especially for large data sets a well-defined and rigorous selection process is required. An artificial intelligence method is proposed and claimed superior to GA, Particle Swarm Optimisation (PSO) and other prevalent algorithms [86], again there is no report on properties of the original variables. The proposed solution, a Teaching-Learning Based Optimisation (TLBO) algorithm. Input parameter settings from the n^{th} iteration are assessed by outcome measurements of quality aspects and thus inform levels set for the $(n+1)^{th}$ iteration. Algorithmic

approaches to input parameter selection in applications to classifier construction are similarly evolved in [2, 49 and 70].

Computational run-time is considerably increased for large data sets and to achieve algorithmic convergence as few as ten input parameters might be an absolute maximum [2]. Selection based on original variable properties either in place of or prior to incorporating an algorithmic application is proposed as an area not currently explored.

Likewise much research is published regarding the pros and cons of analysis in the time or frequency domain but individual variable characteristics are not identified.

Vibration signals are optimally informative when monitoring mechanical industrial processes. In addition signals transformed to the frequency domain have superior deterministic properties over their time domain equivalents. With envelope spectra features highlighting salient signal features and removing extraneous interference. Envelope spectra harmonic amplitudes will form the basis of the novel investigations of this thesis.

2.3 Envelope Spectra and Fast Fourier Analysis

The basic concept of signal analysis is to reduce a complex signal in the time domain to its component parts in the frequency domain. Thus the salient features of the signal become apparent and confusion due to noise is removed.

It is a common practice in condition monitoring to perform analysis in the frequency domain [4, 6, 12, 16, 27 and 46]. Of particular interest, [1, 2, 32, 33, 34, 35 and 39] used spectrum analysis to study vibrations from compressor rigs to detect and diagnose different faults.

By applying the Fourier transform (FT) to a measured signal $x(t)$, such as the vibration from a reciprocating compressor, the repetitive pattern hidden in the data can be highlighted with the Fourier coefficients which allows the key features of vibration to be recognised easily.

The spectrum is the FT coefficients at corresponding frequencies. To find the frequency spectrum of such signals the fast Fourier transform (FFT) is the most efficient calculation method.

The FFT, $X(f)$ a function of the frequency, f , a continuous function of time, t , is given by:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-2\pi jft} dt \quad (2.1)$$

For digital signals, the discrete Fourier transform (DFT) gives a numerical approximation and is widely used in Engineering.

$$X(t) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-j(2\pi kt/N)} \quad (2.2)$$

Where $t = 0, 1, 2, \dots, (N-1)$, N is the number of samples taken, $x(t)$ the value of the signal at time t and k the current frequency, 0 to $(N-1)$ Hz.

Finding the frequency spectrum provides valuable information about the underlying frequency characteristics of signal outputs and so is useful in defining condition characteristics of a system. Hilbert transforms were employed [45, 46 and 104] or wavelet transforms [59 and 111] to similar effect.

Prior research [2] has shown that features extracted from envelope spectra in the frequency domain have superior deterministic properties over their time domain equivalents in monitoring the condition of RCs. Envelope spectra harmonics exhibit a number of discrete components mainly due to the fundamental frequency of the system and its associated harmonics. For the experimental compressor rig employed this fundamental frequency is 7.3Hz, the shaft rotation frequency.

Envelope spectra show only the amplitude profile of original signals and so provide a clearer insight into the underlying behaviour. Signal variations due to noise are filtered out leaving variation due to machine health only. Once the problem becomes too serious nothing is detected but noise, past the point of useful demodulation.

2.4 Condition Monitoring of Reciprocating Compressors

Due to their prevalence and importance in industrial processes there is naturally much research into detection and diagnosis of RC faults [103 and 104] to name but two. The remaining sections of this chapter give a brief overview of the types of RC commonly encountered, the RC operating cycle and common faults associated with RCs.

2.4.1 Types of Compressor

There are three basic types of air compressors namely reciprocating, rotary screw and rotary centrifugal. These types are further classified by the number of compression stages, the cooling method (air, water or oil), type of drive used (electrical motor, engine, steam etc.) and whether lubricated or not (i.e. whether lubricating oil is in contact with the compressed air). A reciprocating compressor increases the pressure in a quantity of air by reducing its volume. This is achieved by a piston encased in a cylinder, the piston compressing and displacing the intake of air [11 and 72].

A single-stage reciprocating compressor is typically used to achieve pressures in the range of 70 to 100psi (4.83 to 6.90bar) whereas a two-stage compressor unit is capable of achieving higher pressures in the range of 100 to 250psi (6.90 to 17.24bar), these higher pressures being achieved by a second piston and cylinder which acts on the air already compressed at the first stage. The rig's simplicity is its strength, weaknesses being the propensity for leaky valves and, being externally cooled, damage due to the interior working parts being subject to extreme operating temperatures [11 and 85].

Rotary air compressors are another positive displacement compressor type with the most common being a single stage spiral lobe oil flooded air compressor consisting of two rotors which compress the air within a casing. A valve less and oil cooled system, working parts are not subjected to extreme temperatures and associated problems. Easy to maintain and operate, smooth pulse-free air outputs of compact size are produced over a long life [11 and 85].

Centrifugal compressors are dynamic and depend on energy transfer from a rotating impeller to the air. High speed rotation is necessary for efficient operation and so this machine type is designed for high capacity and compression of a continuous air flow. The flow, hence capacity, can be regulated by adjusting the inlet guide vanes. An oil free compressor with the lubricated driving gear sealed off from the air and vents [72].

2.4.2 Compression and Expansion Cycle

The reciprocating compressor system is a relatively simple mechanical system which works on a gas, air for example, in order to increase its pressure. Capable of producing high pressures under variable loadings it is widely used [11, 32 and 72].

Air is drawn into stage one through the suction valve and compressed by the piston moving upwards in the cylinder. As the pressure of the air inside the cylinder reaches a critical point the discharge valve is activated and the compressed air is forced out of the cylinder either to a storage tank or onto a second stage for further compression. Once in top dead centre (TDC) position the piston ceases its upwards motion and reverses its direction down the cylinder, the compressed air having been discharged the discharge valve closes [11].

The suction valve opens as the piston travels downwards drawing in more air for compression. Note there is a small cavity between the TDC piston position and the top of the cylinder which houses the valves.

Thus the piston returns to bottom dead centre (BDC) position, the cylinder is again full of non-compressed gas, again the suction valve closes and the cycle is repeated.

Each of the valves allow one way air flow only i.e. the suction valve allows flow into the cylinder only and the discharge valve allows flow out of the cylinder only.

Industrial processes generally require pressures in excess of 3.5bar which is about the limit of a single stage compressor's capabilities thus multiple stage compression systems are the norm. In a two-stage compression system the air at atmospheric pressure is taken into the first stage cylinder and compressed to an intermediate pressure. On discharge the partially compressed air passes through an inter-cooler unit to reduce its temperature before being passed through the intake valve of a second cylinder. At the second stage the air is further compressed hence its pressure further increased after which it is passed in its final pressurised state to a storage tank. The second stage is mechanically identical to the first stage although second stage cylinders are smaller in order to achieve the desired pressure differentials [11 and 72].

Multi-stage air compressors are an obvious extension with multiple cylinders wherein the cooling process becomes increasingly vital and so more sophisticated [11, 32, and 72].

2.4.3 Faults and Implications

There are two major groups of faults the first being failure of mechanical moving parts either through wear due to friction, vibration heat or a combination (for example valves, springs, pistons and piston rings) or parts of the lubrications circuit such as the bearings, crankshafts or connection rods. The second fault group being due to loss of elasticity in sealing components leading to air leaks in the compressor itself or its connective piping.

Approximately 30% of faults are attributed to valve failures and a further 20% to piston ring failure [11]. Hence the focus of this research on valve leakage and related faults.

2.4.4 Applications of Reciprocating Compressors

Reciprocating compressors are an intrinsic part of many industrial processes whose performance and efficiency rely on early detection of RC component deterioration, for example, oil refineries, gas pipelines, chemical plants, natural gas processing plants or refrigeration plants. Consequently component failure or process shutdown is potentially life threatening as well as time consuming and costly.

2.5 Summary

Many advances have been made in the field of process CM over the past fifty years. With the enormous growth in use and power of computers information processing capabilities have vastly increased. However, as demand further increases for greater precision and more immediate responses, the burden in computational time especially for big data sets is becoming unattainable.

Whilst there is much research into modelling classifiers, using both classical statistical methodology and machine learning techniques, there is little evidence of focus on input parameter quality. Computational time saving is much discussed alongside established variable reduction methodologies. Although it is abundantly clear most algorithms have limited process capacity pre selection of input variables is not fully explored. Reduction in complexity being achieved by within model manipulation of

original input variables without clear data properties being identified or preserved. It is the belief of the author that greater explanatory power is achievable if prior inspection of input parameters informs pre modelling variable selection. Whilst also keeping input parameter volume to a minimum.

Theoretical proposals are to be applied to data from an experimental compressor rig. The major considerations being ability to detect the presence of faults and subsequently identify them. The primary research focus being to achieve reduction in input parameter volume. Experimental efficacy demonstrated by the ability of methodological approaches to correctly identify machine condition. Classical statistical methods and specifically developed machine learning techniques are evaluated throughout with all considered input parameter selection criteria. Refinements to current state of the art modelling techniques are thus assessed across both disciplines with respect to classification accuracy. As all these techniques are reliant on advanced multivariate statistical methodology further discussion and specific explanation is developed on the following pages.

Chapter 3, therefore, extends the literature review with a focus on specific mathematical models, their application and potential benefit in CM fault detection and classification for RCs. Also detailing data presentation and manipulation in the MATLAB computer package. To facilitate analysis conducted using the MATLAB computational package data was formatted into structured arrays and meaningful class labelling vectors.

Chapter 3

Multivariate Statistical Techniques

This chapter gives the relevant theoretical background of typical multivariate statistical techniques alongside references for further reading. Each technique is defined then reviewed and its current practice considered with relevance to RC fault analysis. Specifically, its implementation and possible improvements in the field of CM are assessed in the MATLAB platform, specific codes for methods are included in Appendix 2.

3.1 Cluster Analysis

Cluster analysis (CA) creates groups or clusters of data or variables. Clusters are formed in such a way that objects in the same cluster are very similar and objects in different clusters are distinct. Measures of similarity are application dependent and include Euclidean distance and Mahalanobis distance [20, 37 and 42].

Euclidean distance is the Pythagorean metric, the straight-line distance between two points in Euclidean space.

The Euclidean distance, $d(p, q)$, between the points, p and q , being given by

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.1)$$

Mahalanobis distance measures the proximity of a point p to a distribution or cluster mean. A multivariate generalisation of the normal distribution measuring the number of standard deviations point p is from the mean of the cluster. A distance of zero is given if p is at the cluster mean and increases as p moves from the mean along each principal component axis. It should be noted that scaling the axes to unit variance equates the Mahalanobis distance to the standard Euclidean distance. Mahalanobis distance is a function of the data correlations and is scale invariant [20].

The Mahalanobis distance, $D_M(\underline{x})$, of an observation $\underline{x} = (x_1, x_2, \dots, x_n)^T$ from a set of observations with mean $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ and covariance matrix S is defined as

$$D_M(\underline{x}) = \sqrt{(\underline{x} - \underline{\mu})^T S^{-1} (\underline{x} - \underline{\mu})} \quad (3.2)$$

Whilst there are many different CA algorithms there are two main groups, agglomeration techniques whereby all objects start as individuals and are systematically joined until all belong to a common group and division, the reverse process, whereby all objects originate from a single group. Both agglomeration and

division are hierarchic methods which directly facilitate generation of a dendrogram for easy visual group identification [20 and 73]. Hierarchical Clustering groups elements (data or variables) over a variety of scales the resulting cluster tree or dendrogram is not a single set of clusters, but rather a multilevel hierarchy. Applications generally dictate the most appropriate proximity measure. Similarity being assessed either on an individual to individual basis or by comparison of individuals to a group statistic.

Pairwise Euclidean difference d_{ij} , between the i^{th} and j^{th} observations, is given by

$$d_{ij}^2 = (x_i - x_j)(x_i - x_j)' \quad (3.3)$$

A square matrix of order m is generated with each entry (i, j) being the Euclidean distance between the observations i and j . From this an agglomerative hierarchical cluster tree is created using an appropriate linkage method. Agglomerative CA being to select first the two points which are most like i.e. ‘the nearest neighbours’ whereas the division algorithm would first select the ‘farthest neighbours’. Distances calculated according to the linkage method employed; simply the distance between individual data points in the case of single linkage methods; the distance between an individual and the group average for average linkage methods [20 and 73].

Average linkage, $d_{(r,s)}$, being calculated from the average distance between all pairs of objects in any two clusters as given in Equation 3.4.

$$d_{(r,s)} = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \quad (3.4)$$

Where n_r is the number of objects in cluster r , x_{ri} the i^{th} object in cluster r and x_{sj} the j^{th} object in cluster s .

3.1.1 Review and Current Practice

Thus prior clustering of variables is beneficial in determining variable likenesses. Once clustered by similarity it is apparent which variables are homogeneous thus may ‘explain’ the same variation and which are heterogeneous so would be expected to

uniquely account for model variation. CA provides a method of sorting variables into groups with uniform characteristics hence aids input parameter selection. Clustering is currently only regularly used for post hoc classification analysis predominantly in spectral group clustering [31].

It is a novelty of this research that CA is used a priori to determine input parameter characteristics and so inform variable selection pre model construction. Extensions of use in CM for prior evaluation of potential input variables giving a clear indication of variable associations and potential redundancy. Visual output in form of the dendrogram gives an informative visual display showing proximities and interconnections between variables [20 and 97].

Whilst Mahalanobis distance is scale invariant and would be particularly useful should cross signal comparisons be envisaged standard Euclidean distance was used having physically sensible interpretations. Agglomeration was also employed the primary objective being to identify harmonic groups.

Subsequently selected variables were used in constructing classification models to detect abnormal behaviour and identify specific faults.

Models of data with a categorical response, for example the differing states of health in a compressor rig, are known as classifiers. A classifier is built from training data, for which classifications are known. On testing, and in practice, the classifier assigns new test data to one of the categorical levels of the response. Experimentally this is achieved by randomly partitioning data into a training group and one or more test groups. Fundamental to effective model building is the selection of a representative feature set which performs optimally across all cases. The following sections 3.3 to 3.7 and 3.9 describe classifiers from both the multivariate statistical and machine learning fields.

3.2 Discriminant Analysis

Linear discriminant analysis (LDA) is a statistical method used in pattern recognition and machine learning whereby a linear combination of characteristic features is established with the aim of separating two or more classes or events [20 and 73].

Categorical dependent variables are predicted by their scores on a discriminant function established using one or more continuous or binary independent variables.

Originally developed by Fisher in 1936 [41], unlike cluster analysis, discriminant analysis is utilised when groups are known a priori. The LDA technique is very similar to logistic regression or relevance vector machines (RVM) which attempt to express a categorical dependent variable as a linear combination of other measurements [42]. Similar also to analysis of variance (ANOVA) in its attempt to classify cases to known groups except that ANOVA uses a linear combination of categorical independent variables to classify a continuous dependent variable. Should it be unreasonable to assume the independent variables are normally distributed, a necessary assumption for robust LDA modelling, logistic regression offers an alternative method of predicting categorical responses from continuous independent variables. LDA is also similar to principal component analysis [20 and 42] in that a linear combination of continuous variables is sought to best describe the data, however, whilst LDA explicitly models the class differences PCA doesn't take class differences into account.

Discriminant analysis uses training data to estimate the parameters of discriminant functions of the predictor variables. Discriminant functions determine boundaries in predictor space between various classes. The resulting classifier discriminates among the classes i.e. the categorical levels of the response, for example the machine states, 'Healthy', 'ICL' etc. based on the predictor data [20, 37, 42, 57 and 89].

Given a set of observations, \vec{x} , (features, attributes, variables or measurements) on each sample of an event with known class, y , in the training data set; the aim is to identify a good predictor for the class, y , from any similar sample, not necessarily belonging to the training data set, given any observation x .

For the two class case an observation with log likelihood ratio greater than a threshold T is predicted to belong to the first class, Equation 3.5, assuming the conditional probability density functions $p(\vec{x}/y=0)$ and $p(\vec{x}/y=1)$ are normally distributed with means μ_0, μ_1 and covariances Σ_0, Σ_1 respectively. Observations being predicted to belong to the second class if the log of the likelihood ratios are below the threshold, T , resulting in the quadratic discriminant classifier.

$$(\vec{x} - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x} - \vec{\mu}_0) + \ln |\Sigma_0| - (\vec{x} - \vec{\mu}_1)^T \Sigma_1^{-1} (\vec{x} - \vec{\mu}_1) + \ln |\Sigma_1| > T \quad (3.5)$$

If homoscedacity can be assumed then the class covariances can be assumed equal, $\Sigma_0 = \Sigma_1 = \Sigma$ the covariance matrices have full rank and Equation 3.5 simplifies to the decision criterion being based on the dot product

$$\vec{w} \cdot \vec{x} > c$$

for some threshold constant c where:

$$\begin{aligned} \vec{w} &= \sum^{-1} (\vec{\mu}_1 - \vec{\mu}_0) \\ c &= \frac{1}{2} (\vec{\mu}_0^T \sum_0^{-1} \vec{\mu}_0 + \vec{\mu}_1^T \sum_1^{-1} \vec{\mu}_1) \end{aligned} \quad (3.6)$$

Thus the model is a function of a linear combination of the known observations [20, 41, 57, 73, 89 and 91].

For more than two classes the analysis can be extended to defining sub-spaces which appear to contain all the elements of a given class. In practice the class means and covariances are not known but are estimated from the training data set using either the maximum likelihood estimate or maximum a posteriori estimates. Consideration should also be given to the ratio of sample measurements per class to the number of samples in each class, should the former exceed the latter the covariance matrices will be invertible not having full rank hence pseudo inverses or shrinkage estimation would be necessary. Linearity assumptions may also be unreasonable and a kernel function mapping might be more appropriate. Kernel functions are further discussed in the following sections [20, 38, 42 and 61].

3.2.1 Review and Current Practice

Although DA has most commonly been used in its linear form for distinguishing between two groups extensions and generalisations are emergent. Incorporating a non-linear kernel function and more than two groups in small scale categorisation problems has been beneficial particularly where groups are overlapping. [68] used acoustic emissions and LDA to detect bearing faults with very high speed accurate results. Several others have investigated improvements to small-scale categorisation problems using LDA and discriminant partial least squares (DPLS). Fishers DA (FDA) was developed in [21] with DPLS and reported improved classification results over PCA. Most recently [38] proposed a large scale approximation to kernel LDA, individualised learning, in which each test sample is directly compared to all

classification samples. Essentially a one-against-one algorithm which produces large numbers of simple classification problems to solve. A technique more suited to discriminating between two plus groups is Naïve Bayes classification which is discussed in the following section.

3.3 Naïve Bayes Classification

Naïve Bayes is a relatively simple technique for constructing classifiers. Whilst based on Bayes conditional probability it is not strictly speaking a Bayesian statistical method. The Naïve Bayes classifier assumes features are independent within a class although good results are achieved even when the independence assumption is violated. Data is partitioned into training samples and prediction samples. A model is then established using the known classes for the training set and this is applied to the predictor data to ascertain efficacy. Posterior probabilities for each sample dictate group classifications [19 and 91].

The class-conditional independence assumption greatly simplifies the training step allowing individual estimation of the class-conditional density for each feature. Thus the Naïve Bayes classifier can better estimate accurate classification parameters. Deviations from the independence assumption have been shown to have little detrimental effect. Training time is much reduced hence Naïve Bayes offers an advantage where datasets have many parameters or features as for the compressor rig. Variables are assumed to be mutually independent.

Classification is based on estimating the conditional probability $p(C_k / x_1, \dots, x_n)$ for n independent variables or features $\underline{x} = (x_1, \dots, x_n)$

$$p(C_k / \underline{x}) = \frac{p(C_k) p(\underline{x} / C_k)}{p(\underline{x})} \quad (3.7)$$

Since the evidence, $z = p(\underline{x})$, is not dependent on class and is effectively constant under naïve conditional independence assumptions the probability model becomes

$$p(C_k / x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i / C_k) \quad (3.8)$$

Where the evidence, $z = p(\underline{x})$, is a constant scaling factor dependent only on $\underline{x} = (x_1, \dots, x_n)$.

The classifier based on this probability model, the (Naïve) Bayes classifier is given by

$$\hat{y} = \frac{\arg \max_{k \in \{1, \dots, k\}} p(C_k) \prod_{i=1}^n p(x_i / C_k)}{k} \quad (3.9)$$

For some k that assigns the class label $\hat{y} = C_k$.

Different feature distributions are supported, normal and kernel for example. If feature distributions can be assumed normally distributed then each is modelled as such and the Naïve Bayes classifier estimates a further normal distribution for each of the classes by calculating the mean and standard deviation of the training data for the given class [13, 18, 19 and 91].

Features with a continuous distribution can be used along with the Kernel distribution which has the advantage of being robust even in the absence of normal distribution of features. However, greater computational time is required and more memory is utilised if distributions are skewed, multi peaked or multi modal, than with normally distributed variables. The Naïve Bayes classifier calculates a kernel density estimate for each class based on the training data for that class and a width for each class and feature, the kernel is by default normal.

When response data are categorical and the exact nature of the relationship between variables is unknown a non-parametric classification tree can assist interpretation of the system.

The tree predicts the response values at the circular leaf nodes based on a series of questions about the case at the triangular branching nodes. A true answer to any question follows the branch to the left; a false follows the branch to the right.

It should be noted that whilst the tree may fit the training data set well outliers can have a significant effect on the lower branches in particular in which case it is wise to 'prune' the tree [13, 36, 37, 62, 75 and 112].

3.3.1 Review and Current Practice

NB lends itself to increased numbers of groups and input parameters. Although a classification tree becomes overly complex, data is readily presented in matrix or graphical formats. Constructing a confusion matrix (class number by class number) of detailed case allocations records classification patterns. A 3-D bar chart of the confusion matrix data offers a clear visual display. Incorporating all groups with scatter plot visualisation is achievable through variable reduction techniques.

Naïve Bayes is essentially a machine learning technique although less prevalently utilised than ANNs which form the bulk of the learning algorithm methodologies. Authors of [62] use summary statistics (standard error, variance, kurtosis, range, maximum, minimum and sum) extracted from vibration signals to establish a NB classifier with 85% classification success rates.

3.4 Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that generally uses an orthogonal transformation to convert a set of highly correlated variables into a set of linearly uncorrelated variables called principal components (PCs) [20, 42, 73 and 91].

The method is designed to reduce the number of correlated independent variables, X , to a much smaller number of uncorrelated PCs, Z , which are weighted combinations of them. Each case can then be described by a reduced number of PCs which account for most of the variance. The higher the correlations between the original variables the greater the benefit from this method.

For an n by p matrix X consisting of n observations for each of p variables, a set of p -dimensional weights or loadings vectors, $w(p) = (w_1, \dots, w_p)_{(p)}$, map each row vector, $x_{(i)}$, of X , to a new vector of principal component scores, $z_{(i)} = (z_1, \dots, z_p)_{(i)}$, is given by

$$z_{k(i)} = x_{(i)} \cdot w_{(k)} \quad (3.10)$$

The full principal component decomposition of X given by $Z=XW$, where W is the p by p matrix whose columns are the eigenvectors of $X^T X$ and

$$\begin{aligned} z_1 &= w_{11}x_1 + w_{12}x_2 + w_{13}x_3 + \dots + w_{1n}x_n \\ z_2 &= w_{21}x_1 + w_{22}x_2 + w_{23}x_3 + \dots + w_{2n}x_n \\ &\cdot \\ &\cdot \\ &\cdot \\ z_p &= w_{p1}x_1 + w_{p2}x_2 + w_{p3}x_3 + \dots + w_{pn}x_n \end{aligned} \quad (3.11)$$

No data assumptions are required hence its attraction for use with non-interval data or data of unknown distribution [20, 42, 73 and 91].

Initially a set of uncorrelated PCs is produced from the original correlated variables. The first PC accounting for the largest proportion of the variance in the sample; the

second, which must be uncorrelated with the first, the second highest, and so on. Initially as many PCs as original variables are generated together accounting for the total variance in the sample. However, the vast majority of the total variance can be assigned to the first few PCs alone with only a negligible amount ascribed to the remainder. Hence these latter PCs can be dropped from further analysis so reducing the ‘dimensionality’ of the data set. PCA is mostly used as a tool in exploratory data analysis prior to construction of predictive models. Executed in practice either by eigenvalue decomposition of a data covariance or correlation matrix or by singular value decomposition of a data matrix. The later usually after mean centering normalised Z-scores of the data matrix for each attribute. PCA results are usually discussed in terms of their factor scores and loadings. Factor or component scores being the transformed variable coefficients corresponding to particular data points and factor loadings being the weight by which each standardised original variable is multiplied to achieve the component score.

$$\begin{aligned} \text{Var}(Z_i) &= \lambda_i \\ \sum_{i=1}^p \text{var}(Z_i) &= \text{trace}(C) \end{aligned} \tag{3.12}$$

Where $\text{trace}(C)$ is the sum of diagonal elements of matrix C , the covariance matrix, with the corresponding eigenvector Z_i , for each eigenvalue λ_i , given by

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p. \tag{3.13}$$

Operation of PCA can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data [14, 20, 21, 42, 50 and 73].

3.4.1 Factor Analysis

Similar in purpose to PCA, however, distinct in that factor analysis (FA) is founded on a true mathematical model being based on the row ratios of the correlation matrix of a set of original variables. Discounting elements in the leading diagonal, the self-

correlations, correlation matrices have the property that elements in any two rows are almost exactly proportional. Spearman first proposed the model over a hundred years ago (1904) on analysing standardised preparatory school exam scores and finding the common ratio for each of the subjects e.g. Classics and French; French and Music etc. to be approximately equal to 1.2. Hence proposing the model used today [73].

$$X_i = a_i F + e_i \quad (3.14)$$

Where X_i is the i^{th} standardised score, mean zero and standard deviation one; a_i is the factor loading which is a constant, F the factor value and e_i the portion of X_i specific to the i^{th} test only.

Thus there is a constant ratio between the rows of the variable correlation matrix hence this is a plausible model for the data. It also follows that the variance of X_i is given by

$$\begin{aligned} \text{var}(X_i) &= \text{var}(a_i F + e_i) \\ &= a_i^2 + \text{var}(e_i) \end{aligned} \quad (3.15)$$

Further, since the variables are standardised

$$\text{var}(X_i) = 1 = a_i^2 + \text{var}(e_i) \quad (3.16)$$

The square of the factor loading being the proportion of the variance of X_i that is accounted for by the factor. The sum of all the squared factor loadings, $\sum a_i^2$ is the communality of X_i and describes the part of its variance related to the common factors. The remaining part of its variance, which is not accounted for by the common factors, being given by $\text{var}(e_i)$, the specificity of X_i . Although there are no specific or widely accepted guidelines it is a generally accepted rule that loadings between ± 0.3 to ± 1.0 represent salient loadings with the interpretation that the original variable is meaningfully related to that particular factor. Should the factor loadings be difficult to generalise being neither close to zero or ± 1.0 a rotation of the solution could be considered. It should be noted that factor rotation is a mathematical aid to interpretation rather than a refitting of the model hence will not affect the overall

goodness of fit of the model simply the arbitrary axes along which the factors are measured [20].

Whilst FA has its limitations [20] it is of particular benefit in gaining insight into the nature of underlying variables in multivariate data. Its worth being as a descriptive tool for investigation through sensible means to uncover or describe underlying data structures albeit with consideration of methodological limitations. Thus although FA is largely an exploratory technique substantive and practical considerations should strongly guide the analytical process to effective gain.

3.4.2 Review and Current Practice

PCA is frequently used in CM of industrial systems. Prevalent in monitoring chemical processes and often incorporated with wavelet transform methods. Fault classification using PCA extensions [114] and recently emerging Kernel PCA (KPCA) [7, 93 and 113] are increasingly employed. Kernel based PCA extends the method for use in non-linear or overlapping cluster applications. Wavelet transform methods are evolving as an alternative feature extraction method to Fourier much favoured in [93]. Wherein superiority of the KPCA ability to extract higher-order non-linear interrelationships on application to data of higher orders of complexity is also claimed. The paper of [93] further investigates the benefits of using radial based kernel functions with varying parameters to match varying speeds of rotary machinery. Another extension to PCA, multi-scale PCA is considered in section 3.9.

3.5 Support Vector Machines

In machine learning, support vector machines (SVMs) are supervised learning models with associated learning algorithms that analyse data through pattern recognition, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier [19, 26, 108 and 109]. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New

examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on [26].

SVMs can also efficiently perform a non-linear classification using a kernel function, implicitly mapping inputs into high-dimensional feature spaces. Whereas the original problem may be stated in a finite dimensional space, often the sets to discriminate are not linearly separable in that space. Thus the original finite-dimensional space is mapped into a sufficiently higher-dimensional space, making separation easier. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function, $k(x, y)$, selected to suit the problem. The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters α_i of images of feature vectors x_i that occur in the data base. With this choice of a hyperplane, the points x in the feature space that are mapped into the hyperplane are defined by the relation:

$$\sum_i \alpha_i k(x_i, x) = \text{constant} \quad (3.17)$$

Note that if $k(x, y)$ becomes small as y grows further away from x , each term in the sum measures the degree of closeness of the test point x to the corresponding data base point x_i . Hence the sum of kernels give a measure of the relative nearness of each test point to the data points originating in one or other of the sets to be discriminated [26].

Given a set of training data, D , a set of n points of the form

$$D = \{(x_i, y_i) / x_i \in \mathfrak{R}^p, y_i \in \{-1, 1\}\}_{i=1}^n \quad (3.18)$$

where y_i takes the value +1 or -1 indicating which class the point x_i belongs to.

Each x_i is a p -dimensional real vector. The maximum-margin hyperplane dividing points with $y_i = 1$ from those with $y_i = -1$ is given by the set of points x satisfying $w \cdot x - b = 0$.

If the data are linearly separable, hyperplanes can be selected in such a way that they separate the data with no points between them. The region they bound, the margin is then maximised whilst ensuring no points are allowed to fall into it. The planes of the margin are given by $w \cdot x - b = 1$ and $w \cdot x - b = -1$. Samples falling on the margin are called the support vectors. The first class occupies the region $w \cdot x_i - b \geq 1$ and the second class the region $w \cdot x_i - b \leq -1$ jointly described as

$$y_i(w \cdot x_i - b) \geq 1 \quad \text{for all } 1 \leq i \leq n \quad (3.19)$$

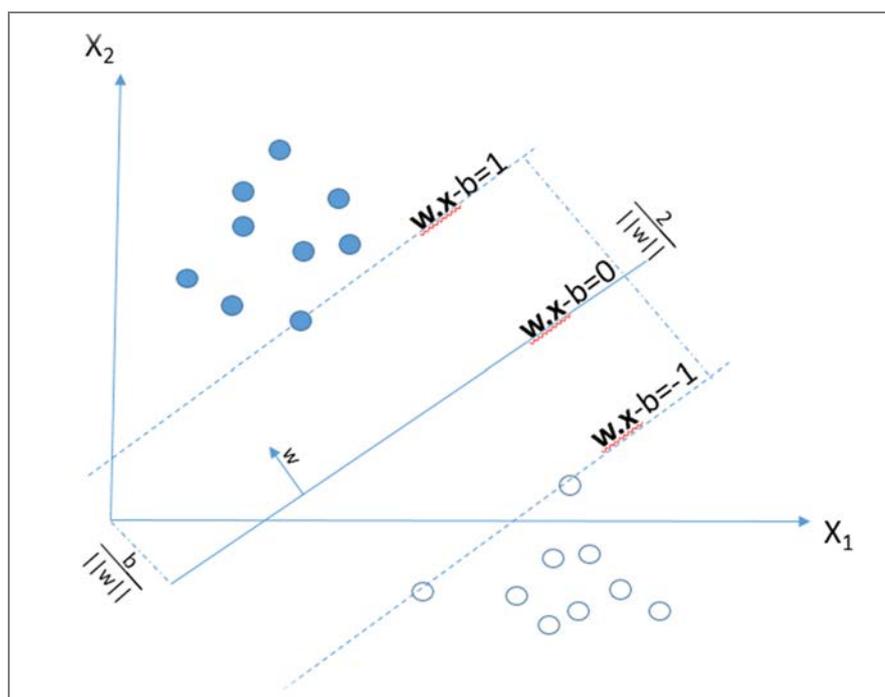


Figure 3.1 SVM Illustration of Margins and Hyperplanes.

3.5.1 Review and Current Practice

Support vectors occur on, therefore define, the boundaries between groups. Vibration sensor outputs were used in [104] with SVM and fast training algorithms were reported

maintaining classification accuracy. Authors in [63] provide an overview of CM artificial intelligence techniques and show less confidence in their accuracy on application to real systems. Stationary features extracted from statistical moments of vibration signals were used in [108] to compare SVMs and ANNs in identifying butterfly valve defects. Massive computational requirements of solving quadratic optimisation problems were overcome using kernel functions. Kernel function selection in [108] was ultimately done by trial and error. Classification rates for their samples being 100% using SVM with chosen kernel function and 80 to 85% using ANNs. [109] extends the model comparisons to small refrigerator compressors further claiming SVMs to be more generalisable than ANNs due to their corresponding risk management strategies. Namely, SVMs minimise an upper bound on the expected risk whereas ANNs traditionally minimise the data training error. Authors of [83] concur with respect to improved generalisability of SVMs. Utilising SVM with a wave matching feature extraction method known as differential evolution (DE) akin to genetic algorithms (GA), (section 3.8). Waveform feature extraction requires prior knowledge of working principles and is prone to distortion due to interfering vibrations. Also SVMs are less successful in the analysis of large group problems. More successful with increased numbers of classes are RVMs (section 3.7) which also require significantly fewer descriptive vectors.

3.6 Relevance Vector Machines

Tipping [101] introduced a machine learning technique which is analogous to SVM. Relevance vector machines (RVM) use an iterative Bayesian inference approach to obtain parsimonious solutions for regression and probabilistic classification. The most probable values of a set of hyper-functions are iteratively estimated from the data by introducing a prior density over all the weights. Sparsity is achieved in practice as the posterior distributions of many of the weights are sharply peaked around zero. Having a Student t-distribution rather than being normally distributed the bulk of the relevance vectors are zero. This is advantageous in reducing the number required in the model but the distribution not probabilistic, hence not directly quantifiable. Furthermore the non-zero weights, the relevance vectors, are not associated with examples close to the decision boundaries as the support vectors are but appear to be representative of prototypical class examples. Its functional form is identical to that of SVM but its probabilistic classification is generally a Gaussian model with covariance function

$$k(x, x^T) = \sum_{j=1}^n \frac{1}{\alpha_j} \phi(x, x_j) \phi(x^T, x_j) \quad (3.20)$$

Where ϕ is a kernel function which is generally Gaussian,

α_j the variances of the prior on the weight vector $w \sim N(0, \alpha^{-1}I)$

x_1, \dots, x_n the training set input vectors.

RVM utilise dramatically fewer kernel functions (the relevance vectors) than SVM whilst maintaining equivalent performance capabilities thus generating sparse computationally efficient models whose non-zero weights are more representative of the classes [2, 19, 101 and 102].

Unlike SVM RVM are probabilistic models. The RVM probabilistic model with x_1, \dots, x_n input training vectors and t_1, \dots, t_n corresponding target values. Target values are real values in applications to regression and are the class labels for classification models. Given a function $y(x)$ the parameters of which are to be inferred or determined in the machine learning process, the output is a linearly weighted sum of m , generally non-linear, fixed basis functions $\varphi(x) = (\varphi_1(x), \dots, \varphi_m(x))^T$

$$y(x; w) = \sum_{i=1}^m w_i \varphi_i(x) = w^T \varphi(x) \quad (3.21)$$

Where $w = (w_1, w_2, \dots, w_m)^T$ are the prior weights and

$\varphi(x) = (\varphi_1(x), \dots, \varphi_m(x))^T$ the basis functions

Only samples associated with non-zero weights, the relevance vectors, contribute to the decision function. RVM having dramatically fewer kernel functions whilst maintaining comparable predictive capabilities consumes far less test time than equivalent SVM models a particularly important feature in on-line fault detection.

3.7 Genetic Algorithms

Genetic algorithm is a method for solving optimisation problems which mimics the process of biological evolution. The algorithm repeatedly modifies a population of individual solutions with each repetition producing a new generation of individuals whose fitness is determined relative to objective function values. In RCs fault analysis, the fitness function uses the classification rate of the training data. Optimal when errors are minimised. This new generation is in turn used in the next algorithmic iteration to produce a further generation and so.

Termination of the process is governed by:

- A solution being found that satisfies minimum criteria
- A fixed number of generation being reached
- An allocated budget or time constraint being reached
- The fitness function plateauing
- Manual inspection
- A combination of the above.

Limitations of the process:

- Highly complex fitness functions require considerable computational time.
- GA with a high number of mutating elements tend to exponentially increasing sample spaces.
- As solution quality is relative to past solutions the fitness requirements are not easily defined a priori.
- Tendency to converge on local optima rather than generating a global solution.
- Operation on dynamic data sets gives rise to shifting targets which hinder convergence.
- Not suitable for decision problems as there is no graded learning (on/ off, right/ wrong).

Evolutionary GA such as grouping GA are used for clustering or partitioning groups or dividing items into distinct groups. Particle Swarm Optimisation (PSO) another population based search method inspired by observation of the collaborative behaviour of biological populations such as birds or bees. Specifically these populations are seen to demonstrate a collective intelligence [49, 70, 86 and 90]. Both GA and PSO are restricted to 10 to 15 input parameters maximum for convergence within reasonable time constraints.

Other algorithmic considerations with respect to computational outputs to be considered [2 and 101] whether to employ a one-against-all (OAA) or a one-against-one (OAO) strategy. Whilst the OAO algorithm generates more individual classifiers to be solved they are of greater simplicity than the fewer more complex classifiers resulting from the OAA method and ultimately OAO algorithms show considerable reductions in overall process time for large data sets.

Many current studies incorporate GA feature selection [2, 49, 70, 86 and 90]. Major advantages being prior knowledge of the process is not required to establish the best fit. Main criticisms being tendency to over fit and limitations of input parameters.

3.8 Multiscale Principal Component Analysis

Wickerhauser is credited with the original idea to jointly use wavelets and PCA [105] with the current model first proposed in [7]. Multiscale PCA (MSPCA) combines the ability of PCA to produce a set of uncorrelated variables with that of wavelet analysis to extract deterministic features. The wavelet coefficients of the PCA are calculated at each scale and selected results are combined. Only those scales showing significant events are included thus the process both de noises and simplifies the original multivariate signal. The technique is appropriate for modelling data with dynamic events due to its multiscale nature, hence its suitability for process fault detection.

PCA captures the correlation and maximum variance between measurements and wavelet analysis captures the within measurement correlation. Thus both the variable correlation and the signal trend are accounted for by MSPCA. The complementary strengths of each procedure resulting in maximum information being extracted from complex multivariate measurements. The aim of multiscale PCA is to reconstruct a simplified multivariate signal, starting from an original multivariate signal and using a simple representation at each of a specified number of resolution levels. Multiscale principal components analysis generalises the PCA of a multivariate signal represented as a matrix by simultaneously performing a PCA on the matrices of details at different levels. A PCA is performed on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting the numbers of retained principal components, interesting simplified signals can be reconstructed. Rules for retention of PCs are akin to those of PCA for example Kaiser's rule retains all PCs with eigenvalues greater than the mean eigenvalue i.e. those contributing greater than average explanatory power.

Each input variable is decomposed at a specified level, L using a discrete wavelet transform (DWT).

The wavelet approximation models, A_L , from each variable are stored in a single matrix (of size m by $\frac{n}{2^L}$) with the wavelet details from each of the L levels stored in matrices of size $m \times \frac{n}{2^i}$, $i = 1, 2, \dots, L$.

A total of $L+1$ matrices (A_L, D_1, \dots, D_L) each representing a different scale and the captured trends at that scale. PCA is performed on all $L+1$ matrices with the objective of extracting correlation across the sensors. Highly correlated variables are then clustered and fed in as sets to local PCA models for sensor validation. Thus a normal PCA model is constructed with this data set being used to build the normalising model [3, 7, 74 and 105].

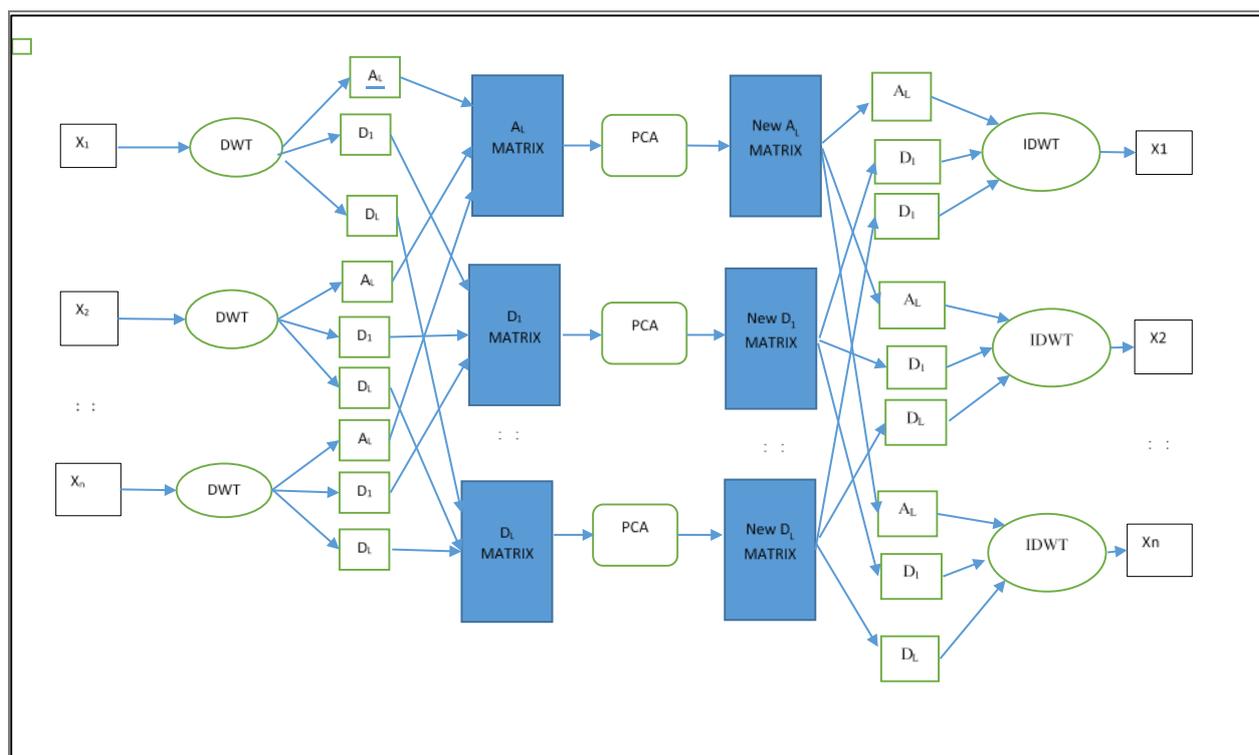


Figure 3.2 Schematic Representation of MSPCA Process Reproduced from [74].

3.8.1. Review and Current Practice

MSPCA is a natural progression in that it combines both waveform feature extraction and PCA variable reduction. Much previous research has emulated this technique but not as an embedded methodology. However, its embedded principle is also a weakness in that specific techniques cannot be tailored to application requirements. A bespoke coupling was used in [103] to diagnose RC valve faults, joining Teager-Kaiser and deep belief networks. The former to estimate the envelope amplitudes which are de-noised by wavelet transform methods and the latter to establish the fault classifier. Likewise wave matching feature extraction methods along with SVM were employed in [83] for classifying vibration signals for fault detection. MSPCA with prior data compression and signal simplification is shown to be effective in reducing input parameter volume [98] with much improved classification rates for identifying RC component faults.

3.9 Andrews Plots of Fourier Profiles

Having identified suitable input parameters for the model building process it is sensible to consider the feasibility of class separation. One useful profile method is found in Andrews plots, a Fourier transform of the signal data. This exploratory data analysis technique attempts to identify structure within the data. If there are distinct data groupings, e.g. the classes examined on the compressor rig, then a Fourier profile plot may highlight differences and so assist in distinguishing between groups (classes).

Andrews' plots of the multivariate data in the matrix X . The rows of X correspond to observations, the columns to variables. Andrews plots represent each observation by a function $f(t)$ of a continuous dummy variable t over the interval $[0,1]$ $f(t)$ is defined for the i^{th} observation in X as

$$f(t) = \frac{X(i,1)}{\sqrt{2}} + X(i,2)\sin(2\pi t) + X(i,3)\cos(2\pi t) + \dots \quad (3.22)$$

Since X contains a large number of observations an Andrews' quantile plot showing only the median and quantiles of $f(t)$ for each t might be utilised to aid interpretation [19, 73, and 91].

3.9.1 Review

Prior assessment of signal profiles for healthy and faulty systems allows realistic understanding of fault detection potential. If the Fourier profiles are not distinct at any point it is unlikely a classifier could be established using envelope spectra harmonics which is capable of differentiating between the classes. However, no evidence was uncovered reporting prior analysis of class profiles in literature searches despite much analysis of this nature.

3.10 Summary of Multivariate Statistical Techniques

CA gives a clear measure of variable connections and dissimilarities whereas RVM and SVM do not, neither do they give a clear indication of variable properties. Variable clustering is employed to gain insight into variable characteristics. Proximity measures in CA are application dependent. Euclidean distance is considered the most apt for these analyses measuring the physical distance between two data points. Agglomerative CA algorithms applied to variable clustering highlights the most like variables. Illustrated on a dendrogram the clustering gives a clear indication of group formation at progressive levels hence is preferable to division in this instance.

Data mining techniques, ANN, SVM and other wavelet methods are discussed in terms of computational efficiency and classification rates applied to wind energy. [63] conclude most controllers are not validated in the field. Hence whilst standard CM approaches may provide accurate results in controlled experimental environments they show weaknesses on application in less predictable circumstances and turbulent winds. Weaknesses which are difficult to evaluate without the ability to identify true relationships to original variable properties.

Class Fourier profiles can be examined for all suitable variables to assess potential for separation. Hence realistic expectations and tolerance setting is feasible.

Construction of appropriate classification models to determine adequacy of input parameters is feasible. CA has the capability to inform input parameter selection in both statistical and machine learning methodologies. In addition, variable reduction techniques offer further illumination with respect to classification success rates.

Chapter 4

Condition Monitoring Data Acquisition from a Compressor Rig

This chapter presents details relating to the test rig specification and data collection process. Functionality of the testing rig is described along with the process of collecting data pertaining to the various faults studied. Experimental procedures and conditions for ensuring reliable readings are also explained. Configuration of the testing rig and the process layout are illustrated verbally, diagrammatically and pictorially. Transducer specifics are stipulated and ranges of capability.

Findings from exploratory investigations of signals captured at strategic points of the compression process are also detailed. A summary of the physical attributes and interrelationships between different signals measured during healthy operation is reported. Salient behaviours are scrutinised and catalogued.

4.1 Reciprocating Compressors

Reciprocating compressors (RC) are vital components of many potentially volatile industrial processes. For example, oil refineries, gas pipelines, chemical plants and refrigeration plants. Component failure is therefore potentially life threatening as well as costly and time consuming. Efficiency and continued performance of these processes rely on early detection of RC component deterioration. Monitoring RCs is difficult mainly due to inaccessibility of component parts hence the importance of performance monitoring through suitably positioned sensors for signal capture. Vibration signals can provide high levels of information facilitating detailed and accurate assessment of system condition. RCs are susceptible to a multitude of faults, both due to mechanical and elastic deterioration, which can occur in isolation or combination. Extracting useful information is often hindered by the large amount of noise captured along with measured signals. For meaningful analysis this extraneous variation needs to be filtered out. In reciprocating machines the problem is all the more apparent than in rotating mechanical systems due to the greater fluctuations in vibration amplitudes and increased complexity of interacting component parts. Thus the potential complexity of RC fault monitoring lends itself perfectly to investigation through complex multivariate statistical models. Techniques which are then easily transferable to other mechanical processes with less intricacy [72 and 85].

Details of the experimental rig employed and the structure of data acquired are given in the following sections.

4.2 Test Facilities and Data Collection

An inherent difficulty in the condition monitoring of RCs is that of accessing component parts which are most apt to fail, the valves for example. This makes direct methods of monitoring more problematic in their application. A suitably monitored test rig was thus utilised with sensors attached at critical stages to record outputs from the process. Thus the process is monitored operating under normal 'healthy' conditions with components in fully operational mode. Resulting measurements provide a bench mark against which to compare the outputs generated once operating under altered conditions simulating various faults which are subsequently introduced to the system [1, 2, 32, 33, 34, 35, 46, 69, 82 and 103].

The ultimate aim being not only to diagnose a non-optimal operational state but to pinpoint the precise problem(s) and predict efficacy so as to avoid inefficient process operation. Ultimately aiming for prognostic specification i.e. predicting precise time to intervention before loss of quality is currently beyond the scope of this research.

4.2.1 Reciprocating Compressor Rig

Output signals were collected from transducers mounted on a two-stage, single-acting Broom Wade TS9 reciprocating compressor rig. A compressor widely utilised in industry. The rig incorporates two cylinders, opposed at 90°, in the form of a 'V'. Intake air passes through a filter to ensure it is sufficiently clean and dry prior to entering the compressor. Air is drawn into the first stage cylinder and compressed, the low pressure and low vibration measurements are monitored at this point via sensors attached to this first stage cylinder. Compressed air is then transferred to the second stage, the second cylinder, and further compressed. High pressure and high vibration measurements are similarly recorded via a second pair of sensors attached to cylinder two. The doubly compressed air is thus transferred to a horizontal holding tank. Temperature sensors also record measurements at both stage 1 and stage 2 of compression. Further transducers are attached to the motor crank shaft to measure electrical current and instantaneous angular speed with a static pressure sensor on the horizontal tank measuring the reservoir pressure. A safety valve on the storage cylinder was incorporated to guard against excessive pressure build up. Splash lubrication protected moving parts. The compressor drives a vane flywheel. Coiled copper piping performs heat removal within the intercooler.

Valve leakage faults were simulated via 2mm diameter drilled holes, an approximate 2% leakage. As joint leakage in the IC is common, the joint adjacent to the second cylinder was loosened by a half turn to simulate malfunction. The LB was simulated by reducing the standard distance between pulley centres by 1.5mm.

The RC was operated under healthy conditions and with four independently seeded faults (suction valve leakage (SVL), discharge valve leakage (DVL), intercooler leakage (ICL) and loose drive belt (LB)), each run being repeated 24 times. Thus a total of 120 observations were recorded at each of six pressure loads.

Sensors attached at appropriate positions of interest yield measurements recorded in Matlab.bin files which in turn are read into the Matlab computing programme. These digital messages are subsequently translated into physical quantities. Trends and interrelationships between the variables during normal running of the system and with

seeded faults were analysed to find identifiable patterns for model building purposes. The ultimate goal being to characterise and so enable identification of fault onset in a timely fashion.

A total of seven measurements were taken during the course of the operation, namely:

1. Low Pressure, pL , the pressure reading at the first stage of compression,
2. High Pressure, pH , the pressure reading at the second stage of compression,
3. Vibration at the first stage of operation, vL .
4. Vibration at the second stage of operation, vH .
5. Angular speed of the crank shaft, ang .
6. Index angular speed, ind .
7. Motor current, i .

These measurements being repeated for each of the six loads, pd . Pressure being directly proportional to the load, as the load increases then so does the pressure in the system. The load effect on the above variables was also investigated. The load measurements varied between 0 (empty cylinder) and 120 (full cylinder), with measurements taken on the seven variables at 10 unit load increments between 70 and 100psi (4.83 and 6.90bar).

Each of the seven variables was recorded in binary form via seven channels linked to a converter connected in turn to a PC thus a seven by 'number of channel points' array of data values was generated. The number of data points per channel depending on the sampling frequency, F_s . The sampling process was repeated three times per simulation thus generating three batches of measurements per channel i.e. 3 x the number of channel points per section, typically 6×10^4 data points in all three sections. It should be noted that whilst these sections will be homogeneous, obvious changes between sections will be observed which should signify only an interruption in measurements taken rather than any significant change in the data readings.

All measurements were initially investigated with respect to their variation and relation to physical measurements. A mainly graphical inspection was made at this stage with each of the channel outputs plotted either against time in seconds or crank shaft angle, θ , in degrees.

$$time (s) = \frac{\text{the number of channel points}}{\text{the sampling frequency}} \quad (4.1)$$

$$\theta = \omega t = \frac{440}{60} \times 2\pi t \quad (4.2)$$

As the motor operates at 440 revolutions per minute.

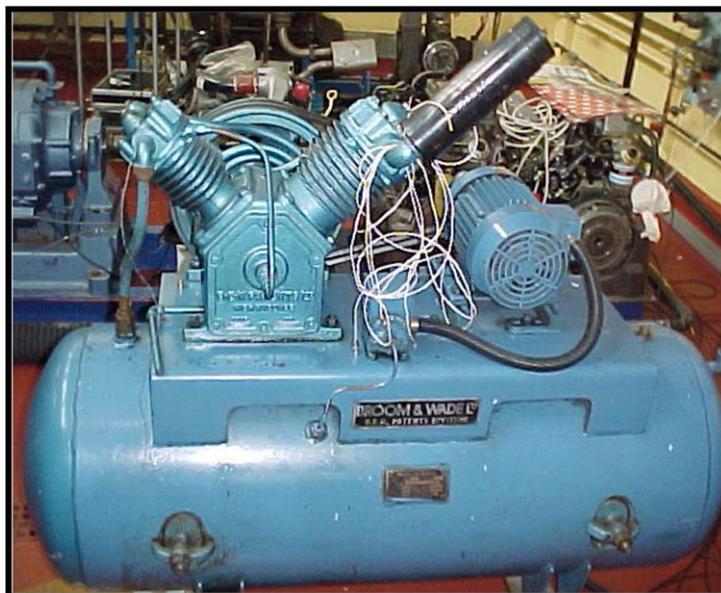


Figure 4.1 Broom WadeTS9 Reciprocating Compressor Rig.

TABLE 4.1 TWO-STAGE BROOM-WADE RECIPROCATING COMPRESSOR SPECIFICATION.

