Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators.

David Shally
Technological University Dublin

Follow this and additional works at: https://arrow.tudublin.ie/engmas

Part of the Controls and Control Theory Commons, Electrical and Electronics Commons, and the Power and Energy Commons

Recommended Citation

This Theses, Masters is brought to you for free and open access by the Engineering at ARROW@TU Dublin. It has been accepted for inclusion in Masters by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.
Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

David Shally, B.Eng.

Submitted for the Qualification of Master of Philosophy

Dublin Institute of Technology

Supervisors:  Mr. Michael Farrell

Mr. Kevin Sullivan

Dr. Michael Conlon

School of Control Systems and Electrical Engineering

May 2012
Abstract

This research project was set up to monitor, on an ongoing basis, the condition of the end windings and their support structures of a 288 MVA 2-pole synchronous generator with a known end winding vibration problem.

Excessive vibration at the end windings was caused by the natural frequencies of individual end bars (local) and of the entire end winding structure (global) being at or close to the magnetic forcing frequency of 100 Hz. Resonant vibration such as this has been a cause of major failures in machines of the same type in the past, resulting in significant down time ranging from a few weeks to a number of months, with the obvious implications in terms of cost, generator availability and revenue loss.

The project covers the installation of an end winding vibration monitoring system, subsequent analysis and testing in order to attempt to lower peak vibration levels, a low tune modification of the end winding support structure (carried out by the OEM) in order to remedy the natural frequency issue, subsequent data analysis and development of regression models to allow the prediction of vibration levels based on plant data and the implementation of the models on PI ACE (Advance Computing Engine). A webpage, available to all plant personnel via the company intranet was then created to display the actual measured vibration data against the model predicted values and other relevant plant data.

With the issue of excessive vibration due to resonance having been remedied by the Original Equipment Manufacturer (OEM) the focus of the project shifted somewhat towards providing an easily interpreted, easily accessible method of monitoring the condition of a now relatively healthy machine well beyond the conclusion of this project. The development of the regression models and web page was seen as the best way of achieving this.
Declaration Page

I certify that this thesis which I now submit for examination for the award of MPhil, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology and has not been submitted in whole or in part for an award in any other Institute or University.

The work reported on in this thesis conforms to the principles and requirements of the Institute’s guidelines for ethics in research.

DIT has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature: [Signature] Date: [Date]

David Shally
Acknowledgements

The following people have been of great help to me in my work:

Mr. Michael Farrell
Mr. Kevin Sullivan
Dr. Joe Condon
Dr. Michael Conlon
Mr. Damien Geraghty
Mr. Mark McKenna
Abbreviations List

EWV    End Winding Vibration
AOH    Annual Overhaul
EOH    Equivalent Operating Hours
EPA    Environmental Protection Agency
IPPC   Integrated Pollution Prevention and Control
FFT    Fast Fourier Transform
FOA    Fibre Optic Accelerometer
GTG    Gas Turbine Generator
CCGT   Combined Cycle Gas Turbine
HRSG   Heat Recovery Steam Generator
VPI    Vacuum Pressure Impregnation
DPT    Dye Penetrant Test
IGV    Inlet Guide Vane
TDC    Top Dead Centre
PGSC   Power Generation Services Company
MSG    Minimum Stable Generation (see Min Gen)
Min Gen Minimum Stable Generation Capability of the Generator/Turbine as declared to the Transmission System Operator (Eirgrid). Generally, this value is the lowest sustainable load that can be achieved by the Turbine without breaching EPA limits for emissions to air set out in the relevant IPPC Licence.
Max Gen Maximum Stable Generation Capability of the Generator/Turbine as declared to the Transmission System Operator. This is usually limited by the temperature withstand limits of the turbine components. These limits feed into generator sizing etc.
CCCW   Closed Circuit Cooling Water
TE     Turbine End
EE     Exciter End
CE     Connected End. The end of the machine where the terminal connections are made. Equivalent to the Turbine End/Drive End of the machine in this case.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCE</td>
<td>Non-Connected End. Equivalent to Exciter End/Non-Drive End of this particular generator</td>
</tr>
<tr>
<td>OEM</td>
<td>Original Equipment Manufacturer</td>
</tr>
<tr>
<td>TSO</td>
<td>Transmission System Operator</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple Linear Regression</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted Least Squares</td>
</tr>
<tr>
<td>LOWESS</td>
<td>Locally Weighted Regression and Scatterplot Smoothing</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>GOF</td>
<td>Goodness of Fit</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
</tr>
<tr>
<td>PD</td>
<td>Polynomial Degree</td>
</tr>
<tr>
<td>DF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Squared Error</td>
</tr>
<tr>
<td>ACE</td>
<td>Advanced Computing Engine</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>OPC</td>
<td>Open Process Control, formerly OLE for Process Control (Pre Active X)</td>
</tr>
</tbody>
</table>
# Table of Contents

**Abstract** .......................................................................................................................... 1

**Declaration Page** ............................................................................................................ 2

**Acknowledgements** ......................................................................................................... 3

**Abbreviations List** ........................................................................................................... 4

**Table of Contents** ............................................................................................................. 6

**Figures, Tables and Equations** ........................................................................................ 13

**Chapter 1**

**Project Introduction** ........................................................................................................ 23

**Chapter 2**

**Literature Review** ............................................................................................................. 27

2.1 End Winding Design ........................................................................................................ 27

2.2 Online End Winding Vibration Monitoring .................................................................. 34

**Chapter 3**

**Gas Turbine Generator End Winding Vibration** ................................................................. 37

3.1 The Power Station ........................................................................................................... 37

3.2 The Gas Turbine Generator ............................................................................................ 37

3.3 Gas Turbine Generator End Winding Vibration ............................................................. 38

3.4 Definition of Natural Frequency .................................................................................... 39

3.5 Static Natural Frequency Testing ................................................................................... 40

3.6 GT Generator End Winding Vibration Timeline ............................................................. 41

3.6.1 Summary of Bump Tests ............................................................................................ 41

3.6.2 Scheduled Major Outage April 2005 (AOH05) ......................................................... 42

3.6.3 Summary of 2005 Bump Test (Post Repair) ............................................................... 44
3.6.4 Scheduled Minor Outage April 2006 (AOH06) ........................................... 45
3.6.5 Quarterly Inspection No.1 (EWV Related) July 2006 ................................. 48
3.6.6 Quarterly Inspection No.2 (EWV Related) September 2006 ......................... 48
3.6.7 Outage for EWV Monitoring System Installation November 2006 .......... 49
3.6.8 Scheduled Major Outage April 2007 (AOH07) ........................................ 49
3.6.9 Forced Outage (Gas Turbine Related) October 2007 ................................. 50
3.6.10 Scheduled Minor Outage April 2008 ..................................................... 50

3.7 Failures of Other Generators of the Same Type Due To Excessive End Winding Vibration ................................................................. 50

3.8 The End Winding Vibration Monitoring System ........................................... 54
  3.8.1 Location of Accelerometers and Plane of Measurement .......................... 55
  3.8.2 End Winding Vibration Monitoring System Overview ............................ 59
  3.8.3 Penetration of Generator for FOA Cabling ............................................. 60
  3.8.4 Replacement of Steel Man Way Covers with Perspex ............................ 60

3.9 Strengths and Weaknesses of the Vibrosyst M EWV Monitoring System ........ 61
  3.9.1 Strengths ................................................................................................. 61
  3.9.2 Weaknesses ............................................................................................ 61

3.10 End Winding Monitoring System Improvements ........................................ 62
  3.10.1 Connection to Delphin Flight Recorder System for High Resolution Data Recording .................................................................................. 62
  3.10.2 Connection to OSISOFT PI Data Acquisition System ......................... 62

Chapter 4

Initial Data Analysis .......................................................................................... 64

4.1 Analysis of Trend Data .................................................................................. 64

4.2 Basic Descriptive Statistics and Correlation ................................................. 68
  4.2.1 Basic Descriptive Statistics ...................................................................... 68
  4.2.2 Correlation Data ..................................................................................... 70

4.3 Measures Taken to Reduce End Winding Vibration Levels and Setting of Alarm Thresholds ................................................................. 72
  4.3.1 Setting of Alarm Thresholds .................................................................... 76

4.4 Experiments Undertaken in Order to Reduce Peak End Winding Vibration Levels ................................. 77
  4.4.1 Peak Vibration Reduction, Option No.1 .................................................... 78
  Introduction of Excitation Limit ...................................................................... 78
Chapter 5
Analysis of Measured Vibration Data Post Low Tune Repair

Chapter 6
Regression Modelling Post Low Tune Repair
6.7 Multiple Linear Regression Models Calculated Using the Stepwise Variable Selection Procedure ................................................................................................................................. 118
  6.7.1 EE27 Multiple Linear Regression (Stepwise Selection, Level 0) .................. 120
  6.7.2 EE27 Multiple Linear Regression (Stepwise Selection, Level 1) .................. 123
  6.7.3 EE27 Multiple Linear Regression (Stepwise Selection, Level 2) .................. 125
  6.7.4 EE27 Multiple Linear Regression (Stepwise Selection, Level 3) .................. 127
  6.7.5 EE27 Multiple Linear Regression (Stepwise Selection, Level 4) .................. 129

6.8 Graphical Assessment of Model Predictive Capability for EE27 L0, L4 and LOWESS Model .......................................................................................................................... 133

6.9 Assessment of Model Error Terms .................................................................................................................................................................................. 135
  6.9.1 Residual Plots .......................................................................................................................... 136
  6.9.2 Model Residual -v- Model Prediction Plots ........................................................................ 138
  6.9.3 Residual Statistics and Distribution Fitting ........................................................................ 141
  6.9.3.1 EE19 L0 MLR Model Residual Statistics ........................................................................ 142
  6.9.3.2 EE19 L4 MLR Residual Statistics ................................................................................. 143
  6.9.3.3 EE19 LOWESS Residual Statistics .............................................................................. 144
  6.9.4 Autocorrelation of the Residuals ...................................................................................... 145

6.10 Model Validation Techniques and Selection of MLR Models for Implementation on PI ACE .......................................................................................................................... 148

6.11 Chapter Summary .......................................................................................................................... 149

Chapter 7
Statistical Analysis of Generator Condition 2006 – 2011 ......................................................... 151
  7.1 Scattergrams ........................................................................................................................ 151
  7.2 Pearson Correlation Tables ............................................................................................... 154
  7.3 Descriptive Statistics ......................................................................................................... 156

Chapter 8
Predictive Performance of MLR And Lowess Models .............................................................. 162
  8.1 2007 Multiple Linear Regression and LOWESS Non-Parametric Regression Models .......................................................................................................................... 164
  8.2 Requirements for Recalculation of Models ..................................................................... 178
    8.2.1 2008 Multiple Linear Regression and LOWESS Non-Parametric Regression Models .......................................................................................................................... 178
8.2.2 2009 Multiple Linear Regression and LOWESS Non-Parametric Regression Models ................................................................. 179
8.2.3 2009/2010 Multiple Linear Regression and LOWESS Non-Parametric Regression Models ............................................................. 180

Chapter 9
Improvement in Model Design due to Availability of Additional Independent Variables .................................................................................. 181
9.1 Construction Data for 2010 Models ................................................................................................................................................ 182
9.2 New Measured and Calculated Independent Variables Made Available for Selection ........................................................................................................ 182
9.3 New Model Independent Variable Arrays/Model Designations .............................................................................................................. 184
9.4 Predictive Performance of New MLR and LOWESS Models .................................................................................................................. 186
  9.4.1 Assessment of Models using Trend Data ......................................................................................................................................... 188
  9.4.2 Assessment of Models using Pearson Correlation Data .................................................................................................................... 191
  9.4.3 Assessment of Models using Goodness of Fit Statistics .................................................................................................................... 194
9.5 Analysis of the Poor Performance of MLR and LOWESS Models for TE25 ....................................................................................... 200

Conclusion ............................................................................................................................................................................................................ 207

References/Bibliography ................................................................................................................................................................................. 211

Appendix 1
End Winding Vibration Monitoring System ................................................................................................................................................... 215
A1.1 Hardware ................................................................................................................................................................................................ 215
  A1.1.1 Fibre Optic Accelerometers (FOA) .................................................................................................................................................. 215
  A1.1.2 The PCU 100 Programmable Units ............................................................................................................................................... 215
  A1.1.3 Fibre Optic Accelerometer Module ............................................................................................................................................. 216
  A1.1.4 Analogue Output Module ............................................................................................................................................................. 216
A1.2 System Installation ................................................................................................................................................................................................ 217
## Appendix 2

**GT Generator End Winding Low Tune Repair** ............................................. 220

A2.1 Requirement for Low Tune Repair of Generator End Windings .................. 220
A2.2 Aim of Low Tune/Detune Repair of GT Generator End Windings ............... 221
A2.2 GT Generator End Winding Low Tune/Detune Measure ................................ 222
   A2.2.1 Schedule of Repair Work ................................................................. 222
   A2.2.2 Initial Inspection .............................................................................. 223
   A2.2.3 Tangential Blocking ........................................................................ 223
   A2.2.4 Axial and Radial Decoupling of Support Ring ................................. 223
   A2.2.5 Drilling of the Support Ring .............................................................. 225
   A2.2.6 Slitting of the Support Braces ......................................................... 225
   A2.2.7 Strapping of Lower Layer Bars to the Support Ring ..................... 226
   A2.2.8 Resin Treatment of End Windings .................................................. 226
   A2.2.9 Bump Test No.1 ............................................................................. 227
   A2.2.10 Secondary Slotting of Support Braces to Fine Tune the ‘Low Tune’ Repair... 228
   A2.2.11 Bump Test No.2 ............................................................................. 229
   A2.2.12 Electrical Testing ........................................................................... 229
   A2.2.13 Final Inspection ............................................................................ 229

## Appendix 3

**Regression** .................................................................................................. 230

A3.1 Simple Linear Regression ......................................................................... 230
A3.2 Multiple Linear Regression ...................................................................... 231
   A3.2.1 The Multiple Regression Model Equation ........................................ 233
   A3.2.2 Calculation of Coefficients ............................................................. 233
   A3.2.3 Regression Model Summary .......................................................... 233
A3.3 Analysis of Regression Model ................................................................ 234
   A3.3.1 ANOVA Table (ANalysis Of VAriance) ........................................... 234
   A3.3.2 Parameter Information/Coefficient Table ....................................... 235
   A3.3.3 Coefficient of Determination ($R^2$) ............................................... 236
A3.4 Weighted Least Squares Regression ....................................................... 238
A3.5 Model Validation ..................................................................................... 239
   A3.5.1 Statistics for Validation of Regression Model .................................. 240
A3.6 Prediction Error ....................................................................................... 241
A3.6.1 Sum of Squares due to Error (SSE) ................................................................. 241
A3.6.2 Mean Squared Error (MSE) ........................................................................... 241
A3.6.3 Root Mean Squared Error (RMSE_{PRED}) ................................................... 242
A3.6.4 Predicted R^2 (R^2_{PRED}) ........................................................................... 243
A3.7 Variable Selection .............................................................................................. 244
A3.8 Regression Model Prediction ............................................................................. 245
  A3.8.1 Prediction Error Bars and Confidence Intervals ........................................... 246
  A3.8.2 Standard Error of Prediction ....................................................................... 246
  A3.8.3 Root Mean Square Error of Prediction, RMSE ............................................ 248
A3.9 Interpolation and Extrapolation in Regression Modelling ................................ 249

Appendix 4

Procedure for Recalculation of MLR Models in Matlab for Implementation on PI

ACE/Process Book .................................................................................................... 250
  A4.1 The Need for Recalculation of Model Coefficients ........................................ 250
  A4.2 Procedure for Calculation/Recalculation and Implementation of New MLR
    Model Coefficients ............................................................................................... 251
    A4.2.1 Calculation/Recalculation of MLR Models using Matlab ......................... 251
    A4.2.2 Updating the Models in PI ACE .................................................................... 257

Appendix 5

MLR and LOWESS Model Predictive Accuracy Data for 2008, 2009 and
2009/2010 Models .................................................................................................... 259
  A5.1 2008 Multiple Linear Regression and LOWESS Non-Parametric Regression
    Models ..................................................................................................................... 259
  A5.2 2009 Multiple Linear Regression and LOWESS Non-Parametric Regression
    Models ..................................................................................................................... 264
  A5.3 2009 - 2010 Multiple Linear Regression and LOWESS Non-Parametric
    Regression Models ............................................................................................... 272

Publications .................................................................................................................. 277
Figures, Tables and Equations

Chapter 2
Equation 2.1. The Biot-Savart Law ................................................................. 27
Figure 2.1 Generator End Winding ................................................................. 28

Chapter 3
Figure 3.1 The Gas Turbine Generator ............................................................. 38
Table 3.1 GT Generator Bump Test Summary ............................................... 42
Figure 3.2 Dusting at GT Generator TE Busbar Connections to Collector Rings at 12 O’
Clock Position ................................................................. 43
Figure 3.3 Crack in Support Bracket No.6 .......................................................... 46
Figure 3.4 Location of Damaged Support Brackets (Z-Profiles) ......................... 46
Figure 3.5 Generator Following Fault to Earth at the Exciter End of the Machine due to
Excessive EWV ................................................................. 51
Figure 3.6 Generator Stator End Winding following End Coil Break due to Excessive
EWV ................................................................. 53
Figure 3.7 Generator Stator Winding following End Coil Break due to Excessive EWV .... 54
Figure 3.8 Gas Turbine Generator: Fibre Optic Accelerometer Locations on Winding
Diagram ................................................................. 57
Figure 3.9 Gas Turbine Generator: Fibre Optic Accelerometer Locations ................ 58
Figure 3.10 GTG EWV Monitoring System Overview ......................................... 59

Chapter 4
Figure 4.1 EE19 and EE35 Measured Vibration (Displacement), Active Power and
Reactive Power 12/12/2006 to 19/12/2006 ....................................................... 65
Figure 4.2 TE17 and TE41 Measured Vibration (Displacement), Active Power and
Reactive Power 12/12/2006 to 19/12/2006 ....................................................... 66
Figure 4.3 GT Generator Reactive Capability Curve ........................................... 69
Figure 4.4 GT Generator Correlation Data for Weeks W003, W010 and W017 ................ 71
Figure 4.5 EE19 and EE35 Measured Vibration (Displacement), Active Power, Reactive
Power, Average Slot Temperature and Ambient Air Temperature
16/01/2007 to 23/01/2007 ................................................................................ 73
Figure 4.6 EE35 Frequency Domain Vibration Data (FFT Result) .................................................. 76
Figure 4.7 Scattergram of Measured Vibration against Reactive for End Coil TE33.
Generator Output between 225 MW and 230 MW and
Ambient Air Temperature Greater than or Equal to 4 °C. .................................................. 78
Table 4.1 R-Squared Values of Measured Vibration against Reactive Power for all End
Coils. Generator Output between 225 MW and 230 MW and
Ambient Air Temperature Greater than or Equal to 4 °C. .................................................. 79
Figure 4.8 Scattergram of Measured Vibration against Reactive Power for End Coil
TE33. ................................................................................................................................. 80
Generator Load 135 MW to 140 MW and Ambient Air Temperature Greater than or
Equal to 4 °C. .................................................................................................................. 80
Table 4.2 R-Squared Values of Measured Vibration against Reactive Power for all End
Coils. Generator Load 135 MW to 140 MW and Ambient Air Temperature
Greater than or Equal to 4 °C. ......................................................................................... 81

Chapter 5
Figure 5.1 Gas Turbine Generator (GTG) Exciter End (EE) Measured Vibration, Active
Power and Reactive Power Post Repair. 05/05/2007 to 14/05/2007.............................. 85
Figure 5.2 Radar Diagram Showing Maximum, Minimum and Average Vibration
Amplitudes Measured at the EE of the GTG 13/03/2007 to 20/03/2007. Pre Low
Tune Repair. ................................................................................................................... 86
Figure 5.3 Radar Diagram Showing Maximum, Minimum and Average Vibration
Amplitudes Measured at the EE of the GTG 05/05/2007 to 14/05/2007. Post Low
Tune Repair. ................................................................................................................... 87
Figure 5.4 Radar Diagram Showing Maximum, Minimum and Average Vibration
Amplitudes Measured at the TE of the GTG 13/03/2007 to 20/03/2007. Pre Low
Tune Repair. ................................................................................................................... 88
Figure 5.5 Radar Diagram Showing Maximum, Minimum and Average Vibration
Amplitudes Measured at the TE of the GTG 05/05/2007 to 14/05/2007. Post Low
Tune Repair. ................................................................................................................... 88
Table 5.1 Table of Maximum Measured Vibration Values for all EE and TE
Accelerometers. Data is filtered for Average Slot Temperature between 67°C and
84°C. GTG Load 166 MW to 230 MW................................................................. 89
Figure 5.6 Correlation Tables for Weeks 03, 10 and 17 (Pre Low Tune Repair) ............. 91
Figure 5.7 Correlation Tables for Weeks 24, 25 and 30 (Post Low Tune Repair)............. 92

Chapter 6
Table 6. 1 TE25 Simple Regression Goodness of Fit Statistics ........................................ 110
Table 6. 2 TE25 Simple Regression Analysis Of Variance ............................................... 110
Table 6. 3 TE25 Simple Regression Model Parameters ...................................................... 110
Equation 6. 1 TE25 Equation of Simple Linear Regression Model .................................. 110
Figure 6.1 TE25 Simple Regression Measured And Predicted Vibration -v- Time .......... 112
Figure 6.2 TE25 Simple Regression Residual Plot .......................................................... 113
Figure 6. 3 TE25 Simple Regression Error Histogram .................................................... 113
Figure 6. 4 TE25 Simple Regression Cumulative Distributions ....................................... 114
Table 6. 4 TE25 MLR Goodness of Fit Statistics .............................................................. 115
Table 6. 5 TE25 MLR Analysis of Variance ................................................................. 115
Table 6. 6 TE25 MLR Model Parameters ........................................................................ 116
Equation 6. 2 TE25 Equation of MLR Model All 10 Plant Variables Included ............... 116
Table 6. 7 TE25 MLR Goodness of Fit Statistics (7 From 10 Selected Explanatory Variables) ............................................................................................................. 117
Table 6. 8 TE25 MLR Analysis of Variance (7 from 10 Selected Explanatory Variables) ... 117
Table 6. 9 TE25 MLR Model Parameters (7 from 10 Selected Explanatory Variables) ...... 117
Equation 6. 3 TE25 Equation of MLR Model (7 Most Significant Explanatory Variables Included) ............................................................................................................. 118
Table 6. 10 EE27 Matlab Stepwise MLR L0 Goodness of Fit Statistics ......................... 120
Table 6.11 EE27 Matlab Stepwise MLR L0 Analysis Of Variance ................................. 120
Table 6.12 EE27 Matlab Stepwise MLR L0 Model Parameters ...................................... 120
Table 6.13 EE27 L0 Correlation Table .......................................................................... 121
Figure 6. 5 Plot of EE27 MLR L0 – L4 RMSE against Stepwise Procedure Step Number ................................................................. 122
Table 6.14 EE27 Matlab Stepwise MLR L1 Goodness of Fit Statistics ......................... 123
Table 6.15 EE27 Matlab Stepwise MLR L1 Analysis Of Variance ................................. 123
Table 6.16 EE27 MLR L1 Model Parameters .................................................................. 124
Table 6.17 EE27 Stepwise MLR L2 Goodness of Fit Statistics ........................................ 125
Table 6.18 EE27 MLR L2 Stepwise ANOVA Table .......................................................... 125
Table 6.19 EE27 MLR L2 Model Parameters .................................................................. 127
Table 6.20 EE27 MLR L3 Goodness of Fit Statistics ...................................................... 127
Table 6.21 EE27 Matlab Stepwise MLR L3 Analysis Of Variance .................................................. 128
Table 6.22 EE27 Matlab Stepwise MLR L3 Model Parameters .................................................. 129
Table 6.23 EE27 Matlab Stepwise MLR L4 Goodness of Fit Statistics ....................................... 129
Table 6.24 EE27 Matlab Stepwise MLR L4 Analysis of Variance .............................................. 130
Table 6.25 EE27 Matlab Stepwise MLR L4 Model parameters ................................................... 130
Table 6.26 EE27 MLR L4 Correlation Table .................................................................................. 131
Figure 6.6 Construction Set Data: EE27 Measured and L0 MLR Model Predicted Vibration Levels ................................................................................................................................. 133
Figure 6.7 Construction Set Data: EE27 Measured and MLR L4 Model Predicted Vibration Levels ................................................................................................................................. 134
Table 6.27 EE27 LOWESS Goodness of Fit Statistics ................................................................. 134
Figure 6.8 Construction Set Data: EE27 Measured and LOWESS Model Predicted Vibration Levels ................................................................................................................................. 135
Figure 6.9 EE19 Multiple Linear Regression L0 Residual Plot .................................................... 137
Figure 6.10 EE19 Multiple Linear Regression (MLR) L4 Residual Plot ........................................ 137
Figure 6.11 EE19 Multiple Linear Regression LOWESS Model Residual Plot ........................... 138
Figure 6.12 EE19 Multiple Linear Regression L0 Residuals -v- L0 Predicted EE19 .................... 139
Figure 6.13 EE19 Multiple Linear Regression Model L4 Residuals -v- L4 Predicted EE19 ......... 139
Figure 6.14 EE19 LOWESS Residuals -v- LOWESS Predicted EE19 ........................................ 140
Table 6.28 EE19 L0 Error Term Statistics .................................................................................... 142
Figure 6.15 EE19 MLR L0 Residuals Histogram .......................................................................... 142
Figure 6.16 EE19 MLR L0 Residuals Cumulative Distributions ..................................................... 142
Table 6.29 EE19 L4 Error Term Statistics .................................................................................... 143
Figure 6.17 EE19 MLR L4 Residuals Histogram .......................................................................... 143
Figure 6.18 EE19 MLR L4 Residuals Cumulative Distributions ..................................................... 143
Table 6.30 EE19 LOWESS Error Term Statistics ........................................................................ 144
Figure 6.19 EE19 LOWESS Residuals Histogram ........................................................................ 144
Figure 6.20 EE19 LOWESS Residuals Cumulative Distributions ................................................... 144
Figure 6.21 EE19 MLR L0, L4 and LOWESS Models Autocorrelation of Error Terms ......... 145
Equation 6.4 The Durbin-Watson Statistic .................................................................................. 146
Equation 6.5 The Durbin-Watson Statistic (Alternative) ............................................................. 146
Table 6.31 EE19 Autocorrelation of Residual Vector Data ............................................................ 147
Chapter 7

Figure 7.1 TE and EE EVW Scattergrams for W060063 .................................................. 151
Figure 7.2 TE and EE EVW Scattergrams for W116119 .................................................. 152
Figure 7.3 TE and EE EVW Scattergrams for W164167 .................................................. 152
Figure 7.4 TE and EE EVW Scattergrams for W216219 .................................................. 153
Table 7.1 Pearson Correlation Table for W6063 (14/01/2008 to 11/02/2008) .................. 154
Table 7.2 Pearson Correlation Table for W116119 (02/02/2009 to 02/03/2009) ............ 154
Table 7.3 Pearson Correlation Table for W164167 (18/01/2010 to 08/02/2010) ............ 155
Table 7.4 Pearson Correlation Table for W216219 (10/01/2011 to 07/02/2011) ............ 155
Figure 7.5 TE09 Descriptive Statistics (Mean, Range, Standard Error and Standard Deviation) 2006 to 2010 ................................................................. 157
Table 7.5 Weekly Data Falling Within the Set Range for Descriptive Statistics Calculations 2006 to 2010 ................................................................. 158
Table 7.6 Old and New Descriptive Statistics Calculation Ranges and Average Value Calculation Intervals ................................................................. 159
Figure 7.6 TE09 Descriptive Statistics Comparison Data for (5 and 15 Minute Averaging Interval) W154 to W171 (26/10/2009 to 08/03/2010) ............ 160
Figure 7.7 TE09 Descriptive Statistics 15 Minute Averaging Interval with New Ranges for Inclusion of Data W154 to W233 (26/10/2009 to 16/05/2011) ............ 160

Chapter 8

Table 8.1 Goodness of Fit Statistics for TE09 from W032035 to W096099 for MLR Models L0 – L4 and the LOWESS Model ................................................................. 166
Figure 8.1 TE09 R-squared values from W032035 to W096099 for MLR L0 – L4 and LOWESS Models ................................................................. 166
Figure 8.2 TE09 Measured and MLR L0 Model Predicted Vibration for W088091 (28/07/2008 to 24/08/2008) ................................................................. 168
Figure 8.3 TE09 Measured and MLR L4 Model Predicted Vibration for W088091 (28/07/2008 to 24/08/2008) ................................................................. 168
Figure 8.4 TE09 Measured and LOWESS Model Predicted Vibration for W088091 (28/07/2008 to 24/08/2008) ................................................................. 169
Table 8.2 TE09 Goodness of Fit Statistics for W088091 (28/07/2008 to 24/08/2008) MLR Interaction Levels L0 – L4 and LOWESS Models ......................... 169
Table 8.3 Correlation Coefficient Table for TE09 Measured against TE09 MLR and LOWESS Model Predicted Values (W088091) ................................................................. 170

Table 8.4 Goodness of Fit Statistics for EE03 from W032035 to W096099 for MLR Models L0 – L4 and the LOWESS Model ................................................................. 171

Figure 8.5 EE03 R-squared values from W032035 to W096099 for MLR L0 – L4 and LOWESS Models ................................................................. 172

Figure 8.6 EE03 Measured and MLR L0 Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008) ........................................................................ 173

Figure 8.7 EE03 Measured and MLR L4 Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008) ........................................................................ 173

Figure 8.8 EE03 Measured and LOWESS Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008) ........................................................................ 174

Table 8.5 EE03 Goodness of Fit Statistics for W088091 MLR Interaction Levels L0 – L4 and LOWESS Models ................................................................. 174

Table 8.6 Correlation Coefficient Table for EE03 Measured against EE03 MLR and LOWESS Model Predicted Values (W088091) ................................................................. 175

Table 8.7 Goodness of Fit Statistics for All End Coils from W032035 to W096099 (07/07/2007 to 20/10/2008) for the Best Performing MLR Models and the LOWESS Model ................................................................. 176

Chapter 9

Table 9.1 Goodness of Fit Statistics for TE01 from W195199 to W232235 for MLR Models L0, L0FOA, L1, L1 FOA, L2 and L2 FOA as well as a LOWESS Model with FOA Data Included in Independent Variable Data ................................................................. 187

Figure 9.1 TE09 Measured and MLR (Best L0 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature W195199 (16/08/2010 to 09/09/2010) ........................................................................ 188

Figure 9.2 TE09 Measured and MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature W195199 (16/08/2010 to 09/09/2010) ........................................................................ 189

Figure 9.3 TE09 Measured and MLR (Best L2 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature W195199 (16/08/2010 to 09/09/2010) ........................................................................ 190
Table 9.2 Pearson Correlation Table for 2010 MLR and LOWESS Model
Construction/Learning Data Measured Variables .......................................................... 191
Table 9.3 Pearson Correlation Table for W208211 (10/01/2011 to 07/02/2011)
Unfiltered .................................................................................................................. 192
Table 9.4 Pearson Correlation Table for W208211.
Data filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C.......................................................... 193
Table 9.5 Goodness of Fit Statistics for W208211 (10/01/2011 to 07/02/2011) Filtered and Unfiltered.......................................................... 195
Figure 9.4 TE09 Measured and MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature (Unfiltered) W208211 (15/11/2010 to 11/12/2010) ......................................................................... 196
Figure 9.5 TE09 Measured Vibration, MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature for W208211. Data Filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C ........................................................................ 197
Table 9.6 TE09 Goodness of Fit Statistics for W208211.
Data Filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C ........................................................................ 198
Table 9.7 Goodness of Fit Statistics for All End Coils from W195199 to W232235 for the Best Performing MLR and LOWESS Models.......................................................... 199
Table 9.8 Pearson Correlation Table Excerpts from 2010 Learning Data, W200203, W208211 and W224227 for Coil Ends TE25, EE03 and TE01........................................ 201
Figure 9.6 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W195199 (August September 2010) ........................................................................ 203
Figure 9.7 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W208211 (November/December 2010) ........................................................................ 204
Figure 9.8 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W208211 (November/December 2010) Data Filtered for Average Cold Air Temperature ≥ 60°C and Ambient Air Temperature ≥ 4°C ........................................................................ 205
Appendix 1

Figure A1.1 Gas Turbine Generator: Sensor Location Exciter End End-Coil Number 3 (EE03) ........................................................................................................................................ 218
Figure A1.2 Gas Turbine Generator (GTG) EWV Monitoring System Protection Box B. ... 219
Figure A1.3 GTG PCU100 and BNC Connector Racks. ............................................................ 219
Figure A2.1 Axial and Radial Decoupling of Support Ring and Slitting of the Support Brace .......................................................................................................................... 221

Appendix 2

Figure A2.2 Radial Decoupling of Support Ring from Phase Lead Connection at 12 O’Clock Position at the Turbine End of the GT Generator .................................................... 224
Figure A2.3 Axial and Radial Decoupling of Support Ring from Support Bracket and Support Brace at the Turbine End of the GT Generator .................................................... 224
Figure A2.4 Through Holes Drilled in the Support Ring at Exciter End of the GT Generator. Tangential Blocking Inserted During High Tune of End Winding at AOH 2005 .................................................................................................................. 225
Figure A2.5 Strapping/Bandaging of Lower Layer End Coil Bars to Support Ring at the EE of the GT Generator ........................................................................................................... 226
Figure A2.6 Setting of the Radial Gap Between the End Winding Support Ring and the Support Brackets (Z-profiles) at the Exciter End of the Generator ................................. 227
Figure A2.7 Slitting of the EE Support Bracing to the Sixth Hole Following the Initial Bump Test .................................................................................................................................. 228
Figure A2.8 Slitting of the TE Support Bracing to the Fifth Hole Following the Initial Bump Test ........................................................................................................................... 229

Appendix 3

Equation A3.1 The Simple Linear Regression Equation .............................................................................. 230
Equation A3.2 The Multiple Linear Regression Equation .............................................................................. 231
Equation A3.3 The Multiple Linear Regression Equation (Vector Matrix Form) .............................................. 231
Equation A3.4 The F-statistic .......................................................................................................................... 235
Equation A3.5 The t-statistic .......................................................................................................................... 235
Equation A3.6 The Simple Linear Regression Equation (No Constant) ....................................................... 236
Equation A3.7 The Multiple Linear Regression Equation (No Constant) ....................................................... 236
Figure A3.1 Venn Diagrams of Regression..................................................................................................... 238
Equation A3.8 Prediction Error ................................................................. 241
Equation A3.9 Sum of Squared Error of Prediction ................................. 241
Equation A3.10 Mean Squared Error ...................................................... 241
Equation A3.11 Root Mean Squared Error .............................................. 242
Equation A3.12 R-squared Predicted (for Data Outside of Construction Set) .................................................. 243
Equation A3.13 R-squared (Construction Set Data) ................................ 243
Equation A3.14 MLR Confidence Intervals .............................................. 246
Equation A3.15 Standard Error of Prediction (Simple Linear Regression) .... 247
Equation A3.16 Standard Error of Prediction (Simple Linear Regression) ... 247
Equation A3.17 Variance of Prediction (Multiple Regression) ...................... 247
Equation A3.17 Variance of Prediction (Alternative) ................................. 248
Equation A3.18 Estimated Standard Error of Prediction ............................ 248

Appendix 4
Figure A4.1 Matlab Stepwise Selection GUI ............................................. 255
Figure A4.2 Microsoft Visual Basic/PI ACE Calculated Variable Project Page .... 257

Appendix 5
Table A5.1 Goodness of Fit Statistics for TE09 from W084087 to W132135 for MLR Models L0 – L4 and LOWESS ................................................................. 260
Figure A5.1 TE09 R-squared Values from W084087 to W132135 for MLR L0 – L4 and LOWESS Models ............................................................................. 261
Figure A5.2 TE09 Measured and Model Predicted Vibration Data (MLR L0 – L4 and LOWESS Models) for W128131 (27/04/2009 to 25/05/2009) ......................... 262
Table A5.2 Goodness of Fit Statistics for all End Coils from W084087 to W132135 for the Best Performing MLR Models and the LOWESS Model. ........................................... 263
Table A5.3 Goodness of Fit Statistics for TE01 from W136139 to W160163 for MLR Models L0 – L4 and LOWESS ........................................................................ 265
Figure A5.3 TE01 Measured and Predicted Vibration for W156159 (09/11/2009 to 13/12/2009) ...................................................................................... 266
Table A5.4 Goodness of Fit Statistics for TE01 W156159 (Unfiltered) for MLR Models L0 – L4 and LOWESS Models using a True setting of 2 % and 6 % with Polynomial Degree set at 2 ......................................................... 266
Table A5.5 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True setting of 2 % and 6 % with Polynomial Degree set at 2. All Data after 30/11/2009 Removed......... 267

Figure A5.4 TE01 Measured and Predicted Vibration for W156159 (09/11/2009 to 13/12/2009). Post Recalculation of Independent Variable Data (Average Stator Current) ................................................................. 267

Table A5.6 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True Setting of 2 %, 6 %, 10 % and 15 % with Polynomial Degree set at 2 and a True Setting of 10 % with Polynomial Degree set at 1. Data filtered for Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C. ................................................................. 268

Table A5.7 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True Setting of 2 %, 6 %, 10 % and 15 % with Polynomial Degree set at 2 and a True Setting of 10 % with Polynomial Degree set at 1. Data filtered for Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C. ................................................................. 268

Table A5.8 Goodness of Fit Statistics for TE01W156159 (Filtered) for LOWESS Models. True Setting of 50 % Polynomial Degree set at 2. Data Filtered for Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C................................................................. 269

Table A5.9 Goodness of Fit Statistics for all End Coils from W136139 to W160163 for the Best Performing MLR Models and the LOWESS Model. .................................................. 270

Table A5.10 Goodness of Fit Statistics for TE01 from W164167 to W176AOH10 for MLR Models L0 – L4 and LOWESS ............................................................................... 273

Figure A5.5 TE09 R-Squared Values for MLR Models L0, L1, L2, L3 & L4 as well as the LOWESS Model W164167 to W176180 (18/01/2010 to 04/05/2010 AOH 10) ................................................................................. 274

Figure A5.6 TE09 RMSE Values for MLR Models L0, L1, L2, L3 & L4 as well as the LOWESS Model W164167 to W176180 (18/01/2010 to 04/05/2010 AOH 10).............. 274

Table A5.11 Goodness of Fit Statistics for all End-Coils from W164167 to W176AOH10 for the Best Performing MLR Models and the LOWESS Model .......... 275
CHAPTER 1
PROJECT INTRODUCTION

This project was undertaken in order to monitor End Winding Vibration of a large turbo generator following the discovery of vibration related damage to its end winding support structure. The damage to this particular generator was uncovered during a scheduled maintenance outage in April of 2005. Reports of similar problems with other machines of the same type, the same age and with a similar number of accumulated equivalent operating hours were circulating through user group forums and user group meetings in the months prior to the discovery. The problem existed across the entire fleet. The scale of the resulting damage to the machines varied from the relatively minor, such as light dusting caused by relative movement of the end coils, to catastrophic failures necessitating off-site repair. The repair time for a machine following a catastrophic failure could be up to eight months depending on the extent of the damage and the proximity and availability of facilities to carry out the required remedial work. Repair of these machines necessitated some redesign of the stator end windings in order to minimise susceptibility to resonance with line frequency vibration and in particular, double line frequency vibration, as well as their harmonics by changing the natural frequency of the end winding structure. Redesign of the end winding structure can involve either an increase or decrease to the stator end winding natural frequency. The commercial implications for the operator of a machine failure due to excessive end winding vibration were great both in terms of repair costs and revenue loss. A number of attempts were made to repair the end winding structure of the generator in the eighteen months prior to the commencement of this project. None were successful in solving the problem in the long term.

At the time of commencement of this project, the selection and purchase of the end winding vibration monitoring system had already been completed. Early work involved planning and supervision, in conjunction with the Engineering Department, the Operations Department and the Electrical, Control and Instrumentation Department, the installation of those parts of the system which could be installed without a shutdown of the machine. Planning for the installation of the sensors and set-up of the system was carried out in conjunction with the generator OEM and the end winding vibration monitoring system provider. The history of the
End Winding Vibration problem at the Power Station, other instances of the same problem occurring on identical machines, as well as the installation and set-up of the End Winding Vibration Monitoring System are dealt with in Chapter 3.

On the return to service of the generator, post installation of the end winding monitoring system, the focus of the project shifted to data analysis and condition monitoring of the machine (Chapter 4). Consultation with the OEM and EWV monitoring system provider failed to provide any meaningful guidance in terms of warning and alarm limit setting. Empirical warning and alarm thresholds for end winding vibration were set based on EWV measured data. Analysis of the end winding vibration data showed that the exciter end of the machine was worst affected with peak vibration values greater than 400µm pk-pk. As recorded peak vibrations were generally considered to be quite high, a number of experiments were undertaken in order to minimise peak vibration amplitudes, with varying degrees of success. These experiments included:

- MVAr limitation,
- Increase of declared Minimum Stable Generation value to Eirgrid National Control Centre (NCC) and,
- Adjustment of generator cooling water flow rates.

A Low Tune Repair (LTR) was carried out during a scheduled outage in April of 2007, the details of which are presented in Appendix 2. Analysis of measured End Winding Vibration (EWV) data post repair and comparison with pre repair data is carried out in Chapter 5.

Once sufficient data was gathered, the development of regression based models for each of the twelve locations monitored commenced. The main aim of this project in terms of modelling is to attempt to develop accurate statistical models of the end winding behaviour of the generator. Models are created for each of the twelve measured end coils individually. It was envisaged that analysis of model predicted data against measured vibration data could allow the analyst to detect and act upon unusual end winding vibration behaviour in a timelier manner than by analysis of measured data alone.

Chapter 6 deals with the development of Regression Models using various methods of calculation and variable selection. The amount, type and quality of independent variables made available for selection and the degree to which the input data is filtered has a major
bearing on the predictive capability of the models used. Once a suitable model for prediction of end winding vibration for a particular point of measurement is developed it will be implemented on the PI database as a new calculated variable and will be automatically calculated at regular intervals. The OSIsoft PI system, as implemented at the power station, consists of two servers with identical databases which have stored to them, data related to all aspects of plant operation, including most data used for plant operation and control systems. Data stored to the PI system is available plant wide for general use via an add-on to Excel and to view online on configurable trends over the plant intranet using RT Webparts.

A basic statistical analysis of measured vibration data from 2007 to 2011 is presented in Chapter 7 to show that the condition of the machine has not altered greatly over the period since the Low Tune Repair was carried out.

Regression models used for prediction require recalculation under normal circumstances in order to ensure predictive capability is maintained. This is mainly due to wear and tear to the machine and its ancillary systems giving rise to minor natural changes in the relationships between variables, which in turn lead to inaccuracy in model prediction. However, other unforeseen factors can necessitate the recalculation of model coefficients.

Utilisation of the plant has changed markedly since the inception of the project, going from essentially a base load station for the first five years of operation to a heavily cycled plant since 2009. The plant Utilisation Factor has dropped from approximately 90% in 2002 to approximately 50% in 2011. Most of that drop has taken place since 2008, coinciding with a massive increase in wind capacity on the grid, the bringing on line of two major gas fired generation stations (>400 MW), one in Dublin and one in Cork as well as downward pressure on demand due to the economic downturn. As a result, the machine is now frequently instructed offline by Eirgrid NCC for a number of days at a time, or is instructed on and off within the same day. This makes the calculation of predictive regression based models for end winding vibration much more challenging, as the amount of plant operating time where non-linear relationships exist between variables used for the calculation of regression models is greatly increased. Therefore where the regression models are to be used in assessing the condition of the generator end windings data related to transient operating conditions must be removed. The application of new data to the Multiple Linear Regression (MLR) and
LOWESS (Locally Weighted Regression and Scatterplot Smoothing) Models, their performance and the problems encountered are discussed in Chapter 8.

Due to poor performance of models for certain measured vibration signals, attempts were made to optimise predictive capability mainly by the introduction of additional independent/explanatory variables and more effective filtration of input data both for model generation and for calculation of predicted values using that model. The attempts made to create more robust MLR and LOWESS models for prediction of End Winding Vibration (EWV) are presented in Chapter 9.
CHAPTER 2

LITERATURE REVIEW

2.1 End Winding Design

The design of electrical machine end windings has developed greatly over the past eighty years or so and particularly in the past thirty to forty years with the availability of ever more powerful computers and software packages.

J.F. Calvert published the first major body of work in the area of End Winding Design in 1931 [1]. His paper covered forces between constructors in the stator slots as well as force calculations for the end windings, using the same method, based on the Biot Savart Law.

The Biot–Savart law can be stated as:

\[
B = \frac{\mu_0}{4\pi} \int_C \frac{I \, dl \times r}{|r|^3}
\]

Equation 2.1. The Biot-Savart Law

The Biot–Savart law is used to compute the resultant magnetic field \( B \) at position \( r \), generated by a steady current \( I \), evaluated over the path \( C \) the electric currents flow, where \( dl \) is a vector whose magnitude is the length of the differential element of the wire, in the direction of conventional current, and \( \mu_0 \) is the magnetic constant. The symbols in boldface denote vector quantities.

Calvert found that forces for the end winding sections were highest at the inner bend of the bottom layer coil as it begins to form the cone shaped basket, see Figure 2.1. The calculated force decreases along the end coil section moving toward the end cap. He found that the forces calculated were sufficient to cause deformation of the end winding basket and possible damage, particularly under short circuit conditions, should the structure be inadequately supported.
The paper “Forces in machine end windings” written by D. Harrington in 1952 [2] built on the work of Calvert in determining the end winding forces of a two pole generator under three phase and single phase short circuit conditions. For his method to be applicable, the angle of the conical section of the end winding basket must be twenty five degrees or less. The calculation of forces is more thorough than that of Calvert and examples are given for calculation of forces on a given coil section due to current flow in another section under a number of different conditions relating mainly to the relative angle between the conductors. The method however, as with Calvert’s, does not take into account the curvature of the end winding basket in calculating the actual distance in space between one coil and another. This would be more significant for 2-pole machines than for those with a greater number of poles, as a coil can span 180 degrees, if the coil is full pitched i.e. coil pitch is equal to the pole pitch.

Following on from Harrington’s work, D.S. Ashworth and P. Hammond (1961) again used current sheets to describe the magnetic field of the end winding [3], in a similar way to Calvert. The paper deals with the effect of leakage reactance at the end windings on the generator magnetic field. It also looks at the effect of the rotor end winding magnetic field on the stator end winding magnetic field, as well as the effect of the stator core. The paper investigates the effect on the end winding magnetic field of varying the angle of the end
winding basket. They found that they were able to calculate a fairly accurate representation of Magnetic Field Strength for the end windings.

In 1965 P.J. Lawrenson wrote a paper [4], on force calculation for electric machine end windings building on previous work using the Biot Savart law to calculate the forces. Lawrenson however, broke down the end winding coils into much smaller segments to improve the accuracy of the calculations. He also took into account the effect of varying the angle of the end winding basket, the spacing between the end coils and the number of coils.

In 1966 J.A. Tegopolous published two papers, [5] [6], on generator end winding forces. The first dealt with the determination of flux densities and the second paper dealt with the calculation of forces. These papers take account of all stator and rotor currents as well as the effects of the rotor retaining ring and stator end plate. The results were in line with those of Calvert although the method of calculation was more comprehensive. Radial forces were found to be greatest where the coil bends to form the cone shape basket as it exits the stator core and they diminish moving toward the coil end caps. Tangential/peripheral force calculations show the greatest values are in the relatively straight sections between the bend leaving/entering the stator and the bend at coil end caps. Calculated radial force values are generally greater than those acting in the tangential direction.

Tegopolous was one of the foremost engineers in the area of end winding magnetics and force calculation. He published numerous papers on these subjects and those relating to them, use of current sheets etc. He dedicated much of his career to the area and published papers on the subject right up until 1996.

D.J. Scott, S.J. Salon, and G.L. Kusic published two papers on forces in generator end windings. The first paper, published in 1981 [7], related to force calculations under steady state conditions. The second, published in 1983 [8], dealt with forces during transient conditions. The technique is based on the work of Tegopolous, but modified due to the availability of what was at the time relatively powerful computer software. The instantaneous time method used by Tegopolous was designed for the calculation of end winding forces under short circuit conditions. Steady state force calculations are, for simplicity calculated using a phasor method rather than an instantaneous time method which is used for the transient condition calculations. For steady state conditions forces are calculated for power factor at 0.9 lagging and 0.95 leading. Force distributions for both the inner and outer layer of the stator end winding structure are displayed separately for sections of the end coils close in
to the stator core. The effect of Power Factor is generally stated to be negligible in terms of forces in the Radial direction but that in the Axial and Peripheral planes loading is greater with leading power factor.

In 1989, Khan, Buckley and Brooks published a paper on force calculation for generator end windings [9], but rather than use the phasor method to calculate the steady state flux densities and force calculations as used by Salon, Scott and Kusic an instantaneous method is used for both steady state and transient conditions. For steady state conditions the quasi-3D Finite Element Model (FEM) analysis was used.

A number of excellent papers were published throughout the 1990’s focusing on the behaviour of stator end windings under steady state and transient conditions. F. Duffeau, N. Richard and N. Szylowicz used a 3D electromagnetic model of the generator to calculate end winding forces [10]. This method allowed them to use real current distributions rather than equivalent current sheet methods as used previously by Tegopolous and others. Their paper focused on the steady state operation of the machine. They present results relating to power outputs from 0.5 $P_n$ to $P_n$ with reactive power varying between $Q_n$ and -$0.8 Q_n$. The model was drawn up for a 3000rpm, 2 pole machine with 48 stator slots and a 5/6 coil pitch, as was the case for the GT Generator at this Power Station. No indication of machine size was given as all values were displayed in PU format. As with any two pole 50 Hz machine the magnetic forcing frequency is 100 Hz. Results for forces calculated in the Axial, Radial and Peripheral (tangential) planes for both the 100Hz and DC components of force showed that amplitudes were greatest in the radial plane followed by the Axial and then the peripheral planes. Radial component forces with $P$ set to $P_n$ were greatest at the point where the coil exits the slot, at the bend after the coil straight section exiting the stator which creates the cone shape of the end winding basket, and at the bend at the coil nose. Force amplitudes at the first bend ($0.4 – 0.5$ p.u.) were approximately 50 % less than at the exit from the stator slot ($0.8 – 1$ p.u.), while forces at the coil nose bends were approximately 50 % less ($0.25 – 0.3$ p.u.) than those at the first bend. The p.u. value for a given position varies depending on power factor with values in general being highest for $Q = Q_n$ and lowest for $Q = -0.8 Q_n$. For the peripheral plane 100 Hz force component behaviour was similar to that of the radial plane, with the variation that force amplitudes from the first bend in the coil to midway along the coil nose bend were greatest with $Q = -0.8 Q_n$. Also the peak force amplitude in the peripheral direction on the coil straight section leaving the stator core was approximately 0.3 p.u as opposed to 1 p.u. in the radial plane. Peak peripheral force amplitudes at the first bend were 0.225 p.u. and for the second
bend were approximately 0.21 p.u. In the axial plane, forces were virtually zero for the straight section leaving the stator core. As for the peripheral plane, force amplitudes are greatest at the first bend (0.25 – 0.37 p.u.) with peak values corresponding to $Q = -0.8 \, Q_n$. As for the straight coil section leaving the stator core force amplitudes are again essentially zero for the straight section of the coil nose nearest the bend, however axial force amplitudes increase sharply again at the tip of the coil nose (0.25 – 0.31 p.u. approximately). For generator power output levels (0.5 – 1 p.u.) and $Q$ held steady at $Q_n$, increased power output gives increased force amplitudes in the axial, radial and peripheral planes. This paper gave an interesting and useful insight into the forces exerted on the various sections of the end winding under steady state conditions at various loads and reactive power settings. As mentioned in the paper, displacement calculations can be calculated using a mechanical model of the end winding structure with force calculations from the electromagnetic model used as inputs.

K. Senske, S. Kulig, J. Hauhoff and D. Wünsch published a paper in 1997 on end winding vibration behaviour under short circuit conditions [11]. Their work involved the development of a very thorough mechanical model of the generator end winding and its support structure including models of the support straps, spacer blocks etc. as well as taking into account the friction between the various components. Rigidity tests were carried out on all components in temperatures from 25 °C up to 80 °C so that variations in natural frequency due to temperature change could be accounted for. Once all of the rigidity data was gathered it was input to the FE model. From the FE models the natural modes and frequencies of the end windings were calculated. For the generator in question, the first natural mode occurred at 77.7 Hz, the second at 84.9 Hz and the third at 103.6 Hz. It was found that the highest vibration amplitudes occurred at the coil end connections (the location where vibration is measured on the GT Generator). The machine described in this paper was a similar machine to the ST generator installed at the power station, which showed no sign of any abnormal EWV behaviour in approximately ten years of service to mid 2010. The authors make the point that in terms of excitation of natural modes of the end winding structure, the natural frequencies near to the system frequency (50 Hz) and magnetic forcing frequency (100 Hz) are especially significant. The main sources of end winding natural frequency excitation are given as:

- Electromagnetic forces due to stator and rotor currents in stationary and transient conditions at simple and double system frequency
• Excitation through the support structure/frame of the machine due to stator core vibrations at double line frequency

• Excitation through the support structure/frame of the machine due to imbalances in the rotating shaft at simple line frequency

In this case the third natural mode, (sixth node) was close to the 100 Hz magnetic forcing frequency, however it not nearly as easily excited as either the first or second natural modes which occur at 77.7 and 84.9 Hz respectively. For the GT generator it was found that the second natural mode was at approximately 100 Hz, and it being easily excited by the rotating magnetic field resulted in unacceptably high levels of vibration/deformation and the resulting damage to the support structure. Presented in the paper were force calculation data at the top bar for a three phase terminal short circuit. Measured forces were greatest in the axial and peripheral planes, however under the same conditions, calculated vibration values were greatest in the radial and peripheral planes, peak axial vibration values being roughly 50% of that calculated for the other planes. The peak amplitude at the coil ends was of the order of 1000 µm. Displacement values for normal operation of the machine were calculated at approximately 30 µm, with greatest amplitudes measured in the radial plane. The similarly sized ST generator of the same records peak to peak vibration values of the order of 50 – 80 µm, which falls in well with the results presented in this paper. This paper provided a very valuable insight into the end winding design of a machine almost identical to those in service at this power station, as well as providing some guidance on expected levels of vibration at the end coils under normal and short circuit conditions.

In 2000, Senske, Kulig and Drubel published a paper on the subject of end winding deformation during steady state and 3-phase short circuit conditions [12]. They calculated electromagnetic forces similarly to the 1997 paper and used those calculations as inputs to a mechanical FE model of the end winding structure. The authors look at the effect of temperature on the natural frequency/eigenfrequencies and amplitudes of the deformations/vibrations. This had particular relevance to this project as it was a problem for the GT Generator at both ends, that static natural frequency measurements, measured on a cold machine gave 2 mode natural frequency values in the range 110 – 120 Hz, but these reduced significantly once the machine was at design temperature, giving resonance with the magnetic forcing frequency and resultant damage to the end winding support structures. The paper includes model generated data for the natural frequencies and deformation amplitudes of the end winding structure with support brackets included and excluded for two machines.
with output capabilities of 120 MW and 170 MW. This too had particular relevance to the problems experienced at the Power Station, namely the final remedy to the natural frequency problem was, essentially, to remove the brackets partially in order to achieve a reduction in end winding static natural frequencies to approximately 90 Hz. The findings of the paper were that for the 120 MW machine with no brackets, the magnitude of the two-mode most excitable eigenfrequency was 30000 µm in the radial plane at 73.9 Hz. The addition of the brackets resulted in the amplitude of the deformation decreasing dramatically to 5000 µm approximately at 138.7 Hz. The stiffening of the end winding basket had a remarkable effect on vibration amplitudes in this case, with the expected consequence of significantly raising the 2-lobe natural frequency. For the 170 MW machine without the brackets fitted, the two-lobe eigenfrequency was at 65.1 Hz with amplitude in the radial plane of almost 16000 µm, significantly better than for the 120 MW machine. When support brackets were added, the 2-lobe eigenfrequency moved to 85.3 Hz with an amplitude of approximately 12000 µm. While the change for the 120 MW machine was more marked, with increasing temperature, the natural frequency of the end winding will drop and over time due to wear and tear could possibly become resonant with the magnetic forcing frequency. Although the radial vibration/deformation amplitude reduction for the 170 MW machine is lesser in magnitude than for the 120 MW machine, the effect of moving the natural frequency from 65.1 Hz to 85.3 Hz is a positive one, as the frequency has been moved away from a possible source of excitation at line frequency (50 Hz) to a mid-range value where excitation by the magnetic forcing frequency is highly unlikely. An experiment was carried out to measure radial and circumferential deformation of the end windings at two positions, one at the end cap and one at the mid-point of the involute section of the end coil for a 94 MVA machine. Data was gathered for the machine without any support brackets fitted. It was found that for the generator without support brackets fitted, the resonant frequency of the end winding structure occurred at 70.5 Hz at 0.8 I_r. It was also found that the resonant frequency can drop quite significantly with temperature, in the case described, from 89 Hz to 79 Hz approximately. No absolute temperature range was supplied.

Measured and Calculated deformation data were presented for the radial and circumferential planes under 3-phase short circuit conditions. For the generator with brackets fitted, only calculated data was presented. In the radial plane the measured and calculated values were similar except at two positions where measured vibration was approximately 50% greater than the calculated values. This was put down to localised issues that were not factored into the calculations. Calculation for deformation of the end winding under 3-phase short circuit
conditions with brackets installed, gave reductions in amplitude of the order of 20 % in the radial plane and approximately 10 % in the circumferential plane.

It was found that a Finite Element (FE) model could be used to determine the general deformation of the end winding structure. However a method for accurately calculating winding deformation was not possible, due to the extremely complex construction of the end winding support structure.

J.T. Park, K.C. Kim and J.I. Lee in a paper published in 2000, again used both the Biot Savart Law and a full 3D Finite Element Model to calculate electromagnetic forces in the end windings [13]. They also considered the reliability of the support ring and insulation material by calculating the stresses placed upon them and comparing those values with yield stress measurements. Also considered was the placement of the support ring on the end winding. A number of locations were considered namely 20 %, 40 %, 60 % and 80 % along the involute section (conical section) of the end winding. It was found that having the support ring as close to the end caps as possible is most advantageous. For the machine in question, having the support ring at the 20 % position is of practically no advantage over an unsupported end winding. If the support ring is positioned 80 % of the way along the involute near the coil end connections then displacement amplitudes are reduced by approximately 45 % and stress by around 12 % when compared with values for the 20 % position. The support ring for the GT generator at the power station is located at approximately the 90 % position along the involute section of the end windings to give the best possible stress and displacement/vibration reduction capability.

2.2 Online End Winding Vibration Monitoring

A number of papers have been published in the area of End Winding Vibration monitoring from the mid-eighties onwards when suitable, reliable technology was developed to monitor end winding vibration on a continuous basis.

T. Fortin and F. Duffeau published a paper on this subject in 1997 [14], which is of great significance in terms of validating the approach adopted for this project. Their vibration monitoring project did however have more instrumentation installed on the end winding structure than is the case for this project.
Their method, as with this project, was to develop a ‘fingerprint’ model of the end winding behaviour of the machine using measured data. The coefficients of the model can then be used with new measured data to predict end winding vibration levels. The measured and predicted model values can then be compared and changes in behaviour be detected at an early stage. The reference (model predicted) signal value for a certain operating condition can be taken as being the vibration signal that should be measured if the end winding structure were in good condition.

They took data from the same machine; 6 months apart, where little difference was seen in the behaviour of the measured values relative to the model predicted vibration amplitudes over the period. They also look at a model used to predict radial vibration on the support ring of a two-pole generator at its exciter end. The model is limited to output power values from 0.5 \( P_n \) to \( P_n \) and for any value of reactive power. Limitations such as this were required in some cases in order to maintain linearity of the relationship between dependent and independent variables included in the model so that predictive accuracy was maintained. This was not usually a problem as in general, enough data remains to generate an accurate model of end winding behaviour. Data is displayed for periods 2 months apart; for the first period, the predicted and measured value signals were almost identical; for the second period however it is seen that the value of the measured signal starts to deviate from the predicted signal. The machine was found to be damaged, but the authors conclude that the damage would have been more severe had it not been detected at an early stage.

In a paper published by engineers from the End Winding Vibration Monitoring System supplier in 2005, M. Cloutier and M. Houghton present case studies of fibre optic accelerometers being used to monitor end winding vibration on large generators [15]. The technology used is identical to that installed at the Power Station in this project. The first study details the use of the fibre optic accelerometers to measure vibration at the phase leads of two large generators with known susceptibility to phase lead failure. The system was installed in order that the machines would be allowed to operate through the peak demand season, unless destructive levels of vibration were recorded. This allowed the machine operator to avoid the prospect of being forced to shut down the plant for four months, at the request of the machine manufacturer, until a repair could be scheduled. Phase lead vibrations levels were monitored over the period and were found to increase steadily however they did not breach the safe limit set by the manufacturer until after the peak demand season.
The second study looks at the installation of fibre optic accelerometers to measure radial plane vibration on the end windings of two 156 MVA machines. As with the installation at the Power Station, twelve sensors were installed per machine, six on each end measuring vibration in the radial plane only.

It was found that vibration at one of the end coils at the turbine end of one of the machines exhibited vibration levels around 250 – 300 % higher than others at the same end of the machine. Consultation with the manufacturer revealed that the coil in question had previously been repaired and would now require ongoing monitoring to ensure that the machine would not be damaged due to excessive vibration at this location.

The paper simply proves the worth of having such a system installed where a known machine flaw exists or where repair works have been carried out previously in order to prevent or minimise damage caused by excessive vibration.
CHAPTER 3

GAS TURBINE GENERATOR END WINDING VIBRATION

3.1 The Power Station

The construction and commissioning of the Power Station was completed in the early 2000s. The station is a natural gas fired Combined Cycle Gas Turbine (CCGT) with Heat Recovery Steam Generator (HRSG) designed to operate in base load or two shifting/cycling modes. The station can also be run on fuel oil at reduced load if required. The station has an overall output capacity of 438 MVA or around 380 MW of active power.

3.2 The Gas Turbine Generator

The GT Generator is a 288 MVA, 2-pole, Air Cooled Synchronous Machine with a rated voltage of 15.75 kV. The stator windings are wound in a two layer lap formation with two coils per winding and eight turns per coil. The stator has 48 slots. The rotor has 16 slots per pole with rated field current for rated output of 1314 ADC at 372 VDC.

The Generator is used as a Synchronous Motor on start-up (to approximately 37 Hz) driven by a Static Frequency Converter (SFC) for torque and speed control.

Figure 3.1 shows a dissected view of the generator. The end windings and their support structure are clearly visible. The configuration of the terminal connections is shown also with electrical connection to the generator made at the top on the Turbine/Drive End.

Figure 3.1 is taken from OEM Generator Manuals 10/11MK [16].
End Winding Vibration (EWV) became an issue for the Gas Turbine Generator after approximately three years of operation (from approximately 25000 EOH). The underlying cause of the excessive vibration was found to be global mode natural frequencies of the end winding structures being at or close to the 100 Hz magnetic forcing frequency of the generator. This resulted in damage to the machine and other similar machines operating globally, the details of which are presented later in this chapter.

Efforts have been made over a number of years to rectify and manage the End Winding Vibration (EWV) problem both by the Original Equipment Manufacturer (OEM) and the company operating the machine. The OEM attempted to solve the resonance problem by increasing global mode natural frequency of the end winding baskets, in 2005, by the addition

---

**Figure 3.1 The Gas Turbine Generator**
of tangential blocking at the coil noses and resin treatment of the entire end winding structure and in 2006, by resin treating the end windings at both end of the machine Having failed to solve the problem through increasing the natural frequency of the end windings it was recommended that a Low Tune/Detune repair be carried out. A ‘Low Tune’ reduces the frequency of the second mode natural frequency of the end windings below the 100 Hz magnetic forcing frequency of the machine. The low tune repair of the generator and its effects on the behaviour of the machine in terms of end winding vibration are dealt with in detail in Appendix 3 and Chapter 5 respectively.

In order to mitigate the risk of catastrophic failure of the generator while repair work was being scheduled, the OEM imposed limitations on the operation of the machine. These conditions although necessary, were not commercially acceptable. An online end winding vibration monitoring system was installed with data from that system provided to the OEM in exchange for relaxation of the imposed running limitations on the machine. The details of the installation of the End Winding Vibration Monitoring System are presented in Section 3.7 and Appendix 1.

3.4 Definition of Natural Frequency

There are numerous definitions of natural frequency and the definition can vary depending upon the field of science the definition is applied in. Physicists define it as the natural mode of vibration of an object in free space.

From an engineering point of view the natural frequency of a system can be defined as, the frequency a system will ‘naturally’ adopt (depending on its structure) when suitably excited. This is also known as normal mode vibration.

If a structure is subjected to forced vibration, and the frequency of that forced vibration is the same as, or close to, that of the natural frequency of the structure, then resonance is set up. Resonant vibration can be an extremely destructive force and must be taken into account in the design of any structure.
3.5 Static Natural Frequency Testing

A Static Natural Frequency Test or Bump Test is used to determine the static natural frequency of a body. The test is carried out in order to determine whether the individual windings (local), or the structure as a whole (global), has a natural frequency at, or close to, a naturally occurring source of excitation within the machine e.g. the 50 Hz turning frequency/electrical line frequency or the 100 Hz magnetic forcing frequency. By far the dominant force at play in the machine is the 100 Hz magnetic forcing frequency. If the local or global mode natural frequencies are found to be at or close to 100 Hz it may require repair and/or redesign work to be carried out to the generator.

A Bump Test is carried out using a calibrated hammer, a FFT (Fast Fourier Transform) analyser, and a single or tri-axial accelerometer. The accelerometer is placed on the component of interest; the structure is then struck with the calibrated hammer which inputs a square wave onto the system (a square wave being made up of an infinite number of sine waves of different frequencies). The component on which the accelerometer is mounted will then vibrate at its natural frequency (excited by a wave of the same frequency). The accelerometer will capture the response of the component, and feed it into the FFT analyser where the Frequency Response Function is calculated.

When conducting a bump test on a generator end winding the OEM recommended that data be collected from at least 12 points on each end. To get a really accurate picture of the whole structure’s behaviour it would require the engineer to measure frequency response at all end bars/coil noses (48 in the case of the GT Generator) in all three planes as well as measuring frequency response at the support brackets (18 in the case of the GT Generator). A bump test of this kind was carried out in September 2006 prior to the installation of the EWV monitoring system in order to aid the decision making process for the placement and plane of measurement of the fibre optic accelerometers.

It should also be noted that the dynamic natural frequencies of the individual bars as well as the structure as a whole will change once the machine is running. Information from the OEM suggests that dynamic natural frequencies are in the range of 3 - 10 Hz lower than static natural frequency values. This means that when measuring static natural frequency, the
desired result is a value well above 100 Hz (ideally in the range 120 - 130 Hz) so that there is no possibility of resonance occurring once the machine has warmed up. The fall in natural frequency arises from thermal expansion of the end basket (mainly axial expansion) during operation.

3.6 GT Generator End Winding Vibration Timeline

3.6.1 Summary of Bump Tests

Table 3.1 displays Static Natural Frequency/Bump Test results for the generator Turbine End and Exciter End End-Winding Structures from 2005, when damage to the machine due to excessive End Winding Vibration (EWV) was first observed, to 2011 where a bump test was carried out as part of the Annual Overhaul (AOH).

In Table 3.1 bump tests carried out by the Original Equipment Manufacturer are denoted by (OEM), while bump tests carried out by outside Power Generation Services Companies are denoted by (PGSC).

For each test, Global Mode 2nd mode and the most critical Local Mode (closest to 100 Hz) natural frequencies are displayed. The 2nd global mode is most likely to cause destructive levels of vibration as it is of the same shape as that of the end winding structure itself under load and in this case is close in value to the double line frequency magnetic forcing frequency.
The OEM Maintenance Division were contracted to carry out maintenance on the Gas Turbine Generator in parallel with the 2005 major outage on the Gas Turbine. An initial inspection of the generator carried out by the Operating Company’s engineers uncovered deposits of white dust (“dusting”) on the end windings at both ends of the machine. Dusting found at the Turbine End (TE) of the machine at approximately Top Dead Centre (TDC) is shown in Figure 3.2.
Inspections revealed wear on the fibreglass spacers between the stator bars as well as between the collector rings at the TE of the machine giving rise to the abrasion dust. Along with wear to fibreglass spacers, cracking of resin and breakages of fibreglass tapes used to bind the stator bars together was also evident. This dusting was due to a poor connection between the bottom (those furthest away from the rotor) stator bars and the support ring.

The underlying reason for the fretting and resultant dusting deposits was determined to be excessive vibration due to resonance caused by local and global mode natural frequencies being at or close to the 100 Hz magnetic forcing frequency of the machine (global mode in this case).

The OEM Engineering Department recommend a “High Tune” of the end winding basket and support structure. The method involves the insertion of tangential blocking near to the winding end caps, as well as resin treatment of the areas where dusting was discovered.

The OEM expedited a repair which lasted 14 days, running 24 hours a day. The repair involved:
• Removal of glass tapes with traces of wear.
• Removal of affected diamond spacers.
• Reinstallation of diamond spacers and tapes.
• Cleaning of areas between bottom layer and support rings where cracks and gaps were observed.
• Build-up of tangential wedging behind the series connections (separate wedging of bottom layer and top layer).
• Filling of gaps and cracks with epoxy resin or epoxy matt and bonding of the bottom layer and support ring.
• Curing of the installed resin components.
• Final measurement of static natural frequencies (bump test).

3.6.3 Summary of 2005 Bump Test (Post Repair)

The tangential wedging and resin treatment raised the global natural frequencies and homogenised the affected ends. The critical 2-lobe vibration at the exciter end after completion of the modification was approximately 118 Hz. The dominant 2-lobe vibration at the turbine end (series connections) after completion of the modification was measured at approximately 107 Hz. The results of the tests of the series connections did not exhibit any independent natural frequencies of the bar ends and the entire end-winding section was classified as homogeneous. The intrinsic stiffness measurements for the coil end caps or local mode natural frequencies of the cantilever like end section of the coil end on the exciter end of the machine were in the range 165 - 173 Hz, with values for the turbine end in the range 140 - 168 Hz. The greater range of natural frequency value for the turbine end is due to the phase connections to the collector rings which are made at every eighth bar and have a lower natural frequency than the other end coils. Refer to Table 3.1.

It was recommended by the Original Equipment Manufacturer (OEM) that an inspection should be performed during the next generator outage. A further inspection and bump test (measurement of natural frequencies) was recommended for the GT at the 50,000 Equivalent Operating Hours (EOH) inspection.
3.6.4 Scheduled Minor Outage April 2006 (AOH06)

Given the findings on the GT Generator in 2005 in relation to the end winding vibration issue, the Generator Operation Company requested that the OEM carry out an inspection of the GT Generator End-Windings and their support structures to confirm if the repair work carried out was successful in preventing excessive vibration. Subsequently the OEM offered to carry out a natural frequency response test (bump test), the global mode results of which are displayed in Table 3.1.

An outside Power Generation Services Company (PGSC) was hired to carry out third party Generator Inspections and Generator Stator “Bump Tests” to check the condition of the end winding baskets. The global mode static natural frequencies recorded by the PGSC are shown in Table 3.1.

Inspection of the generator showed that the repair carried out in 2005 had not fully solved the problem and that there was again the presence of white dusting and lamination damage at the Generator Drive End/Turbine End in the area of the terminal connections (similar to that shown in Figure 3.2). It was also found that the end winding support brackets “z profiles” had been subjected to stresses causing the material and welds (to the stator end plate) at two of the eighteen brackets to fail. The damage occurred on z-profiles 6 and 7, refer to Figures 3.3 and 3.4. Figure 3.4 is taken from the 2006 OEM Consolidated Outage Report [17].

The bump tests carried out both by the OEM and outside PGSC showed that the natural frequency of the end winding on the connected/ drive end had shifted downwards over the year and had at some point passed through the critical two lobe mode of twice line frequency (magnetic forcing frequency) of 100 hertz.

It is of note that all resonant frequencies recorded in 2006 were lower than those recorded by OEM engineers in 2005 post the repair to homogenise and stiffen the end winding baskets.
In consultation with the Operating Company, the OEM initiated a temporary repair procedure to the GT Generator in an effort to homogenise the end winding structure which involved
grinding out of damaged brackets, re-welding of same and (Dye Penetrant Test) DPT of repair, and all other brackets to ensure no other damage existed at the other sixteen support brackets.

Following completion of repair work, a second natural frequency test was carried out by the OEM, the results of which indicated that the 2-lobe natural frequencies of the EE and TE end-winding baskets were at 98 Hz and 108 Hz respectively. The frequency of 108 Hz occurred due to the re-stiffening of the end-winding basket following re-welding of cracked brackets.

It is known that under operation the natural frequency drops by approximately 10 Hz due mainly to thermal expansion. That means the end-winding basket would likely run at, or very close to, the magnetic forcing frequency of 100 Hz. The bump test results of the collector rings showed acceptable natural frequencies of 125 to 130 Hz.

On this occasion the OEM placed running restrictions on the machine post repair:

1. The machine was not to be run in two shifting mode, i.e. the machine was to run as a base load machine only.
2. The machine was to be limited to a maximum of six starts per six month period.
3. The machine was to be stopped at quarterly intervals in order to inspect the condition of the end winding structures.

These measures while necessary were not acceptable from the point of view of the Operating Company, as it placed a great deal of restriction on the running of the machine. The issue of running as a base load station was one that, while not ideal in terms of profit making, was acceptable. However the forced quarterly outages were a much more serious issue as they reduced plant availability and could prove very costly to the company as it takes a significant period of time to get the machine off-line, allow it to cool sufficiently in order to stop the rotor, carry out an inspection and bump test, restart the machine and get back on line again (assuming the machine is required by the TSO). Along with the issue of having to stop the machine, there is always the possibility of incurring problems at restart, which could result in further delay and the associated revenue loss.
At this point the Operating Company engineers decided to look into the possibility of sourcing an online end winding vibration monitoring system. An agreement was reached with the OEM that if a suitable system were installed and the information from it made available to the OEM that the requirement for quarterly inspections could be dropped unless measurements from the installed system rendered it necessary to keep the protective measures in place.

3.6.5 Quarterly Inspection No.1 (EWV Related) July 2006

The first forced quarterly inspection was carried out with no damage of any kind observed. The machine was deemed fit for operation until the next quarterly inspection due in September 2006.

3.6.6 Quarterly Inspection No.2 (EWV Related) September 2006

The September inspection was proposed to allow further release for operation of the generator. An initial inspection was carried out, with no findings of dusting or other damage to the end winding basket. An internal generator inspection and natural frequency response test was carried out by the OEM.

A bump test of the Generator Turbine End and Exciter End Stator End Windings was carried out. This test was conducted firstly to ascertain the condition of the end winding structures and secondly, to provide guidance as to where the accelerometers for the vibration monitoring system might best be placed, and given that they are single axis accelerometers, what plane of measurement would yield the most valuable data (radial, axial or tangential).
3.6.7 Outage for EWV Monitoring System Installation November 2006

The preliminary inspection carried out prior to any contractors gaining entry to the machine uncovered cracking of resin between the bottom layer (furthest from rotor) windings and the end winding support ring. The findings were noted and reported to the OEM. OEM Engineers indicated that some cracking of resin at this location can occur due to thermal expansion of the end coils in the axial direction. The axial expansion of the end coils gives relative movement between the end coils and the fixed support ring (fixed to support brackets which in turn are welded to the stator end plate), resulting in some cracking of the resin at the crossing points of the support ring and the end coils. During this outage the installation of the end winding vibration monitoring system took place (GT Generator only). More information on the system and its installation is provided in Appendix 1.

3.6.8 Scheduled Major Outage April 2007 (AOH07)

Given that the attempts to eliminate the end winding vibration problem up to this point had failed and with both local and global mode natural frequencies of the end winding structures (particularly at the exciter end) hovering perilously close to the 100 Hz magnetic forcing frequency, an agreement was struck between both parties (plant owner/operator and OEM) that the scheduled 30 day outage would be used to attempt another major repair to the end winding structure. This time however, instead of trying to again raise the natural frequency of the structure, a low tune of the end winding basket would be carried out in order to reduce the natural frequency below the critical 100 Hz value.

The main works carried out as part of the low tune repair were as follows:

- Axial de-coupling of support ring.
- Radial de-coupling of support ring.
- Slitting of support braces (18 places).

The Low Tune repair of the generator is described in detail in Appendix 2.
An inspection of the machine carried out prior to any works/inspections etc. being carried out by the OEM discovered that the support bracket/z-profile no.1 on the Turbine End of the machine was cracked. This damage was reported to the OEM Programme Engineer and Operating Company Management and the required repair work was added to the works schedule.

3.6.9 Forced Outage (Gas Turbine Related) October 2007

The plant was forced to shut down due to excessive combustion chamber pressure fluctuation in the Gas Turbine. As the outage was going to be long enough it was agreed that it would be prudent to carry out an inspection of the GT Generator end windings. No damage was apparent, and the machine was returned service once turbine maintenance was complete.

3.6.10 Scheduled Minor Outage April 2008

The inspection of GT Generator uncovered no damage of any kind to the end windings or their support structures. Some minor repair work was required to support strapping for accelerometer cabling.

3.7 Failures of Other Generators of the Same Type Due To Excessive End Winding Vibration

There have been a number of failures of generators of this type due to excessive end winding vibration. In at least two cases this was attributed to the natural frequency of individual stator bars (local mode) or of the end winding structure as a whole (global mode) being at or close to the magnetic forcing frequency of the machine i.e. 100 Hz or 120 Hz depending on the network being supplied.
The first known case occurred in Argentina in June 2005 where a machine failed without warning. At the time the machine had only accumulated 27280 Equivalent Operating Hours (EOH), or about 3 years of continuous operation.

Figure 3.5 is taken from a presentation given by an engineer from a South American operating company of an identical generator to that which is the subject of this project described at a CCGT User Group Meeting held in Dublin in 2006 [18].

The subsequent inspection of the machine found that there had been a fault to earth at the 8 O’clock position (Coils 7, 8 and 9) on the Exciter End End-Winding basket. Along with the damage caused to the windings themselves (in Figure 3.5) there was also extensive damage to the cooling fan blade tips as well as damage to the rotor. Rotor damage was caused mainly by splatters of molten copper and carbon from arcing at the point of the fault.

Figure 3.5 Generator Following Fault to Earth at the Exciter End of the Machine due to Excessive EWV
The OEM proposed two options for the repair of the machine:

- To conduct an on-site repair.
- Remove generator to a manufacturing facility and have the repair work carried out there. Choosing this option would mean the machine would be out of service for 10 months.

To minimise repair costs and down time for the machine owner, a local company was hired to conduct the repair work which included for the stator:

- Rewind of the stator.
- Changing of insulation type from VPI to a standard type (VPI is not an option outside of a factory setting).
- Redesign of stator end winding structure to minimise vibration.
- Cleaning of all ventilation ducts etc.
- General cleaning of machine.
- Electrical Testing etc.

Repair work carried out on the rotor included:

- The extraction of the cooling fan blades.
- Extraction of the rotor retaining rings.
- General cleaning due to the copper and carbon contamination.
- Drying of the winding.
- Electrical testing.
- Replacement of the insulation under the retaining ring.
- Reassembly of the retaining ring and testing and balancing of the rotor.
A second failure of an identical generator occurred at a plant in the Netherlands. The machine again failed due to excessive end winding vibration. The excessive vibration had caused one of the stator bars to break. The damage which is shown in Figures 3.6 and 3.7, while not as extensive as for the machine in Argentina was nonetheless significant, resulting in lengthy downtime for the plant in question. This failure occurred on 11th May 2006. In this case the machine was removed to the OEM manufacturing facility for repair.

Figure 3.6 Generator Stator End Winding following End Coil Break due to Excessive EWV

Figures 3.6 and 3.7 are taken from a presentation given by an engineer from a European operating company of an identical generator to that which is the subject of this dissertation at a CCGT User Group Meeting held in Dublin in 2006 [19].
Figure 3.7 Generator Stator Winding following End Coil Break due to Excessive EWV

3.8 The End Winding Vibration Monitoring System

As mentioned in the previous chapter, the decision by Operating Company to install an End Winding Vibration (EWV) Monitoring System was necessitated by the OEM requirement that the machine be taken off-line at quarterly intervals in order to conduct visual inspections of the end winding structures. That decision, made by the OEM Engineering Department was arrived at based on the damage to the end winding structure discovered during previous outages and also damage to/failure of other machines of the same type and age.

The EWV monitoring system was purchased from a Canadian company specialising in condition monitoring systems for large motors and generators.
3.8.1 Location of Accelerometers and Plane of Measurement

In deciding the location and plane of measurement of the Fibre Optic Accelerometers (FOA) both the generator manufacturer and the end-winding vibration monitoring system provider were consulted on a number of occasions at various face to face meetings and teleconferences.

The decision made following the various consultations was as follows:

On the Exciter End (EE) of the generator, the accelerometers would be placed on bars 03, 11, 19, 27, 35 and 43.

On the Turbine End (TE) of the generator, the accelerometers would be placed on bars 01, 09, 17, 25, 33 and 41.

All sensors measured vibration in the radial plane i.e. perpendicular to the rotor axis. The reason for this was that vibration amplitudes in the radial direction had proved much higher than those in either axial or tangential directions in the September bump test. The sensors would be placed at or near the edge of the end bars where vibration amplitudes are greatest.

The sensor configuration agreed upon resulted in accelerometers being spaced approximately 60 degrees apart in space and also, that there was a sensor on the low and high voltage end of both sections of each winding. For the turbine end the accelerometers were arranged so that they were not placed on a phase lead connection to a collector ring, but were placed on a neighbouring end coil. Phase lead connections are much bulkier than a standard end coil and so behave differently, as the added mass and connection to the collector ring result in a damping effect, giving reduced vibration amplitudes at those locations.

Figure 3.8 is a winding diagram of the Gas Turbine Generator (GTG). The green line running through slot number 35 is the Top Dead Centre (TDC) line of the stator core. The coil end numbers, as they are marked on the machine, are displayed for both the TE (Turbine End) and EE (Exciter End). This machine is configured so that the TE is the Connected End of the machine where the terminal connections are made. The terminal connections are made to the top of the generator.
From the diagram (Figure 3.8) it can be seen that at the Turbine End of the machine, the fibre optic accelerometers are one end coil removed from the phase lead connections. Accelerometers are not placed on the phase lead connections as vibration at those locations is unrepresentative of the end winding as a whole. Vibration at the phase leads is damped due to the greater mass of the end coil as well as being connected to the collector ring. It is also evident that an accelerometer is placed at the high voltage and low voltage end of both coils of each winding.

Figure 3.9 displays the arrangement of the Fibre Optic Accelerometers (FOA) with respect to the winding arrangement in the upper part of the diagram and in 2D space in the lower part of the diagram. In the upper diagram black arrows denote Turbine/Connected End FOAs while orange arrows are used for Non-Connected/Exciter End FOAs. The stator slot numbers are also displayed with the slot number displayed first followed by either a ‘T’ for Top Layer or ‘B’ for Bottom Layer designation. In the spacial (lower) diagram the FOA are shown with the viewer looking inward toward the stator core. Turbine/Connected End FOAs are shown on the outer ring with circular markers, Exciter End/Non Connected End FOAs on the inside using square markers. The colour of the circle indicates the phase (W, V or U) related to that end coil.
TWO LAYER LAP WINDING
3 PHASE CONNECTION YY
2 POLES 48 SLOTS
8 SLOTS PER POLE AND PHASE
CHORD: 20/24 SLOT PITCHES

Figure 3.8 Gas Turbine Generator: Fibre Optic Accelerometer Locations on Winding Diagram
Figure 3.9 Gas Turbine Generator: Fibre Optic Accelerometer Locations
3.8.2 End Winding Vibration Monitoring System Overview

System Overview

Fibre Optic Accelerometers x6
Location: Generator EE
Bar No’s 03,11,19,27,35 and 43.
Optical Signal (Acceleration) x6

Fibre Optic Accelerometers x6
Location: Generator TE
Bar No’s 01,09,17,25,33 and 41.
Optical Signal (Acceleration) x6

Protection Box B
Location: Underneath Generator EE
Integral Feedthrough Connector With Built-in Signal Conditioner.
Conditioners Convert Optical to Electrical. Output 100mVAC/g on a 6VDC bias.

Protection Box A
Location: Underneath Generator TE
Integral Feedthrough Connector With Built-in Signal Conditioner
Conditioners Convert Optical to Electrical. Output 100mV/g on a 6VDC bias.

PCU 100’s (x3)
Location: Main Turbine Hall
Processes up to 8 inputs per unit.
Raw, peak, peak-peak, rms or average values available
Signals Double Integrated to Convert Acceleration to Displacement.
Optional output modules with choice of 0-10V or 4-20mA
Serial RS-485 link to Server for ZOOM Look 5

BNC Connectors allow easy access to RAW data.
DH440 Can be used to analyse raw data, FFT analysis etc.

+6VDC ±4VAC bias voltage output. (50m max)

RS-485 Link

HP Proliant Server
Location: GT PCC
Running Zoom Server, Zoom Look, Zoom
Configuration, Zoom Application
Also Running Sybase SQL Anywhere 8.0.1 Database
Trend data stored to database at 15 minute intervals
(average for the previous fifteen minutes data)

Analogue Output Cards.
Allows processed vibration signals to be used elsewhere.
Choice of 0-10V or 4-20mA

Local Area Network (LAN) Connection

PC5
Location: CCR
Running Zoom Application Only (View Information Only)
Connection to Server via LAN

Figure 3.10 GTG EWV Monitoring System Overview
Figure 3.10 is a System Overview Diagram of the End Winding Vibration Monitoring System installed on the GT Generator. Six Fibre-Optic Accelerometers are assigned to each end of the Generator (Exciter End and Turbine End). The optical signals are processed in the protection boxes and output to the PCU100 Processor Units. The PCU100 Units digitise the signals and the values are forwarded to the HP Proliant Server over A Series RS-485 link. The PCU100 Units have Analogue Output Cards which can be set up to output a number of processed versions of the measured signal or the raw signal depending on the application. The Analogue Output Signal can be transmitted using 0 - 10 V or 4 - 20 mA.

The analogue signals from the accelerometers are each available at BNC connectors mounted underneath the PCU100 Units; these allow the connection of an oscilloscope or other instrument (DI 440) to monitor/record the raw signal data. A dedicated Local Area Network (LAN) is used to connect the Server to a Local PC in the Central Control Room from which the vibration data can be viewed.

### 3.8.3 Penetration of Generator for FOA Cabling

The proposed method of penetration of the generator outer casing was also approved. The method submitted for OEM approval was the use of a steel reinforced flexible conduit of 40 mm diameter which was to be run between the top of the protection box and the floor of the generator, close to the stator end plate. The conduit would be terminated at each end using proprietary couplers which would ensure no leakage of cooling air to atmosphere. The door seals for the protection boxes were also of adequate quality to prevent leakage of cooling air.

### 3.8.4 Replacement of Steel Man Way Covers with Perspex

A proposal made by the Generator Operating Company to the OEM to replace steel man way covers with 20 mm transparent Perspex sheets, to allow visual monitoring with the machine on load, was approved.
This allowed inspection of the end winding structure from the underside at two locations on each end (at approximately 5 and 7 O’clock), from the side at the 10 O’clock position on the Turbine End and 2 O’clock position on the Exciter End as well as at the terminal connections (at the 12 O’clock position) on the Turbine End of the machine.

### 3.9 Strengths and Weaknesses of the VibrosystM EWV Monitoring System

#### 3.9.1 Strengths

Data is stored to the Sybase SQL Anywhere Database. All data since the commissioning of the system is available on the dedicated server.

Once data collection has commenced it can only be stopped by shutting down Zoom Server/Look. These programs are set up to run automatically once the server is booted up. This means that while ZOOM Application is available in the Central Control Room (CCR) so that maintenance staff can view and trend information, the system cannot be shut down, unwittingly or otherwise without going specifically to the Gas Turbine Power Control Centre (GT PCC) to do so.

BNC Connectors allow the connection of the DI440 analysis instrument to take frequency domain information.

Analogue output card allows the connection of the VibrosystM system to other systems e.g. the Delphin flightrecorder. 0 - 10 V or 4 - 20 mA are options available.

Data is exportable in .zoom or .txt format

#### 3.9.2 Weaknesses

Data is stored at relatively low resolution to the database (5 - 15 minute averages). During transient operating conditions (start-up and shut-downs) these periods of averaging are far too long for the data to be representative of the system under these conditions.
Spurious data is recorded from time to time (duplication of time stamps, omission of data etc.). This causes problems when generating analysis spreadsheets where data from the PI system is also used as data arrays for the same period of time are not matched, thus requiring manual filtering of the data.

Zoom files not readily convertible from .txt to useful Excel format without lengthy manual formatting.

In Zoom Application the trend graphic does not update automatically and instead must be manually refreshed.

### 3.10 End Winding Monitoring System Improvements

#### 3.10.1 Connection to Delphin Flight Recorder System for High Resolution Data Recording

Due to deficiencies in the capabilities of the VibrosystM End Winding Vibration Monitoring System, specifically the inability to view data at intervals less than 5 minutes it was decided to use the available analogue output modules to output data to the plant flight recorder system. This system records data at intervals as short as 0.01 s allowing accurate recording of transient vibration behaviour which occurs only at start-up and shut-down.

#### 3.10.2 Connection to OSISOFT PI Data Acquisition System

This improvement was made so that the data recorded could be stored to a database which is accessible from any PC connected to the site network. Data is supplied via an OPC Server connection from the Delphin Flight Recorders. The data is available at intervals down to 1s. PI is available as an add-on to Microsoft Excel and data stored to the PI database is available at any time to the analyst. This makes the analysis of the data infinitely easier as most plant data is available on this system also, and so spreadsheets with EWV and relevant plant data over any period can be generated in a matter of minutes.
The PI package also includes PI Process Book which allows the generation of Graphics to display plant data. Graphics have been created for many of the plant systems, with one created specifically for the monitoring of Generator EWV. These live graphics are available to any PC connected to the plant network and some are permanently on display on in the Central Control Room on dedicated screens.

Another feature of the PI System is RT Webparts which allows the display of configurable live trends on a webpage. Again this is available on any PC connected to the plant network. A webpage with trends related specifically to EWV have been created and are available to view plant wide at any time.

PI ACE (Advanced Computing Engine) allows the creation of new calculated variables on the PI server. The variables are assigned PI Tags and created using Microsoft Visual Basic. This feature of the PI system is used in the calculation of predicted EWV values, using measured plant data and the regression coefficients calculated during the modelling processes discussed later in Chapter 6 of this document.
CHAPTER 4

INITIAL DATA ANALYSIS

4.1 Analysis of Trend Data

Analysis of measured End Winding Vibration (EWV) commenced almost immediately following the installation of the vibration monitoring system. Initial analysis focused on the search for a behaviour pattern or ‘fingerprint’ for the machine in terms of EWV under normal and transient conditions. Maximum values of vibration were also monitored closely for each measurement point in order to set appropriate warning and alarm thresholds.

- At most FOA measurement positions on both the Turbine End and Exciter End, End-Windings Vibration (EWV) levels peaked when the machine operated at, or close to the Minimum Stable Generation (MIN GEN) level.

- Vibration levels at, or close to, Maximum Stable Generation (MAX GEN) were much lower for most measuring points.

- Peak vibration levels (in most cases) occurred during and just after the change from Maximum to Minimum Stable Generation. Vibration levels then tended to fall off after a period and settle at a value about 2 – 10% below the peak value.

- Ambient Air Temperature had a bearing on machine vibration levels. It was noted that vibration peaks tended to be higher on days when the ambient air temperature was relatively low. As the machine is air cooled, variations in ambient air temperature mean that there are corresponding changes in generator Cold Air Temperature and the resulting Hot Air Temperatures.

Note: Maximum Stable Generation (MAX GEN) refers to the maximum output of the machine which is approximately 240 MW depending upon ambient conditions. Minimum Stable Generation (MIN GEN) refers to the minimum output of the machine which is approximately 140 MW depending upon ambient conditions.
Chapter 4

Page 65

Initial Data Analysis

Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

Figure 4.1 EE19 and EE35 Measured Vibration (Displacement), Active Power and Reactive Power
12/12/2006 to 19/12/2006

Referring to Figure 4.1, peak vibration amplitudes occur for Exciter End Bars 19 and 35, where MW output was low (close to minimum declared capability) and when PF was close to unity or leading (the main factor however, was reduction in active power output). As before heavily lagging PF coupled with high/max MW output gave much reduced vibration levels. Peak vibration levels occurred during periods of transition between Maximum Stable Generation (MAX GEN) and Minimum Stable Generation) MIN GEN. Once the transition was made vibration levels dropped off by up to 10% and settled at the new lower level. This behaviour was typical of the vibration patterns seen at each point of measurement on the end windings at the Exciter End of the machine.

Figure 4.2 displays data for Turbine End bars 17 and 41. Again, it is immediately evident that vibration levels were strongly related to generator load. However, the relationship with Power Factor was more pronounced than for the Exciter End vibration signals shown in Figure 4.1. Vibration levels reduced with heavy load and lagging power factor and increased with reduced load or power factor movement from lagging to leading. Vibration values continued to rise as power factor moved towards zero in the leading direction (more MVAr are imported from the network).
During the hottest days of summer, peak Hot Gas Temperatures leaving the core at MAX GEN could be in the region 85 to 94 °C, with temperatures at night when running at MIN GEN in the region of 60 to 70 °C.

During the coldest days of winter, peak Hot Gas Temperatures leaving the stator core at MAX GEN could be in the region 75 to 85 °C, with temperatures at night when running at MIN GEN in the region of 50 to 65 °C.