Type	TS9
Maximum working pressure	1.379MPa (13.8bar)
Number of cylinders	2 (90 degrees opposed in V shape)
First stage piston diameter	93.6mm
Second stage piston diameter	55.6mm
Piston stroke	76mm
Crank speed	440rpm
Motor power	2.2kW
Supply voltage	380/ 420 V
Motor speed	1420rpm
Current	4.1/ 4.8A

4.2.2 Sensor Details and Specification

A three phase, squirrel cage, air cooled driving motor mounted on the receiver transfers power to the compressor through a pulley drive belt system. Motor type KX-C184, transmission ration 3:2 and crank shaft speed 440rpm when operated at 1420rpm. Transducers were connected to the Data Acquisition system (DAS) via coaxial BNC cables to reduce signal noise.

Accelerometers, type YD-5-2, were attached to the head of each cylinder alongside the suction and discharge valves. Each with a frequency range between 0 and 15kHz; sensitivity of 45mVms^{-2} and a temperature tolerance up to 150°C and accelerations up to 2000ms^{-2} . Robust enough to withstand high and fluctuating temperatures and large shock levels yet sufficiently sensitive to measure very low acceleration levels. Affixed via screw threaded brass studs and sealed with suitable ceramic cement offering the additional advantage of buffering heat surges. Signal measurements were relayed to the computerised DAS for storage.

Dynamic strain gauge pressure transducers, GEMs type 2200 were inserted into cylinder pipes also at the cylinder head. With an output of 100mV used with a 10Vd.c. power supply. A range up to 4MPa (600psi) and upper frequency limit of 4kHz. No amplification being required these sensors connect directly to the CED and PC.

Storage tank, static, pressure sensor was recorded by GEM type PS20000. With a maximum outage of 100mV for a 15V supply. Temperature range between -20°C and 105°C and operating pressure range between 0 and 1.35MPa (200psi).

Temperature measurement was via linear response k-type thermocouples positioned inside the cylinder pipe between the pressure sensors and the cylinder head. Responsive between -20°C and 220°C . An operational safety check for both the pressure sensors and the compressor itself.

A Hengstler incremental optical encoder attached to the drive shaft using a spindle adapter monitors IAS to one degree recording 360 pulses per revolution.

A Hall Effect current transducer, RS 286-327, mounted on a printed circuit board measured stator current without necessity to connect to the circuit. Operating temperature between 0°C and 70°C ; bandwidth 100kHz; output voltage 5V; supply voltage $\pm 15\text{Vd.c.}$ ($\pm 5\%$); response time $< 1\mu\text{second}$.

The DAS, CED 1401, a multifunctional data collection interface converts the analogue voltage to a digital value. Recording waveform, digital and marker information and generating output simultaneously for real-time multi-faceted experimental systems such as the CR. With a powerful processor and good memory capabilities high measurement accuracy is achievable.

National Instruments Lab Windows software Tm/ CVI/ version 5.5, written in the C language, enables data storage and conversion into Matlab bin files for export and analysis. A large number of instrument control libraries enable data acquisition and analysis. Eight of the available sixteen data channels were utilised. The set-up panel allows adjustments to specific experimental task. Sampling frequency, F_s , was set to 62.5kHz and the data length to 30.642 thus time between consecutive data points was 0.4903 seconds.

4.3 Experimental Configuration

Data acquisition follows strict procedural steps thus ensuring identical conditions for each run. General safety procedures were routinely performed and maintained isolating the area and ensuring moving parts are suitably covered yet free from interference. Mains connection and all wiring examined to verify correct connectivity and eradicating signs of wear prior to switch on. Access via password protection to the data acquisition channels initiates the second level of safety and procedural checks. Once approved for operation any prior data collection files are deleted and the system allowed to run whilst monitored on line. During this stage each sensor is validated. The CED (Cambridge Electronic Design Ltd.) setup 'trigger' is applied for data saving with 'view' enabled to test signal sensitivity. Lightly tapping cylinders to provoke a response and check connectivity. For null response a reduction in the sampling rate may be required. Seven channels allow simultaneous recording of each of seven signals: the first and second stage cylinder pressures; first and second stage cylinder vibrations; an index signal detailing crank shaft rotations alongside crankshaft rotation variability and the motor current. Each set of signals being measured at six storage tank loads.

The integrated PC has the capacity for one data run alone so each data batch requires external storage and internal deletion prior to a further run. Data collected is binary and is conveyed to the MATLAB software as .bin files for analysis.

Graded faults in components are seeded into the process with data collection repeated as for the healthy system. Signal responses are thus directly comparable.

Reproducibility is checked by repeated running of a given condition.

4.4 Data Structure for MATLAB analysis

Advanced multivariate analysis was conducted on the amplitudes of the envelope spectra harmonics. The first 32 harmonics were collected and their amplitudes for each of 120 samples stored. Measurements comprised 24 repetitions for each of the five classes studied. These 32 harmonics, each a vector of length 120, being the potential input variable set. Collectively a data matrix, X , was constructed of size 120 by 32 i.e. 120 observations by 32 variables. The observations being arranged by class with the first 24 rows corresponding to the healthy class, the next 24 to the first fault and so on.

Hence in this format the classification matrix, X , of observations per class by variables could be directly incorporated in the multivariate analysis. Relevant variables being incorporated in the analysis by selection of corresponding columns of X . Rows of X representing envelope harmonic amplitudes for a particular case i.e. a particular experimental run.

$$X=[s(1).Aeh'; s(2).Aeh'; s(3).Aeh'; s(4).Aeh'; s(5).Aeh']$$

Where $s(i).Aeh$ denotes the set of harmonic amplitudes for the i^{th} class of measurements ($i=1, \dots, 5$) [18, 19 and 44].

4.5 Summary

Clearly compressors form an intrinsic part of many industrial processes and their healthy maintenance is key to continued process efficiency. Deterioration of valves and other soft components has a direct effect on the compressors performance. Detection and identification of these fault types will form the basis of the research problem in this thesis.

A compressor widely used in industry was selected for experimentation. Transducers were chosen to give accurate readings within the desired range but with sufficient capacity to withstand surges in temperature and shocks. Measurement protocols were enforced to ensure direct comparability of all experimental runs. Seeded faults were designed to emulate actual in process faults hence provide realistic and informative data analysis modelling capabilities.

Data taken from a reciprocating compressor rig operating under various conditions of health was analysed to determine key characteristics of decaying parts and sub-par operation. Output data signals collected during operation were investigated to ascertain which most efficiently describe the salient features of machine condition, hence which have the greater worth in determining existence and nature of faults.

4.6 Physical Attributes of Output Signal Data and Measurements

This section investigates the characteristics of the output variables, measured in a healthy compressor, operating normally.

An experimental compressor rig imitates the process of industrial machines, allowing the researcher to observe and record performance by conducting a sequence of controlled trials. Thus a healthy fully operational system can be observed and signal outputs from it compared and contrasted with similarly controlled outputs generated during operation with various seeded faults.

4.6.1 Low Pressure Measurements

Air drawn into the system directly from the atmosphere has a pressure just below 2bar (0.2MPa). The compression stage begins immediately as the piston moves upwards in its cylinder until a pressure of nearly 4bar is achieved as shown in Figure 4.2 and Figure 4.3. The pressurised volume having been discharged a further quantity of air enters the chamber whilst the piston is drawn back down the cylinder. The suction stage is identifiable by a sudden pressure drop back to less than 2bar. This process is repeated seven times per second, each cycle of suction and compression being completed in approximately $\frac{1}{7} = 0.143$ seconds. Clearly the three recorded sections are homogeneous indicating experimental uniformity and operating consistency. The discharge valve opens for approximately 0.025 seconds whilst the suction valve opens for almost three times as long. Intake gas transfers more slowly due to its being of lower pressure and so less agitated. Gas is discharged at greater pressure thus with more forceful expulsion.

Evidently the pressure increases during the compression stage as the piston moves upwards in the cylinder towards the top dead center (TDC) position and the available

volume decreases. The discharge valve oscillates as the pressurised air is forced through it and closes as the piston reverses its direction to travel back down the cylinder towards the bottom dead center (BDC) position. Conversely throughout the expansion stage the pressure is reduced as the available volume increases. During the expansion stage the suction valve opens, also oscillating but less forcibly, and more air is drawn into the cylinder. The compression stage then restarts as the piston again travels towards TDC position and so the cycle continues repeatedly.

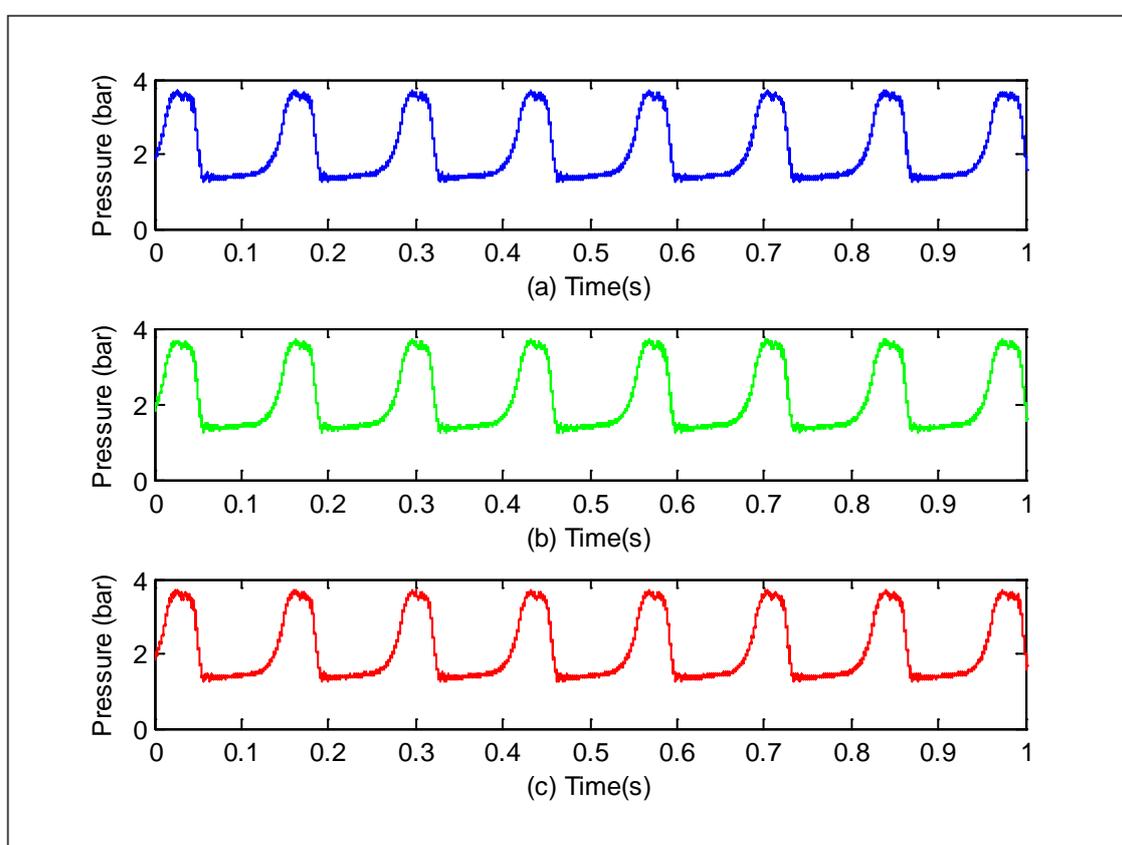


Figure 4.2 Comparison of the Low Pressure Measurements per Section. (a) Section 1, (b) Section 2 and (c) Section 3.

4.6.2 High Pressure Measurements

Again widened fluctuating sections at maximum and minimum pressures indicate opening and closing of valves which are pliant so vibrate against their seal, Figure 4.4. Pressure cycle patterns from each stage are reasonably uniform across the observed

period. Low pressure maximums are approximately 3.75bar with high pressure maximums approaching 7bar. Note the intake pressure at stage two is reduced to around 2.5bar during transition from the first to the second stage. Pressures at each stage exhibit the seven cycles per second expected (440rpm/ 60s). The first stage discharges as the second stage suctions, visibly illustrated in Figure 4.5. Subsequently the second stage discharge valve opens as the first stage cylinder is in its compression stage. Hence the two stages work in harmony and the package of air is efficiently passed through the system under increased pressure. Compression at the first stage takes longer and is less abrupt than that of the second stage whilst discharge for both stages are of comparable duration.

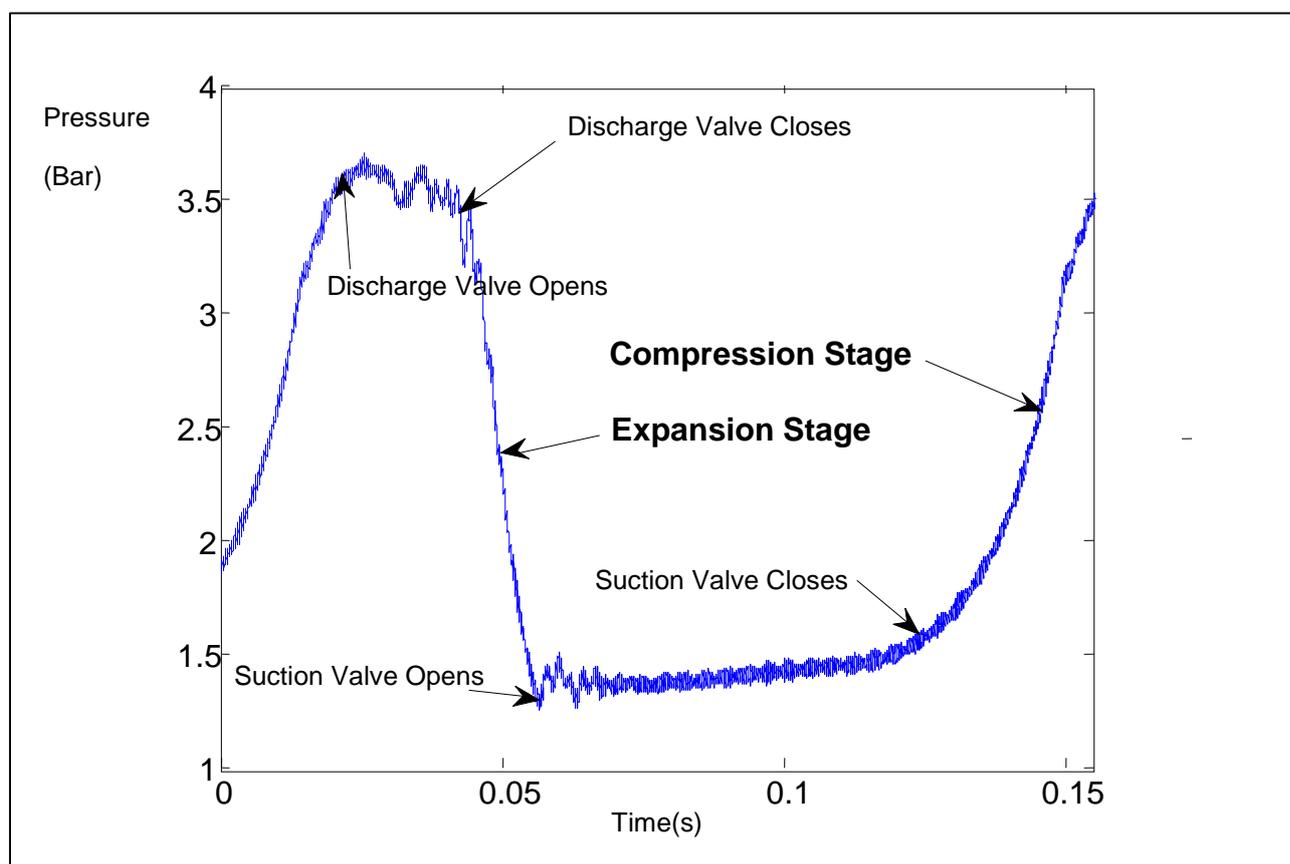


Figure 4.3 Annotated Plot of First Stage (Low) Pressure Cycle.

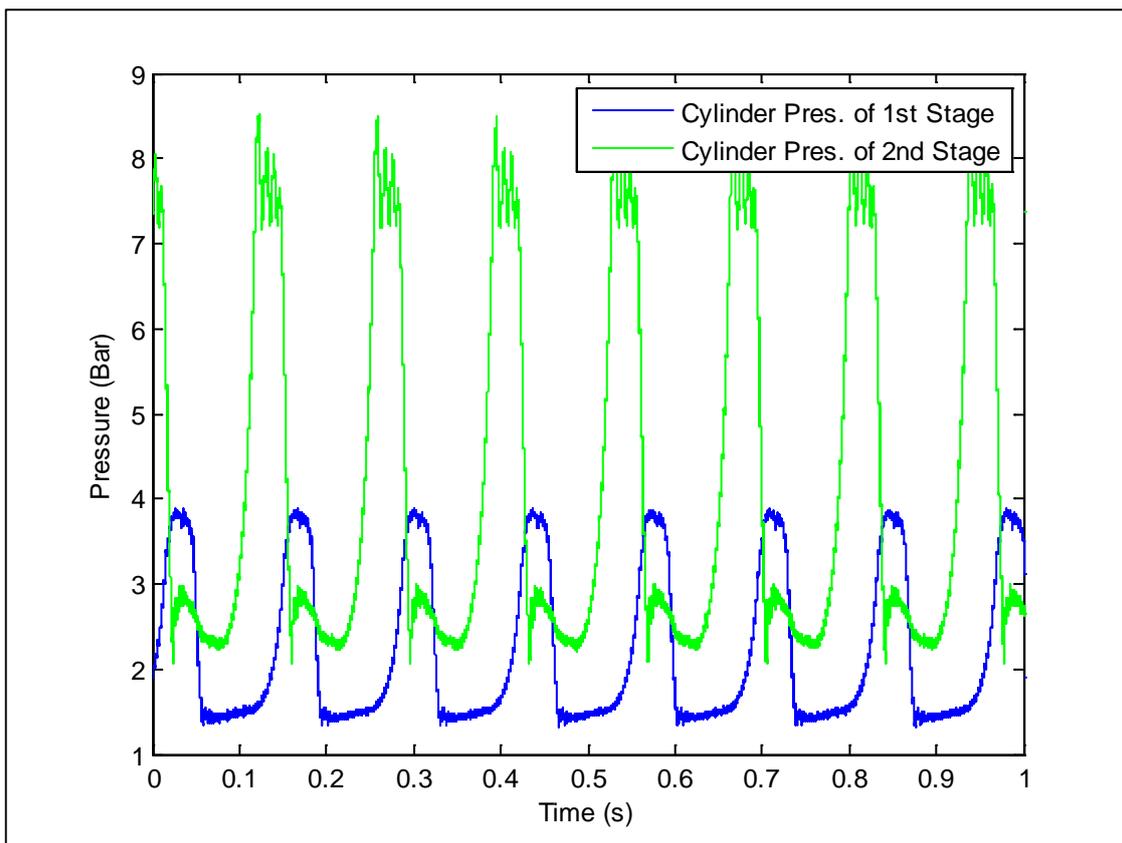


Figure 4.4 Comparison of First and Second Stage Pressure Measurements.

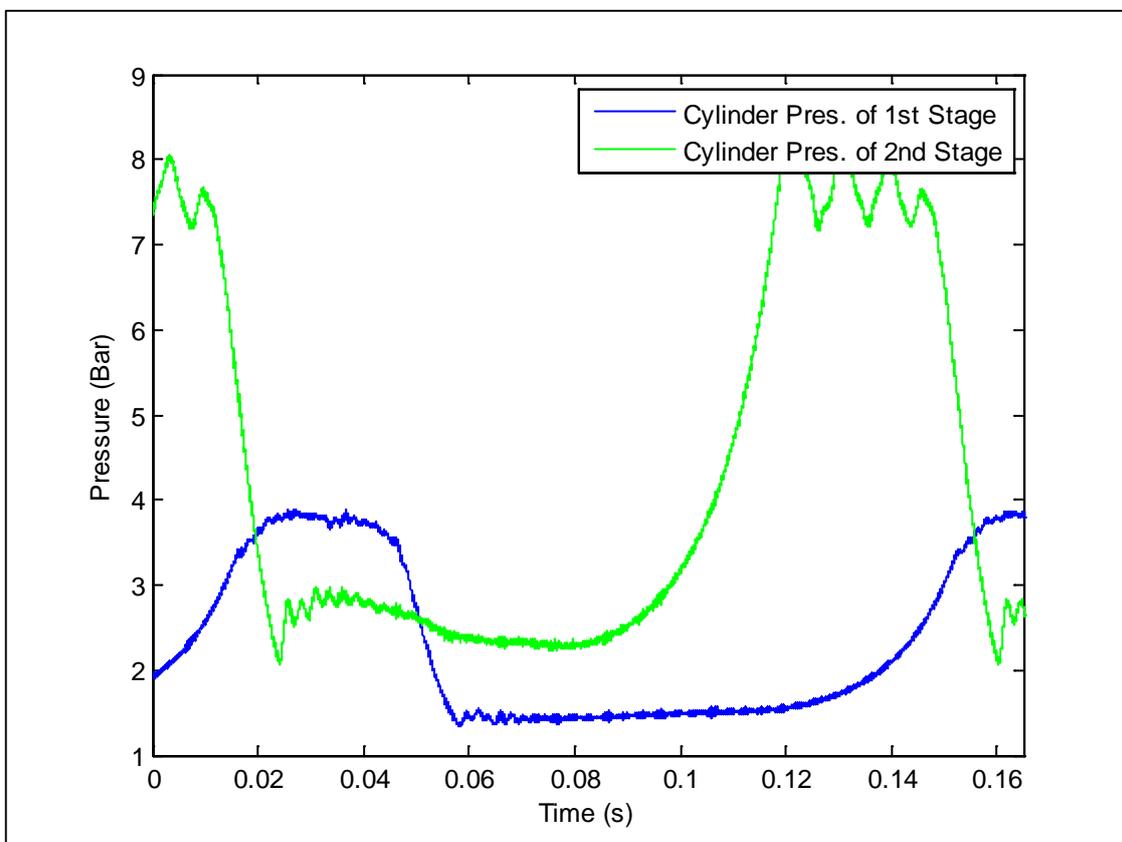


Figure 4.5 Comparison of Compression and Discharge Phases for First and Second Cylinder Pressure Stages.

4.6.3 First Stage Vibration Measurements

Once again signal output is clearly periodic completing seven cycles per second. Inspection of times between peak values highlighted possible inconsistencies in shaft speed. However, comparison across the three sections shows similar amplitudes and periodicity. There is evidence of a trend drift in the third section, Figure 4.6, indicated by the red trace deviating from the horizontal. This is most likely due to external influence, the accelerometers being highly sensitive and so susceptible to interference. Measurements were normalised by the calibrated sensitivities of the transducers. Sensitivity, the root mean square of the signal voltage, was calculated to be $11.0682\text{mV} / \text{ms}^{-2}$ for the vibration at the first stage and $11.7972\text{mV} / \text{ms}^{-2}$ for the vibration at the second stage.

$$\text{Acceleration (m/s}^2\text{)} = \frac{\text{Voltage} \times 1000}{\text{Sensitivity}} \quad (4.3)$$

4.6.4 Second Stage Vibration Measurements

Valve vibrations are influenced by temperatures and by external noise. Extraneous influences induce a trend drift from the zero horizontal thus the de-trended data is analysed to facilitate investigation of compressor vibration inherent features. Whilst the third segment measurements, Figure 4.7, demonstrate an approximately horizontal trend, drift is observed in the other two segments.

Examination of the vibration data over the three segments show clear breaks between the sections this being a discontinuity in measurement recording rather than a significant change in measured trend.

Closer comparison of the first and second stage vibrations over the first 0.2 seconds, Figure 4.8, shows continual noise with intermittent larger peaks which indicate opening and closing of suction and discharge valves. Reassuringly, the first and second stage vibration signals can be seen to follow the same synchronisation as the first and second stage pressures. Peak disturbance corresponding to opening and closing of valves on suction and discharge.

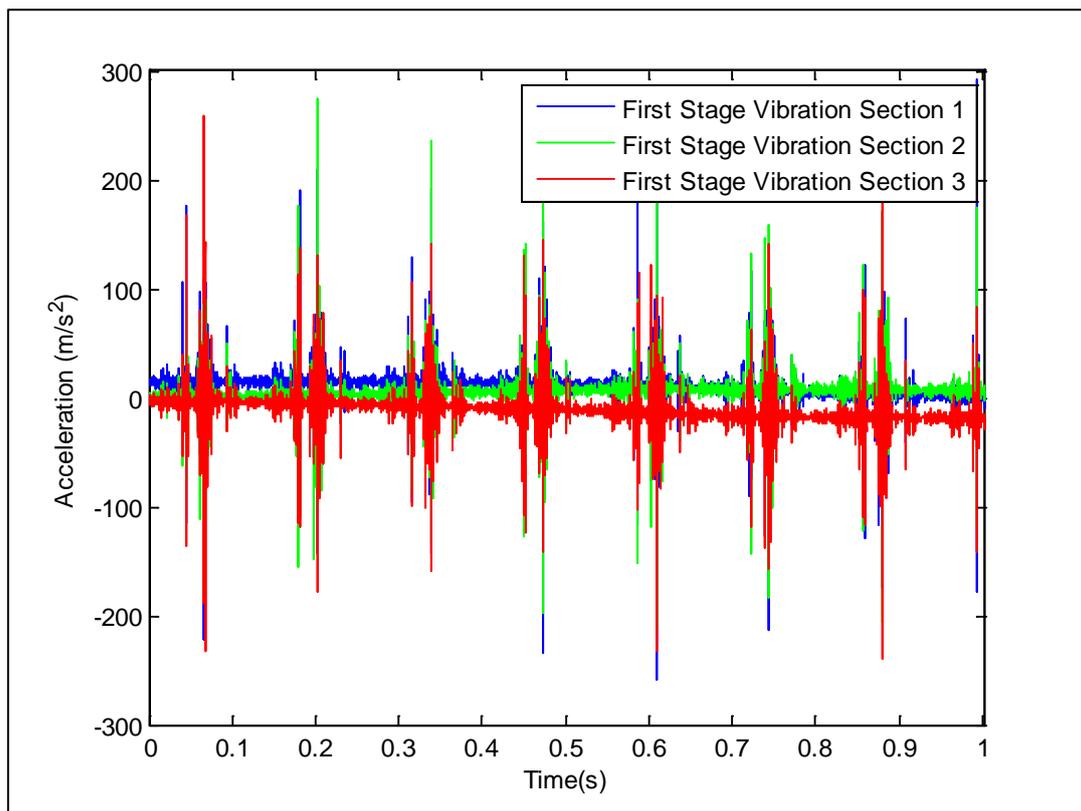


Figure 4.6 First Stage Vibration Measurements of Each Segment.

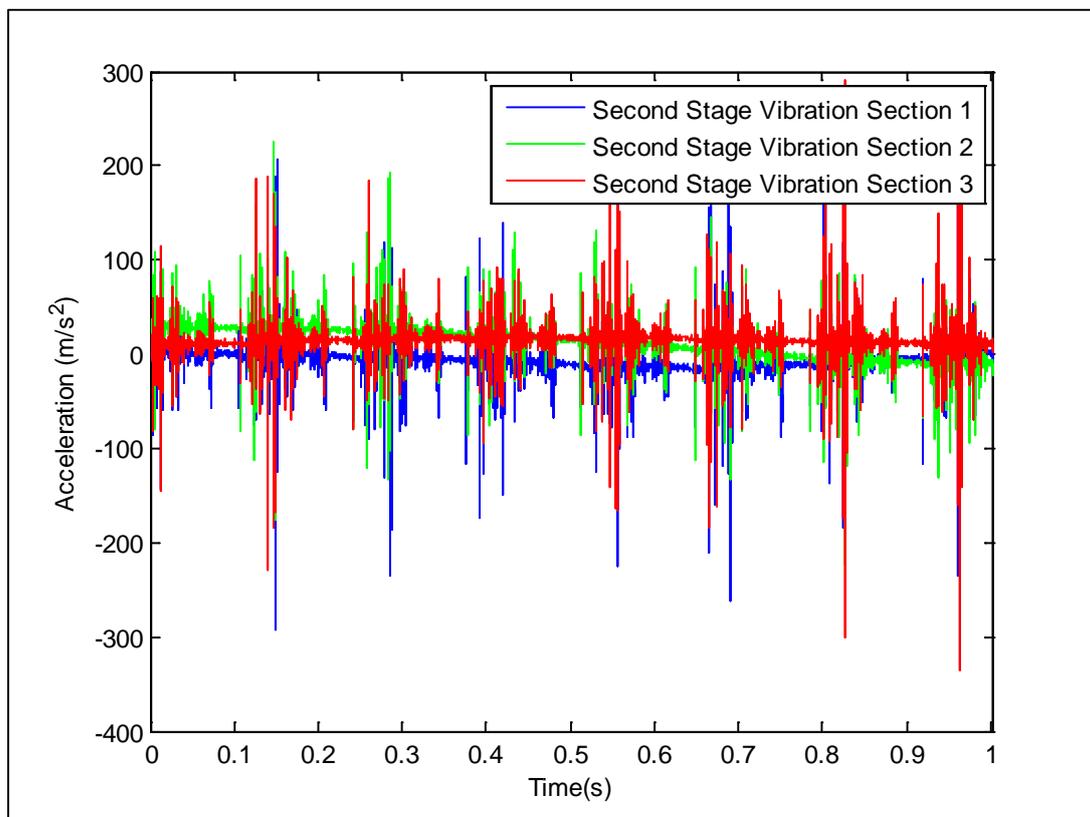


Figure 4.7 Comparison of the Second Stage Vibrations In Each Of the Three Number Segments.

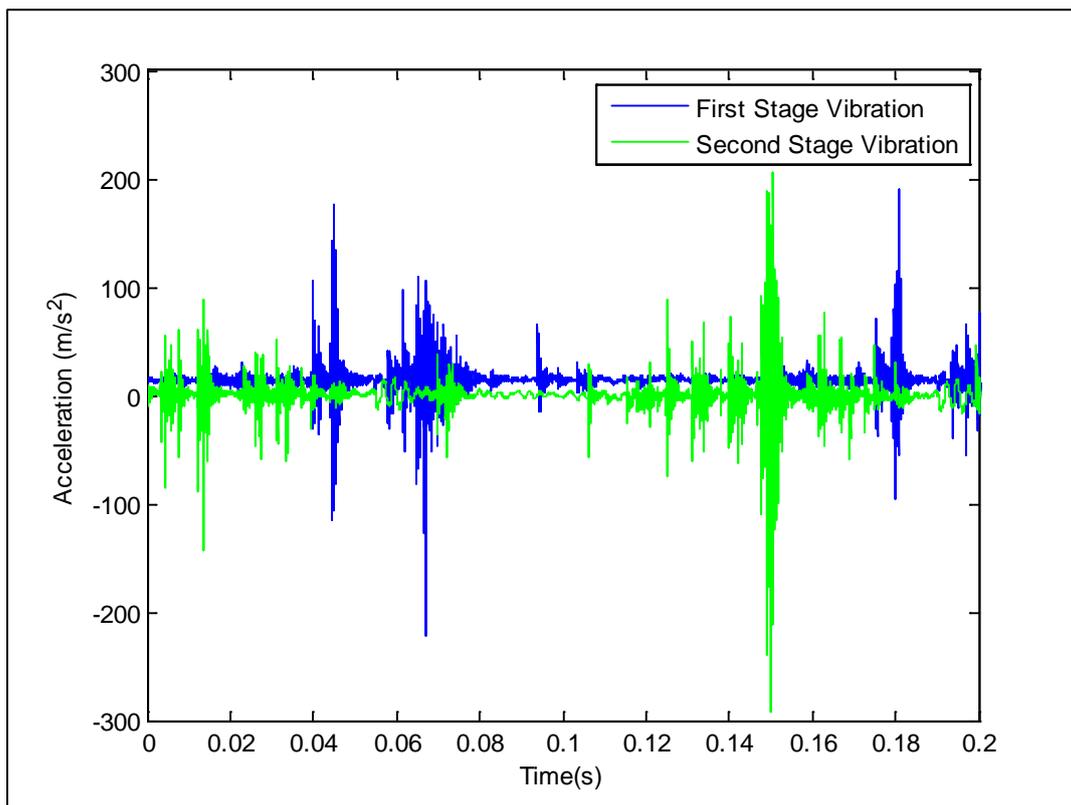


Figure 4.8 First and Second Stage Vibrations Highlighting Offset Actions of Suction and Discharge.

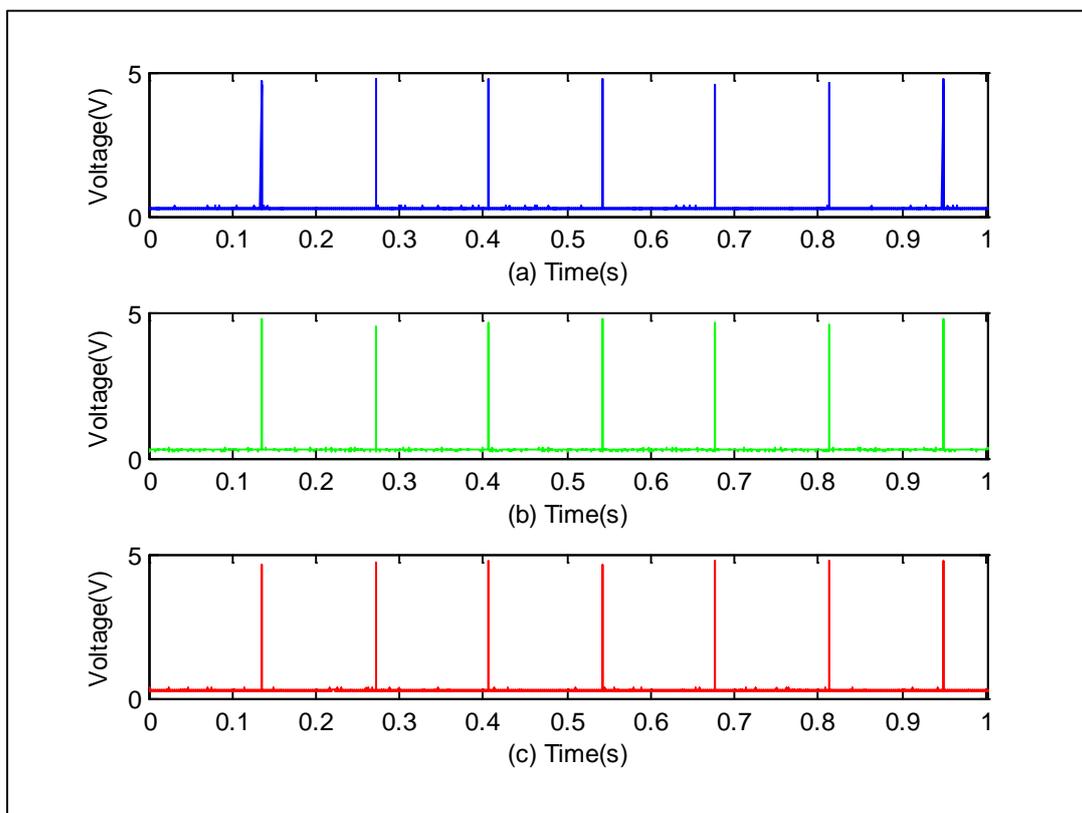


Figure 4.9 Comparison of Speed Measurements by Section Displaying Sensor Pulses. (a) Section 1, (b) Section 2 and (c) Section 3.

4.6.5 Speed Measurements

The speed sensor rotates with the crank shaft recording one pulse per revolution. From Figure 4.9 it is clear that seven revolutions per second are executed. Exact location of these pulses enables calculation of the crank shaft speed and consistency comparisons throughout the duration of operation. The pulse peaks, of approximately 4.5V, being the maximum voltage used by the machine.

To compare the speed across the period of data acquisition the times to each pulse and the average speed per section were calculated, Table 5.1. For example the time in section 1 to the first pulse was 0.1354 seconds i.e. the first revolution took 0.1354 seconds hence the speed for the first revolution was

$$\begin{aligned} \frac{1}{0.1354} &= 7.3831\text{Hz} \\ &= 442.9831 \text{ revolutions per minute (rpm)}. \end{aligned} \quad (4.4)$$

A t-test at the 5% level showed the average revolutions per minute for each section were not significantly different to 440rpm, the stipulated engine speed. The average revolutions per minute being calculated as in Equation 4.4 for each of the seven intervals between pulses.

TABLE 4.2 DISPLAYS THE MEAN REVOLUTIONS PER MINUTE AND CORRESPONDING STANDARD DEVIATIONS FOR EACH OF THE THREE SECTIONS.

Section	Mean rpm	Standard deviation
1	442.9831	0.0608
2	442.9536	0.0443
3	443.9144	0.0443

4.6.6 Angular Speed Measurements

Angular speed is measured via shaft encoder sensors which record 360 pulses per revolution. Initial inspection of the output is hampered by sheer volume of data, however, closer inspection reveals many hundreds of readings which are utilised in precisely calculating the instantaneous angular speed (IAS), Figure 5.9. IAS (rpm.)

fluctuates periodically with the load on the system. The load automatically varies throughout the compression cycle stages being lower during the suction stage and higher during compression.

4.6.7 Current Measurements

Sinusoidal output as expected is evident also exhibiting fluctuations with maximum amplitude ± 0.85 Amps, as shown in Figure 4.11. The amplitude is not constant but also varies sinusoidally. Thus a low frequency cycle is observable corresponding to the 7 cycles per second of the motor speed in addition to the high frequency current cycle at 50Hz.

4.7 Waveform Features

This section investigates the characteristics of the output variables measured and their interrelationships in a healthy compressor operating normally. Waveform feature analysis is conducted in the time domain.

4.7.1 First Stage Pressure and Vibration Measurements in Relation to Index Pulse

All output measurements exhibit about seven cycle per second fluctuation which is a direct consequence of the crankshaft speed of 440rpm (or 7.3Hz). Likewise, each of the outputs exhibit the same periodicity as seen in Figure 4.12. Features are further highlighted in Figure 4.13 where the cycle is displayed between the first two index pulses only. Positions of the index pulse mark the first complete recorded cycle. This first cycle has a duration of approximately $t=0.13s$ falling within the times $t=0.135$ and $t=0.265$ seconds. Vibration measurements can be seen to surge on opening and closing of both the suction and discharge valves. The relationship between the pressure and vibration measurements throughout the cycle is also highlighted. As the cylinder valves are opened for either suction or discharge of air the vibration signals

oscillate with an acceleration of approximately $\pm 100\text{m/s}^2$ during stage 1 and $\pm 200\text{m/s}^2$ on discharge of pressurised air.

4.7.2 Second Stage Pressure and Vibration Measurements in Relation to Index Pulse

Again at the second stage signal outputs display similar behavioural patterns albeit with increased amplitudes. Figure 4.14 and Figure 4.15 show the pressure and vibration are obviously synchronised. The amplitudes of both the high pressure and vibration are greater than those for stage one. With the air more highly pressurised the duration of both suction and discharge stages are reduced and the valves vibrate more aggressively as the air forces its way in and out.

Investigation of IAS in relationship to pressure at both the first and second stages shows clearly that the variables are synchronised with each stage in the cycle. IAS fluctuations corresponding directly to particular stages of the compression and expansion cycle of the first and second stage cylinders.

A normalised plot was considered, demonstrating a more direct comparison of the output signals, Figure 4.17. IAS at its minimum as the first stage cylinder is in the compression phase and the second stage cylinder is in its discharge phase. IAS increases to its maximum as the first stage cylinder discharges and moves back into BDC position taking in a further volume of non-compressed air; meanwhile the second stage cylinder is travelling towards TDC position and maximum pressure in the system. This cycle repeating seven times per second with the IAS returning to its minimum value as the second stage discharges.

4.7.3 Relationship between Cylinder Pressure and Reservoir Pressure

Impact of varying load (reservoir pressure) on cylinder pressure was investigated for a range of loads from 80 to 120psi (5.52 to 8.27bar), values recoded at 10 psi intervals. Both the first and second stage pressures were found to be directly proportional to the applied load. Figure 4.18 and Figure 4.19 illustrate findings with pressure increases at the second stage, which show more noticeable increases with load.

Readings taken from the first cycles of the maximum measurements for the first and second stage pressures and time (seconds) are given in Table 4.3.

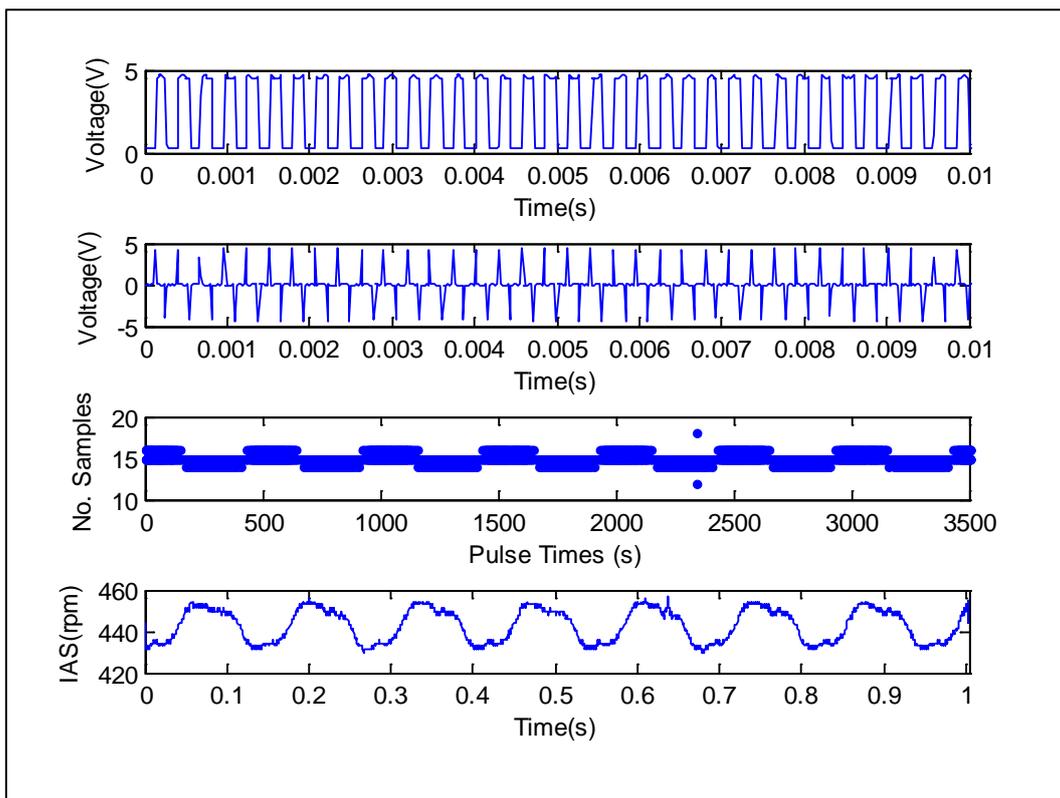


Figure 4.10 From Top: Inspection Of Pulses In First Hundredth Of A Second; The Differences Between Successive Pulses; Number Of Samples Between Successive Pulses; The Instantaneous Angular Speed Which Can Be Seen To Fluctuate Periodically.

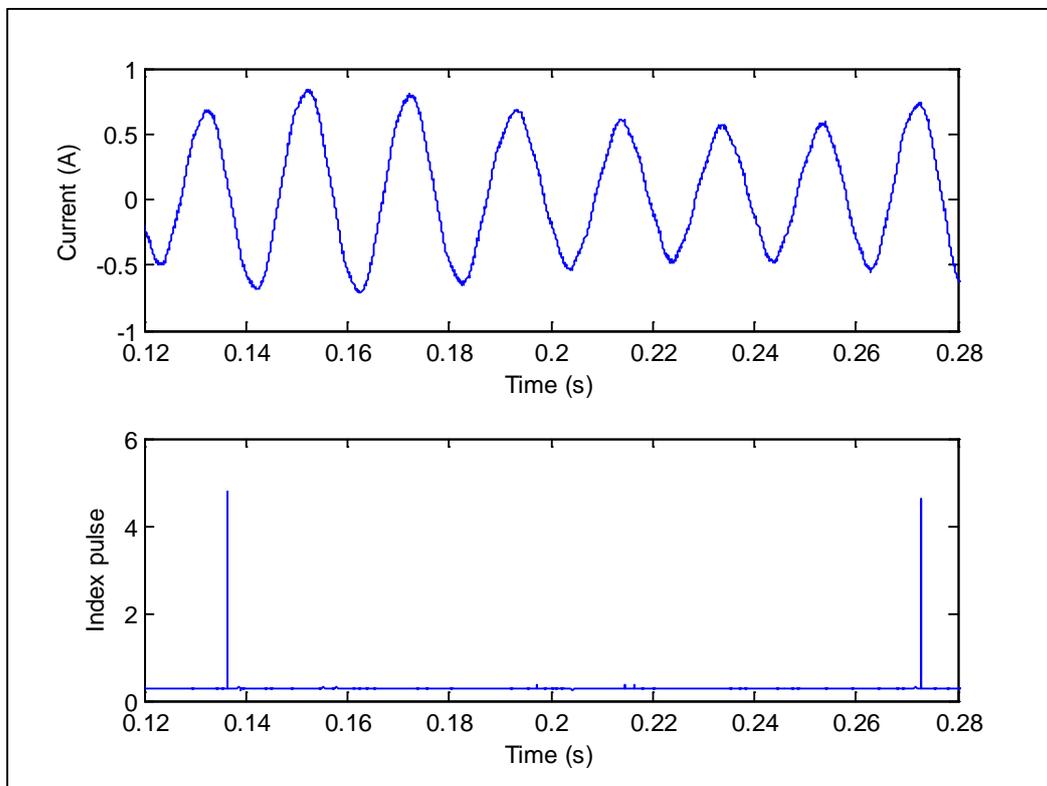


Figure 4.11 Motor Current During First Cycle.

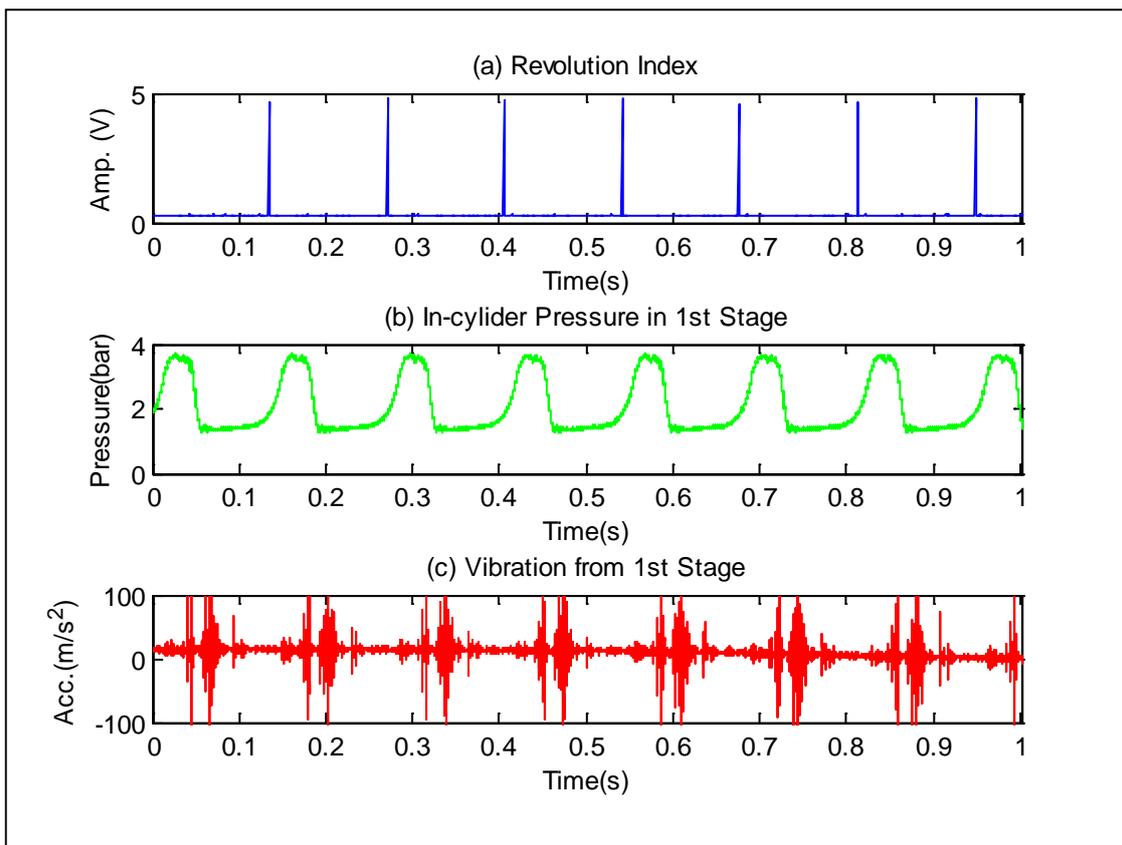


Figure 4.12 Comparison of First Stage Pressure and Vibration Signals in Comparison to Index Marker (First Section).

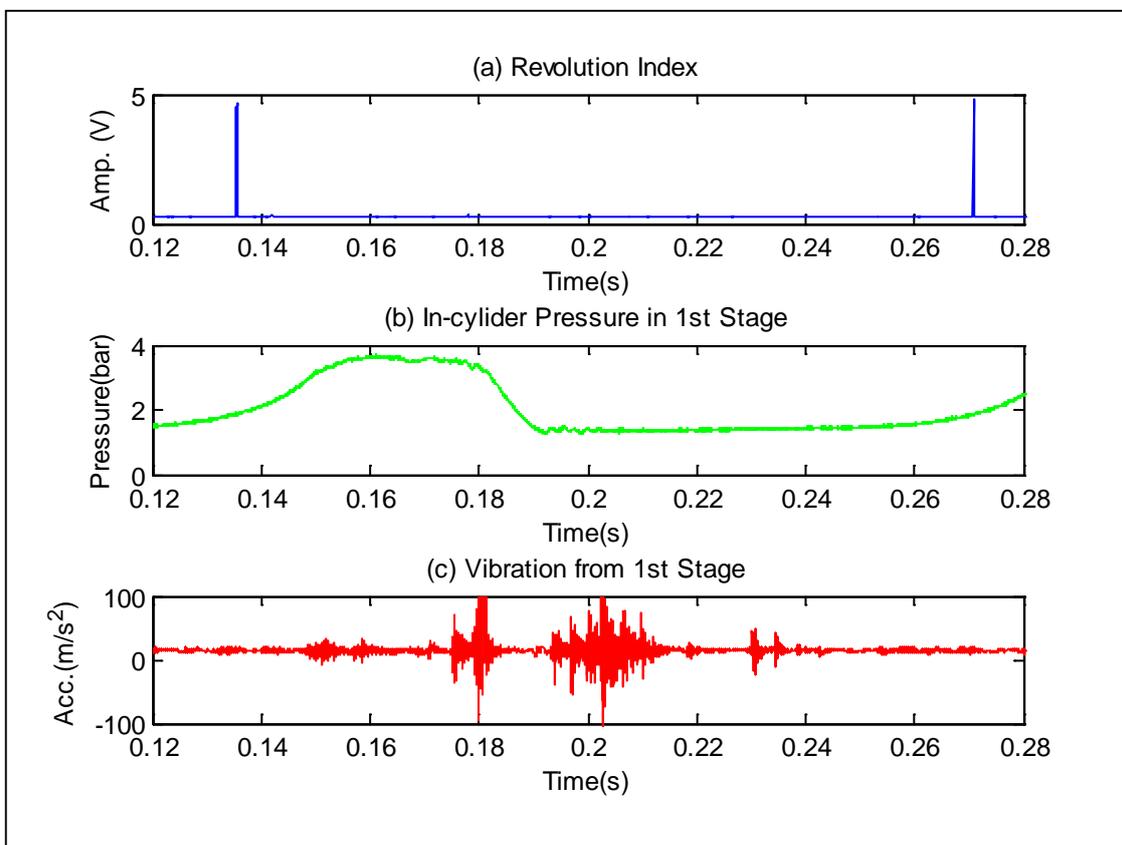


Figure 4.13 Comparison Of First Stage Pressure And Vibration Over First Cycle Relative To Pulse.

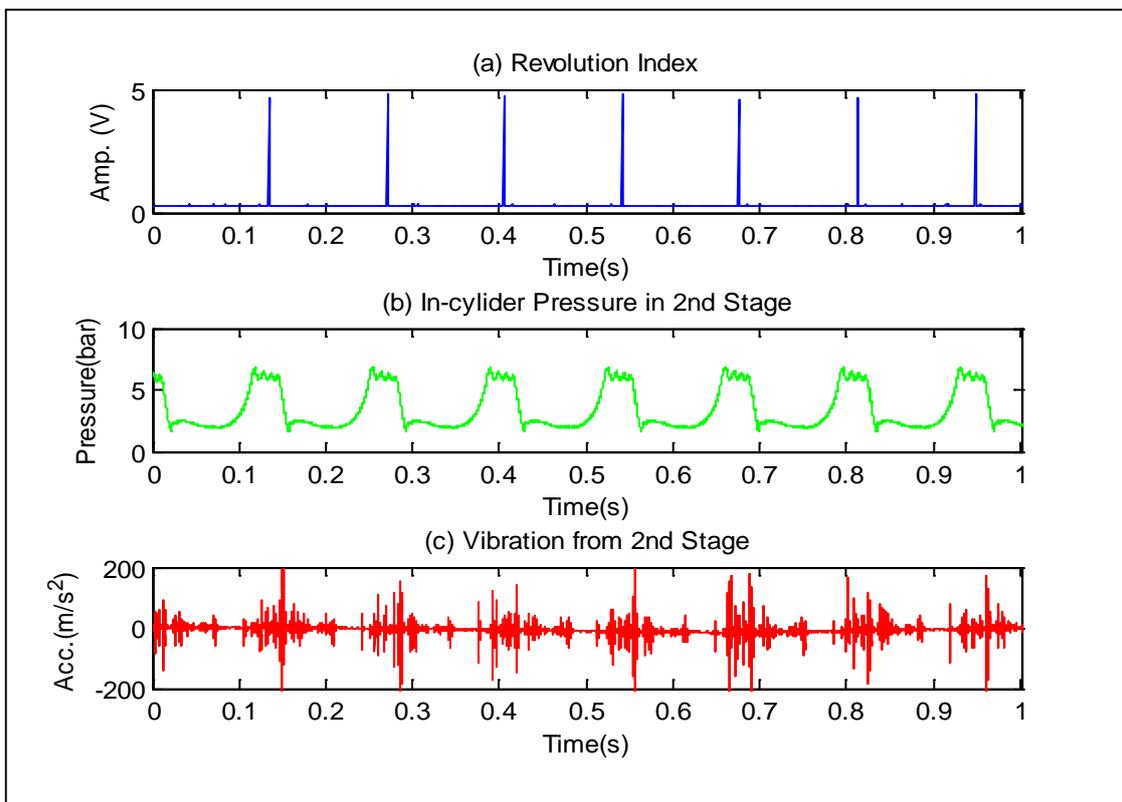


Figure 4.14 Second Stage Pressure And Vibration Cycles With Index Reference.