Most of the change in Stator Core Temperature between Minimum and Maximum Declared Generation Capability could be attributed to the loss or gain of heat energy in the core associated with the loading/de-loading of the generator. For example, at MIN GEN (130 MW approximately) with unity Power Factor there was approximately 4.8 kA flowing in the stator windings, whereas at MAX GEN (235 MW) with a power factor of 0.9 the current flowing in the stator was approximately 9.5 kA. The difference of approximately 5 kA in Stator Current and the related Power Loss (given off as heat) etc. meant that there was a naturally occurring temperature fluctuation between MIN GEN and MAX GEN unless core temperature was...
further controlled. However the seasonal fluctuation of MAX GEN and MIN GEN peak temperatures is due to climatic conditions of the day.

It can also be seen that with the machine operated with a steady active power output, a relationship between reactive power and end winding vibration levels was evident. It can be seen for steady active power output levels at or close to MAX GEN that if the reactive power is heavily lagging, that vibration levels tend to drop and where reactive power is close to zero that vibration levels tend to rise.

Fluctuations in voltage and frequency appeared to have little impact on vibration.
4.2 Basic Descriptive Statistics and Correlation

4.2.1 Basic Descriptive Statistics

Along with looking at the machine “fingerprint”, during the first weeks of monitoring, some basic statistical information was monitored also; basic descriptive statistics such as Maximum, Minimum, Mean, Range, Standard Deviation, and Variance figures.

The generator in normal operation must operate within its capability curve, although the Minimum Stable Generation (MSG) limit which is set by the Gas Turbine, limits the MW aspect of the curve to values greater than 120 MW approximately. Values below this lead to problems with steam temperatures and pressures as well as giving unacceptably high emissions of combustion related gasses such as Nitrous oxide (N₂O), Carbon Dioxide (CO₂) and Carbon Monoxide (CO) among others, particularly Sulphur Dioxide (SO₂) if the machine is running on fuel oil. Emissions are closely monitored by the EPA and breaches can result in severe sanctions. Between MIN GEN and MAX GEN, and maximum leading and lagging reactive power capabilities declared to the system operator, any combination of output power and reactive power is a valid running condition. As the station is instructed by the Transmission System Operator (TSO), Eirgrid, to increase and decrease power output depending on system demand and cost per MW, the ratio of actual output to generator output capability can vary substantially from week to week. This means that the data can be skewed heavily towards the upper or lower capability boundaries of the generator.

In order to have comparable data from one week to the next, the analyst must isolate a particular operating point/region within the capability boundaries and generate statistical data for that point/region only. Selecting a specific operating point leads to there being too little information to draw any reasonable conclusions from. The machine may operate at that point for a number of minutes and may not run at that exact operating point for a long period afterwards. The selection of a suitable region of the capability curve, which gives sufficient data at regular intervals to generate useful statistical information, is the preferred choice.
<table>
<thead>
<tr>
<th></th>
<th>Symbol</th>
<th>Rated (Solid Line B)</th>
<th>Dashed Line A</th>
<th>Dashed Line C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apparent Power Output</td>
<td>S</td>
<td>288 MVA</td>
<td>300 MVA</td>
<td>256 MVA</td>
</tr>
<tr>
<td>Armature Voltage</td>
<td>U</td>
<td>15.75 kV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armature Current</td>
<td>I</td>
<td>10.557 kA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>f</td>
<td>50 Hz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power Factor</td>
<td>P.F.</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Cold Gas Temperature</td>
<td>T&lt;sub&gt;K&lt;/sub&gt;</td>
<td>25°C</td>
<td>15°C</td>
<td>45°C</td>
</tr>
</tbody>
</table>

![Figure 4.3 GT Generator Reactive Capability Curve](image)

Figure 4.3 GT Generator Reactive Capability Curve
The Power Station tends to be instructed, for the most part between two general operating states. During the day, where system demand is high, the station is usually instructed to near MAX GEN, generally with lagging Reactive Power varying between 0 and 100 MVar. At night the machine tends to be operated at, or close to, MIN GEN with leading Reactive Power varying between 0 and 60 MVar.

It was decided that a data set, with boundaries set at 215 - 235 MW and 0 - 50 MVAr Lagging would be used to generate the statistical data. These boundaries are shown in Figure 4.3.

Prior to the “Low Tune” repair, data within these boundaries accounted for an average of 20 - 21% of operating time. This meant that the data was representative of the behaviour of the machine and there was sufficient data to allow meaningful statistical data to be created.

In order to isolate a particular section of the capability plot, boundaries were set for both Active Power and Reactive Power. Statistical information created can then only apply to data falling within a defined area of the graph.

The results of the analysis showed a good degree of variation from week to week. However, it is important to point out that there was no major deterioration of the figures when taken over the entire period. Some sensors showed a slight increase in average vibration levels over the period, at other locations however average vibration levels dropped slightly and some experienced very little change at all. The data shows that variance in the system, pre low tune repair, was quite high.

4.2.2 Correlation Data

Correlation data was also calculated as a means of gaining further understanding of the relationships between the vibration levels seen at the measurement positions and the various plant data. The relationship between vibration signals was also monitored. A good correlation should exist between neighbouring measurement points on each of the end winding structures. The two end winding structures can be considered separate entities and so it would not be expected that there would be any great level of correlation between measurement points on opposite ends of the machine.
Correlation Data prior to Low Tune Repair for Turbine End and Exciter End vibration signals as well as Active Power, Reactive Power, IGV Position, Apparent Power, Average Slot Temperature, Average Stator Current, Exciter Current and Ambient Air Temperature

| Chapter 4 | Page 71 | Initial Data Analysis |

**Correlation Data for Weeks W003, W010 and W017.**

gt generator correlation data for weeks w003, w010 and w017.

<table>
<thead>
<tr>
<th>Pre Low Tune</th>
<th>W003</th>
<th>12/12/2006 to 19/12/2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTG Load Range</td>
<td>135MW to 235MW</td>
<td></td>
</tr>
<tr>
<td>Pre Low Tune</td>
<td>W010</td>
<td>30/01/2007 to 02/02/2007</td>
</tr>
<tr>
<td>GTG Load Range</td>
<td>178MW to 237MW</td>
<td></td>
</tr>
<tr>
<td>Pre Low Tune</td>
<td>W017</td>
<td>20/03/2007 to 27/03/2007</td>
</tr>
<tr>
<td>GTG Load Range</td>
<td>206MW to 230MW</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre Low Tune</th>
<th>W003</th>
<th>12/12/2006 to 19/12/2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTG Load Range</td>
<td>135MW to 235MW</td>
<td></td>
</tr>
</tbody>
</table>

- **Active Power**
- **Reactive Power**
- **IGV Actual Position**
- **Apparent Power**
- **Average Slot Temperature**
- **Average Stator Current**
- **Exciter Current**
- **Ambient Air Temp**

- **EE43 (U Phase, Coil 2, LV)**
- **EE27 (W Phase, Coil 2, LV)**
- **EE11 (V Phase, Coil 1, LV)**
- **TE41 (W Phase, Coil 1, LV)**
- **TE33 (V Phase, Coil 1, HV)**
- **TE17 (W Phase, Coil 2, HV)**
- **TE09 (V Phase, Coil 2, LV)**
- **TE01 (U Phase, Coil 2, HV)**

- **GT Generator Correlation Data**
- **GTG Load Range**
- **Pre Low Tune**
- **Weeks W003, W010 and W017.**

**Figure 4.4 GT Generator Correlation Data for Weeks W003, W010 and W017.**
In each of the three tables presented in Figure 4.4 correlations between vibration signals and field inputs are displayed within the confines of the blue lines. The coefficients values displayed with an orange box denote correlation between the high and low voltage ends of each of the six winding coils.

From Figure 4.4 it is evident that there were strong correlations between vibration signals from the same end of the machine TE or EE. There were also strong correlations between vibration levels and field inputs such as Stator Core Temperature, Active Power, Reactive Power and IGV Actual Position (surrounded by the blue line). The strength of those relationships remained largely the same over time for similar running conditions.

For the three tables shown, the minimum load value varied from 135 MW for W003, where the machine was cycled from Max Gen to Min Gen over a number of days, to 206 MW for W017 where the machine is essentially run at base load 24 hours per day for the entire week.

If the strength of these relationships which were heretofore reasonably steady, suddenly weakened/strengthened then it might reasonably be assumed that the condition of the end winding has deteriorated. This conclusion could only be drawn in conjunction with other available data, most importantly trend data. Faulty measurement would also have to be ruled out as a possible cause as well as any possible spurious data which may have been used etc. Once all other options are exhausted the possibility that the integrity of the end winding support structure has deteriorated must be considered and consultation with both the OEM and others must begin in order to assess the probability of failure and schedule any work if necessary.

4.3 Measures Taken to Reduce End Winding Vibration Levels and Setting of Alarm Thresholds

Although vibration levels were seen to be relatively steady, there was some concern that peak vibration levels typically occurring at low load were excessively high. There was particular concern regarding Exciter End Bars 19 and 35 where peak vibration levels were around 400 µm pk-pk as shown in Figure 4.5. At the time of the vibration peak, ambient air temperature is close to 0 °C, the change in load from approximately 230 MW to 140 MW results in an average slot temperature drop of 21 °C (from 78 °C to 57 °C approximately). It is evident
from the trend that for the previous day (21/01/2007), ambient air temperature was 7 - 8 °C when the machine was instructed from 230 MW to 140 MW, resulting in a vibration peak of approximately 375 μm pk-pk and a minimum average slot temperature value of approximately 60 °C.

Figure 4.5 shows that for the two vibration signals shown (Exciter End End-Coil 19 and Exciter End End-Coil 35) a relationship existed between reactive power and vibration. Where load was held steady, and reactive power varied, it is evident that as reactive power output increased, vibration values decreased and vice versa. This was particularly well demonstrated on 18/01/2007 where a change in MVAr output of 40 MVAr, while power output remains steady at approximately 230 MW, resulted in a drop in measured vibration values of the order of 40 μm pk-pk, falling from around 190 μm pk-pk to 150 μm pk-pk. Vibration values then increased again sharply to approximately 200 μm pk-pk following an instruction to 0 MVAr. The increase in reactive power output also gave rise to a 6 – 7 °C increase in average slot temperature.
Prior to any measures being taken to attempt to minimise EWV amplitudes, both the OEM and the End Winding Vibration System Provider were consulted in an attempt to ascertain whether the vibration levels recorded from the GT Generator were typical of machines of this type and size or whether vibrations levels were excessively high.

Both companies were provided with the relevant data from the machine along with plant data. The response supplied from OEM engineers was:

- That they could not provide definitive alarm thresholds e.g. that any level under 350 µm was safe, 350 - 400 µm was high and above that was dangerous.

- There was no data available to them on end winding vibration levels in those particular machines and so they could not attempt to set such thresholds.

- Normal operating vibration levels would become apparent over time and allow empirical calculation of vibration alarm thresholds but that these would differ from machine to machine.

- The best indicator of a problem would be a sharp change in typical vibration levels.

The EWV Monitoring System Provider was asked to comment on the data recorded thus far which they had been provided with along with the plant data for the same period. In response the EWV Monitoring System Provider engineers said the following:

- Experience would have led them to expect peak vibration levels in the range 200 - 250 µm pk-pk.

- FOA displacement readings were abnormally high when compared with other machines of similar size.

- However, machines with higher levels of End Winding Vibration (up to 600 µm pk-pk) have not failed.
• At a plant in Arizona, an Air cooled machine (200 MW), manufactured by a French company failed with similar EWV levels to those recorded on the GT Generator.

• The EWV Monitoring System Provider had no experience of a failure where the FOA equipment was installed and therefore were not in a position to provide recommendations on limits or deteriorating levels of displacement which may indicate imminent damage to, or failure of the machine.

• The EWV Monitoring System Provider could not provide a ‘risk of failure’ analysis based on their experience.

• Electrical vibration is induced at 100 Hz with mechanical vibration induced at 50 Hz. The generator bearing’s relative vibration of 30 to 40 µm should show up as less than this on the end winding if the vibration was mechanically induced from the rotor.

Their recommendation was to acquire frequency domain spectral data which would detail the main frequency components of the vibration. Testing should be done under various load conditions, particularly when vibrations are known to be very high or very low.

Frequency domain data was recorded using a Diagnostics Instruments DI440. This meter was connected to the system via BNC connectors mounted on the front of the panel where the raw sensor signal was available. Data was collected from all 12 sensors. The frequency domain data showed that the dominant component in every case was, as expected, the double line frequency (100 Hz) component. This served to show that vibrations experienced by the end winding stator end bars were mainly electro-magnetically induced and mechanically induced vibration was, in comparison, of little consequence. This data was sent to both VibrosystM and the OEM for their consideration. The Frequency Domain Data (FFT) for Exciter End Coil-End 35 is shown in Figure 4.6.
The response from the EWV Monitoring System Provider to the FFT Data gathered was that vibration levels were generally very high and that the fact that the 100 Hz component was so dominant was of great concern. That differed from the opinion of the Operator Engineers and the OEM where the consensus was that the results were confirmation of a known problem. Also it would be expected that the 100 Hz component would dominate in that area of the machine as the main forces the end windings are subject to are double line frequency (100 Hz) excitation forces. Both the OEM and EWV Monitoring System Provider engineers agreed that any course of action that might reduce vibration levels should be explored.

4.3.1 Setting of Alarm Thresholds

Using data gathered from the first few weeks of operation, warning and alarm thresholds for the EWV Monitoring System were set as follows:
The Warning EWV limit was set at 350 µm pk-pk as peak EWV values rarely breached this threshold and if the limit were exceeded it was for a short period of time only.

The Alarm limit was set at 400 µm pk-pk. This value was chosen as it was significantly higher than the warning limit but not so much higher that it reduces the significance of the warning limit value or is unlikely to ever be triggered. At the time of setting the alarm limit value peak recorded EWV had never breached the 400 µm pk-pk threshold.

Following the consultations mentioned earlier in the chapter with the OEM and the End Winding Vibration Monitoring System Provider with regard to the setting of such limits it was decided that the limits already set should remain in place.

### 4.4 Experiments Undertaken in Order to Reduce Peak End Winding Vibration Levels

Given that vibration levels were strongly linked to active power/stator current/stator core temperature and reactive power/exciter current it was decided that there were three possible courses of action available in order to reduce peak vibration levels. These were

1. To introduce an excitation limit on the GT Generator which meant declaring at 0 MVAr leading capability to the system operator. This action was suggested based on data which showed a possible relationship between higher vibration levels and leading power factor.

2. To throttle down valves on the cooling water supply to the generator fan banks, thereby increasing the cold gas temperature and as a result, raising stator core temperature, thereby resulting in reduced levels of end winding vibration

3. To raise the MIN GEN level declared to the System Operator (Eirgrid) by about 20 % in order to limit maximum vibration levels to around 350 µm pk-pk.
4.4.1 Peak Vibration Reduction, Option No.1.

Introduction of Excitation Limit

Given that vibration levels were dependant on load, and that variation in MW output could not be avoided, the focus moved to the relationships between Reactive Power and vibration and also, Temperature and Vibration.

The relationships between measured vibration at most end coils and Reactive Power were quite strong. Figure 4.7 is a scattergram of measured vibration at end coil TE33 against Reactive Power. The data comes from the period 05/12/2006 to 14/01/2007 and is filtered for Active Power between 225 MW and 230 MW and Ambient Air Temperature greater than 4 °C. A linear trendline is shown (red), as is the equation that describes it (located at the top right hand corner of Figure 4.7). The R-Squared goodness of fit statistic (Coefficient of Determination) value is also displayed.

![Figure 4.7 Scattergram of Measured Vibration against Reactive for End Coil TE33. Generator Output between 225 MW and 230 MW and Ambient Air Temperature Greater than or Equal to 4 °C.](image)
Table 4.1 displays the R-Squared Goodness of Fit statistic (Coefficient of Determination) for measured EWV all End Coils against Reactive Power. Values are fairly uniform for the Turbine End but are much more erratic for the Exciter End.

The three coils with below average Coefficient of Determination values (EE11, EE27 and EE43) are the three Low Voltage EWV measurement signals for the Exciter End. It is not thought that this is of any great significance as the Coefficient of Determination values related to the LV EWV measured signals from the TE are in line with values for the HV End Coils.

<table>
<thead>
<tr>
<th>End Coil</th>
<th>R Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE01</td>
<td>0.4344</td>
</tr>
<tr>
<td>TE09</td>
<td>0.6275</td>
</tr>
<tr>
<td>TE17</td>
<td>0.6244</td>
</tr>
<tr>
<td>TE25</td>
<td>0.5463</td>
</tr>
<tr>
<td>TE33</td>
<td>0.6683</td>
</tr>
<tr>
<td>TE41</td>
<td>0.6057</td>
</tr>
<tr>
<td>EE03</td>
<td>0.5846</td>
</tr>
<tr>
<td>EE11</td>
<td>0.0027</td>
</tr>
<tr>
<td>EE19</td>
<td>0.7471</td>
</tr>
<tr>
<td>EE27</td>
<td>0.3363</td>
</tr>
<tr>
<td>EE35</td>
<td>0.8410</td>
</tr>
<tr>
<td>EE43</td>
<td>0.0757</td>
</tr>
</tbody>
</table>

Table 4.1 R-Squared Values of Measured Vibration against Reactive Power for all End Coils.

Generator Output between 225 MW and 230 MW and Ambient Air Temperature Greater than or Equal to 4 °C.

The data displayed in Figure 4.7 for TE33 and the coefficient of determination data tabulated in Table 4.1 for all end coils suggested that a move to 0 MVAR leading capability might prove beneficial in lowering EWV amplitudes. This decision was backed up by submissions from the EWV Monitoring System Provider and a Power Generation Services Company (PGSC) hired to review and comment on some of the EWV data.

This measure which came into effect on 18th January ’07 proved unsuccessful. The strong relationship mentioned related to the base load (MAX GEN) condition. The relationship did not hold for operation at or close to MIN GEN. When the generator ran at MSG the effect of
the drop in MW output (and the associated stator core temperature drop) far outweighed the effect any change reactive power might have had.

Figure 4.8 is a scattergram of TE33 Measured Vibration against Reactive Power from the period 05/12/2006 to 14/01/2007 for generator load between 140 MW and 145 MW with Ambient Air Temperature greater than or equal to 4 °C. From the figure it is evident that although the trendline shows a slight increase in vibration as leading reactive power increases, the increase is negligible 3 – 4 µm pk-pk. The R-squared value for the data is 0.0023, implying no statistically significant relationship between the two variables in question. Linear trend lines applied to all of the TE data show a slight increase in vibration for increased leading reactive power. R-squared values across all TE end coils are only slightly above zero indicating that there is no significant relationship between the two variables. Linear trend lines applied to all of the EE EWV data show a slight decrease in vibration for increased leading reactive power. R-squared values for vibration against reactive power across all EE end coils are only slightly above zero signifying that again, there is no significant relationship between the two variables.
Table 4.2 displays the R-Squared Goodness of Fit statistic (Coefficient of Determination) for measured EWV all End Coils against Reactive Power for Active Power values between 140 MW and 145 MW, with ambient air temperature greater than or equal to 4 °C. R-Squared values are uniformly poor across all end coils, implying little or no relationship between Reactive Power and measured vibration.

<table>
<thead>
<tr>
<th>End Coil</th>
<th>R Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE01</td>
<td>0.1860</td>
</tr>
<tr>
<td>TE09</td>
<td>0.0072</td>
</tr>
<tr>
<td>TE17</td>
<td>0.0526</td>
</tr>
<tr>
<td>TE25</td>
<td>0.2574</td>
</tr>
<tr>
<td>TE33</td>
<td>0.0023</td>
</tr>
<tr>
<td>TE41</td>
<td>0.0129</td>
</tr>
<tr>
<td>EE03</td>
<td>0.0639</td>
</tr>
<tr>
<td>EE11</td>
<td>0.0371</td>
</tr>
<tr>
<td>EE19</td>
<td>0.0598</td>
</tr>
<tr>
<td>EE27</td>
<td>0.0040</td>
</tr>
<tr>
<td>EE35</td>
<td>0.0172</td>
</tr>
<tr>
<td>EE43</td>
<td>0.0468</td>
</tr>
</tbody>
</table>

Table 4.2 R-Squared Values of Measured Vibration against Reactive Power for all End Coils.
Generator Load 135 MW to 140 MW and Ambient Air Temperature Greater than or Equal to 4 °C.

Between Maximum Stable Generation (Max Gen) and Minimum Stable Generation (Min Gen) conditions Average Stator Slot Temperature can drop by as much as 35 °C. This has a major effect on vibration levels. Peak vibration values remained at roughly the same level as had been the case previously, however the machine was now forced to run at MSG with reactive power 0 MVAr which proved to be more onerous than running with a heavily leading power factor, which would give higher stator current, a rise in temperature and thus a reduction in vibration levels.

The results of the experiment were not as expected and although there was slight improvement, at some running conditions, they did not merit retaining a restriction on operation. Normal reactive power capability was declared to Eirgrid NCC on 26th January 2006.
The station is paid a fixed sum per month for reactive power capability on both generators; with the restriction regarding excitation in effect on the GT Generator most of this revenue would be lost. The restriction, having been lifted after only 8 days ensured revenue loss was relatively minor.

### 4.4.2 Peak Vibration Reduction, Option No.2.

#### Experiments with Generator Cooling System

The purpose of these experiments was to ascertain whether or not adjustments to the Generator CCCW flow rate could maintain Generator Cold Gas Temperature at approximately 30 °C by varying the position of the four cooling water return flow butterfly valves. It was anticipated that accurate control would be difficult as flow rates do not change in a linear fashion as the valve is moved from one position to another. Each valve has nine open positions.

The CCCW system is shared with the GT bearing lubrication system. The flow rate of water around the system is approximately 280 m³/hr, 190 m³/hr being fed to the Generator and 90 m³/hr to the GT Lubrication Oil Cooler. Any adjustment in cooling water flow rate made to one system directly affects the flow rate of cooling water to the other.

The first experiment was conducted on the 18th January 2007. This experiment was adversely influenced by changes in load occurring at around the same time as valves were being closed down on the cooling system. This led to apparently positive results but these changes could not be attributed solely to the cooling system adjustments.

A second experiment was carried out on the 22nd January 2007. This experiment proved more successful as load remained steady throughout. Results were positive, although to get large changes in temperature, cooling fan banks needed to be switched off.

A second noticeable factor in this experiment was the differing responses of the TE and EE measured vibration signals. These differing response characteristics would make design of a suitable temperature controller difficult. It was decided that it would be prudent to monitor
behaviour after the low tuning of the generator end windings, and make a decision on the possible implementation of an automatic controller at that stage.

The adjustments to be made to the end winding structure could result in changes to their respective temperature characteristics that may:

1. Bring the temperature characteristics of both ends closer together, because of this the design of an automated temperature controller may have proven worthwhile as both ends would react similarly to changes in temperature.

2. The temperature characteristics of the generator diverge, in which case implementing a controller would prove almost impossible.

Any implementation of a temperature controller would have required that due regard been given to the effects on the GT lube oil system. A minimum temperature of 60 °C was maintained in the GT Lubrication Oil Reservoir by an electric heating system. High temperature control was limited by the capability of the GT Lubrication Oil Cooler (Heat Exchanger) which shared cooling water with the GT Generator.

There were protection trips associated with high GT Bearing Temperatures and high GT Generator Cold Air Temperature. Bearing temperatures in excess of 110 °C trigger a warning alarm for abnormally high bearing temperature. A GT Trip signal was triggered at 120 °C. If either of the Generator Cold Air Temperature values reaches 40 °C an alarm was triggered, with a GT Trip signal activated at 45 °C.

The bearing temperature trip limit was not crucial as cooling water would be diverted away from the generator to the lubrication system for the most part. The generator cold air temperature trip limit would be of great concern, as during periods of warm weather this value could easily be breached. The generator cold air temperature setpoint of any automated control system would be 30 °C.
4.4.3 Peak Vibration Reduction, Option No.3

Raising of Declared Gas Turbine Minimum Stable Generation Value to 160 MW (CCGT MSG 250 MW)

This measure was taken on the 23rd January 2007 in response to a peak vibration level of 409 µm pk-pk on the previous night during very cold weather conditions. It was hoped that the plant would be instructed to minimum declared MW output in the early morning of the 23rd January in order to assess the effect of the changes to the cooling system separately. This did not occur however, and so the raised MSG measure was introduced in conjunction with the changes to generator closed circuit cooling water flow.

The effect of the two measures combined, raised MIN GEN and the throttling back of cooling water valves on the generator cooling system led to a much improved situation in terms of vibration levels. Once these measures were introduced, there were no excursions over 350 µm pk-pk by any of the end bars at any time. After two weeks operation at these settings, it was decided that they should remain in place until the scheduled outage in April 2007.

As stated, any further closing down of valves would most likely have resulted in a move past the current state of convergence of the vibration behaviour, and into a state of divergence, with the TE this time having the higher levels of vibration under most operating conditions.

Any further increase of the MSG level of the station would certainly have led to problems in terms of flexibility. The station must operate at base load in so far as is possible (OEM instruction), however, if the declared MSG level were raised too high, it could have resulted in the station being instructed off-line by Eirgrid at times of minimum demand. This would have been unacceptable as the station might then have been instructed to two-shift again, which would almost certainly have resulted in a breach of the three starts per quarter restriction imposed by the OEM in April of 2006,
CHAPTER 5

ANALYSIS OF MEASURED VIBRATION DATA

POST LOW TUNE REPAIR

In this Chapter, data collected in the weeks immediately after the Low Tune Repair will be analysed, and compared with data recorded pre repair. Any changes in the EWV fingerprint of the machine will have a bearing on setting of alarm thresholds, peak vibration reduction measures and calculation of statistical models.

Vibration data gathered in the days immediately after the repair showed a marked change in the behaviour of the end windings. Peak vibration amplitudes which previously occurred during operation at, or close to minimum stable generation (MAX GEN) capability (see Figures 4.1 and 4.3), (a feature of the machine's behaviour prior to the repair) were eliminated in some cases and much reduced for the remainder see Figure 5.1.
5.1 Radar Charts of Measured Vibration Pre and Post Low Tune Repair

At the Exciter End (EE) of the Generator, peak vibration levels now occurred when load was near maximum capability of the machine rather when operating at minimum declared output capability. Peak end winding vibration levels for all locations were now below 250 µm pk-pk for all but the most onerous of operating conditions, compared with over 400 µm pk-pk prior to the repair.

The behaviour of the entire Exciter End End-Winding Structure became more uniform with all sensor positions behaving very similarly. Figures 5.2 and 5.3 are Radar Graphs of Maximum, Minimum and Average measured vibration values from before and after the Low Tune Repair.

Figure 5.2 Radar Diagram Showing Maximum, Minimum and Average Vibration Amplitudes Measured at the EE of the GTG 13/03/2007 to 20/03/2007. Pre Low Tune Repair.

Figure 5.2 covers the period 13/03/2007 to 20/03/2007 prior to the repair. Data are displayed for all FOAs at the Exciter End of the generator. From the diagram it can be seen that for end coil bars EE19, EE35 and EE43 peak vibration values were above 300 µm pk-pk, while for bars EE03 and EE27 peak values were approximately 250 µm pk-pk.
Post repair (see Figure 5.3), the behaviour of all of the end windings were very similar, with peak amplitudes across the six points of measurement varying from 156 µm pk-pk to 174 µm pk-pk approximately (previously 134 µm pk-pk to 311 µm pk-pk). The similarity of the values indicates that the end winding basket was behaving as a monolithic structure. The resin treatment of the end winding will also have contributed significantly to the reduction in vibration amplitudes.

At the Turbine End (TE) of the Generator, as with the Exciter End (TE), peak vibration levels for comparable operating conditions were significantly reduced from those recorded prior to the Low Tune Repair, (see Figures 5.4 and 5.5). The relative magnitude of the reduction in vibration amplitudes at the TE of the machine were similar to those recorded at the EE, (see Table 5.1), however the behaviour of the TE did not appear to be as homogenous as that of the EE, due mainly to below average improvement (reduction) in peak measured vibration values at bars TE01 and TE41. The greatest improvement in maximum vibration values was seen at bars TE09, TE25 and TE33. Bar TE41 remained the TE end coil with the highest peak and average vibration values (for the conditions specified) at 222 µm pk-pk approximately.
Figure 5.4 Radar Diagram Showing Maximum, Minimum and Average Vibration Amplitudes Measured at the TE of the GTG 13/03/2007 to 20/03/2007. Pre Low Tune Repair.

Figure 5.5 Radar Diagram Showing Maximum, Minimum and Average Vibration Amplitudes Measured at the TE of the GTG 05/05/2007 to 14/05/2007. Post Low Tune Repair.
### 5.2 Change in Maximum Recorded Vibration Values

Table 5.1 displays peak amplitude vibration values for all end winding vibration measurement locations prior to and post completion of the LTR at the TE and EE of the generator. The magnitude of the changes, in percentage terms, relative to the data measured prior to the repair are also displayed. The overall average change in peak vibration measurement over the entire TE and EE is also shown.

<table>
<thead>
<tr>
<th>Exciter End</th>
<th>Max Amplitude Prior To LTR</th>
<th>Max Amplitude Post LTR</th>
<th>% Change</th>
<th>Average % Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>03</td>
<td>257.73</td>
<td>155.90</td>
<td>-39.51</td>
<td>-32.55</td>
</tr>
<tr>
<td>11</td>
<td>134.21</td>
<td>159.27</td>
<td>18.67</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>311.36</td>
<td>162.20</td>
<td>-47.91</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>246.89</td>
<td>160.59</td>
<td>-34.95</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>344.62</td>
<td>173.04</td>
<td>-49.79</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>299.63</td>
<td>174.36</td>
<td>-41.81</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turbine End</th>
<th>Max Amplitude Prior To LTR</th>
<th>Max Amplitude Post LTR</th>
<th>% Change</th>
<th>Average % Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>225.35</td>
<td>185.93</td>
<td>-17.49</td>
<td>-30.57</td>
</tr>
<tr>
<td>09</td>
<td>272.23</td>
<td>141.98</td>
<td>-47.85</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>283.96</td>
<td>184.18</td>
<td>-35.14</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>189.45</td>
<td>149.16</td>
<td>-21.27</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>206.89</td>
<td>124.10</td>
<td>-40.02</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>283.37</td>
<td>221.98</td>
<td>-21.66</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 Table of Maximum Measured Vibration Values for all EE and TE Accelerometers. Data is filtered for Average Slot Temperature between 67°C and 84°C. GTG Load 166 MW to 230 MW.

For both sets of data used in creating this table, Gas Turbine Generator (GTG) load varies between 166 MW and 230 MW. The lower limit is set at 166 MW as this is the minimum GTG load over one of the periods in question. The same rationale was used to set the upper limit. As vibration levels are sensitive to temperature, upper and lower limits (67 °C and 84 °C respectively in this case) were set so that vibration amplitudes were not skewed too severely towards the upper or lower end of the scale for a given generator load.

It is apparent that the reduction in peak vibration was quite marked with the average drop at the EE about 32.5 % and at the TE 30.5 %. A notable exception is EE11 which had increased levels of vibration since the completion of the generator repair. With the exception of EE11, the reduction in peak vibration levels seen at the EE varied from 35 % to 50 % approximately. The reduction in peak vibration at the TE of the generator is in the range 17.5 % to 48 % approximately.
5.3 Correlation Tables

The Correlation Tables, shown in Figures 5.6 and 5.7, display correlation coefficients for all Turbine End (TE) and Exciter End (EE) Fibre Optic Accelerometers (FOA) as well as for Active Power, Reactive Power, IGV Position, Apparent Power, Average Stator Current, Average Slot Temperature, Exciter Current and Ambient Air Temperature.

Note: for the remainder of this document, so that time periods are made easier to reference in the text as well as in trends etc. a method of referencing is applied as follows:

Data in respect of a specific week will be prefixed with a ‘W’ followed by a three digit week number e.g. 001. W001 refers to the first seven days’ of data recorded post installation of the End Winding Vibration Monitoring System. Weeks begin on Monday at 00:00 and run to Sunday at 23:59.

Data in respect of a number of weeks presented together (four grouped together is standard throughout the document) will again be prefixed with a ‘W’ followed instead by a six digit number made up of the start week number and finish week number (inclusive). W001004 refers to data for the four week period W001 to W004.

This convention is used extensively throughout the remainder of the document in order to improve its readability, however dates and times are also provided where required to give clarity.

The Tables in Figure 5.6 are from weeks 3 (W003), 10 (W010) and 17 (W017), prior to the Low Tune Repair. The Tables from Figure 5.7 relate to weeks 24 (W024), 25 (W025) and 30 (W030), following the return to service of the generator post Low Tune Repair.

All coefficients with values between 0.5 and 0.6999 or -0.5 to -0.6999 are highlighted in green. All coefficients with values between 0.7 and 0.9999 or -0.7 to -0.9999 are highlighted in red. Correlation coefficients of vibration signals to measured plant variable data are shown within the confines of the blue lines.
Correlation Data prior to Low Tune Repair for Turbine End and Exciter End vibration signals as well as Active Power, Reactive Power, IGV Position, Apparent Power, Average Slot Temperature, Average Stator Current, Exciter Current and Ambient Air Temperature

Pre Low Tune

GTG Load Range 235MW to 235MW

<table>
<thead>
<tr>
<th>Pre Low Tune</th>
<th>GTG Load Range 235MW to 235MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>W003</td>
<td>12/20/2006 to 19/12/2006</td>
</tr>
<tr>
<td>TE01</td>
<td>TE09</td>
</tr>
<tr>
<td>1.00</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Active Power

Reactive Power

IGV Actual Position

Apparent Power

Average Slot Temperature

Average Stator Current

Exciter Current

Ambient Air Temp

Pre Low Tune

GTG Load Range 178MW to 237MW

<table>
<thead>
<tr>
<th>Pre Low Tune</th>
<th>GTG Load Range 178MW to 237MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>W000</td>
<td>30/01/2007 to 06/03/2007</td>
</tr>
<tr>
<td>TE01</td>
<td>TE09</td>
</tr>
<tr>
<td>1.00</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Active Power

Reactive Power

IGV Actual Position

Apparent Power

Average Slot Temperature

Average Stator Current

Exciter Current

Ambient Air Temp

Pre Low Tune

GTG Load Range 206MW to 230MW

<table>
<thead>
<tr>
<th>Pre Low Tune</th>
<th>GTG Load Range 206MW to 230MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>W017</td>
<td>20/03/2007 to 27/03/2007</td>
</tr>
<tr>
<td>TE01</td>
<td>TE09</td>
</tr>
<tr>
<td>1.00</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Active Power

Reactive Power

IGV Actual Position

Apparent Power

Average Slot Temperature

Average Stator Current

Exciter Current

Ambient Air Temp

Figure 5.6 Correlation Tables for Weeks 03, 10 and 17 (Pre Low Tune Repair)
### Correlation Data Prior to and Post Low Tune Repair for TE and EE Accelerometers as well as Active Power, Reactive Power, IGV Position, Apparent Power, Average Slot Temperature, Average Stator Current, Exciter Current and Ambient Air Temperature.

<table>
<thead>
<tr>
<th>Post Low Tune</th>
<th>W24</th>
<th>05/05/2007 to 14/05/2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtered for GT Generator Load between 150MW and 230MW</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post Low Tune</th>
<th>W25</th>
<th>14/05/2007 to 21/05/2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtered for GT Generator Load between 150MW and 230MW</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post Low Tune</th>
<th>W30</th>
<th>16/05/2007 to 25/05/2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtered for GT Generator Load less than 150MW, Max GT Generator Load 223MW.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 5.7 Correlation Tables for Weeks 24, 25 and 30 (Post Low Tune Repair)**

---

Chapter 5  
Page 92  
Analysis of Data Post LTR
A change in behaviour of the machine was clearly evident when the tables from W003, W010 and W017 (pre repair) were compared against those from W024, W025 and W030 (post repair). Strong relationships between vibration signals from the same end of the machine, particularly at the TE (e.g. between TE01 and TE09, TE17, TE33 and TE41 coefficient values are all above 0.7) that existed prior to the Low Tune Repair, were diminished or eliminated post repair (no coefficient values above 0.5 between TE01 and any vibration signal from the TE).

At the EE the relationships between vibration signals were diminished post repair but less so than for the TE. Prior to the LTR strong relationships existed between most plant variables and most vibration variables (for the eight plant variables shown, all EE vibration signals have correlation values greater than 0.5 for at least seven), generally the strength of these relationships were diminished post repair (particularly at EE03, EE11 and EE19).

For the TE, pre and post repair, most correlation coefficients for plant data against vibration variable data were below 0.5. The lower the value of the correlation coefficient, the weaker and less reliable the relationship between the two variables is. This can lead to difficulty in generating accurate regression models as discussed in later chapters.

For the three tables displaying correlation coefficient data post low tune repair (W024, W025 and W030) it is evident that for the conditions specified, the coefficient values remained fairly constant, allowing for slightly different ambient conditions and variation in the operating regime of the plant by Eirgrid NCC.

**5.4 Revised EWV Warning and Alarm Settings applied Post LTR**

As mentioned in Section 5.2, peak EWV values recorded at both ends of the machine were much reduced post Low Tune Repair. In response to the new ‘normal’ vibration condition of the end winding, the Warning and Alarm settings were changed.

A new High Warning threshold was set at 300 µm pk-pk with the Alarm threshold set at 350 µm pk-pk for both ends of the generator, in the days following the LTR.
A further adjustment was made to the EE Warning and Alarm limits, reducing the High Warning threshold to 250 µm pk-pk and the Alarm setpoint to 300 µm pk-pk.
CHAPTER 6

REGRESSION MODELLING POST LOW TUNE REPAIR

6.1 Introduction

This Chapter explains some of the regression model types used to model measured EWV at each of the twelve stator coil noses fitted with an accelerometer, the assumptions made in using those models and the methods of calculation.

The creation of new independent variables, the creation of various independent variable data sets such as the L0, L1 and L4 groups of independent variables used for the calculation of MLR models are described in this chapter.

Variable selection techniques such as ‘All Possible Regressions’ and ‘Stepwise’ used to create the best possible regression model from a set of independent variables ,by selecting the most statistically significant variables in explaining variation in the dependent variable and excluding the remainder, are also covered in this chapter.

Model parameters and the statistics related to them are presented and explained for some of the models.

The performance of various regression model types in predicting measured vibration (using Construction Data Only) is examined using Goodness of Fit Statistics, Analysis of Variance Tables, Graphical Assessment as well as Residual Statistics and Distribution Fitting.

6.2 Regression Model Types Employed

A number of regression model types were used as part of this project, each with varying degrees of success in terms of accurately predicting measured End Winding Vibration at the twelve stator coil noses/end turns where accelerometers had been fitted.
6.2.1 Simple Regression

Simple Regression uses a single independent variable to describe the behaviour of a data set (dependent variable/predictand). The calculation uses the least squares method.

There are some assumptions associated with the calculation of the Simple Regression Model which must be adhered to in order to achieve accurate results:

- The Error \( (e_i) \) is normally distributed
- \( \sum (e_i) = 0 \) There is no reason to believe that the error terms will tend to be positive rather than negative or vice versa.
- \( \sum (e_i^2) = \sigma^2 \), all observations have equal (constant) variances (“Homoscedasticity”)
- \( \sum (e_i e_j) = 0(i <> j) \) errors in one observation are uncorrelated with error in other observations (Correlated errors in time series are referred to as “serial correlation” or “autocorrelation”)
- Variables are measured without error (nonstochastic)

More information on Simple Linear Regression is available in Appendix 3. Information for this section and for the relevant section of Appendix 3 was taken from a number of sources including [20] and [21].

6.2.2 Multiple Linear Regression

Multiple Linear Regression (MLR) is similar in concept to Simple Linear Regression but uses multiple independent variables to explain variance in the dependent variable.

The five assumptions required for accurate prediction using Simple Linear Regression, covering specification of the error term and measurement variables, apply also to Multiple Linear Regression models.

However, as there are now multiple independent variables, two further assumptions apply:
- The number of observations must be greater than the number of coefficients to be calculated.
- No exact linear relationship can exist between any subset of explanatory variables (perfect multicolinearity).

More information on Multiple Linear Regression is available in Appendix 3. Information for this section and for the relevant section of Appendix 3 was taken from a number of sources including [20], [21] and [22].

### 6.2.3 Weighted Least Squares Regression

Weighted Least Squares Regression is a method of calculating efficient regression coefficients where there is a problem with heteroskedasticity (unequal scatter) in error variance.

This method of residualisation was used in an effort to improve the performance of regression models, particularly for some end winding positions where accurate modelling was proving difficult. The method has four major stages:

- Conduct an Ordinary Least Squares (OLS) Regression in order to obtain the residuals/errors and squared residuals/errors.
- Having now created a squared error variable, regress the squared error variable on the original independent variables using the OLS method.
- Calculate the reciprocal of the square root of the absolute squared residuals. These “weights” are then multiplied by all of the variables in the regression model.
- Finally, obtain the OLS regression of the weighted dependent variable on the weighted independent variables.

Results using this method proved no better than the method mentioned previously and so WLS was not employed beyond the early stages of the model development process. More information on Weighted Least Squares Regression is available in Appendix 3.
6.2.4 LOWESS

The LOWESS (Locally Weighted Regression and Scatterplot Smoothing) regression modeling method was originally developed by Cleveland in 1979 [23], with further work published by among others, H. G. Muller in 1987 [24], Cleveland and Devlin in 1988 [25] Cleveland, Devlin and Grosse again in 1988 [26] and Ruppert and Wand in 1994 [27]. The method operates by applying a low-order (usually 1 - 4) polynomial to a subset of the data around each point in the data set. The polynomial is calculated using explanatory variable data corresponding to the given segment. The polynomial is fitted using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The value of the regression function for the point in question is then obtained by evaluating the local polynomial using the explanatory variable values for that data point. This procedure is carried out for all data points in the data set. Many aspects of the LOWESS Regression modeling method, such as the degree of the polynomial function applied to each segment, the size of the segments of the data set used in calculating the response of each point and the weight function (Kernel), can be changed and optimized by the analyst. More information on LOWESS Regression modelling is available in Appendix 3.

6.3 Independent/Explanatory Variables used for Model Calculation

The availability, quality and relevance of suitable independent variables for use in the calculation of regression model coefficients are of paramount importance if satisfactory results, in terms of model predictive capability, are to be achieved.

6.3.1 Available Plant Data for use as Independent Variables

Available plant data that might be used as independent variable data, relevant to the generator includes the following:

- Active Power Output (MW).
- Reactive Power (MVAr).
- Stator Core Slot Temperatures (°C) measured at 9 locations.
- Gas Turbine IGV (Inlet Guide Vane) Actual Position (0- 100%).
- Exciter Current (A DC).
- Stator Currents (kA) for each phase L1, L2 and L3.
- Stator Line Voltages (kV) i.e. L1-L2, L1-L3 and L2-L3.
- Frequency (Hz).
- Apparent Power (MVA) calculated from active and reactive power.
- Ambient Air Temperature (°C).

### 6.3.2 Independent Variables used for Regression Models

The following independent variables were made available for selection

- Active Power (MW).
- Reactive Power (MVAr).
- IGV Actual Position (%).
- Apparent Power (MVA).
- *Average* Slot Temperature (°C).
- *Average* Stator Current (kA).
- Exciter Current (A DC).
- Ambient Air Temperature (°C).
- *Average* Stator Voltage (kV).
- Frequency (Hz).

Due to there being very little temperature difference between measuring points along the stator core length, it was decided to average the nine variables and create one slot temperature variable. As the nine sensors are divided into three sets of three with a set located close to each end of the stator as well as a set at the centre, consideration was given to creating three variables, the average of the three sensors measuring at each point along the generator's length. However, as again there was so little variation between the newly created variables, it was decided that the creation of a single variable, the average of all nine sensor values, would be created.
It was decided, again due to there being little difference between recorded values, that stator voltages and currents would also be averaged to give one variable for each.

6.3.3 Creation of New Explanatory Variables

Models developed using only measured plant variables (i.e. independent variable data taken directly from the plant), proved to be of reasonably good quality for some end coil vibration signals, but quite poor for the remainder, when data from outside of the construction set was applied to them. It was required therefore that a method be found to better explain variance in the dependent variable, with no availability of new relevant plant data. The application of the models to independent variable data from outside of the construction data set is dealt with in detail in Chapter 8

The $R^2$ and RMSE values for the regression models (particularly those related to the Turbine End) were acceptable for the construction data set. However, for data sets outside of the construction data sets the results proved less promising. The $R^2_{PRED}$ values were far below the $R^2$ values for the construction data set and RMSE values were much higher. Some reduction in the $R^2$ value is natural as the model is “tuned” to the construction data set. The model should however be robust enough to accurately model vibration behaviour outside of the construction data set.

There are a number of reasons why this may not have been the case, namely:

- Overfitting.
- Explanatory variable data outside of the ranges used in the calculation of the model.
- Omission of significant variable(s) (more serious).
- Inclusion of insignificant variable(s) (less serious).
- Inappropriate regression model used.

In order to create a better model the cause(s) of inaccuracies in model predictions must be identified. Once the problem is identified remedial action can be taken.
In the case of the models using a proportion of the measured input variables only, poor performance of the models was attributed to some or all of the following:

- Models were ‘tuned’ too tightly to the construction data (Overfitting).
- Poor performance of some particular models pointed to the omission of a significant input variable.
- Lack of independent variable/variables capable of explaining the variation in vibration at that location.

In order to develop better regression models, particularly in terms of robustness when faced with new data, a number of options were available, such as:

- The introduction of more relevant explanatory variables (not possible in this case).
- The creation of “new” variables using powers, interactions and transforms of the original explanatory variables.
- The use of different regression model types such as WLS or LOWESS.

As there were no more relevant field inputs available the option of simply introducing more variables to the explanatory variable set was not available.

Creating new variables from the existing set by using factors, interaction and transforms was a viable option. Explanatory variable data sets were created using the following combinations of the ten variables mentioned in Section 5.3.2

1. Interaction Level 1 i.e. the original data set giving 10 explanatory variables
2. Interaction Level 2 e.g. a*b giving 45 explanatory variables
3. Interaction Level 3 e.g. a*b*c giving 120 explanatory variables
4. Interaction Level 4 e.g. a*b*c*d giving 210 explanatory variables
5. Quadratic powers/indices e.g. a^2, b^2 etc. giving 10 extra variables
6. Cubic powers/indices e.g. a^3, b^3 etc. giving 10 extra variables
7. Quartic powers/indices a^4, b^4 etc. giving 10 extra variables
Models were developed using all of the possibilities shown. It was found however that better results were attained by having all combinations of variables available for inclusion to the same model, rather than having interactions and quadratic multiples for instance, available to separate models only. For this reason, analysis moved forward using the largest explanatory data sets i.e. those containing the original set of 10 explanatory variables with interactions and factors included as well as interactions between factors and interactions.

The following sets of variables were used in the creation of the MLR models shown later in the chapter:

**L0 (Level 0)** 10 Independent Variables mentioned in 5.3.2 available for selection.

**L1 (level 1)** 10 original independent variables  
+ a*b (Level 2) Interactions (45 additional variables)  
+ Quadratic Powers of the 10 original independent variables
This gave a total of 65 Independent variables available for selection

**L2 (Level 2)** 10 original independent variables  
+ a*b (Level 2) Interactions (45 additional variables)  
+ Quadratic Powers of the 10 original independent variables  
+ Quadratic Powers of a*b interactions (45 additional variables)  
This gave a total of 110 Independent Variables available for selection

**L3 (Level 3)** 10 original independent variables  
+ a*b (Level 2) Interactions (45 additional variables)  
+ Quadratic Powers of the 10 original independent variables  
+ a*b*c (Level 3) Interactions (120 additional variables)  
+ Cubic Powers of the 10 original independent variables  
+ Unique interactions between Quadratic Powers of the original 10 independent variables and the 10 original independent variables e.g. a*a*b, a*a*c etc. (45 additional variables)
This gave a total of 240 Independent Variables available for selection

**L4 (Level 4)** 10 original independent variables  
+ a*b (Level 2) Interactions (45 additional variables)
+ Quadratic Powers of the 10 original independent variables
+ a*b*c (Level 3) Interactions (120 additional variables)
+ Cubic Powers of the 10 original independent variables
+ a*b*c*d (Level 4) interactions (210 additional variables)
+ Quartic Powers of the 10 original independent variables
+ Unique interactions between Quadratic Powers of the original 10 independent variables and the 10 original independent variables e.g. a*a*b, a*a*c etc. (45 additional variables)
+ Unique interactions between Cubic Powers of the original 10 independent variables and the 10 original independent variables e.g. a*a*a*b, a*a*a*c etc. (45 additional variables)
+ Unique interactions between Quadratic Powers of the original 10 independent variables and Level 2 interactions e.g. a*a*b*c, a*a*b*d etc. (239 additional variables)

This gave a total of 789 Independent Variables available for selection

6.4 Variable Selection with Large Numbers of Explanatory Variables

There are issues with having large numbers of independent variables available for selection using any variable selection procedure (overfitting, inclusion of insignificant or exclusion of significant independent variables etc.), but particularly the stepwise procedures, as the chance of developing a “false” model increases with increasing number of explanatory variables.

6.4.1 The ‘All Possible Regressions Technique’ for Independent Variable Selection

For data sets using up to around 20-30 variables, the preferred selection procedure is the “best model” or “all possible regressions” selection procedure. The use of the ‘All Possible Regressions’ technique as an alternative to the Stepwise selection method is discussed by Leach in “Alternatives to Stepwise Regression” [28]. The method simply generates a model
for every possible combination of the explanatory variables and chooses the best model either in terms of $R^2$ or some measure of goodness of fit (usually either MSE or RMSE).

Generally the user is allowed to select a maximum and minimum number of variables for inclusion in the model and the analysis software will produce a “best model” result for each, e.g. if there are ten independent/explanatory variables to choose from, and the best model with between 3 and 6 explanatory variables is required, then the best models with 3, 4, 5, and 6 independent variables are displayed.

For the model using 3 explanatory variables the software will have calculated all of the possible 120 models, for the model with 6 explanatory variables will have calculated all of the possible 210 models. It is easy to see why the method is not suitable for models where the pool of explanatory variables is large, as the number of possible models rises sharply with increased numbers of explanatory variables.

The use of this method becomes too time consuming and computationally intensive with large numbers of independent variables. This method worked quite well when used with the original explanatory variables only.

For the MLR models there were a number of choices to be made in terms of explanatory variable selection methods. Given the relatively high number of variables for selection the “all possible regressions” method would prove unviable. The next obvious choice would be some form of stepwise selection procedure, either forward selection, backward elimination or true stepwise (using both forward selection and backward elimination). The preferred variable selection procedure for this project was the “true” stepwise procedure.

### 6.4.2 Regression Models using The Stepwise Procedure in Matlab

Using Matlab 7 as the analysis software, the dependent variable and selected explanatory variable data sets were used to calculate the regression model. The true stepwise function was used with the significance values for the coefficient t-statistics set at a maximum 0.05 for inclusion and a minimum of 0.1 for exclusion. These significance figures can be changed if required. For this project however they are satisfactory settings.
Note: There are three types of Stepwise Variable Selection Procedure

1. Forward Selection: Includes none of the variables at the start of the procedure and continues to add (in order of significance) variables until such time as no improvement is seen in the resulting models RMSE figure.

2. Backward Elimination: Includes all of the variables at the start of the procedure and continues to remove (in reverse order of significance) variables until such time as no improvement is seen in the resulting models RMSE figure.

3. True Stepwise: Combines the Forward Selection and Backward Elimination techniques. The procedure starts with no independent variables included in the model as per Forward Selection, and then adds and removes variables as required until such time as no improvement is seen in the resulting models RMSE figure for the addition or removal of more variables.

For this project the True Stepwise variable selection procedure was used exclusively.

The Stepwise GUI displays all steps of the stepwise procedure. The variable coefficients t-statistics and p-values are also displayed. The procedure continues taking steps (adding and removing variables) until an optimum model is arrived at, where $R^2$ is maximised, RMSE is minimised and the p-values of all included variables are <0.1. This does not however guarantee the best model when the analysis is geared toward predicting the behaviour of the dependent variable for data sets outside of the construction data set.

The stepwise procedure has a tendency to “overfit” i.e. to tune the regression model too finely to the construction set data thereby making the model almost uniquely accurate for that data set. For this project it was required that the model be robust enough to predict accurately the behaviour of the dependent variable for a long period outside of the construction data set. For this reason a model other than the optimum model chosen by the stepwise procedure could be selected by the analyst. The operation of the stepwise procedure itself lends itself to error other than overfitting and inclusion of insignificant variables. The stepwise procedure is set up firstly to include any variables with a t-stat not equal to zero and a corresponding p-value less than 0.05. Only when all of the inclusions are made, are any variables considered for
exclusion regardless of their p-values. This can lead to a less than optimal final model selection. The reason for this is that as every variable is added or removed from the model, new coefficient values, t-stats and p-values are calculated for each variable, included and excluded. If p-values for variables already included in the model are significantly above the 0.1 threshold for consideration for exclusion and are then not removed, the calculation of significances for the variables inside and outside of the models are then incorrect as there are insignificant variables included in the model itself.

In the case of regression model solutions calculated using the Stepwise Variable Selection Method, in order that the integrity of the models created using larger pools of explanatory variables be verifiable, models were created from each of the data sets L0 to L4 detailed in 5.3.3 and their performance compared using both the construction data and new input data. The model solutions calculated from the larger explanatory variable sets would be expected to outperform those calculated from a smaller pool of explanatory variables when applied to construction set data. When the models are applied to new data however this is not necessarily the case as the more complex models can prove less robust in dealing with new data for reasons already mentioned.

Information on the use of the Stepwise Procedure in Matlab for this particular project is available in Appendix 4.

### 6.5 Construction Data Sets

The initial construction data sets used four weeks’ of data recorded immediately after the machine returned to service in May 2007. Construction data sets were made up of some or all of the ten independent variables from Section 6.3.2 as well as the dependent variable/vibration signal to be modelled. However, as the machine ran at near to 100 % output for the entire time the data was not fully representative of all possible normal running conditions.

New construction data sets, using the same variables as the initial constructions data sets, were created using eight weeks’ of data following the Low Tune Repair (LTR). These data sets were a better representation of the normal operation of the generator and models created from them were used for a number of months.
When the Single Energy Market (SEM) for the entire island of Ireland came into operation on November 1st 2007, the mode of operation of the plant was changed for a number of weeks. The generator was operated in cycling/two-shifting mode i.e. the generator operated only during times of peak demand and otherwise was instructed offline. Although this mode of operation in itself was not unusual and indeed had been typical in earlier years, since the imposition of operating restrictions (mentioned in previous chapters) by the OEM in the previous year, the machine had been run almost continuously, cycling between min and max stable generation depending on demand.

The two-shifting mode of operation was not represented in the construction data set and so some models’ predictions became erratic when faced with data pertaining to those conditions (see Figure 6.1). For this reason a recalculation of the models was carried out. However, as discussed in Chapter 8, simply generating new construction data sets and models, incorporating data relating to specific operating conditions not encountered up to that point in time, does not guarantee success in accurately predicting variation in the dependent variable under those conditions. The measured variation in the dependent variable must comply with the assumptions made in using the specified model (e.g. linear relationships between variables for the MLR models). If this is not the case then the new model may not prove any more accurate than the previous version.

To generate new models, taking account of the two-shifting operation, the options available in terms of developing new construction sets and models therefrom were:

1. Create a separate batch of models to deal exclusively with the “cycling” mode of operation.
2. Create a new construction data set incorporating all data since the low tune repair. (Seven months’ worth of data amounting to some 14000+ observations per variable)
3. Create a new construction data set using the original eight weeks post repair data and some data (four weeks’ worth), from later in the year when the machine was run in “cycling” mode (i.e. two non-sequential data sets). The assumption would have to be made that there was no deterioration in machine condition in the interim. Data from the intervening months could be used to validate the model solutions calculated from the new construction data set.
In the interests of having as simple a system as possible it was decided to create a new construction data set using the original eight weeks post repair data as well as four weeks’ worth of data from when the machine was run in cycling mode. This raised the number of observations per variable from 5204 to 6943. In the next section the first Simple and MLR models are calculated using the eight week (5204 observation) data set. The Construction data sets used with stepwise selection use the larger 6943 observation construction/learning data sets.

The additional data means that winter running conditions were factored in to the calculation of the regression model. This was at least as important, if not more so, in the development of robust regression models than the inclusion of two-shifting data which was transient and therefore, for the most part, filtered from the construction data.

6.6 Regression Model Calculation

Initial Simple and MLR models were calculated using Microsoft Excel. Using only the ten plant variables listed in Section 6.3.2, results for the learning data proved satisfactory. However when faced with data outside of the construction/learning data set the quality of the predictions fell away sharply.

Inspection of the trend data shows that there is a strong relationship between active power/average stator current etc. and measured vibration levels (at most sensor positions). There are also strong correlations between variables such as Active Power, Average Stator Current and Average Slot/Core Temperature (e.g. see Tables 6.13, 6.17, 6.21 or 6.24). For this reason care must be exercised when selecting variables for inclusion in the models.

For regressions carried out on construction data where there are a large number of observations, cross correlation (between explanatory variables) is not as significant a potential problem as it would be for smaller data sets, as the amount of data available means there is enough difference between variables to allow the effect of each independent variable to be determined separately.
Where there are large numbers of explanatory variables the “all possible regressions” or “best model” selection of variables is not suitable as the number of possible models is too large to allow efficient calculation (with the software packages available). This means that the stepwise procedure was the most viable variable selection procedure.

As mentioned in Section 6.4.2, care must be taken when using the stepwise procedure. Applying the procedure to its end will yield a model equation with all coefficients within the set significance levels. However, while this final model may prove accurate against the construction data set, this may not be the case for new data applied to it. The reason for this is the stepwise procedure’s tendency to “overfit”. A model must be chosen that is accurate for the construction data but also robust enough to give a good approximation of measured vibration when presented with data from outside of the construction data set.

For this project the regression models were used to describe as accurately and as robustly as possible the variance in the vibration signals measured at twelve coil noses of the stator end windings of the GT Generator. The models were not used for hypothesis testing and so some of the model assumptions mentioned in Section 6.2.2 and Appendix 3 did not need to be applied as rigorously as would otherwise be the case.

### 6.6.1 Simple Regressions

In developing a regression model capable of predicting the measured vibration behaviour of a stator end turn during normal operation of the generator, the use of Simple Regression is too simplistic an approach. Conducting a simple regression for each explanatory variable does however give the analyst a picture of how significant an individual explanatory variable might be in explaining variance in the dependent variable. This information provides a useful background for choosing explanatory variables for the calculation of Multiple Linear Regression (MLR) models.

The results for a Simple Regression for Turbine End End-Coil Number 25 (TE25), using eight weeks’ worth of post low tune data as the construction data set, are shown in Tables 6.1 to 6.3. The model equation (Equation 6.1) is also displayed.
Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

Regression Modelling

**Goodness Of Fit Statistics**

<table>
<thead>
<tr>
<th>Source</th>
<th>Observations</th>
<th>Sum of weights</th>
<th>DF</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>MSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>5204</td>
<td>5204</td>
<td>5202</td>
<td>0.767</td>
<td>0.767</td>
<td>138.67</td>
<td>11.78</td>
</tr>
</tbody>
</table>

Table 6.1 TE25 Simple Regression Goodness of Fit Statistics

**Analysis Of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>2378429</td>
<td>2378429</td>
<td>17152</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>5202</td>
<td>721351</td>
<td>139</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5203</td>
<td>3099781</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2 TE25 Simple Regression Analysis Of Variance

**Model Parameters**

<table>
<thead>
<tr>
<th>Source</th>
<th>Value</th>
<th>Standard error</th>
<th>t</th>
<th>Pr &gt;</th>
<th>Lower bound (95%)</th>
<th>Upper bound (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>16.972</td>
<td>0.800</td>
<td>21.22</td>
<td>&lt;0.0001</td>
<td>15.40</td>
<td>18.54</td>
</tr>
<tr>
<td>Active Power (MW)</td>
<td>0.492</td>
<td>0.004</td>
<td>130.97</td>
<td>&lt;0.0001</td>
<td>0.48</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Table 6.3 TE25 Simple Regression Model Parameters

**Equation Of The Model**

TE25 = 16.9723520658666 + 0.492083303360637*Active Power (MW)

Equation 6.1 TE25 Equation of Simple Linear Regression Model

It is evident from the R² data in Table 6.1 that a large proportion of the variance in measured vibration at TE25 was explained by variance in Active Power. A figure of 11.78 for RMSE
was acceptable given that only one explanatory variable was being used and approximated to 10% of mean vibration levels for the data set.

The ANOVA table with the F-statistic value of 17152 and an associated p-value for the F-statistic <0.0001 means that the explanatory variable was indeed significant in explaining variance in the measured vibration signal at TE25.

The Model Parameters Table allows the reader to assess the significance of the individual coefficients in predicting TE25. The t-statistic values for both the intercept and the Active Power coefficients were not equal to zero and were therefore deemed to have some level of significance in explaining the variation in TE25. The p-values associated with the t-statistics were <0.0001 so the variable and intercept could be deemed significant. Indeed even if the p-value for the intercept were high, it would still have been included (in a hierarchical model such as this) as it is the simplest term available for inclusion.

The regression results for other locations were broadly similar to those presented for TE25, above. However they also suffer from the same weakness, namely that once data from outside of the construction data set was applied, the accuracy of the model reduced to unacceptable levels as measured by RMSE or $R^2_{\text{PRED}}$.

Along with Goodness of Fit Statistics, ANOVA Statistics and Model Parameter Values such as those presented in Tables 6.1 to 6.3, various graphs can be used to analyse the results of the regression. Figure 6.1 shows measured vibration at TE25 (blue) against a series calculated using a Simple Linear Regression Model. The TE25 Predicted series (red) was calculated using the Model Equation 6.1. The same independent variable data used to calculate the regression coefficient and intercept was input to the model and the TE25 Predicted series was the result.

The model was calculated using a construction data set made up of only GT Active Power as the independent variable and TE25 measured vibration as the dependent variable. The data used was from the first eight weeks of operation post LTR and was unfiltered except for any duplicated, spurious or corrupted/invalid observation data.
From Figure 6.1 it is evident that the model predictions were some way from being accurate. The Mean value of the vibration was well approximated but the majority of the variance in the measured signal was not matched by the model.

In Figure 6.2, a clear pattern can be seen in the error data corresponding to the peaks and troughs of the measured vibration signal, most of which are not mimicked by the model.

Figure 6.3 shows a histogram of the error terms for TE25. It is evident that the distribution was close to being of a normal distribution, but had a higher concentration of error terms in the -10 to +10 range than would be the case for a normally distributed data set. For that reason the distribution fails the normality tests. Adjusting the set-up of the histogram can sometimes alter the distribution enough to pass the normality test.

Figure 6.4 is a cumulative frequency distribution plot of the error term for the TE25 regression model. The measured and ideal (Normal) plots are well matched however there is some deviation from the Normal curve, particularly on the positive side (right.)
Figure 6.2 TE25 Simple Regression Residual Plot

Figure 6.3 TE25 Simple Regression Error Histogram
In order to achieve better quality models, Multiple Linear Regression models were required. MLR models allow a number of input variables to be used in order to maximise the accuracy and efficiency of the models.

**6.6.2 Multiple Linear Regression Models**

The models presented were created from the independent variables discussed in Section 6.3.2 only i.e. no interactions etc. were available for selection to these models.

It is evident from the Goodness of Fit data presented in Table 6.4 that for a Multiple Linear Regression (MLR) model using 10 explanatory variables there was a significant increase in the $R^2$ value (as will always be the case for the inclusion of more variables) as well as a corresponding reduction in the RMSE value (11.78 to 9.16 µm pk-pk) when compared to the Simple Regression model.
From the ANOVA Table (Table 6.5) the F statistic value was reasonably high at 3179 with an associated p value of <0.001, meaning that the model was statistically significant in predicting variance in TE25.

Looking at the P-values for the t-statistics in the Model Parameter Table (Table 6.6) it is evident that for this regression there were three potentially insignificant explanatory variables included, namely; IGV Actual Position, Average Stator Current and Frequency. The p-value for the Intercept was also greater than zero but this can be ignored as an intercept, being the simplest term, will always be included in the model.

### 6.6.2.1 TE25 Multiple Linear Regression

#### Goodness Of Fit Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>5204</td>
</tr>
<tr>
<td>Sum of weights</td>
<td>5204</td>
</tr>
<tr>
<td>DF</td>
<td>5193</td>
</tr>
<tr>
<td>R²</td>
<td>0.860</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.859</td>
</tr>
<tr>
<td>MSE</td>
<td>83.81</td>
</tr>
<tr>
<td>RMSE</td>
<td>9.16</td>
</tr>
</tbody>
</table>

Table 6.4 TE25 MLR Goodness of Fit Statistics

#### Analysis Of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>10</td>
<td>2664534</td>
<td>266453</td>
<td>3179</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>5193</td>
<td>435247</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5203</td>
<td>3099781</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.5 TE25 MLR Analysis of Variance
Model Parameters

<table>
<thead>
<tr>
<th>Source</th>
<th>Value</th>
<th>Standard Error</th>
<th>t</th>
<th>Pr &gt;</th>
<th>Lower bound (95%)</th>
<th>Upper bound (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>325.22</td>
<td>136.85</td>
<td>2.38</td>
<td>0.02</td>
<td>56.92</td>
<td>593.53</td>
</tr>
<tr>
<td>Active Power</td>
<td>0.97</td>
<td>0.08</td>
<td>12.13</td>
<td>&lt;0.0001</td>
<td>0.81</td>
<td>1.13</td>
</tr>
<tr>
<td>IGV Actual Position</td>
<td>0.00</td>
<td>0.01</td>
<td>0.26</td>
<td>0.80</td>
<td>-0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Average Stator Current</td>
<td>0.91</td>
<td>3.80</td>
<td>0.24</td>
<td>0.81</td>
<td>-6.54</td>
<td>8.35</td>
</tr>
<tr>
<td>Average Slot Temp</td>
<td>0.25</td>
<td>0.05</td>
<td>5.18</td>
<td>&lt;0.0001</td>
<td>0.15</td>
<td>0.34</td>
</tr>
<tr>
<td>Ambient Air Temp</td>
<td>0.42</td>
<td>0.06</td>
<td>7.18</td>
<td>&lt;0.0001</td>
<td>0.31</td>
<td>0.53</td>
</tr>
<tr>
<td>Apparent Power</td>
<td>-1.34</td>
<td>0.14</td>
<td>-9.39</td>
<td>&lt;0.0001</td>
<td>-1.62</td>
<td>-1.06</td>
</tr>
<tr>
<td>Exciter Current</td>
<td>0.15</td>
<td>0.04</td>
<td>4.39</td>
<td>&lt;0.0001</td>
<td>0.09</td>
<td>0.22</td>
</tr>
<tr>
<td>Average Stator Voltage</td>
<td>6.22</td>
<td>1.15</td>
<td>5.42</td>
<td>&lt;0.0001</td>
<td>3.97</td>
<td>8.47</td>
</tr>
<tr>
<td>Frequency</td>
<td>-7.59</td>
<td>3.04</td>
<td>-2.49</td>
<td>0.01</td>
<td>-13.55</td>
<td>-1.62</td>
</tr>
<tr>
<td>Reactive Power</td>
<td>-0.44</td>
<td>0.06</td>
<td>-6.73</td>
<td>&lt;0.0001</td>
<td>-0.56</td>
<td>-0.31</td>
</tr>
</tbody>
</table>

Table 6.6 TE25 MLR Model Parameters

Equation Of The Model

$$TE25 = 325.221203932723 + 0.969582746832633 \times \text{Active Power (MW)} + 3.57826101067264E^{-03} \times \text{IGV Actual Position} + 0.247413314469877 \times \text{Average Stator Current} + 0.419588509561296 \times \text{Average Slot Temp} + 0.419588509561296 \times \text{Ambient Air Temp} -1.33943904821299 \times \text{Apparent Power} + 0.15425301562678 \times \text{Exciter Current (A)} + 6.21965721449547 \times \text{Average Stator Voltage} - 7.5860834151923 \times \text{11MKA01CE005.XQ01} - 0.436032016910436 \times \text{Reactive Power (MVAr)}$$

Equation 6.2 TE25 Equation of MLR Model All 10 Plant Variables Included

If another MLR is carried out with the statistically insignificant variables (IGV Actual Position, Average Stator Current and Frequency) removed from the explanatory variable matrix then the model will improve in terms of efficiency and may also improve in terms of accuracy/goodness of fit. This proved to be the case (efficiency gain through lower explanatory variable count and no dis-improvement in terms of error or $R^2$) though the improvement in this particular case was negligible.
6.6.2.2 TE25 MLR Using the Seven Most Significant Measured Variables

Goodness Of Fit Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>5204</td>
</tr>
<tr>
<td>Sum of weights</td>
<td>5204</td>
</tr>
<tr>
<td>DF</td>
<td>5196</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.859</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.859</td>
</tr>
<tr>
<td>MSE</td>
<td>83.88</td>
</tr>
<tr>
<td>RMSE</td>
<td>9.16</td>
</tr>
</tbody>
</table>

Table 6.7 TE25 MLR Goodness of Fit Statistics (7 From 10 Selected Explanatory Variables)

Analysis Of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>2653962</td>
<td>380566</td>
<td>4537</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>5196</td>
<td>435819</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>5203</td>
<td>3095781</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8 TE25 MLR Analysis of Variance (7 from 10 Selected Explanatory Variables)

Model Parameters

<table>
<thead>
<tr>
<th>Source</th>
<th>Value</th>
<th>Standard error</th>
<th>t</th>
<th>Pr &gt;</th>
<th>Lower bound (95%)</th>
<th>Upper bound (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-15.71</td>
<td>2.42</td>
<td>-6.50</td>
<td>&lt; 0.0001</td>
<td>-20.45</td>
<td>-10.97</td>
</tr>
<tr>
<td>Active Power</td>
<td>0.98</td>
<td>0.08</td>
<td>12.77</td>
<td>&lt; 0.0001</td>
<td>0.83</td>
<td>1.13</td>
</tr>
<tr>
<td>Reactive Power</td>
<td>-0.43</td>
<td>0.06</td>
<td>-7.49</td>
<td>&lt; 0.0001</td>
<td>-0.55</td>
<td>-0.32</td>
</tr>
<tr>
<td>Average Slot Temp</td>
<td>0.24</td>
<td>0.05</td>
<td>5.06</td>
<td>&lt; 0.0001</td>
<td>0.15</td>
<td>0.33</td>
</tr>
<tr>
<td>Ambient Air Temp</td>
<td>0.42</td>
<td>0.06</td>
<td>7.32</td>
<td>&lt; 0.0001</td>
<td>0.31</td>
<td>0.53</td>
</tr>
<tr>
<td>Apparent Power</td>
<td>-1.30</td>
<td>0.09</td>
<td>-15.00</td>
<td>&lt; 0.0001</td>
<td>-1.48</td>
<td>-1.13</td>
</tr>
<tr>
<td>Exciter Current</td>
<td>0.15</td>
<td>0.03</td>
<td>4.86</td>
<td>&lt; 0.0001</td>
<td>0.09</td>
<td>0.21</td>
</tr>
<tr>
<td>Average Stator Voltage</td>
<td>3.80</td>
<td>0.46</td>
<td>8.22</td>
<td>&lt; 0.0001</td>
<td>2.90</td>
<td>4.71</td>
</tr>
</tbody>
</table>

Table 6.9 TE25 MLR Model Parameters (7 from 10 Selected Explanatory Variables)
Equation Of The Model

$$
TE25 = -15.7102149509884 + 0.976045267078095 \times \text{Active Power (MW)} - 0.43200697354615 \times \text{Reactive Power (MVAr)} + 0.240370032515298 \times \text{Average Slot Temp} + 0.418984857371505 \times \text{Ambient Air Temp} - 1.30483598768302 \times \text{Apparent Power} + 0.152741180856561 \times \text{Exciter Current (A)} + 0.80439208091904 \times \text{Average Stator Voltage}
$$

Equation 6.3 TE25 Equation of MLR Model (7 Most Significant Explanatory Variables Included)

The progression from Simple Regression to Multiple Linear Regression (MLR) resulted in a marked improvement in the quality of potential regression models. However, the inclusion of all available input variables for each sensor is usually not the correct action. A better model is achieved by choosing only statistically significant independent variables and omitting the remainder.

The construction data is also source of weakness for these models; the inclusion of shutdown and start-up data means that there will be some outliers in the data used to calculate the model although the effect for the example model presented is not very severe. The learning data used covered a period where ambient conditions were mild and in general the machine operated at close to full load almost continuously. Models generates subsequent to those presented in this section had filters applied to eliminate shut-down and restart data from the construction data and from any data used to calculate goodness of fit statistics or any other evaluative statistics.

6.7 Multiple Linear Regression Model Solutions Calculated Using the Stepwise Variable Selection Procedure

Regression models generated to predict measured vibration at the twelve measurement locations will all be slightly different in configuration. Variation in vibration at one location may be more sensitive to temperature than at another location etc. As a result, for each vibration signal, a number of variables will most likely be found to be statistically insignificant in predicting its variance. In order to produce the best possible representation of the dependent variable using the available explanatory variables a number of selection procedures can be used. The two main procedures used were the ‘All Possible Regressions’ technique and the Stepwise selection procedures. For reasons discussed in Section 6.4.1 the All Possible Regressions method cannot be used for regressions using a large number of
independent variables. The True Stepwise Procedure for variable selection was therefore the method of choice for this application despite its shortcomings (discussed in Section 6.4.2).

The Stepwise Variable selections should be tested for Multicolinearity, particularly for data with large numbers of independent variables. This can be done using either Matlab or XLSTAT. Usually this was carried out in XLSTAT simply because it was an add-on to Microsoft Excel and the data is presented as a new Excel Worksheet.

**Note 1:** For this section, EE27 is used as an example; the same procedures are applied for each of the twelve vibration signals.

**Note 2:** For this section, Goodness of Fit, Analysis of Variance, Model Parameter and Correlation Data is presented for the L0 and L4 models only. A plot of RMSE against Stepwise Step Number is also shown for the L0 and L4 models. Goodness of Fit and Analysis of Variance data is displayed for MLR Models L1 to L3.

**Note 3:** The references L0, L1, L2, L3 and L4 in relation to MLR Models are used widely in this document. These refer to the make-up of the construction set data used to calculate MLR models. The independent variables, both measured and calculated, available for selection at each level (L0 to L4) are described in Section 6.3.3. Models referenced L0 had only the original ten measured plant variables available for selection for the model, whereas a model referenced L4 had over 700 independent variables available for selection in order to describe variance in the dependent variable.

The results for the stepwise procedure carried out with the 10 original plant variables for Exciter End End-Coil Number 27 (EE27) MLR Level 0 Model are displayed in Tables 6.10 to 6.13.

From the Goodness of fit data shown in Table 6.10 it is evident that the $R^2$ value was satisfactory at 0.869 ($R^2$ values above 0.7 are generally considered acceptable where there is little data. However where data is plentiful, as was the case for this project, values in excess of 0.85 for the construction set data should easily be attained) but that the RMSE figure was relatively high at 8.24 indicating that the variance of the measured signal was not being predicted by the L0 MLR Model as well as it might have been.
6.7.1 EE27 Multiple Linear Regression (Stepwise Selection, Level 0)

**Goodness of Fit Statistics:**

<table>
<thead>
<tr>
<th>Source</th>
<th>Observations</th>
<th>Sum of weights</th>
<th>DF</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>MSE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6943</td>
<td>6943</td>
<td>6937</td>
<td>0.869</td>
<td>0.869</td>
<td>67.94</td>
<td>8.24</td>
</tr>
</tbody>
</table>

Table 6.10 EE27 Matlab Stepwise MLR L0 Goodness of Fit Statistics

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>3125977</td>
<td>625195</td>
<td>9202</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>6937</td>
<td>471319</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected</td>
<td>6942</td>
<td>3597296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.11 EE27 Matlab Stepwise MLR L0 Analysis Of Variance

**Model Parameters**

<table>
<thead>
<tr>
<th>Source</th>
<th>Value</th>
<th>Standard Error</th>
<th>t</th>
<th>Pr &gt;</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-42.07</td>
<td>5.29</td>
<td>-7.95</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>MW</td>
<td>-0.46</td>
<td>0.05</td>
<td>-9.10</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>MVAR</td>
<td>-0.53</td>
<td>0.03</td>
<td>-15.79</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>IGV</td>
<td>0.48</td>
<td>0.02</td>
<td>25.14</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Exciter_I</td>
<td>0.18</td>
<td>0.02</td>
<td>10.18</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Core.Temp</td>
<td>0.32</td>
<td>0.03</td>
<td>34.52</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.12 EE27 Matlab Stepwise MLR L0 Model Parameters
Table 6.13 EE27 L0 Correlation Table

From the ANOVA Table (Table 6.11), it is apparent that five variables of the ten available for selection were included in the model (Model DF = 5). The independent variables MVA, Stator Current, Ambient Air Temperature, Voltage and Frequency were deemed insignificant in explaining variance in the EE27 EWV Signal and so were excluded from the model. The F statistic value of 920 was acceptable and indicated that the model was capable of explaining a significant proportion of the variance in the measured signal.

From the Model Parameters Table (Table 6.12), it is evident that the p-values for the independent variable t-statistic values were all <0.001 indicating that they were significant in explaining the variation in the EE27 measured signal.

In the EE27 L0 Correlation Table (Table 6.13), the strength of the relationships between the independent variables included in the model, and the strength of the relationships between the independent variables and the dependent variables are displayed. It is evident that some independent variables were highly correlated e.g. MW and IGV, but none were perfectly correlated and so did not need to be excluded from the model.
Figure 6.5 is a plot of model RMSE against Matlab True Stepwise Procedure Step Number for MLR model solutions calculated from each of the construction data sets L0 to L4. In Figure 6.5 the Stepwise Procedure had run to its natural conclusion for each of the five sets of construction data. For the L0 data (blue) it is evident that from Step 7 (of 11) onward, there was very little improvement in the RMSE value of the model for the inclusion of additional independent variables. For this reason the model at Step 7 was selected. At step seven, six independent variables were included in the model namely MW, MVAr, IGV, MVA, Exciter_I and Core_Temp. The RMSE value for the model was 8.24272, the $R^2$ value was 0.868901 and the F-statistic value was 7668.37 with $p = 0$. MVA was subsequently removed as its P-value for the t-statistic was 0.3101, above the threshold of 0.1 for exclusion, giving the model presented. It is immediately evident that the removal of the insignificant independent variable had little effect on either the RMSE of the $R^2$ values, but did give an improvement in the F-statistic value of approximately 20%.
6.7.2 EE27 Multiple Linear Regression (Stepwise Selection, Level 1)

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observations</strong></td>
</tr>
<tr>
<td><strong>Sum of weights</strong></td>
</tr>
<tr>
<td><strong>DF</strong></td>
</tr>
<tr>
<td><strong>R</strong>²</td>
</tr>
<tr>
<td><strong>Adjusted R</strong>²</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
</tr>
</tbody>
</table>

Table 6.14 EE27 Matlab Stepwise MLR L1 Goodness of Fit Statistics

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>Corrected</td>
</tr>
</tbody>
</table>

Table 6.15 EE27 Matlab Stepwise MLR L1 Analysis Of Variance

From the Goodness of fit data shown in Table 6.14 it is apparent that the R² value was an improvement on the L0 model at 0.932 and that the RMSE had improved significantly from 8.24 to 5.95 indicating that the variance of the measured signal was better predicted by the L1 MLR Model than by the L0 model (over the period covered by the Construction Data).

From the ANOVA Table 6.15, it is evident that ten independent variables of the 65 available for selection were included in the model (Model DF = 10). The F statistic value of 9641 (a slight improvement on the L0 model) was acceptable and indicates that the model was capable of explaining a significant proportion of the variance in the measured signal.

Analysing the Model Parameter data in Table 6.16, it is seen that nine of the ten P-values for the t-statistic values were below the 0.05 level required for inclusion in the model. The P-value for the independent variable MW.*Stator_I of 0.082, while above the threshold for inclusion, is not above the threshold for exclusion from the model i.e. 0.1 and was therefore not removed.
Model Parameters

| Source               | Value  | Standard Error | t      | Pr > |t| |
|----------------------|--------|----------------|-------|-----|---|
| Intercept            | 310.45 | 12.43          | 24.98 | < 0.0001 |
| MW                   | -2.16  | 0.13           | -17.02| < 0.0001 |
| MVA                  | 0.48   | 0.06           | 7.47  | < 0.0001 |
| IGV                  | -1.27  | 0.09           | -13.40| < 0.0001 |
| Exciter_I            | 0.09   | 0.01           | 9.40  | < 0.0001 |
| Core_Temp            | -3.78  | 0.14           | -27.44| < 0.0001 |
| MW.*Core_Temp        | 0.02   | 0.00           | 36.21 | < 0.0001 |
| MW.*Exciter_I        | 0.00   | 0.00           | -3.71 | 0.00 |
| MW.*Stator_I         | 0.02   | 0.01           | 2.39  | 0.02 |
| MW.*IGV              | 0.01   | 0.00           | 16.21 | < 0.0001 |
| MVAR.*Exciter_I      | 0.00   | 0.00           | -3.75 | 0.00 |

Table 6.16 EE27 MLR L1 Model Parameters

If Matlab were allowed to carry out the Stepwise procedure to its end then there would be 26 steps with 23 independent variables included in the final model. The RMSE value for that model would be 5.54007, the $R^2$ figure would be 0.9340012 with a F-statistic value of 3561. However, it is unlikely that the Matlab calculated final model would prove useful when applied to data outside of the construction data due to the tendency of the procedure to overfit models to the dependent variables.

Observing the EE27 RMSE -v- Stepwise Procedure Step Number plot of Figure 6.5 (L1 values displayed as red squares) it can be seen that a point was reached where the addition of more explanatory variables to the model (approximately Step 13) resulted in minimal further reduction in RMSE. At step thirteen, twelve independent variables were included in the model namely: MW, MVAR, MVA, IGV, EXI, Core_Temp, Voltage, MW.*Core_Temp, MW.*Exciter_I, MW.*Stator_I and MVAR.*Exciter_I.

Two independent variables were included at this point despite having high P-values for their t-statistic values namely MW and voltage, at 0.3915 and 0.1296 respectively. It was decided to remove the MVA variable first as it was shown to have been insignificant for the L0 model where less independent variables were available for selection. With MVA removed, the P-value for MW reduced to 0, and so MVA was permanently eliminated. The removal of MVA had little effect on the P-value for Voltage and so it too was permanently excluded from the model having breached the exclusion threshold of 0.1.
Again, as with the L0 model, the removal of the two independent variables had little effect on the goodness of fit statistics’ values. Similar to the L0 model, there was an increase in the F-statistic value, in this case, from approximately 8116 to 9641 due mainly to the reduction in the number of explanatory variables included in the model. The MSE figure however rose slightly from 34.481 for the model with 12 variables included to 34.809 for the model with the MVA and Voltage explanatory variables removed. The effects of this increase on model performance were minimal.

### 6.7.3 EE27 Multiple Linear Regression (Stepwise Selection, Level 2)

**Goodness of Fit Statistics**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>6943</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of weights</td>
<td>6943</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td>6929</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.935</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.935</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>33.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>5.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.17 EE27 Stepwise MLR L2 Goodness of Fit Statistics

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>13</td>
<td>3365009</td>
<td>258847</td>
<td>7721</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>6929</td>
<td>232287</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6942</td>
<td>3597296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.18 EE27 MLR L2 Stepwise ANOVA Table

Level 2 models had 110 explanatory variables available for selection. For the model data shown, there were many other independent variables available for selection with t-statistic p-values less than the 0.05 value required for selection. However, considering the RMSE -v- Step Plot in Figure 6.5 for the L2 MLR Model (green triangles), it was evident that the
addition of more than approximately 20 explanatory variables to the model resulted in minimal reduction in RMSE.

If Matlab was allowed to carry out the Stepwise procedure to its end then there would have been 38 steps with 31 independent variables included in the final model. The RMSE value for that model would have been 5.47, the $R^2$ figure would have been 0.943 with a F-statistic value of 3655. However, it is unlikely that the Matlab calculated final model would prove useful when applied to data outside of the construction data due to the tendency of the procedure to overfit models to the dependent variables.

It was decided to pick a model from the mid-point of the RMSE -v- Step plot (Figure 6.5) and tune it slightly, removing any insignificant variables so that a more robust version of the model could be developed for application to data from outside of the construction data set. This involves a sacrifice in terms of performance on construction set data, which will be reduced by removing some independent variables from the model.

Initially a model with 15 variables included was chosen. The independent variable MW was removed due the P value for its t-statistic being greater than 0.1 (0.1009 in this case). The independent variable Voltage.*Voltage was removed due to multicolinearity. No other variables were removed giving the model presented a DF of 13.

The $R^2$ value was an improvement on the L1 model at 0.935 as shown in Table 6.16. The RMSE value had improved slightly from 5.95 to 5.79 indicating that the variance of the measured signal was marginally better predicted by the L2 MLR Model than by the L1 model.

From the ANOVA Table (6.18) it is evident that 13 variables of the 110 available for selection were included in the model (Model DF = 13). The F statistic value of 7721 was less than the L0 and the L1 models. This was mainly related to the increase in the number of independent variables used in the model compared to the L0 and L1 models. The improvement in the MSE value from 34.809 for the L1 model to 33.524 for the L2 model was not large enough to keep the F-statistic value at around 9000 given that the model Sum of Squares is steady at roughly 3350000.
Model Parameters

| Source           | Value  | Standard error | t     | Pr > |t| |
|------------------|--------|----------------|-------|------|------|
| Intercept        | 188.57 | 29.23          | 6.45  | <0.0001 |
| MVAR             | -2.03  | 0.08           | -25.64| <0.0001 |
| IGV              | -1.22  | 0.10           | -11.71| <0.0001 |
| MVA              | -2.51  | 0.16           | -15.49| <0.0001 |
| Exciter_I        | 0.67   | 0.02           | 28.19 | <0.0001 |
| Core_Temp        | -2.47  | 0.17           | -14.67| <0.0001 |
| Voltage          | -4.53  | 1.77           | -2.55 | 0.01 |
| MW.*Core_Temp    | 0.03   | 0.00           | 35.88 | <0.0001 |
| MW.*Exciter_I    | 0.00   | 0.00           | -28.26| <0.0001 |
| MW.*Stator_I     | -0.07  | 0.02           | -4.47 | <0.0001 |
| MW.*IGV          | 0.01   | 0.00           | 15.15 | <0.0001 |
| MW.*MVA          | 0.01   | 0.00           | 10.54 | <0.0001 |
| MVAR.*Exciter_I  | 0.00   | 0.00           | 31.08 | <0.0001 |
| Core_Temp.*Core_Temp | -0.01 | 0.00 | -5.79 | <0.0001 |

Table 6.19 EE27 MLR L2 Model Parameters

Model Parameter data for the L2 MLR model is presented in Table 6.19. All of the P-values for the t-statistic figures were below the 0.05 level required for inclusion in the model. For some models, in order to minimise the number of independent variables included, variables such as Voltage in this case, with P-values related to their t-statistics figures only marginally greater than 0 were considered for exclusion (manually).

6.7.4 EE27 Multiple Linear Regression (Stepwise Selection, Level 3)
Regression Modelling

Table 6.21 EE27 Matlab Stepwise MLR L3 Analysis Of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>11</td>
<td>3300408</td>
<td>305492</td>
<td>8938</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>6931</td>
<td>236887</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6942</td>
<td>3597296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Level 3 models had 240 explanatory variables available for inclusion. As with the previous models, there were many other independent variables available for selection with t-statistic p-values less than the 0.05 value required for selection. Indeed if the Matlab Stepwise Procedure were carried out to its fullest extent, 105 steps in this case, 81 independent variables would be included in the final model. The RMSE value for that model would be 4.9256 with a R² figure of 0.95372, F-statistic value of 1746 and associated P-value of 0.

In line with the L0, L1 and L2 models an initial model was chosen based on the RMSE -v- Stepwise Step Number plot (Figure 6.5, L3 model data shown in purple). When using the L3 and L4 construction data sets, given that they are so large, with so many viable independent variables to choose from, it was necessary to pick a point on the RMSE -v- Step plot that was before the point where the curve flattens out. A good starting point was generally between step 20 and step 40 depending upon the model (TE01, EE03 etc.). The same procedure as with the L0 – L2 models applies i.e. remove any insignificant independent variables or those with near perfect multicolinearity. The effective addition and removal of significant and insignificant independent variables respectively became more and more complex as the number of independent variables available rose. The inclusion/exclusion of a particular variable can have a marked impact on the significance within the model of many other independent variables some of which may, as a result, have to be excluded. This can make the set-up time of the higher level models quite lengthy. The model eventually arrived at in this case had a model DF of 11.

From the Goodness of Fit Data shown in Table 6.20 it is apparent that the R² value was a very slight dis-improvement over the L2 model at 0.934 and that the RMSE had dis-improved from 5.79 to 5.85.
From the ANOVA Table (6.21) it is evident that 11 of the 240 independent variables for selection were included in the model (Model DF = 11). The F statistic value of 8938 was greater than for the L2 model (7721), due mainly to the reduction in the number of independent variables included (13 for L2, 11 for L3) and was similar to the values for the L0 and the L1 models (9201 and 9641 respectively).

### Model Parameters

| Source                | Value   | Standard Error | t       | Pr > |t| |
|-----------------------|---------|----------------|---------|------|---|
| Intercept             | 236.65  | 15.34          | 15.43   | <0.0001 |
| MW                    | -1.42   | 0.13           | -10.93  | <0.0001 |
| MVAr                  | -0.44   | 0.08           | -5.55   | <0.0001 |
| IGV                   | -1.21   | 0.14           | -8.52   | <0.0001 |
| Exciter_I             | 0.23    | 0.03           | 6.93    | <0.0001 |
| Core_Temp             | -3.33   | 0.19           | -17.52  | <0.0001 |
| MW.*Core_Temp         | 0.02    | 0.00           | 11.97   | <0.0001 |
| MW.*Exciter_I         | 0.00    | 0.00           | -7.52   | <0.0001 |
| MW.*IGV               | 0.01    | 0.00           | 11.49   | <0.0001 |
| MVAr.*Exciter_I       | 0.00    | 0.00           | 3.67    | 0.00  |
| MW.*MVAr.*Core_Temp   | 0.00    | 0.00           | 4.44    | <0.0001 |
| Core_Temp.*Core_Temp.*Core_Temp | 0.00 | 0.00 | -10.40 | <0.0001 |

Table 6.22 EE27 Matlab Stepwise MLR L3 Model Parameters

L3 MLR Model Parameters data are presented in Table 6.22. It is evident that of the P-values for the t-statistic figures were well below the 0.1 threshold for exclusion from the model.

6.7.5 EE27 Multiple Linear Regression (Stepwise Selection, Level 4)

### Goodness of Fit Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>6943</td>
</tr>
<tr>
<td>Sum of weights</td>
<td>6943</td>
</tr>
<tr>
<td>DF</td>
<td>6927</td>
</tr>
<tr>
<td>R²</td>
<td>0.940</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.939</td>
</tr>
<tr>
<td>MSE</td>
<td>31.38</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.60</td>
</tr>
</tbody>
</table>

Table 6.23 EE27 Matlab Stepwise MLR L4 Goodness of Fit Statistics
Table 6.24 EE27 Matlab Stepwise MLR L4 Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>15</td>
<td>3379921</td>
<td>225328</td>
<td>7180</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>6927</td>
<td>217375</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>6942</td>
<td>3557296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.25 EE27 Matlab Stepwise MLR L4 Model parameters

| Source                                | Value       | Standard error | t        | Pr > |t| |
|---------------------------------------|-------------|----------------|----------|------|----------|
| Intercept                             | -178.176239 | 33.427614      | -5.330179 | < 0.0001 |
| MW                                    | 3.850476    | 0.287434       | 13.430817 | < 0.0001 |
| MVAr                                  | -1.673002   | 0.116640       | -14.343306 | < 0.0001 |
| MVA                                   | -1.851124   | 0.135405       | -13.670883 | < 0.0001 |
| IGV                                   | -1.331836   | 0.135110       | -9.857451  | < 0.0001 |
| Exciter_I                             | 0.732160    | 0.038156       | 19.183602  | < 0.0001 |
| Core_Temp                             | -4.033920   | 0.812569       | -4.964400  | < 0.0001 |
| MW.*Core_Temp                         | -0.026403   | 0.002896       | -9.147641  | < 0.0001 |
| MW.*Exciter_I                         | -0.003291   | 0.000183       | -17.752828 | < 0.0001 |
| MW.*Stator_I                          | -0.023589   | 0.008111       | -2.908414  | 0.003644 |
| MW.*IGV                               | 0.008146    | 0.000634       | 12.840051  | < 0.0001 |
| MVAr.*Exciter_I                       | 0.001289    | 0.000158       | 8.144930   | < 0.0001 |
| Core_Temp.*Core_Temp                  | 0.074918    | 0.007383       | 10.175017  | < 0.0001 |
| MW.*MVAr.*Stator_I.*Core_Temp         | 0.000003    | 0.000000       | 3.663347   | < 0.0001 |
| MW.*MVA.*Core_Temp.*Core_Temp         | 0.000001    | 0.000000       | 15.433170  | < 0.0001 |
| Core_Temp.*Core_Temp.*Core_Temp       | -0.000004   | 0.000000       | -18.437724 | < 0.0001 |

Level 4 models had 789 explanatory variables available for selection. In this case (for EE27) if the Matlab Stepwise Procedure were carried out to its fullest extent, 162 steps, 118 independent variables would be included in the final model. The RMSE value for that model would be 4.9256 with a $R^2$ figure of 0.95372, F-statistic value of 1746 and associated P-value of 0. A model with 118 independent variables included would almost certainly be overfitted to the data, resulting in poor performance when data from outside of the construction data was applied to it. Therefore, as with the L0 to L3 models the analyst chose a model from the RMSE -v- Stepwise Plot on the stepwise GUI with the aim of attaining a $R^2$/RMSE value as close to the optimal model (0.95372/4.9256) calculated by Matlab as possible, using a
minimal number of independent variables i.e. best possible RMSE/R² with fewest possible variables.