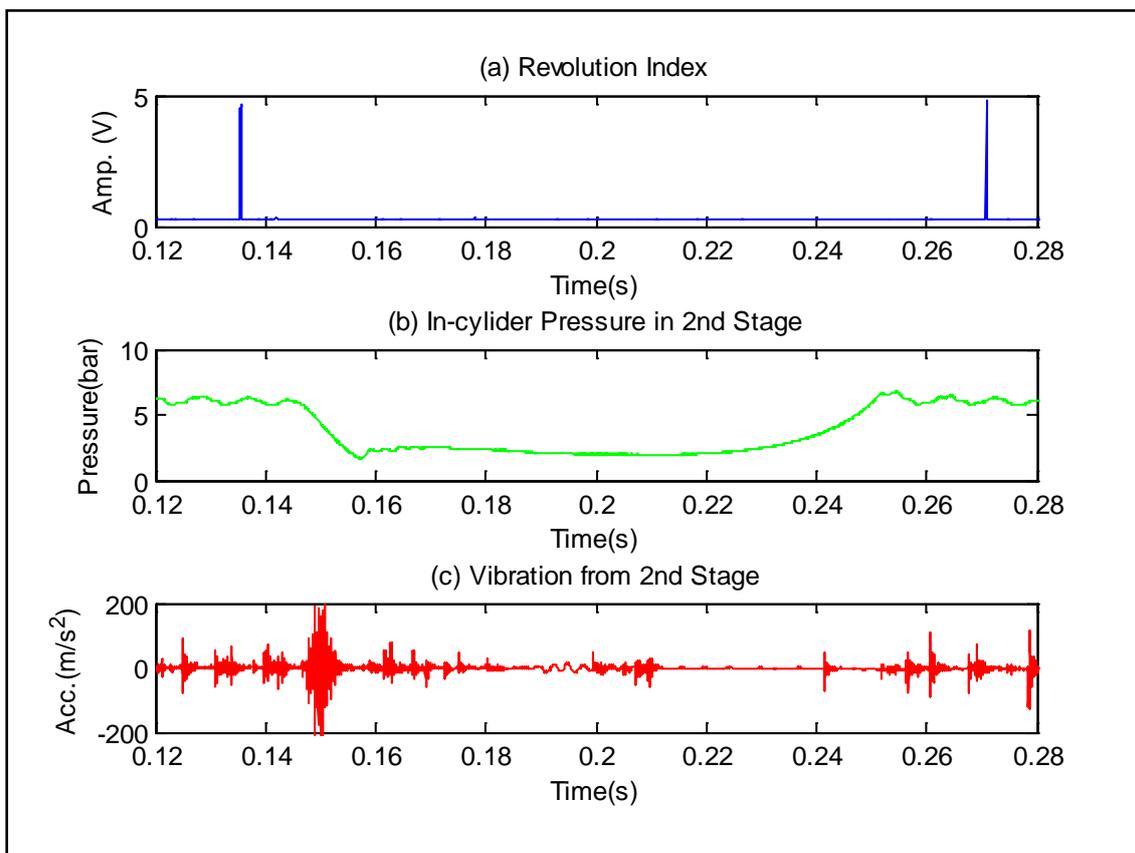


Figure 4.15 Comparison Of Second Stage Pressure And Vibration Within First And Second Pulses In Relation To Index Pulse.

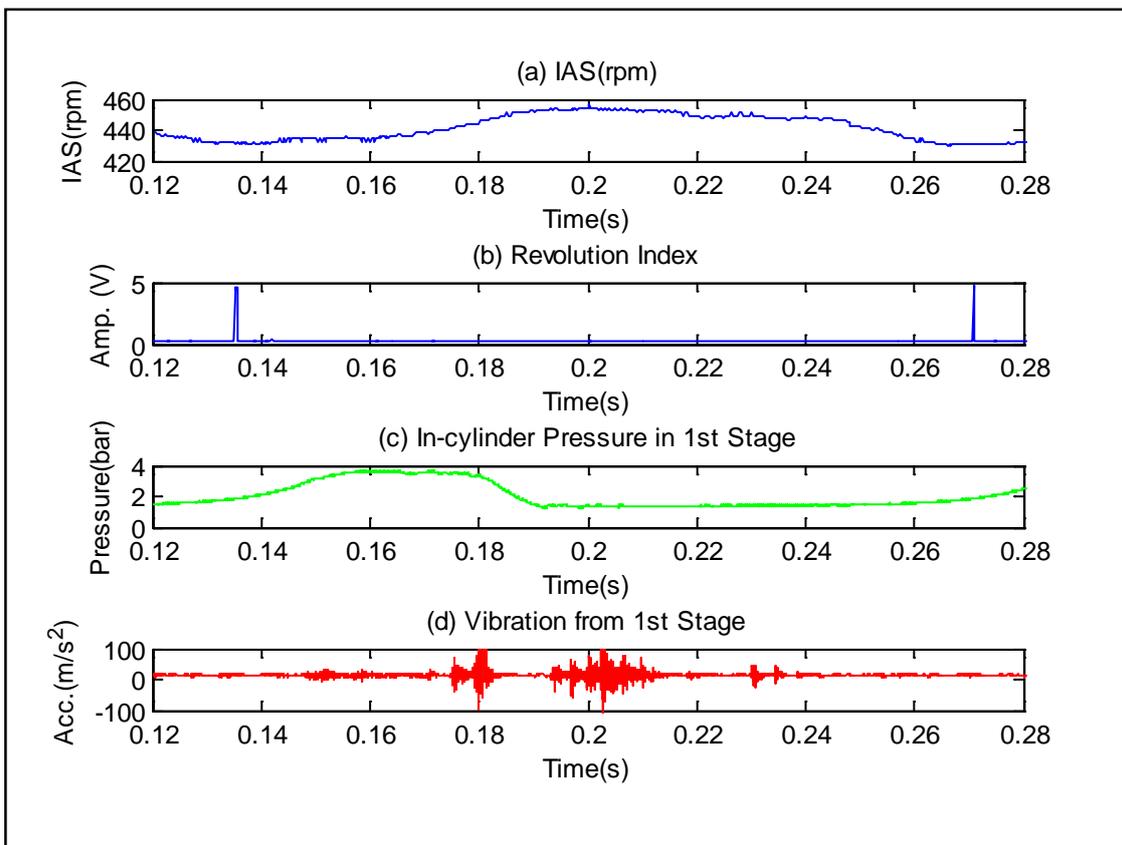


Figure 4.16 Relative Values of IAS, First Stage Pressure and Vibrations In Comparison to the Index.

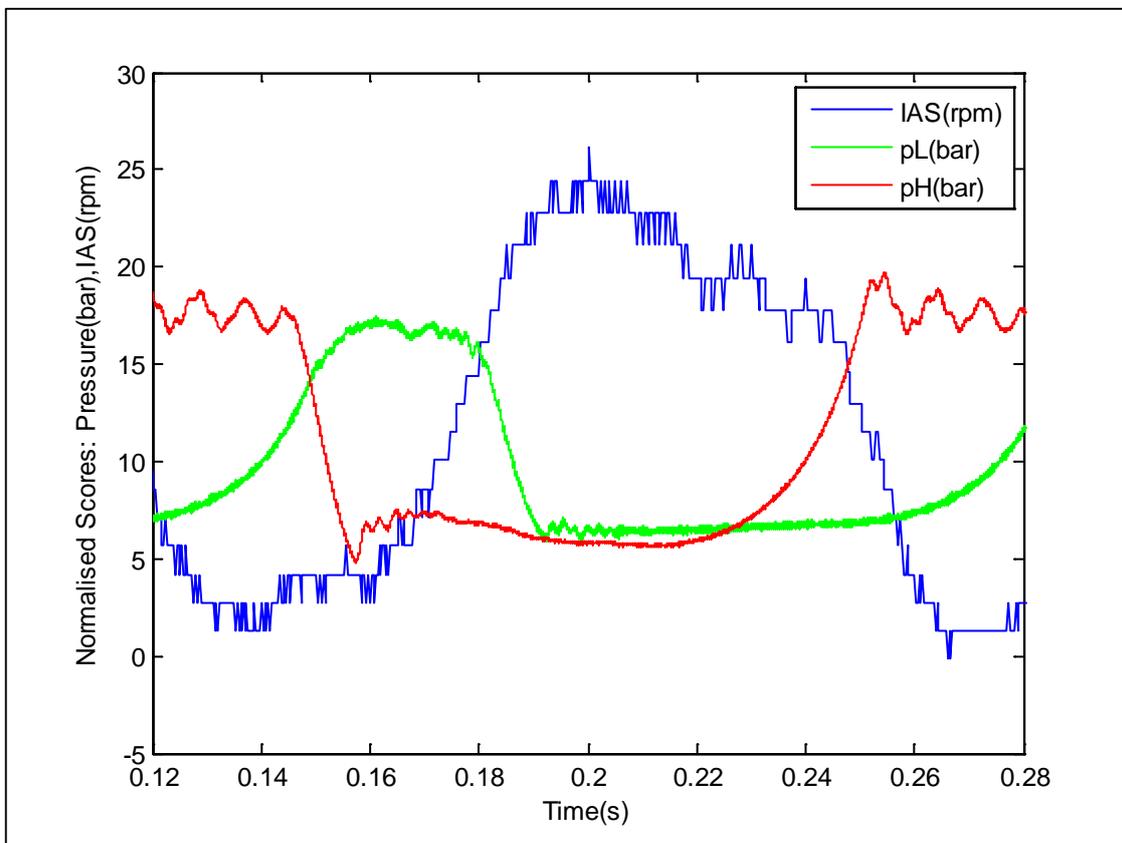


Figure 4.17 Standardised Values for Direct Comparison [IAS-430; $10 \cdot pL / \text{Mean}(pL)$; $10 \cdot pH / \text{Mean}(pH)$].

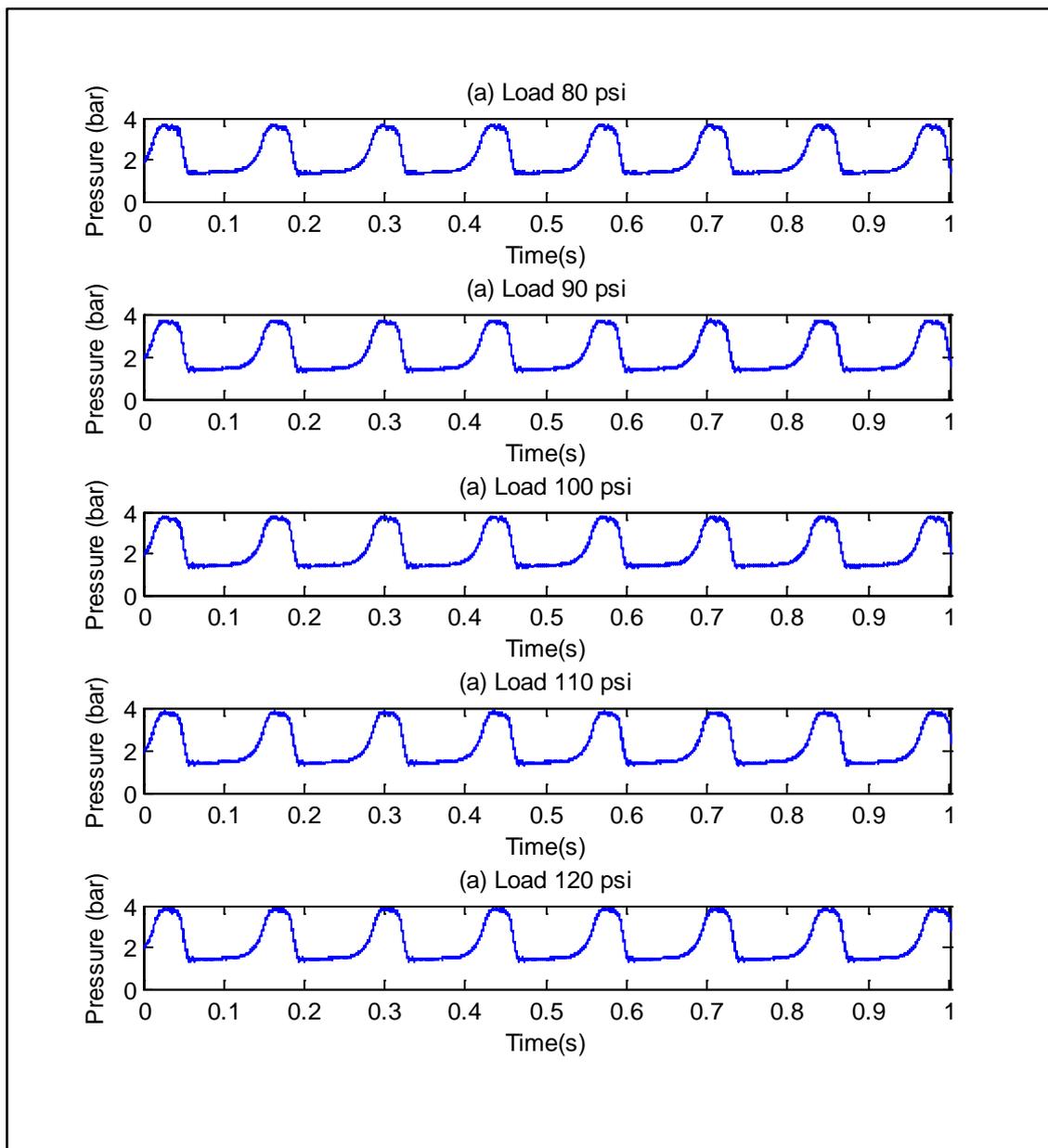


Figure 4.18 Comparison Of First Stage Cylinder Pressure By Reservoir Load. Note: 80psi=5.52bar, 90psi=6.21bar, 100psi=6.90bar, 110psi=7.58bar, 120psi=8.27bar.

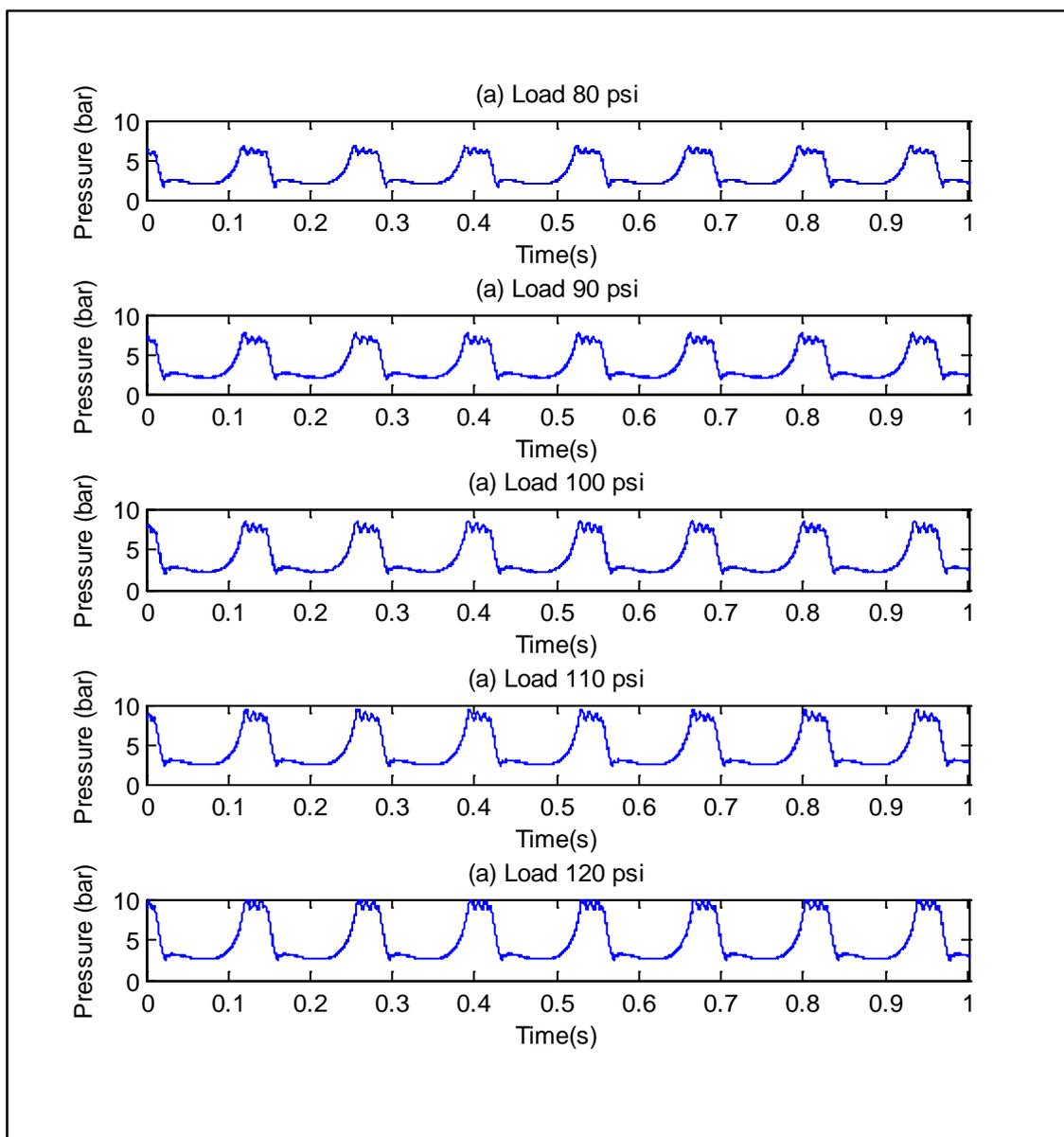


Figure 4.19 Comparison Of Second Stage Cylinder Pressure By Reservoir Load. Note: 80psi=5.52bar, 90psi=6.21bar, 100psi=6.90bar, 110psi=7.58bar, 120psi=8.27bar.

TABLE 4.3 PEAK VALUES OF FIRST AND SECOND STAGE PRESSURES.

Load	80psi (5.52bar)	90psi (6.21bar)	100psi (6.90bar)	110psi (7.58bar)	120psi (8.27bar)
<i>pH</i> max	6.6650	7.6190	8.4790	9.2960	9.9070
Time (s)	0.1167	0.1199	0.1210	0.1225	0.1222
<i>pL</i> max	3.6420	3.6670	3.7380	3.7950	3.8600
Time(s)	0.1576	0.1585	0.1585	0.1611	0.1614

Strong positive linear relationships between both the first and second stage maximum pressures and load applied are revealed.

Correlation coefficient matrix, $Pd, pLmax$, for the first stage maximum low pressure measurements with respect to load is

$$\begin{aligned}
 Pd, pLmax &= \begin{matrix} 1.0000 & 0.9914 \\ 0.9914 & 1.0000 \end{matrix} \\
 \text{Regression model} & \quad pLmax = 3.1764 + 5.64 \times 10^{-3} Pd \qquad (4.5)
 \end{aligned}$$

Correlation coefficient matrix, $Pd, pHmax$, for the second stage maximum low pressure measurements with respect to load is

$$\begin{aligned}
 Pd, pHmax &= \begin{matrix} 1.0000 & 0.9970 \\ 0.9970 & 1.0000 \end{matrix} \\
 \text{Regression model} & \quad pHmax = 0.2322 + 0.0816 Pd \qquad (4.6)
 \end{aligned}$$

Thus for unit increase in the load the first stage pressure increases by 5.64mbars and the second stage pressure increases by 0.0816bar.

Both models demonstrate near perfect positive correlation between the applied load and the pressure with goodness of fit statistics of almost 100% implying variations in the pressure can be almost entirely explained by variations in applied load. Thus confirming the systems efficiency.

4.7.4 Relationship between Speed and Reservoir Pressure

The motor speed is unaffected by the load applied as illustrated in Figure 4.20, the seven cycle pattern being identical for each of the five loads applied. Hence the crank shaft speed and so the index position is comparable for all signal outputs throughout the compression cycle. Direct comparisons are thus feasible.

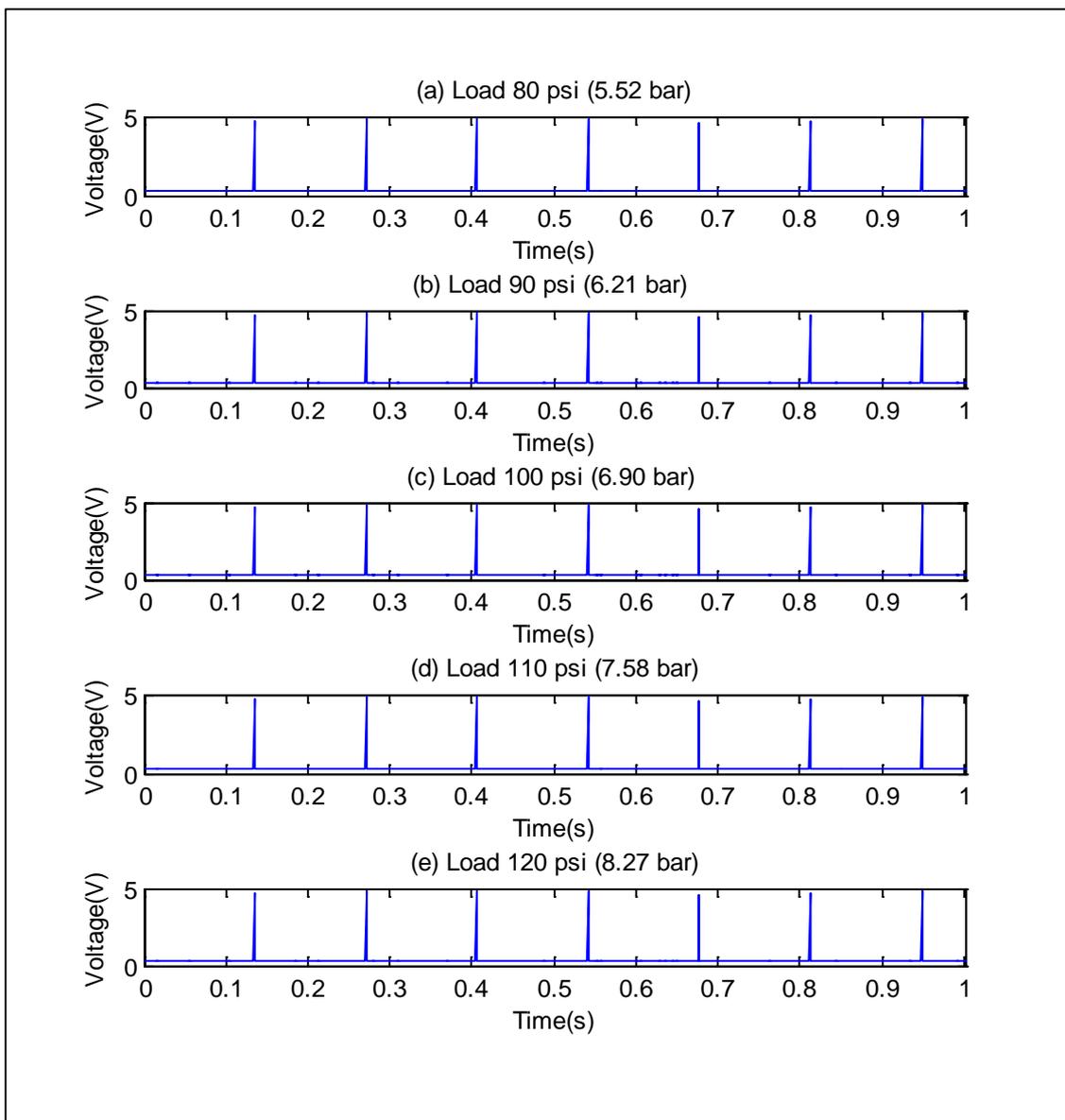


Figure 4.20 Speed of Motor under Varying Loads (psi). Note: 80psi=5.52 bar, 90psi=6.21 bar, 100psi=6.90 bar, 110psi=7.58 bar, 120psi=8.27 bar.

Chapter 5

Exploratory Data Analysis

Extended experimentation incorporating system faults being similarly analysed emergent blueprints for system health are presented.

Elementary models in both the time and the frequency domains are considered and show promising fault detection potential.

Lastly the research prompting the original motivation for much of this thesis work is presented and appraised.

5.1 Fault Classification Using Statistical Modelling and Graphical Means in the Time Domain

This section focuses on the classification of fault types by identifying abnormal behaviour patterns in the time domain.

Prior to any complex analysis or model building it is essential to gain insight into the data and its distributional patterns. Exploratory data analysis through visual or graphical displays offers global inspection and an opportunity to identify and further investigate any anomalies. Variable profiling in the form of a simple time series plot of measurements by classification group allows direct comparisons. Analysis of output signals across a range of faults with a view to identifying any characteristic features which may best describe behaviour so providing a means to identify faults.

Figure 5.1 shows the pressure at the first stage by fault type (load fixed at 100psi = 6.90bar). Whilst the periodicity is consistent evidently there are some striking visual differences in the amplitudes of these measurements across the five classes. A leak in the inter-cooler system would appear to be easily distinguishable from the healthy fault free case. The first stage pressure in the faulty intercooler simulation has a far narrower band of values (max +0.9184bar, min -0.4243bar) than in the healthy case (max +1.7182bar, min -0.8873bar). For the loose belt and leaky suction valve faults there is little visible difference in comparison to the healthy case whereas more extreme amplitudes (-0.9987, 2.0621) were observed when the discharge valve leaked. Although the simple time series plot of the variables provides a useful check of signal properties with respect to frequency and amplitude it does not of course provide a robust method for fault identification in itself.

Figure 5.2 gives a magnified view of the first cycle and clearly shows that whilst first stage pressure readings for the SVL and LB are little removed from those of the healthy system there are obvious differences in the case of both DVL and ICL. For the DVL the pressure exceeds its healthy maximum by 0.5bar (healthy maximum 4.0bar, DVL 4.5bar) whereas with an ICL the maximum pressure achieved is barely greater than ± 0.5 bar. In addition the corresponding ICL pressure trace continues to be considerably closer to zero than those of the other classes.

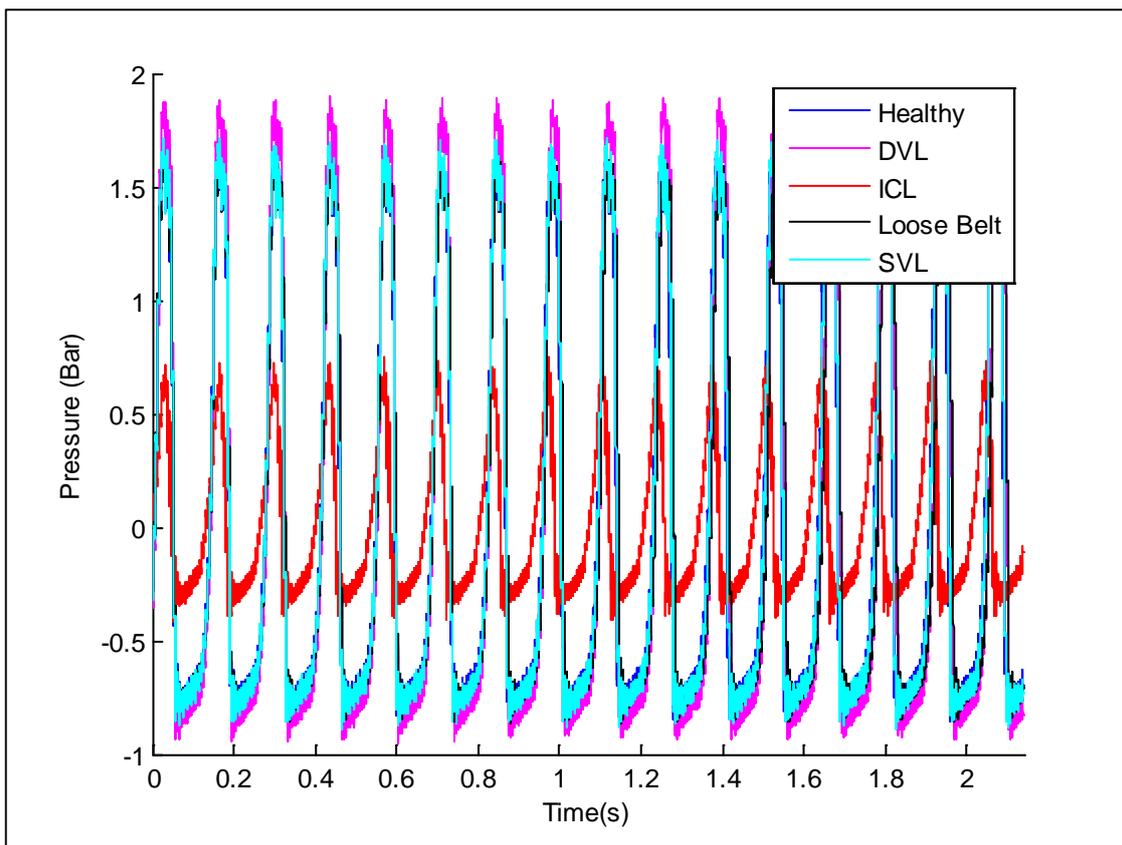


Figure 5.1 Healthy and Faulty Signals Comparison: First Stage Pressure Measurements (load 100psi).

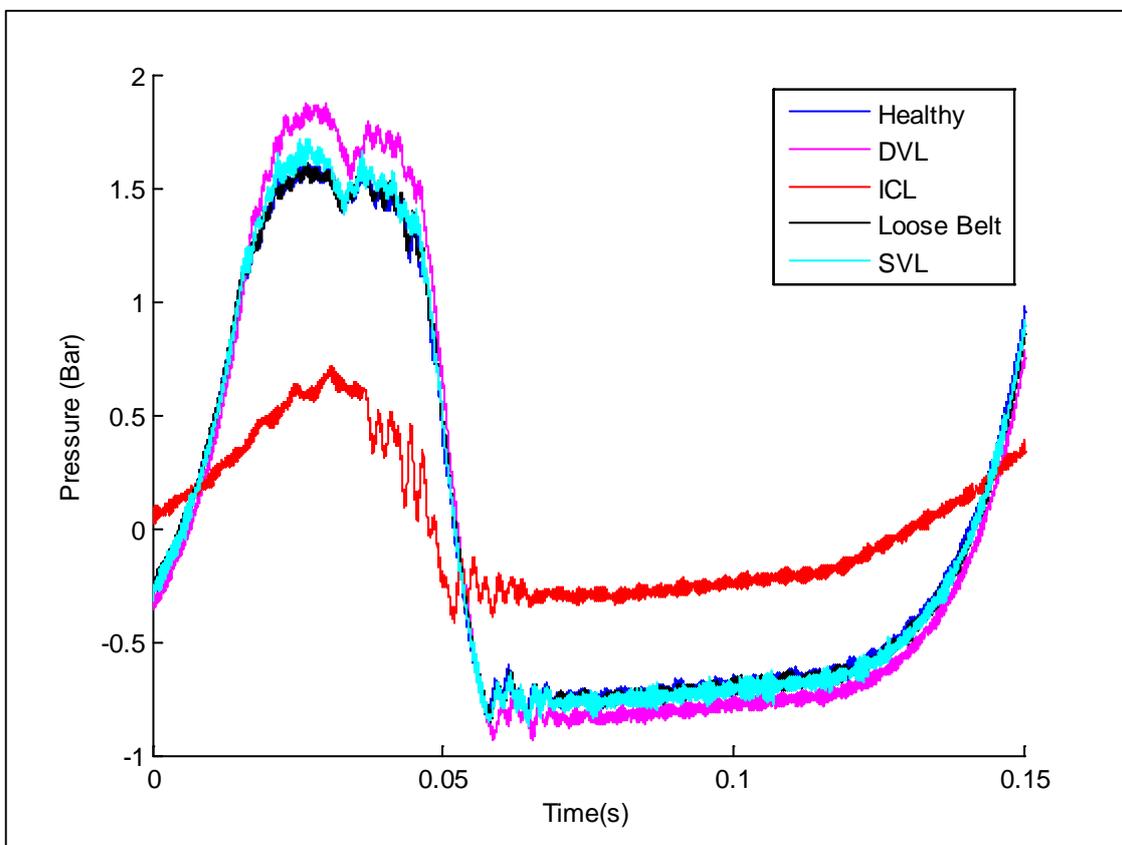


Figure 5.2 First Stage Pressure for Each of the Five Single Fault Cases.

Inspection of data distributions is essential prior to numerical analysis. Distributions by fault are illustrated by the multiple boxplots Figure 5.3. All distributions are clearly negatively skewed; again it is the case of the faulty intercooler seal which is most obviously different. ICL median value is also significantly different to the other four groups. Having detected a deviation from the norm during operation accurate diagnosis of the source is critical and the more extreme the fault characteristics the easier to pinpoint the fault. Obviously the more timely and precise the diagnosis of a fault, the greater the opportunity for planned maintenance potentially reducing the impact of disruption due to need for emergency intervention. Symmetry, location and spread of the distributions are immediately comparable on inspection of the boxplot series and should the pressure at first stage not exceed one bar a leak in the intercooler is highly probable. Profile and box plots are ideal for identifying pattern differences between fault distributions and highlighting rogue measurements but may not provide criteria for fault identification and a more comprehensive rule is necessary if each fault is to be accurately identified.

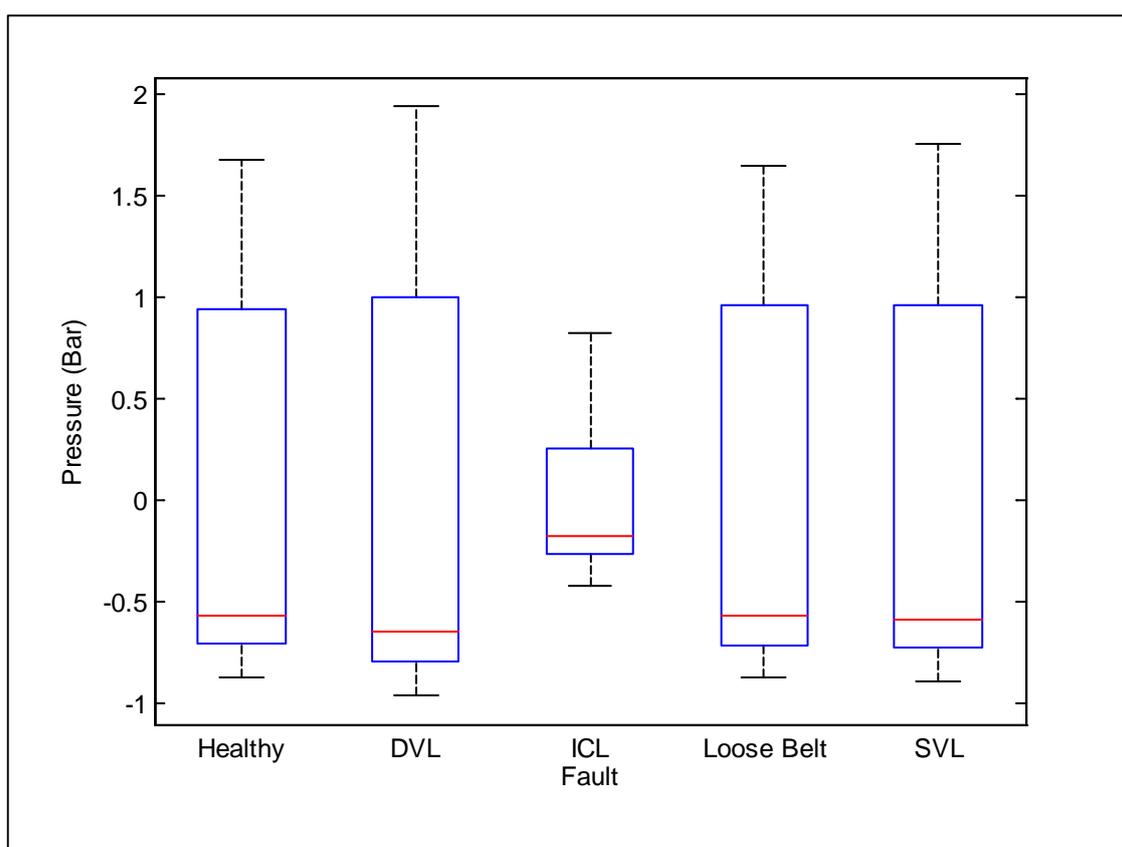


Figure 5.3 Boxplots displaying first stage pressure characteristics by fault (load 100).

Characteristics of each fault type were compared and contrasted by their summary statistics, Table 5.1.

Figure 5.4 shows a clustering of the mean pressure values for each of four data segments plotted against corresponding kurtosis values. It can be seen that each observation could be classified with relative ease to the correct case by defining suitable kurtosis boundary values. The greatest difficulty will arise in separating those with a SVL from the healthy case since both the mean and kurtosis values for these two groups are very similar. A rule of the form:

Kurtosis < 1.9 indicates a loose belt;

Kurtosis >2.2 implies an intercooler leak;

Kurtosis = 2.0 ± 0.01 suggests a DVL (5.1)

TABLE 5.1: SUMMARY STATISTICS FOR FIRST STAGE PRESSURE MEASUREMENTS (LOAD 100PSI=6.90BAR).

Low pressure	Healthy	DVL	ICL	Loose Belt	SVL
Median	-0.5866	-0.6751	-0.1764	-0.5845	-0.6049
50% range	[-0.7277, 0.9698]	[-0.6751, 1.0221]	[-0.2650, 0.2515]	[-0.7340, 0.9926]	[-0.7536, 0.9848]
Min	-0.8873	-0.9987	-0.4243	-0.8886	-0.9176
Max	1.7182	2.0621	0.9184	1.6930	1.8071
Mean	-8.7413e-015	3.7512e-015	5.2778e-015	1.1843e-014	1.3790e-014
RMS	0.9287	1.0833	0.3256	0.9295	0.9618
Kurtosis	1.9229	2.0024	2.2163	1.8873	1.9310

To adequately separate the healthy cases and those with leaking suction valves the overall kurtosis values for the groups 1.9229 (healthy) and 1.9310 (SVL) could be used with suitable cut off values for the mid-range. Thus allocate to the healthy group if $1.9 < \text{kurtosis} < 1.9270$ and for $1.9270 < \text{kurtosis} < 1.95$ allocate to the leaking suction valve group. It should be noted that this rule applies only in consideration of the first stage pressure with the load restricted to 100psi.

Since the distribution of the vibration variable is considerably different to that of the first stage pressure this simple rule could not be directly applied in this format.

Although some clustering is evident, Figure 5.5, groups are not sufficiently separate to assign cases one or two-dimensionally based on mean and kurtosis measurements. Closer inspection of the summary statistics of the first stage vibration measurements and the upper and lower warning bands (calculated at $\bar{x} \pm 2\sigma$) indicate significant differences in the width of the warning bands with the DVL giving rise to far greater range of measured vibrations. Again this in itself proved insufficient to devise a rule for separation of each of the fault types from the healthy case.

Partitioning pL measurements by number segment proved effective in determining an elementary rule for assigning to classes. However, this technique was ineffective using vL signals. Although plots of segment means against kurtosis showed case clustering for both variables, the technique would not be robust should a sequence of faults develop. On the other hand clustering algorithms based on the raw data would, if computationally possible, put too great a burden on resources. Data and variable reduction techniques such as PCA provide alternative means of scrutiny. Rules for group allocations via multivariate analysis of the variables as a whole offering greater model sophistication. The most effective parameter combinations being determined to distinguish between individual faults and fault combinations. Chapters 6 to 8 explore the potential.

Establishing a statistically robust model often requires strict adherence to the underlying model assumptions particularly in terms of data distributions. Rarely does experimental data bear close resemblance to the normal distributions assumed in parametric statistics. If these deviations are slight a useful model may still be developed, however, great care should be taken to assess model accuracy through appropriate residual analysis. Figure 5.6 highlights the distributional differences between fault runs even for apparently identical simulations.

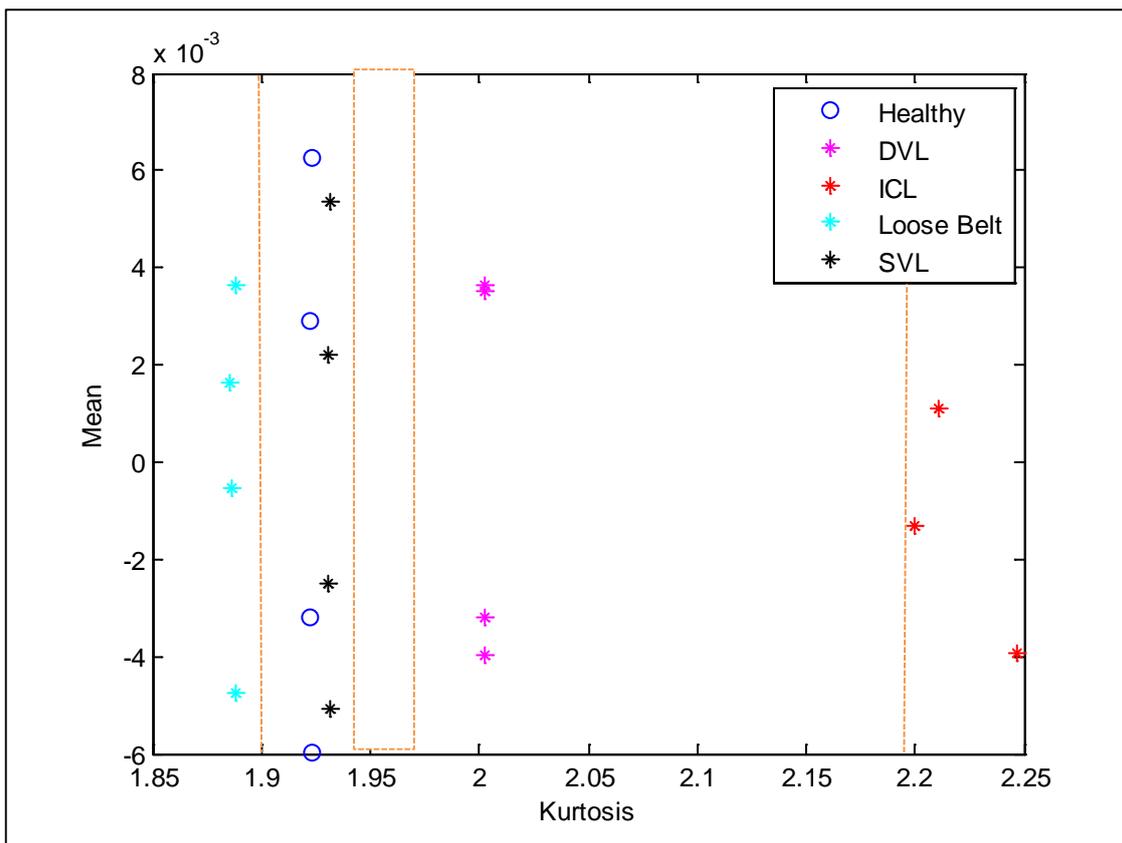


Figure 5.4 Scatter Plot to Show the Mean Vs Kurtosis of First Stage Pressure Per Number Segment.

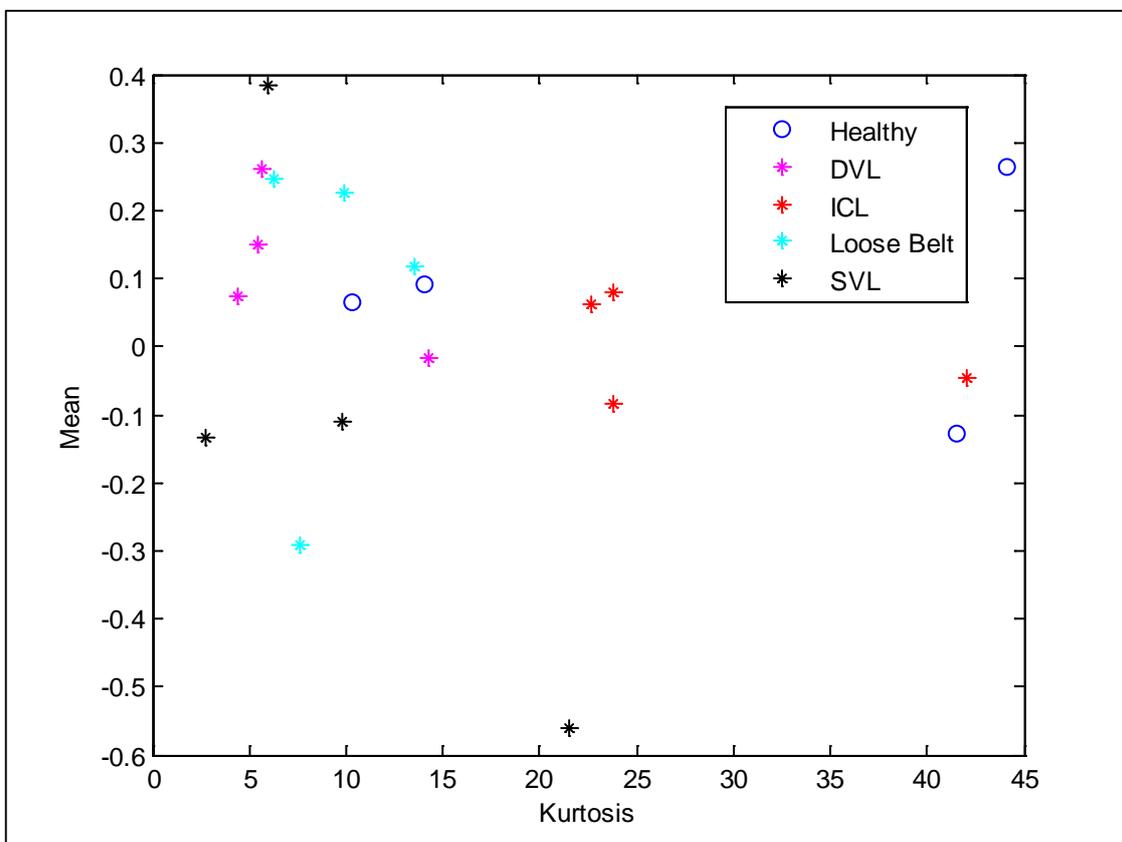


Figure 5.5 Plot of Mean and Kurtosis for the Vibration Measurements at First Stage with Load 100psi (6.90 bar).

TABLE 5.2 SUMMARY STATISTICS FOR FIRST STAGE VIBRATION (LOAD 100).

First stage vibration measurements	'Healthy	DVL	ICL	Loose belt	SVL
Mean	0.0691	-0.0349	0.0080	-0.0638	0.0413
RMS	0.2430	0.3235	0.2479	[0.2605	0.2338
Kurtosis	5.4286	2.8116	16.1214	3.2880	4.0229
Lower limit/ warning band	-0.4169	-0.6819	-1.1338	-0.5130	-0.4263
Upper limit/ warning band	0.5551	0.6121	0.1422	0.5290	0.5089
Difference between warning bands	0.9720	1.2940	0.9916	1.0420	0.9352

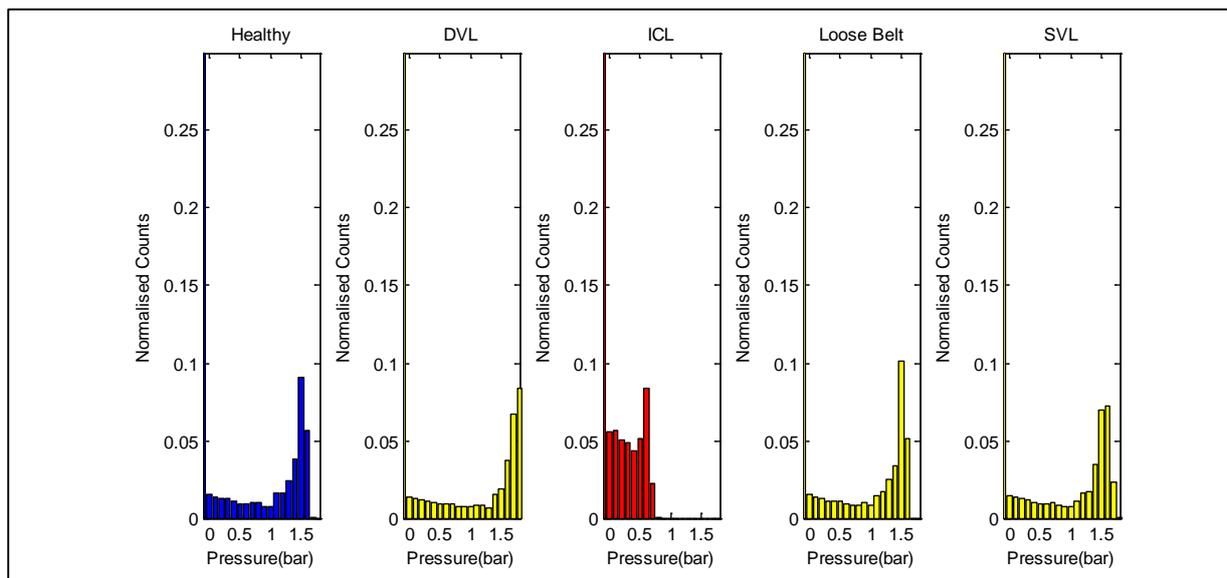


Figure 5.6 Class Distributions in Healthy Compressor Rig.

5.2 Signal Analysis in the Frequency Domain

The fast Fourier transform (FFT) of the envelope signals from vibration sensors were calculated by applying 87835 point filter with Hanning windows. The magnitude of the FFT being taken as the amplitude or FT spectrum in the spectrum analysis. Envelope spectra show only the amplitude profile of original signals and so provide a clearer insight into underlying behaviour of the compressor [45, 46, 88 and 104]. Measurements for the seven variables were recorded for each of five machine conditions, one healthy and four with seeded faults. Variable measurements included the vibration at the first and second stage of the compressor cycle. A demodulated signal was considered to 32 harmonics by passing through a Hilbert filter and applying an envelope technique, the *env1* MATLAB function.

5.2.1 Frequency Spectra of the First Stage Vibration Signals

Figure 5.7 and Figure 5.8 illustrate the frequency spectra of the first stage RC vibration signals. By plotting the spectra of all five classes simultaneously they can be compared directly to reveal their unique signal characteristics. The harmonics, displayed by the tick marks at 7.3Hz intervals, clearly pick out the subsequent signal peaks which are coincident for all faults. Hence, once rules are assigned, these signal characteristics can be used to differentiate between healthy and faulty systems and so identify different faults.

The spectrum plots clearly show that the first fundamental frequency, due to the shaft rotation speed per second, occurs at 7.3Hz regardless of fault type. Subsequent peaks are at integer multiples of this fundamental frequency.

Whilst positioning of the fundamental frequency and its higher order harmonics are the same across all classes, the spectra display differing amplitudes. In particular the amplitude of the ICL signal is less than half that of all other signals at the first harmonic. This could provide an effective means to differentiate this fault from the others. However, it is difficult to see substantial differences between others in these original spectra, therefore a universal rule is unattainable.

To identify clear differences between all other faults, the amplitudes of their harmonic components up to the 32nd order were extracted. Focus on fault amplitudes at specific harmonic orders offers insight into machine condition. Thus even with limited data input, rules are established to provide a means for efficient exploration.

Figure 5.8 highlights two major spikes in the first stage vibration signal spectrum, the first at the 18th and the second at the 22nd harmonic component. Each class amplitude being visibly distinct at the 22nd harmonic, the first stage vibration signal frequency spectrums would appear to offer a means of distinguishing between machine states at this point. Differences in class amplitudes would be statistically significant if each amplitude variance be sufficiently low that the confidence intervals are non-overlapping. However, this is not the case for the ICL which clearly displays amplitudes very similar to the healthy signal. Harmonic spectra from vibration signals are obviously information rich. Amplitudes also visibly varied with load throughout the cycle. Although interesting this is not currently relevant the aim of this research being to provide a non-intrusive diagnostic for each of the considered simulated faults.

5.2.2 Frequency Spectra of the Second Stage Vibration Signals

The spectra for the second stage vibration signals, Figure 5.9, give similar results to those at the first stage. But it seems that by using the values at the first harmonic component it is possible to detect differences between all classes.

For each of the five cases the mean amplitude per harmonic and associated standard deviations were calculated. Thus 95% confidence intervals of the spectrum amplitudes at the first harmonic were calculated per class, Table 5.3. Clearly the ICL amplitudes are significantly different to all other cases. Note the confidence interval for ICL is not overlapping with those of any other distributions.

Confidence intervals of spectrum amplitudes at other notable harmonics were investigated, however, despite apparent visual differences at particular harmonics only the ICL was found to have a significantly different distribution at the 5% level.

5.2.3 Envelope Spectrum Analysis of Vibration Signals

As shown in section 5.2.2, spectral analysis of the original signals is not sufficiently sophisticated to differentiate between different classes. Therefore, to pursue accurate identification of machine faults more advanced analysis is required.