The EE27 L4 MLR final model Goodness of Fit Data is shown in Table 6.23. It is apparent that for the selected EE27 L4 model, the R² value was an improvement over the L3 model at 0.940 and that the RMSE had improved slightly from 5.846 to 5.602 indicating that the variance of the measured signal was better predicted by the L4 MLR Model than by the L3 model (for the construction data).

From the ANOVA Table (Table 6.24) it is apparent that 15 of the 789 independent variables for selection were included in the final model (Model DF = 15). The F statistic value of 7180 was lowest of the five MLR models, mainly due to the L4 model having the largest number of independent variables included.

From the Model Parameters Table (Table 6.25) the analyst can see that the p-values for the independent variable t-statistic values were all well below the 0.1 threshold for exclusion from the model. Indeed, with the exception of MW.*Stator_I, all P values were <0.0001 indicating an acceptable level of significance in explaining the variation in the EE27 measured signal.

### Table 6.26 EE27 MLR L4 Correlation Table
The strength of the relationships between the independent variables included in the model, and the strength of the relationships between the independent variables and the dependent variable (EE27 in this case) are shown in the Correlation Table (Table 6.26). It is evident that some independent variables are highly correlated e.g. MW and MW*Exciter_I, but none are deemed perfectly correlated and so need not be excluded from the model.

From Figure 6.5 RMSE -v- Matlab True Stepwise Procedure Step Number (L4 Model Data is displayed in orange), It is evident that from approximately Step 70 onward, there was very little improvement in the RMSE value of the model for the inclusion of additional independent variables. However, as discussed earlier in this section, with large numbers of independent variables available a model was selected from the range Step 20 to Step 40 so as to limit the number of independent variables (preferably <40) included in the final model.

Initially when the L4 model was selected from the Matlab Stepwise GUI by selecting the model with an acceptable RMSE value, with a tolerable number of independent variables included, 18 in this case, two variables were picked out as having unacceptable multicolinearity (MW.*MVA and MW.*MVA.*Core_Temp) and one variable (Voltage) had a p-value for t-statistic >0.1. These variables were removed, giving a model with all independent variables being statistically significant. This may not necessarily have been the case, the removal of one independent variable could lead to another becoming statistically insignificant (P >0.1) when the model parameters are recalculated with the lesser number of variables included. Any added variables could have the same effect, resulting in a number of inclusions/exclusions before a model is arrived at where all of the independent variables have P-values related to their t-statistics of less than 0.1.

The application of the L0 – L4 MLR models to new data is covered in Section 8.1 2007 MLR and LOWESS Non-Parametric Regression Models. Goodness of fit statistics are presented for L0 – L4 MLR Models as well as for the LOWESS Model over a number of months. The Stator End Coil used as an example in Section 8.1 was TE09, nonetheless the procedure used to derive each of the twelve models, as described in this chapter, was essentially the same.
6.8 Graphical Assessment of Model Predictive Capability for EE27 L0, L4 and LOWESS Model

Figures 6.6, 6.7 and 6.8 show measured vibration data for Exciter End Bar No.27 plotted along with the predicted vibration data for MLR Models L0, L4 and the LOWESS Model.

The data relating to two-shifting mode is easily distinguished from the original data set owing to its lower mean value and higher variance.

![EE27 Measured and MLR L0 Model Predicted Vibration](image)

**Figure 6.6 Construction Set Data: EE27 Measured and L0 MLR Model Predicted Vibration Levels**

The improvement in performance of the L4 model over the L0 model is clearly visible from Figures 6.6 and 6.7 with a much greater proportion of the variance explained by the L4 model. The improvement for this particular end Coil (EE27) was not as marked as at some other end coils where the lower level models (L0 – L2) could be quite inaccurate.
The LOWESS model outperformed all of the stepwise generated MLR models mentioned in the previous section with a $R^2$ value of 0.967 and a RMSE of 4.118 (see Table 6.27). It is evident also from Figures 6.8 and 6.9 that the variance in the measured EE27 signal was better approximated by the LOWESS model than by either of the MLR Models for which data are displayed.

The LOWESS model in this case was calculated using the LOWESS method, a ‘True’ value of 5% (polynomial applied to 5% of the data series at a time), polynomial degree of 2 using the Triweight Kernel.

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

Table 6.27 EE27 LOWESS Goodness of Fit Statistics
6.9 Assessment of Model Error Terms

As well as analysing the $R^2$ and RMSE quality of fit statistics, error term data should also be considered when assessing regression models. In this regard the main things to take into account are:

- Patterns within the error.
- Linearity of the regression mean.
- Normal distribution of the error terms.
- Skewness.
- Kurtosis.
- Autocorrelation of error.
Note: For this section of the document, Error/Residual data will be displayed for Exciter End End-Coil 19 (EE19) for the L0 and L4 Multiple Linear Regression (MLR) Models and the LOWESS Model. A similar analysis of error terms is carried out for each model predicting each of the 12 measured signals.

6.9.1 Residual Plots

Error/Residual data for Exciter End End-Coil Number 19 (EE19) L0, L4 and LOWESS regression models are presented in Figures 6.9 to 6.11. The L0 model residual plot of Figure 6.9 bears some visual resemblance to the measured vibration data, particularly towards the latter end of the series which is not ideal. The error term should be random in nature with no discernible pattern. The variance of the error term was quite large with most of the data falling within the range -30 to +30 µm pk-pk. The trend line (shown in red) drawn through the data is a second order polynomial approximation of the data. Ideally the residual data should average zero and so the trendline should follow the 0.00 gridline on the trend. This was not the case however, as there was a negative average residual value at the beginning of the series with a slightly positive average toward the end of the series. This again points to the model probably being unable to accurately predict vibration data at EE19.

The EE19 L4 Model Residual Plot, Figure 6.10, shows an improvement on the L0 (Figure 6.9) version with most points within the +/- 10 bands and very few excursions beyond the +/- 30 bands. The pattern visible in the L0 residual plot was not present in the L4 residual data. In addition, the polynomial trendline fitted to the data lay almost flat along the 0.00 gridline. These factors would point to the L4 being reasonably effective in accurately predicting variance in the EE19 vibration signal when presented with new data from outside of the construction/learning data.
Figure 6.9 EE19 Multiple Linear Regression L0 Residual Plot

Figure 6.10 EE19 Multiple Linear Regression (MLR) L4 Residual Plot
The LOWESS residual plot (Figure 6.11) was an almost ideal residual plot with the vast majority of points falling inside the +/-10 bands, and just a few erratic values. The error plot has no discernible pattern implying that the model is of a reasonable quality. A second order polynomial trendline fitted to the data (shown in red) lies almost exactly on the 0.00 gridline which again is a positive result.

6.9.2 Model Residual -v- Model Prediction Plots

Linearity of the Regression Means can be assessed visually by plotting the residuals against any of the independent variables or against the predicted values. As an example, Residual -v- Model Predicted Plots were used in order to assess the linearity of the mean of the EE19 L0, L4 and LOWESS models and are shown in Figures 6.12 – 6.14.
Figure 6.12 EE19 Multiple Linear Regression L0 Residuals -v- L0 Predicted EE19

Figure 6.13 EE19 Multiple Linear Regression Model L4 Residuals -v- L4 Predicted EE19
It is immediately clear that there was significant non-linearity in the regression mean for the L0 model as shown in Figure 6.12. A second order polynomial was again fitted to the data (shown in red) and acted to emphasise the non-linearity in the data. A result such as this suggests that the L0 model will not accurately predict vibration at EE19.

For the L4 model, the results (displayed in Figure 6.13) are much better than for the L0 model, with almost perfect linearity of the mean. The polynomial trendline (shown in red) fitted to the data lies almost flat along the 0.00 gridline. As with the other residual analysis tests the results of the Linearity of Regression Test for the EE19 L4 MLR Model were positive.

![Figure 6.14 EE19 LOWESS Residuals -v- LOWESS Model Predictions](image)

**Figure 6.14 EE19 LOWESS Residuals -v- LOWESS Predicted EE19**

The result of the Linear of Regression Test for the LOWESS model was inferior to the L4 model. From Figure 6.14 it is apparent that the spread of the residuals was similar to the L4 model, however the fitted trendline deviated from the 0.00 gridline for predicted values greater than 120 μm pk-pk approximately. For predicted values less than 120 μm pk-pk, the
trendline lay on, or close to, the 0.00 gridline. This result was broadly positive but suggested some vulnerability to inaccurate prediction at the upper end of the range.

6.9.3 Residual Statistics and Distribution Fitting

Distribution fitting data for the three error vectors shown previously (EE19 L0 MLR, EE19 L4 MLR and EE19 LOWESS) are displayed in Tables 6.28 - 6.30 and in Figures 6.15 - 6.20.

The distributions for the L1 and L4 models failed the Kolmogorov-Smirnov test for normality due to both skewness and kurtosis as evidenced by Tables 6.28 and 6.29. The Normal Distribution parameters were 0 for both skewness and kurtosis. The L0 model error term was however a closer approximation to a normal distribution than the L4 model for which the skewness value was similar to the L0 model, but the kurtosis value is much higher owing to the concentration of the error values around 0.

The error data for the LOWESS model did not pass a normality test and also had a slight negative deviance in the Mean value away from zero (-0.248). The skewness data for the LOWESS model (presented in Table 6.30) were similar to that of the L4 MLR model however the kurtosis of the model was even more pronounced with a large amount of the data concentrated around the 0 point thus breaching the requirement that the distribution be mesokurtic.

Note: A distribution with a positive kurtosis is known as leptokurtic, a distribution with a negative kurtosis is known as platykurtic and a distribution with zero kurtosis mesokurtic.

The variance figure for the L0 error term (71.17, from Table 6.28) was quite large relative particularly to the LOWESS residual data (31.62, from Table 6.29). Indeed it can be seen that the variance figure was reduced by approximately 40 % between the L1 and L4 models and reduced by 26 % between the L4 and LOWESS model. This reduction in the variance value was a positive reflection on the accuracy of the models, meaning that the error term data was more tightly bound to the 0 error line, giving rise to the higher kurtosis figures mentioned earlier.
6.9.3.1 EE19 L0 MLR Model Residual Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Data</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Variance</td>
<td>71.173</td>
<td>71.173</td>
</tr>
<tr>
<td>Skewness (Pearson)</td>
<td>0.400</td>
<td>0.000</td>
</tr>
<tr>
<td>Kurtosis (Pearson)</td>
<td>0.513</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 6.28 EE19 L0 Error Term Statistics

Figure 6.15 EE19 MLR L0 Residuals Histogram

Figure 6.16 EE19 MLR L0 Residuals Cumulative Distributions
6.9.3.2 EE19 L4 MLR Residual Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Data</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Variance</td>
<td>42.943</td>
<td>42.943</td>
</tr>
<tr>
<td>Skewness (Pearson)</td>
<td>0.447</td>
<td>0.000</td>
</tr>
<tr>
<td>Kurtosis (Pearson)</td>
<td>1.649</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 6.29 EE19 L4 Error Term Statistics

Figure 6.17 EE19 MLR L4 Residuals Histogram

Figure 6.18 EE19 MLR L4 Residuals Cumulative Distributions
6.9.3.3 EE19 LOWESS Residual Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Data</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.248</td>
<td>-0.248</td>
</tr>
<tr>
<td>Variance</td>
<td>31.620</td>
<td>31.620</td>
</tr>
<tr>
<td>Skewness (Pearson)</td>
<td>0.496</td>
<td>0.000</td>
</tr>
<tr>
<td>Kurtosis (Pearson)</td>
<td>4.540</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 6.30 EE19 LOWESS Error Term Statistics

![EE19 LOWESS Residuals Histogram](image1)

Figure 6.19 EE19 LOWESS Residuals Histogram

![EE19 LOWESS Cumulative Distributions](image2)

Figure 6.20 EE19 LOWESS Residuals Cumulative Distributions
### 6.9.4 Autocorrelation of the Residuals

Autocorrelation is a special case of correlation which instead of referring to the relationship between two or more variables, refers to the relationship between successive values of the same variable. Autocorrelation in the regression error term is generally caused by one or all of the following:

- A significant independent variable has been excluded.
- Misspecification of the functional form of the model e.g. where a linear form is applied when the actual form of the model is non-linear.
- Measurement errors in the dependent variable. Interpolation for instance can lead to loss of valuable data by averaging out true disturbances.
- Misspecification of the Error. Successive error values may be related due to purely random factors or events.

![Figure 6.21 EE19 MLR L0, L4 and LOWESS Models Autocorrelation of Error Terms](image)
The first 100 autocorrelation coefficients for the EE19 L0, L4 and LOWESS Models as well as the upper and lower 95% confidence bounds are shown in Figure 6.21. Autocorrelation values for the construction set residuals were calculated using the Matlab autocorr function. The Matlab autocorr function calculates the 95% upper and lower confidence bound for the autocorrelation as $\pm 2/\sqrt{N}$.

Information on Autocorrelation was gleaned from a number of sources e.g. [29], [30] and [31].

**The Durbin-Watson Test for Autocorrelation**

The Durbin-Watson statistic ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation. A d value toward 0 indicates positive autocorrelation while a d value toward 4 indicates negative autocorrelation.

The Durbin-Watson test for autocorrelation can be calculated in two ways firstly by using:

$$
\text{Equation 6.4 The Durbin-Watson Statistic}
$$

Or alternatively using the value of the first order autocorrelation coefficient:

$$
\text{Equation 6.5 The Durbin-Watson Statistic (Alternative)}
$$
‘d’ values for EE19 L0, L4 and LOWESS Models are as follows:

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Regressors</th>
<th>First Order Autocorrelation Coefficient</th>
<th>Durbin-Watson Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE19 L0</td>
<td>9</td>
<td>0.335</td>
<td>1.33</td>
</tr>
<tr>
<td>EE19 L4</td>
<td>34</td>
<td>0.143</td>
<td>1.71</td>
</tr>
<tr>
<td>EE 19 LOWESS</td>
<td>8</td>
<td>0.007</td>
<td>1.99</td>
</tr>
</tbody>
</table>

Table 6.31 EE19 Autocorrelation of Residual Vector Data

Table 6.31 shows the first order autocorrelation coefficients of the L0, L4 and LOWESS regression model residual vectors, the number of regressors included in each model and the Durbin Watson statistic associated with the error vector of each model.

The results for the L4 MLR model and the LOWESS model are good. The result for the L0 model indicates a problem with positive autocorrelation of the error term.

The Upper and Lower 95 % confidence bound for the Durbin Watson Statistics are found in a set of tables but are limited to a sample size of 200 and 20 regressors. Analysing the tables however, the 95 % Upper and Lower confidence bounds for positive correlation converge towards 2 as n increases. For very high n, as is the case for this data, the L4 model would likely be classed as having an unacceptable level of autocorrelation in the error term.

An alternative autocorrelation test statistic called the Durbin h-Test designed for large sample sizes is not applicable in this case.

**Alternative to Durbin-Watson**

Some analysts do not like to use the Durbin-Watson statistic and instead prefer to use the rule of thumb that if the first order autocorrelation of the regression residual vector is greater than 0.3 then the regression method used is not the best estimation method. From Table 6.28, as with the Durbin Watson statistic, L0 is seen to be substandard as a model while the L4 and LOWESS models perform well.
6.10 Model Validation Techniques and Selection of MLR Models for Implementation on PI-ACE

Verification of the accuracy of the Regression Models may be carried out in two ways:

1. A random sample of the construction data can be applied to the model and the predictive capability of the model assessed.

2. New data is applied to the calculated models, predictions are calculated and the accuracy of the model predictions measured using a number of Goodness of fit statistics.

For this application, as there is no shortage of data, new data are applied to the model for verification purposes. The application of new data to the models is dealt with in Chapter 8.

The best performing MLR Model for each sensor position is chosen for application on the PI-ACE system. The models are chosen based on the following criteria:

- Highest average/ most consistent $R_{pred}^2$
- Lowest average RMSE values
- Simplest model i.e. if there is little improvement in performance between a Level 2 (L2) and Level 4 (L4) model then the simpler model should be selected.
- This is repeated for each sensor location, giving 12 models for implementation on PI-ACE.

The application of new data to the MLR and LOWESS Regression Models and their performance is presented in Chapter 8, with further information available in Appendix 5.
Information on problems with using various regression techniques and on model validation was garnered from a number of sources e.g. [33] and [34].

6.11 Chapter Summary

Some basic information on the various regression modelling techniques employed on this project are presented in Section 6.2.

The measured plant variables available for use as model independent variables, the creation of new variables and the development of a number of Construction/Learning Data Sets (input data for the regression models) are presented in Section 6.3.

Variable selection techniques for the MLR models particularly, such as all possible regressions and the various Stepwise variations, their strengths and weaknesses are discussed in Section 6.4.

The make-up of the construction data in terms of size (number of observations) and how the data was filtered prior to model calculation are discussed in Section 6.5.

The Simple and Multiple Regression Models presented in Sections 6.6.1 and 6.6.2 respectively were created using construction data made up of measured plant variables only i.e. those listed in Section 6.3.2. The construction data for these models was made up of a number of weeks’ worth of unfiltered sequential plant data. The simple model performed reasonably well given that only one independent variable was used. The MLR model was initially displayed with all available independent variables included, and subsequently with only statistically significant variables included, following the removal of three statistically insignificant variables. Goodness of Fit, ANOVA, Model Parameter data are presented and discussed for these models.

In section 6.7 MLR models are calculated using the construction data sets L0 to L4 presented in Section 6.3.3. The construction data in this case was filtered for shut downs and start-ups and data from a wider range of operating conditions are included as discussed in Section 6.5. The performance of the models developed from the L0, L1 to L4 construction/learning data
sets are presented and compared to ascertain the best performing models as defined in Section 6.10. LOWESS model data is also presented for comparison against the MLR models.

Assessment of model performance is discussed in Sections 6.8 and 6.9. Section 6.8 deals with graphical/visual assessment of the models by plotting measured vibration against model predicted vibration for the construction set data. GOF scores of some of the MLR models and the LOWESS model are also discussed. In Section 6.9 assessment of model performance was carried out by analysis of the error term/residual data. Techniques utilised include Error Plots, Residual -v- Predicted Plots, Statistics related to the Model Error Term, Distribution Fitting and checks for Autocorrelation of the residuals.

Techniques for model validation and the criteria for selection of MLR models for implementation on PI ACE is discussed in Section 6.10.
CHAPTER 7

STATISTICAL ANALYSIS OF GENERATOR CONDITION

2006 – 2011

In this Chapter a basic assessment of the behaviour of the machine with respect to End Winding Vibration from the time of installation of the accelerometers to mid-2011 is conducted. Particular attention is paid to data from May 2007 onwards (Post Low tune Repair). The assessments are made using basic graphical and statistical analysis methods applied to the measured vibration data.

7.1 Scattergrams

Scattergrams of the measured vibration data from W060063 (Jan/Feb 2008), W116119 (February 2009), W164167 (Jan/Feb 2010) and W216219 (Jan/Feb 2011) are presented in Figures 7.1 to 7.4. Mean and Median values displayed in each of the charts show that there has been no major change in the levels of vibration on either end of the machine in the time since the “Low Tune” repair was carried out (April/May 2007).

Figure 7.1 TE and EE EVW Scattergrams for W060063
Figure 7.1 shows vibration data scattergrams for each of the twelve points of measurement with their Mean and Median values marked with a red cross and bar respectively. For the period in question (14/01/2008 to 11/02/2008) the average generator load was 183.97 MW, the average ambient air temperature was 9.5 °C with a minimum value of 0.59 °C.

Figure 7.2 TE and EE EVW Scattergrams for W116119

![Figure 7.2 TE and EE EVW Scattergrams for W116119](image)

Figure 7.3 TE and EE EVW Scattergrams for W164167

![Figure 7.3 TE and EE EVW Scattergrams for W164167](image)

Figure 7.2 displays measured vibration data scattergrams for each of the twelve accelerometers with their Mean and Median values marked with a red cross and bar respectively. For the period
in question (02/02/2009 to 12/03/2009) the average generator load was 188.45 MW, the average ambient air temperature was 8.66 °C with a minimum value of 0.76 °C.

Figure 7.3 displays vibration data scattergrams for each of the twelve points of measurement along with their Mean and Median values. For the period in question (18/01/2010 to 08/02/2010) the average generator load was 192.85 MW, the average ambient air temperature was 5.76 °C with a minimum value of -1.29 °C.

![Figure 7.3 Scattergrams for TE and EE EVW](image)

Figure 7.4 TE and EE EVW Scattergrams for W216219

Figure 7.4 shows measured vibration data scattergrams for each of the twelve points of measurement along with their Mean and Median values. For the period in question (10/01/2011 to 07/02/2010) the average generator load was 179.57 MW, the average ambient air temperature was 6.04 °C with a minimum value of -2.40 °C.

The scattergrams shown in Figures 7.1 to 7.4 demonstrate that measured vibration levels at the generator end coils did not change significantly over the measurement period. Peak vibration levels for the Turbine End are in the region 200 - 225 μm pk-pk during normal conditions but can rise to 275+ μm pk-pk during periods of extremely low ambient temperature such as the winter periods of 2009 and 2010. Exciter End vibration levels peaked at approximately 140 μm pk-pk, rising to 175 μm pk-pk and were not as adversely affected by changes in Cold Air Temperature/Ambient Air Temperature.
7.2 Pearson Correlation Tables

The following are Pearson Correlation Tables for the same data sets used in Section 7.1. Tables 7.1 and 7.2 show data for the 12 measured vibration signals and the 10 relevant explanatory variables at the time the data was recorded. Tables 7.3 and 7.4 display values for 13 independent variables as the data for CCCW Outlet Temperature, Generator Cold Gas Temperature (one measurement from each end of the machine) as well as Generator Hot Gas Temperature were made available.

Table 7.1 Pearson Correlation Table for W6063 (14/01/2008 to 11/02/2008)

<table>
<thead>
<tr>
<th>Variables</th>
<th>TE 00</th>
<th>TE 01</th>
<th>TE 17</th>
<th>TE 25</th>
<th>TE 33</th>
<th>TE 41</th>
<th>EE 01</th>
<th>EE 11</th>
<th>EE 19</th>
<th>EE 27</th>
<th>EE 35</th>
<th>EE 43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Power (MVA)</td>
<td>-0.10</td>
<td>0.81</td>
<td>0.63</td>
<td>0.86</td>
<td>0.80</td>
<td>0.84</td>
<td>0.87</td>
<td>0.89</td>
<td>0.89</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Reactive Power (MVA)</td>
<td>0.82</td>
<td>0.96</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Generator V1 (V)</td>
<td>0.75</td>
<td>0.78</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Generator V2 (V)</td>
<td>0.74</td>
<td>0.77</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>Generator V3 (V)</td>
<td>0.73</td>
<td>0.76</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>Generator V4 (V)</td>
<td>0.72</td>
<td>0.75</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
<td>0.81</td>
</tr>
<tr>
<td>Generator V5 (V)</td>
<td>0.71</td>
<td>0.74</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Generator V6 (V)</td>
<td>0.70</td>
<td>0.73</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 7.2 Pearson Correlation Table for W116119 (02/02/2009 to 02/03/2009)

<table>
<thead>
<tr>
<th>Variables</th>
<th>TE 00</th>
<th>TE 01</th>
<th>TE 17</th>
<th>TE 25</th>
<th>TE 33</th>
<th>TE 41</th>
<th>EE 01</th>
<th>EE 11</th>
<th>EE 19</th>
<th>EE 27</th>
<th>EE 35</th>
<th>EE 43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Power (MVA)</td>
<td>0.84</td>
<td>0.87</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>Reactive Power (MVA)</td>
<td>0.87</td>
<td>0.90</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Generator V1 (V)</td>
<td>0.88</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Generator V2 (V)</td>
<td>0.88</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Generator V3 (V)</td>
<td>0.89</td>
<td>0.91</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Generator V4 (V)</td>
<td>0.90</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Generator V5 (V)</td>
<td>0.90</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Generator V6 (V)</td>
<td>0.90</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Chapter 7 Page 154 Analysis of Generator Condition
Table 7.2 shows Pearson Correlation Data for all measured dependent and independent variables for W16119 (02/02/2009 to 02/03/2009).

Table 7.3 Pearson Correlation Data for W161467 (18/01/2010 to 08/02/2010)

Table 7.3 displays Pearson Correlation Data for all measured dependent and independent variables for W161467 (18/01/2010 to 08/02/2010).

Table 7.4 Pearson Correlation Table for W216219 (10/01/2011 to 07/02/2011).
Table 7.4 shows Pearson Correlation Data for all measured dependent and independent variables for W216219 (10/01/2011 to 07/02/2011).

The Pearson Coefficient of Determination Data shown in Tables 7.1 - 7.4 again show that there was no major shift in the behaviour of the machine as evidenced by very similar coefficient values describing the relationships between the various variables used in the calculation of the regression models over a long period of time (2008 - 2011). The Data shown was deliberately chosen from roughly the same period of each year so that the data was a “like for like” comparison in so far as was possible. Relationships between some of the various variables changed seasonally, due mainly to the prevailing ambient conditions.

### 7.3 Descriptive Statistics

In assessing the condition of the machine, as well as developing Stepwise MLR models and LOWESS models some basic descriptive statistics were calculated. As mentioned in Chapter 4, statistics were calculated for the generator under certain operating conditions so that the results were comparable from one period to the next. From Week 1 (W001) to Week 208 (W208) statistics were calculated for data falling within the bounds 215 - 235 MW, 0 - 50 MVAr Lagging and Average Slot Temperature ≥ 78 °C.

Due to changes in the running regime of the plant, the bounds of generator capability and temperature used for the calculation of descriptive statistics were changed to 205-215 MW, -15 ≥ MVAr ≤ +15 and 70 °C ≥ Average Slot Temperature ≤ 85 °C from W154 (26/10/2009) to the present. Descriptive statistics were calculated for both sets of criteria from W154 to W208.

It is evident from Figure 7.5 that the condition of the machine had not changed greatly over time. The change in behaviour resulting from the Low Tune Repair in 2007 is clearly visible. Changes in vibration behaviour related to seasonal ambient conditions are also clearly visible.
Since the inception of this project the normal operating regime of the machine has changed considerably. When the area of the generator capability used to calculate descriptive statistics was originally selected, about 22% of the weekly operating values fell within those constraints.

Due to the massive growth in wind generation capability, the bringing on line of two new gas powered generating plants in Cork and a drop in system demand, the machine now tends to operate at a lower output during times of peak demand. Typically the plant is instructed to around 300 MW during the morning and evening peak demand periods. This equates to approximately 200 MW provided by the GT Generator and 100 MW from ST Generator, leaving a spinning reserve of approximately 30 MW. The maximum output from the plant can vary by up to 15% from plated maximum capability depending on prevailing ambient conditions.
The reduction in running time and the increased number of start-ups and shut-downs resulted in the average number of valid points available to calculate the descriptive statistics falling from 147 of 672 (21.8%) in 2007 to 27 of 672 (4%) in 2010. For some weeks there were no data falling into the required range.

<table>
<thead>
<tr>
<th>Year</th>
<th>Average Weekly Data Points (n of 672)</th>
<th>Percentage of Valid Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>147</td>
<td>21.8</td>
</tr>
<tr>
<td>2008</td>
<td>56</td>
<td>8.3</td>
</tr>
<tr>
<td>2009</td>
<td>39</td>
<td>5.8</td>
</tr>
<tr>
<td>2010</td>
<td>27</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Table 7.5 Weekly Data Falling Within the Set Range for Descriptive Statistics Calculations 2006 to 2010.

However, even with the reduced sample sizes, the data shows that the behaviour of the machine had been consistent from when the repair was carried out in 2007 as shown for TE09 in Figure 7.5.

The mean values varied from 100 µm pk-pk to 150 µm pk-pk. Seasonal variations were clearly visible with a peak and trough for every year. Some of the remaining variation stems from operating conditions skewed heavily toward one end of the acceptable range or from data sets with very few points included for calculation of statistics.

Due to the changed operating regime of the generator the descriptive statistic values started to become a little erratic. It was particularly noticeable in the range values which dropped off to almost zero in some instances, especially in the 2009 and 2010 data. This suggested that the data pertaining to those points was from a very narrow band of data within the acceptable range and generally related to calculations where the number of points (count) was very low.

To address this issue it was decided that another data range should be considered for calculation of descriptive statistics. It was decided to change the data selection criteria for the calculation of descriptive statistics as shown in Table 7.6.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Old (15m)</th>
<th>New 1 (15m)</th>
<th>New 2 (5m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Power (MW)</td>
<td>215-235</td>
<td>205-215</td>
<td>205-215</td>
</tr>
<tr>
<td>Reactive Power (MVAr)</td>
<td>0-50 Lagging</td>
<td>-15 ≥ MVAr ≤ +15</td>
<td>-15 ≥ MVAr ≤ +15</td>
</tr>
<tr>
<td>Average Slot Temperature (°C)</td>
<td>≥78°C</td>
<td>70 ≥ T ≤ 85</td>
<td>70 ≥ T ≤ 85</td>
</tr>
</tbody>
</table>

Table 7.6 Old and New Descriptive Statistics Calculation Ranges and Average Value Calculation Intervals.

Due to all vibration measurement and plant data being made available to the PI System, data for all variables are available at intervals of as little as 1s depending on how tag parameters were set up. It was decided that, as the data is available at shorter intervals, it would be prudent to investigate whether a reduction in sample interval would lead to any improvement in the quality of regression models or in stability of descriptive statistics data. Any change in this regard resulted in an increase in computational effort to generate models and calculate statistics which must be kept in mind.

Figure 7.6 shows some descriptive statistics (Mean, Range, Standard Deviation and Standard Error) for End Coil TE09 relating to the period 26/10/2009 to 01/03/2010 (W154 to W171) with interval settings of 5 minutes and 15 minutes. No data is displayed for weeks 162 and 164 as the machine was instructed off-line by Eirgrid NCC.

Mean values for both interval settings were, as expected almost identical and average at approximately 110 µm pk-pk. Values for Standard Deviation and Standard Error were also broadly similar for both interval settings over the period, with Standard Deviation averaging slightly higher for the 5m interval setting and Standard Error averaging slightly higher for the 15m interval setting. The most significant change was in the Range values with the 5 minute interval data averaging about 10 µm pk-pk greater than the values for the 15 minute interval setting.

It was decided to persist with the use of 15m intervals as the benefits of using the shorter interval were not significant enough to warrant the change being made a permanent one.
Figure 7.6 TE09 Descriptive Statistics Comparison Data for (5 and 15 Minute Averaging Interval) W154 to W171 (26/10/2009 to 08/03/2010)

Figure 7.7 TE09 Descriptive Statistics 15 Minute Averaging Interval with New Ranges for Inclusion of Data W154 to W233 (26/10/2009 to 16/05/2011)
In Figure 7.7 descriptive statistic data for TE09 relating to the period 26/10/2009 to 01/03/2010 (W154 to W234) with the interval setting at 15 minutes are displayed. A dip in mean vibration values is clearly evident during the periods of colder weather which are highlighted in green.

From the data presented in Chapter 7 it is apparent that the condition of the machine remained relatively stable over the period covered by this project. Some seasonal variation in average recorded vibration values, due to the effects of ambient conditions and varying running regimes imposed by Eirgrid NCC are visible in the data. Overall however, it is clear that since the Low Tune Repair carried out in April of 2007 measured vibration at the end windings stabilised, and remained relatively stable over an extended period. The data presented in Chapter 7 is complemented for the most part by the MLR and LOWESS model data presented in Chapter 8 i.e. the condition of the machine end winding had not deteriorated greatly since the Low Tune Repair was carried out. This assertion was verified by annual Static Natural Frequency Tests/Bump Tests and visual inspections as detailed in Chapter 3.
CHAPTER 8

PREDICTIVE PERFORMANCE OF MLR AND LOWESS MODELS

In this Chapter, model solutions calculated using the methods set out in Chapter 6 (more specifically the MLR models from Section 6.7) and explained further in Appendix 3, will have data from outside of the learning/construction data applied to them. The predictions of the various models will be analysed and compared against actual measured vibration data.

This machine, like any other, is subject to wear and tear and so its behaviour in terms of vibration has changed slightly over time. The calculation of a regression model cannot account for change in machine behaviour due to wear and tear or other events such as maintenance activities etc. Changes such as those mentioned will therefore necessitate that the model parameters be recalculated in order to maintain predictive accuracy. Any recalculation of the models due to a fall-off in accuracy is carried out only after it has been established that the machine is healthy. This is done by conducting a visual inspection, Non-Destruct Testing (NDT) and a “Bump Test” on the end winding structure.

The data to be applied to the 2007 models will be filtered in the same way as the data used to calculate them i.e. all data where GT Generator Active Power output is above 120 MW with frequency between 49.5 and 50.5 Hz, and Stator Core Temperature > 50 °C. This simply ensures that the generator is synchronised and operating conditions are normal.

Over time the models were modified by recalculation of coefficients and addition/creation of new explanatory variables. A number of changes have taken place in terms of declared capability of the unit, the main one being a change to the Minimum Stable Generation value from 216 MW to 184 MW in 2009, and to 171 MW in 2011. These adjustments must be assessed for any effect they might have on the normal behaviour of the generator, in terms of vibration, temperature etc. Any change in the relationship between generator temperatures and load or vibration will generally require a recalculation of the regression models, if the change is a permanent one. It was found that most data below 216MW (142MW GT Output) needed
to be filtered out due to nonlinearities in variable relationships involving generator temperatures.

The regression models used in this section were calculated using the True Stepwise method (Backward and Forward) or the LOWESS method. The details of how these methodologies work and how they were applied in this case is dealt with in Chapter 6 and explained further in Appendix 3.

As mentioned previously, due to the behaviour of the End Coils being non-uniform it was not possible to generate an “average” model for the entire end winding basket on each end of the machine, instead a model was generated for each of the 12 points where vibration is measured.

Depending on the End Coil in question a calculated model may prove accurate over quite a long period (up to 11 months) or on the other hand may require recalculation due to a reduction in predictive accuracy after 3 – 6 months. For most models relatively steady predictive capability was maintained over 9 – 10 months.

Predictive capability of the models can degrade for a number of reasons:

- A change in the vibration behaviour of a specific End Coil/number of End Coils in the End Winding Basket due to breakage, wear and tear etc.

- A change in the vibration behaviour of a specific End Coil/number of End Coils in the End Winding Basket due a change in the set-up of the machine. Generally the only change made is to cooling water flow rates.

- A change in the vibration behaviour of a specific End Coil/number of End Coils in the End Winding Basket due a change in ambient condition. This affects the temperature of the cooling medium as well as the temperature of the generator building, all of which leads to changes in vibration amplitudes at a given load. The extremely cold temperatures experienced over the winter periods of 2009 and 2010 led to the predictive capability of the Regression Models being compromised.
Database errors mainly of two distinct types have also lead to problems with the predictive capability of the models namely:

1. Errors due to flat lining of variable data i.e. no change in the value of a variable over an extended period. This can equally affect either dependent or independent variable data.
2. Errors in calculated values e.g. incorrect results for average value of two or more variables or other calculated data.

With models using only the original ten input variables, it usually requires that the model be at least L2 or L3 to attain a good quality predictive model. At L2 and L3 the number of independent variables is quite high (110 and 240 independent variables respectively) and as a result the stepwise procedure will be lengthy and include many variables which are of statistical significance in predicting construction data dependent variable values but do not increase the predictive capability of the model to any great degree.

It was found also that even when considered manual intervention was made in order to limit the number of variables included in a given model, that the model with more variables included in the calculation and with the best goodness of fit score on the construction/learning data set did not necessarily perform better that those with less dependent variables available for selection. This again was due mainly to the inclusion of statistically insignificant variables or the exclusion of significant variables.

### 8.1 2007 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

The first reliable Multiple Linear Regression models generated were calculated using some or all of the available measured plant variables only, see Section 6.6.2. In order to provide more variables for use in model calculation it was decided to use interactions and power multiples of the independent variable data. Obviously this leads to the creation of many statistically insignificant variables, but the generation of even one or two significant variables can greatly improve the predictive capability of the MLR models. The use of interactions etc. results in
the creation of large amounts of independent variables. The calculation of these variables is dealt with in Chapter 6.

The Models were calculated using input data sets L0, L1, L2, L3 and L4, L0 being the ten original variables only, L4 being an input data set with an interaction level of 4. i.e. 4 independent variable values were used to calculate the new variables these were added to the L3 list to give the L4 independent variable array. For more information refer to section 6.3.3 Creation of New Explanatory Variables

The construction/learning data set was created using data from autumn and winter of 2007. This ensured that data was provided from both warm and cold ambient conditions to allow a better (year round) approximate model to be generated. The construction/learning data set used approximately 10 weeks’ worth of data (6944 data points) in order to calculate the model coefficients.

The LOWESS Model in this case was calculated using the LOWESS method (Robust LOWESS is also an option) with a Polynomial Degree setting of 2, the percentage of the Learning data over which the polynomial was implemented at a time is 2 % (True = 2 %). 2 % of the Learning Sample of 6943 observations/data points equates to 139 observations approximately. The kernel used was the Triweight kernel.

Table 8.1 shows goodness of fit statistics for TE09 from W032035 to W096099 (02/07/2007 to 20/10/2008) for the five MLR models L0 to L4 and the LOWESS model. It is apparent that the accuracy of prediction varied seasonally. The R-squared value for the better models averaged at approximately 0.7 which is an acceptable score.

The most accurate models (as defined in Section 6.10) over the 68 week period were the L4 model (with an average R-squared value of 0.67 and average RMSE of 16.87 µm pk-pk) and the LOWESS model (with an average R-squared value of 0.64 and average RMSE of 17.74 µm pk-pk)
### Table 8.1 Goodness of Fit Statistics for TE09 from W032035 to W096099 for MLR Models L0 – L4 and the LOWESS Model

![Figure 8.1 TE09 R-squared values from W032035 to W096099 for MLR L0 – L4 and LOWESS Models](image-url)

![Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators](image-url)
Figure 8.1 shows R-squared values for Turbine End End-Coil Number 9 (TE09) from W032035 to W096099 (02/07/2007 to 20/10/2008) for the five MLR (Multiple Linear Regression) models L0 to L4 and the LOWESS model. The values vary depending on the running regime prevailing at the time. For instance from W032035 to W056059 the machine was operated in two-shifting/cycling mode where the Power Plant was instructed off overnight on a regular basis. This leads to less reliable predicted values due to the cooling and reheating of the generator on shut-down and restart respectively. It can take up to 6 hours for the system to return to a completely normal state following a start-up of the machine. This warm-up time varies depending upon the length of time for which the machine has been offline and ambient conditions. The repeated stop-start cycle means that there are long periods of transient operating conditions where measured vibration values are more erratic than normal, and where the relationships between the various independent variables and the measured vibration values are not the same as for normal operation. Diminished predictive capabilities as well as more erratic measured vibration values combined to give reduced goodness of fit scores for the data in question.

Figures 8.2 to 8.4 show Measured and MLR L0, MLR L4 and LOWESS Model predicted vibration amplitudes respectively for the period W088091 (28/07/2008 to 24/08/2008). This period was chosen as an example as at this point the Goodness of Fit (GOF) scores of some of the models had begun to deteriorate, in some cases quite dramatically. Choosing data from this period allowed more in depth examination of why the deterioration in GOF scores occurred.

Measured vibration, TE09, is shown in dark blue in each of the figures (8.2 to 8.4). It is apparent that the predicted values gave a reasonable representation of the measured signal, although the LOWESS model appeared to perform quite poorly. The W088091 etc. designation for time periods is explained in Section 5.3.
Figure 8.2 TE09 Measured and MLR L0 Model Predicted Vibration for W088091 (28/07/2008 to 24/08/2008)

Figure 8.3 TE09 Measured and MLR L4 Model Predicted Vibration for W088091 (28/07/2008 to 24/08/2008)
Chapter 8

Performance of Regression Models

Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

Figure 8.4 TE09 Measured and LOWESS Model Predicted Vibration for W088091
(28/07/2008 to 24/08/2008)

TE09 Goodness of fit statistics for the period W088091 (28/07/2008 to 24/08/2008) are shown in Table 8.2 (extracted from Table 8.1).

<table>
<thead>
<tr>
<th>TE09 Goodness of Fit Statistics for W088091</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
</tr>
<tr>
<td>Predicted Mean</td>
</tr>
<tr>
<td>R-Squared</td>
</tr>
<tr>
<td>Matlab Fit</td>
</tr>
</tbody>
</table>

Table 8.2 TE09 Goodness of Fit Statistics for W088091 (28/07/2008 to 24/08/2008)

MLR Interaction Levels L0 – L4 and LOWESS Models

From Table 8.2 it is evident that for this particular data, the best approximation of the measured signal was given by the L0 model. On this occasion the L0 model gave a predicted series with an R-squared value of 0.73 with the next best model being the L4 model (0.71). The L0 RMSE value which mirrors the R-squared value in terms of ranking a model’s
goodness of fit was the lowest of the six models at 12.55 µm pk-pk with the L4 model again next best with a RMSE value of 13.06 µm pk-pk. This would not normally be the case. In general, at least one of the more complex models would provide a better approximation of the measured signal than the L0 model.

However, the difference between the measured and predicted mean values was smallest for the more complex L4 model at 2.87 µm pk-pk, with the next best being the L1 model at 2.9 µm pk-pk, followed by the L0 model at 3.01 µm pk-pk. The difference between the measured and predicted mean can be a good indicator of how well the predicted values match the measured values. However, some models produce a predicted value series with a mean value very close to that of the measured signal but with a lesser degree of variance (particularly true for simple regression models i.e. one independent variable).

For the period W088091 this was the case with the L4 prediction particularly, which produced a predicted value series with a difference between the measured and predicted means of 2.870 µm pk-pk but did not match the measured signal variance as successfully as the L0 model. This resulted in an R-squared value of 0.71 as opposed to 0.73 for the L0 model.

The goodness of fit statistic values while reasonably good for the better performing L0 and L4 models were relatively poor for the others. It is noteworthy that W088091 was the 15th data set (of 17) applied to the model and as expected the predictive capability of the models was starting to wane.

### Correlation Matrix

<table>
<thead>
<tr>
<th>Variables</th>
<th>TE09</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE09</td>
<td>1</td>
</tr>
<tr>
<td>TE09 L0 Predicted</td>
<td>0.881</td>
</tr>
<tr>
<td>TE09 L1 Predicted</td>
<td>0.857</td>
</tr>
<tr>
<td>TE09 L2 Predicted</td>
<td>0.859</td>
</tr>
<tr>
<td>TE09 L3 Predicted</td>
<td>0.863</td>
</tr>
<tr>
<td>TE09 L4 Predicted</td>
<td>0.883</td>
</tr>
<tr>
<td>TE09 LOWESS Predicted</td>
<td>0.780</td>
</tr>
</tbody>
</table>

Table 8.3 Correlation Coefficient Table for TE09 Measured against TE09 MLR and LOWESS Model Predicted Values (W088091)
Correlation coefficient values for TE09 measured vibration against MLR L0 to L4 and LOWESS model predicted vibration are displayed in Table 8.3. It is apparent that the model predicted data for all models were highly correlated with the measured data, the average correlation value being 0.854 approximately. The LOWESS model performs poorly when compared with the rest of the models with a correlation coefficient of 0.78. The correlation figures generally complement the results of the Goodness of Fit statistics.

<table>
<thead>
<tr>
<th>Learning Data</th>
<th>Goodness of Fit Statistics EE03 from W032035 to W096099 (02/07/2007 to 20/10/2008)</th>
<th>Goodness of Fit Statistics EE03 from W032035 to W096099 (02/07/2007 to 20/10/2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE03 I1 R Squared</td>
<td>Predicted Mean 0.76 ± 0.00 RMSE 6.84 µm pk-pk Measured Mean 0.80 ± 0.00 Matlab Fit 0.56 ± 0.00</td>
<td></td>
</tr>
<tr>
<td>EE03 I2 R Squared</td>
<td>Predicted Mean 0.80 ± 0.00 RMSE 6.99 µm pk-pk Measured Mean 0.80 ± 0.00 Matlab Fit 0.56 ± 0.00</td>
<td></td>
</tr>
<tr>
<td>EE03 I3 R Squared</td>
<td>Predicted Mean 0.80 ± 0.00 RMSE 6.99 µm pk-pk Measured Mean 0.80 ± 0.00 Matlab Fit 0.56 ± 0.00</td>
<td></td>
</tr>
<tr>
<td>EE03 I4 R Squared</td>
<td>Predicted Mean 0.80 ± 0.00 RMSE 6.93 µm pk-pk Measured Mean 0.80 ± 0.00 Matlab Fit 0.59 ± 0.00</td>
<td></td>
</tr>
<tr>
<td>EE03 I5 R Squared</td>
<td>Predicted Mean 0.80 ± 0.00 RMSE 6.84 µm pk-pk Measured Mean 0.80 ± 0.00 Matlab Fit 0.59 ± 0.00</td>
<td></td>
</tr>
</tbody>
</table>

Table 8.4 Goodness of Fit Statistics for EE03 from W032035 to W096099 for MLR Models L0 – L4 and the LOWESS Model

Table 8.4 displays Goodness of Fit statistics for end coil EE03 from W032035 to W096099 (02/07/2007 to 20/10/2008) for the five MLR models L0 to L4 and the LOWESS model. It is evident that the accuracy of prediction varies seasonally but less so than is the case for TE09.

The best performing models (as defined in Section 6.10) over the 68 week period were the L3 model (with an average R-squared value of 0.37 and average RMSE of 6.84 µm pk-pk) and the LOWESS model (with an average R-squared value of 0.45 and average RMSE of 6.35 µm. |
pk-pk). Although the Goodness Of Fit (GOF) scores were lower than for the better performing models, they were at least consistent. This still gave the analyst a reasonable indicator of machine behaviour although the GOF scores were lower than might have been desired. In general a score of 0.7 is taken to be a reasonable fit.

Figure 8.5 shows R-squared values for Exciter End End-Coil Number 3 from W032035 to W096099 (02/07/2007 to 20/10/2008) for the five MLR models L0 to L4 and the LOWESS model. It is immediately evident that the goodness of fit values were lower for EE03 than for TE09 across all of the models. However, the R-squared values were consistently in the region of 0.35 – 0.55 which suggests that the behaviour EE03 had been stable over the period in question, up to W088091 at which point the R-squared values collapse into negative figures.

Figure 8.5 EE03 R-squared values from W032035 to W096099 for MLR L0 – L4 and LOWESS Models

Figures 8.6 to 8.8 show EE03 measured against MLR L0, MLR L4 and LOWESS model predicted vibration series’ respectively, for the period 28/07/2008 to 24/08/2008. EE03 measured vibration is shown in blue in each case. It is apparent that the predicted values gave a reasonable representation of the measured signal although the match was not as good as for TE09.
Figure 8.6 EE03 Measured and MLR L0 Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008)

Figure 8.7 EE03 Measured and MLR L4 Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008)
Figure 8.8 EE03 Measured and LOWESS Model Predicted Vibration for W088091 (28/07/2008 to 25/08/2008)

Table 8.5 EE03 Goodness of Fit Statistics for W088091

<table>
<thead>
<tr>
<th></th>
<th>L0</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>LOWESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measured Mean</td>
<td>100.42</td>
<td>100.42</td>
<td>100.42</td>
<td>100.42</td>
<td>100.42</td>
<td>100.42</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>113.09</td>
<td>107.82</td>
<td>107.82</td>
<td>106.86</td>
<td>105.87</td>
<td>107.44</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.90</td>
<td>10.76</td>
<td>10.76</td>
<td>8.40</td>
<td>8.69</td>
<td>8.71</td>
</tr>
<tr>
<td>R-Squared</td>
<td>-2.21</td>
<td>-0.93</td>
<td>-0.93</td>
<td>-0.17</td>
<td>-0.26</td>
<td>-0.26</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>-0.79</td>
<td>-0.39</td>
<td>-0.39</td>
<td>-0.08</td>
<td>-0.12</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

Table 8.5 displays Goodness of Fit data for EE03 MLR and LOWESS Models for the period W088091. It shows that the $R^2_{\text{PRED}}$ and Matlab Fit values were negative for the period indicating a loss of predictive capability for the models. However, looking at the trends of Figures 8.6 to 8.8 the model approximations to the measured signal, while not a match, appear to be a reasonably good fit. However, with RMSE figures greater than 8.00 across the board, achieving satisfactory GOF scores was unlikely and this is borne out in the GOF data.
Table 8.6 displays correlation coefficients for each of the EE03 MLR Models, L0 to L4, and the LOWESS Model against EE03 measured vibration data. In this case the correlation coefficient values do not complement the GOF data, as was the case for TE09, in terms of best and worst performing models. The average correlation coefficient value across the six models for EE03 was 0.817. This value is comparable to the values calculated for TE09, where all GOF scores were positive for W088091 (the average correlation coefficient values across the six models for TE09 was 0.854).

Considering the trend data for both TE09 and EE03, it would appear that both should perform similarly well in terms of GOF scores etc. but this was not the case as shown in Figures 8.1 and 8.5 respectively. One reason behind the relatively large difference in GOF performance was a slight positive bias that the models have developed over the measured signal. This bias led to the mean values for the predicted series’ being significantly higher than the measured series mean, particularly for the L0 model, although for the L4 model the result was reasonable.

Another reason behind the poorer GOF performance of the EE03 model compared with the TE09 models relates to overall variance of the measured signal i.e. for EE03 for the period W088091 the maximum measured vibration value for the period was 129.23 µm pk-pk and minimum value is 81.90 µm pk-pk giving a range of 47.33 µm pk-pk whereas for TE09 the maximum recorded value is 154.58 µm pk-pk, the minimum value was 59.63 µm pk-pk thereby gave a range of 94.95 µm pk-pk. For GOF statistics such as $R^2$ and the Matlab Fit statistics used in this document where the calculation of Goodness of Fit is based around
deviance from the mean measured series value, having a narrow range means that for the same Mean value the effect of any error is magnified. For instance a 3 µm pk-pk prediction error on a 120 µm pk-pk measured will have a higher GOF score, the further from the mean the error occurs. This is the reason why, for what would seem like a similar result in terms of prediction for TE09 and EE03 when looking at the trend or correlation data, the GOF values were so poor for the EE03 models relative to the models for TE09.

### Table 8.7 Goodness of Fit Statistics for All End Coils from W032035 to W096099 (07/07/2007 to 20/10/2008) for the Best Performing MLR Models and the LOWESS Model.

| Chapter 8 | Page 176 | Performance of Regression Models |
Table 8.7 shows Goodness of Fit statistics (RMSE and R-Squared only) over the period W032 to W099 (02/07/2007 to 20/10/2008) for the best performing MLR model (as defined in Section 6.10) and LOWESS Model related to each of the twelve end coils where vibration was measured.

From Table 8.7 it is evident that predictive capability was good across a majority of the end coils with R² values averaging better than 0.7 for most points of measurement.

Models relating to EE03 perform well over the first 3 to 6 months after which the GOF scores (R-Squared) fall away from an average of approximately 0.6 to approximately 0.35. From W088091 onwards the GOF scores for both the MLR and LOWESS models turned negative and did not recover.

Models for TE25 performed even more poorly, averaging at approximately 0.45 from W032035 to W056059 with GOF scores after that being very low or negative over the remainder of the period covered, with the exception of W088091 and W092095 where the GOF scores recovered to approximately 0.45 before returning to negative figures for W096099.

Vibration measured at TE25 is of particular interest because damage had occurred at this location previously; the same is true of TE33. More information on the reasons for poor performance of MLR and LOWESS models for TE25 is presented in Chapter 9.
8.2 Requirements for Recalculation of Models

8.2.1 2008 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

Models were recalculated following the 2008 Annual Overhaul. This operation was carried out for two reasons:

1. Due to natural degradation of model predictive capability (MLR and LOWESS).
2. The timing was opportune. The models could be recalculated in the knowledge that the machine had been electrically and mechanically inspected and tested, and was found to be in excellent condition.

Minor issues with the regression models over the 44 weeks they were used included:

1. Loss of accuracy due to the machine being heavily cycled for long periods. This data was filtered from data used to calculate descriptive statistics etc.
2. Loss of accuracy due to low Ambient Temperature, which in turn affects Generator Cooling Water Temperature, Generator Cold Gas Temperature etc. Again this data was filtered from data used to calculate descriptive statistics etc.
3. Natural degradation of the model predictive capability of some models (TE17, TE25 and EE03 particularly) which began to become problematic after approximately 24 to 28 weeks of implementation.

The poor predictive capability of models or their speed of degradation in some cases is down to the calculation of models not being optimal, while in other cases it can be down to a lack of significant variables to explain variation in the dependent variables (this is the case for TE25 in particular, more information is available in Chapter 9).

More information on the performance of the 2008 MLR and LOWESS models for each vibration signal is available in Appendix 5.1.
8.2.2 2009 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

Model coefficients were recalculated post AOH 2009 again due to degradation of the predictive accuracy of the models. A number of issues arose over the period of service of the 2009 models namely:

1. Database calculation error of on the Average Stator Current calculated variable in PI. The miscalculation led to minor differences in the resultant calculations in comparison with calculations carried out manually. However even though the discrepancy was relatively minor, for models sensitive to Average Stator Current (large model coefficient) the effect on predicted values was profound. The error was corrected and predicted values returned to normal.

2. Extremely cold weather in the period October 2009 to January 2010 resulted in loss of model predictive accuracy, particularly for vibration signals where models were heavily dependent on temperature variables for prediction.

3. For the LOWESS Model the reduced number of observations included in the construction data set and the maintenance of a 2 % ‘True’ setting (the size of the subset of data to which the low order polynomial is applied at a time) led to poor model accuracy. This setting had to be changed in order to develop accurate models. The best model was found by trial and error by gradually increasing the True value. Predictive accuracy will improve with increasing true value setting up to a point and will then drop-off again. For this case it was found that the optimal setting was 6% (141 observations).

More information on the performance of the 2009 MLR and LOWESS models for each vibration signal is available in Appendix 5.2.
8.2.3 2009/2010 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

The model solutions were recalculated after a reduction in the minimum stable generation capability of the Gas Turbine from approximately 143 MW to 132 MW approximately (this was subsequently further reduced to 114 MW).

It was found that relationships between variables (mainly temperature related) at loads below 140MW were not linear, or were linear only under certain ambient conditions. This required the data for loads below the original Min Gen value to be filtered from both the construction data and the data applied to the model.

The filtering of data must go beyond the time of instruction below the original Min Gen as once instructed above that load level it still takes some time for the system temperatures to return to normal (giving linear relationships between variables).

More information on the performance of the 2009/2010 MLR and LOWESS models for each vibration signal is available in Appendix 5.2.

Nearing the Annual Overhaul of 2010 it was becoming clear that some innovation was required with regard to how model solutions were being calculated. It was hoped that new plant variables data which were to be made available on the PI database, would provide some statistically significant independent variables, particularly for modelling of vibration signals which had proven difficult to model accurately up to that point. The development of the new MLR and LOWESS models from new independent variable data are discussed in Chapter 9.
CHAPTER 9

IMPROVEMENT IN MODEL DESIGN DUE TO

AVAILABILITY OF ADDITIONAL INDEPENDENT VARIABLES

From the commencement of this project the running regime of the Power Station changed gradually from a predominantly base load generator with some two-shifting during the summer months, to the situation at the time of writing, where the machine could be stopped and restarted 4 - 5 times per week, be called upon to provide power for 4 - 5 hours during evening peak demand times or be instructed off for weeks at a time particularly during the summer months. Utilisation Factor (UF) for the Power Station has dropped significantly over the period covered by this project but the most acute drop occurred from 2010 onwards. The resulting relatively low Utilisation Factor came about in three ways, firstly by being instructed off for extended periods, secondly by being used as a large peaking plant for 4-5 hours per day and thirdly, being operated at part load almost exclusively whilst instructed online by Eirgrid NCC.

These issues have had a uniformly negative effect on the ability of the MLR and LOWESS models to accurately predict end-coil vibration. Start-up and shut down data had to be filtered out as transient behaviour at these times cannot be predicted accurately by the MLR or LOWESS models. After a start, there is a period of time, depending upon how long the machine has been instructed offline, where the end-coil vibration/temperature relationships of the generator are abnormal. It can take a number of hours for all temperatures to settle into their normal operating ranges; this data must also be filtered out when calculating goodness of fit statistics.

To improve the predictive capability if the MLR and LOWESS models, some innovation was required with regard to calculation of the models. New plant measured variables were made available to PI and by extension to the independent variable selection process of the various models. Some differential temperature values were added to the list of independent variables.
made available for selection. Measured vibration signals from other end-coils (i.e. the eleven end-coils other than the one for which the model was being generated) were also made available for selection as independent variables should they prove statistically significant.

9.1 Construction Data for 2010 Models

New Multiple Linear Regression (MLR) and LOWESS models were calculated using construction data made up of 1978 data points (3 weeks) taken from W168 (08/02/2010 – 15/02/2010), 177 (12/04/2010 – 19/04/2010) and W194 (09/08/2010 – 16/08/2010). Taking data from non-sequential weeks was done in order to ensure that as wide a range of operating and ambient conditions as possible were included in the construction data set. No start-up or shut down data was included in the construction data. For models created from 2010 onwards data was filtered for GT Load >130 MW.

It should be noted that the number of data points included in a typical learning data set was, over time, reduced. This change was made to speed up calculation of the model coefficients etc. which proved quite time consuming where the learning data set was large. This is particularly true for the LOWESS procedure.

9.2 New Measured and Calculated Independent Variables Made Available for Selection

A number of new independent variables, both measured and calculated were introduced for the 2010 Models. The new measured variables added were as follows (Reference designations are shown in brackets):

GT Closed Cooling Water Outlet Temperature (CCCWT)

This is the temperature of the outlet/supply side of the Gas Turbine Closed Cooling Water System feeding the GT Generator and Lube Oil/Jacking Oil heat exchangers of the Gas Turbine.
Average Generator Cold Gas Temperature (ACAT)

The Average Temperature of the Supply Cooling Air to the GT Generator. This figure is made up of six temperature measurements, three from the exciter end and three from the turbine end of the machine. The cooling system is set up such that there is no significant difference in the Cold Air Supply Temperatures at either end of the machine. The temperature differential is usually < 1 °C.

Hot Air Temperature (HAT)

The temperature of the hot/return cooling air as it leaves the GT Generator.

Differential Temperatures

The new independent variables created using differential temperatures and made available to the Stepwise selection process are as follows (Reference designations are shown in brackets):

1. Average Stator Slot Temperature - Hot Air Temperature (SLOTT_HAT)
2. Average Stator Slot Temperature - Average Cold Air Temperature (SLOTT_ACAT)
3. Average Stator Slot Temperature - Closed Circuit Cooling Water Outlet Temp (SLOTT_CCCWT)
4. Average Stator Slot Temperature - Ambient Air Temperature (SLOTT_AMBT)
5. Hot Air Temperature - Average Cold Air Temperature (HAT_ACAT)
6. Hot Air Temperature - Closed Circuit Cooling Water Outlet Temperature (HAT_CCCWT)
7. Hot Air Temperature - Ambient Air Temperature (HAT_AMBT)
8. Average Cold Air Temperature - Closed Circuit Cooling Water Outlet Temperature (ACAT_CCCWT)
9. Average Cold Air Temperature - Ambient Air Temperature (HAT_AMBT)
10. Closed Circuit Cooling Water Outlet Temperature - Ambient Air Temperature (CCCTW_AMBT)
It was expected that some of these variables would prove significant enough statistically to be included in at least some of the Stepwise Models.

FOA measured values were added to the list of independent variables available for selection. Each Model had the data from the eleven FOAs (not currently being modelled) made available for selection should they prove to be statistically significant.

### 9.3 New Model Independent Variable Arrays/Model Designations

New arrays of independent variables were created to replace the L0 to L4 arrays used earlier in this document. In some of the arrays, the eleven measured vibration signals other than the one being modelled, were made available as independent variables to the model, the model construction data arrays were therefore specific to each dependent variable.

Fourteen Independent Variable Data Arrays were created for each measured vibration signal using varying numbers of independent variables, varying levels of interaction etc. MLR models were generated from each of the fourteen independent variable arrays for each of the 12 measured vibration signals. Their performance was assessed and compared against measured vibration data so that the best possible model could be generated.

The revised Regression Model/Independent Variable Array designations break down was as follows:

1. A two letter designation relating to the end of the generator to which the Regression Model/Independent Variable Array refers (TE or EE).
2. The End-Coil Number at the end of the generator under consideration
3. An alphanumeric reference to the Interaction level of the model (L0, L1 or L2).
4. FI denotes Full Interaction i.e. Interactions between all measured independent variables, including the eleven eligible vibration signals are included in, and available for selection from the Independent Variable Array.
5. Reference to inclusion of FOA Measured Vibration Data as Independent Variables. ‘FOA’ is added to the model designation if the measured vibration independent variables are included.
6. If the FOA designation is followed by OSO then vibration signal independent variables from the same end of the machine as the signal being modelled are included e.g. if TE09 is being modelled then TE01, TE17, TE25, TE33 and TE41 are made available for inclusion in the model from the independent variable array (no Exciter End vibration data is included).

The following are independent variable arrays developed for use in MLR models. Data relating to MLR Models generated from these independent variable arrays are presented for TE09 in the next section of this Chapter (9.4).

**TE/EE## L0**
The original ten independent variables plus the three new measured temperature values plus the newly created differential temperature variables only (L0 = 23 independent variables)

**TE/EE## L0FOA**
L0 plus the eleven eligible FOA data series (L0FOA = 34 independent variables)

**TE/EE## L1**
Independent variables made available for selection for this model were; the original ten independent variables, plus the three new measured temperature variables, plus the newly created differential temperature variables, plus first level interactions of the thirteen measured independent variables (this included squared values of each of the 13 measured independent variables). (L1 = 114 independent variables)

**TE/EE## L1FOA**
For this model the independent variables made available for selection by the Stepwise selection process were; all of those included in L1, plus the eleven eligible FOA variables, plus the squared values of the eleven eligible FOA variables. (L1FOA = 136 independent variables)

**TE/EE## L2**
The independent variables available for selection for this model were; the original ten independent variables, plus the three new measured temperature variables, plus the
differential temperature variables, plus first level interactions of the thirteen measured independent variables, plus second level interactions of the thirteen measured independent variables. \((L^2 = 493\) independent variables\)

**TE/EE## L2FOA**

The independent variables made available for selection in the L2FOA models were; all of those included in the L2 model independent variable data set, plus the eleven eligible FOA variables, plus the squared values of the eleven eligible FOA variables, plus the cubed values of the eleven eligible FOA variables. \((L2FOA = 502\) independent variables\)

A number of models were created to allow an assessment of the improvement or otherwise (relative to other models of the same type e.g. MLR) in predictive capability of models where more independent variables were made available for selection. Creating large amounts of new variables can bring problems with multicollinearity between explanatory variables; this again can lead to less than optimal model variable selection.

LOWESS Models were also developed using measured variables only i.e. the \(L0\) data, or a selection of variables therefrom. LOWESS models were generated for various numbers and combinations of independent variables until an optimal model for each vibration signal was found i.e. best accuracy with fewest independent variables included. For all vibration signals on both ends of the machine, the LOWESS models performed better with the eligible vibration signal data made available to the model as independent variables.

**9.4 Predictive Performance of New MLR and LOWESS Models**

In this section data relating to the performance of the various models is presented. From Table 9.1 it is evident that an increase in the number of independent variables made available to the stepwise selection process did not necessarily lead to the generation of a more robust MLR model of the system. In this case, for TE01, the best performing MLR model was the L1FOA model which had an average \(R^2\)-squared value over the entire period of 0.858. The LOWESS Model also performed well over the period with an average \(R^2\)-Squared value of 0.881. The LOWESS model however, achieved the high average \(R^2\)-squared values while having a
significantly lower RMSE value, averaging 14.75 μm pk-pk over the ten months approximately, compared with 15.88 μm pk-pk for TE01L2FOA.

| Goodness of Fit Statistics TE01 from W195199 to W232235 (16/08/2010 to 30/05/2011) |
|---------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| TE01 L0                         |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 121.99 | 112.91 | 112.79 | 109.89 | 105.61 | 118.74 | 115.51 | 120.23 |
| RMSE L0                         | 18.37  | 31.89  | 29.62  | 32.13  | 31.72  | 24.61  | 28.59  | 31.56  |
| TE09 L0 R Squared               | 0.55   | 0.36   | 0.26   | 0.25   | 0.21   | 0.41   | 0.33   | 0.34   |
| Matlab Fit                      | 0.80   | 0.59   | 0.45   | 0.44   | 0.37   | 0.65   | 0.55   | 0.56   |

| TE01 L0 FOA                     |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 121.36 | 105.62 | 103.56 | 98.97  | 93.27  | 109.93 | 102.84 | 108.87 |
| RMSE L1                         | 14.40  | 31.66  | 30.96  | 32.37  | 37.11  | 22.42  | 31.33  | 30.76  |
| TE09 L1 R Squared               | 0.65   | 0.37   | 0.27   | 0.25   | 0.23   | 0.46   | 0.26   | 0.35   |
| Matlab Fit                      | 0.88   | 0.60   | 0.46   | 0.43   | 0.41   | 0.71   | 0.45   | 0.58   |

| TE01 L1                         |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 126.69 | 108.61 | 110.11 | 106.68 | 108.82 | 125.06 | 122.13 | 120.68 |
| RMSE L2                         | 19.39  | 33.67  | 30.78  | 30.20  | 33.59  | 24.86  | 30.00  | 30.15  |
| TE09 L2 R Squared               | 0.53   | 0.31   | 0.23   | 0.30   | 0.16   | 0.40   | 0.29   | 0.36   |
| Matlab Fit                      | 0.78   | 0.52   | 0.41   | 0.51   | 0.30   | 0.64   | 0.50   | 0.60   |

| TE01 L1 FOA                     |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 126.69 | 113.89 | 110.26 | 103.70 | 102.00 | 114.63 | 109.43 | 115.11 |
| TE09 L3 R Squared               | 0.77   | 0.69   | 0.58   | 0.54   | 0.50   | 0.70   | 0.55   | 0.61   |
| Matlab Fit                      | 0.95   | 0.90   | 0.82   | 0.79   | 0.75   | 0.91   | 0.80   | 0.85   |

| TE01 L2                         |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 121.57 | 111.36 | 115.55 | 111.21 | 113.05 | 128.75 | 127.92 | 130.82 |
| RMSE L4                         | 18.00  | 32.39  | 27.97  | 29.28  | 30.38  | 25.52  | 29.84  | 29.61  |
| TE09 L4 R Squared               | 0.56   | 0.35   | 0.30   | 0.32   | 0.24   | 0.38   | 0.30   | 0.38   |
| Matlab Fit                      | 0.81   | 0.58   | 0.51   | 0.54   | 0.42   | 0.62   | 0.50   | 0.61   |

| TE01 L2 FOA                     |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 124.39 | 117.18 | 110.48 | 98.95  | 92.04  | 111.33 | 104.78 | 108.95 |
| RMSE LOESS                      | 11.74  | 11.43  | 27.76  | 29.36  | 35.10  | 18.94  | 28.30  | 27.78  |
| Matlab Fit                      | 0.71   | 0.74   | 0.31   | 0.32   | 0.27   | 0.54   | 0.33   | 0.41   |
| TE09 L2 R Squared               | 0.92   | 0.93   | 0.52   | 0.53   | 0.47   | 0.79   | 0.55   | 0.66   |

| TE01 LOESS FOA                  |        |        |        |        |        |        |        |        |
| Measured Mean                   | 125.84 | 119.68 | 118.83 | 114.18 | 109.28 | 117.69 | 118.60 | 124.71 |
| Predicted Mean                  | 126.05 | 114.94 | 112.32 | 107.59 | 104.49 | 113.48 | 114.94 | 119.23 |
| Matlab Fit                      | 0.79   | 0.74   | 0.64   | 0.66   | 0.64   | 0.66   | 0.61   | 0.70   |
| TE09 LOESS R Squared            | 0.96   | 0.93   | 0.87   | 0.88   | 0.87   | 0.88   | 0.85   | 0.91   |

Table 9.1 Goodness of Fit Statistics for TE01 from W195199 to W232235 for MLR Models L0, L0FOA, L1, L1 FOA, L2 and L2 FOA as well as a LOESS Model with FOA Data Included in Independent Variable Data.
9.4.1 Assessment of Models using Trend Data

Figures 9.1 to 9.3 display trend data for TE09 Measured and MLR Model Predicted vibration for the period W195 to W199 (16/08/2010 to 09/09/2010). Average Stator Core Temperature and Ambient Air Temperature data are also displayed. From Figure 9.1 it is evident that the model with measured vibration data made available as independent variables (TE09L0FOA) outperformed the model which does not include those variables (TE09L0). It is apparent that the variance in the measured signal was matched almost exactly by the TE09L0FOA MLR Model predicted values. For W195199 the $R^2$ value for the L0 model was 0.89 with a RMSE value of 8.04. For the TE09L0FOA model the $R^2$ value was 0.92 with the RMSE figure being 6.99.

From Figure 9.2 it is evident that the L1 models gave a better approximation of the measured signal than the L0 models shown in Figure 9.1. For both the L0 and L1 models, the models with the measured vibration data made available as independent variables (TE09L0FOA and TE09L1FIFOA) outperformed the models which did not include those variables (TE09L0 and...
TE09L1). It is apparent that the variance in the measured signal was matched almost exactly by the TE09L1FIFOA MLR Model predicted values shown in green in Figure 9.2. For W195199 the $R^2$ value for the TE09L1 model was 0.876 with a RMSE value of 8.723, while for the TE09L1FIFOA model the $R^2$ value was 0.956 with a RMSE figure of 5.172.

![Figure 9.2 TE09 Measured and MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature W195199 (16/08/2010 to 09/09/2010).](image)

In Figure 9.3 it is notable that the L2 model gives quite a good approximation of the measured signal. However, as was the case for the L0 and L1 models, the model with the measured vibration data made available as independent variables (TE09L2FOA) outperformed the model which did not include those variables (TE09L2). It is apparent that the variance in the measured signal was matched almost perfectly by the TE09L2FOA MLR Model predicted values. For W195199 the $R^2$ value for the TE09L2 MLR model was 0.892 with an associated RMSE value of 8.16. The $R^2$ value for the TE09L2FOA MLR model was 0.968 with a RMSE figure of 4.402.

The results presented here for TE09 are also typical of the performance of the regression models for the remaining eleven vibration signals. In general the addition of FOA data to the
list of independent variables available for selection to the stepwise process has led to a marked improvement in the quality of models generated.

The main issue encountered with the revised 2010 models, as with the 2009 and 2009/2010 models was a fall-off in Goodness of Fit (GOF) scores of all models during periods of very low ambient temperature.

![Figure 9.3 TE09 Measured and MLR (Best L2 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature W195199 (16/08/2010 to 09/09/2010).](image)

The reduced predictive capability at low ambient temperature was caused not by a deficiency in the predictive capability of the model however, under these conditions the correlation between dependent variables and independent variables as well as between dependent variables and other dependent variables changed significantly. The solution to the problem, as presented later in this chapter, was to filter out data related to periods where relationships between variables become non-linear. For MLR models, if the assumption of linearity is not met predictions are therefore inaccurate and so goodness of fit scores are meaningless.
9.4.2 Assessment of Models using Pearson Correlation Data

Pearson Correlation Tables for measured vibration and plant variables are presented in this section. Correlation tables are presented for the Construction Data Set and W208211 (10/01/2011 to 07/02/2011). W208211 being a period of extremely harsh weather, thus giving a comparison of correlation coefficients calculated under normal and abnormal conditions respectively. Further tables are presented with various filters applied to show the effect of removing “bad” data.

<table>
<thead>
<tr>
<th>Variables</th>
<th>TE01</th>
<th>TE09</th>
<th>TE17</th>
<th>TE25</th>
<th>TP59</th>
<th>TE63</th>
<th>TE64</th>
<th>EE11</th>
<th>EE19</th>
<th>EE27</th>
<th>EE33</th>
<th>EE40</th>
<th>EE41</th>
<th>EE58</th>
<th>MV01</th>
<th>MV03</th>
<th>MV05</th>
<th>MV07</th>
<th>MV09</th>
<th>MV11</th>
<th>MV13</th>
<th>MV15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Power (MW)</td>
<td>0.83</td>
<td>0.84</td>
<td>0.44</td>
<td>0.80</td>
<td>0.00</td>
<td>0.87</td>
<td>0.79</td>
<td>0.76</td>
<td>0.44</td>
<td>1.00</td>
<td>0.65</td>
<td>0.59</td>
<td>1.00</td>
<td>0.64</td>
<td>0.51</td>
<td>0.45</td>
<td>0.59</td>
<td>0.55</td>
<td>0.55</td>
<td>0.56</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>Reactive Power (MVAR)</td>
<td>0.49</td>
<td>0.38</td>
<td>0.40</td>
<td>0.70</td>
<td>0.71</td>
<td>0.46</td>
<td>0.46</td>
<td>0.70</td>
<td>0.00</td>
<td>1.00</td>
<td>0.64</td>
<td>0.45</td>
<td>0.70</td>
<td>0.60</td>
<td>0.76</td>
<td>0.76</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.71</td>
<td>0.69</td>
<td>0.69</td>
</tr>
<tr>
<td>KVG Actual Position (%)</td>
<td>0.86</td>
<td>0.88</td>
<td>0.40</td>
<td>0.70</td>
<td>0.71</td>
<td>0.46</td>
<td>0.72</td>
<td>0.79</td>
<td>0.76</td>
<td>0.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Apparent Power (MVAR)</td>
<td>0.84</td>
<td>0.85</td>
<td>0.42</td>
<td>0.70</td>
<td>0.61</td>
<td>0.38</td>
<td>0.47</td>
<td>0.81</td>
<td>0.77</td>
<td>0.44</td>
<td>1.00</td>
<td>0.61</td>
<td>0.95</td>
<td>0.90</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>Average Slot Temp (°C)</td>
<td>0.53</td>
<td>0.55</td>
<td>0.42</td>
<td>0.70</td>
<td>0.61</td>
<td>0.46</td>
<td>0.47</td>
<td>0.70</td>
<td>0.80</td>
<td>0.54</td>
<td>0.80</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Average Motor Current (A)</td>
<td>0.84</td>
<td>0.84</td>
<td>0.42</td>
<td>0.70</td>
<td>0.61</td>
<td>0.46</td>
<td>0.47</td>
<td>0.70</td>
<td>0.80</td>
<td>0.54</td>
<td>0.80</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Exciter Current (A)</td>
<td>0.77</td>
<td>0.78</td>
<td>0.48</td>
<td>0.70</td>
<td>0.61</td>
<td>0.38</td>
<td>0.47</td>
<td>0.70</td>
<td>0.80</td>
<td>0.54</td>
<td>0.80</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>Ambient Air Temp (°C)</td>
<td>0.93</td>
<td>0.97</td>
<td>0.52</td>
<td>0.70</td>
<td>0.58</td>
<td>0.36</td>
<td>0.71</td>
<td>0.67</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
<td>0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>Average Motor Voltage (V)</td>
<td>0.25</td>
<td>0.26</td>
<td>0.08</td>
<td>0.56</td>
<td>0.32</td>
<td>0.20</td>
<td>0.10</td>
<td>0.10</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Frequency (Hz)</td>
<td>0.04</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>CCOW Outlet Temp (°C)</td>
<td>0.23</td>
<td>0.28</td>
<td>0.08</td>
<td>0.10</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Average Cold AirTemp (°C)</td>
<td>0.41</td>
<td>0.50</td>
<td>0.10</td>
<td>0.12</td>
<td>0.04</td>
<td>0.25</td>
<td>0.32</td>
<td>0.55</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Net Air Temp (°C)</td>
<td>0.66</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Table 9.2 Pearson Correlation Table for 2010 MLR and LOWESS Model Construction/Learning Data:

Measured Variables

Table 9.2 shows Pearson Correlation coefficient data relating to the Construction/Learning data measured variables. The column related to End Coil TE01 is highlighted in green. It is immediately evident that there were strong correlations between TE01 and independent variables such as Active Power, Average Slot Temperature and Exciter Current which did not exist in the correlation coefficient for W208211 (Table 9.3). Correlations between the end coil vibration series were generally also stronger for the construction data than for data from W208211.
Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

Chapter 9

Page 192

Improvement in Model Design

Table 9.3 Pearson Correlation Table for W208211 (10/01/2011 to 07/02/2011) Unfiltered.

<table>
<thead>
<tr>
<th>Variables</th>
<th>TE01</th>
<th>TE09</th>
<th>TE17</th>
<th>TE25</th>
<th>TE33</th>
<th>TE53</th>
<th>TE57</th>
<th>TE61</th>
<th>TE63</th>
<th>EE03</th>
<th>EE19</th>
<th>EE43</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE01</td>
<td>1.00</td>
<td>0.49</td>
<td>0.48</td>
<td>-0.21</td>
<td>0.03</td>
<td>0.64</td>
<td>0.44</td>
<td>0.01</td>
<td>0.47</td>
<td>0.38</td>
<td>0.03</td>
<td>0.54</td>
</tr>
<tr>
<td>TE09</td>
<td>0.49</td>
<td>1.00</td>
<td>-0.74</td>
<td>-0.53</td>
<td>-0.79</td>
<td>-0.19</td>
<td>0.35</td>
<td>-0.17</td>
<td>0.83</td>
<td>0.60</td>
<td>0.58</td>
<td>0.03</td>
</tr>
<tr>
<td>TE17</td>
<td>0.48</td>
<td>-0.74</td>
<td>1.00</td>
<td>0.72</td>
<td>0.77</td>
<td>0.17</td>
<td>-0.43</td>
<td>0.01</td>
<td>-0.75</td>
<td>-0.73</td>
<td>-0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>TE25</td>
<td>0.21</td>
<td>0.53</td>
<td>0.72</td>
<td>1.00</td>
<td>0.67</td>
<td>0.38</td>
<td>0.03</td>
<td>0.32</td>
<td>-0.57</td>
<td>-0.50</td>
<td>-0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>TE33</td>
<td>0.03</td>
<td>0.75</td>
<td>0.77</td>
<td>0.67</td>
<td>1.00</td>
<td>0.67</td>
<td>-0.22</td>
<td>0.26</td>
<td>-0.48</td>
<td>-0.63</td>
<td>-0.18</td>
<td>-0.24</td>
</tr>
<tr>
<td>TE53</td>
<td>0.64</td>
<td>0.19</td>
<td>0.17</td>
<td>0.58</td>
<td>0.67</td>
<td>1.00</td>
<td>0.20</td>
<td>0.24</td>
<td>-0.06</td>
<td>-0.11</td>
<td>-0.15</td>
<td>-0.85</td>
</tr>
<tr>
<td>TE57</td>
<td>0.44</td>
<td>0.35</td>
<td>0.43</td>
<td>0.01</td>
<td>0.22</td>
<td>0.29</td>
<td>1.00</td>
<td>0.93</td>
<td>0.36</td>
<td>0.15</td>
<td>0.35</td>
<td>0.15</td>
</tr>
<tr>
<td>EE03</td>
<td>1.01</td>
<td>0.17</td>
<td>0.26</td>
<td>0.52</td>
<td>0.28</td>
<td>0.24</td>
<td>0.39</td>
<td>1.00</td>
<td>0.04</td>
<td>0.33</td>
<td>0.57</td>
<td>-0.14</td>
</tr>
<tr>
<td>EE19</td>
<td>0.47</td>
<td>0.83</td>
<td>0.75</td>
<td>0.57</td>
<td>0.68</td>
<td>0.06</td>
<td>0.06</td>
<td>0.04</td>
<td>1.00</td>
<td>0.75</td>
<td>0.65</td>
<td>-0.16</td>
</tr>
<tr>
<td>EE43</td>
<td>0.38</td>
<td>0.60</td>
<td>0.75</td>
<td>0.30</td>
<td>0.63</td>
<td>-0.11</td>
<td>0.56</td>
<td>0.33</td>
<td>0.75</td>
<td>1.00</td>
<td>0.27</td>
<td>0.01</td>
</tr>
<tr>
<td>Rotor Freq (Hz)</td>
<td>0.03</td>
<td>0.58</td>
<td>0.34</td>
<td>-0.29</td>
<td>-0.58</td>
<td>-0.31</td>
<td>0.05</td>
<td>0.37</td>
<td>0.65</td>
<td>0.97</td>
<td>1.00</td>
<td>0.36</td>
</tr>
<tr>
<td>Ambient Temp (°C)</td>
<td>0.54</td>
<td>0.03</td>
<td>0.28</td>
<td>0.20</td>
<td>-0.24</td>
<td>-0.65</td>
<td>-0.15</td>
<td>-0.04</td>
<td>0.16</td>
<td>0.01</td>
<td>0.36</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The Pearson Correlation Table 9.3 displays correlation coefficients for all measured variables used in the MLR and LOWESS models for the period W208211 (10/01/2011 to 07/02/2011). Ambient conditions over the period in question were particularly harsh with temperatures well below 0°C for prolonged periods. The column related to TE01 is highlighted in red. It is immediately evident that the coefficient values were almost uniformly very low (<0.20).

Correlation values for some vibration signals (TE09, TE17, TE41, EE03, EE19 and EE43) remained reasonably strong.

Correlation values for TE09, EE03 and EE19 remain positive but are reduced in strength by on average 37 % for W208211, before any filtering of data is carried out. The deterioration in the correlation values between TE01 and these end coils proved to be temperature sensitive and so with judicious filtering the correlation values recovered to near learning data values for the correlation table of the filtered data shown in Table 9.4.

Table 9.4 displays Pearson correlation coefficients for all 25 measured variables for W208211 with a filter applied for Average Slot Temperature ≥60 °C and Ambient Air Temperature ≥4 °C. The column related to TE01 is highlighted in green.
Improvement in Model Design

Table 9.4 Pearson Correlation Table for W208211.

Data filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C.

Note: Figures shown in bold, with green shaded cells, are different from zero with a significance level alpha = 0.05.

Correlation values between the variables other than the measured vibration variables remained stable throughout, whether a filter was applied or not. Taking Active Power as an example, it can be seen that the correlations between Active Power and most of the independent variables shown were almost exactly the same (with the exception of Frequency which was not statistically significant from zero in any case) for W208211 as for the learning data. However, with the filter applied, some of the correlation values (particularly related to temperature variables) were diminished, so care was required when carrying out such an operation.
9.4.3 Assessment of Models using Goodness of Fit Statistics

This section shows goodness of fit data for 2010 models using TE09 as an example. Statistics are presented with the data both unfiltered and filtered for certain conditions. W208211 is used as an example as conditions experienced during that time were extreme and as such required the analyst to spend more time filtering and analysing data than would normally be the case. Goodness Of Fit (GOF) data is presented for six of the best performing MLR models, (two L0, two L1 and two L2 models) and the LOWESS model. The definition of ‘best performing model’ for the purposes of this project is specified in Section 6.10. Some trend data for TE09 models is also shown for the same period for visual comparison of the performance of the various models. Finally, a table of goodness of fit statistics for the best performing MLR and LOWESS model for each end coil over the forty weeks from W195 to W235 (16/08/2010 to 30/05/2011) are presented.

Table 9.5 shows Goodness of Fit Statistic values for W208211 with three filters applied to the data. The first filter applied was for the data outside of the first 500 recorded values thus removing data related to the period of extreme weather completely. The second was for Average Slot Temperature (SLOTT) ≥60 °C and Ambient Air Temperature ≥4 °C thus removing data related to the coolest times of day or where generator temperatures are very low. The third filter was for Average Slot Temperature (SLOTT) ≥60 °C and Ambient Air Temperature ≥6 °C which further filters the data for low ambient temperature and indeed had the effect of creating a very similar set of data to that of the first filter. For TE09 the L1FOA model was the best performing MLR model and for W208211 was the most consistent and best performing model (as defined in Section 6.10) with a filter of any kind applied, but only ranked third behind the LOWESS model and the L2 FOA model for the unfiltered data. Taking just the MLR models into account, there was a relatively insignificant difference in performance between the L1FOA and L2FOA models in terms of goodness of fit scores. There was a very noticeable difference in the performance of models where vibration data has been added to the pool of independent variables against models where vibration data was not made available. The superior performance of the FOA models was particularly clear with the data filters applied.
The filter for Average Slot Temperature (SLOTT) ≥60 °C and Ambient Air Temperature ≥4 °C was chosen as the standard filter for application during periods of cold weather. It was decided that a ≥6 °C filter on Ambient Air Temperature removed too much sample data and so should not be used.
Figure 9.4 TE09 Measured and MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature (Unfiltered) W208211 (15/11/2010 to 11/12/2010)

Figure 9.4 shows TE09 measured vibration against the best performing (as defined in Section 6.10) L1 MLR Models predicted vibration values (on the primary vertical axis) as well as Average Slot Temperature and Ambient Air Temperature (on the secondary vertical axis). It is evident from the plot that for this period the variance in the measured signal (TE09) was not approximated particularly well by L1 MLR Models. The deterioration in model predictive capability corresponded to Ambient Temperature falling below 0°C for prolonged periods (between 25/11/2010 and 10/12/2010). Before and after the cold weather the models performed satisfactorily.

With filters applied to predictive data for periods of very low ambient air temperature, the problem of inaccurate predictions was reduced significantly. This is easily observed for the TE09 filtered data for W208211 displayed in Figure 9.5 where a filter has been applied for Average Stator Slot Temperature >60 °C and Ambient Temperature >4 °C. Both the MLR model (TE09L1FIFOA) and the LOWESS model (TE09 LOWESS FOAOSO) generate relatively good predictions for the measured values. Both models however failed to predict
values at the lower end of the range with the same accuracy as at the upper end, although this does correspond to ambient temperature averaging below 5 °C from approximately point 320 onwards. Similar results were recorded for the L0 and L2 models.

![Figure 9.5 TE09 Measured Vibration, MLR (Best L1 Models) Predicted Vibration, Average Slot Temperature and Ambient Air Temperature for W208211. Data Filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C](image)

It was noted that sharp changes in measured vibration amplitudes at the Turbine End Coils during periods of cold weather were not matched by the data from the Exciter End. It was thought therefore, that for the Models where FOA data was made available as independent variable data the effect of having an FOA independent variable from the Turbine End included in a model for an Exciter End End-Coil (or vice versa) might be to adversely affect the predictive capability of the model under these conditions. This led to the development of the independent variable arrays which used vibration data from the same end of the generator only, as the signal being modelled (referenced OSO).

The problem of inaccurate prediction at low ambient temperature cannot be solved as such given the constraints of a linear model. In order that the models were valid, a filter had to be applied to the prediction data to remove data pertaining to Ambient Air Temperatures below 4
°C and Average Slot Temperatures below 60 °C. This gave a marked improvement in Goodness of Fit (GOF) statistics.

In order to eliminate or limit the possibility of FOA independent variable data (if included in the model in question) from one end of the machine adversely affecting the predictive capability of models for End Coil Vibration on the other end, it was decided to generate models where the inclusion of aforementioned FOA independent variable data would be limited to the remaining five eligible variables from the same end as the FOA Model being calculated. Using TE09 as an example, the TE09L0FOAOSO, TE09L1FOAOSO and TE09L1FOAOSO models had the FOA Vibration Variables relating to the Exciter End of the machine removed from the list of independent variables available for selection by the Stepwise selection process. As a result only 5 End Coil Vibration measured variables were included in the independent variable data for these models.

Results for these models were mixed, but overall proved not to be any better than existing models. Goodness of fit statistics for the Models using the FOA data from the same end of the machine as the FOA being modelled (TE09 in this case) are presented in Table 9.8.

<table>
<thead>
<tr>
<th>TE09 Model Prediction Goodness of Fit Statistics W208211 Filtered</th>
<th>TE09I0FOA</th>
<th>TE09I0OSOFOA</th>
<th>TE09I1FOA</th>
<th>TE09I1FIOSOFOA</th>
<th>TE09I2FOA</th>
<th>TE09I2OSOFOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>11.36</td>
<td>15.56</td>
<td>9.42</td>
<td>8.42</td>
<td>7.00</td>
<td>7.46</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.56</td>
<td>0.40</td>
<td>0.63</td>
<td>0.67</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td>R Squared</td>
<td>0.80</td>
<td>0.63</td>
<td>0.87</td>
<td>0.89</td>
<td>0.93</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 9.6 TE09 Goodness of Fit Statistics for W208211.
Data Filtered for Average Stator Slot Temperature >60°C and Ambient Temperature >4°C.

Table 9.7 displays R-Squared and RMSE values for the best performing (as defined in Section 6.10) MLR Model and the LOWESS models at all 12 End-Coils. It is apparent that the performance of most the models across all end coils was broadly similar, with the notable exception of TE25, where, as expected the R-squared value fails to stay positive throughout. Recalculation of the TE25 LOWESS model had to be carried out twice in order to maintain
Table 9.7 Goodness of Fit Statistics for All End Coils from W195199 to W232235 for the Best Performing MLR and LOWESS Models.
Both MLR and LOWESS models for EE03 and EE11 performed poorly. For EE03 and EE11 construction data sets reverted to the pre 2010 L0 to L4 types, as models generated from these construction set produced better results than the later versions. Variance in the measured vibration signal at these locations is low compared with other locations which results in low R-Squared GOF scores despite the predicted series being, at least visually, a reasonable match to the measured values. For these models the correlation between the measured and predicted vibration series could be quite strong (>0.8), despite poor GOF scores. This is discussed in detail in Chapter 8 using EE03 as an example. Recalculation of the LOWESS model for EE03 and EE11 resulted in stable GOF scores (approximately 0.6 on average) for the remainder of the period up to AOH11.

9.5 Analysis of the Poor Performance of MLR and LOWESS Models for TE25

Modelling of the Turbine End End-Coil Number 25 (TE25) vibration signal proved to be the most difficult of all over the term of this project. In this section the issues surrounding the relatively poor level of predictive accuracy of both the MLR and to a lesser extent the LOWESS models will be discussed.

The behaviour of TE25 changed greatly over the normal operating temperature range of the machine. The relationship between Average Slot Temperature (and by association most other generator related temperatures) and measured vibration was non-linear over the entire range of both variables. A reasonably linear relationship existed where Average Cold Air Temperature remained above 30 °C or below 25 °C, however the correlation values at the upper end were positive while at the lower end they were negative. The middle of the range (Average Cold Air Temperature in the range 25 - 30 °C) acted as a transition range for the change from positive to negative coefficients and vice versa. Therefore a single MLR or LOWESS model using any of the generator temperatures as independent variables will be unlikely to be capable of producing an accurate representation of the measured values over a prolonged period of time as evidenced by Goodness of Fit values for TE25 presented in Section 9.4.
Under certain conditions the behaviour can be modelled successfully e.g. W195199 where Ambient Temperature is high, Stator Core Temperature remains relatively high and by association Hot Air Temperature, Average Cold Air Temperature etc. However over a number of weeks and months (particularly given how the machine was usually instructed) these conditions did not exist for very long.

Correlation coefficients of all measured variables and differential temperatures for the periods covered by Learning Data, W200203, W208211 and W224227 for End Coils TE25, EE03 and TE01 are displayed in Table 9.8.

| Excerpts from LD, W200203, W208211 and W224227 Correlation Tables relating to End Coils TE25, EE03 and TE01 |
|--------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| TE01 (µm pk-pk)                                  | Learn D W200203 W224227        | Learn D W200203 W208211        | Learn D W208211 W224227        | Learn D W200203 W208211 W224227 |
| -0.182                                           | 0.402                          | -0.258                          | -0.182                          | -0.622                          |
| -0.233                                           | 0.375                          | -0.524                          | -0.233                          | -0.094                          |
| 0.333                                            | -0.280                         | 0.713                           | 0.333                           | 0.721                           |
| 1.000                                            | 1.000                          | 1.000                           | 1.000                           | 0.580                           |
| -0.381                                           | -0.342                         | 0.662                           | -0.336                          | -0.688                          |
| 0.580                                            | 0.370                          | 0.333                           | 0.333                           | 0.187                           |
| 0.580                                            | 0.808                          | -0.022                          | 0.187                           | 1.000                           |
| 0.511                                            | 0.203                          | 0.327                           | 0.360                           | 0.483                           |
| 0.442                                            | 0.242                          | -0.561                          | -0.249                          | 0.680                           |
| 0.408                                            | 0.550                          | -0.291                          | -0.043                          | 0.724                           |
| 0.599                                            | 0.208                          | -0.280                          | -0.153                          | 0.542                           |
| 0.431                                            | 0.593                          | 0.200                           | 0.089                           | 0.672                           |
| 0.423                                            | 0.289                          | -0.490                          | -0.345                          | 0.668                           |
| 0.076                                            | 0.280                          | -0.410                          | -0.356                          | 0.252                           |
| 0.399                                            | 0.344                          | -0.512                          | -0.317                          | 0.709                           |
| 0.422                                            | 0.295                          | -0.508                          | -0.334                          | 0.678                           |
| 0.452                                            | 0.633                          | -0.497                          | -0.069                          | 0.787                           |
| 0.420                                            | 0.288                          | -0.509                          | -0.333                          | 0.676                           |
| 0.339                                            | 0.300                          | -0.490                          | -0.377                          | 0.581                           |
| -0.153                                           | 0.586                          | -0.094                          | 0.237                           | 0.295                           |
| 0.180                                            | 0.204                          | -0.028                          | -0.003                          | 0.229                           |
| 0.136                                            | 0.132                          | -0.081                          | 0.061                           | 0.022                           |
| -0.099                                           | 0.667                          | -0.147                          | 0.253                           | 0.357                           |
| 0.121                                            | 0.694                          | -0.201                          | -0.229                          | 0.545                           |
| 0.310                                            | 0.733                          | -0.371                          | -0.118                          | 0.704                           |
| 0.561                                            | 0.520                          | -0.454                          | -0.127                          | 0.714                           |
| 0.477                                            | 0.389                          | -0.092                          | -0.104                          | 0.104                           |
| 0.495                                            | 0.369                          | -0.541                          | -0.269                          | 0.684                           |
| 0.523                                            | 0.428                          | -0.523                          | -0.230                          | 0.704                           |
| 0.582                                            | 0.394                          | -0.482                          | -0.256                          | 0.602                           |
| 0.596                                            | 0.462                          | -0.481                          | -0.231                          | 0.602                           |
| 0.433                                            | -0.183                         | 0.011                           | -0.289                          | 0.020                           |
| 0.348                                            | 0.457                          | -0.130                          | 0.051                           | 0.250                           |
| 0.618                                            | 0.406                          | -0.285                          | -0.202                          | 0.409                           |
| 0.624                                            | 0.544                          | -0.326                          | -0.137                          | 0.419                           |

Table 9.8 Pearson Correlation Table Excerpts from 2010 Learning Data, W200203, W208211 and W224227 for Coil Ends TE25, EE03 and TE01.