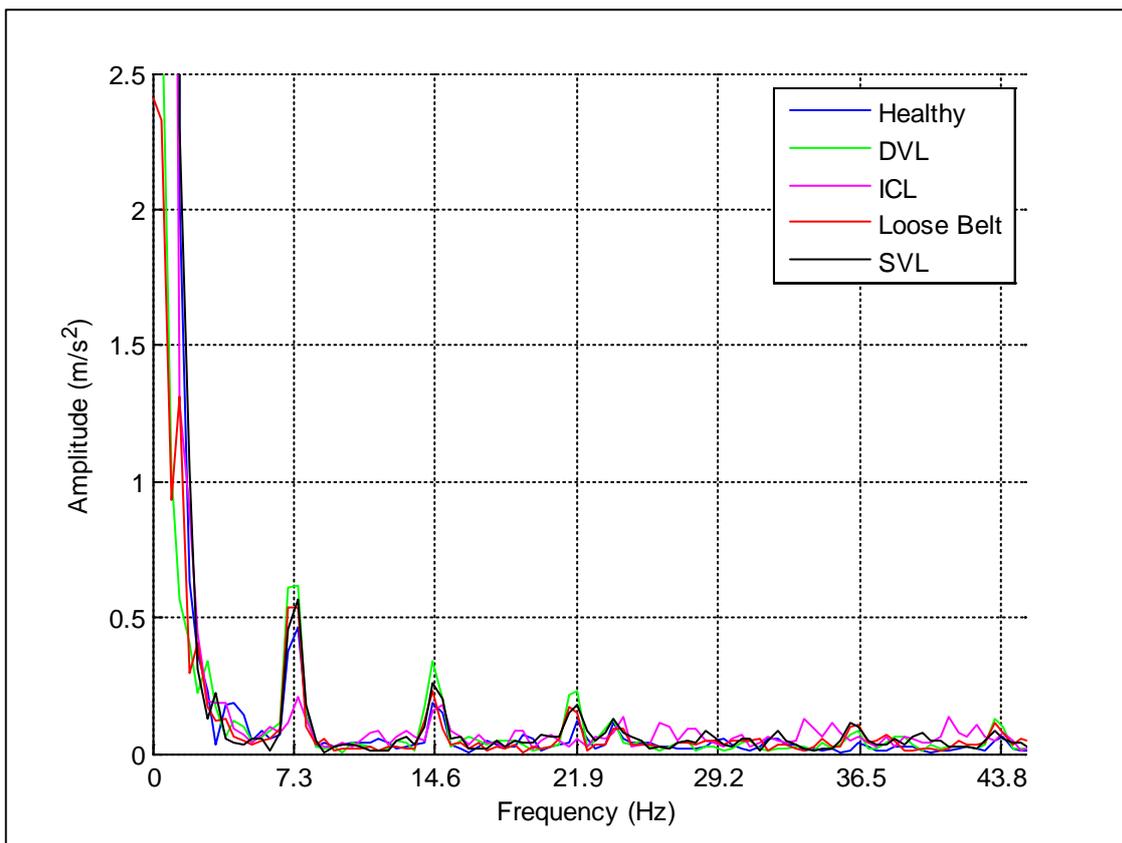


Figure 5.7 Spectrum Plot: Fundamental Frequency For Each Machine State, First Stage Vibration Signal.

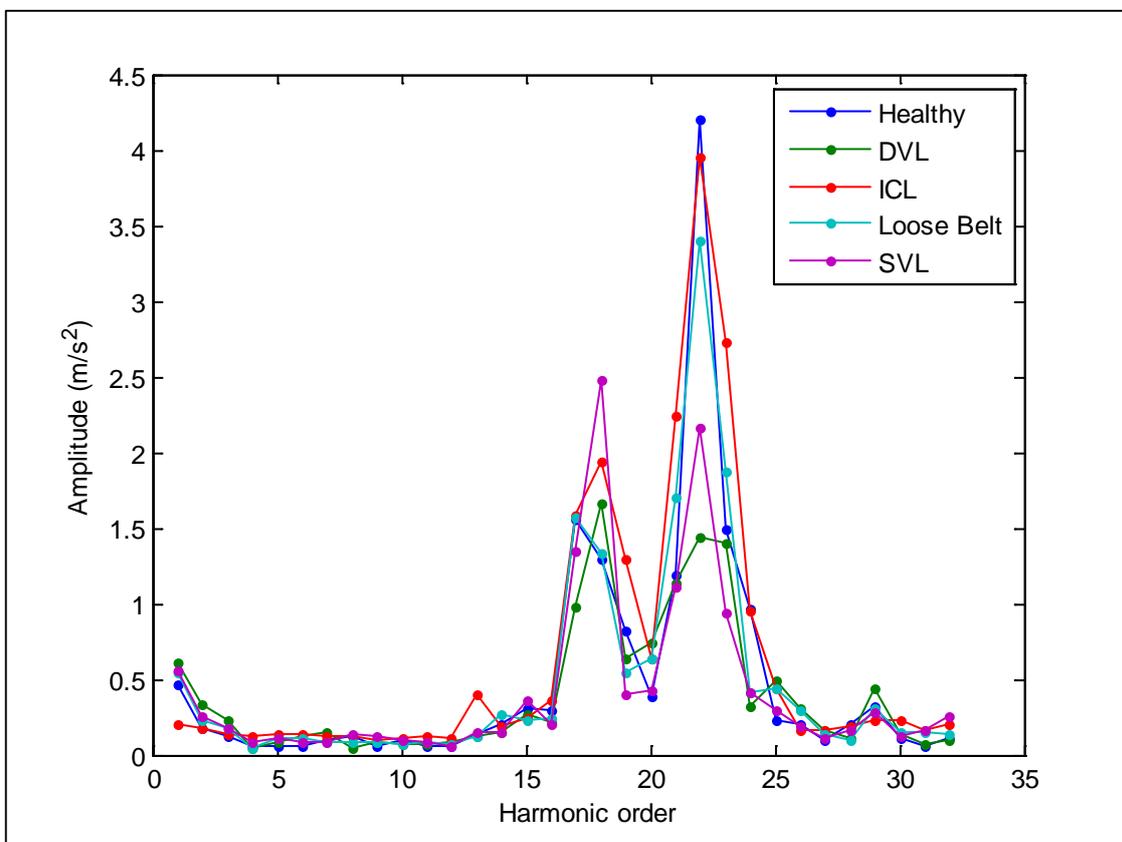


Figure 5.8 Frequency Spectra for First Stage Vibration Measurements to 32nd Harmonic (Fundamental Frequency 7.3Hz).

Through spectrum analysis it was found that the main changes occur at the operating frequency and its high order harmonics thus envelope analysis is used to enhance this feature.

Envelope spectra, as previously stated, show only the amplitude profile of original signals and so provide a clearer insight into the underlying behaviour. A demodulated signal is considered to 32 harmonics by passing through a Hilbert filter and applying an envelope technique, env1 Matlab function. Signal variations due to noise are filtered out leaving variation due to machine health only. Once the problem becomes too serious nothing is detected but noise, past the point of useful demodulation.

Envelope spectra of the first and second stage vibration signals are displayed in Figure 5.10 and Figure 5.11 respectively. Envelope spectra at the second stage with amplitude labelling at decisive harmonics are illustrated in Figure 5.12.

TABLE 5.3 SUMMARY OF FREQUENCY SPIKES AND AMPLITUDES BY FAULT TYPE,

	First major spike at the 18 th harmonic	Second major spike at the 22 nd harmonic
Healthy	1.333	4.202
ICL	1.938	3.956
Loose belt	1.333	3.406
SVL	2.482	2.163
DVL	1.663	1.435

TABLE 5.4 95% CONFIDENCE INTERVALS OF THE FIRST HARMONIC AMPLITUDES.

	Mean	St. Dev.	95% Confidence Interval	Significant at 5%
Healthy	0.3269	0.0927	0.1452, 0.5086	
DVL	0.3454	0.1042	0.1412, 0.5496	
ICL	1.1658	0.1971	0.7795, 1.5521	*
Loose Belt	0.3769	0.1186	0.1444, 0.6094	
SVL	0.3805	0.0772	0.2292, 0.5318	

TABLE 5.5 95% CONFIDENCE INTERVALS OF AMPLITUDE SPECTRA AT GIVEN HARMONICS.

	9th Harmonic				11th Harmonic			
	Mean	St. Dev.	95% CI		Mean	St. Dev.	95% CI	
Healthy	0.3269	0.0927	0.1452	0.5086	0.1676	0.1681	-0.1619	0.4971
DVL	0.3454	0.1042	0.1412	0.5496	0.1645	0.1424	-0.1146	0.4436
ICL	1.1658	0.1971	0.7795	1.5521	0.2403	0.2594	-0.2681	0.7487
LB	0.3769	0.1186	0.1444	0.6094	0.2560	0.1578	-0.0533	0.5653
SVL	0.3805	0.0772	0.2292	0.5318	0.1777	0.1276	-0.0724	0.4278
	15th Harmonic				17th Harmonic			
	Mean	St. Dev.	95% CI		Mean	St. Dev.	95% CI	
Healthy	0.2615	0.3265	-0.3784	0.9014	0.1365	0.0189	0.0995	0.1735
DVL	0.3396	0.3332	-0.3135	0.9927	0.1539	0.0215	0.1118	0.1960
ICL	0.2679	0.4142	-0.5439	1.0797	0.5750	0.0474	0.4821	0.6679
LB	0.2737	0.4037	-0.5176	1.0650	0.1849	0.0285	0.1290	0.2408
SVL	0.3262	0.3577	-0.3749	1.0273	0.2690	0.0188	0.2322	0.3058

As seen it is possible to separate the four faulty cases from the healthy through inspection of the envelope spectrum amplitudes at given harmonics. Note a series of harmonics are required to fully identify all four faults. Specifically:

- At the 9th harmonic the spectra for the DVL and SVL are of obviously different magnitudes to that of the healthy case. The amplitude for the ICL is too similar to that of the healthy case to identify at this point.
- However, at the 15th harmonic the ICL has an amplitude well in excess of the healthy case.
- The loose belt fault lags the healthy case with a peak at the 7th harmonic (4.071) following a peak of similar amplitude in the healthy case at the 6th harmonic.
- The loose belt displays a slightly larger amplitude at the 11th harmonic than the healthy signal (2.956 and 2.441 respectively). More obviously the loose belt has a far larger amplitude (2.624) at the 17th harmonic where the healthy peak is less than 2.
- The healthy signal envelope spectrum rises sharply from the 19th harmonic and is clearly greater than all other spectra between the 21st and 27th harmonics. It's amplitude of 2 being double those of all the fault cases.

Clearly the condition of a system can be ascertained by analysing the vibration signals at strategic points, here at the first and second stage of the operation. Through comparison of the envelope spectrum for each of the four faults and the healthy case a non-intrusive method of diagnosis is established. The results show that it is possible to both distinguish between healthy and faulty systems and to accurately identify the particular fault. Further investigations into distributional characteristics of the data show the ICL to be significantly different thus identifiable, Table 5.8 summarises key findings.

TABLE 5.6 SPECTRUM AMPLITUDES FOR THE SECOND STAGE VIBRATION SIGNAL.

Machine Condition	Amplitude of Envelope Spectra at Stipulated Harmonics		
	9 th Harmonic	15 th Harmonic	17 th Harmonic
Healthy	2.014	1.553	<2
DVL	3.86		
SVL	1.117		
ICL	1.838	2.843	
LB			2.624

5.2.4 Envelope Spectra for the Motor Current Signals

Envelope spectra for the motor current signal are particularly well defined Figure 5.13 and Figure 5.14, with the fundamental frequency at 7.3Hz as anticipated. Each harmonic peak again occurs at multiples of the fundamental frequency and is easily identifiable with sidebands also clearly present. However, due to excessive noise motor current spectra are only really useful up to about the tenth harmonic i.e. to approximately 73Hz. Further investigations are suspended at this point to focus on the vibration signal readings which offer more profitable developments with respect to input parameter selection and manipulation.

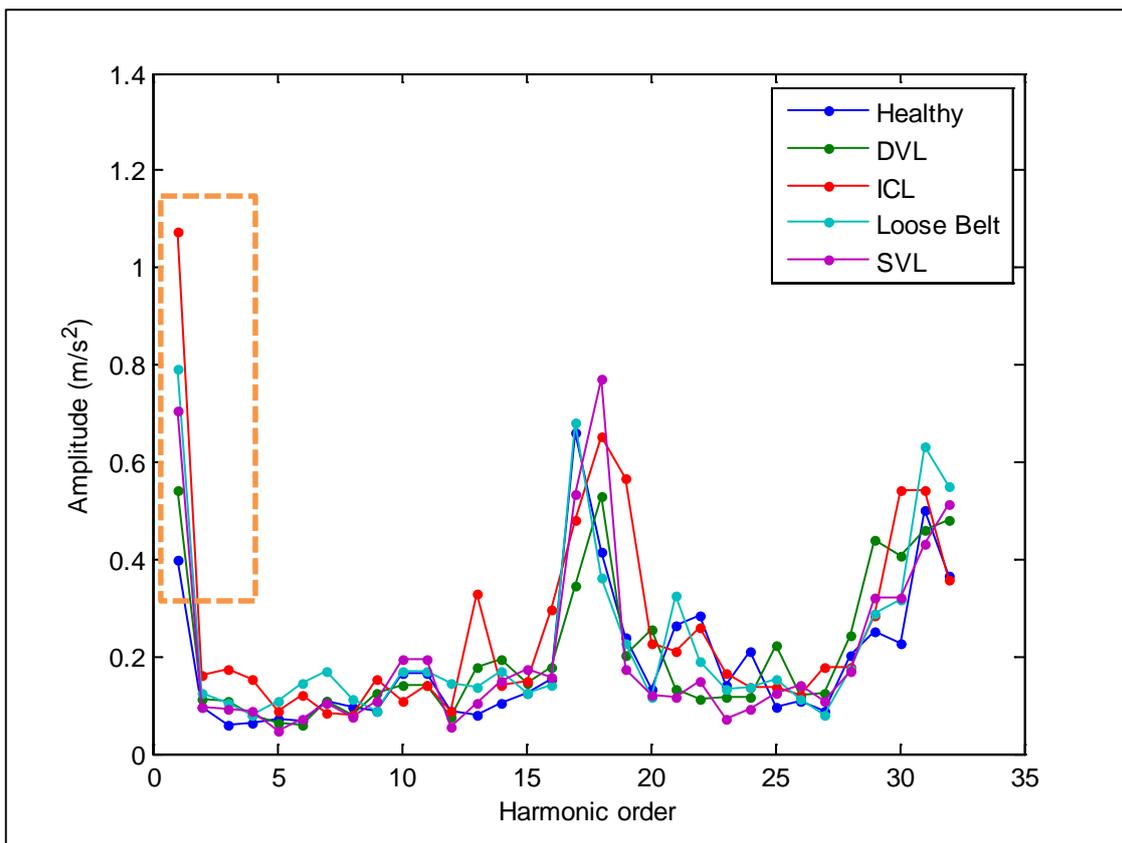


Figure 5.9 Second Stage Frequency Spectra for Healthy System and Four Machine Faults.

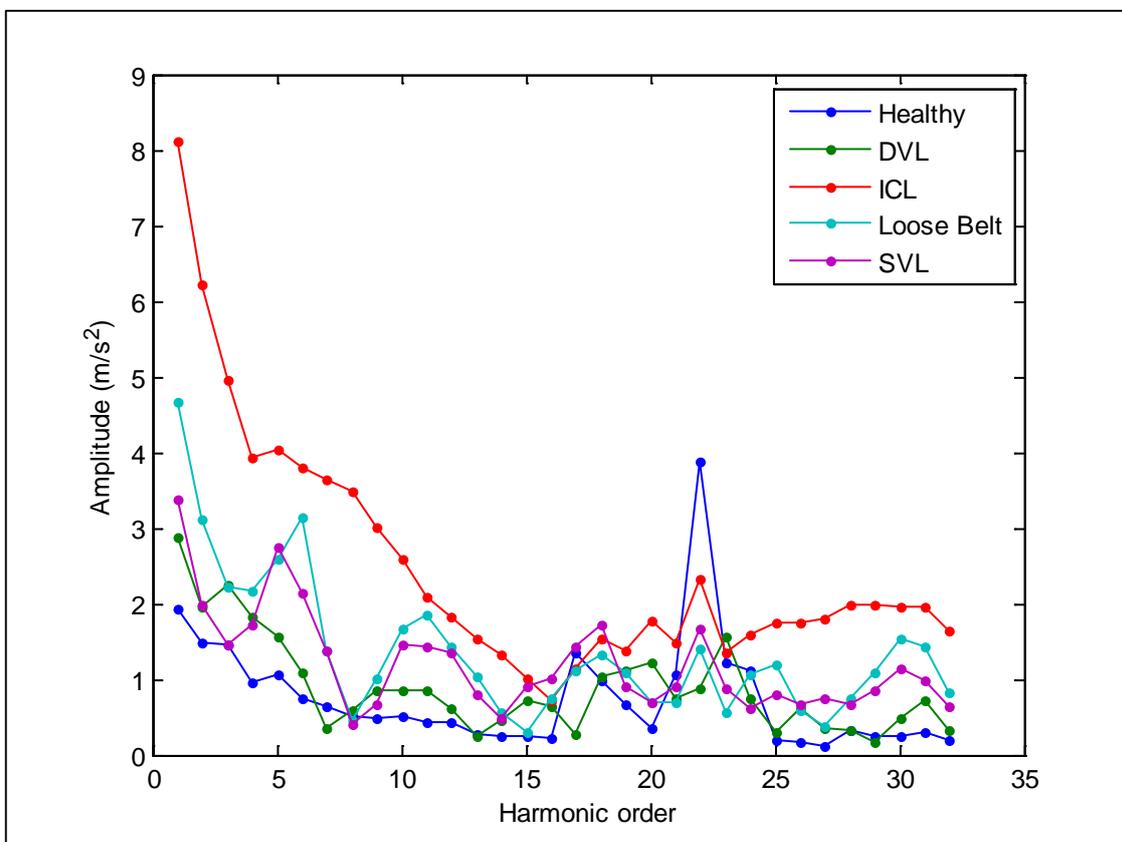


Figure 5.10 Envelope Spectrum of First Stage Vibration Signal.

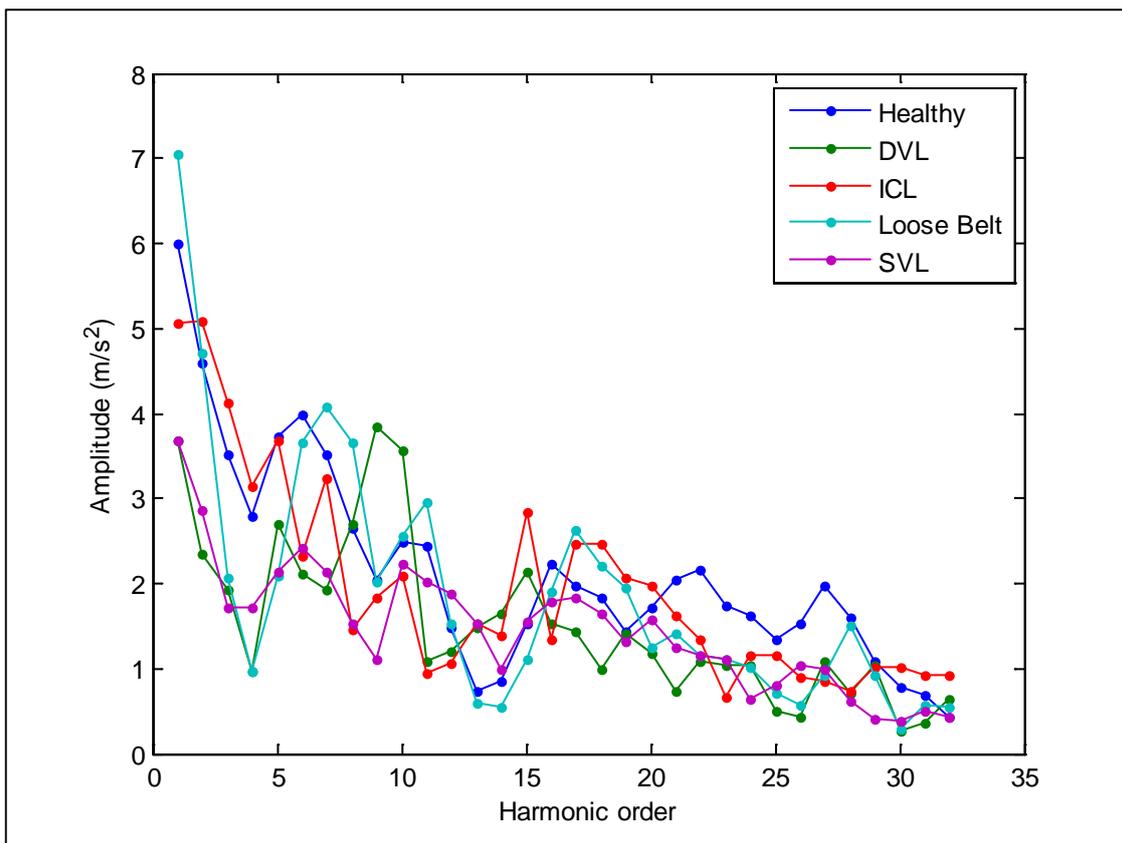


Figure 5.11 Envelope Spectra Of Second Stage Vibration Signal.

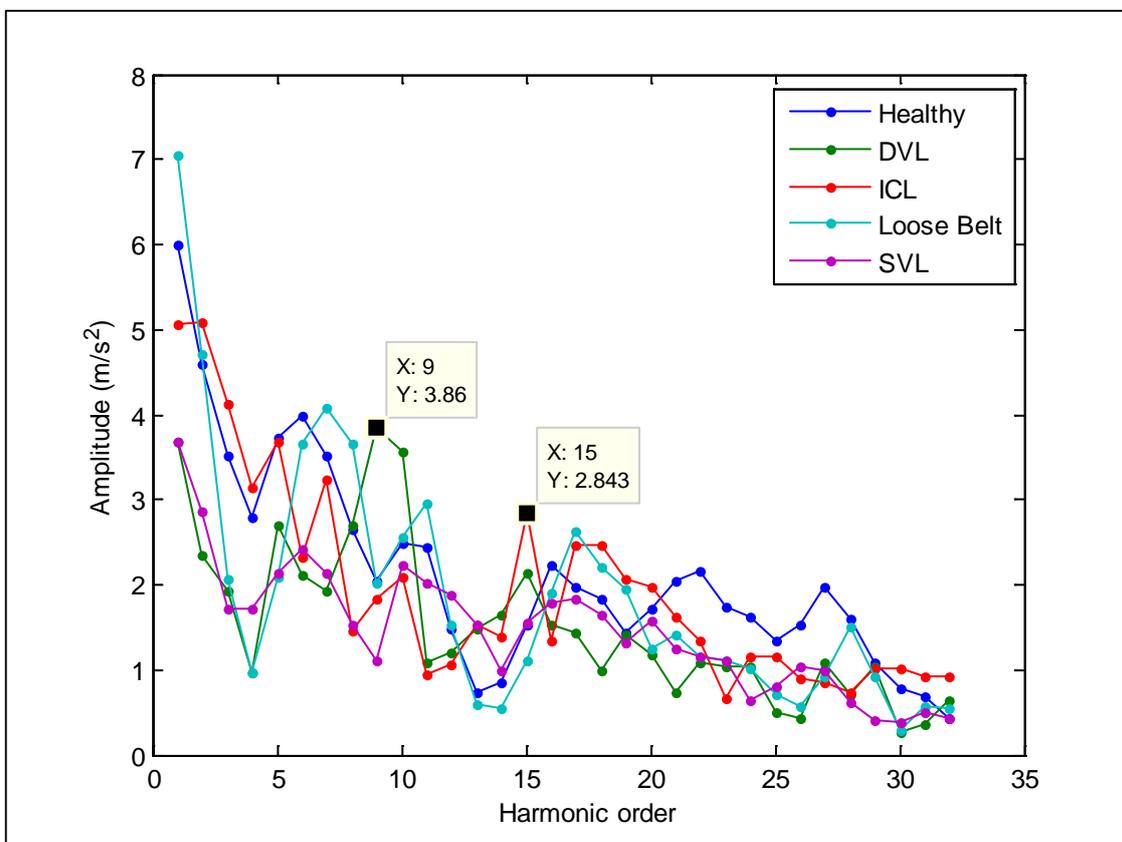


Figure 5.12 Envelope Spectra Displaying Amplitudes At Decisive Harmonics In Second Stage Vibration Signal.

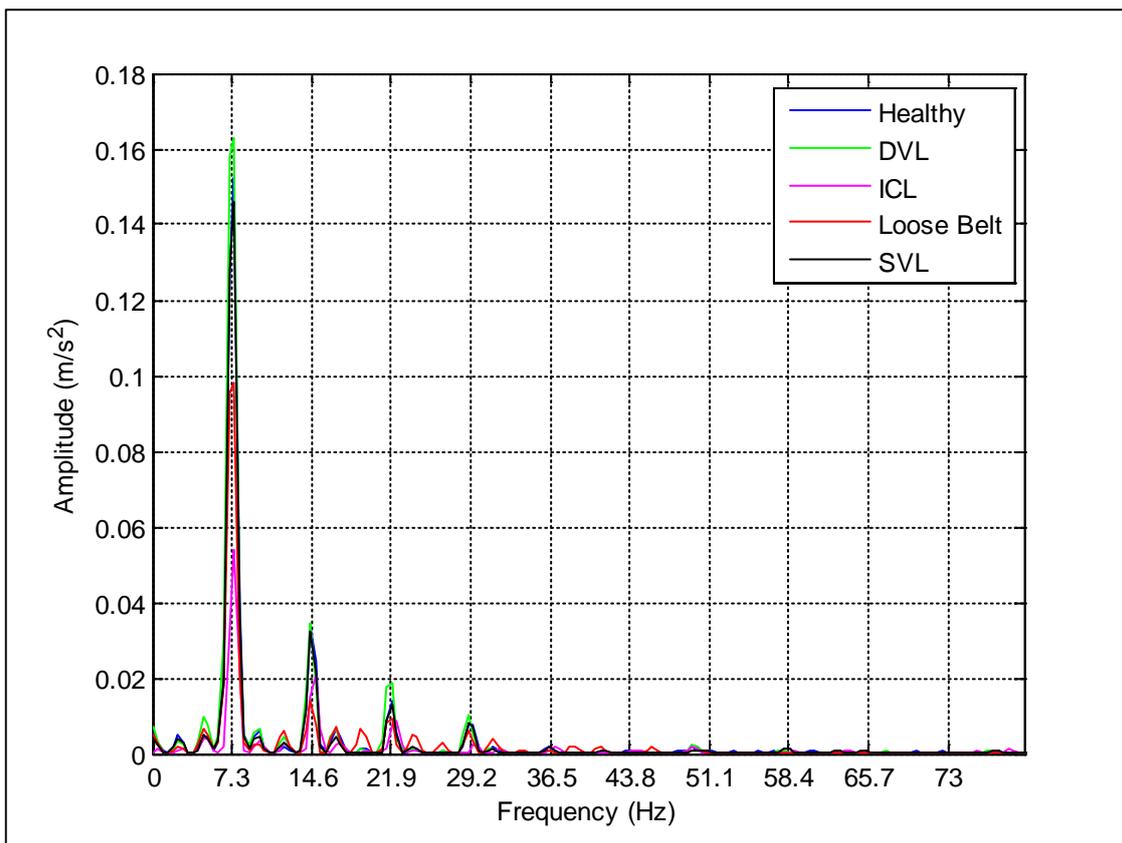


Figure 5.13 Envelope Spectrum for Motor Current Showing Peaks at Fundamental Frequency Multiples.

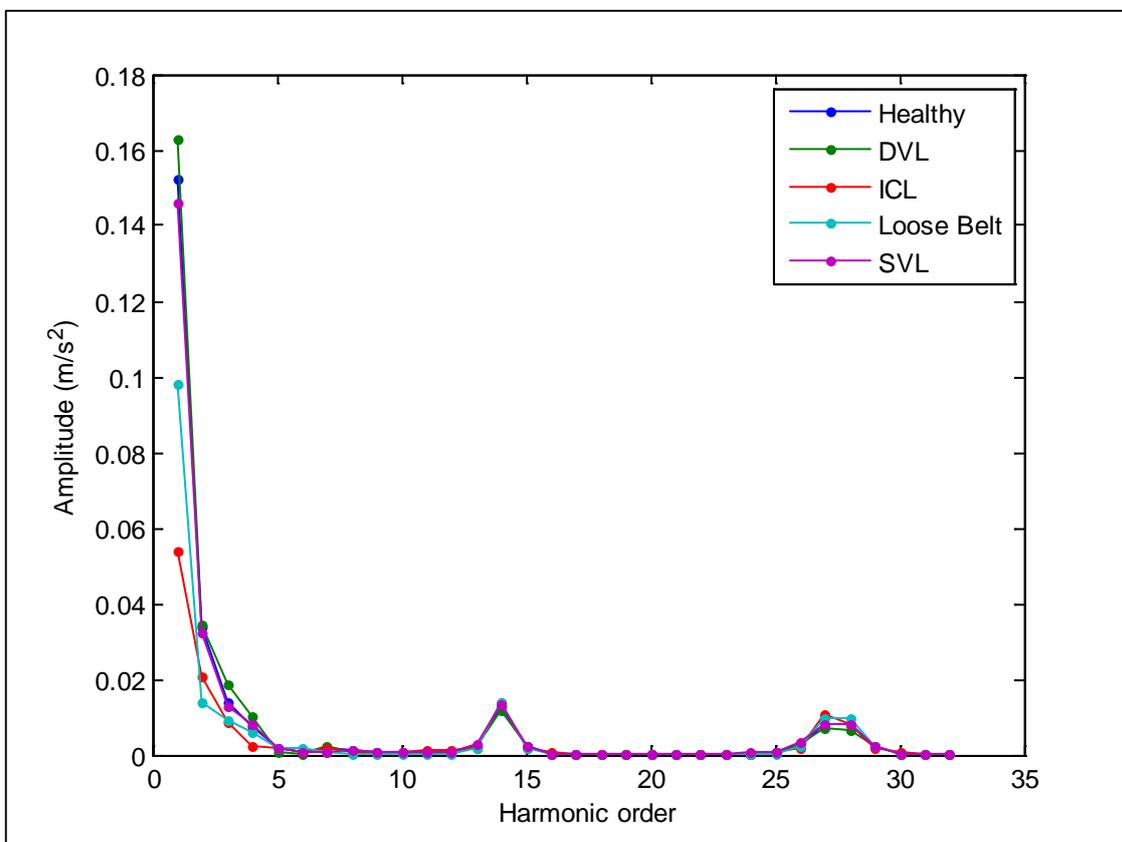


Figure 5.14 Envelope Spectrum for Motor Current Against Harmonic Order.

5.3 Preliminary Investigation into Fault Diagnosis in the Frequency Domain

This section details a collaborative investigation which was felt to be optimal at time of publication [2]. The subsequent analysis and findings serving as motivation for the undertakings within this thesis. Although the study made major contributions in the field it is now considered a less than optimal solution.

Data from the compressor rig was analysed in the frequency domain with the overall aim of detecting and identifying faults induced. The major focus of the study to achieve computational savings and training algorithm convergence alongside classification accuracy.

Through previous analysis it had been shown that a classifier using the RVM technique was far more efficient than using SVM sample wise hence computational time wise.

Using just two input parameters, features 4 and 7, it was demonstrated an SVM model used 24 samples whilst a comparable RVM model required just 4, Figure 5.15 and Figure 5.16 respectively.

Envelope harmonics extracted from the frequency domain were utilised rather than their time domain equivalent as they most efficiently describe modulations due to fault presence as discussed in section 5.2. Figure 5.17 gives a direct comparison of the envelope frequency spectra of different machine states, clearly showing the compressor's fundamental frequency to be 7.3Hz and highlighting the differences in both size and location of harmonic amplitudes for faulty and healthy systems. Hence the envelope harmonic amplitudes were included as input parameters.

Feature set size was determined by the limitations of the RVM algorithm which could accommodate a maximum of 16 input variables to remain stable hence the harmonic components 2 to 15 from the envelope spectra were used. Feature 1 had previously been shown to have little explanatory value and was thus eliminated.

80 data samples for each of 8 classes were selected and randomly assigned to one of two groups of 40 which formed the training data set and the validation data set respectively.

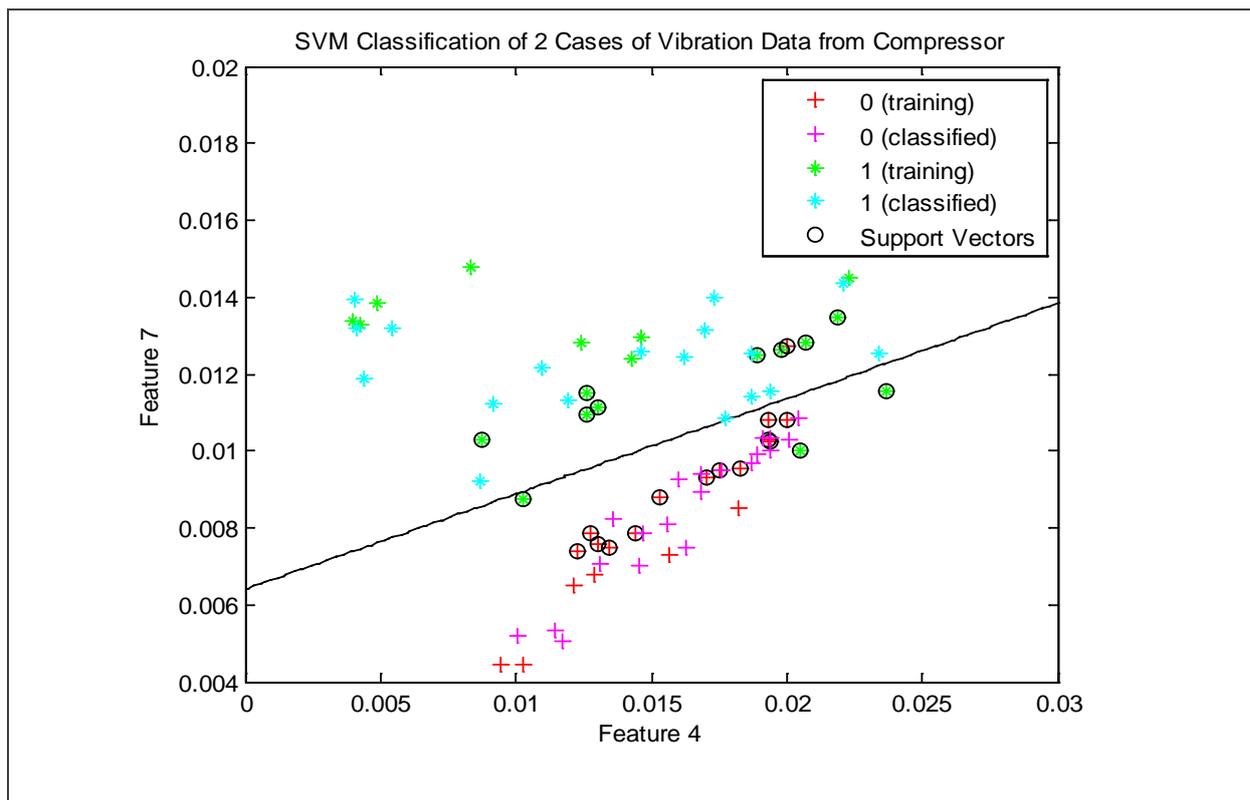


Figure 5.15 Comparison of Model Complexity 24 Sample SVM.

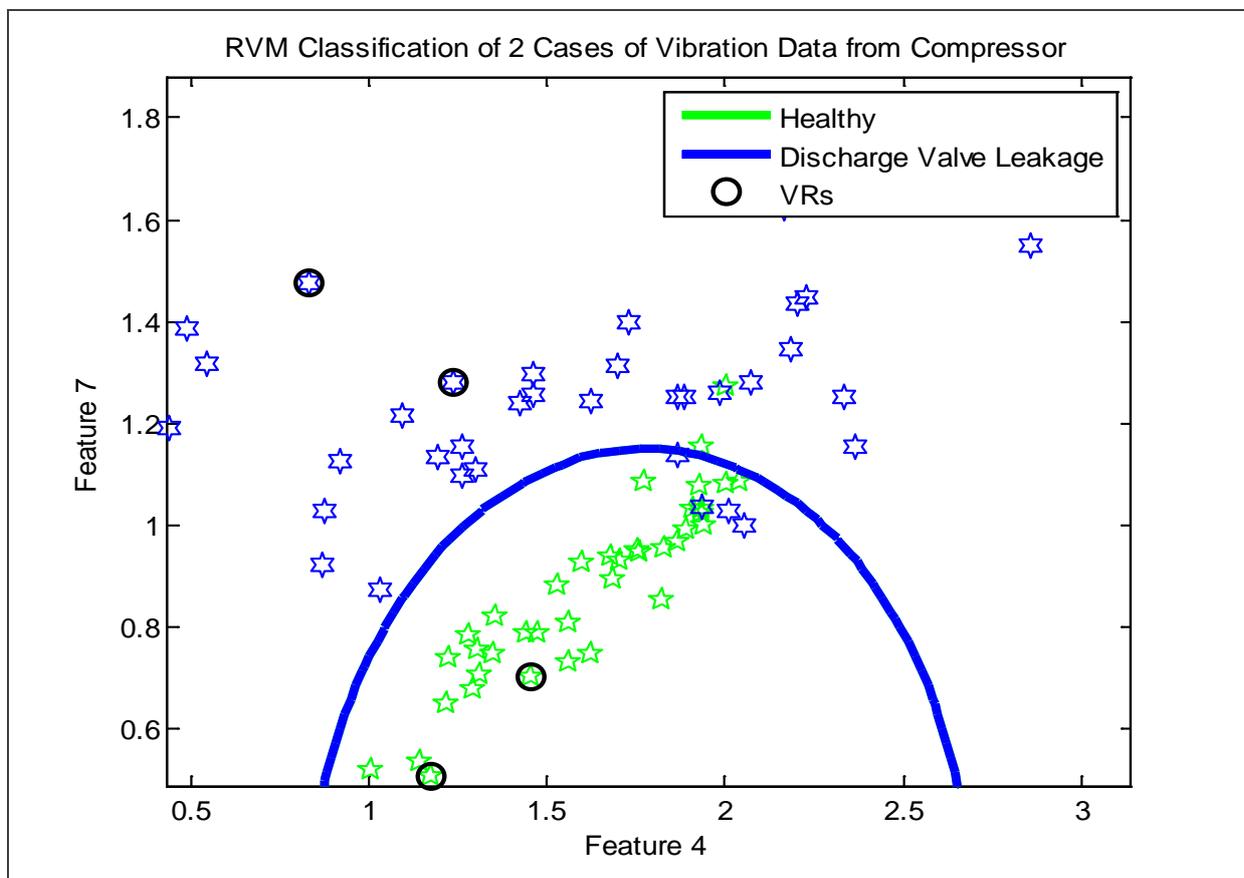


Figure 5.16 Comparison of Model Complexity 4 Sample RVM.

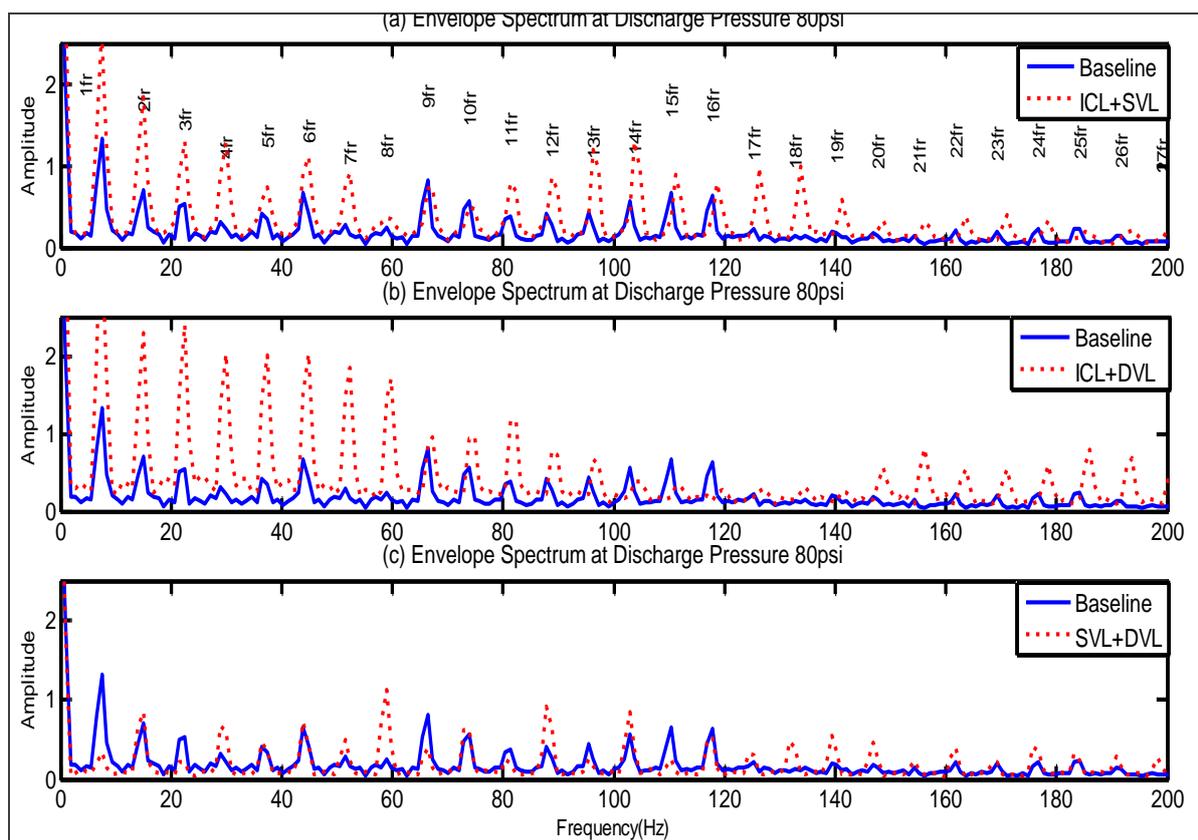


Figure 5.17 Comparison of Envelope Spectra of Faults to Healthy (baseline).

Although a one-against-one (OAO) training algorithm leads to far more individual classifiers than a one-against-all (OAA) algorithm the OAO option was selected as the preferable training approach since the overall computational time to solution was vastly reduced. Classifiers generated through OAA algorithms are typically highly complex and generally require much greater computational efforts than the equivalent multiple OAO classifiers.

Hence RVM models were trained and evaluated using the partitioned data set and a OAO training algorithm first without genetic algorithm (GA) feature selection then with GA feature selection [Section 3.8, 49, 70, 86 and 90]. A GA is an automated algorithm for selecting a reduced number of parameters. Gradient based searches highlight feature usage patterns and so offer a practical means for selecting the most prevalent parameters. Eight classes in total were seeded into the rig, four single and four combined as detailed in Table 5.7.

Incorporating GA feature selection algorithms into the model showed improved classification success across the classes. A visual summary of the harmonic features selected per class is given in Figure 5.17 and appears to show a strong association

between the two variables. Average classification rates are summarised in Table 5.8. A numerical summary of feature usage by class is given in Table 5.9.

TABLE 5.7 FAULTS INDUCED IN SIMULATIONS.

Class Number	Class Code	Class Abbreviation	Class Description
1	C ₁	H	Healthy
2	C ₂	DVL	Discharge Valve leakage
3	C ₃	SVL	Suction Valve leakage
4	C ₄	LB	Loose Drive Belt
5	C ₅	IL	Intercooler Leakage
6	C ₆	DVL+SVL	Discharge Valve leakage with Suction Valve leakage combined fault
7	C ₇	SVL+IL	Suction Valve leakage with Intercooler combined fault
8	C ₈	DVL+IL	Discharge Valve leakage with Intercooler combined fault

TABLE 5.8 AVERAGE CLASSIFICATION RATES.

Cases	Input Harmonics	Classification Rate (test data) (%) RVM-OAO Without GA	Classification Rate (test data) (%) RVM-OAO With GA
Overall	2-15	95.95	97.00
Healthy	2-15	90.00	93
DVL	2-15	100.00	97.5
SVL	2-15	97.50	93
LB	2-15	95.00	100
ICL	2-15	95.00	100
DVL+SVL	2-15	100.00	100
SVL+ICL	2-15	90.00	93
DVL+ICL	2-15	100.00	100

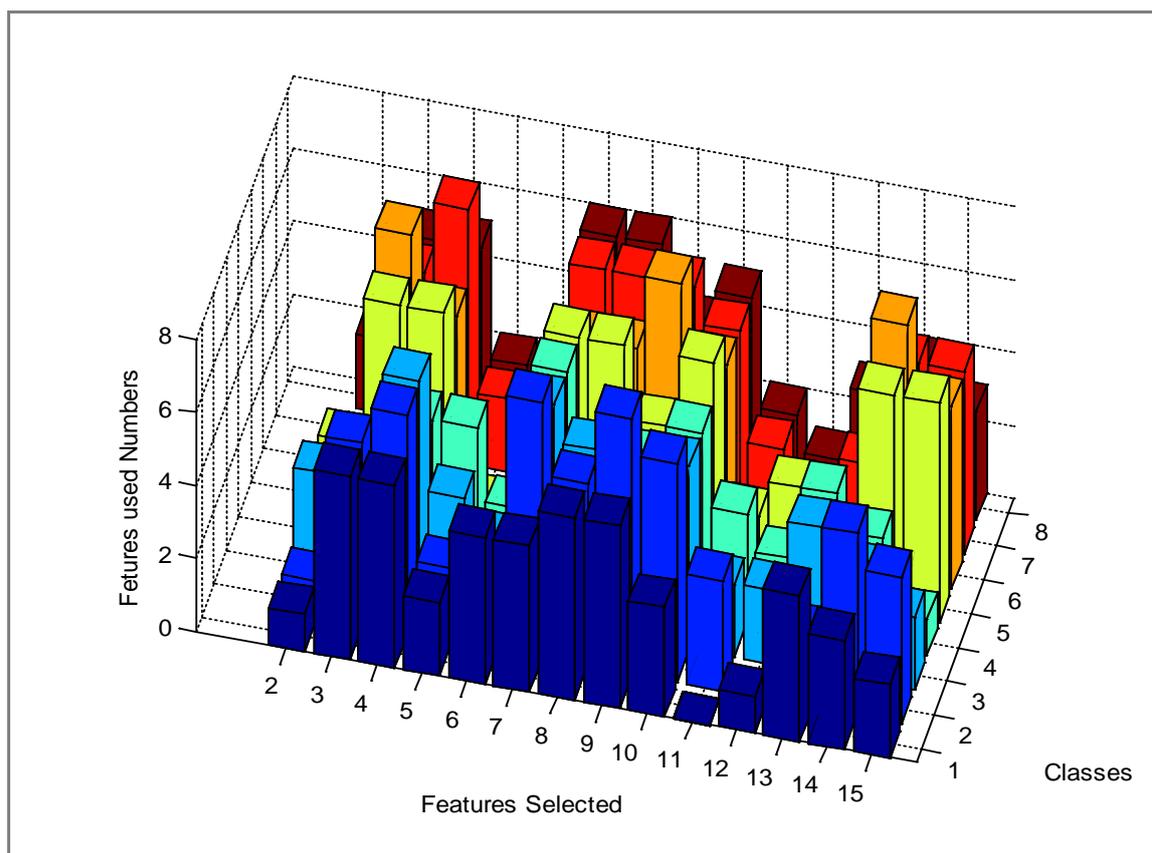


Figure 5.18 Features Selected by Class.

TABLE 5.9 FEATURE USAGE BY CLASS.

	<i>f</i> ₂	<i>f</i> ₃	<i>f</i> ₄	<i>f</i> ₅	<i>f</i> ₆	<i>f</i> ₇	<i>f</i> ₈	<i>f</i> ₉	<i>f</i> ₁₀	<i>f</i> ₁₁	<i>f</i> ₁₂	<i>f</i> ₁₃	<i>f</i> ₁₄	<i>f</i> ₁₅
<i>C</i> ₁	1	5	5	2	4	4	5	5	3	0	1	4	3	2
<i>C</i> ₂	1	5	6	2	1	7	5	7	6	3	0	2	5	4
<i>C</i> ₃	3	4	6	3	2	6	5	3	5	2	2	4	3	2
<i>C</i> ₄	2	3	4	4	2	6	2	4	5	3	2	4	3	1
<i>C</i> ₅	2	6	6	1	1	6	6	4	6	2	3	1	6	6
<i>C</i> ₆	1	7	5	0	2	5	5	7	5	0	1	0	7	5
<i>C</i> ₇	0	5	7	2	1	6	6	6	5	2	0	2	5	5
<i>C</i> ₈	2	5	5	2	1	6	6	4	5	2	1	3	4	3

Consequently a further summary of the dominant features present for each of the classes was constructed, Table 5.10. Note each of the classes is seen to contain a dominant pair of features, 4 and 7, which form the ‘base’ model for the following analysis and later experimental classification models.

TABLE 5.10 DOMINANT FEATURES USED IN CLASSIFICATION.

Class	Cases	Features Used
C_1	Healthy	3,4,6,7,8,9,13
C_2	DVL	3,4,7,8,9,14
C_3	SVL	3,4,7,8,10,13
C_4	LB	4,5,7,9,10,13
C_5	ICL	3,4,7,8,9,14,15
C_6	DVL+SVL	3,4,7,8,9,10,14
C_7	SVL+ICL	3,4,7,8,9,10,14,15
C_8	DVL+ICL	3,4,7,8,9,10,14

Finally a multiclass multi-kernel RVM (mRVM) was trained for four classes (H, ICL, DVL+ICL and SVL+DVL) again using the two input parameters envelope features 4 and 7.

The main purpose of mRVM algorithms is to gain high predictive accuracy rates whilst maintaining computational efficiency. mRVMs produce more sparse solutions both sample and kernel wise enabling application to large-scale multi-feature multinomial classifications. Performance across the four faults, Table 5.11, Figure 5.19, for the two parameter mRVM was extremely good with 100% for binary and 97.5% accuracy rates for combined faults. However, it has since been demonstrated, for example in the DA in section 6.3, that the two envelope features 4 and 6 have still greater explanatory power.

A mainly intuitive experimental exploration of the effect on mRVM classifications of using differing numbers of classes and input harmonics was undertaken, Table 5.12. It became apparent that the physical mechanisms were potentially related to input features. Further that particular harmonic features were strongly associated with specific fault characteristics being repeatedly used in their presence. Some features were found to have strong associations with more than one fault, for example, the

'base' model of features 4 and 7. Feature 9 was used repeatedly in the presence of the DVL whilst the SVL diagnosis relied heavily on features 8 and 10. Confounding factors were also obvious with the ICL emulating some of the characteristics of other faults. Detection and accurate diagnosis of an increased number of faults whilst maintaining computational efficiency is only achievable through optimal input parameter selection.

TABLE 5.11 CLASSIFICATION RATES FOR MRVM.

State	Number of Samples	Correct classification	Accuracy	Error Rate
Healthy	40	40	100%	0.0125%
ICL	40	40	100%	
SVL+IL	40	39	97.5%	
ICL+DVL	40	39	97.5%	
Overall			98.75%	

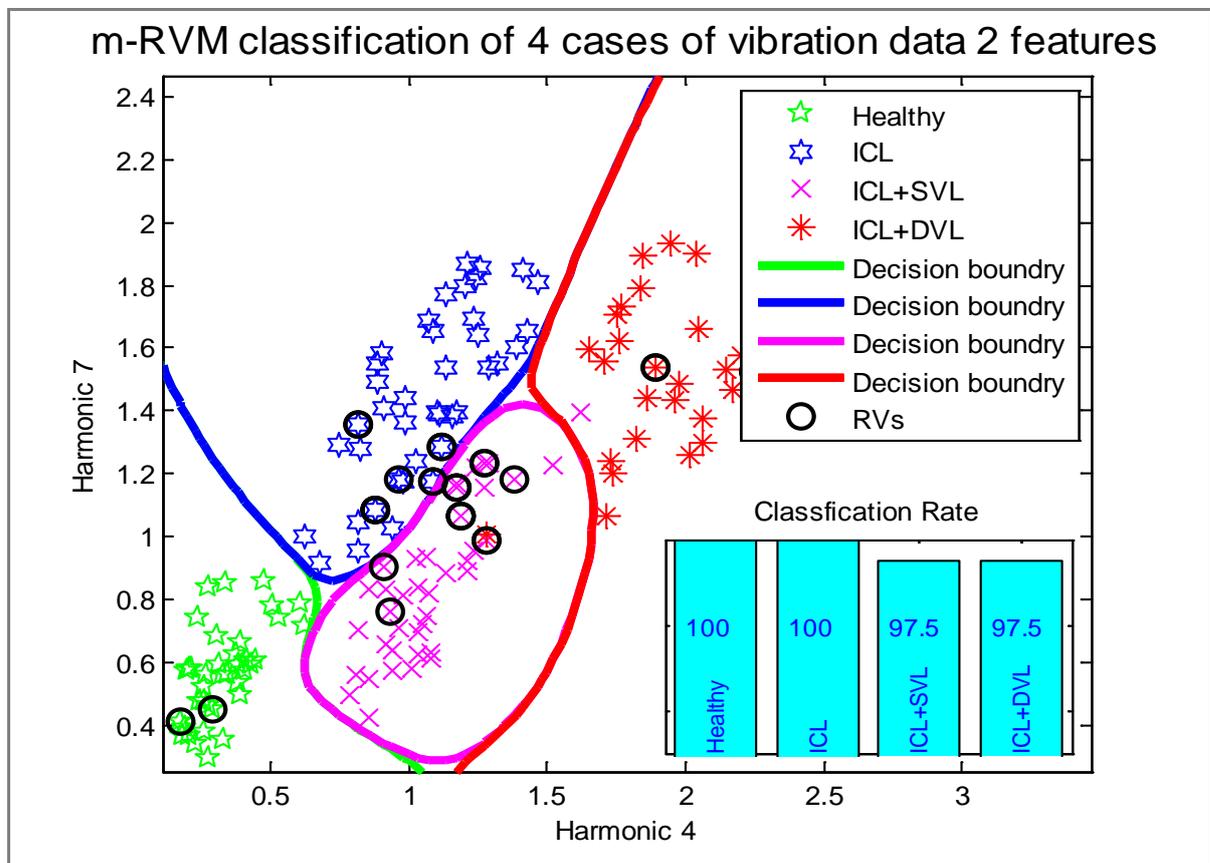


Figure 5.19 mRVM Classification Plot and Decision Boundaries.

Whilst a GA based approach optimised input feature selection with improved classification accuracy across a large number of classes when compared to RVMs without GA, using mRVM with 2 input features (4 and 7) achieved near perfect explanation of fundamental behaviour in all cases. However, it has now been demonstrated that more optimum solutions still are possible with respect to explanatory power of individual harmonic features. Hence the motivation for further investigating input parameter properties and their interrelationships. Thus enabling optimal input parameter selection in generating models with further inflated powers of separation. It should be noted that a feature being ‘little used’ does not render it useless being potentially of vital importance in fine tuning the classification process.

TABLE 5.12 CLASSIFICATION RESULTS USING MRVM WITH VARYING NUMBERS OF INPUT PARAMETERS AND CLASSES.

Classes	Harmonic Orders	Accuracy	Error Rate
1,4,6,7	4-15	98.75%	0.0125
1,4,5,7,8	4-15	86%	0.160
1,2,4,5,7,8	2-15	73.75%	0.2625
1,2,3,4,5	2-15	77.0%	0.23
1,2,3,4,5,6,7	2-15	59.29%	0.4071
1-8	4,5,6,10,15	24.16 %	0.3969
1-8	2-15	50%	0.50

5.4 Summary

Each signal measurement from the RC rig was seen to have a cyclic output of approximately seven cycles per second. Interrelationships between signals were also well defined and could be directly attributed to position in the two-stage compressor cycle.