For each variable, if the row is highlighted in green it signifies that the correlation coefficient for that variable stayed either positive or negative for the four periods covered.
If the figures in the row are highlighted in orange it indicates that the coefficient for that variable stayed either positive or negative with the exception of W208211 where ambient temperatures were very low over a prolonged period.

The columns highlighted in red/orange are those relating to W208211 (15/11/2010 to 11/12/2010). It is immediately evident that TE25 had very few (3/32) independent variables that hold their sign over the four sets of data. This is very poor when compared with EE03 where 25 of the 32 variables available for selection in MLR and LOWSS Models had either a positive or negative correlation coefficients over the four sets of data selected; while for TE01 24 of 32 independent variables met this criterion.

EE03 was itself not one of the easier End Coils to generate an accurate MLR or LOWESS model for; however its correlation coefficient values were quite steady over the entire period. There were sharp drops in coefficient values (across the board with the possible exception of TE17) and some sign changes (5 variables) for data related to W208211. Otherwise, the correlation coefficient values were relatively steady throughout. The same can be said of TE01 and the remaining nine End coils.

The fall-off in coefficient values for W208211 can for the most part be put down to adverse ambient conditions. Conditions in the winters of 2010 and 2011 were unusually harsh resulting in generator temperatures being below normal over a prolonged period of time. The ability of the Generator Closed Cooling Water System to control the Cold Air Temperature of the Generator was limited. The system is designed to keep Cold Gas Temperature below a certain limit. The only options available to the control system to limit low cold gas temperature was to switch off two banks of cooling fans on the CCCW (this was generally only done when ambient temperature was very low). Cooling water flow to the generator could be adjusted manually, however this operation had to be carried out with extreme caution as the CCCW system was shared between the Generator and the GT Lube Oil system and any adjustment to cooling water flow on one system was offset by a corresponding increase or decrease in cooling water flow to the other.

The result of this was that the relationship between generator EWV and associated temperatures become non-linear, or existing non-linearity was worsened where ambient temperatures were low. The regression models could extrapolate in order to generate a
predicted value, but the extrapolation was based on an assumption of linearity in the relationships between the dependent and independent variables over their entire range. Predictions under these conditions were therefore unlikely to prove accurate however, if the data was filtered prior to or after the calculation to remove data related to low ambient air temperature then the Goodness of Fit scores for the various models returned to values approaching ‘normal’.

Figure 9.6 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W195199 (August/September 2010)

In Figure 9.6 data relating to End Coil TE25 are displayed for W195199 (August September 2010). The series’ shown are End Coil TE25 Measured Displacement/Vibration (µm pk-pk), GT Generator Active Power (MW), Average (Stator Core) Slot Temperature (°C), Hot Air Temperature (°C), Average Cold Air Temperature (°C) and Ambient Air Temperature (°C).

It is evident that vibration values and load were positively correlated (0.5199), particularly where ambient air temperatures remained consistently above 15 °C (Average Slot Temperature >65 °C, Average Cold Air Temperature >30 °C or Hot Air Temperature >60 °C.
approximately). However it is evident also that this relationship did not hold once temperatures began to fall. Referring again to Figure 9.6, between points 1101 and 1201 approximately where daytime temperatures were relatively low, although load level was typical, vibration amplitudes were lower due to the fall off in generator temperatures. It was also clearly visible that spikes to the downside which occur when the machine was instructed to load up quickly (load and vibration were negatively correlated) from <200 MW were more pronounced where Average Cold Air Temperatures fell below 28 °C. Due to the strong correlations between Active Power (and by association Average Slot Temperature and all other generator related temperatures) and measured vibration amplitudes for most of the learning data and for the period covered by W195199, the average MLR and LOWESS predicted model Goodness of Fit scores as measured using the R-squared statistic were approximately 0.7.

Figure 9.7 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W208211 (November/December 2010)
In Figure 9.7 data relating to End Coil TE25 are displayed for W208211 (November/December 2010). The series’ displayed are the same as in the previous figure. It is evident immediately that the correlation between load and measured vibration was negative (-0.496). It is evident that the average vibration amplitude had dropped significantly compared with W195199 (from 79.69 µm pk-pk for W195199 to 54.39 µm pk-pk for W208211). The average ambient temperature for the period was 2.89 °C down from 16.41 °C for W195199, with a minimum value for the period of -5.444 °C. The average ambient temperature between the points 901 and 1901 (27/11/2010 to 11/12/2011) was approximately 0 °C (-0.094 °C). It can be seen that vibration spikes which occurred during large downward load changes were far more pronounced where the ambient temperature was at, or close to, 0 °C (giving Average Cold Air Temperature of approximately 20 °C or below, along with reduced Stator Core Slot and Hot Air Temperatures).

![Figure 9.8 TE25 Measured Vibration, Active Power, Average Slot Temperature, Hot Air Temperature, Average Cold Air Temperature and Ambient Air Temperature for W208211 (November/December 2010). Filtered for ACAT ≥60°C and AMBT ≥4°C.](image-url)
With a filter applied to the data (as in Figure 9.8) for Average Slot Temperature ≥ 60 °C and Ambient Temperature ≥ 4 °C the correlation between measured vibration and load (as well as all generator related temperature variables) improved to -0.732.

While this was a good correlation value in and of itself, it strongly opposed the correlation values for the learning data set, for W195199 and most of W200203.
Conclusion

The aim of this project was to develop analytical tools to monitor and assess the condition of the end windings of a large 2-pole synchronous generator with a known end winding vibration problem. A number of analysis methods were used to assess the behaviour of the end windings e.g. correlation, descriptive statistics, and most importantly regression modelling of the end winding vibration data. Models were created using various methods including Multiple Linear Regression (MLR), Weighted Least Squares (WLS), Locally Weighted Regression and Scatterplot Smoothing (LOWESS). Multiple models were generated for each point of measurement (12 points of measurement in total, six on the Exciter End and six on the Turbine End) and their predictive performance assessed against the measured values and against each other.

The use of the LOWESS model for this type of analysis is novel and has proved very successful not withstanding its limitations. Its use of low order polynomials and weighting factor (Kernel function) applied to segments of the data at a time gives a flexibility that is not available when using the MLR type models. Thus LOWESS models created using only the measured independent variables (8 out of 10 in most cases) as inputs performed as well as, or better than, the higher order MLR models which may have hundreds of independent variables available for inclusion to the model.

The MLR models have proven reasonably accurate and indeed for long term accuracy of prediction have outperformed the LOWESS models for certain end coils. The use of the true stepwise selection procedure has also been a success for data sets with large numbers of independent variables, with the caveat that the Matlab result must be manually checked for insignificant variables still included (in the main) and must also be assessed by the analyst with regard to the number of independent variables included in the final model. A reduction in the number of independent variables can usually be carried out without any loss of predictive capability and indeed can yield a more robust model. This is due to the stepwise procedures’ tendency to overfit data to the dependent variable thus giving a model that is too tightly tailored to the construction set data and so is practically incapable of long term prediction of vibration values when new data is applied.
The regression models were developed with the intention of running the best performing model (as defined in Section 6.10) as a calculated variable on the PI system, thus having live measured and model predicted data available for comparison. The most accurate model in terms of prediction was the LOWESS model however it is not easily implementable on PI ACE. Prediction using LOWESS is very computationally intensive and as the demands on the particular servers were already great it was not viable to use this method for ‘live’ predictions of vibration.

Over the first two years the “live” calculated (MLR) models worked quite well (start-up and shut-downs apart) as the generator was generally operated continuously, being cycled between 140 and 230 MW on a daily basis. However the change of operating regime (discussed in the introduction) imposed on the plant, from 2009 onwards in particular, has meant that the plant operates under transient conditions much more regularly than was previously the case. This has led to a marked reduction in available valid input data to the models and so has made the generation of accurate models much more difficult. The usability of the live predicted models as an analysis method diminished over time due to the changes in operating regime mentioned already, although primarily the analysis was carried out retrospectively in any case. The availability of the live model predicted vibration values is a feature rather than a cornerstone of the analytical process. A potentially catastrophic EWV condition can develop quite quickly, but is generally present for some time prior to a failure, usually unbeknown to the operators; therefore retrospective analysis on a weekly basis is sufficient, along with daily monitoring of vibration levels by station staff. The warnings and alarms set up for the system will also alert staff immediately to any abnormally high vibration values.

Model coefficients were recalculated on a number of occasions for various reasons. The first recalibration of MLR coefficients followed a sharp fall-off in predictive capability of models for certain locations. By coincidence the deterioration in capability occurred following a scheduled outage. It is believed that the fall off in predictive capability may have resulted from a setting change to a variable in PI to which some of the end coils are sensitive (the variable is included in some models but not in others e.g. deterioration occurs at EE03 but not at TE09 because the problem exists on an input variable selected for inclusion in the EE03 models but not for the TE09 models).
A number of operational changes as well as two unusually harsh winters, have resulted in the predictive capability of the regression models being diminished. The main operational changes made were to the Minimum Stable Generation value declared to Eirgrid NCC. Over approximately 18 months Minimum Stable Generation was reduced by approximately 20%. This resulted for example in previously linear relationships between load and stator core temperature becoming non-linear. For most end windings, data for generator output below the original 140MW Min Gen value is filtered out due to non-linearity in relationships between measured vibration and explanatory/independent variable data. Changes to the normal operating regime of the plant, dictated by Eirgrid NCC, influenced by issues such as changes to the commercial and technical offers made by the plant operator (e.g. changes in start-up costs, changes to Minimum Stable Generation value, ramp rates etc.), the addition of new generators/interconnectors, available renewable energy etc. have also forced changes in how models are calculated and how output data from those models are analysed. The same is true of severe weather conditions (low temperature), for instance, with all generator cold gas cooling fan banks switched off, cold gas supply cooling water temperature control can no longer be maintained, resulting in lower than normal stator core temperatures, particularly at part load. At the upper range of generator capability the effects are less marked as supply air cooling is required and so some degree of control can be exerted over generator cold gas temperature.

In the main all of the modelling problems encountered have been overcome. The use of regression based modelling to predict end winding vibration has been successful. However, while the models are very useful tools in predicting vibration and monitoring for any changes in behaviour, they cannot be used in isolation to determine the health of the end windings. Rather, the models must be used as part of a suite of tools to monitor and analyse machine condition. Trends, basic descriptive statistics and correlation data are used to complement regression model prediction data. The setting up of a dedicated trend page in PI RT-Webparts that is viewable from any PC connected to the site intranet has also proven useful. A dedicated page has been developed in PI Process Book, permanently on display in the Central Control Room, with a mimic diagram of the end windings showing live vibration values as well as having warning and alarm limit visual alerts.

The addition of the new variables (particularly vibration data) greatly improved performance of the MLR and LOWESS models for the most part. Poor predictive capability at TE17 has
effectively been eliminated. However it was found that the existing independent variable sets might suit certain End Coils somewhat better, namely EE03 and EE11 (each of which performed poorly for 2010) and these will have the L3 (EE03 best performing MLR for 2009/2010 model) and L4 (EE03 best performing MLR for 2009/2010 model) data sets made available for calculation of models again from this point on.

Overall, by using Regression Modelling, Descriptive Statistics and Correlation Data an accurate and reliable assessment of machine behaviour can be developed and monitored over a long period. Due to the more complex operating regime of the plant in recent years however it does require more work/knowledge on behalf of the analyst to use these methods to ascertain/monitor the condition of the machine than was envisaged at the start of the project.

Further work is under way to generate efficient models for the current operating regime and the development of new live calculated versions is also underway using PI Process Book as the platform for calculation rather than PI ACE. Although this limits flexibility somewhat, for MLR models at least, it is sufficiently powerful.

Additional work will be carried out to investigate the possibility of predicting vibration during start-up which is extremely stressful to the machine. This will require models to be generated from data of much shorter interval (1 – 10 seconds) as the machine moves from barring speed to Minimum Stable Generation output (112MW) in approximately seven minutes. Modelling of vibration during start-up conditions would be extremely beneficial due to the running regime imposed by Eirgrid NCC since early 2012 whereby the machine tends to be instructed ON, infrequently and for short duration.

Modelling techniques such as Artificial Neural Networks (ANN) may be applicable in this case. Should techniques such as this prove to be applicable, then the calculation of models capable of accurately predicting EWV over the entire operating range of the generator may be possible. Techniques such as artificial neural networks may also prove to be better suited to prediction of vibration for start-up than the more standard regression modelling techniques described in this thesis.
References/Bibliography


Appendix 1

End Winding Vibration Monitoring System

A1.1 Hardware

The hardware purchased included the following main components:

- 24 single axis fibre-optic accelerometers c/w converters from optic to electrical signal.
- 4 junction boxes c/w power supply for accelerometers, mounting plate for 6 signal converters and terminals for cable connections.
- 3 PCU 100 processing units capable of processing 8 vibration channels each.
- Dedicated HP Proliant Server.

A1.1.1 Fibre Optic Accelerometers (FOA)

The FOA100 accelerometer is a single axis accelerometer designed for operation in areas with high levels of electromagnetic interference (EMI). Its construction is wholly non-metallic and non-conductive making it immune to EMI. Its construction makes it ideal for operation in areas where standard piezo-electric and piezo-resistive accelerometers cannot operate reliably and is well suited to the turbo-generator environment.

The accelerometer signal is fed into a conditioner which converts the optical signal to an electrical one. It has a measuring range of 0-40 g and, the output signal has sensitivity of 100 mV/g on a bias voltage of 6 VDC.

A1.1.2 The PCU 100 Programmable Units

The PCU100 is a multi-channel, digital processing unit. It has on-line programmability allowing changes to be made without loss of information gathering function. Its modular
design allows it to be set up for various input/output configurations depending on the application. It can accept up to 8 input signal and via analogue output card output up to 8 processed analogue outputs. The raw signals from the input card can also be used as input to devices such as oscilloscopes etc.

A1.1.3 Fibre Optic Accelerometer Module

This card accepts the 100 mV/g signal from the accelerometer conditioner converts A/D and processes the signal as required by the user. The output signal can be acceleration, velocity or displacement. This card can accept a maximum of two accelerometer signals. Two RAW analogue output channels are available.

A1.1.4 Analogue Output Module

The analogue output modules are four-channel electronic cards which can convert signals based on user requirements. The PCU100 can accept a maximum of three analogue output modules and a single module can accept four signals from the vibration input modules for processing and output up to four analogue outputs through I/O ports at the rear. The analogue outputs are available as 0 – 10 V and 4 – 20 mA signals. The module can also be set up to monitor alarm threshold limits and detect violations and can also detect the whether the alarm is due to a rising or falling value.
A1.2 System Installation

The installation of the GT Generator EWV monitoring system was carried out between the 24\textsuperscript{th} and the 27\textsuperscript{th} of November 2006.

In the weeks running up to the outage the following preparatory work was undertaken:

- Purchase and set-up of server.
- Installation of junction boxes underneath generator.
- Fabrication of flange plates for penetration point to generator.
- Installation of cabling between junction boxes and cabinet containing PCU100s.
- Sourcing of rack cabinet for PCU100s.
- Installation of cabinet in the Steam Turbine Hall approximately half way between the two generators.
- Installation of power supplies etc. for PCU100s and signal conditioners in the junction boxes.
- Installation of cabling between PCU100 cabinet and Server located in the GT PCC.
- Testing of cables etc.

Prior to the work commencing, a final review of the installation, sensor locations etc. was carried out.

Prior to the commencement of the Fibre Optic Accelerometers (FOA) installation, an inspection of the end winding support structure was carried out. As there were no major findings from the inspection, the installation of the accelerometers was allowed to commence.

The installation of the sensors was completed in the following steps:

The sensors were placed on the required bar close to the edge and temporarily secured with tape. The sensor had to be facing in the correct direction so that measurements were made in the correct plane i.e. radial.

When all six sensors were temporarily affixed the cabling was routed carefully keeping required minimum bending radius requirements in mind.
The sensors were then fixed to the coils using a two-part epoxy paste (EP8117). The accelerometers were then tied down securely with mica tape which was treated with an epoxy varnish and hardener (Epoxylite 230).

The cabling was carefully affixed to the support structure using mica tape and varnish.

At the exit point of the generator the loom of cables was affixed to the generator casing using a strapping and the Epoxy paste.

Once the FOA installation was complete, 10-12 hours of forced ventilation through the generator was required to allow the epoxies/resins to cure.

![Figure A1.1 Gas Turbine Generator: Sensor Location Exciter End End-Coil Number 3 (EE03)](image)

The next step was to terminate the cabling at the protection boxes underneath the generator (see Figure A1.2) and at the PCU100 cabinet (see Figure A1.3) as well as test the accelerometers for function to ensure no damage had occurred during the installation. Once the hardwiring of the system was complete the PCU 100’s were powered up and programmed. All aspects of the set up could be implemented at the PCU100, however, a set up file created in ZOOM Configuration 5 could be stored to a flash card and downloaded to the PCU100. This method was much less time consuming.
Figure A1.2 Gas Turbine Generator (GTG) EWV Monitoring System Protection Box B.

Figure A1.3 GTG PCU100 and BNC Connector Racks.
Appendix 2

GT Generator End Winding Low Tune Repair

A2.1 Requirement for Low Tune Repair of Generator End Windings

The Low Tune Repair carried out by the OEM during the scheduled 30 day outage of April/May 2007 was recommended in order to finally solve the problem of excessive end winding vibration.

Previous attempts at “High Tuning” or raising the natural frequency of the end winding structures had proved unsuccessful. Increased static natural frequency values achieved after high tuning and resin treatment proved to be temporary with the natural frequency values creeping steadily downwards over a period of months back toward potentially dangerous values.

The attempts to raise natural frequency and homogenize the response of the entire structure also failed to prevent further damage to support brackets etc.

The OEM put forward the option of a Low Tune Repair/Detune of the TE and EE End Windings as final solution to the problem. At the time of the proposal however there were very few machines in operation having had this repair carried out. And among the ones in operation there was only a few thousand hours operation between them post repair. This was a cause of concern to the plant operator as the repair could not really be considered proven at the time.

After much in house deliberation, and consultation with both the OEM and with other plant owners it was decided to go ahead with the proposed repair.
**A2.2 Aim of Low Tune/Detune Repair of GT Generator End Windings**

The stator end caps behave, in some respects, similarly to a cantilever beam fixed at one end. The natural frequency of the beam is then dependent on its length, breadth and the material used. The longer the beam the lower the natural frequency will be. By cutting the support braces back and decoupling the support ring from the support brackets, the length of the “cantilever-like” section of the end winding is increased, resulting in a lowering of the natural frequency of the end bars see Figure A2.1 taken from an OEM end winding repair proposal document from September 2006 [34].

The purpose of the repair was to move the static natural frequency of the end winding structures away from the 100Hz magnetic forcing frequency. The last available static natural frequency values available for the machine pre repair were 111 Hz on the Exciter End and 99.7Hz on the Turbine End. The aim of the “Low Tune” was to take the static natural frequency values for both ends well below the 100 Hz level to around 90 Hz approximately. This would drop even further (3 - 10 Hz) when the machine was running due mainly to the effects of heating.

![Low Tuning Measure](image)

Figure A2.1 Axial and Radial Decoupling of Support Ring and Slitting of the Support Brace
A2.2 **GT Generator End Winding Low Tune/Detune Measure**

### A2.2.1 Schedule of Repair Work

1. Initial Inspection.

2. Repair to z-profile no.1. (TE Only)

3. **TE/EE Low Tune Works**
   - Axial and Radial decoupling of Support Ring
   - Slitting of Support Braces
   - Drilling of Support Ring
   - Addition of Tangential Blocking at the winding series connections/ end caps (Carried during High Tune repair in 2005)
   - Strapping of bottom layer bars to support ring
   - Resin treatment of cracks in resin at the intersection of all top layer bars and the support ring on the TE and all bars from 24 to 44 on the EE reported following a previous inspection in November 2006.
   - Resin treatment of new strapping.
   - Static natural frequency test (‘bump-test’)
   - Further slitting of support bracing
   - Second ‘bump-test’

4. Preparation of the Generator for electrical testing (removal of flexible links between busbars and generator terminal connections).

5. Electrical Testing (High Voltage Test, PI Test etc.).
6. Final Inspection.

**A2.2.2 Initial Inspection**

An initial inspection carried out prior to the commencement of repair works uncovered cracking of support bracket (z-profile) No.1. With the exception of the cracked support bracket a the TE of the machine, The EE and TE end winding structures appeared in good conditions with no visible signs of excess stress or wear. There were no signs of any fresh dusting deposits caused by excess relative movement of end coils at either end. The repair to z-profile no.1 delayed the commencement of low tuning work on the TE of the generator by three days.

**A2.2.3 Tangential Blocking**

Tangential blocks, shown in Figure 4.4, which would ordinarily be installed as part of the Low Tune Repair were installed on this GT Generator during a ‘High-Tune’ repair carried out in 2005. The blocks are inserted at the straight section of the end coils just before the series connections. Similar machines have had the LTR carried out as a first measure or purely as a precautionary measure and would not have had this additional support blocking installed.

**A2.2.4 Axial and Radial Decoupling of Support Ring**

The works began with the axial and radial decoupling of the support rings. Radial decoupling was achieved by removing the packing between the support ring and the support brackets (z-profiles), and removal of bandages and packing supporting the phase leads against the support ring at the TE of the machine, see Figure A2.2. Axial decoupling involved the cutting of the support bracing along the support ring and the removal of strapping between the support brace and the support brace as shown in Figure A2.3
Figure A2.2 Radial Decoupling of Support Ring from Phase Lead Connection at 12 O’ Clock Position at the Turbine End of the GT Generator.

Figure A2.3 Axial and Radial Decoupling of Support Ring from Support Bracket and Support Brace at the Turbine End of the GT Generator.
A2.2.5 Drilling of the Support Ring

Holes were drilled in the support ring, parallel to the rotor. These holes were drilled in order to allow strapping of the bottom/lower layer bars to the support ring.

![Figure A2.4 Through Holes Drilled in the Support Ring at Exciter End of the GT Generator. Tangential Blocking Inserted During High Tune of End Winding at AOH 2005.](image)

A2.2.6 Slitting of the Support Braces

The support braces are fibreglass plates which are bolted to the support brackets. They fit against the end winding cone in the axial direction and allow the bottom layer bars to be strapped to the support brackets. The straps are run through holes bored through the support brace along its axial length. In order to increase the amount of the end basket that can move freely the support braces were slotted axially through the holes used for strapping. Initially the support braces at each end were slotted to the third hole, see Figure 4.3. Further slotting could be carried out if necessary after the completion of works and a ‘bump-test’.
A2.2.7 Strapping of Lower Layer Bars to the Support Ring

The bottom layer bars are strapped to the support ring in order to further homogenise the whole end winding structure. Multiple layers of strapping are used and the entire structure is resin treated afterwards.

A2.2.8 Resin Treatment of End Windings

All new support bandages/strapping were resin treated to strengthen and homogenise the end winding structure. It had been noticed during a previous inspection in November 2006 that cracks had begun to appear in the resin at the point of intersection of the bottom layer end winding bars and the support ring. The cracks affected all bars from 24 to 44 at the EE of the machine and all TE bars. The areas of the TE and EE end windings affected were also resin treated.
The gap between the outside (radially) of the support ring and the support brackets (z-profiles) was closed to between 0.5 mm and 1 mm. This was carried out using shims made of insulating material. The shims are glued to the support ring. They are also strapped to the support ring for added security. The gap is maintained at 0.5 to 1 mm so that the end winding basket is free to move and deform slightly during normal operation. However, during fault conditions the end winding structure will not be allowed to deform enough to cause damage to the structure as the support ring will rest against the support brackets.

**A2.2.9 Bump Test No.1**

With the major works complete a ‘bump test’ was carried out at both ends in order to assess the effectiveness of the repair thus far. Results indicated that while the natural frequencies were lower, the drop was not satisfactory. Further slotting of the support braces would be required. OEM engineers reported that the global mode natural frequencies of the TE and EE
end winding baskets with the support braces slotted to the third hole were 94.5 Hz and 99.2 Hz respectively.

**A2.2.10 Secondary Slotting of Support Braces to Fine Tune the ‘Low Tune’ Repair**

In consultation with their main engineering headquarters, engineers from the generator manufacturer decided that the support braces on the Exciter End (EE) would be slotted to the sixth hole. The support braces on the Turbine End (TE) would be slotted to the fifth hole. The Exciter End required the support bracing to be slotted to a greater degree that the Turbine End support braces as the static natural frequencies at the Exciter End were higher than at the Turbine End.

![Image of slitting support brace to sixth hole at EE of the machine](image)

*Figure A2.7 Slitting of the EE Support Bracing to the Sixth Hole Following the Initial Bump Test.*
A2.2.11 Bump Test No.2

A second bump test was carried out after the completion of the additional slotting of the support braces. Static natural frequency values were found to be in the range 89.5 Hz to 91.7 Hz at the Turbine End and 92.3 Hz to 92.5 Hz at the Exciter End. This was deemed a satisfactory result. The machine was then cleaned and preparations made for electrical testing.

Figure A2.8 Slitting of the TE Support Bracing to the Fifth Hole Following the Initial Bump Test.

A2.2.12 Electrical Testing

Electrical testing was carried out successfully and the machine was found to be in good order.

A2.2.13 Final Inspection

A final visual inspection was carried out with all personnel and materials removed from the machine. The generator was found to be in good order and fit for service
Appendix 3

Regression

A3.1 Simple Linear Regression

Simple Linear Regression generates a linear trend line using the least squares method in order to describe the behaviour of a data set (dependent variable/predictand). The calculation is made using the dependent variable and another variable which has a relationship with the dependent variable, referred to as the independent variable, explanatory variable or predictor. For example: if we wish to develop a prediction technique which relates vibration to active power. Vibration is referred to as the dependent variable while Active Power is referred to as the independent variable.

The simple linear regression equation can be expressed as follows:

\[ y = \beta_0 + \beta_1 x_1 + e \]

*Equation A3.1 The Simple Linear Regression Equation*

Where:  
\( y \) is the column vector of dependent variable values,  
\( \beta_0 \) is the constant/y-axis intercept,  
\( \beta_1 \) is the slope/regression coefficient,  
\( x_1 \) is a column vector of independent/explanatory variable values and,  
\( e \) is a column vector of error/residual values.

All three column vectors are of the same length.
Assumptions are made in the use of the simple regression method which the researcher must try to meet in order to achieve accurate results. The assumptions are as follows:

1. The error \( e_i \) is normally distributed
2. \( \sum (e_i) = 0 \) There is no reason to believe that the error terms will tend to be positive rather than negative or vice versa.
3. \( \sum (e_i^2) = \sigma^2 \), all observations have equal (constant) variances (“Homoscedasticity”)
4. \( \sum (e_i e_j) = 0 \) \((i \neq j)\) errors in one observation are uncorrelated with error in other observations (Correlated errors in time series are referred to as “serial correlation” or “autocorrelation “)
5. Variables are measured without error (nonstochastic)

### A3.2 Multiple Linear Regression

Multiple Regression is a logical extension of the two variable or ‘simple’ regression. With Multiple Linear Regression there are a number of independent variables used. The Multiple Linear Regression Model is of the form:

\[
y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_K x_K
\]

**Equation A3.2 The Multiple Linear Regression Equation**

Where \( y \) is the predicted value of the dependent variable, \( b_0 \) is the y-axis intercept/constant i.e. the value for \( y \) when all explanatory variable values are set to 0, \( b_k \) is the coefficient of regression for the \( K^{th} \) explanatory variable, \( x_k \) is the \( K^{th} \) explanatory variable.

The MLR model can be written in vector matrix form as:

\[
Y = X\beta + e
\]

**Equation A3.3 The Multiple Linear Regression Equation (Vector Matrix Form)**
Where: Y is the column vector of the dependent variable for the data set, X is the matrix of explanatory variables for the period, $\beta$ is a row vector of regression coefficients with the regression constant first e is a column vector of residuals/error values.

There are a number of assumptions made in the use of the multiple regression method which must be met in order to achieve accurate results as follows:

1. The error $e_i$ is normally distributed
2. $\sum(e_i) = 0$ There is no reason to believe that the error terms will tend to be positive rather than negative or vice versa.
3. $\sum(e_i^2) = \sigma^2$, all observations have equal (constant) variances (“Homoscedasticity”)
4. $\sum(e_i e_j) = 0 (i \neq j)$ errors in one observation are uncorrelated with error in other observations (Correlated errors in time series are referred to as “serial correlation” or “autocorrelation “)
5. Variables are measured without error (nonstochastic)
6. The number of observations exceeds the number of coefficients to be estimated.
7. No exact linear relationship exists between any subset of explanatory variables (perfect “multicolinearity”)

Assumptions 1 – 4 involve specification of the error term, while 5- 7 refer to explanatory variables.

Assumption 6 guarantees that there is enough variance to estimate coefficients. If the number of coefficients to be calculated were the same or greater than the number of observations, then the coefficients could take on any values and therefore could not be identified. For example, two points are required to fit a simple regression (two points give a straight line). If there were only one observation (data point) any intercept and any slope is possible.

Assumption 7 guarantees that two variables are not identical (do not measure the same thing). If the two variables are identical then it is impossible to identify the independent effect of either of them. Perfect multicolinearity is a virtual non-entity in real world situations.
Violation of this assumption is usually down to mistakes made by the analyst rather than any problem with the data set.

A3.2.1 The Multiple Regression Model Equation

The intercept is the predicted value of Y when each of the explanatory variables (X₁…Xₖ) are equal to zero. The kᵗʰ expression slope is the predicted change in Y with a unit change in the kᵗʰ explanatory variable, with the remaining variables held constant. For this reason the regression slopes in a multiple regression are often referred to as net regression coefficients. That is, they correspond to a controlled experiment in which all other explanatory/independent variables are held at certain fixed values.

A graphical representation of multiple regression is not easy to create as it yields a k-dimensional surface (or regression plane) rather than a single line which results from a simple regression.

A3.2.2 Calculation of Coefficients

As is the case with simple regression, when calculating a multiple regression the least squares criteria are applied. The error eᵢ is the distance from the regression surface to the actual/measured/true values. Calculating the regression coefficients involves finding the surface which minimises these distances. The statistic used is SSE (Sum of Squares due to Error). For a perfect regression SSE will equal 0.

A3.2.3 Regression Model Summary

The measure R is the positive multiple correlation coefficient. It represents the correlation between the dependent variable yᵢ and the fitted value ŷᵢ.
A3.3 Analysis of Regression Model

A3.3.1 ANOVA Table (ANalysis Of VAriance)

The ANOVA table provides an overall summary of the model. The row labelled Regression shows the variation explained by the model. The row labelled Residual shows the remaining or residual variation after fitting the model. The column headed Degrees of Freedom (DF) contains the degrees of freedom associated with each term. If the model contains only 1 explanatory variable then the degrees of freedom associated with the regression term is 1. The total degrees of freedom equal the number of observations minus 1. The Residual Degrees of Freedom (DFE) equal the total degrees of freedom minus the Regression Degrees of Freedom (DFR). The Mean Square (MS) is equal to the Sum of Squares (SS) divided by the value of DF. The general version of the ANOVA table is shown below.

<table>
<thead>
<tr>
<th></th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Stat</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>DFR = k</td>
<td>SSR</td>
<td>MSR</td>
<td>F</td>
<td>P-value</td>
</tr>
<tr>
<td>Residual</td>
<td>DFE = n-k-1</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>DFT = n-1</td>
<td>SST</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where:

DF = Degrees of Freedom
SS = Sum of Squares
MS = Mean Square (=SS/DF)
F = Value of F-statistic (=MSR/MSE)
P-value = significance level related to F-statistic
k = number of explanatory/independent variables

The F statistic is equal to the Regression Mean Square (MSR) divided by the Residual Mean Square (MSE). For any multiple regression model, the F-statistic provides a test of the null hypothesis that none of the explanatory/independent variables are significant in explaining variance in the dependent variable. Under the null hypothesis ($\beta_1 = \beta_2 = \cdots = \beta_{k-1} = 0$) the F statistic follows Fisher’s F distribution and has an expected value of 1. The higher the value
of $F$ the more doubt there is about the null hypothesis. The $F$ statistic equation is expressed as follows

$$
F = \left( \frac{\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 / (k - 1)}{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / (n - k)} \right)
$$

Where: $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$

Equation A3.4 The $F$-statistic

The p-value associated with the $F$ statistic is the probability of seeing the value of the $F$ statistic for a given linear regression if the null hypothesis ($\beta_1 = \beta_2 = \cdots = \beta_{k-1} = 0$) is true.

### A3.3.2 Parameter Information/Coefficient Table

Further data including parameter estimates, parameter estimate standard errors and associated $t$ statistics and $p$ values are also usually given.

The Standard errors are the standard errors of the regression coefficients. The regression coefficient for the predictor/explanatory variable is the difference in response per unit difference in the Explanatory variable.

The $t$-statistic tests the hypothesis that a population regression coefficient $\beta$ is 0, i.e. $H_0 : \beta = 0$

It is the ratio of the sample regression coefficient ($\beta$) to its standard error. The $t$-statistic for the regression coefficients is calculated as follows

$$
t \text{ statistic} = \frac{\beta_i}{SSE_{\beta}}
$$

Equation A3.5 The $t$-statistic
The statistic has the form \((\text{estimate} - \text{hypothesised value})/\text{SE}\). However, since the hypothesised value is zero the statistic reduces to \(\text{estimate}/\text{SE}\).

The p-values or observed significance levels for the t-statistics are also displayed.

The degrees of freedom used to calculate the p-values are given by the Error DF (Degrees of Freedom) from the ANOVA. The p-value indicates whether or not the independent variables have statistically significant predictive capabilities.

Theoretically, the p-value for the constant could be used to determine whether the constant could be removed from the model. In practice this not done for two reasons
- When there is no constant the simple regression model is given by:
  \[
  y = b_1 x
  \]
  \text{Equation A3.6 The Simple Linear Regression Equation (No Constant)}

  And for the multiple regression model is
  \[
  y = b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_k x_k
  \]
  \text{Equation A3.7 The Multiple Linear Regression Equation (No Constant)}

  This forces the \(y\) term to 0 when the \(x\) terms are 0. Even when the condition is appropriate it is often incorrect to place this constraint on the regression.

  Standard practice (hierarchical modelling) is to include all simpler terms when a more complicated term is added to a model. Nothing is simpler than a constant so if an independent variable is to be added to the model, the constant should be included also.

**A3.3.3 Coefficient of Determination (\(R^2\))**

The coefficient of determination measures the combined contribution of all included independent variables in explaining the variation in the dependent variable. \(R^2\) is calculated as \(\text{SSR}/\text{SST}\). The inclusion of another explanatory variable will always increase the \(R^2\) figure.
This will occur even when the added variable has no relationship to changes in the dependent variable. For this reason, $R^2$ should not be used to compare regressions using different numbers of independent variables.

A Venn Diagram explanation of $R^2$ is shown in Figure A3.1. The upper diagram represents a simple regression where variance in the dependent variable, vibration, is explained by a single independent variable, active power. The lower part of the diagram represents a multiple regression where variance in vibration is explained by two independent variables, active power and reactive power.

*Note: the sum of areas from simple regression will be greater than the sum from partial by the size of the intersection.*
Area in common is variation in vibration explained by active power

Area used to estimate the significance of Active Power

Area of intersection adds to variance explained jointly by both explanatory variables, but does not add to their individual significance.

Figure A3.1 Venn Diagrams of Regression

A3.4 Weighted Least Squares Regression

For regressions with cross-section data, it is usually safe to assume the errors are uncorrelated, but often their variances are not constant. This problem is known as heteroskedasticity (“unequal scatter”) as opposed to the usual assumption of constant error variance referred to as homoscedasticity. Although the mean of the dependent variable may be a linear function of the explanatory variables, the variance of the error terms may also depend on those same
explanatory variables, so that the observations might fan out when plotted on a scatter diagram.

There are a number of methods of dealing with heteroscedasticity, namely:

- For data where there is known heteroskedasticity, the use of weighted least squares regression (WLS) is recommended to obtain unbiased estimates.
- Test for heteroskedasticity using a squared residual regression
- Estimate the unknown heteroskedasticity parameters using a squared residual regression, then use the estimated variances in the WLS formula to get efficient estimates of regression coefficients
- Use least squares estimators and get estimates of standard errors which are correct under arbitrary heteroskedasticity. This method is relatively inefficient.

For this project the method of residualisation mentioned was used in an effort to improve the performance of regression models, particularly for some end winding positions where accurate modelling was proving difficult. The method has four major stages:

- Conduct an Ordinary Least Squares (OLS) Regression in order to obtain the residuals/errors and squared residuals/errors
- Having now created a squared error variable, regress the squared error variable on the original independent variables using the OLS method.
- Calculate the reciprocal of the square root of the absolute squared residuals. These “weights” are then multiplied by all of the variables in the regression model.
- Finally, obtain the OLS regression of the weighted dependent variable on the weighted independent variables.

**A3.5 Model Validation**

$R^2$, even if adjusted for loss of degrees of freedom due to the number of predictors in the model, can give a misleading, overly optimistic view of the accuracy of prediction when the model is applied to data outside of the construction/learning data set. For this project application outside the calibration period is the rule rather than the exception. The learning
data/calibration period statistics are typically biased as the model is fine-tuned for maximum agreement with the calibration period data. Sometimes too large a pool of predictors is used in automated procedure (e.g. stepwise) to select the best predictors. The calibration period/learning data itself may not be representative of the normal relationships between the variables. This can result in the models being accurate under one operating condition or period of time and inaccurate for another.

Therefore it is advisable to validate the regression model by testing it using data from outside of the learning data/calibration period. A number of methods are available such as cross validation and split-sample validation but because to the amount of available data for the project we do not require to use either of these, instead we can simply apply the model to data not used in the calculation of the model to validate its accuracy.

It must also be kept in mind that for all regressions, modelled relationships may not be valid for periods where the predictors/independent variables operate outside of their normal ranges as used in the calculation of the model. Some data from outside of the construction set data may have a multivariate distribution of predictors not matched exactly in the calibration data period, thereby requiring the model to approximate independent variable data in order to arrive at a value. Making the distinction as to whether a prediction is arrived at through interpolation or extrapolation is beneficial. Interpolation is more likely to be accurate as there are boundaries on minimum and maximum values.

**A3.5.1 Statistics for Validation of Regression Model**

Validation statistics measure the accuracy or inaccuracy of the prediction for a given validation period/validation data set.
A3.6 Prediction Error

All of the statistics described in this section are computed as some function of the error, which is the difference between the observed and predicted values.

\[
\hat{e}_{\text{PRED}(i)} = y_i - \hat{y}_{(i)}
\]

Equation A3.8 Prediction Error

Where: \( y_i \) and \( \hat{y}_{(i)} \) are the observed and predicted values of the dependent variable at point \( (i) \).

A3.6.1 Sum of Squares due to Error (SSE)

SSE_{\text{PRED}} is the sum of the squared differences of the observed and predicted values.

\[
SSE = \sum_{i=1}^{n} (\hat{e}_{(i)}^2)
\]

Equation A3.9 Sum of Squared Error of Prediction

Where the summation is over \( n \) data points, making up the validation period/data set.

A3.6.2 Mean Squared Error (MSE)

MSE_{\text{PRED}} is the average squared error for the validation data, or the sum of squared errors divided by the length of the validation period

\[
MSE = \frac{SSE}{n}
\]

Equation A3.10 Mean Squared Error

The better the accuracy of the model predictions, the lower the MSE value.
A3.6.3 Root Mean Squared Error (RMSE\textsubscript{PRED})

The RMSE\textsubscript{PRED} is a measure of the average size of the prediction error for the validation period, and is computed as the square root of the mean squared error of the prediction for the particular period.

$$RMSE = \sqrt{MSE} = \left( \frac{\sum_{i=1}^{n} (\hat{e}(i))^2}{n} \right)^{\frac{1}{2}}$$

Equation A3.11 Root Mean Squared Error

The relative advantage of RMSE over MSE is that it expressed in the units of the dependent variable. The equivalent of the RMSE within the learning data is the standard error of the estimate \( s_e \). RMSE\textsubscript{PRED} will generally be greater than \( s_e \) because \( s_e \) reflects the tuning of the model to the data in the learning data set. The difference between RMSE\textsubscript{PRED} and \( s_e \) is a measure of the importance of this tuning of the model. If the difference is small, the model is said to be validated, or to verify well. The definition of small depends greatly on the specific project.
A3.6.4 Predicted $R^2$ ($R^2_{\text{PRED}}$)

This statistic measures the predictive ability of the model. This is similar to the RE (Reduction of Error statistic) in that it measures the ability of the model to predict accurately for data outside of the learning data set.

For this project $R^2_{\text{PRED}}$ is calculated as

$$R^2_{\text{PRED}} = 1 - \frac{SSE_{\text{PRED}}}{SST_{\text{PRED}}} = 1 - \left[ \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \right]$$

Equation A3.12 R-squared Predicted (for Data Outside of Construction Set)

The RE function uses the original learning data dependent variable mean $\bar{y}_c$ in place of the prediction data set mean $\bar{y}_i$, which is not suitable for use in this project.

Like the RE statistic $R^2_{\text{PRED}}$ has a possible range of - $\infty$ to 1. An $R^2_{\text{PRED}}$ of 1 indicates a perfect prediction of the data set, and can only be achieved if all residuals are equal to 0 ($SSE_{\text{PRED}} = 0$). The minimum possible value cannot be specified as $R^2_{\text{PRED}}$ can be negative and arbitrarily large if $SSE_{\text{PRED}}$ is much greater than $SST_{\text{PRED}}$. A positive $R^2_{\text{PRED}}$ can be taken to mean that the model has some ability to accurately predict the behaviour of the dependent variable. On the other hand, if the $R^2_{\text{PRED}}$ value is 0 or negative it can be said that the model has no ability to predict accurately the behaviour of the dependent variable.

In general it can be taken that if $R^2_{\text{PRED}}$ values for the validation data set/new data set are close to the $R^2$ values attained for the learning data that the model can be described as validated. Where $R^2$ is defined as

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \left[ \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \right]$$

Equation A3.13 R-squared (Construction Set Data)
A3.7 Variable Selection

With automated entry of independent variables (stepwise methods etc.) into the model comes high risk of overfitting, the inclusion of insignificant variables and/or the exclusion of significant variables.

There are a number of options when deciding how to limit the number of variables used in the model. Dealing firstly with overfitting; as the $R^2$ value will rise with the addition of more independent variables, regardless of their relevance. The $R^2_{ADJ}$ statistic could be used as an alternative criterion to identify when to stop adding variables to the model. This however has a couple of major flaws. The $R^2_{ADJ}$ statistic assumes that all of the input variables are independent, whereas in practice the variables are often inter-correlated. Secondly, the $R^2_{ADJ}$ statistic does not solve the problem of picking variables from a pool of explanatory variables. If the pool of possible explanatory variables is large, $R^2$ can become seriously biased (high). This bias will not be accounted for by the adjustment for the number of variables in the model used by the algorithm for $R^2_{ADJ}$.

Another technique for tackling overfitting of the regression model is to use cross-validation as a guide for stopping entry of additional explanatory variables. By evaluating the performance of the model on data withheld from calibration at every step of the stepwise procedure, the level of complexity (number of explanatory variables) above which the model is over-fit can be estimated. Graphs of change in calibration data set and validation data set accuracy statistics as a function of step in a stepwise selection scheme can be used as a guide for cutting off entry of predictors to the model. For example using RMSE and $R^2$ of the learning data set models can be selected based on minimisation of RMSE and maximisation of $R^2$ using the fewest possible explanatory variables. Once there is no reasonable gain to be made (in terms of minimisation of RMSE or maximisation of $R^2$) by the inclusion of additional variables, no more explanatory variables should be included. Persistence with entry of further variables into the model amounts to over-fitting. Overfitting refers to the tuning of the model to noise rather than any real relationship between the variables. An extreme example of overfitting would be where a model has the same number of explanatory variables as data points/observations in the learning data. This model would explain 100% of the variance in the dependent variable even if the explanatory variable data is random noise.
Stepwise methods are prone to including variables which have limited or no statistical significance for reasons already mentioned, and so care must be taken when using this method of variable selection. Thorough testing of chosen models is a must.

For this project models are calculated using 12 weeks’ worth of data (6943 observations) and have been used to predict 40 (27000+ observations) weeks’ worth of new data not included in the calculation of the models.

**A3.8 Regression Model Prediction**

Model predicted values are obtained by applying the prediction equation to data from outside of the learning data set. Once the model has been developed, the generation of the prediction is a simple mathematical step, but important assumptions are made in taking it.

The multivariate relationship between the dependent variable and the explanatory variables in the learning data period is assumed to apply into the future. This assumption could be violated for many possible reasons. For example, the generator may require further repair work be carried out to the stator end windings which could result in a change to the behaviour of the end winding structure (e.g. further high or low tuning, resin treatments etc.). Wear and tear to the machine over an extended period will result in the behaviour pattern of the machine altering somewhat. The duty of the machine may change so that predictions must be made for explanatory variable data which is outside of their range used for the calculation of the model.

Many other scenarios could be envisaged that would invalidate the application of the regression model to predict future vibration data. For time series in general, regardless of the system, it is important to check the ability of the model to predict outside of its learning data period as described previously.
A3.8.1 Prediction Error Bars and Confidence Intervals.

Along with a prediction, an estimate of the uncertainty surrounding its accuracy should be provided. Error bars are frequently used to summarise this uncertainty on a time series plot of the prediction. Differing methods can be used to calculate error bars, such as those discussed by Nicolich and Jorgensen [35].

Standard error of the estimate, \( s_e \), is the square root of the mean squared residuals, MSE i.e. RMSE. Assuming that the error term is random a 95% confidence interval of roughly

\[
CI \approx \hat{y}_i \pm 2x_e
\]

Equation A3.14 MLR Confidence Intervals

Confidence band by this method are the same width for all predicted values of vibration. The \( \pm 2x_e \) band is often a good approximation to the 95% confidence interval, especially if the learning data set is large as is the case for this project.

However, due to uncertainty in the sample mean of the dependent variable and in the estimation of the coefficients of regression, variance in the predicted vibration values for data not used in the calculation of the model is larger than is indicated by the MSE and is not the same for all predicted values. To combat this, another, slightly more precise estimate of prediction error called the standard error of prediction may be used.