Analysis in the time domain showed reasonable separation potential across the different classes with the second stage pressure amplitudes showing quite different profiles. Most noticeably, although not surprisingly, the ICL signal had very low second

stage pressure amplitude. However, there were no clear rules for separation of all faults simultaneously.

Analysis of vibration signals in the frequency domain highlights the unique signal characteristics. Frequency spectra clearly show the fundamental frequency of the rig and higher harmonics at its integer multiples. However, whilst the positioning of the fundamental frequency is the same for all classes position and size of amplitudes varies across the classes. This forms the basis of reliably distinguishing between fault types. Further, harmonics extracted from the envelope spectra of vibration signals in the frequency domain are seen to have superior deterministic properties over their time domain equivalents. Amplitudes of the envelope harmonics being specific to process condition with a greater amplitude or displaced amplitude implying presence of a fault. Envelope spectra show only the amplitude profile of original signals and so provide a clearer insight into the underlying behaviour of a process having extraneous noise removed.

Findings from the preliminary investigation into fault diagnosis in the frequency domain showed how great the potential for detecting and identifying faults using envelope spectra from the second stage vibration signals. High classification success rates were achieved using RVM identifying single and compound faults. The limiting factor of the study being the algorithmic restrictions on numbers of input parameters for model convergence. Subsequent investigations revealed strong associations between specific fault characteristics and particular envelope harmonics. Whilst incorporating a GA was seen to further improve classification success rates the underlying input parameter structure was unknown as physical characteristics were not preserved. Findings motivating continued research, the following chapter begins with the exploratory investigation of the envelope spectra harmonic traits.

Chapter 6

Multivariate Classifiers Using Reduced Number of Input Parameters

Novel methods for inspection of fault signal profiles to aid identity of separable classes prior to analysis is demonstrated. Variable clustering techniques and input parameter grouping are introduced at the outset of this chapter. Followed by construction and appraisal of multivariate classifiers established using reduced numbers of input variables determined through variable clustering and the efficacy of the multivariate models constructed. Discriminant analysis and Naïve Bayes classification models are established for both two and five classes.

6.1 Identification of Separable Classes

6.1.1 Andrews Plots of the Second Stage Vibration Signal

In addition to identifying suitable input parameters for the model building process, it is sensible to consider the feasibility of class separation. One useful profile method is found in Andrews plots, a Fourier transform of the signal data.

Greatest profile differences occur at approximately $t=0.4s$ as shown in Figure 6.1 and Figure 6.2. Although the healthy case is difficult to decipher across the range, the faulty cases do appear to form distinct groups and as such should be identifiable. Obviously the greater the differences in profile plots the more easily the cases are separated.

6.1.2 Andrews Plots of the Motor Current Signal

The Andrews plots based on the motor current signal show mostly distinct profiles for each of the classes, thus models with high classification capabilities are anticipated as shown in Figure 6.3 and Figure 6.4. Most closely woven classes being the DVL and SVL along with the Healthy system so it might be expected these would be more troublesome to differentiate between. On the other hand the ICL profile suggests a far more straightforward classification to be achievable. Indeed classifications based on the motor current envelope features gave high rates of success for both discriminant analysis and Naïve Bayes classifiers. However, since the envelope spectra for the motor current consists mainly of noise from around the tenth harmonic this line of enquiry was not pursued and modelling focused on the second stage vibration envelope features

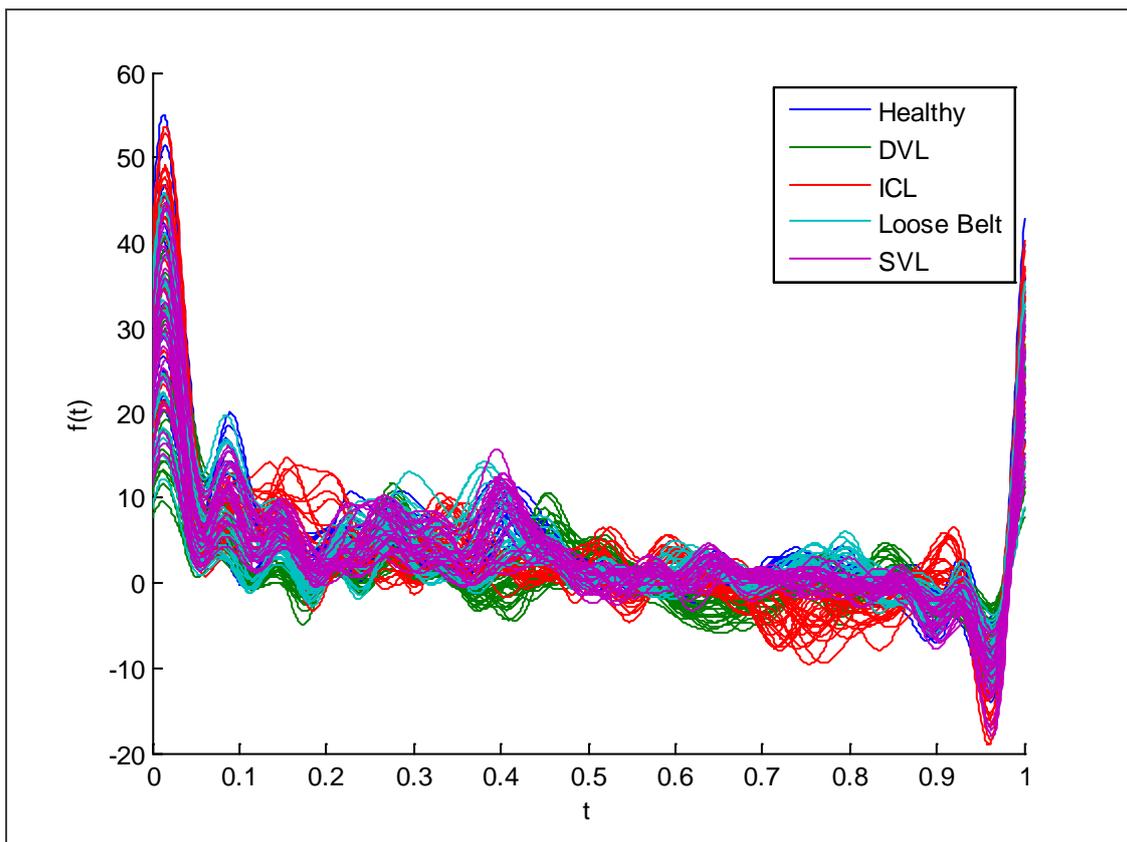


Figure 6.1 Andrews Plot Showing the Feature Profiles by Class Against time(s).

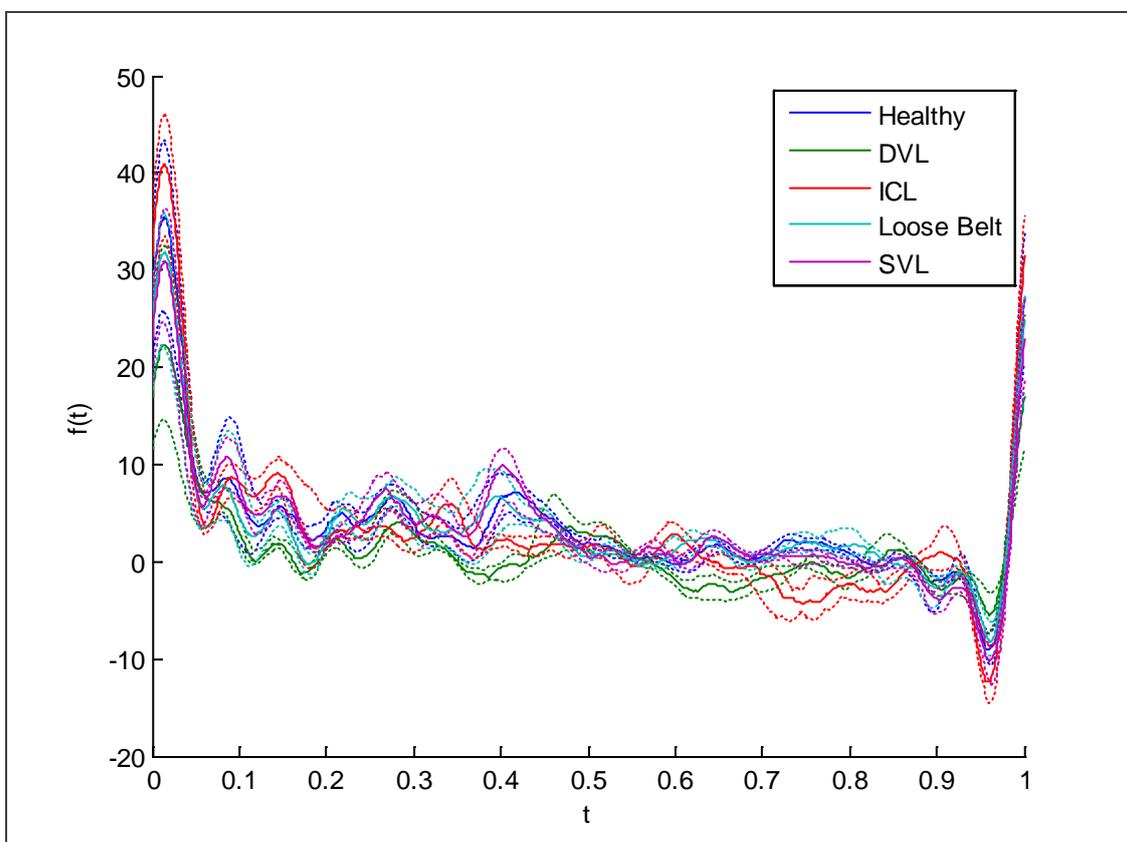


Figure 6.2 Andrews Class Profiling using Second Stage Vibration Signal Against time (s).

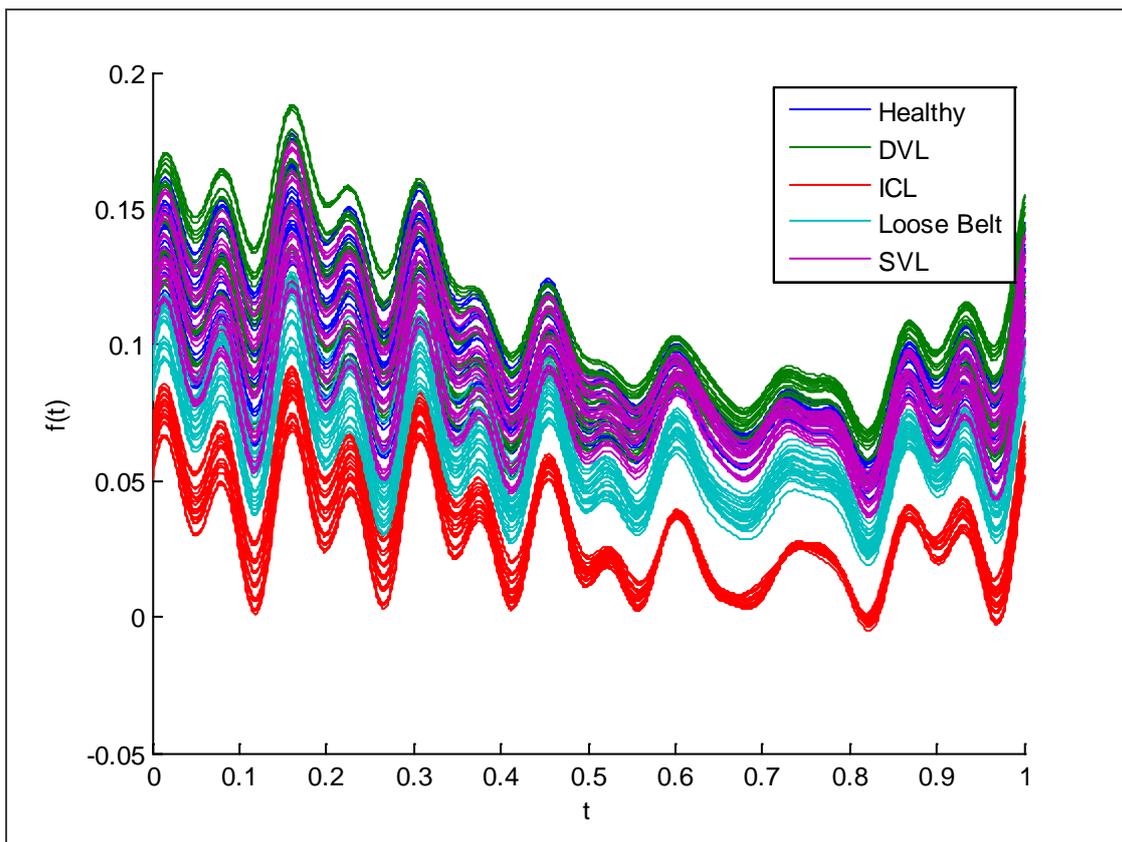


Figure 6.3 Andrews Plots for each of the Five Cases using Motor Current Signal Against time(s).

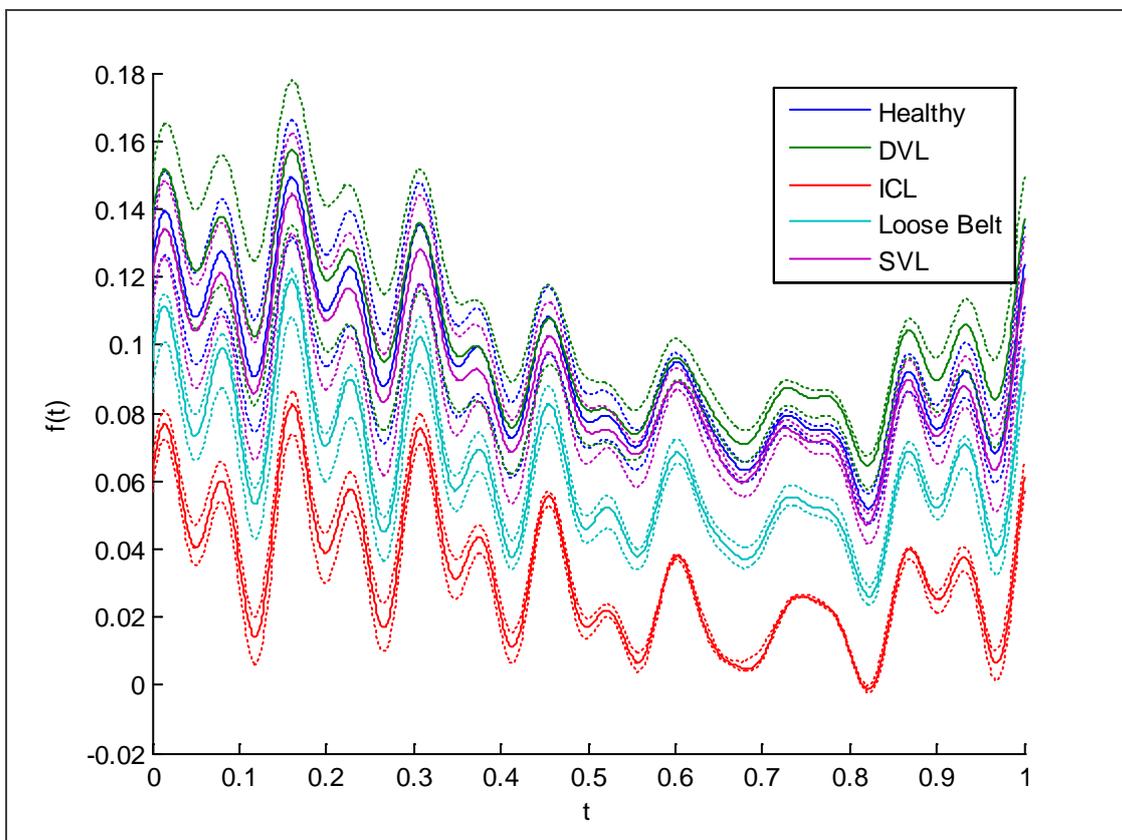


Figure 6.4 Andrews Plots using Motor Current: Medians and Quartiles Against time(s).

6.2 Variable Clustering

As the main focus of the CA was to identify variable similarity an agglomerative method was employed thus group formation and structure were easily observable. Euclidean distance was used as proximity measure since the measurements represent harmonic amplitudes.

Investigations into feature similarities through clustering are displayed in the dendrogram Figure 6.5. Classification matrix, X (120 by 32) comprised 120 observations, 24 repeated runs of all 5 classes or machine states, at each of the first 32 harmonic features extracted from the envelope spectrum analysis. Dendrograms were based on the pairwise Euclidean distance between each of the variables generating a differences vector of length $m(m-1)/2 = 32*31/2=496$.

6.2.1 Clustering of Second Stage Vibration Signal Envelope Harmonics

A summary of the second stage vibration cluster results is given in Table 6.1. Harmonic features 23, 24, 27 and 28 were found to be most similar whilst the group containing features 3 and 5 is least like all others Figure 6.5. There are several other early groupings of harmonic features which would be expected to explain similar variation between cases. Other harmonics would also appear to readily form group pairs. For example, harmonics 13 and 14 and harmonics 11 and 16. Interestingly many of these early groupings are between harmonic features within the range selected in the preliminary investigation, discussed in section 5.5, which would suggest discriminating power may have been lost due to duplication.

Extended inspection highlights the existence of 3 main harmonic groups plus 3 independent harmonic features. The five 'semi-independent' features (6, 7, 8, 10 and 15) shown outside the group 3 brackets not being included in any grouping until the distance, $T>14$ suggests they are heterogeneous and potentially provide unique information with respect to separation of cases. Envelope harmonics are grouped by similarity thus forming homogeneous sets. A representative feature was selected from each group through inspection of the envelope spectra with the criteria of maximum case separation i.e. maximum amplitude discrepancy. Group 3 had a number of late joining features which can be considered as partial outsiders which are likely to provide additional explanatory power to the model.

It is reasonable to expect the variation within each group might be explained by one representative feature from that group since within group characteristics are

homogeneous. Between group differences confirm findings from prior research showing cases to be associated with specific harmonic features.

Prior research has shown RVM based on envelope features to be effective in separating classes. The authors in [2] employed genetic algorithms for feature selection with near perfect classification results. Clustering of features here highlights similarities within groups of features and heterogeneity between groups. Subsequently cluster scatter diagrams were constructed using single group representatives to explore visual separation of observations per class.

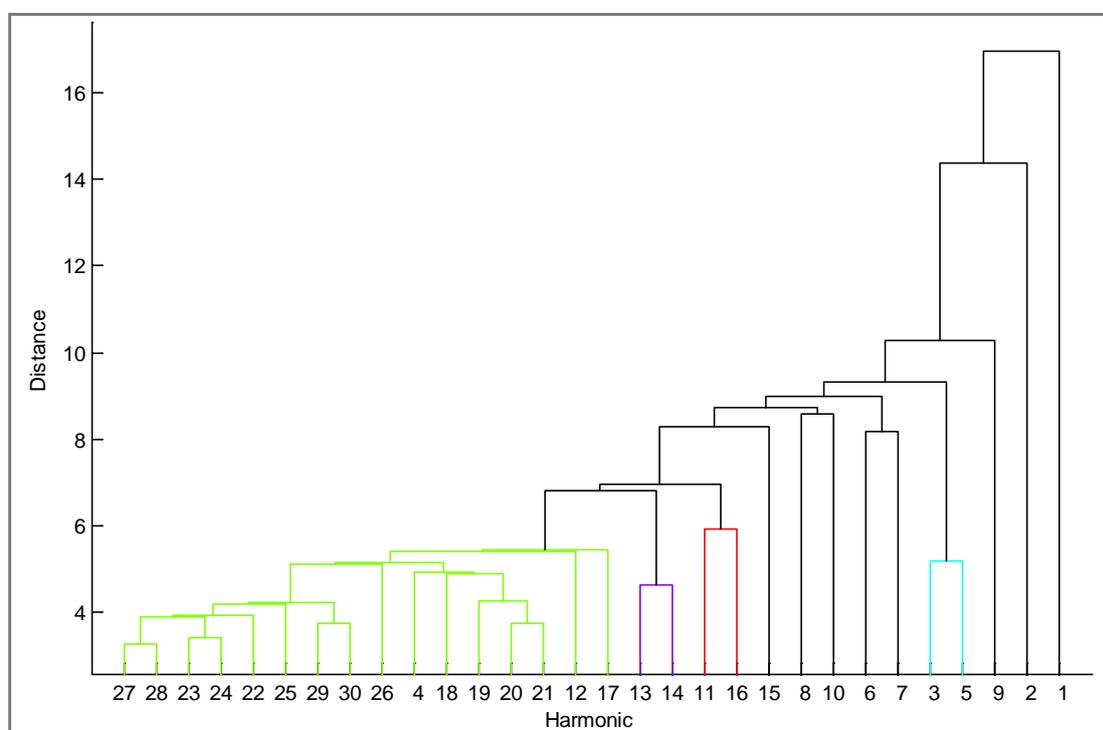


Figure 6.5: Dendrogram Displaying Feature Similarity (colour threshold $T=6.1$) Second Stage Vibration Envelope Feature. Clustering Based on Euclidean Distance between Pairwise Observations.

TABLE 6.1 SECOND STAGE VIBRATION ENVELOPE FEATURE CLUSTER GROUPS (GROUP THRESHOLD SET AT 6.0 WITH SUB-SETS SHOWN IN BRACKETS FORMING PRIOR TO THIS DISTANCE).

	Group Members
Group 1	4 selected as representative of large group of like features.
Group 2	(13,14), (11,16)
Group 3	(3, 5) 6, 7, 8, 10, 15
Independent features	2, 9 (1 omitted as has been shown to have little discriminating power).

6.2.2 Cluster Analysis of the Motor Current Signal Envelope Harmonics

Motor current envelope features formed three main groups with a further five features joining those groups at a greater distance. Features 27 and 28 although very similar to each other are seen to be least like most of the other envelope features as shown in Figure 6.6. Feature 2 being the greater outsider by far.

Envelope features are grouped by similarity thus forming homogeneous sets of features. A representative feature was selected from each group through inspection of the envelope spectra with the criteria of maximum case separation. Features late joining groups can be considered as partial outsiders, namely 3, 4, 5, 7 and 14, and as such are likely to provide additional explanatory power to the model.

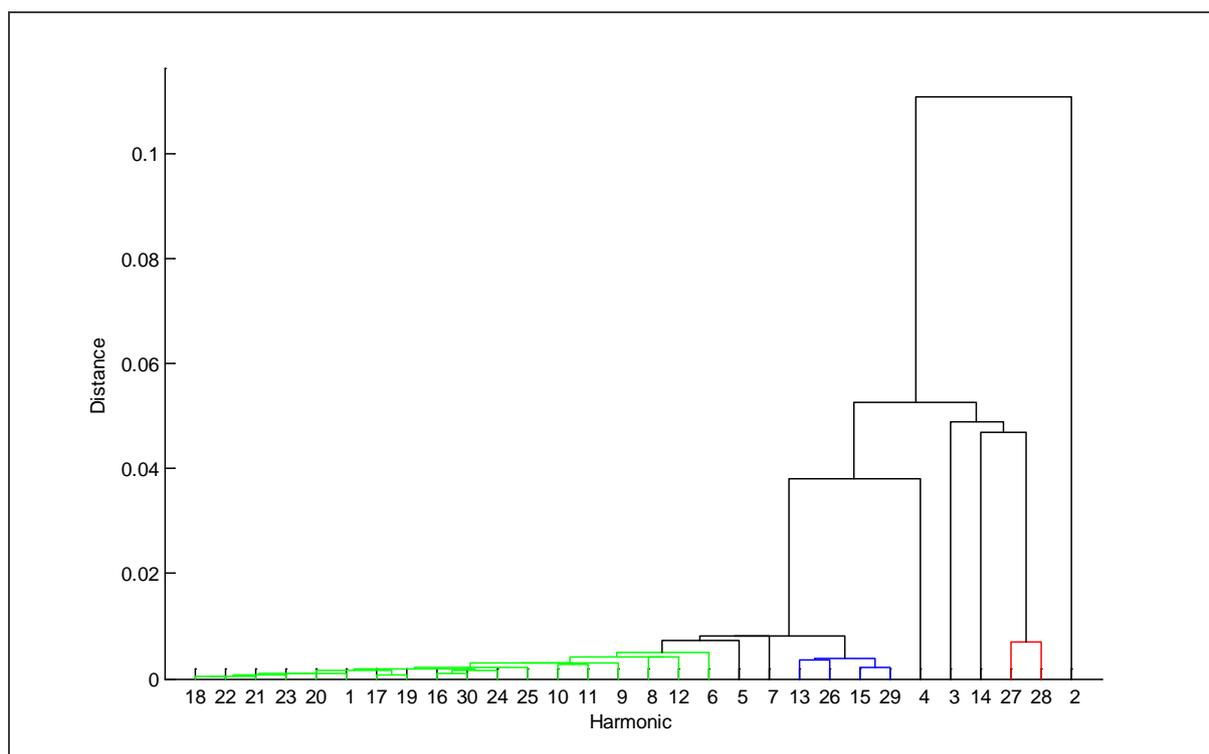


Figure 6.6 Dendrogram of Motor Current Feature Clustering Based on Euclidean Distance Between Pairwise Observations.

TABLE 6.2 MOTOR CURRENT HARMONIC FEATURE CLUSTER GROUPINGS

	Group Members
Group 1	Large number of features including [1 6 8 9 10] 5
Group 2	[13 15 26] 7 4
Group 3	[27 28] 3 14
Independent	2

6.2.3 Cluster Analysis of the Second Stage Pressure Signal Envelope Harmonics

Clustering of the second stage pressure output signal envelope harmonics, demonstrates high levels of uniformity in group formation. Findings clearly support the selection of relevant model input parameters through clustering to identify variable duplication so reduce numbers of input variables in model construction.

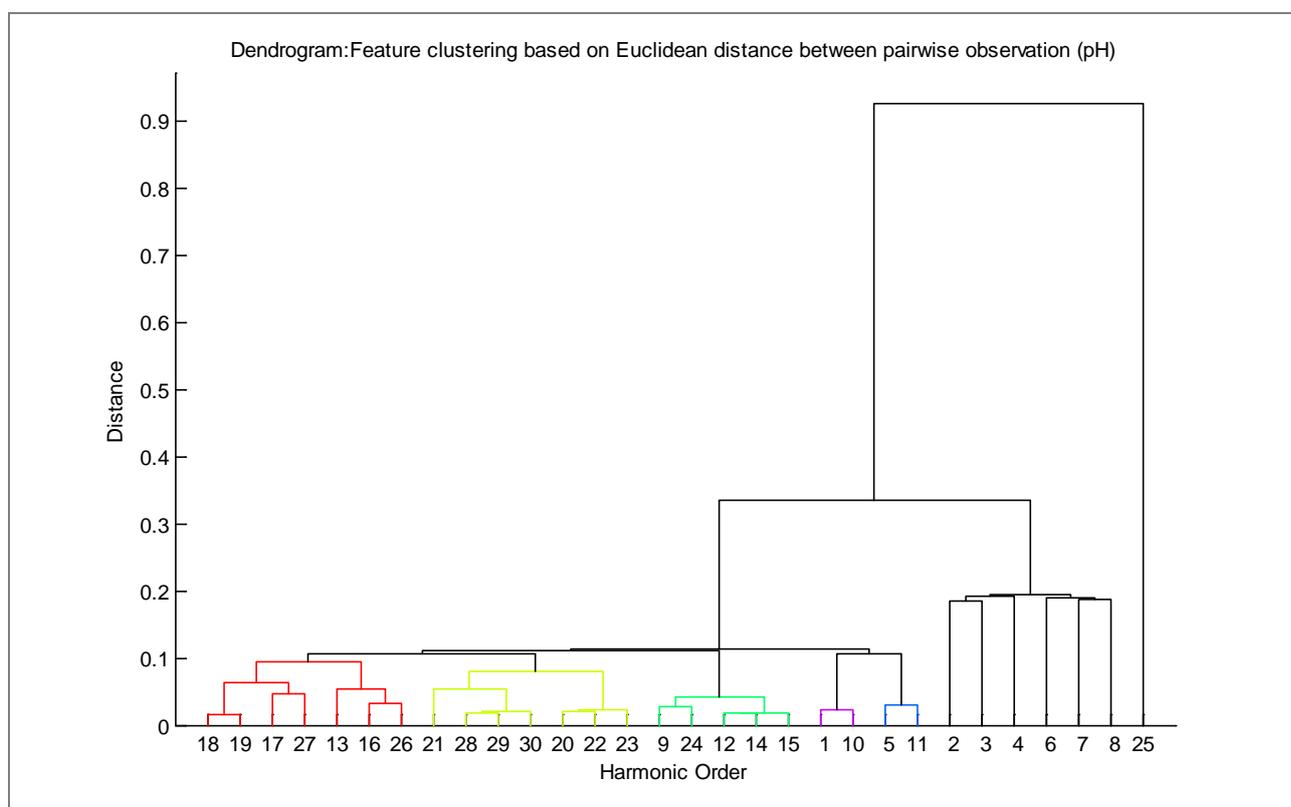


Figure 6.7 Dendrogram for Second Stage Pressure Envelope Features.

Homogeneous feature groupings can be identified from the dendrogram, Figure 6.7, with selection of optimum group representatives via inspection of the envelope frequency spectrum diagram. Ideally the harmonic feature should illustrate effective separation for all classes. Pressure signal groups are formed in a particularly balanced manner. The first three groups being formed evenly within a distance of 0.1 and as such are especially homogeneous.

Figure 6.8 highlights within group similarity with practically identical scatter plots using either harmonic feature 3 or 5. Similarly identical plot series are achieved using a chosen feature, f_i , as a representative of group i plotted against any feature f_j from any other group, j . For example feature 6 plotted against any of the group of features 9, 12, 14, 15 or 24. Clearly this indicates the repeated capabilities of group members and suggests a reduced parameter model is practical. Harmonic 25 is the only exception being unlike all others and should be considered an individual.

It should be noted that Euclidean distance is variable dependent hence the difference in scales. Further the same harmonics are not grouped together for each of the signal measurements considered despite their having been generated simultaneously. Also of note, whilst the pressure signal envelope harmonics form particularly uniform homogeneous groupings their use is not pursued due to the invasive nature of signal collection.

6.2.4 Investigative cluster scatter plots

Proximity of measurements using envelope harmonic features 4 and 7 alone were investigated as they fell into different cluster groups, hence are assumed heterogeneous. Previously features 4 and 7 have been shown to have superior discriminating powers in construction of mRVM models [2]. Figure 6.9 shows clear grouping of observations by class using these two harmonic features alone. Further 2-D plots illuminate the separation potential of different envelope harmonics.

Figure 6.10 highlights the similarities between harmonic features 6 and 7 alongside the obvious differences in harmonic 12 whereas Figure 6.11 suggest features 4 and 6, both from the 'independent' group of features, would be ideal for separation of the healthy and ICL classes.

Clearly the healthy and ICL measurements are readily formed into distinct groups so highlighting the possibility of discrimination between these classes using just two envelope features derived from the second stage vibrations of the compressor.

Each class exhibits fundamental characteristics which can be utilised in differentiating between machine states. Class profiles can be further visually investigated using multivariate graphical techniques, Andrews plots for example.

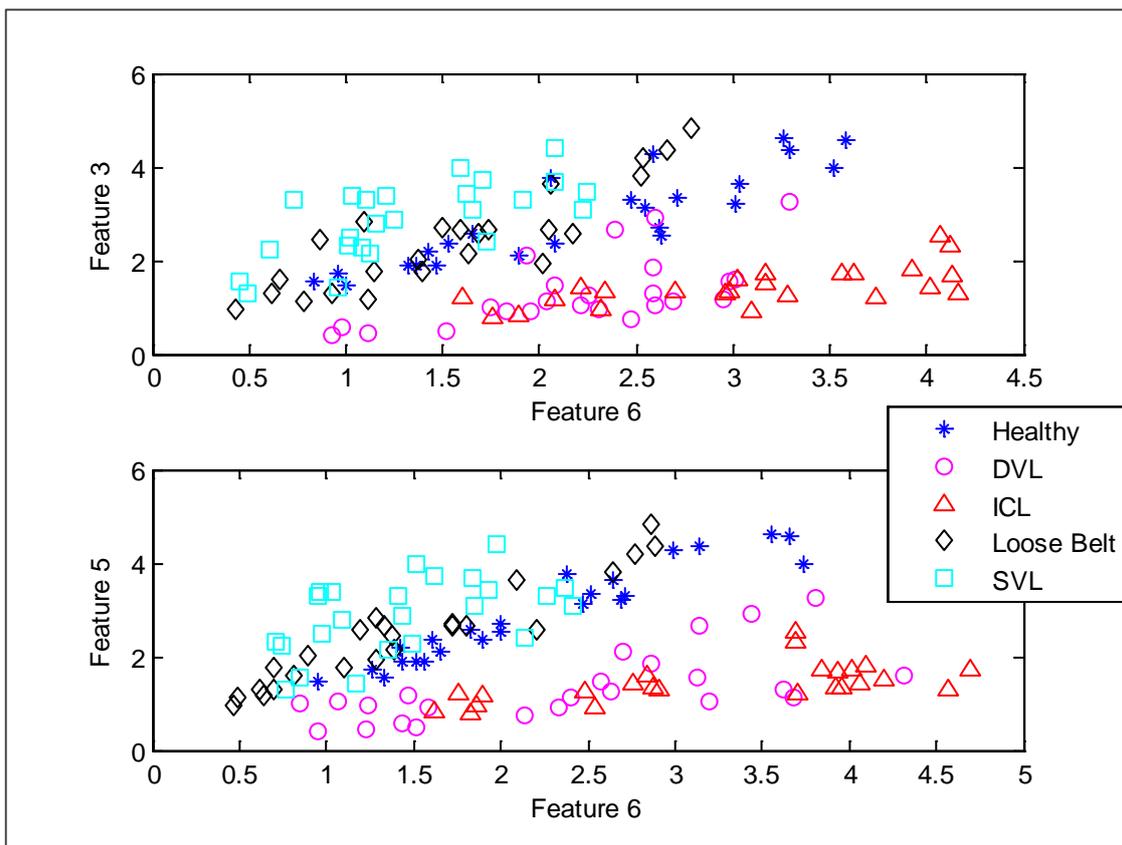


Figure 6.8 Scatter Plots Showing Classification Similarities for Homogeneous Harmonics 3 and 5 from the Second Stage Vibration Signal.

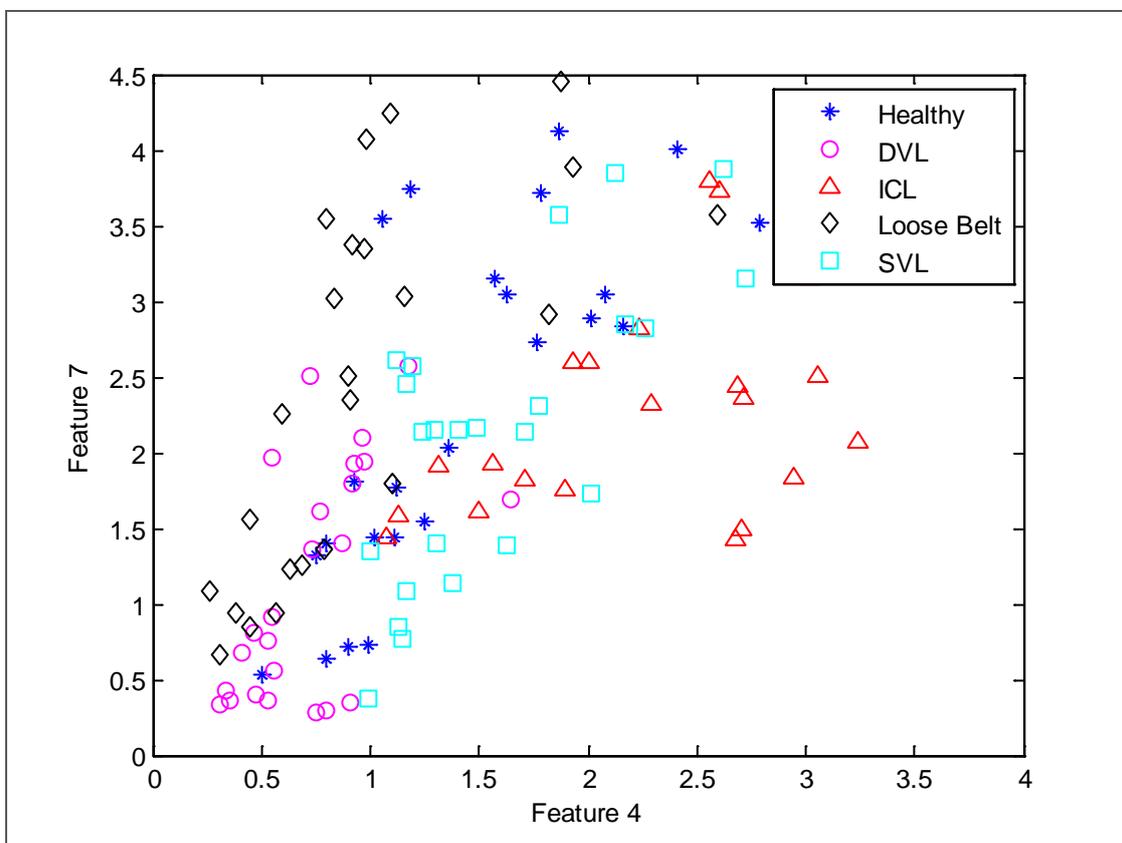


Figure 6.9 Scatter Plot: All Cases Using Envelope Spectrum Features 4 and 7 from the Second Stage Vibration Signal.

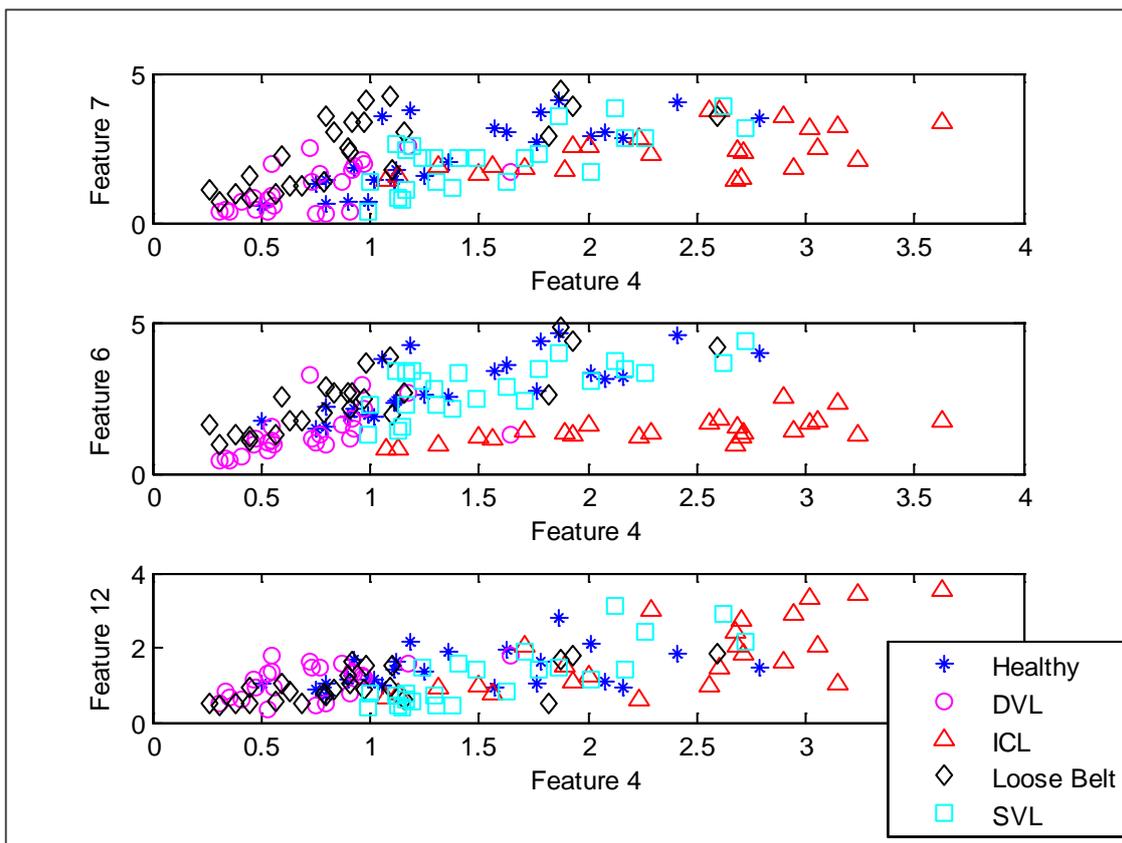


Figure 6.10 Comparison of Homogeneous and Heterogeneous Feature Pairings.

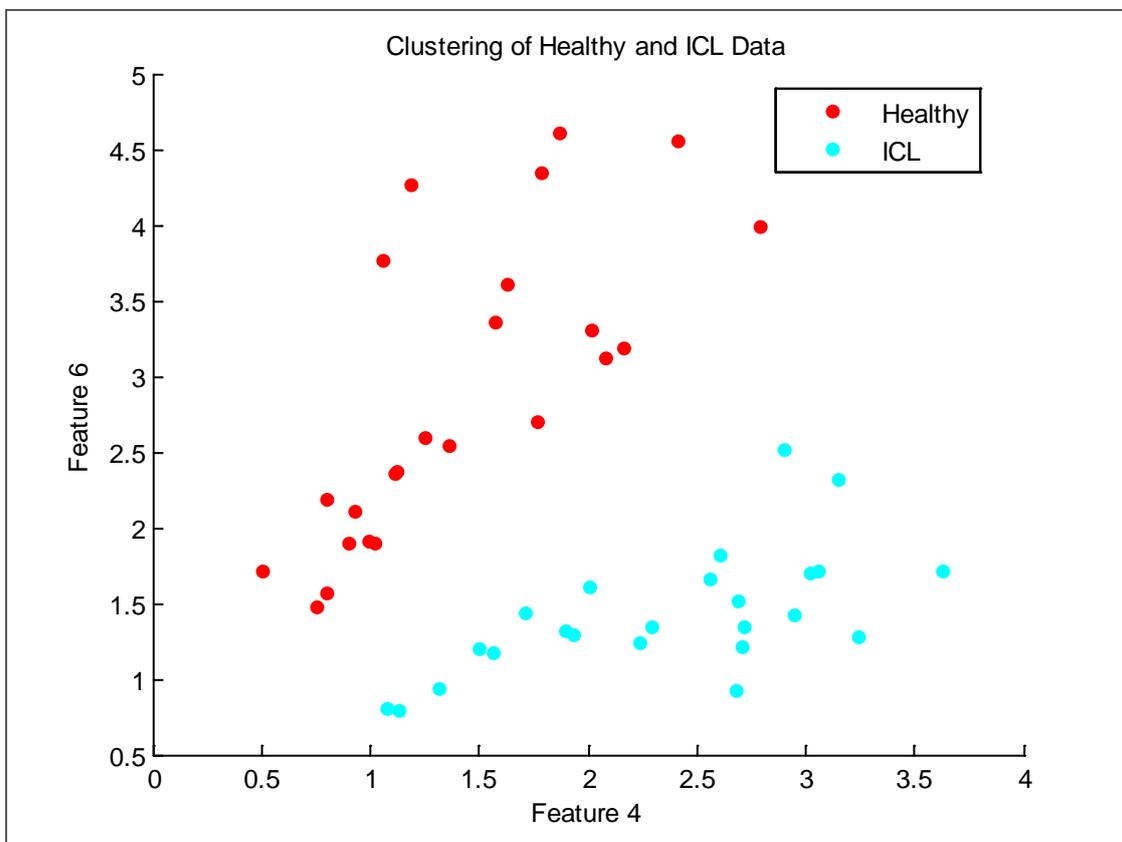


Figure 6.11 Two-Dimensional Cluster Plot Displaying the Distinct Groups of Healthy and ICL Using Envelope Features 4 and 6. (Second Stage Vibration Signal)

6.3 Discriminant Analysis

Choice of input parameter is of paramount importance hence the value of optimum selection. Previously, although highly informative, harmonic features 4 and 7 proved insufficient to fully separate the two class groups 'healthy' and 'ICL', Figure 6.12. Whereas Figure 6.13 displays a particularly successful discrimination between the same two classes using harmonic features 4 and 6 and is offered here to visually demonstrate the principle for the two dimensional two group case.

Ideally a discriminant function would be developed for separation of all classes. However, since this would necessarily be as high a dimension as 12, suggested by the envelope feature clustering in the dendrograms it is not possible to offer a visual representation in this format.

All cases are clearly grouped by categorical class using only features 4 and 6, Figure 6.14, with the ICL cases once again almost entirely separate. However, further information is required to achieve complete class isolation. Addition of features to represent each of the cluster groups identified through the CA plus all the 'independent' set supplies greater explanatory power to the discriminant function. Achievement measured through classification success rate is akin to the goodness of fit improvements in multivariate regression analysis [37, 42, 73 and 91]. Additional measures of overall fit of the discriminant function are available. Most commonly stipulated values being Wilkes Lambda or D^2 which measure the degree to which the group means differ in multivariate analysis of variance (MANOVA). Additionally the partial F-values indicate which variables have the greatest impact on a model. A higher partial F-value implying a variable has greater impact on the discriminant function [37].

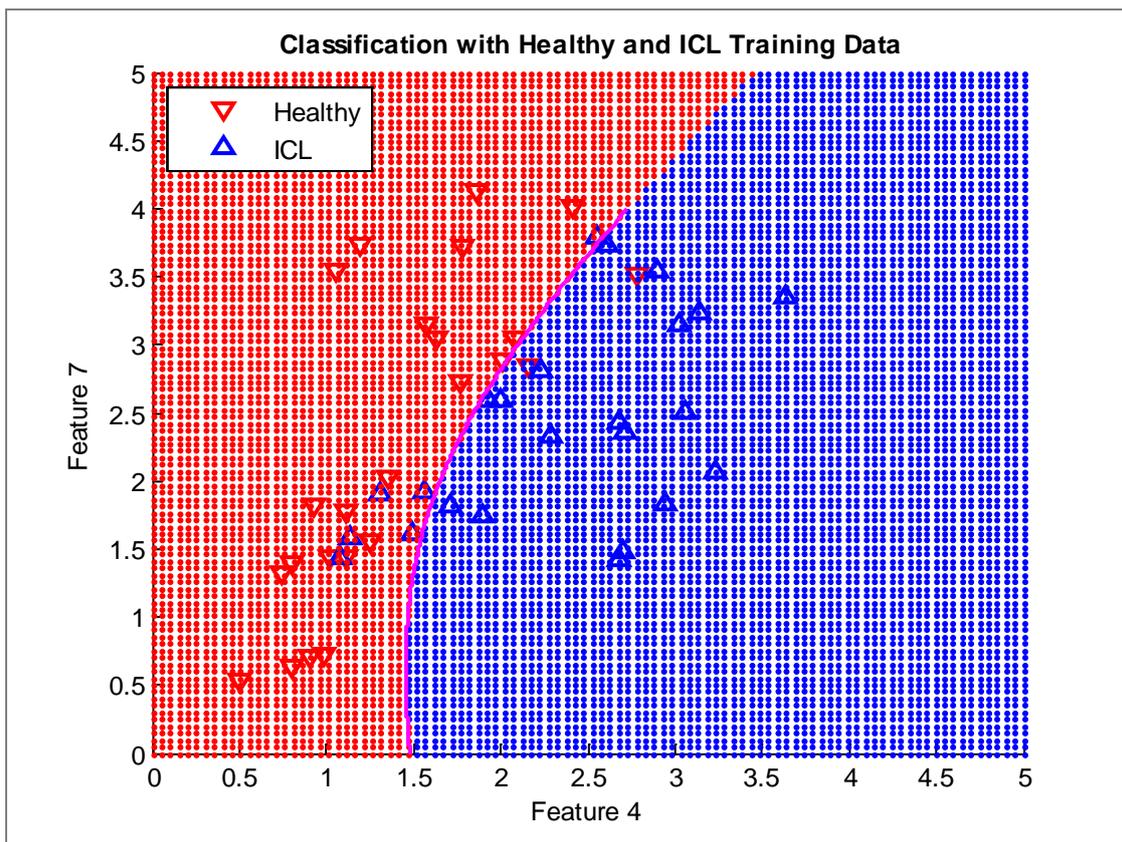


Figure 6.12 Discriminant analysis Using Features 4 and 7 To Separate the Healthy and Inter-Cooler Leak Classes.

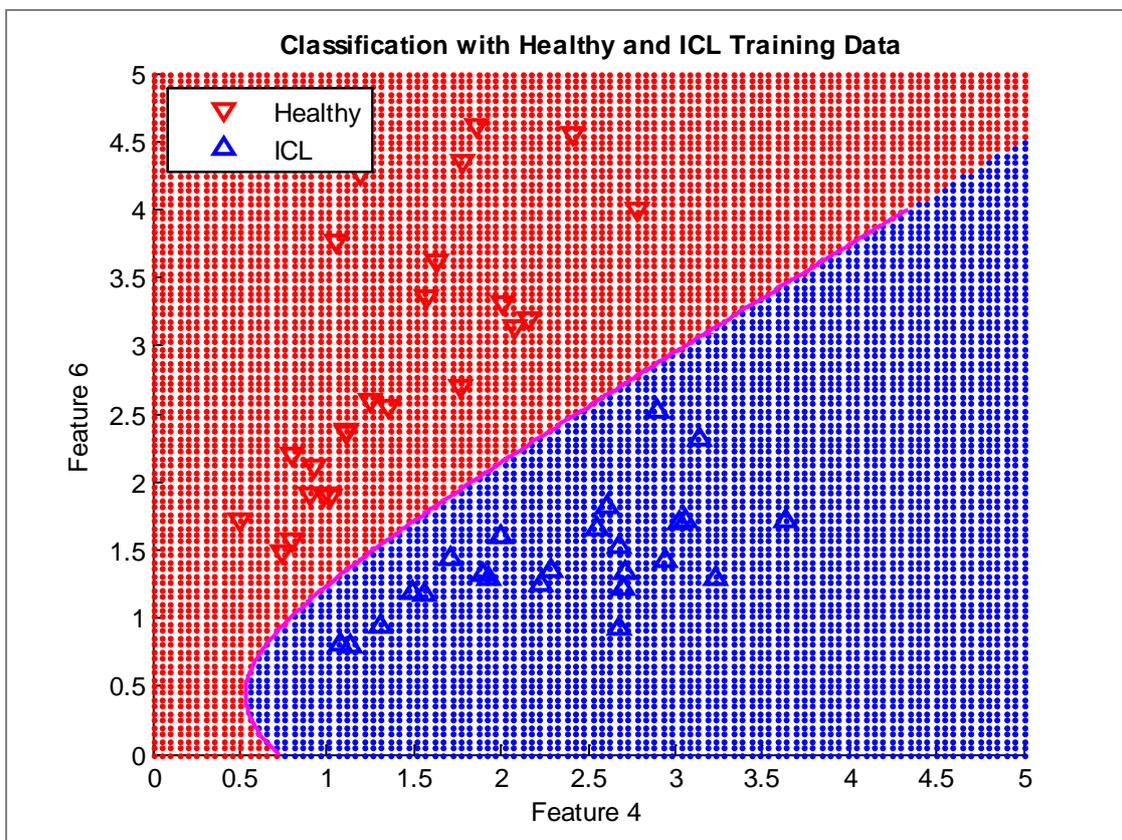


Figure 6.13 Successful Discrimination between the Healthy and ICL using Features 4 and 6.

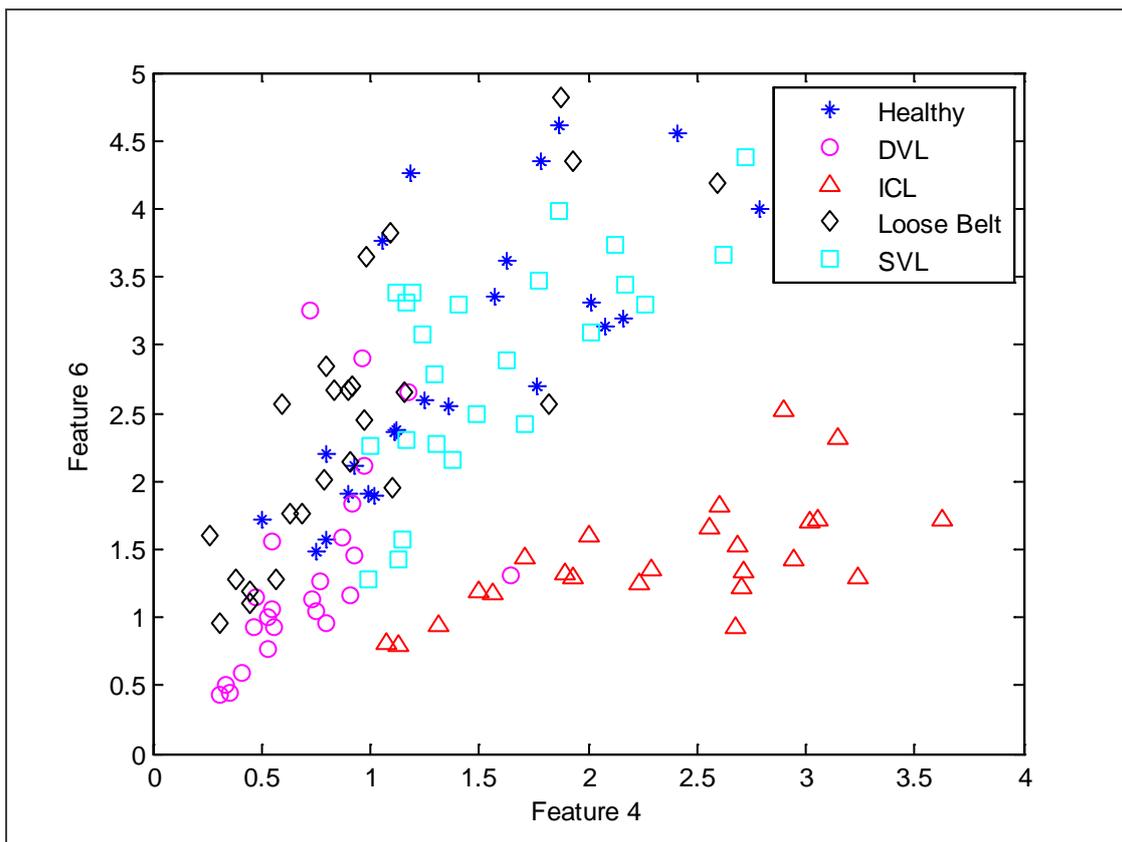


Figure 6.14 Scatter Plot of All Five Classes using Features 4 and 6 Alone.

6.4 Naïve Bayes Classification

This section explores the potential of the NB technique to correctly classify cases firstly considering the two class instance then all five classification groups.

Consider a NB classification system using just two input parameters from the second stage vibration signal. A simple classification rule was established using harmonic features 4 and 6 alone, although not highly successful it illustrates the NB technique. The classification tree, Figure 6.15, provides a useful visual method of classifying any further sample using measurements on features 4 and 6.

For example a case with a feature 4 amplitude of 0.8 and a feature 6 amplitude of 2.2, i.e. $f_4 = 0.8$ and $f_6 = 2.2$, would be allocated to the healthy group after passing through six decision nodes, detailed in Table 6.3. Allocation paths are easily identified on the tree plot Figure 6.15.

TABLE 6.3 CLASSIFICATION STEPS USING MEASUREMENTS ON FEATURES 4 AND 6

Step 1	$f_4 < 0.985049$	True (move left)
Step 2	$f_6 < 1.17358$	False (move right)
Step 3	$f_6 < 2.32193$	True (move left)
Step 4	$f_4 < 0.717411$	False (move right)
Step 5	$f_6 < 1.47048$	False (move right)
Step 6	$f_4 < 0.834252$	True (move left)
Decision:		Allocated to the Healthy group.

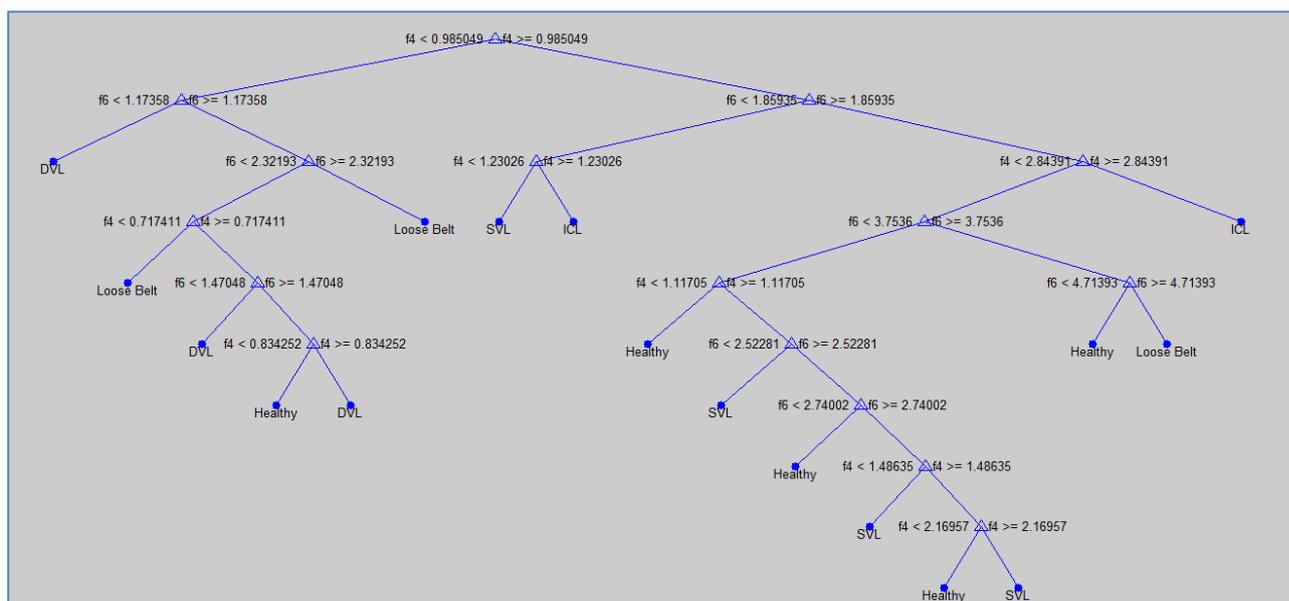


Figure 6.15 Naïve Bayes Classification Tree Using Two Input Parameters (Features 4 and 6).

Overall classification rates using just two input parameters even when restricted to the two group case were relatively poor with approximately 60% success rates.

However, for the two group case 100% success rate was achieved using five input parameters. Envelope features [2, 7, 9, 12 and 17], from the second stage vibration signal were 100% successful in separating the two classes 'healthy' and 'ICL'.

Consideration of all five classes required a much larger set of input parameters to achieve reasonable classification success rates. Classification success rates were calculated for a number of models constructed using the features indicated through the cluster analysis,

Table 6.4. Exploratory classification analyses were performed using incremental feature sets based on the cluster analysis groupings. Initially the group one harmonic, 4, and a group three representative, 6, were added as the previously demonstrated effective two parameter model. Further parameters were iteratively included to represent each group. However, whilst inclusion of harmonic 9 from the outsider group resulted in a 16% improvement in the overall classification rate on the original two parameter model, addition of certain parameters was found to reduce the overall success rates. Closer inspection showed that these parameters may be of considerable benefit in identifying a particular fault but to the detriment of other group classifications. The highest classification rate achieved was 82% using both the 10 parameter model [3, 4, 6, 7, 8, 9, 10, 11, 14 and 15] and the 12 parameter model [adding 13 and 14]. Further inspection of the scatter plot, Figure 6.14, and the cross

classification bar chart, Figure 6.16, reveal the ICL group to be almost entirely separate with a single DVL case embedded in the ‘wrong’ cluster. Classification confusion between the overlapping groups of healthy, DVL and LB is similarly obvious. As expected overlapping groups with very similar measurements on the parameters used are far more difficult to distinguish between. Individual parameters have greater capacity to detect specific deviations and a balance of capabilities is required for the best overall modelling success. For example inclusion of harmonic 14 appears to increase the models ability to correctly classify SVL faults although a greater number of incorrect assignments to SVL are also observed in its presence.

Perfect 100% classification rate is indicated if the classification matrix equates to 24I (I being the 5 by 5 identity matrix). Specific model comparisons are given in Table 6.4 and Table 6.6. Input parameter set [3, 4, 6, 9 and 11] gave 81% overall classification rate. Adding 14 to the input parameter set [3, 4, 6, 9, 11 and 14] gave 78% classification rate, an overall reduction in correct allocations with only one misallocated SVL but fewer correct allocations to all other groups except DVL. Further evidence input parameter capabilities are complex and their effects interconnected.

Overall success rates increased across all five classes once the full ten feature model was established as seen Figure 6.16 in and Table 6.5. Ironically inclusion of all 32 envelope harmonics as input parameters resulted in an overall reduction in classification rate across all five classes. Table 6.6 gives a direct comparison per class to the 10 parameter model.

TABLE 6.4 EFFECT ON CLASSIFICATION RATES OF INCLUSION OF HARMONIC 14

Class	Input parameter set [3, 4, 6, 9 and 11] (81%)	Input parameter set [3, 4, 6, 9, 11 and 14] (78%)
Healthy	17 3 0 4 0	16 3 0 5 0
DVL	1 23 0 0 0	1 23 0 0 0
ICL	1 2 20 0 1	1 3 18 0 2
LB	4 2 0 16 2	6 2 0 13 3
SVL	1 1 0 1 21	1 0 0 0 23

TABLE 6.5: CLASSIFICATION RATES FOR ALL 5 CLASSES USING THE SECOND STAGE VIBRATION ENVELOPE HARMONICS.

Number of model parameters.	Harmonics	Miss classification rate	Classification success
2	[4, 6]	48/120	60%
3	[4, 6, 9]	29/120	76%
4	[3, 4, 6, 9]	22/120	82%
5	[3, 4, 6, 9, 11]	23/120	81%
6	[3, 4, 6, 9, 11, 14]	27/120	78%
10	[3, 4, 6, 7, 8, 9, 10 11, 14, 15]	22/120	82%
12	[3, 4, 6, 7, 8, 9, 10 11, 13, 14, 15, 16]	22/120	82%
14	[2, 3, 4, 6, 7, 8, 9, 10 11, 13, 14, 15, 16]	22/120	82%
15	[2, 3, 4, 6, 7, 8, 9, 10 11, 12 13, 14, 15, 16]	24/120	80%
32		30/120	75%

TABLE 6.6 COMPARISON OF CLASSIFICATION BY CLASS

Class	10 parameter model classification matrix	32 parameter model classification matrix
Healthy	17 4 0 3 0	16 4 0 4 0
DVL	0 24 0 0 0	1 23 0 0 0
ICL	0 3 19 0 2	0 2 20 1 1
LB	5 1 0 17 1	6 6 0 12 0
SVL	1 2 0 0 21	1 4 0 0 19

TABLE 6.7 CLASSIFICATION SUCCESS RATES PER NUMBER OF GROUPS AND PER MODEL

Model type and input parameters utilised.	2 groups (Healthy and ICL)	5 groups
DA using 2 input parameters [4, 6]	100%	
NB using 2 input parameter [4, 6]	94%	53%
NB using 5 input parameters [2, 7, 9, 12, 17]	100%	64%
NB using 15 input parameters [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]	n/a	80%

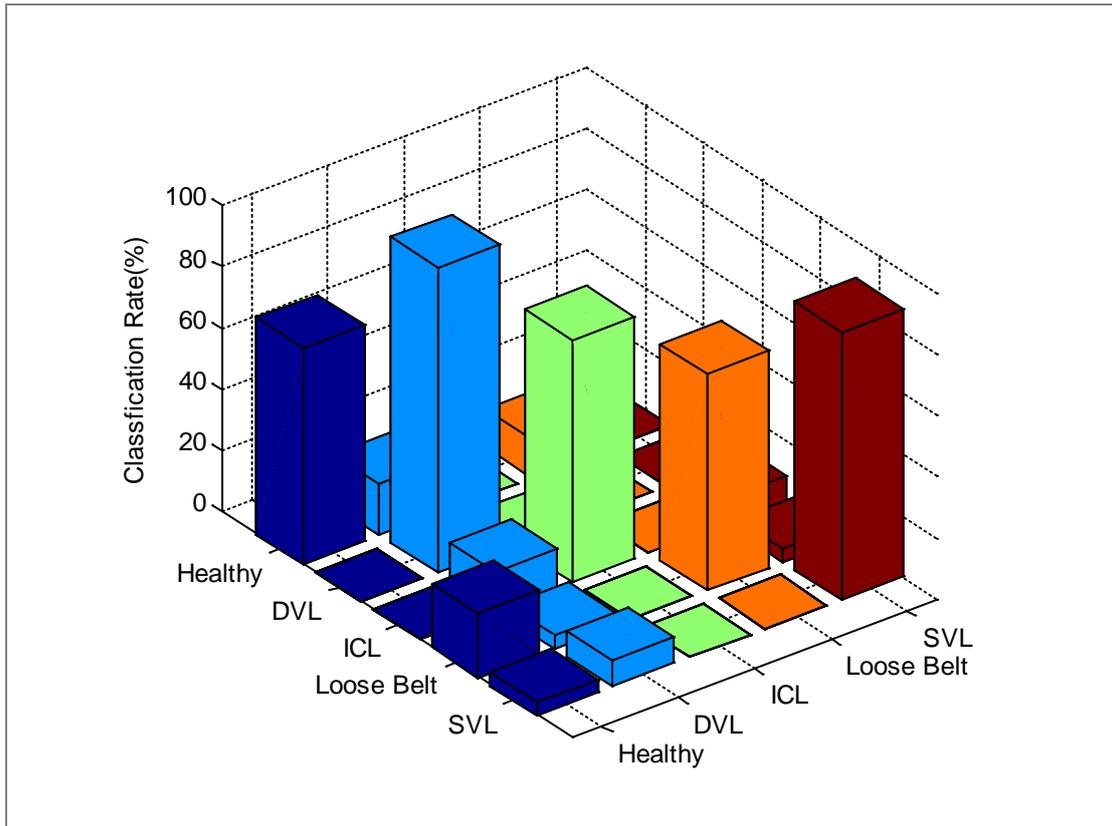


Figure 6.16 Second Stage Vibration, NB 10 Parameter Model With 82% Successful Classification Across All 5 Groups.

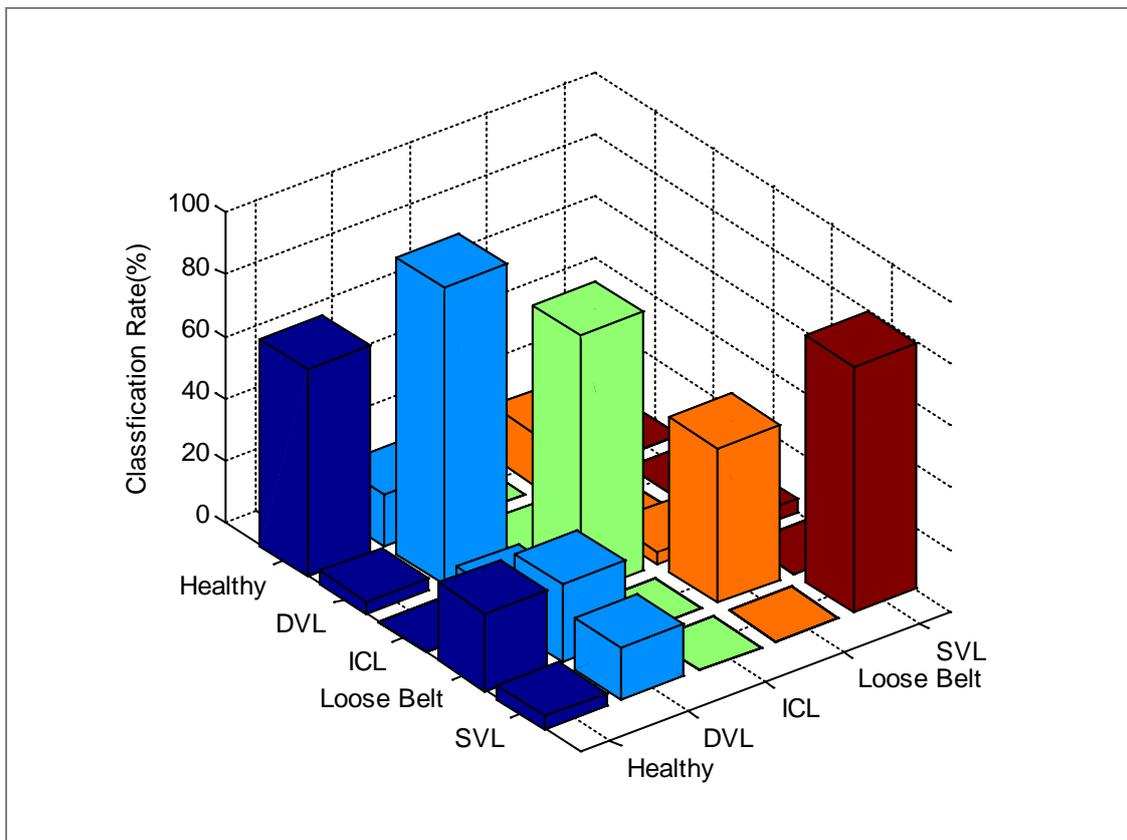


Figure 6.17 Second Stage Vibration, NB 32 Parameter Model With 75% Successful Classification Across All 5 Groups.

6.5 Summary

Inspection of Fourier profiles allows a visual appreciation of the potential for class separation. Class profiles which are most distinct being more readily identifiable. Fourier profiling using Andrews' plots gave a clear indication which classes would be most easily separable and which might be more problematic.

Clustering variables by proximity measures such as Euclidean distance gives insight into variable properties and similarities. Like variables being allocated to homogeneous groups. Variable clustering using different output signal harmonic amplitudes highlighted both cluster distances and cluster patterns were signal dependent.

Within cluster variables were demonstrated to have almost identical explanatory powers with superior parameter sets being selected across the range of heterogeneous clusters. Notably scatter plots constructed with the sixth harmonic amplitudes and either 16th or 17th harmonics were practically identical. Nevertheless, replacing harmonic 7 with harmonic 6 alongside harmonic 4 delivered superior results.

Classification success varies extensively depending on the number of groups considered and the number of parameters incorporated in the model. A two-dimensional DA gave perfect classification in the two group case (healthy and ICL) as did a five parameter NB model. However, considering all five classes simultaneously requires far greater model complexity and was not achieved using DA although a 10 parameter NB model accomplished 82% success rate, Table 6.7.

Again model complexity hence computational efforts are significantly reduced by prior evaluation of variables. Enabling selection of a reduced number of heterogeneous input parameters to ensure maximum explanatory power across all classes. The NB classification tree established using two input parameters provides a useful visual method for future sample classifications. Classification success rates were calculated for a number of NB models using the features indicated through the cluster analysis. The highest classification rate achieved across all five classes was 82% using various input parameter sets. Although on one occasion requiring 15 input features, it should be noted that 15 parameters exceeds the maximum number permitted for many algorithms. However, since the full 32 parameter model realised reduced overall classification rates it would seem prudent to focus on parameter set quality rather than improved means of manipulating large numbers of variables.

Chapter 7

Multivariate Classifiers Using Variable Reduction

Methods

This chapter investigates the efficiency of multivariate classifiers established through the variable reduction methods of principal component analysis (PCA), factor analysis (FA) and support vector machines (SVM).

Previously input parameters have been selected through variable clustering techniques prior to model building i.e. a reduced number of heterogeneous variables were selected upon which to establish classifiers.

This next section focusses on classifiers constructed through variable reduction techniques. Firstly using PCA, whereby all variables, the 32 envelope harmonics, are input and a number of PCs are established based on these original variables. Secondly through construction of SVMs.

The purpose being to compare and contrast efficiency of the two methods of variable selection prior to establishing a classifier, as discussed in Chapter 6, and variable reduction techniques, Chapter 7.

7.1 Principal Component Analysis of Envelope Harmonics

The focus of the analysis in this section was to seek underlying principal components (PCs) to define the variation in the system which could then be used as input variables in construction of classifiers to identify machine health. As emphasised in Chapter 3, ideally a smaller number of PCs will account for the vast majority of the total variance. Hence a reduced number of highly representative new variables are established each of which incorporates elements of all the original variables.

Envelope harmonics 1 to 32 for the second stage vibration signal were stored for each of the 120 observations across the 5 classes. PCA was performed on the subsequent covariance matrix. Whilst all 32 PCs were retained relatively few were expected to contribute towards construction of the diagnostic model. Decisions as to how many of the constructed PCs to include being application dependent in part. For example it may be felt that a model must account for a minimum percentage of the variation hence the number of PCs is chosen to reflect this. Alternatively there may be limits placed on model dimensionality. Often a diagrammatic representation of the PC variance is used to assist decisions. Using a scree plot of the eigenvalues the gradient change is generally considered as the cut-off point. A reasonable assumption as the gradient change represents a more abrupt reduction in the eigenvalue size. Since only the first three PCs had eigenvalues greater than one, hence contribute substantially towards the total variation in the system, a two and three PC model were investigated in this analysis. Any PC with an eigenvalue greater than one is considered to contribute 'more than its' share' towards explaining the variance in the system. With $\lambda_1 = 12.3416$ the majority of the variance, almost 60%, was incorporated in PC1 with an additional 14% from the second PC and approximately 8% from the third, results are summarised in Table 7.1.

Thus the first two PCs accounted for approximately 73% of the total variation in measurements. Clearly when all the cases are plotted against these first two PCs the SVL group, Figure 7.1, is seen to be entirely separate from all other classes having the lowest scores on both the 1st and 2nd principal components. For further information PC scores by class are summarised in Table 7.3.

Identifying the SVL fault is particularly straightforward the first two PCs forming a sufficiently sophisticated model for successful classification. Even using just two PCs all other cases are reasonably well grouped by class and with the addition of a third PC, which increases the cumulative sum of the model to 81%, classification rates improve further still. The fourth PC has a variance very close to one and could reasonably be incorporated to further improve model accuracy. However, the

remaining PCs all have variances less than one and so offer increasingly negligible contributions in deterministic terms thus further classification improvements are not realistic using PCA. The cumulative sums for the first 14 principal components are reported in Table 7.2. Whilst 81% of the total variation in the system is accounted for by the first three PCs alone the first ten PCs are required to achieve 95%.

TABLE 7.1 SUMMARY OF EIGENVALUES AND PC VARIANCE FOR THE FIRST THREE PCs.

Variance (PC1)	$\lambda_1 = 12.3416$
Variance (PC2)	$\lambda_2 = 2.9228$
Variance (PC3)	$\lambda_3 = 1.7662$
Total variance (PC1+PC2+PC3)	$\lambda_1 + \lambda_2 + \lambda_3 = 17.0321$
Total variance in system	$\sum \lambda_i = 21.201$
Cumulative sum of variance	$17.0321 / 21.201 = 0.8103$

TABLE 7.2 CUMULATIVE SUMS FOR THE FIRST 14 PRINCIPAL COMPONENTS

(1) 0.5872	(2) 0.7262	(3) 0.8103	(4) 0.8556	(5) 0.8824	(6) 0.9037	(7) 0.9198
(8) 0.9328	(9) 0.9449	(10) 0.9545	(11) 0.9616	(12) 0.9673	(13) 0.9713	(14) 0.9750

TABLE 7.3 CLASS SCORES ON THE FIRST TWO PRINCIPAL COMPONENTS.

Class	Range of scores on principal component 1	Range of scores on principal component 2
Healthy	[-4.351, -1.326]	[0.262, 0.576]
SVL	[-5.980, -4.576]	[-1.595, -0.280]
DVL	[-2.926, 1.291]	[0.225, 0.818]
LB	[2.304, 6.850]	[0.394, 2.379]
ICL	[1.301, 4.677]	[0.445, 2.033]

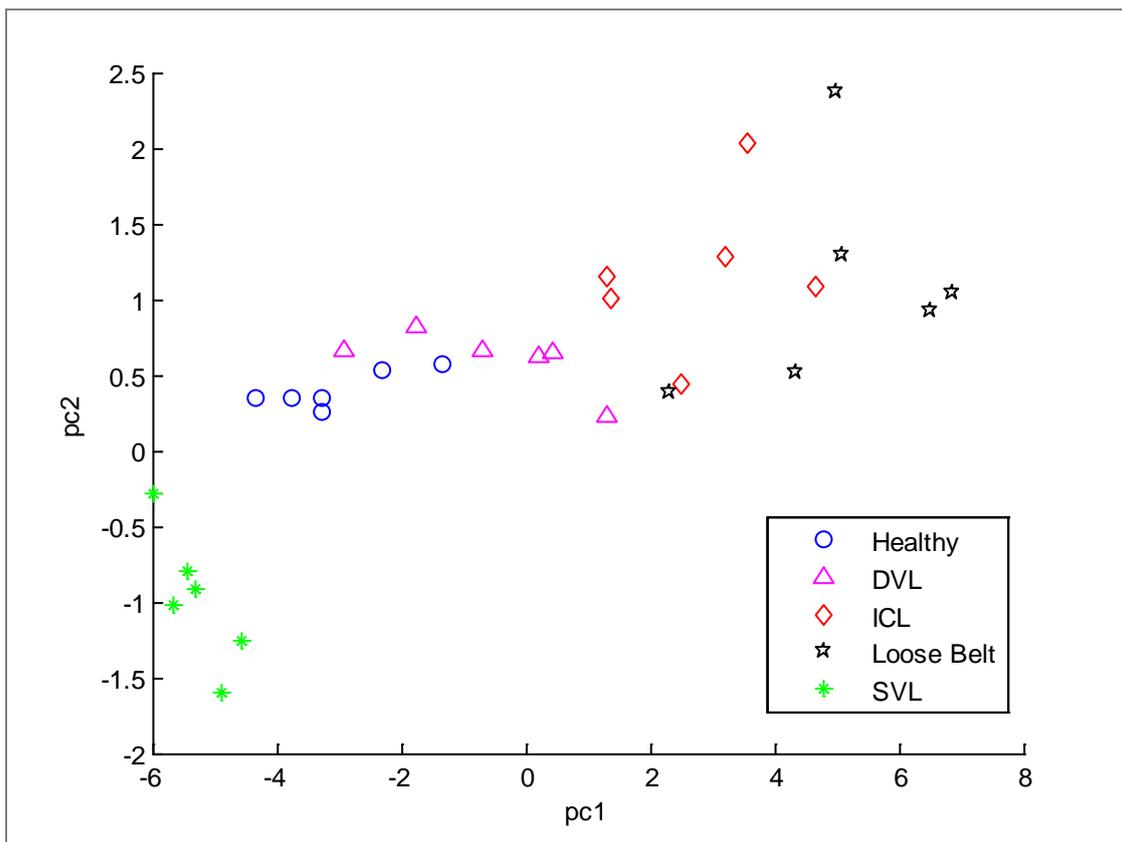


Figure 7.1 Fault Clustering Using the First Two Principal Components.

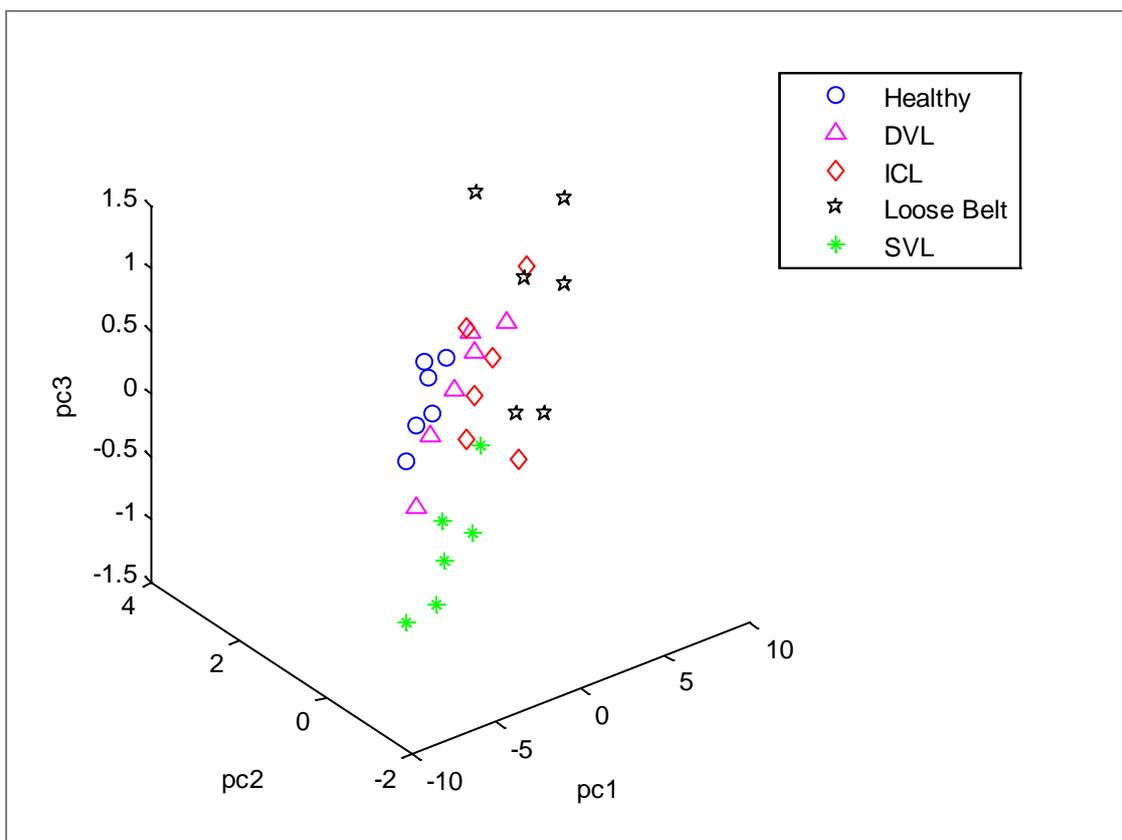


Figure 7.2 Fault Clustering Using the First Three Principal Components Accounting for 81% of the Total Variation.

7.2 Confirmatory Factor Analysis

Since there appear to be underlying generic health conditions governed by collective groups of harmonic features a confirmatory factor analysis was conducted.

Inspection of the factor loadings on the first two factors shows high factor 1 loadings for harmonic features 6 and 7 thus these two harmonics are highly correlated as might be expected, also both have negligible factor 2 loadings. On the other hand harmonic feature 4 has a high factor 2 loading with much lower factor 1 loading so is less correlated with features 6 and 7 but more highly correlated with features 3 and 5.

The specific variance of harmonic 6 is 0.0632 which is close to zero so implies the variable is almost entirely determined by its common factors, in fact 93% of the variance of harmonic 6 is accounted for by factor one hence its superior deterministic power. 70% of the variance of harmonic 7 is explained by factor one. In comparison harmonic 9 has a specific variance of 0.9710 almost 100% which implies there is practically no common factor component in the variable. Indeed harmonic 9 possesses just 2% common variance in factor one and considerably less in factor two.

TABLE 7.4 FACTOR LOADINGS AND SPECIFIC VARIANCE FOR KEY HARMONICS.

Factor loadings	Envelope harmonic									
	2	3	4	5	6	7	9	12	13	14
Factor 1	0.6679	0.1804	0.3615	0.1735	0.9659	0.8343	0.1459	0.2905	-0.1865	-0.1239
Factor 2	0.6448	0.7963	0.7810	0.8083	-0.0622	0.3980	-0.0877	0.7282	0.8267	0.7816
Specific variance	0.1381	0.3333	0.2593	0.3166	0.0632	0.1455	0.9710	0.3854	0.2817	0.3738
T ² =	0.7707	0.6372								
		-0.6372	0.7707							
Log likelihood= -24.6256 with 433 degrees of freedom and highly significant association										
Chi-square = 2.6144e+003, p: 1.2702e-307.										

Hotelling's multivariate T² test is highly significant, summary details Table 7.4, thus rejecting the null hypothesis of equal means. Concluding there are significant differences in the mean envelope harmonic factor loadings.

7.3 Support Vector Machine Classification

An established machine learning technique, SVM uses data partitioning to first train the algorithm then classify the remaining test data set. Samples are randomly allocated to one data set only. Again the method highlights the benefit of careful input parameter selection.

Cases were randomly assigned to either the training or test data set, data was equally partitioned between the two. The classifier structure established first incorporated a linear kernel function using the second stage vibration envelope harmonics 4 and 7. The two group instance 'healthy' and 'faulty' was initially considered, Figure 7.3. Note for computational purposes data belong to one of two binary groups 0 or 1, healthy and not healthy respectively i.e. the 'healthy' class were represented by 0 and the 'faulty' class by 1. Whilst 77% of cases were correctly classified it is clear a very high proportion of samples were incorporated in the model which is likely to be over specified. Including a hard-margin condition in the analysis, Figure 7.4, proved equally high in sample use but with classification rate reduced to 68% no improvement. Interestingly repeat analysis using harmonic features 4 and 6 had little impact on the classification rates although the classifier format was clearly altered. Combining all faults into one group clouding interpretation and whilst the classification success might be enhanced through the use of quadratic or radial based kernel functions the primary focus of this thesis is the impact of the input parameters on model supremacy.

Further analysis was conducted with the data groups classified by 'SVL' and 'all others' since it had previously been shown sensible to assume the groups were roughly linearly separable hence inclusion of the second model embedding a hard margin hyperplane is justified. Using harmonics 4 and 7 gave improved and identical classification rates for both models of 88.33%, however, the number of samples employed in the model specification was still extremely high with 35 and 41 samples respectively. Repeat analysis using harmonics 4 and 6 gave further improved and again identical classification rates of 93.33% in addition both showed a dramatic reduction in the number of support vector samples required to specify the classifier, Figure 7.5, Figure 7.6, just three in the case of the hard-margin hyperplane. Classification success of 98% was realised in the three parameter model utilising harmonics 4, 6 and 7.

7.4 Summary

Variable reduction techniques utilise all variables creating a new set each a weighted sum of the originals. General guidance on the number of PCs to retain is application specific to an extent although usually at most two to three are preferable.

For the RC, the first two PCs accounted for approximately 73% of variation in the system and were sufficient to identify and classify the SVL. 81% classification success was achieved across all 5 classes when the first 3 PCs were retained. Directly comparable to the previous NB models for all 5 classes.

Confirmatory FA revealed underlying harmonic characteristics suggested in CA and subsequent modelling successes. Both harmonics 6 and 7 had very high loadings on factor 1 with negligible loadings on factor 2 confirming the high correlation between them. Harmonic 4 having the opposite loading pattern being uncorrelated with both 6 and 7 i.e. harmonic 4 is heterogeneous to harmonics 6 and 7 and hence explains different system variation. Heterogeneous harmonics are vital as efficient input parameter pairings. With negligible specific variance (0.06) 93% of harmonic 6 is accounted for by factor 1 hence its superior deterministic properties. Similarly with a specific variance of 0.97, harmonic 9 contributes only 2% towards the common variance hence its poor explanatory power.

SVM offer an alternative variable reduction method. High classification success rates were achieved in the studied two group cases. Unsurprisingly the greater the explanatory power of the individual input parameter the greater the classification success rate. This also resulted in fewer SVs required for specification therefore reduced computational efforts in the classification rules. 98% classification success rate was achieved for the two group case using just three input harmonic variables.

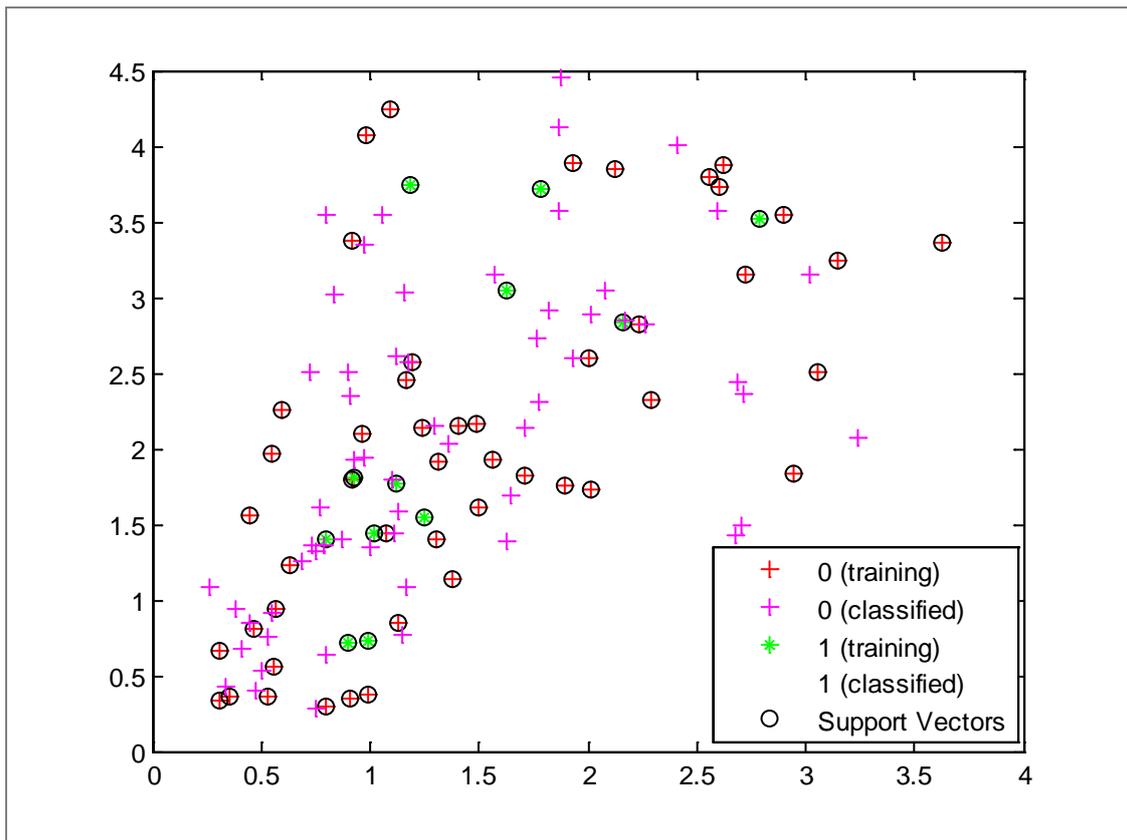


Figure 7.3 SVM Classification Linear Kernel Function ('Healthy' and 'Not healthy') Using Harmonics 4 and 7.

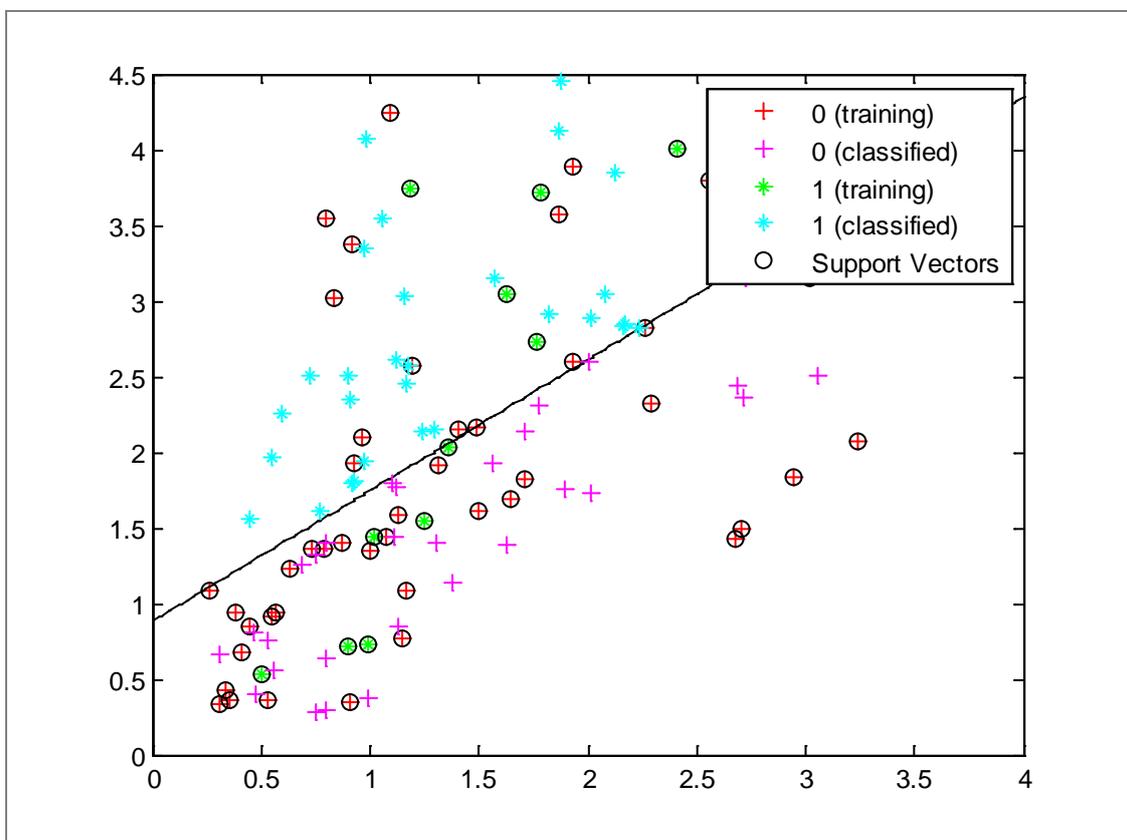


Figure 7.4 SVM Including Hard-Margin, Using Harmonics 4 and 7 ('Healthy' and 'Not Healthy').

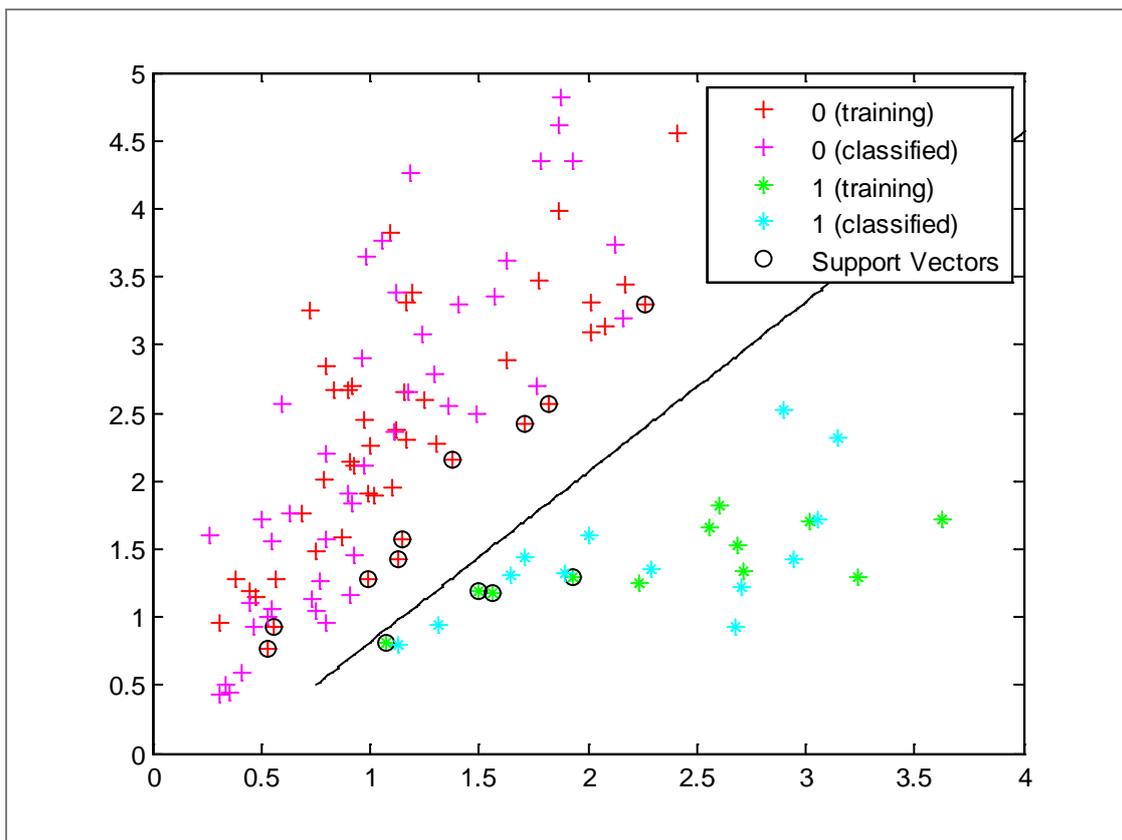


Figure 7.5 SVM Linear Kernel Using Harmonics 4 and 6 (SVL Identified).

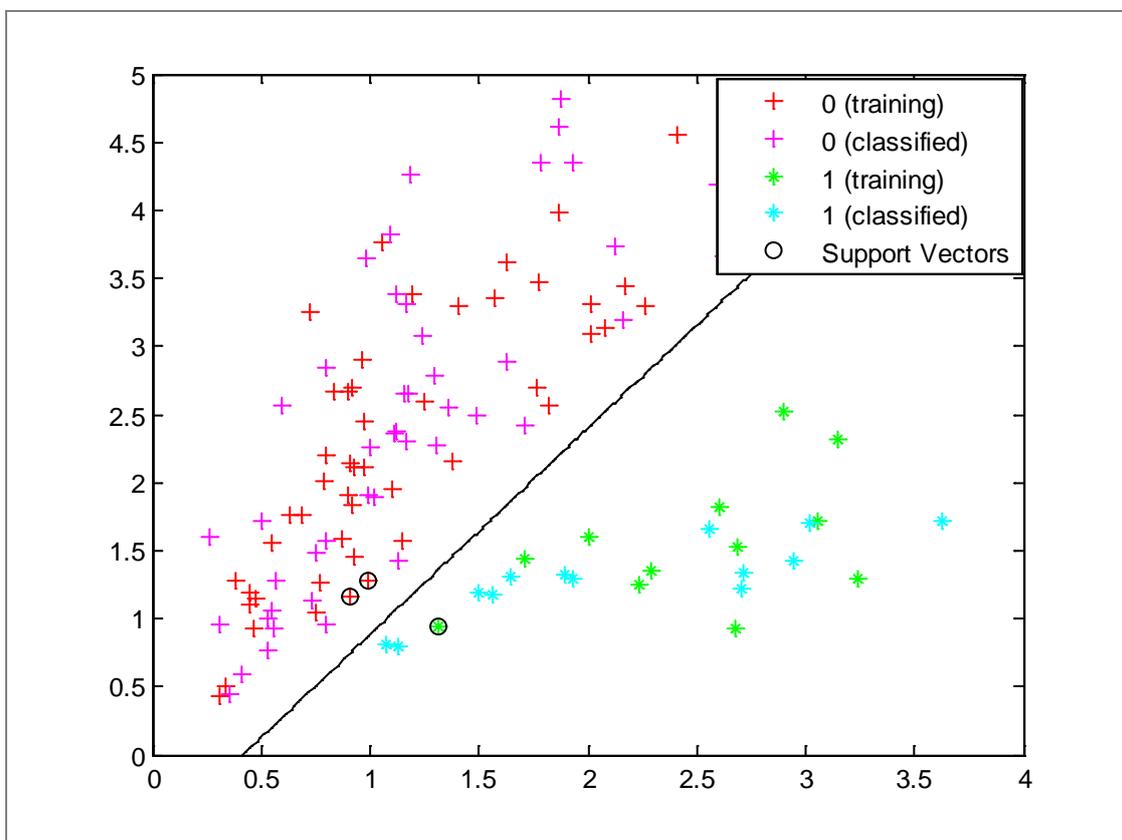


Figure 7.6 SVM Including Hard-Margin Using Harmonics 4 and 6 (SVL Identified)

Chapter 8

A Combined Approach to Variable Reduction Incorporating Signal Compression

During this chapter elements analysed in the preceding chapters are contemplated as the basis for further work beyond the remit of this thesis. The concepts studied separately in chapters 6 and 7 are jointly considered in a proposed combined variable reduction technique.

The benefits of inputting reduced numbers of variables into models in a controlled manner along with variable reduction techniques have separately been explored in the preceding chapters of this thesis. Should it be possible not only to effectively combine these two techniques but also to trim or de-noise the data signals or envelope harmonics using data compression a much reduced input parameter volume might be achievable. Thus reducing computational efforts whilst maintaining model efficiency and avoiding classifier bias.

Possible extensions to approaches are investigated in this chapter. Selection of a reduced number of input parameters being utilised in conjunction with variable reduction techniques. These conjoined methods are considered along with prior data compression. Data compression is a wavelet technique which can be 'lossy' or 'lossless' [48 and 77] terms which refer to the statistical information contained in the signal. Should data compression be accomplished which has negligible effect on the explanatory power of an input variable then its contribution to input volume reduction is clearly valuable.

8.1 Combined Variable Reduction Technique

Comparative PCA analyses were conducted using the reduced sets of input parameters suggested through CA. Ten envelope harmonics [3, 4, 6, 7, 8, 9, 10, 11, 14 and 15] were incorporated in a PCA model with visibly improved classifications. Although the first eigenvalue was not nearly so dominant as the model using the entire set of envelope harmonics, the variance accounted for by the first 3 PCs was increased to 81.45% and only the first 6 PCs were required to account for over 95%. Figure 8.1 shows each class is visibly separated with the SVL, DVL and healthy groups being particularly tightly clustered.

Including all 15 dominant envelope harmonics from the CA (the ten parameter set plus 2, 5, 12, 13 and 16) PCA gave slightly improved results both visually and cumulative percentage wise. PCs one to three accounting for 82.23% of the system variance but the first 7 PCs were required to exceed 95%.

Controlled selection of the most pertinent input parameters along with variable reduction techniques appears to be a realistic proposition. The computational savings along with improved classification accuracy being evident albeit not overly dramatic. However, further analysis incorporating CA selections into algorithms with restricted input capabilities would be expected to result in desirable outcomes. Exploratory analysis proposed for future work.

In the next two sections, 8.2 and 8.3, data compression is investigated as a further means of reducing input parameter volume.

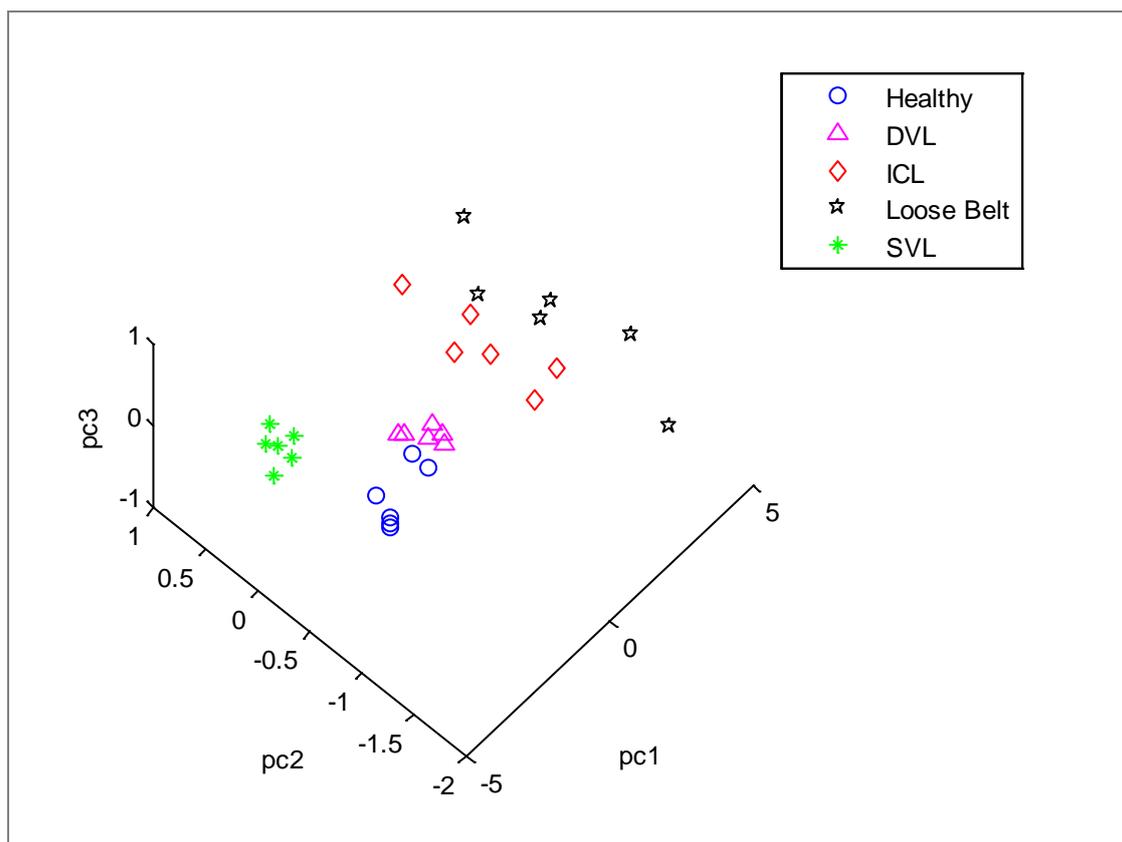


Figure 8.1 PCA Using Reduced Number of 10 Input Parameters Indicated by CA

8.2 Multiscale PCA

The aim of multiscale PCA is to reconstruct a simplified multivariate signal, starting from an original multivariate signal and using a simple representation at each of a specified number of resolution levels. Multiscale principal components analysis generalises the PCA of a multivariate signal represented as a matrix by simultaneously performing a PCA on the matrices of details of different levels. A PCA is performed on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting the numbers of retained principal components, interesting simplified signals can be reconstructed.

Signal data from the RC is used in the form of a 120 by 32 matrix of cases (24 repetitions of each of 5 operating classes) by envelope spectra harmonics. A multiscale PCA at level 5 using Matlab 'sym4' was executed with PCs retained according to Kaiser's rule. Kaiser's rule retains PCs with eigenvalues greater than the mean eigenvalue value. Signal compression results are shown for harmonics 4, 6, 7

and 9 for both the initial signal simplification and improvements on reducing the number of retained PCs, Figure 8.2 and Figure 8.3.

From a compression perspective the results are good. The percentages reflecting the quality of column reconstructions given by the relative mean square errors are close to 100%.

The cumulative percentage variation for the first seven PCs is very good from a compression perspective all values being close to the maximum 100% i.e. 97.2383 98.3299 87.5594 93.8073 87.0964 92.6310 and 97.2287.

Initially 7 components were retained according to Kaiser's rule $\lambda > \bar{\lambda}$ the vector of components retained, npc , being $npc = [1, 1, 1, 1, 1, 2, 2]$. Results illustrated in Figure 8.2. However, these initial results are improved by retaining fewer principal components, removing the first three PCs which are primarily composed of noise with small contributions to the signal is an effective albeit rather crude de-noising process, Figure 8.3.

The npc vector, the number of retained principal components selected by Kaiser's rule is updated to [0 0 0 2 2 1 3]

As expected, the rule keeps two principal components, both for the PCA approximations and the final PCA, but one principal component is kept for details at each level.

Demonstrated in Figure 8.3, results are visibly greatly improved with the noisy original signal being smoothed and simplified on reconstruction.

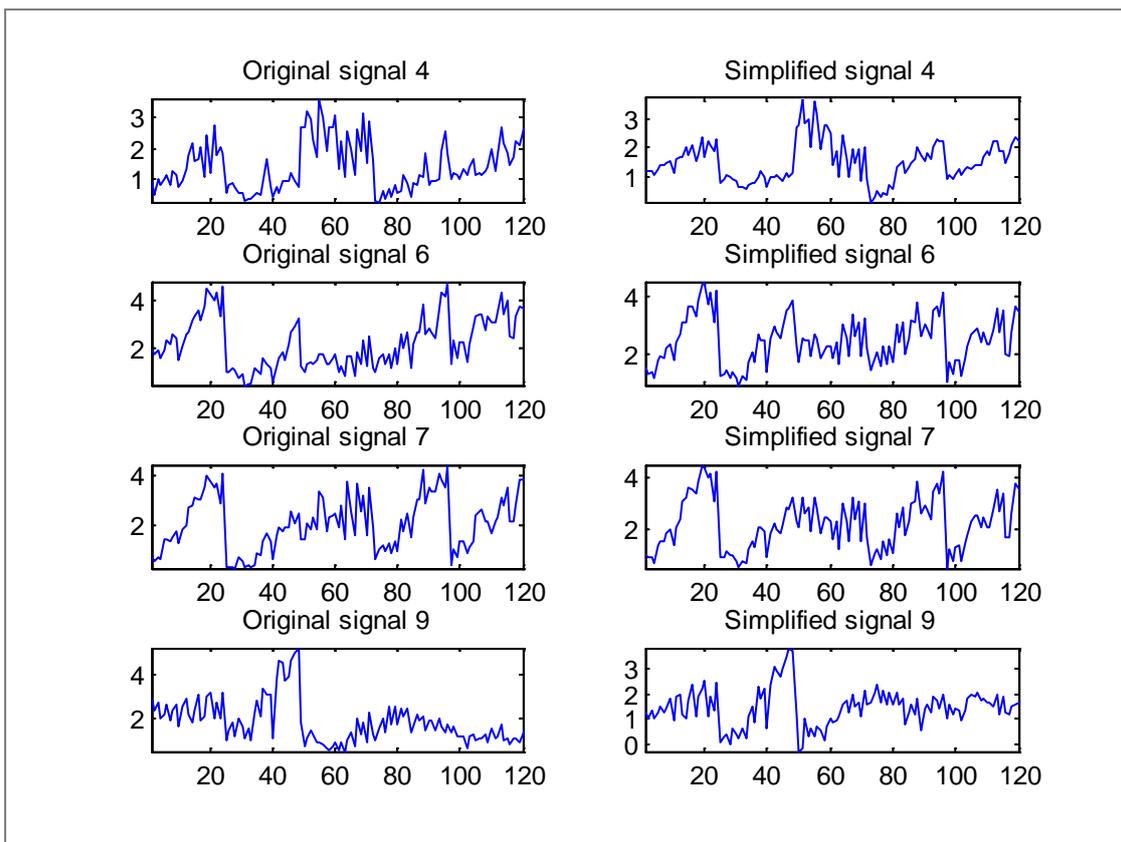


Figure 8.2 First Stage Multiscale PCA Results

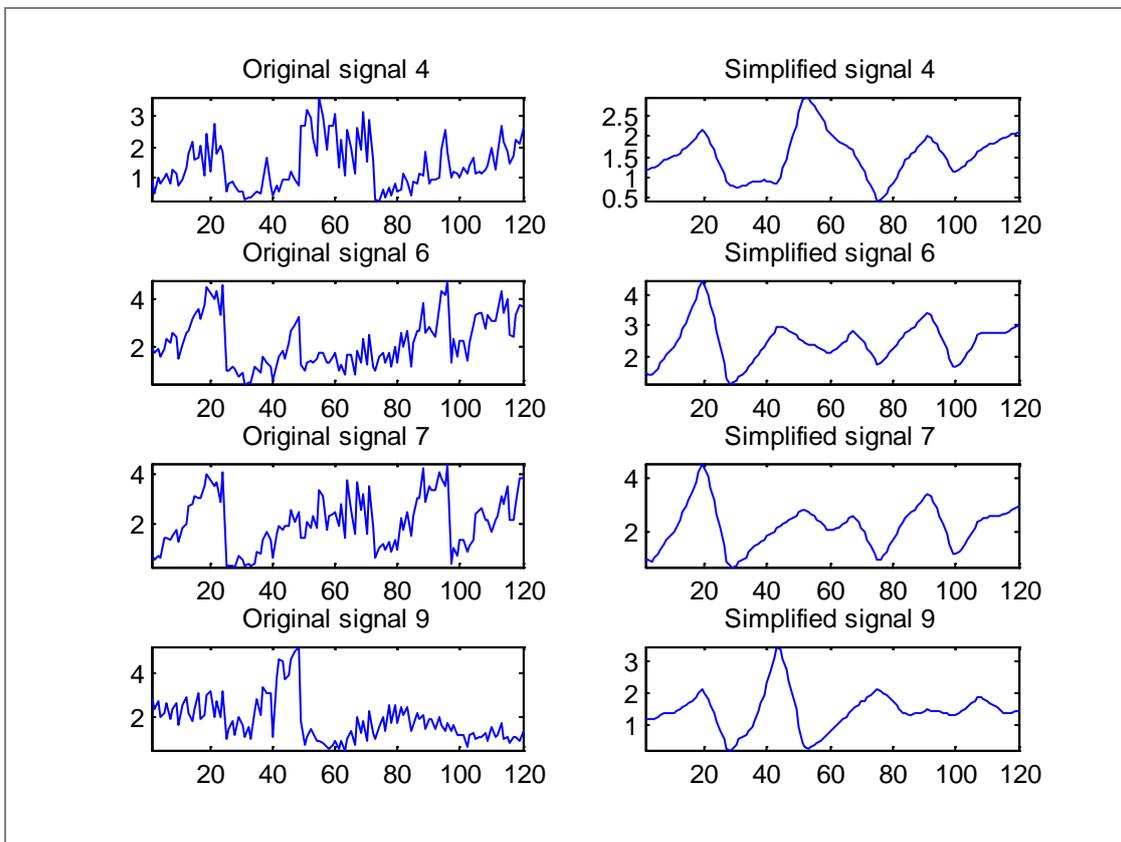


Figure 8.3 Improved Multiscale PCA Signal Comparisons.

8.3 Data Compression

Initial investigation into the potential benefits of signal compression methods was focussed on a small sample of harmonics. Previous analysis has shown that up to 32 envelope harmonics may be incorporated into the model building process in one form or another thus if signal compression is a realistic a priori de-noising technique huge cumulative volume savings can reasonably be expected overall.

Analysing the 4th envelope harmonic signal a comparison of the original signal, Figure 8.5, and the compressed signal, Figure 8.6, visibly highlights the capabilities of compression. Clearly the salient features of the original signal are preserved but minor fluctuations are smoothed and simplified. Should the smoothed compressed signal preserve the explanatory power of the original then the potential for great volume reduction in input parameters especially for large data sets is vast.

Compression appears to have an impact on the signal distribution with the compressed 4th harmonic signal being more positively skewed than the original, Figure 8.4. One-way analysis of variance (ANOVA) is significant at 97%, however, the correlation between the original and compressed signal being 0.8003 the goodness of fit statistic $R^2=64\%$ is much lower.

TABLE 8.1 ONE-WAY ANOVA 4TH HARMONIC.

Source	SS	df	MS	F	Prob.>F
Columns	0.001	1	0.00058	0	0.9732
Error	121.687	238	0.51129		
Total	121.688	239			

TABLE 8.2 ONE-WAY ANOVA 6TH HARMONIC.

Source	SS	df	MS	F	Prob.>F
Columns	0	1	0.00035	0	0.985
Error	238.24	238	1.00101		
Total	238.24	239			

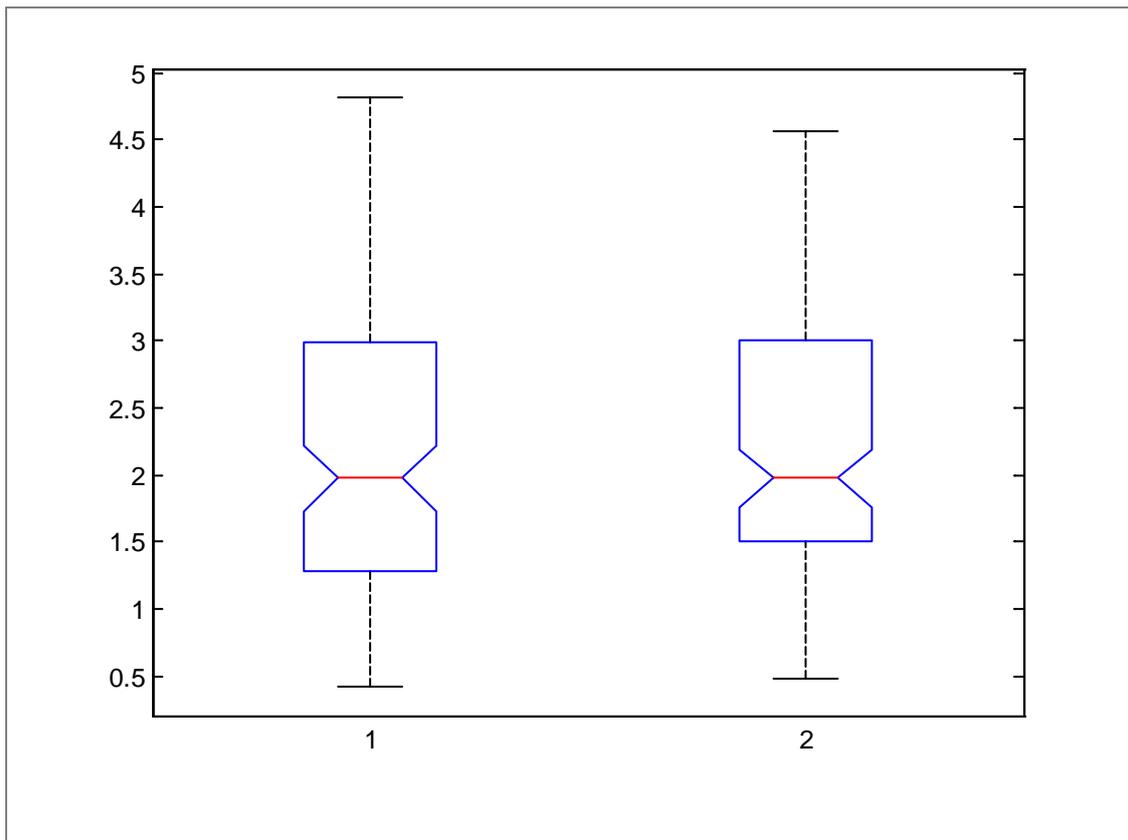


Figure 8.4 Distributional Comparison of 1. 4th Harmonic Amplitudes and 2. Compressed 4th Harmonic Amplitudes.

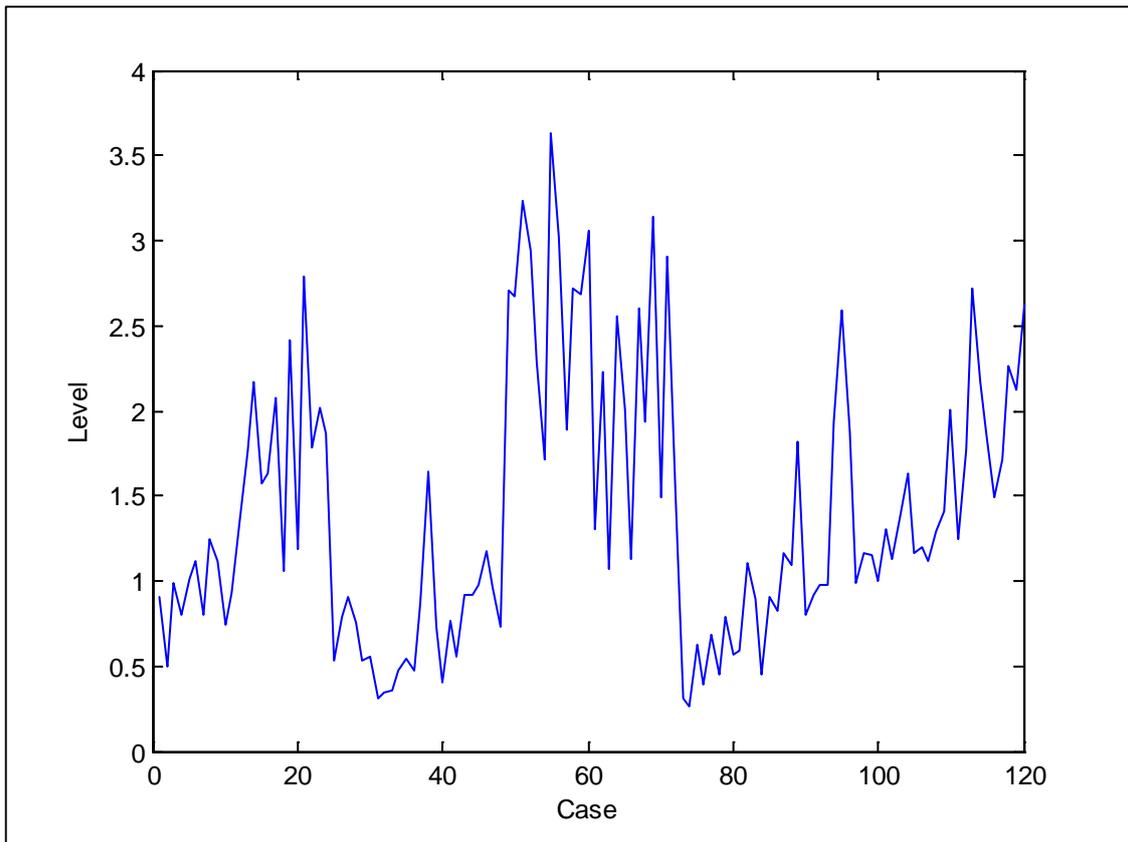


Figure 8.5 Original Signal Envelope Harmonic 4.

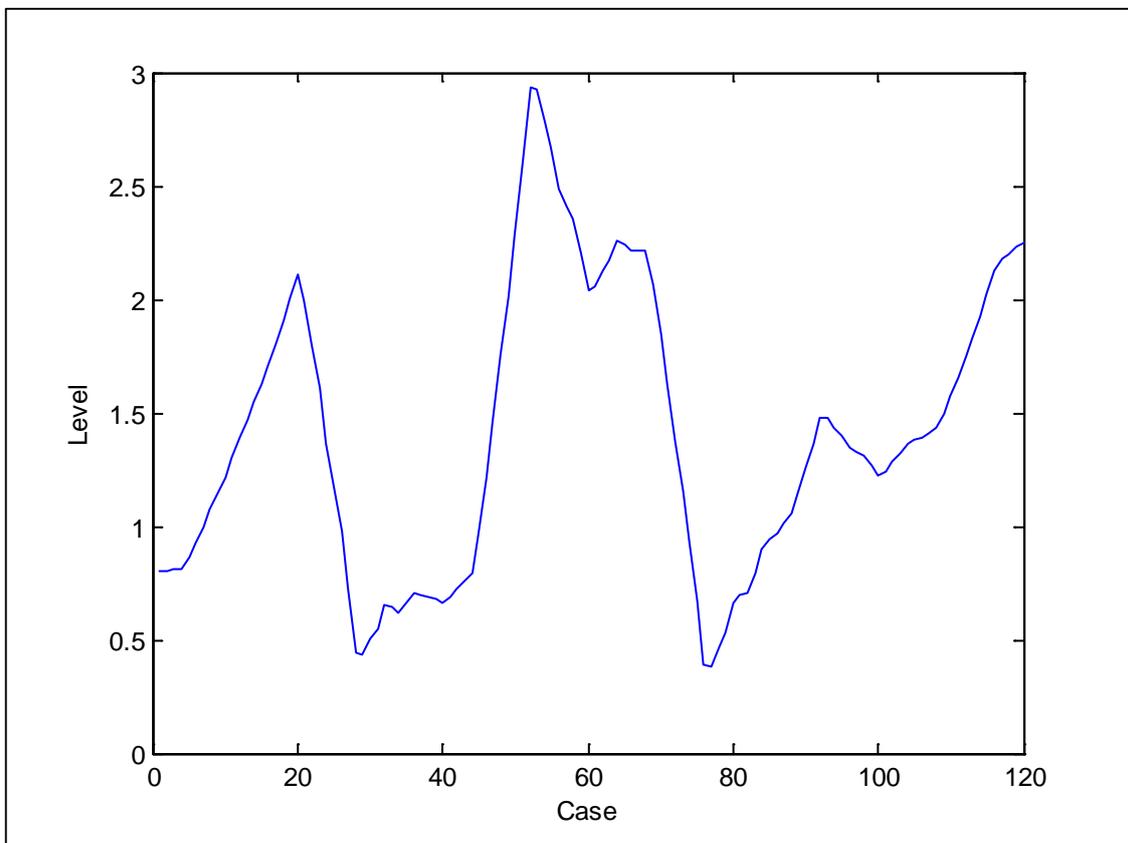


Figure 8.6 Compressed Signal Envelope Harmonic 4.

Analysis of the summary statistics for the 6th harmonic and its compressed signal show the mean values to be almost identical, 2.2023 and 2.2047 respectively. With a correlation coefficient of 0.8493 between them hence a goodness of fit statistic $R^2=72\%$. A high proportion of information was removed which was clearly due to the local fluctuations. The key question is how relevant are these local fluctuations to the explanation of variation in fault amplitudes. A one-way ANOVA is significant at 98.5% as shown in Table 8.2, confirming the null hypothesis of equal group means for the original 6th harmonic and its compressed signal.

Distribution wise again the signals are not alike with the compressed signal of harmonic 6 showing near perfect normal distribution whilst the original signal is positively skewed and has a far upper outlier, Figure 8.7.

Thus signal compression does not appear to have a uniform effect on the signal distribution, Figure 8.8.

However, a scatter plot of the cases by class as shown in Figure 8.9, using the compressed signals shows a marked improvement from the original model with most classes clearly depicted in near linear tracks.

Outstandingly a NB classification model for all classes, using compressed 4th and 6th harmonics alone, realised an 83.3% classification success which far exceeds the prior rates. Previously comparable rates requiring 10 plus input parameters. Refer to Table 6.5, Section 6.4 for a summary of NB models using uncompressed envelope harmonics. Recall also the 2 parameter NB using the original signal harmonics 4 and 6 resulted in just 60% successful classification.

The 3-D classification plot is illustrated in Figure 8.10. The classification matrix, below, gives specific numerical details of group allocations. The healthy group, row 1, having the greater number of cases allocated to other groups with 3 to the DVL and 5 to the LB groups. All DVL cases were correctly allocated. False positive allocations of healthy to fault states are inconvenient although less critical than false negative classifications.

H	16	3	0	5	0
DVL	0	24	0	0	0
ICL	0	1	22	0	1
LB	3	2	0	17	2
SVL	1	1	0	1	21

False negative errors, or type I errors, give a measure of the significance level, α . Sensitivity of a test being defined as $(1 - \alpha)$ i.e. the proportion of faults correctly identified as faults. Specificity or power of a test being the proportion of healthy cases correctly identified as healthy. Sensitivity and specificity are complementary measures intrinsic to the test not dependent on fault prevalence. A balance between the two is sought to maximise information gain. Affording equal weight to the true positive and false positives rates optimises test information and leads to a convenient measure of worth, the information gained. Table 8.3 displays the test summary information from which the information gain is calculated to be 0.484, by definition an informed decision, Equation 8.1.

Note equivalence to zero implies chance-level performance and <0 perverse use of information.

$$\begin{aligned} \text{Information gain} &= \text{specificity} + \text{sensitivity} - 1 \\ &= 0.913 + 0.571 - 1 = 0.484 > 0 \end{aligned} \tag{8.1}$$

TABLE 8.3 CALCULATION OF TEST SENSITIVITY AND SPECIFICITY

		Predicted Condition	
		Healthy	Faulty
True Condition	Healthy	TP = 16	FP = 8 Type 11 Error (β)
	Faulty	FN = 12 Type 1 Error (α)	TN = 84
		Sensitivity (power) $1 - \beta = \frac{16}{28} = 0.571$	Specificity $1 - \alpha = \frac{84}{92} = 0.913$

Successful classification of the DVL class was indeed expected, on inspection of the Fourier profiles of faults Figure 8.11, its profile being entirely separate in the first phase. The Andrews plot, Figure 8.8, based on the compressed 4th and 6th signals alone in fact shows great promise with respect to separable classes with a clearly defined sinusoidal pattern. Over the range $0 < t < 0.5$ the DVL is distinct with the LB and Healthy signals very well grouped. The second phase, $0.5 < t < 1.0$ shows clear ICL

separation with promising indication for SVL which is only slightly overlapped with the LB.

Compression would appear to have trimmed output signals reducing profile overlap, a feature further illuminated in the scatter plot Figure 8.9, thus isolating salient class characteristic trends.

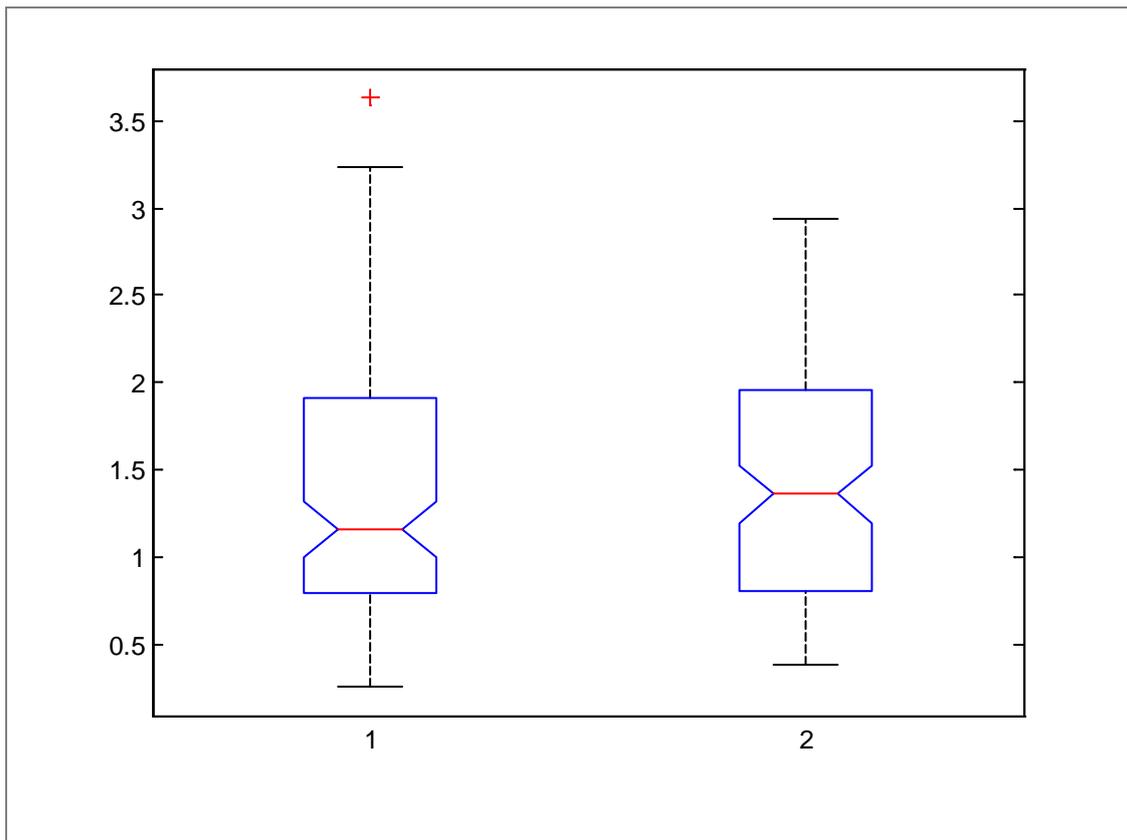


Figure 8.7 Distributional Comparison of 1. 6th Harmonic Amplitudes and 2. Compressed 6th Harmonic Amplitudes.

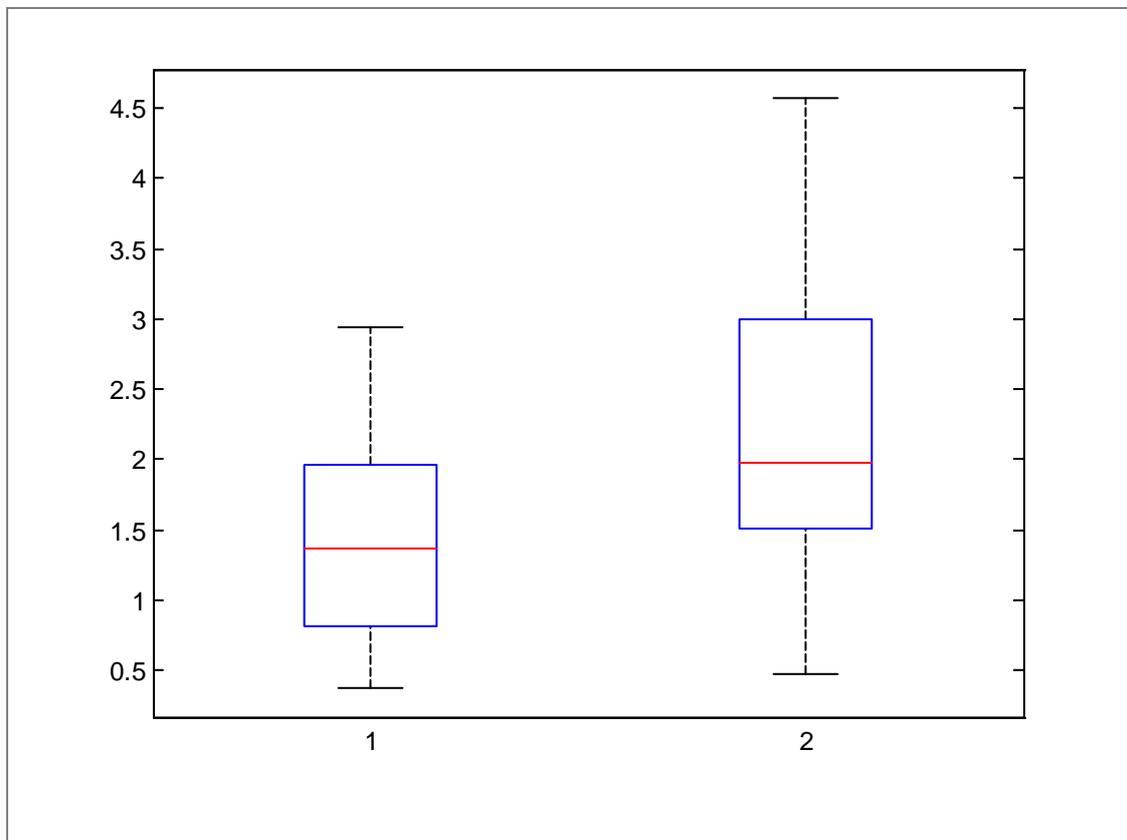


Figure 8.8 Comparison of Distributions of Compressed Signals 1. 4th Harmonic 2. 6th harmonic.

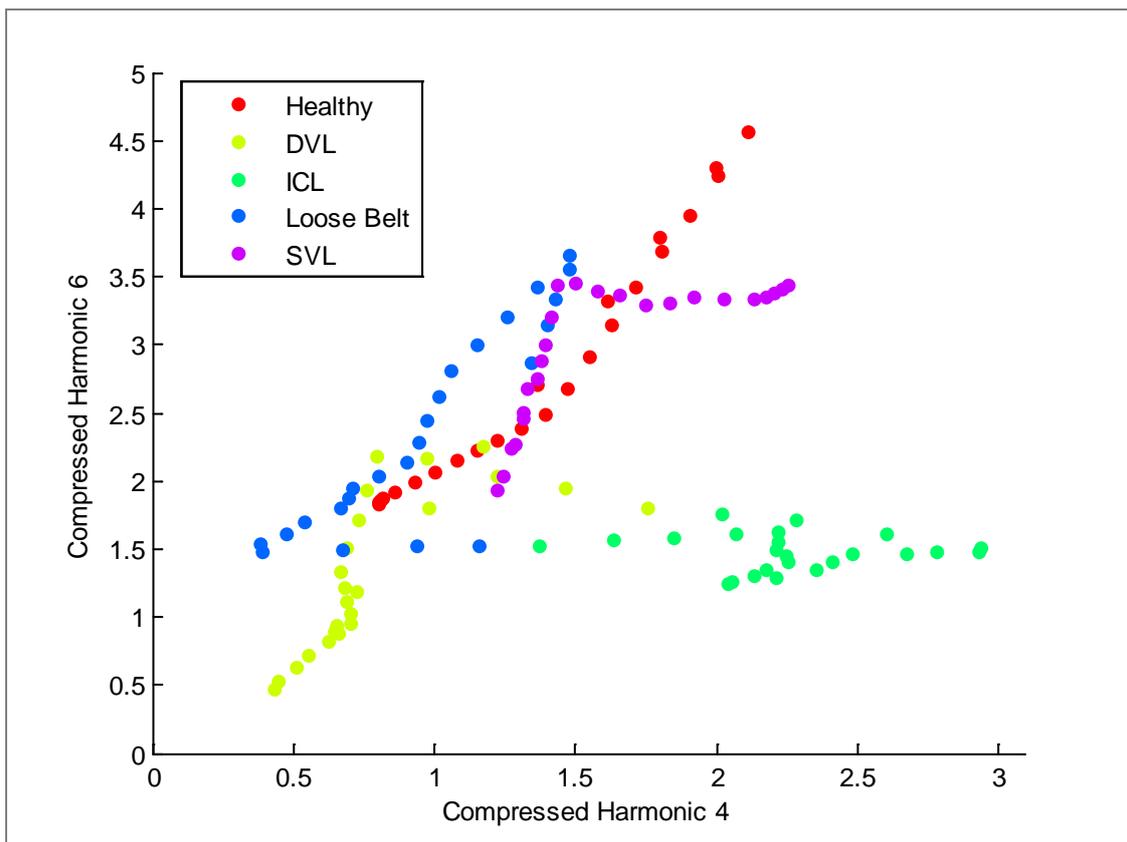


Figure 8.9 Scatter Plot by Class Using Compressed Signals.

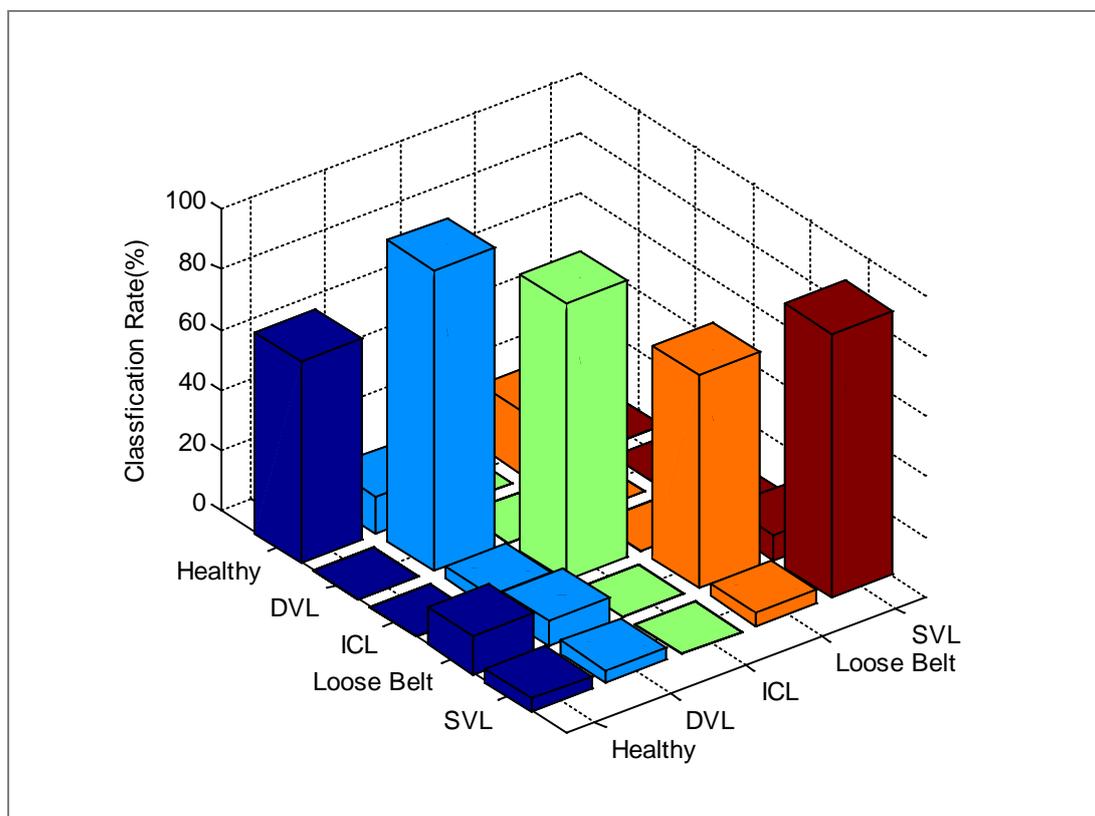


Figure 8.10 NB Classification for all 5 Classes using Two Compressed Harmonics (4 and 6).

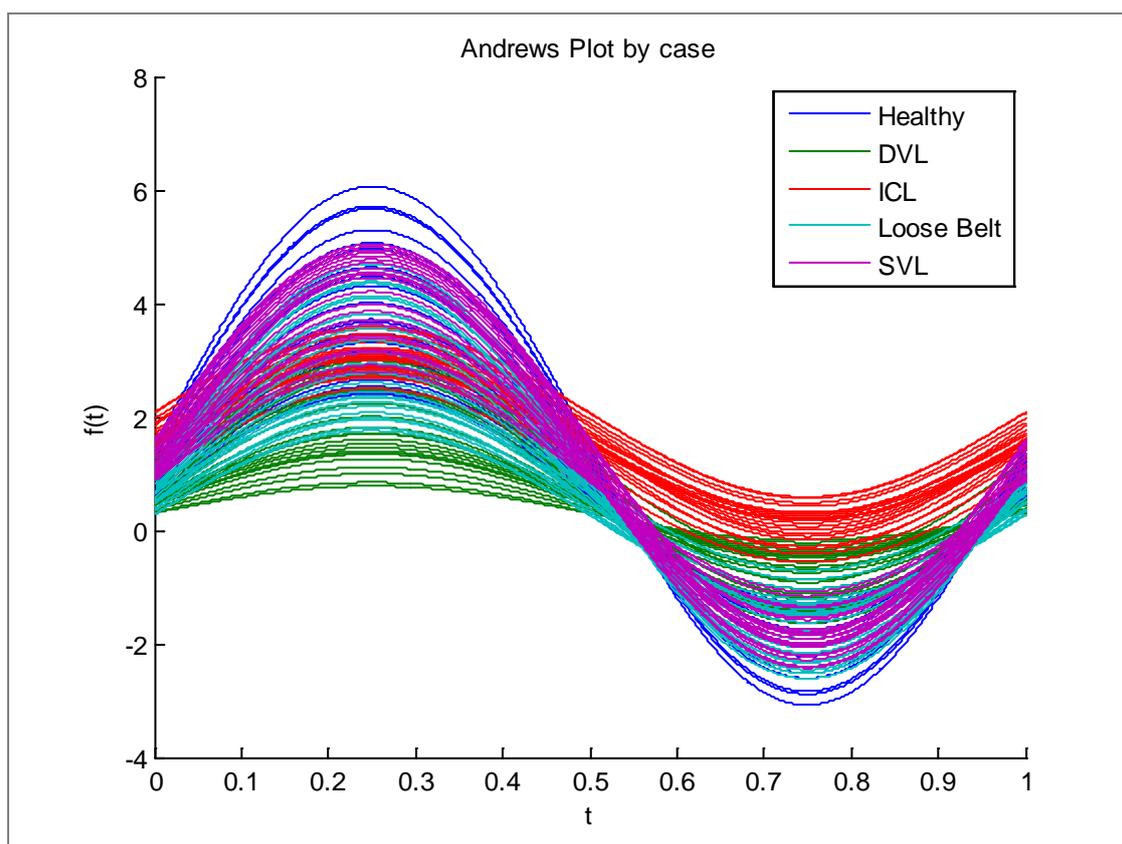


Figure 8.11 Andrews Plot using Compressed 4th and 6th Harmonics Only.

A natural progression would be to extend the single signal compression to a universal compression of all the stored harmonics prior to analysis of input parameters i.e. a multi-signal compression and wavelet analysis.

Analysis was extended by production of a three dimensional surface representation of the two dimensional Andrews plot. Thus a more detailed visual inspection of behaviour within and between faults was possible, Figure 8.12. Clearly the signals are locally irregular and noisy but nevertheless five different general shapes can be distinguished relating to the class profiles.

Surface representation of the original data using both de-noising and compression, Figure 8.13, retains the five class groups distinction thus highlighting the periodicity of the multi signal, however, the reconstruction is not as complete as desired and the residuals retain much of the original variations. Greater sophistication is required in the wavelet decomposition to gain further insight. Another area which shows promise for further investigations.

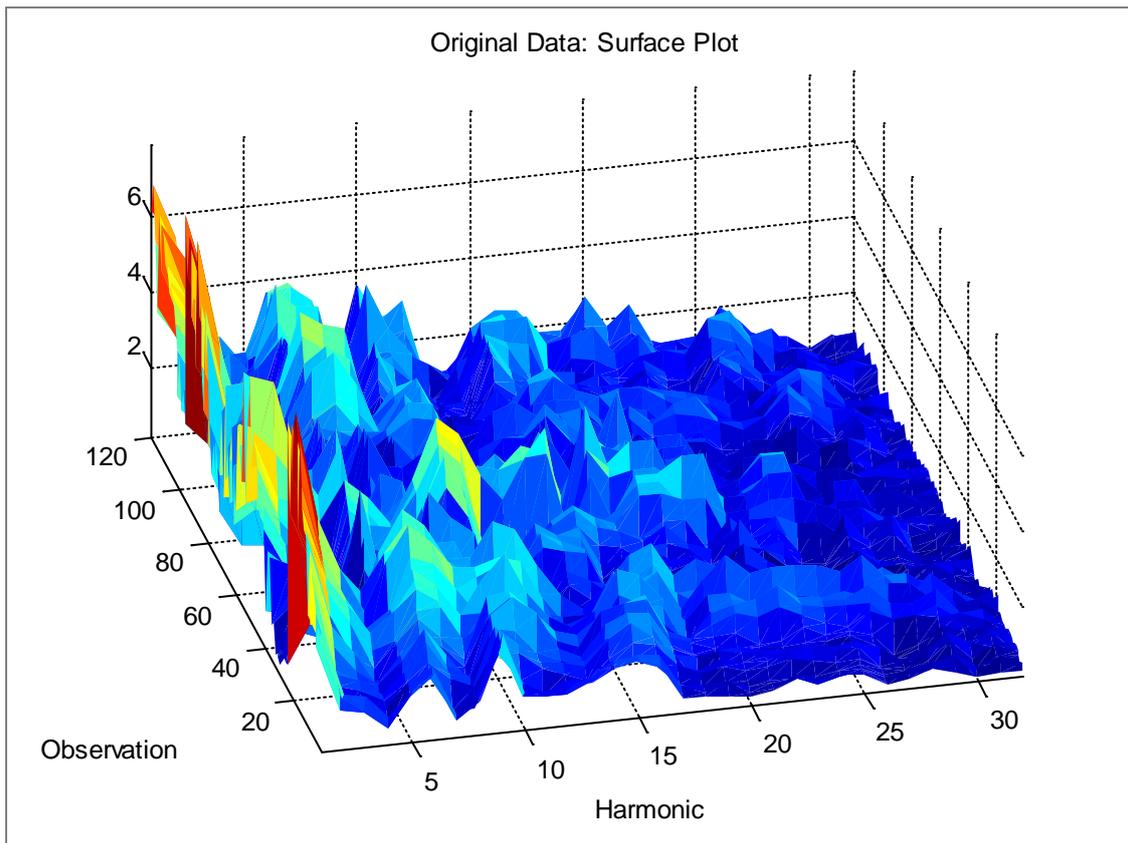


Figure 8.12 Surface Plot of Original Envelope Harmonics.

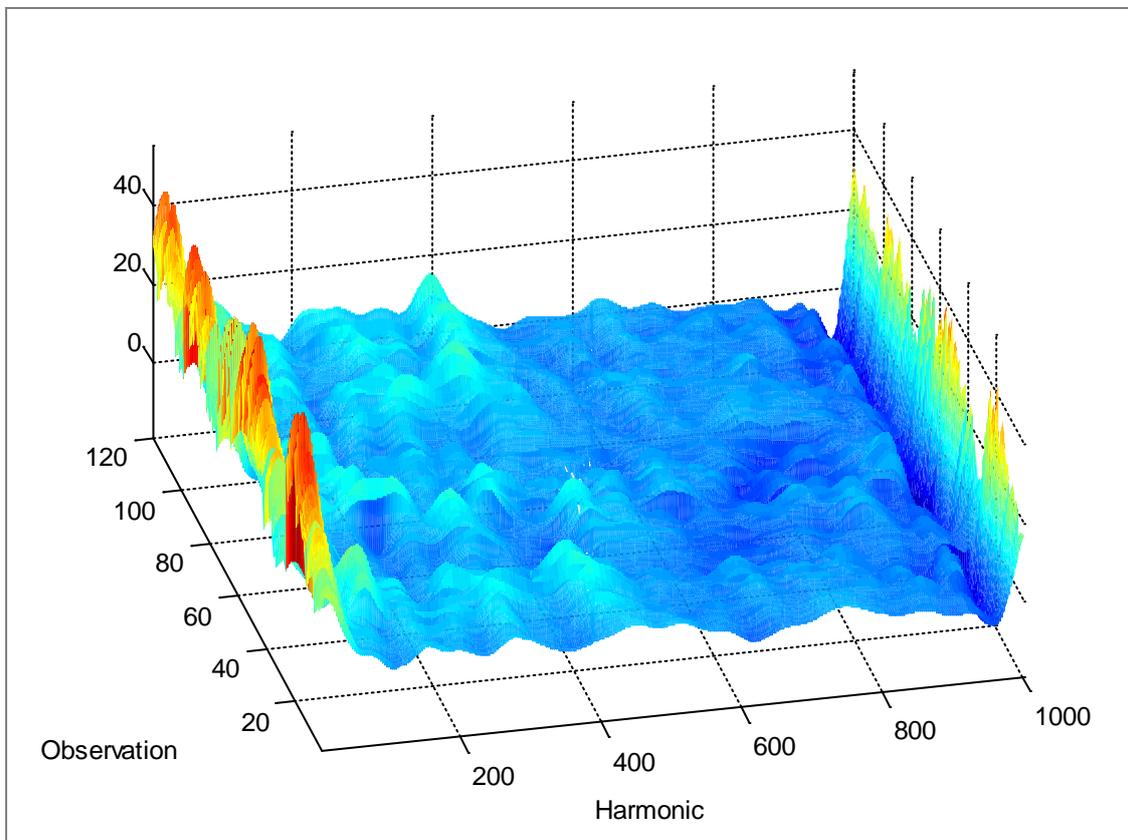


Figure 8.13 Wavelet Decomposition of Fault Profiles.

8.4 Summary

With both variable reduction and pre-selection of input parameters being beneficial in classifier construction a combined analysis was anticipated to give still greater classification success. Indeed improved group clustering was evident in both the ten and fifteen input parameter PCA models along with higher proportions of the system variance being explained by fewer PCs.

Compression of input parameters was shown to preserve mean values whilst smoothing localised fluctuations. Classifiers constructed using the compressed signals gave much improved successful classification rates using far fewer input parameters. It would thus seem the information loss on compression was not pertinent to fault detection and identification. The compressed two parameter NB model having a classification rate of 83% across all five classes compared to 60% for the non-compressed model. Visually displayed through Fourier profiling and scatter plots which would appear to demonstrate unequivocally the signals are trimmed to the benefit of fault determination.

Chapter 9

Review of Thesis Achievements

This chapter gives a summary of the work undertaken throughout the PhD and its worth. Contributions to knowledge are detailed and evidenced as are the research novelties. Thesis achievements are reviewed and critically appraised against the initial objectives. Achievements are consolidated and form the basis for suggested extended analysis.

9.1 Review of Thesis Objectives

9.1.1 Time Domain

Measurements captured via sensors attached at strategic points of the process were converted to MATLAB read inputs, translated into physical quantities and subsequently analysed. Each signal output was shown to demonstrate the same cyclic pattern of approximately seven cycles per second. Points on cyclic traces relating directly to stage of compression process.

Signal synchronisation was readily identified whilst the compressor rig operated in a healthy state through relating physical characteristics of each signal relative to the revolution index pulse. Visual displays of signal outputs describing all aspects of the physical process.

Trends and interrelationships between measured variables during normal running of the system and with faulty components introduced were analysed. Subsequently condition specific patterns were identified for model building purposes. Thus enabling identification of fault characteristic patterns through signal inspection during process operation.

The ultimate aim being to characterise and so enable identification of fault onset in a timely fashion. Signal analysis in the time domain across all fault classes showed plausible separation potential.

Whilst the second stage pressure signal readily highlighted ICL characteristics it was less useful in simultaneously determining other compressor rig defects.

Underlying distribution patterns for each fault were compared with respect to symmetry and skew via multiple boxplots and once again exposed the ICL as significantly different to all other conditions. All distributions were skewed to varying degrees and plots of amplitude arithmetic mean against kurtosis provided a crude classification rule utilising the first stage pressure measurements. Extreme deviations from the norm are most readily recognised, however, detection and accurate diagnosis of performance flaws at early onset reaps far greater benefit. Hence further scrutiny was necessary.

Specific details of the experimentation and analysis in the time domain are reported in section 4.6 'Physical Attributes of Output Signal Data and Measurements'; 4.7 'Waveform Features' and 5.1 'Fault Classification Using Statistical Modelling in the Time Domain'.

9.1.2 Frequency Domain

FT spectra for all captured signals operating under each machine state, healthy and the four seeded faults, were computed. Subsequently the FFT, FT spectrum amplitudes, were calculated and their profiles analysed. Demodulated signals were considered to 32 harmonics and demonstrated to display individual characteristics.

Spectrum plots, regardless of process health or variable inspected, clearly indicated the fundamental frequency to occur at 7.3 Hz. A direct consequence of the shaft rotation speed per second. Also subsequent spectral peaks occurred at integer multiples of this fundamental frequency. Defining characteristics due to type of fault present were indicated by the harmonic amplitudes. Thus harmonic amplitudes offer the key to machine condition and so enable a fingerprinting database to be established.

Envelope spectrum plots of the second stage vibration signal for all five classes simultaneously clearly display unique condition characteristics. Hence credible rules for discriminating between healthy and faulty systems can be established and individual faults identified. Envelope spectra harmonics were demonstrated to possess superior deterministic power over their time domain equivalents hence were favoured for subsequent analysis.

Summary of analysis and findings from section 5.2 'Signal Analysis in the Frequency Domain'.

9.1.3 Variable Selection

Through Fourier profiling using Andrews plots, differences in fault characteristics are highlighted hence suitability for class separation is clearly illustrated.

Whilst the pressure signal envelope harmonics form particularly uniform homogeneous groupings their use is not pursued for modelling purposes due to the invasive nature of signal collection. Hence, due to ease of continual monitoring and uninterrupted signal collection along with its outstanding ability to describe the process condition, the second stage vibration output signal was utilised for classifier construction.

Agglomerative CA and the subsequent visual illustration via dendrograms illuminated variable proximities and highlighted group formation. Both within and between group

proximities are readily identifiable hence homogeneous groupings ascertained with ease.

Once clustered heterogeneous variables were selected to represent each group and demonstrated, through simple scatter plots, to allocate cases to distinct groupings in two and three dimensional space. Thus highlighting the potential discriminatory powers of reduced sets of input parameters.

Full details of the variable selection process are contained in sections 6.1 'Identification of Separable Classes' and 6.2 'Variable Clustering'.

9.1.4 Cluster Construction

It should be noted that Euclidean distance is variable dependent hence the difference in dendrogram threshold scales. Further the same harmonics are not grouped together for each of the variables considered despite their having been generated simultaneously. However, for whichever signal considered, variable clustering was employed, using Euclidean distance as a proximity measure, to sift harmonics according to spectrum amplitudes. Resultant groupings of like harmonics were then readily displayed in dendrograms with colour threshold labelling added for easy cluster associations.

Using harmonic features identified through CA multivariate classifiers were established to discriminate amongst the classes. Two class models being highly successful and giving 100% classification accuracy using just two input parameters. Discriminatory power decreased as the number of classes increased necessitating greater numbers of input parameters to better describe variation among cases.

Similar trends were disclosed utilising Naïve Bayes although NB was better able to accommodate the increased complexity of simultaneously differentiating between up to five different classes.

Utilising all original variables a reduced weighted parameter set was constructed using PCA. The variation in the entire system was hence explored incorporating just a subset of the new weighted variables.

PCA models utilising the first two and first three PCs gave classification success results which were directly comparable to the NB models across all five classes.

Likewise SVM gave high classification success rates in the two group cases. Not surprisingly, incorporating input parameters previously found to possess high explanatory powers resulted in greater successful allocations with fewer SVs being required for specification.

Confirmatory FA added further evidence to the belief that not all input parameters are equally capable of discerning fault characteristics. Highly successful input parameters displaying very high factor loadings on factors one or two with minimal loadings on the other. Highly correlated, homogeneous, harmonics having high loadings on the same factor with optimal harmonic pairings, heterogeneous harmonics, having high loadings on opposite factors.

Further details and graphical outputs in sections 7.1 'Principal Component Analysis of Envelope Harmonics', 7.2 'Confirmatory Factor Analysis' and 7.3 'Support Vector Machine Classification'.

9.1.5 Extended Analysis

Data compression was incorporated as a means to further reduce input parameter volume hence lessen computational requirements.

Compressing envelope harmonics 4 and 6 appeared to retain information relevant to classification success. The localised irregular fluctuations being eliminated the input parameter volume was considerably diminished yet marked improvements were obvious with clear case grouping when viewed as scatter plots. Indeed the two parameter NB using compressed harmonics 4 and 6 achieved a classification success rate of 83.3% greatly improving on the prior rate of 60%.

A PCA using a reduced set of input parameters gave some visible improvements in classification success. Whilst the first eigenvalue wasn't nearly so dominant, hence the first PC didn't account for as large a proportion of the variance, overall the variance accountability was increased. Additionally only 6 PCs were required to explain over 95% of the process variability compared to 10 PCs in the non-restricted model. Additional modelling added weight to the benefits of pre selecting input parameters both with and without data reduction techniques.

Chapter 8 contains an expansion of these findings and further preliminary investigations considering total compression for all input variables alongside potential developments with MSPCA.

9.2 Contributions to Knowledge

- 9.2.1 Development of a methodology to identify feasibility of class separation a priori through application of Fourier profiles in the form of Andrews plots. Separable classes being visibly identifiable. Reciprocating compressor fault Fourier profiles exposed the likelihood of distinguishing between different fault types. Evidenced in section 6.1 where fault profiles were compiled using the envelope spectra from second stage vibration, pressure and motor current spectra.
- 9.2.2 Potential input parameters were demonstrated to form homogeneous cluster groupings, within group harmonics seen to demonstrate comparable properties. Assimilation of envelope spectra harmonic features into homogeneous cluster groups enabled selection of heterogeneous representatives as a sparse input parameter set. Restricting models to fewer highly informative variables enables algorithmic execution and convergence within manageable computational expenditures. Reducing input parameter volume benefits algorithmic convergence. Model efficiency was maintained as only duplicitous variables are eliminated. Reduced complexity of input parameter set results in reduced computational efforts keeping both time and costs at a minimum. Highly laudable classifiers were fabricated. Refer to section 6.2 for details of application to RC envelope harmonics.
- 9.2.3 Confirmatory factor analysis was utilised to reveal underlying harmonic characteristics suggested through variable cluster analysis. A pair of heterogeneous parameters each with high specificity established an extremely effective SVM classification rule for RC faults. Confirmatory factor analysis also revealed high correlations between harmonic pairs through like extreme loadings on factors 1 or 2. Successful heterogeneous pairings were demonstrated to be uncorrelated each with a high specific score on opposing factors. SVM were seen to give much increased classification success rates for the two input parameter case when heterogeneous harmonics with high specificity were utilised. Further details in Chapter 7 section 7.1.1.
- 9.2.4 Compression of signal data resulted in further input parameter volume reduction with vastly improved classification success rates. Repeated modelling using compressed signals proffered greatly improved precision in all cases. Accuracy in classification of RC faults improving dramatically. In

particular a two parameter NB attained 83% successful classifications across five classes compared to the 60% success for the non-compressed model. Further information in section 8.2.

9.3 Conclusions

The purpose of this thesis was to compare and contrast statistical methods for optimal selection of model input parameters. Comparing and contrasting findings for each subsequent model with respect to classification accuracy.

Over the past fifty years as computational power and processing capabilities have vastly increased the exponential growth in data collection has been accommodated with relative ease. However, as demand for ever greater precision and more immediate feedback continues to grow alongside increasing structural complexity the burden on computational time is telling. A new approach to data validation wherein data is pre-sorted and filtered to eliminate superfluous elements has been demonstrated and shown to be effective.

9.3.1 Initial Exploratory Analysis Findings

Each measured signal taken directly from the compressor rig displayed the same cyclic output regardless of point of collection or variable type. Thus first and second stage pressures and vibration measurements had a common seven cycle per second profile. Inter-relationships between these outputs were seen to be clearly defined and directly attributable to collection point in the compression cycle. Distinctive pattern changes were observable in the presence of faults. Moreover, groups of pattern changes were fault specific thus fault blueprints could be established hence the suitability of this method for CM purposes.

Time domain analysis was found to give rudimentary explanation of pattern differences albeit generally isolating a smaller number of faults. However, scrutiny failed to form clear rules of separation for all faults simultaneously.

Consideration of signals in the frequency domain highlighted the unique signal characteristics with the frequency spectra clearly showing the fundamental frequency of the rig and higher harmonics at its integer multiples. However, whilst the positioning of the fundamental frequency was the same for all classes, position and magnitude of higher order amplitudes varied. This formed the basis of reliably distinguishing between fault types. Being non-intrusively collected and robust in mechanical applications, if external interference be controlled, the second stage vibration signals were deemed the most informative source for continual monitoring purposes.

Moreover, harmonics extracted from the envelope spectra of vibration signals in the frequency domain were seen to have still more superior deterministic properties over their time domain equivalents. Amplitudes of the harmonics being specific to process condition with an altered amplitude size or displaced amplitude implying presence of a fault. Envelope spectra showing only the amplitude profile of original signals provided a clearer insight into the underlying behaviour of a process having extraneous noise removed.

Envelope spectra taken from the second stage vibration signals revealed the potential for detecting and identifying faults. High classification success rates were achieved using RVM to identify both single and compound faults. Algorithmic capabilities with respect to numbers of input parameters being the limiting factor. Model convergence was achievable for reduced numbers of parameters only.

Subsequent investigations revealed strong associations between specific fault characteristics and particular envelope harmonics. Incorporating GA to aid input selection was seen to further improve classification success rates. However, the underlying input parameter structure was then unknown as physical characteristics were not preserved. Hence this preliminary study fuelled the motivation for further exploratory investigations into both the individual merits of each envelope spectra harmonic and parameter selection criteria.

Most algorithms have the capability to process only a limited number of input variables. Computational time saving is of course a priority and established variable reduction methodologies offer input parameter simplifications. However, pre selection of input variables is not yet recognised practice. Reduction in complexity, if any, being achieved by within model manipulation of original input variables without clear data properties being identified or preserved. Greater explanatory power is achievable whilst keeping input volume to a minimum if prior inspection of available parameter properties informs pre modelling variable selection.

9.3.2 Class Profiling and Variable Selection

Group profiling through the use of Andrews Fourier plots indicated the potential for class separation thus gave a measure of the likely classification success of any subsequently developed classifier. Class Fourier profiles can be examined for all suitable output variables to assess potential for separation. Hence realistic expectations and tolerance setting is feasible.

Investigation of variable interdependencies through data clustering techniques provides a robust method for exploratory variable analysis. Deepening understanding of variable properties facilitates reduction in the number employed in establishing classifiers. A major advantage in condition monitoring being that optimum feature sets are reliably selected prior to incorporation in model algorithms. Reducing run time and ensuring convergence whilst maintaining explanatory power. Simultaneously algorithm sensitivity is maximised thus increasing opportunities for intervention and health prognosis.

Model efficiency was optimised through clustering potential input parameters with respect to similarities leading to enhanced understanding of variable characteristics so defining an optimal set of variables for compressor fault classification.

CA revealed underlying variable characteristics highlighting connections and dissimilarities. Homogeneous groups of variables were easily identifiable and a heterogeneous set selected for modelling purposes. Variable reduction techniques and machine learning methodologies such as RVM and SVM do not preserve original variable features hence the models cannot be directly related to the physical aspects of the application. Whilst data mining techniques support high computational efficiency the impact of controlled experimental measures is unattainable and not validated in the field.

Classifiers constructed using a simplified input parameter set pertinent to point of interest was demonstrated to benefit both accuracy and computational savings. Highly successful classifiers being developed using both discriminant analysis and naïve Bayes methods. A method generalisable to all mechanical process monitoring applications.

For the two group case, both DA and NB models achieved 100% classification success with two and five input parameters respectively. Across all five fault classes a ten parameter NB model was seen to be highly efficient demonstrating the efficacy of input parameter selection from heterogeneous groups.

Variable reduction techniques provided an alternative approach wherein all the original variables were reconstructed as a smaller number of new variables. Again computational savings were made due to reduced input volume and resulting classification success rates were high although the original variable characteristics are not preserved.

9.3.3 Classifier Construction

Classification success was shown to vary extensively depending on the number of fault groups considered and the number of parameters incorporated in the model.

Again model complexity hence computational efforts were significantly reduced by prior evaluation and selection of a reduced number of heterogeneous input parameters to ensure maximum explanatory power across all classes.

Naïve Bayes was shown to lend itself to classification of increased numbers of groups and input parameters.

Reductions in successful classification rates on inclusion of all 32 envelope harmonics compared to restricted input parameter sets determined by CA is indicative of the necessity for combined techniques.

9.3.4 Extended Analysis

Compression of input parameters was demonstrated to preserve mean values whilst smoothing localised fluctuations. Classifiers constructed using the compressed signals gave much improved successful classification rates using far fewer input parameters. Any information loss not being pertinent to fault detection and identification. The compressed two parameter NB model having a classification rate of 83% across all five classes compared to 60% for the non-compressed model. However, greater precision and process control is achievable if compression thresholds be theoretically determined. Han et al. (2016) propose an automated empirical formula based on auto-regressive moving average models with 'Swing Door Trending' which they claim provides a dynamic self-regulatory feature. Certainly pre-defining expectations through threshold specification promises greater controls over potential information loss.

9.4 Novelties

- 9.4.1 Application of Fourier profiles in the form of Andrews plots to determine a priori feasibility of class separation applied to reciprocating compressor faults. Plot of Fourier profiles of output signal envelope spectra against time revealed class separation potential. Should Fourier based fault profiles not display separation at least in part then attempts to distinguish between the classes using FFT harmonics is futile [Section 6.1].
- 9.4.2 Focus on pre-evaluation of potential input parameters by clustering variables offers a novel approach to reducing algorithmic volumes. Excess bulk eliminated pre computation hence duplicity in variability accounted for per input parameter avoided. Organisation of envelope harmonics into homogeneous groupings both informs model construction and offers greater understanding of individual parameter associations and complexities. Key to maximising explanatory power. Sifting potential input variables prior to incorporating in computational algorithms affords real advantages in controlled classifier construction [Section 6.2].
- 9.4.3 Heterogeneous input parameters which complement one another were pre-selected with the priority of optimising explanatory power of classifiers. This informed choice ensuring complete coverage of all eventualities studied whilst reducing computational requirements. Properties and characteristics of input parameters were preserved unlike standard executions of SVM and RVM. Thus meaningful interpretation of outcomes is facilitated. RC faults can be identified and associated with original input parameters. Visual output in form of the dendrogram gives a clear visual mapping showing proximities and interconnections between variables [Section 6.2]. Surges in computational capabilities have facilitated massive increases in data processing potential although ever increasing data volumes foil progress. Research in this thesis reveals the underlying characteristics of the input variables and offers a viable solution to the big data problem [Chapters 6, 7 and 8].
- 9.4.4 Harmonic characteristics were further explored and validated through confirmatory factor analysis. Resultant identification of high energy input parameter pairings a major boost to modelling efficiency and reduced model complexity. Individual parameter worth assessed by specificity thus identifying unique attributes of individual variables [Section 7.1.1].

- 9.4.5 Compressed signal harmonics offered still greater input parameter volume reduction. Repeated modelling using compressed signals proffered greatly improved precision in all cases. Whilst local fluctuations were removed on compression signal means were preserved [Section 8.2].
- 9.4.6 Generalisation to wide range of on-line monitoring systems. Classical statistical and machine learning techniques are merged to mutual benefit with input variables pre-selected, compressed and input into both multivariate statistical and wavelet models with increased fault classification capabilities [Chapter 8].

References

- [1]. Ahmed, M., Gu, F. and Ball, A. (2012). Fault Detection of Reciprocating Compressors using a Model from Principles Component Analysis of Vibrations. *Journal of Physics: Conference Series*, 364, 012133. ISSN 1742-6596.
- [2]. Ahmed, M., Smith A., Gu F. and Ball A.D. (2014). Fault Diagnosis of Reciprocating Compressors Using Relevance Vector Machines with A Genetic Algorithm Based on Vibration Data. *Proceedings of the 20th International Conference on Automation & Computing*.
- [3]. Akinduko, A.A. and Groban, A.N. (2013). Multiscale Principal Component Analysis. In *Journal of Physics: Conference Series Vol. 490 (1)*. 012081. IOP Publishing.
- [4]. Amirat, Y., Benbouzid M. E. H., Al-Ahmar E., Bensaker, B. and Turri S. (2009). A Brief Status on Condition Monitoring and Fault Diagnosis in Wind Energy Conversion Systems. *Renewable and Sustainable Energy Reviews*, vol. 13, 2629-2636, 12.
- [5]. Andrews, D. F. (1972). Plots of High Dimensional Data. *Biometrics*, 28:125–136.
- [6]. Baron, R. (1996). *Engineering Condition Monitoring Practice, Methods and Applications*. Addison Wesley Longman. ISBN 0-582-24656-3.
- [7]. Bakshi, B. R. (1998). Multiscale PCA with Application to Multivariate Statistical Process Monitoring. *AIChE Journal*, 44(7), 1596-1610.
- [8]. Bensasi, A., Gu, F. and Ball, A D. (2004). Instantaneous Angular Speed Monitoring of Electric Motors. *Journal of Quality in Maintenance Engineering (JQME)*, Vol. 10 (2), 1355-2511.
- [9]. Bernal-de-Lázaro, J. M., Llanes-Santiago, O., Prieto-Moreno, A., Knupp, D. C., and Silva-Neto, A. J. (2016). Enhanced Dynamic Approach to Improve the Detection of Small-Magnitude Faults. *Chemical Engineering Science*, 146, 166-179.
- [10]. Bhattacharya, A. and Pranab K. D. (2014). Recent Trend in Condition Monitoring for Equipment Fault Diagnosis. *International Journal of System Assurance Engineering and Management* 5 (3): 230-44.
- [11]. Bloch, H. P. and Hoefner, J. J. (1996). *Reciprocating Compressors: Operation and Maintenance*. Houston, Texas: Gulf Pub. Co.

- [12]. Bond, R.R. (2010). *Vibration Based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. Wiley. ISBN 9780470747858.
- [13]. Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1993). *Classification and Regression Trees*. Chapman and Hall, Boca Raton.
- [14]. Bro, R. and Age K. S. (2014). Principal Component Analysis. *Analytical Methods* 6.9: 2812-2831.
- [15]. Campbell, G. (2009). *Guide to Writing an M.Eng. / MSc Dissertation*. School of Chemical Engineering and Analytical Science.
- [16]. Carden, E.P. and Fanning, P. (2004). Vibration Based Condition Monitoring: A Review. *Structural Health Monitoring*, 3(4), 355-377.
- [17]. C´esar Garc´ia-Osorio, C. F. (2005). Visualization of High-Dimensional Data via Orthogonal Curves. *Journal of Universal Computer Science*, vol. 11 (11), 1806-1819.
- [18]. Chapman, S.J. (2009). *Essentials of MATLAB Programming*. 2nd edition. Centage Learning. ISBN 13:978-0-495-29570-9.
- [19]. Chapman, S.J. (2013). *MATLAB Programming with Applications for Engineers*. Centage Learning International Edition. ISBN 13:978-0-495-66808-4.
- [20]. Chatfield, C. and Collins, A.J. (1980). *Introduction to Multivariate Analysis*. Chapman and Hall. ISBN 0-412-16030-7.
- [21]. Chiang, L.H., Russell, E.L. and Braatz, R.D. (2000). Fault Diagnosis in Chemical Processes Using Fisher Discriminant Analysis, Discriminant Partial Least Squares, and Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, Vol. 50 (2), 243-252, ISSN 0169-7439.
- [22]. Chiang, L. H., and Braatz, R. D. (2003). Process Monitoring Using Causal Map and Multivariate Statistics: Fault Detection and Identification. *Chemometrics and Intelligent Laboratory Systems*, 65(2), 159-178.
- [23]. Chiang, L. H., Kotanchek, M. E., and Kordon, A. K. (2004). Fault Diagnosis Based on Fisher Discriminant Analysis and Support Vector Machines. *Computers and Chemical Engineering*. Vol. 28(8), 1389-1401.
- [24]. Chiang, L. H., Jiang, B., Zhu, X., Huang, D., and Braatz, R. D., (2015). Diagnosis of Multiple and Unknown Faults Using the Causal Map and Multivariate Statistics. *Journal of Process Control*, Vol. 28, 27-39.
- [25]. Cohen, A. (1995). *Wavelets and Multiscale Signal Processing*. Chapman and Hall.
- [26]. Cristianini, N., Shawe-Taylor, J. and Books, I. (2000). *An introduction to Support Vector Machines and other Kernel-based Learning Methods (Corrected edition.)*. Cambridge: Cambridge University Press.

- [27]. Davies, A. (1998). Handbook of Condition Monitoring: Techniques and Methodology. Chapman and Hall. ISBN 9780412613203.
- [28]. De Botton, G., Ben-Ari, J. and Sher, E. (2000). Vibration Monitoring as a Predictive Maintenance Tool for Reciprocating Engineering. Proceedings of Mechanical Engineering 214 (D8), 895-903.
- [29]. Deming, W E. (1982). Out of the Crisis: Quality, Productivity and Competitive Position. ISBN 0-521-30553-5.
- [30]. Ding, S. X. (2008). Model-based Fault Diagnosis Techniques. Design Schemes, Algorithms and Tools. Springer.
- [31]. Dong, L., Li, X. and Xie, G. (2014). Nonlinear Methodologies for Identifying Seismic Event and Nuclear Explosion using Random Forest, Support Vector Machine and Naïve Bayes Classification. Abstract and Applied Analysis, 2014, 1-8.
- [32]. Elhaj, M., Gu, F., Shi. J. and Ball. A. (2001a). Early Detection of Leakage in Reciprocating Compressor Valves using Vibration and Acoustic Continuous Wavelet Features. Proceedings of the 14th International COMADEM Conference, Manchester University, 749 -756.
- [33]. Elhaj, M., Gu. F., Shi. J. and Ball. A. (2001b). A Comparison of the Condition Monitoring of Reciprocating Compressor Valves Using Vibration, Acoustic, Temperature and Pressure Measurements. Electronic Proceedings of the 6th Annual Maintenance and Reliability Conference (MARCON), Gatlinburg, Tennessee, USA.
- [34]. Elhaj, M., Gu, F. and Ball, A. (2001c). The Airborne Acoustic Condition Monitoring of Reciprocating Compressors. Electronic Proceedings of the 5th Annual Maintenance and Reliability Conference (MARCON), Gatlinburg, Tennessee, USA.
- [35]. Elhaj, M., Gu. F. and Ball, A. (2004). Numerical Simulation Study of a Two Stage Reciprocating Compressor for Condition Monitoring. Proceedings of the 17th International Congress on Condition Monitoring and Diagnostic Engineering Management (COMADEM), Cambridge, 602-611.
- [36]. Elangovan, M., Ramachandran, K. I. and Sugumaran, V. (2010). Studies on Bayes Classifier for Condition Monitoring of Single Point Carbide Tipped Tool Based on Statistical and Histogram Features. Expert Systems with Applications, 37(3), 2059-2065.
- [37]. Everitt, B. S. and Dunn, G. (2001). Applied Multivariate Data Analysis. 2nd Edition. Arnold, London, ISBN 0 340 74122 8.
- [38]. Fan, Z., Xu, Y., Ni, M., Fang, X. and Zhang, D. (2016). Individualized Learning for Improving Kernel Fisher Discriminant Analysis, Pattern Recognition, Vol. 58, 100-109. ISSN 0031-3203.

- [39]. Feng, G., Hu, N., Mones, Z., Gu F. and Ball, A. D. (2016). An Investigation of the Orthogonal Outputs from an On-Rotor MEMS Accelerometer for Reciprocating Compressor Condition Monitoring. *Mechanical Systems and Signal Processing*, vol. 76–77, 228-241.
- [40]. Fernandez-Temprano, M., Gardel-Sotomayor, P.E., Duque-Perez, O. and Morinigo-Sotelo, D. (2013). Broken Bar Condition Monitoring of an Induction Motor under Different Supplies Using a Linear Discriminant Analysis. 9th IEEE SDEMPED 162-168.
- [41]. Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7 (2), 179-188.
- [42]. Flury, B., Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach*. Chapman and Hall. ISBN 0-412-30030-3.
- [43]. Gajjar, S. and Palazoglu, A., (2016). A Data-Driven Multidimensional Visualization Technique for Process Fault Detection and Diagnosis. *Chemometrics and Intelligent Laboratory Systems*, vol. 154, 122-136.
- [44]. Gilat, A. (2008). *Matlab: An Introduction with Applications*. John Wiley and Sons. ISBN 978-0-470-10877-2.
- [45]. Gong, Y. (2009). Functional Calculi Generated by Clifford-Hermite Wavelets. *Computational Methods and Function Theory*, 9(2), 365-377.
- [46]. Gu, F. and Ball, A. (2002). Instantaneous Angular Speed Signature Extraction through Hilbert Transform and Fourier Transform. Technical Report: MERG- 0402, Manchester University.
- [47]. Gu, F., Ball, A.D. (1995). Use of the Smoothed Pseudo-Wigner-Ville Distribution in the Interpretation of Monitored Vibration Data Maintenance. vol. 10, 16-23.
- [48]. Han, S., Liu, X., Chen, J., Wu, J. and Ruan, X. (2016). A Real-time Data Compression Algorithm for Gear Fault Signals. *Measurement*, 88, 165-175.
- [49]. Hassan R., Cohanin B., De Weck O. and Vente G. (2005). A Comparison of Particle Swarm Optimization and the Genetic Algorithm. In *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*. 18-21.
- [50]. He, Q., Yan, R., Kong, F., and Du, R. (2009). Machine Condition Monitoring Using Principal Component Representations. *Mechanical Systems and Signal Processing*, 23(2), 446-466
- [51]. Hogg, R.V. and Craig, A.T. (1978). *Introduction to Mathematical Statistics*. 4th edition. Collier Macmillan. ISBN 0-02-978990-7.
- [52]. Hoel, P.G. (1971). *Introduction to Mathematical Statistics*. 4th edition. Wiley.
- [53]. Huang, P. and Gao, G. (2016). Parameterless Reconstructive Discriminant Analysis for Feature Extraction. *Neurocomputing*, vol. 190, 50-59.

- [54]. Hywel, J., P. R. Drake, and A. Davies. (1994). Condition-Based Maintenance and Machine Diagnostics.
- [55]. Isermann, R. (2001). Fault-Diagnosis Applications. Model-based condition Monitoring; Actuators, Drives, Machinery, Plants, Sensors, and Fault-Tolerant Systems. Springer. ISBN 978-3-642-12766-3 Springer Berlin Heidelberg New York.
- [56]. Isermann, R. (2006). Fault-Diagnosis Systems an Introduction from Fault Detection to Fault Tolerance. Springer. ISBN-10 3-540-24112-4 Springer Berlin Heidelberg New York.
- [57]. Johnson, A. and Wichern, D.W. (2002). Applied Multivariate Statistical Analysis. 5th edition. Prentice Hall. ISBN 0-13-121973-1.
- [58]. Kalpakjian, S., and Schmid, S. (2000). Manufacturing Engineering and Technology (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- [59]. Kankar, P. K., Sharma, S. C. and Harsha, S. P. (2011). Fault Diagnosis of Ball Bearings Using Continuous Wavelet Transform. Applied Soft Computing Journal, 11(2), 2300-2312.
- [60]. King, S., Bannister, P. R., Clifton, D. A., and Tarassenko, L. (2009). Probabilistic Approach to the Condition Monitoring of Aerospace Engines. IMechE Vol. 223 Part G: J. Aerospace Engineering.
- [61]. Kruger, U. and Xie, L. (2012). Statistics in practice. ISBN 97804 700 28193.
- [62]. Kumar, M.G., Hemanth, K., Gangadhar, N., Kumar, H. and Krishna, P. (2014). Fault Diagnosis of Welded Joints through Vibration Signals using Naïve Bayes Algorithm. Procedia Materials Science, 5, 1922-1928.
- [63]. Kusiak, A., Zhang, Z. and Verma, A. (2013). Prediction, Operations, and Condition Monitoring in Wind Energy. Energy, 60, 1-12.
- [64]. Laslett, O.W., Mills, A.R., Zaidan, M.A. and Harrison, R.F. (2014), Fusing an Ensemble of Diverse Prognostic Life Predictions. Aerospace Conference IEEE, 1-10.
- [65]. Lee, D. S., Park, J. M. and Vanrolleghem, P. A. (2005). Adaptive Multiscale Principal Component Analysis for On-line Monitoring of a Sequencing Batch Reactor. Journal of Biotechnology 116, 195-210.
- [66]. Lee, S. (2008). An Introduction to Mathematics for Engineers: Mechanics. Hodder Education. ISBN 978-0-340-965528.
- [67]. Lei, Y., Lin, J., Zuo, M. J., and He, Z., (2014). Condition Monitoring and Fault Diagnosis of Planetary Gearboxes: A review. Measurement, Vol. 48, 292-305.
- [68]. Li, J. C. Li, S.Y. (1995). Acoustic Emission Analysis for Bearing Condition Monitoring. Wear, Vol. 185 (1), 67-74, ISSN 0043-1648.

- [69]. Li, D., Wu, H. and Gao, J. (2013). Experimental Study on Stepless Capacity Regulation for Reciprocating Compressor Based on Novel Rotary Control Valve. *International Journal of Refrigeration*, Vol. 36, 1701-1715.
- [70]. Lin, Y. H., Lee, W. S. and Wu, C.Y. (2014). Automated Fault Classification of Reciprocating Compressors from Vibration Data: A Case Study on Optimisation using Genetic Algorithm. *Procedia Engineering*, Vol. 79, 355-361.
- [71]. Liu, H., Gegov, A., & Cocea, M. (2015). Rule based systems for big data. Springer International Publishing.
- [72]. Lüdtke, K. (2004). *Process Centrifugal Compressors: Basics, Function, Operation, Design, Application*. Berlin, London. Springer Science and Business Media.
- [73]. Manley, B. F. J. (1986). *Multivariate Statistical Methods: A Primer*. 3rd Edition. Chatfield and Collins ISBN 9781584884149.
- [74]. Misra, M., Yue, H., Qin, S. and Ling, C. (2002). Multivariate Process Monitoring and Fault Diagnosis by Multi-scale PCA. *Computers and Chemical Engineering* 26 1281-1293.
- [75]. Muralidharan, V. and Sugumaran, V. (2012). A Comparative Study of Naïve Bayes Classifier and Bayes Net Classifier for Fault Diagnosis of Monoblock Centrifugal Pump Using Wavelet Analysis. *Applied Soft Computing Journal*, 12(8), 2023-2029.
- [76]. Murray-Smith, D. J. (2012). *Modelling and Simulation of Integrated Systems in Engineering: Issues of Methodology, Quality, Testing and Application*. Cambridge: Woodhead. ISBN 085709078X, 9780857090782.
- [77]. Natarajan, B.K. (1995). Filtering Random Noise from Deterministic Signals via Data Compression. *IEEE Transactions on Signal Processing*, 43(11). 2595-2605.
- [78]. Oakland, J (2003). *Statistical Process Control*. 5th edition. Oxford: Butterworth-Heinemann. ISBN 0750657669.
- [79]. O'Conner, D. T. (2010). *Practical Reliability Engineering*. 4th edition, Wiley, ISBN 13:978-0-470-84463-2.
- [80]. Osarenren, J.O. (2015). *Integrated Reliability: Condition Monitoring and Maintenance of Equipment*. CRC Press, Taylor and Francis Group. ISBN 1482249405.
- [81]. Pan, S., Han, T., Tan, A.C. and Lin, T. R. (2016). Fault Diagnosis System of Induction Motors Based on Multiscale Entropy and Support Vector Machine with Mutual Information Algorithm. *Shock and Vibration*
- [82]. Pichler, K., Lughofer, E., Pichler, M., Buchegger, T., Klement, E. P., and Huschenbett, M. (2016). Fault Detection in Reciprocating Compressor

- Valves Under Varying Load Conditions. *Mechanical Systems and Signal Processing*. Vol. 70–71, 104-119.
- [83]. Qin, Q., Jiang, Z.-N., Feng, K. and He, W., (2012). A Novel Scheme for Fault Detection of Reciprocating Compressor Valves Based on Basis Pursuit, Wave Matching and Support Vector Machine. *Measurement*, Vol. 45, 897-908.
- [84]. Ragno, G., De Luca, M. and Loele, G. (2007). An Application of Cluster Analysis and Multivariate Classification Methods to Spring Water Monitoring Data. *Microchemical Journal*, Vol 87 (3), 119-127.
- [85]. Rao, B. (1996). *Handbook of Condition Monitoring*. Elsevier Science Ltd, Oxford, UK, ISBN 1 85617 2341.
- [86]. Rao, V. R. and Kalyankar, V.D. (2013). Parameter Optimization of Modern Machining Processes Using Teaching–Learning-Based Optimization Algorithm. *Engineering Applications of Artificial Intelligence*, Vol. 26 (1), 524-531, ISSN 0952-1976.
- [87]. Rao, J.S. (2000). *Vibratory Condition Monitoring of Machine*. Alpha Science International Ltd. ISBN 9781842650103.
- [88]. Ruqiang, Y. and Gao, R. X. (2006). Hilbert–Huang Transform-Based Vibration Signal Analysis for Machine Health Monitoring. *Instrumentation and measurement*, *IEEE Transactions on* 55.6: 2320-2329.
- [89]. Safo, S.E. and Ahn, J., (2016). General Sparse Multi-Class Linear Discriminant Analysis. *Computational Statistics & Data Analysis*, Vol. 99, 81-90.
- [90]. Samanta, B. and Nataraj, C. (2009). Use of Particle Swarm Optimization for Machinery Fault Detection. *Engineering Applications of Artificial Intelligence*. Vol. 22(2), 308-316.
- [91]. SAS Institute (1985). *SAS user's guide: statistics* (Vol. 1 and 2). SAS Institute.
- [92]. Scholkopf, B., Smola, A. and Muller, K. R. (1996). A Technical Report No 44 *Nonlinear Component Analysis as a Kernel Eigenvalue Problem*.
- [93]. Shao, R., Hu, W., Wang, Y. and Qi X. (2014). The Fault Feature Extraction and Classification of Gear Using Principal Component Analysis and Kernel Principal Component Analysis Based On The Wavelet Packet Transform. *Measurement*, Vol. 54, 118-132, ISSN 0263-2241,
- [94]. Silverman, B. W. (1991). [Choosing a Kernel Regression Estimator]: Comment: Should we use kernel methods at all? *Statistical Science*, 6(4), 430-433.
- [95]. Silvey, S.D. (1979). *Statistical Inference: Monographs on Applied Probability and Statistics*. Chapman and Hall. ISBN 0-412-13820-4.

- [96]. Slack, N., Chambers, S. and Johnston, R. (2007). *Operations Management* 5th edition. Harlow: Pearson Education.
- [97]. Smith, A., Gu, F. and Ball, A. D. (2015). Selection of Input Parameters for Multivariate Classifiers in Proactive Machine Health Monitoring by Clustering Envelope Spectrum Harmonics. *Applied Mechanics and Materials*, 798, pp. 308-313. ISSN 1662-7482
- [98]. Smith, A., Gu, F. and Ball, A. (2016). Maintaining model efficiency, avoiding bias and reducing input parameter volume in compressor fault classification. In: *2016 7th International Conference on Mechanical and Aerospace Engineering*, 18th to 20th July 2016, London, UK , pp. 196-201
- [99]. Smith, J. D. (1989). *Vibration measurement and analysis*. London: Butterworths. ISBN 9780408041010, 0408041013.
- [100]. Smith, D. J. (2005). *Maintainability and Risk. 7th Edition-Practical Methods for Engineers Including Reliability Centred Maintenance and Safety Related Systems*. Butterworth-Heinemann. ISBN 9780080458939.
- [101]. Tipping, M.E. (2001). Sparse Bayesian Learning and the Relevance Vector Machine. *The Journal of Machine Learning Research*, Vol. 1, 211-244.
- [102]. Tipping, M.E. and Faul, A. (2002). Analysis of Sparse Bayesian Learning. *Advances in Neural Information Processing Systems*, Vol. 14, 383-389.
- [103]. Tran, V. T., Al Thobiani, F. and Ball, A. D. (2014). An Approach to Fault Diagnosis of Reciprocating Compressor Valves Using Teager–Kaiser Energy Operator and Deep Belief Networks. *Expert Systems with Applications*, Vol. 41, 4113-4122, 7.
- [104]. Wang, Y. Xue, C. Jia, X. and Peng, X. (2015). Fault Diagnosis of Reciprocating Compressor Valve with the Method Integrating Acoustic Emission Signal and Simulated Valve Motion. *Mechanical Systems and Signal Processing*, Vol. 56–57, 197-212.
- [105]. Wickerhauser, M. V. (1994). *Adapted Wavelet Analysis from Theory to Software*.
- [106]. Widodo, A. and Yang, B. (2007). Support Vector Machines in Condition Monitoring and Fault Diagnosis. *Mechanical Systems and Signal Processing*. 21(6), 2560-2574.
- [107]. Williams, J.H. (1994). *Condition-Based Maintenance and Machine Diagnostics*. Chapman and Hall. ISBN 10:0412465000.
- [108]. Yang, B. S., Hwang, W. W., Ko, M. H. and Lee, S. J. (2005a). Cavitation Detection of Butterfly Valve Using Support Vector Machines. *Journal of Sound Vibration*, Vol 287, 25-43.
- [109]. Yang, B. S., Hwang, W. W., Kim, D. J., Chit Tan, A. (2005b) *Condition Classification of Small Reciprocating Compressor for Refrigerators Using*

- Artificial Neural Networks and Support Vector Machines. *Mechanical Systems and Signal Processing*, Vol. 19, 371-390 ISSN 0888-3270.
- [110]. Yardley, E. D. (2002). *Condition Monitoring: Engineering the Practice*. Professional Engineering. ISBN 186058361x.
- [111]. Yuan, H., Gao, J., Guo, H. Z., and Lu, C. (2011). An Efficient Method to Process the Quantised Acoustoelectric Current: Wavelet transform. *IEEE Transactions on Instrumentation and Measurement*. Vol. 60(3), 696-702.
- [112]. Zaidan, M. A, Mills, A. R. and Harrison, R.F. (2013). Bayesian Framework for Aerospace Gas Turbine Engine Prognostics. *IEEE Aerospace Conference Proceedings*. Vol. 1, 43.
- [113]. Zhang, Y., Li, S. and Hu, Z., (2012). Improved Multi-Scale Kernel Principal Component Analysis and its Application for Fault Detection. *Chemical Engineering Research and Design*. Vol. 90 (9), 1271-1280, ISSN 0263-8762.
- [114]. Zhang, Y., Bingham, C. M. and Gallimore, M., (2013). Fault Detection and Diagnosis Based on Extensions of PCA. *Advances in Military Technology*, Vol. 8 (2), 27-41.
- [115]. Zheng, L., Li, T. and Ding, C. (2014). A Framework for Hierarchical Ensemble Clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(2), 1-23.
- [116]. Zio, E. (2013). *PHM for Nuclear Applications*. Politecnico di Milano, Centrale Superlec, Paris.
- [117]. Zvokrlj, M., Zupan, S. and Prebil, I. (2010). Multivariate and Multiscale Monitoring of Large-Size Low-Speed Bearings using Ensemble Empirical Mode Decomposition Method Combined with Principal Component Analysis. *Mechanical Systems and Signal Processing*. 24(4), 1049-1067.

Publications

2016

Smith, A., Gu, F., & Ball, A. (2016, July). Maintaining Model Efficiency, Avoiding Bias and Reducing Input Parameter Volume in Compressor Fault Classification. In: Mechanical and Aerospace Engineering (ICMAE), 2016 7th International Conference on (pp. 196-201). IEEE.

2015

Smith, A., Gu, F. and Ball, A. (2015). Selection of Input Parameters for Multivariate Classifiers in Proactive Machine Health Monitoring by Clustering Envelope Spectrum Harmonics. Applied Mechanics and Materials, 798, pp. 308-313. ISSN 1662-7482

Smith, A., Gu, F. and Ball, A. (2015). Selection of Input Parameters for Multivariate Classifiers in Proactive Machine Health Monitoring by Clustering Envelope Spectrum Harmonics. In: 6th International Conference on Mechanical and Aerospace Engineering, 16th - 17th July 2015, Rome, Italy

Abusaad, S., Benghozzi, A., Smith, A., Gu, F. and Ball, A. (2015). The Detection of Shaft Misalignments using Motor Current Signals from a Sensorless Variable Speed Drive. In: Vibration Engineering and Technology of Machinery. London: Springer. pp. 173-183. ISBN 978-3-319-09918-7

2014

Ahmed, M., Smith, A., Gu, F. and Ball, A. (2014). Fault Diagnosis of Reciprocating Compressors Using Relevance Vector Machines with A Genetic Algorithm Based on Vibration Data. In: The 20th International Conference on Automation and Computing (ICAC'14), 12-13 September 2014, Cranfield University

2013

Smith, A. and Gu, F. (2013). Distributional Considerations in Inference Based Condition Monitoring Stages: Detection, Diagnosis And Prognosis. In: Proceedings of Computing and Engineering Annual Researchers' Conference 2013: CEARC'13. Huddersfield: University of Huddersfield. p. 238. ISBN 9781862181212

Appendices

A1. Experimental Compressor Rig

Experimental Study



1. Compressor Specification

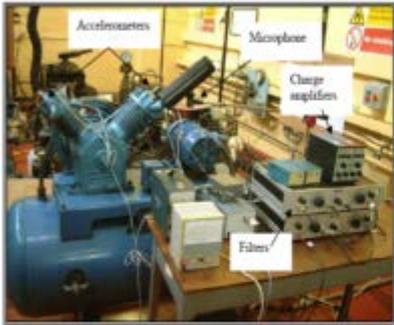
➤ Two Stage Broom-Wade Reciprocating Compressor

– Type	TS9
– Max working pressure	200 lbs/in ²
– Number of cylinders	2 (90° Vee)
– Piston stroke	76 mm
– Crank Speed	420 rpm
– Motor power	2.2 kW
– Supply Volts	380v
– Motor Speed	1425 rpm

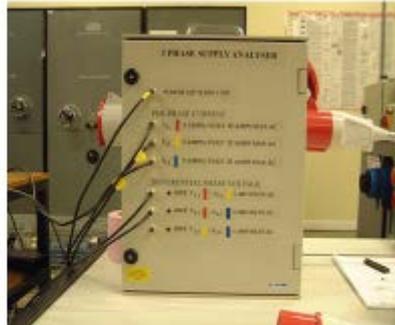


Broom-Wade Compressor

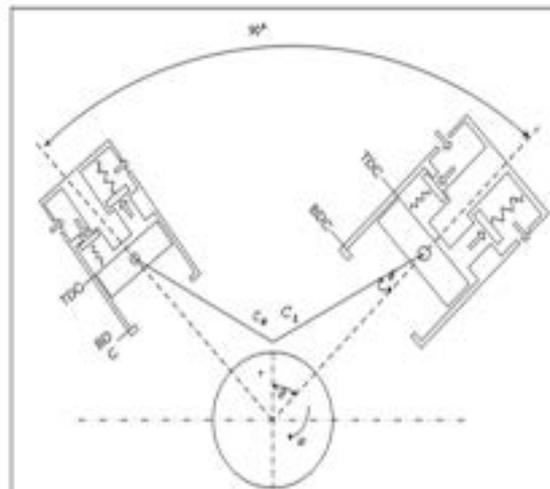
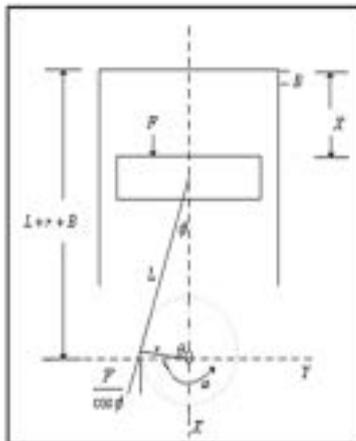
2. Test Rig Facilities



Test system includes in-cylinder pressure, speed and vibration measurements



Three-phase power supply analyser includes phase voltage, current and power measurements

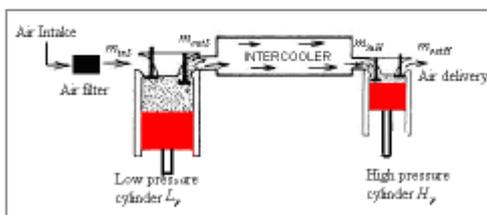


Cylinder pressures

Low Pressure Cylinder

$$\dot{p}_{cL} = \frac{1}{V_{cL}} [C_{dL}^2 \dot{m}_{vdL} - C_{cL}^2 \dot{m}_{vdL} - \gamma P_{cL} v_{cL}]$$

$$\dot{m}_{vdL} = \beta_{dL} C_{dL} A_{dL} \sqrt{2 P_{at} |P_{cL} - P_{at}|}$$



High Pressure Cylinder

$$\dot{p}_{cH} = \frac{1}{V_{cH}} [C_{dH}^2 \dot{m}_{vdH} - C_{cH}^2 \dot{m}_{vdH} - \gamma P_{cH} v_{cH}]$$

$$\dot{m}_{vdH} = \beta_{dH} C_{dH} A_{dH} \sqrt{2 P_{at} |P_{cH} - P_{at}|}$$

Fault Models



➤ Valve Leakage Model

$$\dot{p}_{cH} = \frac{1}{V_{cL,H}} [C_{dL}^2 \dot{m}_{vdL} - C_{cH}^2 \dot{m}_{vdH} - \gamma P_{cH} v_{cH} - C_{dH}^2 \dot{m}_{dH} - C_{cH}^2 \dot{m}_{dH}]$$

The flow rate of discharge leakage \dot{m}_{dH} can be calculated

$$\dot{m}_{dH} = \beta_{dH} C_{dH} A_{dH} \sqrt{2 P_{at} |P_{cH} - P_{at}|}$$

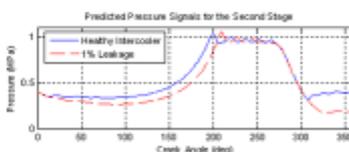


➤ Intercooler Leakage Model

$$\dot{p}_{in} = \frac{1}{V_{in}} [C_{dL}^2 \dot{m}_{vdL} - C_{dH}^2 \dot{m}_{vdH} - C_{in}^2 \dot{m}_{in}]$$

The flow rate of intercooler leakage \dot{m}_{in} can be calculated

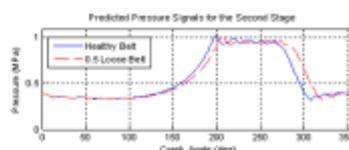
$$\dot{m}_{in} = \beta_{in} C_{in} A_{in} \sqrt{2 P_{in} |P_{in} - P_o|}$$



➤ Belt Transmission Model

$$J_c \frac{d\omega_c}{dt} = [C_b \Delta \dot{\theta} + (K_b - \delta K_b) \Delta \theta] / v_0 - T_c$$

δK_b = is the change of the belt stiffness due to belt deterioration



A2 MATLAB Code

A2.1 MATLAB Code for Cluster Analysis

Using the classification matrix, X: [observations-by-variables]organised by class e.g. the first 24 measurements on the healthy system given by s(1).Aeh ('Aeh'the amplitude of the envelope harmonics) and so on for each class of fault.

```
X=[s(1).Aeh'; s(2).Aeh'; s(3).Aeh'; s(4).Aeh'; s(5).Aeh'];

% cluster analysis. 1/10/14
% data is now arranged in a matrix X (obs by features) (120 by 32)
d = pdist(X'); % the pairwise Euclidean distance(default measure) between
each of the obs.
% 1 by 496 = 32*31/2=m(m-1)/2
% Z = linkage(d); % an (m-1) by
% c = cluster(Z,'maxclust',3:5);
% construct a dendrogram for clustering of variables (features)
%re order matrix into 32-by-120 (amplitudes for each load by harmonic
features)
z = linkage(d);
c = cluster(z,'maxclust',6);
crosstab(c);
[H,T]=dendrogram(z, 'COLORTHRESHOLD',3.1); %Colour threashold is set to 3.1
title ('Dendrogram: Feature clustering based on Euclidean Distance Between
Pairwise Observations');
xlabel('Harmonic');
ylabel('Distance');
```

A2.2 MATLAB Code for Discriminant Analysis

```
DISCRIMINANT ANALYSIS-CLASSIFICATION (12Nov14)
% 32 COLUMNS OF 120 MEASUREMENTS [MATRIX X]
% select two variables for 2-d plot
% e.g. X(:,4) for harmonic 4 etc
% GROUP = label
% DISCRIMINANT ANALYSIS-CLASSIFICATION

%Define the variables within the matrix X
%each being 24 by 32 sub matrix
%X=[s(1).Aeh'; s(2).Aeh'; s(3).Aeh'; s(4).Aeh'; s(5).Aeh'];
H= s(1).Aeh';
DVL= s(2).Aeh';
ICL= s(3).Aeh';
LB=s(4).Aeh';
SVL=s(5).Aeh';

figure(16),clf
```

```

X2Gp=[s(1).Aeh'; s(3).Aeh']; %just the healthy and ICL data
Gps2={Group{: ,1} Group{: ,3}}';
%group = 2Groups %label
%label={Group{: ,1} Group{: ,2} Group{: ,3} Group{: ,4} Group{: ,5}}';
%eg for healthy and SVL use 2Groups={Group{: ,1} Group{: ,5}}'

Plot the two groups (H and ICL) data using harmonics 4 and 6

h1 = gscatter(X2Gp(:,4),X2Gp(:,6),Gps2,'rb','v^',[],'off');
set(h1,'LineWidth',2)
title('Scatter Plot of Healthy and ICL');
xlabel('Feature 4 ');
ylabel('Feature 6');
legend('Healthy','ICL',...
       'Location','NW')

%to classify the groups:
N = size(X,1); %120 (24 measurements on each of 5 classes)

%Linear discriminant analysis using harmonics 4 and 6
ldaClass = classify([X(:,4) X(:,6)],[X(:,4) X(:,6)],label5);

bad = ~strcmp(ldaClass,label5);
ldaResubErr = sum(bad) / N;
[ldaResubCM,grpOrder] = confusionmat(label5,ldaClass)
hold on;
plot(X(bad,1), X(bad,2), 'kx');
hold off;
[x,y] = meshgrid(4:.1:8,2:.1:4.5);
x = x(:);
y = y(:);
j = classify([x y],[X(:,4) X(:,6)],label5);
gscatter(x,y,j,'grb','sod')
qdaClass = classify([X(:,4) X(:,6)],[X(:,4) X(:,6)],label5,'quadratic');
bad = ~strcmp(qdaClass,label5);
qdaResubErr = sum(bad) / N
s = RandStream('mt19937ar','seed',0);
RandStream.setDefaultStream(s);
cp = cvpartition(label5,'k',10)
ldaClassFun= @(xtrain,ytrain,xtest)(classify(xtest,xtrain,ytrain));
ldaCVerErr = crossval('mcr',[X(:,4) X(:,6)],label5,'predfun', ...
ldaCVerErr = crossval('mcr',[X(:,4) X(:,6)],label5,'predfun', ... %'label'
in nb code change my label to label 5
        ldaClassFun,'partition',cp)
        qdaClassFun
=
@(xtrain,ytrain,xtest)(classify(xtest,xtrain,ytrain,...
        'quadratic'));
qdaCVerErr = crossval('mcr',[X(:,4) X(:,6)],label5,'predfun',...

```

```

qdaClassFun, 'partition', cp)

[X,Y] = meshgrid(linspace(0,5),linspace(0,5));
X = X(:); Y = Y(:);
[C,err,P,logp,coeff] = classify([X Y],[X2Gp(:,4) X2Gp(:,6)],...
                                Gps2 , 'quadratic');

%To Visualize the classification:
hold on;
gscatter(X,Y,C, 'rb', '.',1, 'off');
K = coeff(1,2).const;
L = coeff(1,2).linear;
Q = coeff(1,2).quadratic;
% Plot the curve K + [x,y]*L + [x,y]*Q*[x,y]' = 0:
f = @(x,y) K + L(1)*x + L(2)*y + Q(1,1)*x.^2 + ...
        (Q(1,2)+Q(2,1))*x.*y + Q(2,2)*y.^2
h2 = ezplot(f,[0 5 0 4]);
set(h2, 'Color', 'm', 'LineWidth', 2)
axis([0 5 0 5])
xlabel('Feature 4')
ylabel('Feature 6')
title('\bf Classification with Healthy and ICL Training Data')

```

A2.3 MATLAB Code for Naïve Bayes Classification

```

%Naive Bayes Classification investigations 30th Jan 2015
%Initial investigations assuming a multivariate normal
%distribution

N = size(X,1); %120 (24 measurements on each of 5 classes)

%Linear discriminant analysis using harmonics 4 and 6
ldaClass = classify([X(:,4) X(:,6)],[X(:,4) X(:,6)],label5);
%Classification using naive Bayes with assumption variables are multivariate
%normal dist (mnd)
bad = ~strcmp(ldaClass,label5);
ldaResubErr = sum(bad) / N;
[ldaResubCM,grpOrder] = confusionmat(label5,ldaClass)
hold on;
plot(X(bad,1), X(bad,2), 'kx');
hold off;
[x,y] = meshgrid(4:.1:8,2:.1:4.5);
x = x(:);
y = y(:);
j = classify([x y],[X(:,4) X(:,6)],label5);
gscatter(x,y,j, 'grb', 'sod')
qdaClass = classify([X(:,4) X(:,6)],[X(:,4) X(:,6)],label5, 'quadratic');

```

```

bad = ~strcmp(qdaClass,label5);
qdaResubErr = sum(bad) / N
s = RandStream('mt19937ar','seed',0);
RandStream.setDefaultStream(s);
cp = cvpartition(label5,'k',10)
ldaClassFun= @(xtrain,ytrain,xtest)(classify(xtest,xtrain,ytrain));
ldaCVERr = crossval('mcr',[X(:,4) X(:,6)],label5,'predfun', ...   '%label'
in nb code change my label to label 5
                ldaClassFun,'partition',cp)
                qdaClassFun
                =
@(xtrain,ytrain,xtest)(classify(xtest,xtrain,ytrain,...
                'quadratic'));
qdaCVERr = crossval('mcr',[X(:,4) X(:,6)],label5,'predfun',...
                qdaClassFun,'partition',cp)

nbGau= NaiveBayes.fit([X(:,4) X(:,6)], label5);
nbGauClass= nbGau.predict([X(:,4) X(:,6)]);
bad = ~strcmp(nbGauClass,label5);
nbGauResubErr = sum(bad) / N   %N=120 declared above
nbGauClassFun = @(xtrain,ytrain,xtest)...
                (predict(NaiveBayes.fit(xtrain,ytrain), xtest));
nbGauCVERr = crossval('mcr',[X(:,4) X(:,6)],label5,...
                'predfun', nbGauClassFun,'partition',cp)

%Alternative approach (not assuming mnd) using non-parametric kernel density
estimation

nbKD= NaiveBayes.fit([X(:,4) X(:,6)], label5,'dist','kernel');
nbKDClass= nbKD.predict([X(:,4) X(:,6)]);
bad = ~strcmp(nbKDClass,label5);
nbKDResubErr = sum(bad) / N
nbKDClassFun = @(xtrain,ytrain,xtest)...
                (predict(NaiveBayes.fit(xtrain,ytrain,'dist','kernel'),xtest));
nbKDCVERr = crossval('mcr',[X(:,4) X(:,6)],label5,...
                'predfun', nbKDClassFun,'partition',cp)

% Note: kernel error identical to Gaussian 0.0417

%Decision Trees offer useful visual and direct classification of further
%data

t = classregtree([X(:,4) X(:,6)], label5,'names',{'f4' 'f6' });
[grpname,node] = t.eval([x y]);
gscatter(x,y,grpname,'grb','sod')
view(t);

```

```

%Compute the resubstitution error and the cross-validation error for decision
tree.
dtclass = t.eval([X(:,4) X(:,6)]);
bad = ~strcmp(dtclass,label5);
dtResubErr = sum(bad) / N

dtClassFun
=
@(xtrain,ytrain,xtest)(eval(classregtree(xtrain,ytrain),xtest));
dtCVerErr = crossval('mcr',[X(:,4) X(:,6)],label5, ...
    'predfun', dtClassFun,'partition',cp)

%% Naive Bayes classification matrix (confusion matrix) and visual output as
3-d bar chart.
%myfNo=[1:32];
myfNo= [2 3 4 6 7 8 9 10 11 14 15]%input parameter set
%[2 3 4 5 6 7 8 9 10 11 12 13 14 15 16];% %[4 6]%[2 7 9 12 17]%
O1 = NaiveBayes.fit(X(:,myfNo),label5);
C1 = O1.predict(X(:,myfNo));
cMat1 = confusionmat(label5,C1) %
figure(5),clf
bar3(cMat1/24*100)
set(gca,'xticklabel',CaseStr)
set(gca,'yticklabel',CaseStr)
zlabel('Classification Rate(%)')

```

A2.4 MATLAB Code for PCA

```

figure(1),clf
    plot3(pk',sigma',kur','*')
xlabel('Peak'),ylabel('rms'),zlabel('kurtosis');
grid on
legend(CaseStr)
title('Plot of variables by case');

% organise pca data matrix

pk=pk';sigma=sigma';kur=kur'; skew=skew';
pcadata=[pk(:) sigma(:) kur(:) skew(:)];

[pc,score,latent,tsquare] = princomp(pcadata);

cumsum(latent)./sum(latent)

```

```

xmarker={'ob','^m','dr','pk','*g','*c','*m','*r'};
figure(2),clf
hold on
for icase=1:numfaults
    plot(score((icase-1)*numLoad+1:icase*numLoad,1),score((icase-
1)*numLoad+1:icase*numLoad,2),xmarker{icase})
end

legend(CaseStr)
Title('Principal Component Analysis based on standard deviation, kurtosis,
peak value and skewness.');
```

```

xlabel('pc1');
ylabel('pc2');
```

```

figure
pcclusters = clusterdata(score(:,1:2),2);
gscatter(score(:,1),score(:,2),pcclusters)
xlabel('First Principal Component');
ylabel('Second Principal Component');
title('Principal Component Scatter Plot with 2 Coloured Clusters');
```

```

%Biplot--axis                                rotation:
biplot(pc(:,1:2),'Scores',score(:,1:2),'VarLabels',...
    % {'X1' 'X2' 'X3' 'X4'})
```

```

%Factor analysis: note pca matrix is pcadata=[pk, sigma, kur]
% maximum likelihood estimate, lambda, of the factor loadings matrix, in a
% common factor analysis model with m common factors. X is an n-by-d matrix
where each row is an observation of d variables.
lambda = factoran(pcadata,1)
% gives one factor with loadings lambda =    0.4477    0.9975   -0.5152
```

A2.5 MATLAB Code for SVM

Classify data using support vector machine 17/08/15

```

% Syntax
% SVMStruct = svmtrain(X, Group)
% Group = svmclassify(SVMStruct, X)
% Group = svmclassify(SVMStruct, X, 'Showplot', ShowplotValue)
%
```

```

% Description
% Group = svmclassify(SVMStruct, Sample) classifies each row of the data in
% Sample using the information in a support vector machine classifier
structure SVMStruct,
%created using the svmtrain function. Sample must have the same number of
columns as the data used to train
```

```

%the classifier in svmtrain. Group indicates the group to which each row of
Sample has been assigned.
% Group = svmclassify(SVMStruct, Sample, 'Showplot', ShowplotValue)
%controls the plotting of the sample data in the figure created using the
Showplot property with the svmtrain function.
% Examples
%
%
data=[X(:,4), X(:,7)]; %Using harmonics 4 and 7

% From the label vector, [label5] create a new column vector, groups, to
classify data into two groups:
%Classify 'healthy' and 'faulty'
groups = ismember(label5, 'Healthy'); %'Group' here may need replacing with
'CaseStr' ???

SVMStruct = svmtrain(X, groups)
Group = svmclassify(SVMStruct, X)
%Group = svmclassify(SVMStruct, X, 'Showplot', ShowplotValue)
%?ShowplotValue ?

% Randomly select training and test sets.

[train, test] = crossvalind('holdOut',groups);
cp = classperf(groups);
%
% Use the svmtrain function to train an SVM classifier using a linear kernel
function and plot the grouped data.
svmStruct = svmtrain(data(train,:),groups(train),'showplot',true);

% Add a title to the plot, using the KernelFunction field from the svmStruct
structure as the title.
title(sprintf('Kernel Function: %s',...
             func2str(svmStruct.KernelFunction)),...
       'interpreter','none');

% Classify the test set using a support vector machine.

classes = svmclassify(svmStruct,data(test,:), 'showplot',true);

% Evaluate the performance of the classifier.

classperf(cp,classes,test);
cp.CorrectRate

```

```

% Use a one-norm, hard margin support vector machine classifier by changing
the boxconstraint property.
figure
svmStruct = svmtrain(data(train,:),groups(train),...
                    'showplot',true,'boxconstraint',1e6);

classes = svmclassify(svmStruct,data(test,:), 'showplot',true);

% Evaluate the performance of the classifier.

classperf(cp,classes,test);
cp.CorrectRate

```

A2.6 MATLAB Code for MSPCA

```

%Multiscale PCA.
%first set the wavelet parameters
level = 5;
wname = 'sym4';
npc = 'kais'; %Use Kaiser's rule: retains PCs with eigenvalues > mean value

[x_sim, qual, npc] = wmspca(X ,level, wname, npc); % multiscale PCA:

%Display the original and simplified signals:
figure(5),clf
kp = 0;
for i = 1:4
    subplot(4,2,kp+1), plot(X (:,i)); axis tight;
    title(['Original signal ',num2str(i)])
    subplot(4,2,kp+2), plot(x_sim(:,i)); axis tight;
    title(['Simplified signal ',num2str(i)])
    kp = kp + 2;
end

% Are results from a compression perspective good?
%The percentages reflecting the quality of column reconstructions
%given by the relative mean square errors are close to 100%.
qual

%The output argument npc is the number of retained principal components
selected by Kaiser's rule:
npc

%To suppress the details at levels 1 to 3, update the npc argument as follows:
% npc(1:3) = zeros(1,3);
% npc
% npc =      0      0      0      1      1      2      2

```

```

%Supress details at levels 1 to 3:
npc = [0 0 0 2 2 1 3];
%npc

%Repeat the multiscale PCA using reduced PC set:
[x_sim, qual, npc] = wmspca(X, level, wname, npc);

%Display the Original and Final Simplified Signals
figure(6),clf
kp = 0;
for i = 1:4
    subplot(4,2,kp+1), plot(X (:,i)); axis tight;
    title(['Original signal ',num2str(i)])
    subplot(4,2,kp+2), plot(x_sim(:,i)); axis tight;
    title(['Simplified signal ',num2str(i)])
    kp = kp + 2;
end

% 16/2/16 The simplified signals are in matrix x.
%The parameters of multiscale PCA are available in PCA_Params:
%PCA_Params

%1x7 struct array with fields:
    pc
    variances
    npc

%PCA_Params is a structure array of length d+2 (here, the maximum
decomposition level d=5)
%such that PCA_Params(d).pc is the matrix of principal components.
%The columns are stored in descending order of the variances.
%PCA_Params(d).variances is the principal component variances vector, and
%PCA_Params(d).npc is the vector of selected numbers of retained principal
components

%Display original and simplified signals for harmonics 4 6 7 9
%Display the original and simplified signals:
figure(7),clf
kp = 0;
for i = [4, 6, 7, 9]
    subplot(4,2,kp+1), plot(X (:,i)); axis tight;
    title(['Original signal ',num2str(i)])
    subplot(4,2,kp+2), plot(x_sim(:,i)); axis tight;
    title(['Simplified signal ',num2str(i)])
    kp = kp + 2;
end

```

```

%Supress details at levels 1 to 3:
npc = [0 0 0 2 2 1 3];
%npc

%Repeat the multiscale PCA using reduced PC set:
[x_sim, qual, npc] = wmspca(X, level, wname, npc);

%Display the Original and Final Simplified Signals
figure(8),clf
kp = 0;
for i = [4, 6, 7, 9]
    subplot(4,2,kp+1), plot(X (:,i)); axis tight;
    title(['Original signal ',num2str(i)])
    subplot(4,2,kp+2), plot(x_sim(:,i)); axis tight;
    title(['Simplified signal ',num2str(i)])
    kp = kp + 2;
end

% Andrews plot of simplified signal profiles by group, figure 12
figure(12), clf
%subplot (2,1,1)
andrewsplot((x_sim(:,i)), 'group', label)% ,co{icase}
title ('Andrews Plot by case');
xlabel('t');
ylabel('f(t)');

%Andrews plot of original signal profiles by group, figure 13
figure(13), clf
%subplot (2,1,1)
andrewsplot((X (:,i)), 'group', label)% ,co{icase}
title ('Andrews Plot by case');
xlabel('t');
ylabel('f(t)');

harmonic=1:32;
[Loadings1,specVar1,T,stats] = factoran(X,2);
scatter(Loadings1(:,1),Loadings1(:,2), harmonic)% how to label data points
1 to 32
xlabel('Factor 1')
ylabel('Factor 2')
harmonic=1:32;

A2.7 MATLAB Code for Andrews plots
andrewsplot((X (:,i)), 'group', label)% ,co{icase}
title ('Andrews Plot by case');
xlabel('t');
ylabel('f(t)');

```