Refer to [35] for more information on the calculation of prediction error bars and confidence intervals.

A3.8.2 Standard Error of Prediction

The standard error of prediction is proportional to \( s_e \), but it also takes into account the uncertainty in the estimated mean of the dependent variable and in the estimates of the regression coefficients. Due to these extra factors the prediction error is larger when the explanatory variable data is a long way from their learning data means, and vice versa. The calculation of the standard error of prediction for a simple linear regression is as follows:
Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

\[ s_{\hat{y}} = s_e \left[ 1 + \frac{1}{n} + \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right] \]

**Equation A3.15 Standard Error of Prediction (Simple Linear Regression)**

Where:
- \( s_e \) is the standard error of the estimate,
- \( n \) is the no of observations/data points in the learning data set,
- \( x_i \) is the value of the explanatory variable corresponding to observation \( i \),
- \( \bar{x} \) is the calibration period/learning data set mean of the explanatory variable,
- \( x_p \) is the value of the explanatory variable for the prediction observation/data set, and the summation in the denominator is over the \( n \) observations in the learning data set.

The most important thing to note here is that \( s_{\hat{y}} > s_e \), and that the difference has contributions from the two terms on the right, inside the square root.

The calculation of this statistic for MLR is more complicated than the method for simple regression. The model predicted dependent variable data for MLR is then given by:

\[ \hat{y}_p = x_p^T \hat{\beta} \]

**Equation A3.16 Standard Error of Prediction (Simple Linear Regression)**

Where:
- \( \hat{y}_p \) is the column vector of model predicted dependent variable values,
- \( x_p^T \) is the transposed explanatory variable matrix,
- \( \hat{\beta} \) is a column vector of regression coefficients.

If the linear model is correct the model predicted value is an unbiased estimate of the measured dependent variable value. The variance of the prediction is given by

\[ s_{\hat{y}}^2 = \sigma^2 \left( 1 + x_p^T \left( x_p \right)^T \right) \]

**Equation A3.17 Variance of Prediction (Multiple Regression)**
\[ \sigma^2 \left(1 + h_p \right) \]

Equation A3.17 Variance of Prediction (Alternative)

Where: \( \sigma^2 \) is estimated as the residual mean square, or \( s_e \).

The estimated standard error of prediction is the square root of the conditional variance figure given by Equation A3.17.

\[ s_\hat{y} = \sigma \sqrt{1 + h_p} \]

Equation A3.18 Estimated Standard Error of Prediction

**A3.8.3 Root Mean Square Error of Prediction, RMSE**

Another way to create a confidence interval is to use the validation error as an estimate of the expected error of prediction. For example, with leave 1 out cross validation, or the PRESS procedure, \( \text{RMSE}_v = \sqrt{\text{PRESS} / n_v} \) is the validation equivalent of the standard error of prediction, and if normality is assumed can be used in the same way as described for \( s_e \) or \( s_\hat{y} \) to place confidence bands at a desired significance level around the predictions. For example, an approximate 95\% confidence interval is \( \hat{y}_i \pm 2 \text{RMSE}_v \). For this project we will use the averaged RMSE values for a number periods of prediction outside of the construction data set as the \( \text{RMSE}_v \) value, and apply the intervals at \( \hat{y}_i \pm 2 \text{RMSE}_v \) as mentioned above to give an approximate 95\% confidence interval.
A3.9 Interpolation and Extrapolation in Regression Modelling

The regression equation is developed using a data set called the construction data set, calibration data set or learning data set. For this construction set the explanatory variables have a defined range. For example, in regressing end winding vibration on active power, the end winding vibration data may have a range between 110 and 200µm pk-pk. The relationship between the explanatory variable and the dependent variable expressed by the regression applies only when the explanatory variable values are similar to values present in the learning data set. When the model is applied to generate predictions for data outside of the construction data set an important question is how different can the explanatory variable data be from its values in the construction data set before the predictions are considered unreliable/invalid. Generally speaking, when the explanatory data are within the ranges of the construction data set but do not exactly match a particular condition, the predictions are called interpolations. If the explanatory data lies outside the ranges of the construction data set then the predictions are called extrapolations. Interpolations tend to be more accurate as there are similar data to estimate from within the construction data set. This may not necessarily be the case for extrapolation where there may be a completely different relationship between the dependent variable and the explanatory variables for that particular condition. Extrapolations, depending upon the specific investigation may give the most interesting results. In the case of this project extrapolations occur mainly during start-up, run-down and extraordinary operating duties which occur from time to time. However, instead of dumping the extrapolated values, these predictions can simply be flagged as prediction during known abnormal conditions and so can be subject to different criteria for analysis.
Appendix 4

Procedure for Recalculation of MLR Models in Matlab for Implementation on PI ACE/Process Book

A4.1 The Need for Recalculation of Model Coefficients

Recalculation of the regression models may be required when the predictive capabilities of the models begin to deteriorate, as indicated by some or all of the following symptoms:

1. Reduction in $R_{PRED}^2$ value over a defined period.
2. Increase in MSE/RMSE above an acceptable threshold.
3. Significant deterioration in the error spread (heteroskedasticity).
4. Increased bias in the error.
5. Increased error variance.

If the quality of the prediction of the regression models start to decay as detailed above a number of actions should be taken

1. Firstly look in the data for any abnormal operating conditions that would not be usual for the machine and that may not have been filtered from the prediction independent variable data.

2. Rule out a faulty FOA or other hardware problem.

3. Ensure that the data gathered from PI is good quality i.e. vibration value has not flat lined due to a communication breakdown or corruption of the database.

4. Ensure that any values calculated in PI ACE are being calculated properly and are not in error.
5. If the machine is two shifting very regularly this will lead to increased variance in the error as the machine is in a state of change on a very regular basis. Running at sub optimal temperatures etc.

6. If none of the conditions 1 - 4 exist i.e. the machine is running normally under standard operating conditions, then the deterioration may be down to a change in behaviour of the machine end windings which may be down to either, general wear and tear and ageing, or a problem has developed.

If all conditions listed 1-5 have been ruled out and no increase in average vibration levels is evident throughout the machine, and no physical damage is evident in the machine then a recalculation of the regression models is necessary to take account of the new machine fingerprint of behaviour.

When models start to perform poorly, a comparison with the other model types should be carried out to ascertain whether the drop off in performance is evident for all model types or just one. If the others are still performing that the poorly performing model must have some deficiency in terms of variable selection, coefficient calculation or possibly implementation (mistake in calculation).

**A4.2 Procedure for Calculation/Recalculation and Implementation of New MLR Model Coefficients**

**A4.2.1 Calculation/Recalculation of MLR Models using Matlab**

Gather data for the learning data set. The learning data should cover as wide a range of operating conditions as possible with the following constraints (which apply to data applied to the model when it is to be used for the purpose of predicting vibration levels also).

1. The data should cover all load points between minimum and maximum stable generation. Note that the machine may operate a loads slightly below the declared minimum stable generation value under some operating conditions i.e. min load
operation with high system frequency would mean that frequency influence/governor control of the machine would reduce load below Min Gen in order to reduce system frequency.

2. The data should include as wide a range of values for MVAr as is possible.

The two conditions already mentioned mean that the model covers as much of the normal operating conditions as possible

3. Data relating to Stat-up and Shut-down of the machine should be filtered out along with some data after start-up where the machine has not reached normal operating temperature if necessary.

4. Due to the reduction in the generator minimum stable generation (Min Gen) capability from approximately 140 MW to 112 MW the input data, particularly for the MLR models must be filtered as (in general) nonlinearities in the relationships between certain variables (temperature related) tend to appear at loads below 140 MW.

Once gathered arrange the data in column sin EXCEL and note the position of each in the array. The data labels are not imported into the data array in Matlab. Also ensure that the data is stored in Sheet 1 as this is where Matlab will by default import the data from (alternatively specify the sheet in the Matlab command).

Open a new workspace in Matlab and save it as GTGStepwise02022008 or similar. **Note: the file name must begin with a letter.**

In the Command window import the learning data excel file to Matlab using the xlsread function as follows:

```matlab
[learndata, headertext] = xlsread('D:\GTGStepwise\GTGlearndata.xls', 'Data', 'a1:x6944');
```

The data will appear as two arrays, one with the header text and one with the data array.
The array will now have to be broken up into its constituent variables. To do this, open the array and in the Array Editor right click the column header of the variable you want to create. From the menu click ‘Create Variable from Selection’. A new variable will appear in the workspace called ‘unnamed’. Rename the variable by right clicking on it and selecting ‘Rename’.

The names of the explanatory/independent variables must be renamed as follows:

- MW         GTG Load
- MVar       GTG Reactive Power
- IGV        GT Inlet Guide Vane Position
- Amb_Temp   Ambient Air Temperature
- Core_Temp  Average Stator Core Temperature
- Exciter_I  Rotor Excitation Current
- Stator_I   GTG Average Stator Current
- Voltage    Average Stator Voltage
- Frequency  System Frequency

The reasoning behind using these exact designations for the dependent variables is that interaction arrays were created using these ten variables with these designations assigned.

*Note: The designations for the new independent variables are listed in Section 9.2.*

The calculations for each Interaction Array (L1 – L4) are stored in a word document and can simply be copied and pasted into the Command Window in order to create the array, assuming of course that the names of the variables in the Matlab Workspace are exactly as set out above.

*Note: the L0 array is made up of the measured variables only i.e. no calculated variables.*

Otherwise the designations need to be changed in the word document to reflect the variable names used in the Matlab Workspace.
The next step is to copy and paste the interaction array calculations into the Command Window.

Four new arrays should result named Interactions1, Inteactions2, Interactions3 and Interactions4. The new arrays will have 65, 110, 240 and 789 columns respectively, each column corresponding to a dependent variable. All of the arrays include the 10 original variables listed above in columns 1 – 10.

With all of the data now in place the stepwise selection procedure can be employed to calculate a multiple linear regression model for each dependent variable for each of the interaction levels (L0 – L4).

Taking Connected End/Turbine End Coil Nose No.01 as an example,

In the Command Window enter the following:

```
stepwise(Interactions1,CE01);
```

The Stepwise GUI will now open

The blue dot in the Model History Plot shows the model with only the simplest term included i.e. the intercept.

There are two options for adding/removing independent variables from the model, ‘Next Step’ simply takes the next recommended step be that the addition or removal of a certain variable while ‘All Steps’ takes all recommended steps.

The best option is to select ‘All Steps’ and once the procedure is complete analyse the Model History Plot. The plot will show that for each step the RMSE value for the model reduces. As more steps are taken the improvement becomes less significant. The job of the analyst is now assess whether or not the model generated by taking all recommended steps is the ‘best model’ for implementation on new data. While the obvious answer to this question may seem to be yes, the reality is quite different.
The models to be applied to the new input data (data from outside of the learning data set) must be of good quality. Generally speaking, the models generated using the stepwise procedure (particularly when large numbers of variables are available for selection), using ‘all recommended steps’, tend to overfit the model to the data.

This overfitting, results in a model that is unstable when exposed to new data from outside of the construction data set.

It is best for the analyst to choose a model with a low RMSE value, but one that is also efficient for calculation.

Figure A4.1 Matlab Stepwise Selection GUI

Figure A4.1 depicts a stepwise selection of variables for MLR for TE/CE01 with an interaction level of 4 (789 variables available for selection).

It is evident that for the addition of the first 10 or so variables there is a marked improvement in the RMSE value of the model. There is then a slow but steady improvement in the RMSE figure for the addition of new independent variables until about 40 are included.

The RMSE figure with 39 variables included from the active window shown in Figure A4.1 is 5.93 while the RMSE figure for the model with 220+ variables included is 5.08. There is some improvement for between 40 and 180 variables included but it is minor. For the case
shown in Figure 4.1 the model number 20 was chosen. It was chosen for the following reasons:

- Improvement in accuracy for additional variables beyond this point leads to overfitting of the data

- The intercept values of the more complex models tend to be very large in magnitude, this model was chosen as the intercept is not extremely large.

From 180 variables on there is negligible improvement in the predictive capability of the model (for the construction data set)

Once a model has been chosen the data relating to it can be exported to the workspace by clicking on export and selecting the model parameters to be exported. As a minimum, coefficients (beta) and statistics (stats) should be selected so that the intercept, and coefficients for each independent variable included in the model are available in the workspace.

To apply the model to new data simply copy the statistics and coefficient data into the workspace of the Matlab file for the new (outside of construction set) data.

Ensure the variable designations of the new data variables are in exactly the same format as those of the learning data.

Paste in the interaction calculations as per the calculation of the initial model.

```
CE01Pred=(Interactions4*CE01I1beta)+CE01I1stats.intercept;
```

This will yield a prediction for CE01.
A4.2.2 Updating the Models in PI ACE

To change the model in PI ACE, write the model equation in the required form (as per the existing model e.g. beta*Pltagname^2 etc. and simply paste it into PI ACE in place of the current model equation.

The format of the equation will be as follows:

\[ \text{TE25Pred.Value()} = \text{intercept} + ((\text{Beta1}\times\text{Variable1}) + (\text{Beta2}\times\text{Variable2})) \]

Add conditions to the particular model as required e.g. **For/ If** PI11MBY10CE901.XQ01 **greater than** 140 **Then** apply the model, **Else** Model = 0 etc.

**Note:** _PI11MBY10CE901.XQ01 is the PI designation for GT Generator Power Output (MW)._
The condition for implementation of the model shown in Figure A4.2 is simply that the system frequency is greater than 49 Hz i.e. the generator is online. Normally a number of conditions apply to each model which are specifically tailored to that model.

Note: With the stepwise selection complete and checked for insignificant variables (P>0.05), a check for multicolinearity in the selected variables should be carried out. This check is not carried out automatically as part of the stepwise selection procedure in Matlab. The selected variables can be input to EXCEL and a Multiple Linear Regression calculated using (using XLSTAT) the stepwise selected variables which will automatically check for multicolinearity and remove from the regression the variable in question. Alternatively a manual check can be carried out in either EXCEL or Matlab.
Appendix 5


A5.1 2008 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

The Multiple Linear Regression models used for 2008 were of the same type as were used in 2007 i.e. the same measured independent variables are used and the same calculated independent variable data sets/arrays (i.e. L0, L1, L2, L3 & L4) are used. Table A5.1 displays goodness of fit data for TE09 Multiple Linear Regression (MLR) models for interaction levels L0 to L4 as well as for the LOWESS Model.

The LOWESS Model in this case was calculated using the LOWESS method (Robust LOWESS is also an option) with a Polynomial Degree setting of 2, the percentage of the Learning data over which the polynomial is implemented at a time is 2% (True = 2%). 2% of the Learning Sample of 9615 observations/data points equates to 192 observations approximately. The kernel used is the Triweight kernel.

It can be seen that the accuracy of prediction is relatively good throughout but with a notable trough in performance for the periods W100103, W112115 and W116119. For the Period W100103 the machine was regularly operated in cycling/two-shifting mode. During the periods W112115 and W116119 (cover 05/01/2009 to 02/02/2009 and 02/02/2009 to 02/03/2009 respectively) ambient temperatures were low enough to cause the relationship between measured temperature values and other measured dependent and independent variable values to change significantly. The assumption on which the models were calculated therefore no longer hold and so predictive accuracy is lost. As a result the data must be filtered to remove the data where low ambient temperatures exist.
### Goodness of Fit Statistics TE09 from W084087 to W132135 (30/06/2008 to 22/06/2009)

<table>
<thead>
<tr>
<th></th>
<th>W084087</th>
<th>W088091</th>
<th>W092095</th>
<th>W096099</th>
<th>W100103</th>
<th>F W112115</th>
<th>F W116119</th>
<th>F W120123</th>
<th>F W124127</th>
<th>W128131</th>
<th>W132135</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L0 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>105.71</td>
<td>99.31</td>
<td>101.66</td>
<td>99.31</td>
<td>99.02</td>
<td>99.11</td>
<td>106.05</td>
<td>109.04</td>
<td>114.55</td>
<td>112.97</td>
<td>114.31</td>
</tr>
<tr>
<td>RMSE</td>
<td>5.39</td>
<td>7.62</td>
<td>9.01</td>
<td>10.41</td>
<td>18.03</td>
<td>15.79</td>
<td>16.20</td>
<td>11.69</td>
<td>13.57</td>
<td>15.42</td>
<td>11.62</td>
</tr>
<tr>
<td>TE09 L0 R Squared</td>
<td>0.95</td>
<td>0.90</td>
<td>0.89</td>
<td>0.87</td>
<td>0.60</td>
<td>0.64</td>
<td>0.67</td>
<td>0.82</td>
<td>0.80</td>
<td>0.72</td>
<td>0.80</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.79</td>
<td>0.69</td>
<td>0.66</td>
<td>0.63</td>
<td>0.37</td>
<td>0.40</td>
<td>0.43</td>
<td>0.58</td>
<td>0.55</td>
<td>0.47</td>
<td>0.55</td>
</tr>
<tr>
<td><strong>L1 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>105.50</td>
<td>97.82</td>
<td>100.79</td>
<td>99.37</td>
<td>99.29</td>
<td>101.19</td>
<td>107.84</td>
<td>106.91</td>
<td>115.06</td>
<td>114.42</td>
<td>113.33</td>
</tr>
<tr>
<td>TE09 L1 R Squared</td>
<td>0.95</td>
<td>0.86</td>
<td>0.87</td>
<td>0.86</td>
<td>0.62</td>
<td>0.76</td>
<td>0.67</td>
<td>0.78</td>
<td>0.77</td>
<td>0.69</td>
<td>0.76</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.78</td>
<td>0.62</td>
<td>0.64</td>
<td>0.63</td>
<td>0.39</td>
<td>0.51</td>
<td>0.43</td>
<td>0.53</td>
<td>0.52</td>
<td>0.44</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>L2 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>105.75</td>
<td>98.64</td>
<td>101.29</td>
<td>99.57</td>
<td>99.23</td>
<td>101.06</td>
<td>107.40</td>
<td>107.15</td>
<td>115.29</td>
<td>114.49</td>
<td>113.11</td>
</tr>
<tr>
<td>TE09 L2 R Squared</td>
<td>0.96</td>
<td>0.88</td>
<td>0.88</td>
<td>0.86</td>
<td>0.62</td>
<td>0.75</td>
<td>0.67</td>
<td>0.77</td>
<td>0.77</td>
<td>0.69</td>
<td>0.76</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.79</td>
<td>0.65</td>
<td>0.65</td>
<td>0.63</td>
<td>0.38</td>
<td>0.50</td>
<td>0.42</td>
<td>0.52</td>
<td>0.52</td>
<td>0.44</td>
<td>0.51</td>
</tr>
<tr>
<td><strong>L3 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>106.07</td>
<td>98.91</td>
<td>101.32</td>
<td>99.08</td>
<td>98.75</td>
<td>99.02</td>
<td>103.89</td>
<td>106.55</td>
<td>114.16</td>
<td>113.73</td>
<td>114.06</td>
</tr>
<tr>
<td>RMSE</td>
<td>5.03</td>
<td>8.02</td>
<td>8.55</td>
<td>10.03</td>
<td>18.32</td>
<td>11.54</td>
<td>14.49</td>
<td>13.16</td>
<td>13.82</td>
<td>15.84</td>
<td>12.40</td>
</tr>
<tr>
<td>TE09 L3 R Squared</td>
<td>0.96</td>
<td>0.89</td>
<td>0.90</td>
<td>0.87</td>
<td>0.59</td>
<td>0.81</td>
<td>0.74</td>
<td>0.76</td>
<td>0.79</td>
<td>0.70</td>
<td>0.77</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.80</td>
<td>0.67</td>
<td>0.68</td>
<td>0.65</td>
<td>0.36</td>
<td>0.56</td>
<td>0.49</td>
<td>0.51</td>
<td>0.54</td>
<td>0.45</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>L4 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>105.92</td>
<td>98.86</td>
<td>102.18</td>
<td>99.64</td>
<td>98.81</td>
<td>99.49</td>
<td>104.93</td>
<td>107.31</td>
<td>113.99</td>
<td>111.90</td>
<td>114.12</td>
</tr>
<tr>
<td>RMSE</td>
<td>5.37</td>
<td>7.81</td>
<td>8.00</td>
<td>9.06</td>
<td>17.64</td>
<td>10.96</td>
<td>14.27</td>
<td>14.10</td>
<td>13.77</td>
<td>14.01</td>
<td>13.14</td>
</tr>
<tr>
<td>TE09 L4 R Squared</td>
<td>0.95</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
<td>0.62</td>
<td>0.83</td>
<td>0.75</td>
<td>0.73</td>
<td>0.79</td>
<td>0.77</td>
<td>0.75</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.79</td>
<td>0.68</td>
<td>0.70</td>
<td>0.68</td>
<td>0.38</td>
<td>0.58</td>
<td>0.50</td>
<td>0.48</td>
<td>0.54</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>LOWESS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>105.03</td>
<td>104.27</td>
<td>107.54</td>
<td>103.89</td>
<td>107.28</td>
<td>95.15</td>
<td>103.82</td>
<td>101.17</td>
<td>108.27</td>
<td>102.63</td>
<td>109.76</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>105.37</td>
<td>99.95</td>
<td>103.20</td>
<td>100.98</td>
<td>99.14</td>
<td>99.21</td>
<td>105.70</td>
<td>107.07</td>
<td>114.55</td>
<td>112.10</td>
<td>114.22</td>
</tr>
<tr>
<td>RMSE</td>
<td>4.17</td>
<td>7.69</td>
<td>8.00</td>
<td>9.48</td>
<td>17.66</td>
<td>11.26</td>
<td>16.03</td>
<td>12.10</td>
<td>12.51</td>
<td>13.43</td>
<td>12.20</td>
</tr>
<tr>
<td>TE09 LOWESS R Squared</td>
<td>0.97</td>
<td>0.90</td>
<td>0.91</td>
<td>0.89</td>
<td>0.62</td>
<td>0.82</td>
<td>0.68</td>
<td>0.74</td>
<td>0.83</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.83</td>
<td>0.68</td>
<td>0.70</td>
<td>0.67</td>
<td>0.38</td>
<td>0.57</td>
<td>0.43</td>
<td>0.49</td>
<td>0.58</td>
<td>0.53</td>
<td>0.53</td>
</tr>
</tbody>
</table>

### Table A5.1 Goodness of Fit Statistics for TE09 from W084087 to W132135 for MLR Models L0 – L4 and LOWESS

The R-squared value for the TE09 models average at approximately 0.79 which is a good overall score.

The best performing model (as defined in Section 6.10) over the 64 week period was the L4 model (with an average R-squared value of 0.81 and average RMSE of 11.93 µm pk-pk) and the LOWESS model (with an average R-squared value of 0.81 and average RMSE of 11.32 µm pk-pk) followed closely by the L3 MLR model (with an average R-squared value of 0.80 and average RMSE of 11.93µm pk-pk).
Figure A5.1 shows R-squared values for Turbine End End-Coil Number 09 MLR and LOWESS predicted vibration data. The best performing (as defined in Section 6.10) MLR model (L4) and the LOWESS model values are displayed in red and orange respectively.

![TE09 R-squared Statistic for MLR Models L0, L1, L2, L3, L4 and the LOWESS Model](image)

Figure A5.1 TE09 R-squared Values from W084087 to W132135 for MLR L0 – L4 and LOWESS Models

Figure A5.2 shows Turbine End End-Coil Number 9 (TE09) Measured and Model Predicted Vibration Data using MLR L0 – L4 and LOWESS Models for W128131 (27/04/2009 to 25/05/2009). It is evident that the measured vibration data series is closely matched by the regression model predicted data series’.

Table A5.2 shows goodness of fit statistics (RMSE and R-Squared only) over the period W084087 to W132135 for the best performing MLR model and LOWESS model for each of the twelve end coils where vibration levels are measured.
From Table A5.2 it is apparent that predictive capability is satisfactory for the twelve end coils, EE03, TE17 and TE25 apart. As with the 2007 data, models relating to EE03 perform well over the first 3 to 6 months after which the GOF scores (R-Squared) fall away from an average of approximately 0.75 to approximately 0.45. From W124127 onwards the GOF scores for both the MLR and LOWESS models deteriorate to near zero falling into negative figures coming up to the 2009 Annual Overhaul at W132135.

Models for TE17 perform similarly to EE03 with R-Squared values over the first five months averaging about 0.7. Average values beyond W112115 are approximately 0.4. Most of the remaining end coils have relatively consistent GOF scores across the period covered by the 2008 model.

Goodness of fit scores for TE25 were relatively consistent (but low) over the entire period, averaging approximately 0.3 for the MLR model and 0.4 for the LOWESS model.
## 2008 MLR and LOWESS Best Performing Models Goodness of Fit Statistics (RMSE and R-Squared Only)

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
<th>RMSE</th>
<th>R Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE01 LOWESS</td>
<td>0.73</td>
<td>0.91</td>
<td>0.93</td>
<td>0.90</td>
<td>0.87</td>
<td>0.70</td>
<td>0.60</td>
<td>0.61</td>
<td>0.70</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE09 L4</td>
<td>5.37</td>
<td>7.81</td>
<td>8.00</td>
<td>9.06</td>
<td>17.64</td>
<td>10.96</td>
<td>14.27</td>
<td>14.10</td>
<td>13.77</td>
<td>14.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE09 LOWESS</td>
<td>0.95</td>
<td>0.90</td>
<td>0.91</td>
<td>0.90</td>
<td>0.62</td>
<td>0.83</td>
<td>0.75</td>
<td>0.73</td>
<td>0.79</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE17 LOWESS</td>
<td>0.62</td>
<td>0.76</td>
<td>0.71</td>
<td>0.69</td>
<td>0.71</td>
<td>0.59</td>
<td>0.47</td>
<td>0.33</td>
<td>0.27</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE25 L1</td>
<td>4.11</td>
<td>5.31</td>
<td>6.53</td>
<td>7.01</td>
<td>8.59</td>
<td>8.79</td>
<td>8.51</td>
<td>7.46</td>
<td>7.72</td>
<td>6.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE25 LOWESS</td>
<td>0.78</td>
<td>0.54</td>
<td>0.35</td>
<td>0.27</td>
<td>0.27</td>
<td>0.23</td>
<td>0.28</td>
<td>0.40</td>
<td>0.37</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE33 L4</td>
<td>4.79</td>
<td>5.93</td>
<td>8.35</td>
<td>10.58</td>
<td>12.44</td>
<td>18.80</td>
<td>9.15</td>
<td>7.44</td>
<td>8.92</td>
<td>8.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE33 LOWESS</td>
<td>0.85</td>
<td>0.77</td>
<td>0.62</td>
<td>0.51</td>
<td>0.66</td>
<td>0.41</td>
<td>0.65</td>
<td>0.33</td>
<td>0.67</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE41 L3</td>
<td>6.73</td>
<td>8.08</td>
<td>8.61</td>
<td>8.00</td>
<td>11.30</td>
<td>14.80</td>
<td>12.89</td>
<td>12.93</td>
<td>18.09</td>
<td>19.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE41 LOWESS</td>
<td>0.96</td>
<td>0.93</td>
<td>0.91</td>
<td>0.91</td>
<td>0.84</td>
<td>0.69</td>
<td>0.43</td>
<td>0.42</td>
<td>0.25</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE03 L4</td>
<td>4.51</td>
<td>5.12</td>
<td>5.11</td>
<td>5.68</td>
<td>6.94</td>
<td>6.62</td>
<td>6.67</td>
<td>7.29</td>
<td>7.71</td>
<td>7.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE03 LOWESS</td>
<td>0.81</td>
<td>0.76</td>
<td>0.76</td>
<td>0.70</td>
<td>0.43</td>
<td>0.48</td>
<td>0.47</td>
<td>0.37</td>
<td>0.14</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE11 L4</td>
<td>3.62</td>
<td>4.72</td>
<td>5.33</td>
<td>7.30</td>
<td>6.84</td>
<td>7.09</td>
<td>12.66</td>
<td>12.71</td>
<td>7.35</td>
<td>6.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE11 LOWESS</td>
<td>0.90</td>
<td>0.74</td>
<td>0.71</td>
<td>0.61</td>
<td>0.69</td>
<td>0.66</td>
<td>0.50</td>
<td>0.50</td>
<td>0.26</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE19 L4</td>
<td>3.62</td>
<td>4.72</td>
<td>5.33</td>
<td>7.30</td>
<td>6.84</td>
<td>7.09</td>
<td>12.66</td>
<td>12.71</td>
<td>7.35</td>
<td>6.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE19 LOWESS</td>
<td>0.90</td>
<td>0.74</td>
<td>0.71</td>
<td>0.61</td>
<td>0.69</td>
<td>0.66</td>
<td>0.50</td>
<td>0.50</td>
<td>0.26</td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE27 L2</td>
<td>4.50</td>
<td>5.90</td>
<td>5.52</td>
<td>6.62</td>
<td>5.72</td>
<td>10.77</td>
<td>NO DATA</td>
<td>7.04</td>
<td>11.22</td>
<td>5.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE27 LOWESS</td>
<td>0.94</td>
<td>0.83</td>
<td>0.89</td>
<td>0.81</td>
<td>0.89</td>
<td>0.77</td>
<td>NO DATA</td>
<td>0.77</td>
<td>0.57</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE35 L4</td>
<td>5.35</td>
<td>5.90</td>
<td>5.39</td>
<td>7.29</td>
<td>7.29</td>
<td>6.08</td>
<td>7.35</td>
<td>10.23</td>
<td>5.64</td>
<td>7.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE35 LOWESS</td>
<td>0.79</td>
<td>0.66</td>
<td>0.81</td>
<td>0.80</td>
<td>0.84</td>
<td>0.64</td>
<td>0.57</td>
<td>0.81</td>
<td>0.88</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE43 L3</td>
<td>4.47</td>
<td>5.34</td>
<td>5.57</td>
<td>7.85</td>
<td>7.11</td>
<td>9.96</td>
<td>8.34</td>
<td>12.74</td>
<td>8.76</td>
<td>8.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE43 LOWESS</td>
<td>0.95</td>
<td>0.91</td>
<td>0.92</td>
<td>0.81</td>
<td>0.84</td>
<td>0.69</td>
<td>0.81</td>
<td>0.22</td>
<td>0.76</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A5.2 Goodness of Fit Statistics for all End Coils from W084087 to W132135 for the Best Performing MLR Models and the LOWESS Model.
A5.2 2009 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

Models were generated on two separate occasions in 2009. The first, post AOH 09 and the second following a change in generation capability in October 2009 the details of which are discussed later in the chapter.

The Multiple Linear Regression models used for 2009 were of the same type as were used in 2008 i.e. the same measured independent variables are used and the same calculated independent variable data are used. Table A5.3 displays goodness of fit data for the Multiple Linear Regression (MLR) models for interaction levels L0 to L4 as well as for the LOWESS Model.

The LOWESS Model in this case was calculated using the LOWESS method (Robust LOWESS is also an option) with a Polynomial Degree setting of 2, the percentage of the Learning data over which the polynomial is implemented at a time is 2 % (True = 2 %). 2 % of the Learning Sample of 2346 observations/data points equates to 47 observations approximately. The kernel used is the Triweight kernel.

It can be seen from the data displayed Table A5.3 that goodness of fit scores are relatively good over the first 5 data sets (W136159 to W152155) but then deteriorate dramatically. This was caused by a calculation error on the PI database.

In particular the LOWESS data does not recover even when the data is filtered to remove the affected data. This is due to the percentage of the learning data that the polynomial is applied to is set at 2 % with a much smaller learning sample size than for previous models. The setting should be set at 5 – 6 % for this learning data set to give 117 or 141 observations respectively.
Table A5.3 Goodness of Fit Statistics for TE01 from W136139 to W160163 for MLR Models L0 – L4 and LOWESS

Figure A5.3 shows measured against MLR predicted EWV for Turbine End End-Coil Number 1 (located at the bottom dead centre position approximately). The very erratic MLR calculated data values starting from 01/12/2009 (highlighted in green) are caused by an error in the Average Stator Current calculation on the PI system.

No data is displayed for the period highlighted in red as the machine was instructed OFF for a number of days.
The reason behind the very poor performance of some models (e.g. TE01 L0 and L3) and the better performance of others (e.g. TE01 L1) in this case comes down to variable selection in the stepwise process. If Average Stator Voltage is included as an independent variable in a given MLR model, then predictions related to that model will become inaccurate if the Average Stator Voltage data is compromised in any way. The degree to which the model predictive capability is reduced depends upon the significance (weighting relative to the other independent variables) of the variable within the model.

Table A5.4 Goodness of Fit Statistics for TE01W156159 (Unfiltered) for MLR Models L0 – L4 and LOWESS Models using a True setting of 2% and 6% with Polynomial Degree set at 2.

| Goodness of Fit Statistics TE01 for W156159 Unfiltered |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                  | TE01 L0         | TE01 L1         | TE01 L2         | TE01 L3         | TE01 L4         | TE01 LOWESS    |
|                  |                 |                 |                 |                 |                 | TRUE 2% PD 2   |
| Measured Mean    | 103.87          | 103.87          | 103.87          | 103.87          | 103.87          | 103.87          |
| Predicted Mean   | 101.57          | 101.19          | 101.19          | 94.09           | 98.58           | 99.11           |
| RMSE             | 101.19          | 101.19          | 94.09           | 98.58           | 99.11           | 91.40           |
| R Squared        | -5.21           | 0.44            | 0.44            | -6.07           | 0.15            | -12.27          |
| Matlab Fit       | -1.49           | 0.25            | 0.25            | -1.66           | 0.08            | -2.64           |

Appendix 5  Page 266  Model Performance Data
From Tables A5.4 and A5.5 it is evident that the miscalculation of one of the independent variable values leads to a dramatic change in the predicted values of the models where it is one of the most significant independent variables included. For models where the particular variable is excluded or is less significant in terms of calculating the predicted vibration value the change to predictive capability is less pronounced.

Table A5.5 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True setting of 2% and 6% with Polynomial Degree set at 2.

| Goodness of Fit Statistics TE01 for W156159 Filtered (09/11/2009 to 30/11/2009) |
|----------------------------------|-------|-------|-------|-------|-------|-------|-------|
|                                  | TE01 L0 | TE01 L1 | TE01 L2 | TE01 L3 | TE01 L4 | TE01 LOWESS TRUE 2% PD 2 | TE01 LOWESS TRUE 6% PD 2 |
| Measured Mean                    | 100.80  | 100.80  | 100.80  | 100.80  | 100.80  | 100.80                        | 100.80                        |
| Predicted Mean                   | 96.88   | 98.00   | 98.00   | 96.27   | 98.46   | 104.75                        | 99.79                        |
| RMSE                              | 21.47   | 17.61   | 17.61   | 18.31   | 19.36   | 31.91                         | 18.24                        |
| R Squared                         | 0.16    | 0.44    | 0.44    | 0.39    | 0.32    | -0.85                         | 0.40                         |
| Matlab Fit                        | 0.09    | 0.25    | 0.25    | 0.22    | 0.18    | -0.36                         | 0.22                         |

All Data after 30/11/2009 Removed.

Figure A5.4 TE01 Measured and Predicted Vibration for W156159 (09/11/2009 to 13/12/2009).

Post Recalculation of Independent Variable Data (Average Stator Current).
Table A5.6 displays goodness of fit data for TE01 for W156159 with Average Stator Current values corrected and no filter applied. As would be expected the goodness of fit scores are significantly better than for the unrepaired data whose GOF scores are displayed in Table A5.4. However, the scores overall are quite low relative to previous data for the same end coil.

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics TE01 for W156159 (Unfiltered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Measured Mean</td>
</tr>
<tr>
<td>Predicted Mean</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>R Squared</td>
</tr>
<tr>
<td>Matlab Fit</td>
</tr>
</tbody>
</table>

Table A5.6 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True Setting of 2 %, 6 %, 10 % and 15 % with Polynomial Degree set at 2 and a True Setting of 10 % with Polynomial Degree set at 1.

The reason behind the lower than average GOF scores was again low ambient temperature which in turn lead to lower than normal generator slot temperatures and deviance from the normal relationships between measured temperature variables and other measured variables which were used to calculate the model coefficients, polynomials etc.

<table>
<thead>
<tr>
<th>Goodness of Fit Statistics TE01 for W156159 (Filtered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Measured Mean</td>
</tr>
<tr>
<td>Predicted Mean</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>R Squared</td>
</tr>
<tr>
<td>Matlab Fit</td>
</tr>
</tbody>
</table>

Table A5.7 Goodness of Fit Statistics for TE01W156159 (Filtered) for MLR Models L0 – L4 and LOWESS Models using a True Setting of 2 %, 6 %, 10 % and 15 % with Polynomial Degree set at 2 and a True Setting of 10 % with Polynomial Degree set at 1. Data filtered for Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C.

Table A5.7 shows GOF statistics for W156159 with a filter applied to data where Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C. It is immediately evident that the GOF statistics values improve very significantly. Leaving out values for L0 which improved by 1550 % and TE01 LOWESS TRUE 2 % PD 2 which gave negative R-squared and Matlab Fit values throughout, the average improvement in GOF scores was
approximately 63 % (the highest being TE01 L4 at 122 %, the lowest being TE01 LOWESS TRUE 6 % PD 2 at 34.1 %).

Tables A5.4 through to A5.7 also display GOF statistics values for LOWESS Models using different TRUE and PD settings than those used to give LOWESS GOF statistics values in Table A5.3. The data shows that there is a marked improvement in model performance when the TRUE setting is changed from 2 % up to 6 %. For TRUE setting set at 10 % and above the GOF scores continue to improve but the level of improvement reduces sharply as the TRUE value increases further. If the TRUE setting is increased further the improvement in performance of the model eventually gives way to a collapse in predictive capability as shown in Table A5.8 where the TRUE setting is at 50 %.

<table>
<thead>
<tr>
<th>TE01 Goodness of Fit Statistics W156159</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
</tr>
<tr>
<td>Predicted Mean</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>R Squared</td>
</tr>
<tr>
<td>Matlab Fit</td>
</tr>
</tbody>
</table>

Table A5.8 Goodness of Fit Statistics for TE01W156159 (Filtered) for LOWESS Models.

True Setting of 50 % Polynomial Degree set at 2.

Data Filtered for Average Slot Temperature ≥ 60 °C and Ambient Air Temperature ≥ 6 °C.

The TRUE setting value at which the predictive capability of the model will collapse varies depending upon the number of observation in the learning data set. The larger the learning data set used the smaller the TRUE value used should be. For instance, for this project the setting is set at 1 - 2 % for large learning data sets (5000 - 10000 observations approximately) and 5 – 6 % for smaller learning data sets (2000 - 3000 observations approximately)
### Table A5.9 Goodness of Fit Statistics for All End Coils from W136139 to W160163 for the Best Performing MLR Models and the LOWESS Model.

<table>
<thead>
<tr>
<th>Coils</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
<th>RMSE</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>W136139</td>
<td>11.47</td>
<td>0.92</td>
<td>19.15</td>
<td>0.78</td>
<td>7.16</td>
<td>0.93</td>
<td>8.24</td>
<td>0.91</td>
<td>8.48</td>
<td>0.43</td>
<td>7.70</td>
<td>0.53</td>
<td>5.64</td>
<td>0.36</td>
</tr>
<tr>
<td>W140143</td>
<td>11.50</td>
<td>0.92</td>
<td>16.64</td>
<td>0.85</td>
<td>7.53</td>
<td>0.94</td>
<td>14.26</td>
<td>0.72</td>
<td>8.74</td>
<td>0.39</td>
<td>8.31</td>
<td>0.54</td>
<td>0.78</td>
<td>0.36</td>
</tr>
<tr>
<td>W144147</td>
<td>14.99</td>
<td>0.88</td>
<td>26.49</td>
<td>0.63</td>
<td>8.77</td>
<td>0.93</td>
<td>13.49</td>
<td>0.72</td>
<td>9.10</td>
<td>0.59</td>
<td>9.38</td>
<td>0.54</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>W148151</td>
<td>15.33</td>
<td>0.79</td>
<td>20.72</td>
<td>0.62</td>
<td>10.44</td>
<td>0.88</td>
<td>11.52</td>
<td>0.72</td>
<td>9.28</td>
<td>0.31</td>
<td>9.37</td>
<td>0.54</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>W152155</td>
<td>17.42</td>
<td>0.67</td>
<td>21.72</td>
<td>0.49</td>
<td>12.08</td>
<td>0.78</td>
<td>14.66</td>
<td>0.67</td>
<td>12.97</td>
<td>0.21</td>
<td>7.94</td>
<td>0.42</td>
<td>0.63</td>
<td>0.36</td>
</tr>
<tr>
<td>W156159F</td>
<td>18.40</td>
<td>0.42</td>
<td>18.07</td>
<td>0.48</td>
<td>14.24</td>
<td>0.57</td>
<td>17.87</td>
<td>0.30</td>
<td>7.23</td>
<td>0.15</td>
<td>7.65</td>
<td>0.02</td>
<td>0.73</td>
<td>0.30</td>
</tr>
<tr>
<td>W160163F</td>
<td>16.62</td>
<td>0.70</td>
<td>13.05</td>
<td>0.75</td>
<td>10.52</td>
<td>0.69</td>
<td>12.32</td>
<td>0.02</td>
<td>7.84</td>
<td>0.06</td>
<td>7.65</td>
<td>0.02</td>
<td>0.73</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Note: RMSE and R-Squared values are calculated for each coil using the best performing MLR models and the LOWESS model. The values are rounded to two decimal places for clarity.
From Table A5.9 it is apparent that the predictive capability for the chosen MLR models at each of the twelve measurement points, as evaluated by $R^2$ and RMSE Goodness of Fit (GOF) scores, were strong over the period covered i.e. W136 to W163 (22/06/2009 to 11/01/2010). $R^2$ values for most measurement points averaged better than 0.7, with the exception of TE17 and TE25. The performance of the LOWESS models were less impressive with $R^2$ values generally lower and RMSE values generally higher than for the MLR models. In contradiction to data for 2007 and 2008, models relating to EE03 perform well over the entire period with the notable exception of W156159 where weather conditions were harsh. Indeed all models for all end coils see a dip in GOF scores for W156159.

Models for TE17 and TE25 perform similarly to each other, performing well over the first four to five periods before gradually falling away to zero over the remainder of the period. R-Squared values over the first four to five months average about 0.6, but fall away sharply after W152155 and fail to recover.

Most of the remaining end coils have relatively consistent GOF scores across the period covered by the 2009 model.
A5.3 2009 - 2010 Multiple Linear Regression and LOWESS Non-Parametric Regression Models

Multiple Linear Regression and LOWESS models were recalculated using new learning data. Due to the need for more flexibility from generators onto the national grid some improvements were made to HNC in order to bring the machine into line with Grid Code requirements relating to Minimum Stable Generation capability e.g. the station must have a MSG capability of ≤50% of plated maximum generation capability. Not having the ability to achieve this value results in a charge to the generator by the Transmission System Operator (TSO) i.e. Eirgrid. The change made on this occasion, which directly affects operation of the Generator was to reduce the Minimum Stable Generation capability of the overall plant from 216 MW to 200 MW (143 MW to 132 MW approximately for the GT Generator). These reduced values for minimum generation will for the most part be filtered out, particularly during the winter months due to increased non-linearity in the relationships between vibration and other generator related measured variables used in creating the MLR and LOWESS models.

The main reason behind the recalculation on this occasion was to generate new models to better predict the behaviour of TE25 and to a lesser degree TE17 and EE03 from January 2010 through to AOH10. The models are created using the same methodology as mentioned previously in the chapter. The period covered by these models is from 18/01/2010 to 04/05/2010. The 2010 Annual Overhaul (AOH10) began on 04/05/2010.

The performance of the MLR and LOWESS models was in line with those created previously. From Table A5.10 it is evident that all of the MLR models perform strongly over the entire period. The best performing (as defined in Section 6.10) MLR model was the L4 model with an average R-Squared value of approximately 0.79. The LOWESS model also performs reasonably well with an average R-Squared value of 0.72. It is also evident that for some TE01 models (L3, L4 and LOWESS in particular) there is degradation in the GOF Statistic values for W167171. The main reason for this is low ambient temperature. Ambient temperature for W167171 is below 4 °C for approximately 47% of the period as opposed to 28% for W164167, 15.6% for W172175 and 2.5% for W176AOH10.
Table A5.10 Goodness of Fit Statistics for TE01 from W164167 to W176AOH10 (18/01/2010 to 14/05/2010)

<table>
<thead>
<tr>
<th></th>
<th>W164167</th>
<th>W168171</th>
<th>W172175</th>
<th>W176AOH10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L0 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>94.85</td>
<td>107.17</td>
<td>93.71</td>
<td>101.03</td>
</tr>
<tr>
<td>RMSE</td>
<td>11.44</td>
<td>9.75</td>
<td>9.03</td>
<td>11.24</td>
</tr>
<tr>
<td>TE09 I0 R Squared</td>
<td>0.74</td>
<td>0.62</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.49</td>
<td>0.39</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>L1 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>97.37</td>
<td>105.03</td>
<td>93.66</td>
<td>105.70</td>
</tr>
<tr>
<td>RMSE</td>
<td>7.73</td>
<td>10.25</td>
<td>9.87</td>
<td>11.04</td>
</tr>
<tr>
<td>TE09 I1 R Squared</td>
<td>0.57</td>
<td>0.59</td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.34</td>
<td>0.36</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>L2 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>97.37</td>
<td>105.03</td>
<td>93.66</td>
<td>105.70</td>
</tr>
<tr>
<td>RMSE</td>
<td>14.87</td>
<td>10.23</td>
<td>9.87</td>
<td>11.04</td>
</tr>
<tr>
<td>TE09 I2 R Squared</td>
<td>0.57</td>
<td>0.59</td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.34</td>
<td>0.36</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>L3 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>97.31</td>
<td>105.37</td>
<td>94.36</td>
<td>105.39</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.95</td>
<td>10.00</td>
<td>9.31</td>
<td>10.19</td>
</tr>
<tr>
<td>TE09 I3 R Squared</td>
<td>0.62</td>
<td>0.60</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.38</td>
<td>0.37</td>
<td>0.56</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>L4 Interaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>95.34</td>
<td>104.97</td>
<td>93.81</td>
<td>102.07</td>
</tr>
<tr>
<td>RMSE</td>
<td>9.83</td>
<td>8.95</td>
<td>8.62</td>
<td>10.96</td>
</tr>
<tr>
<td>TE09 I4 R Squared</td>
<td>0.81</td>
<td>0.68</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.57</td>
<td>0.44</td>
<td>0.60</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>LOWESS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured Mean</td>
<td>85.19</td>
<td>103.63</td>
<td>95.17</td>
<td>105.70</td>
</tr>
<tr>
<td>Predicted Mean</td>
<td>93.08</td>
<td>105.12</td>
<td>93.52</td>
<td>101.98</td>
</tr>
<tr>
<td>RMSE</td>
<td>10.45</td>
<td>10.04</td>
<td>10.33</td>
<td>13.81</td>
</tr>
<tr>
<td>TE09 LOWESS R Squared</td>
<td>0.79</td>
<td>0.60</td>
<td>0.77</td>
<td>0.73</td>
</tr>
<tr>
<td>Matlab Fit</td>
<td>0.54</td>
<td>0.37</td>
<td>0.52</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Displayed in Figures A5.4 and A5.5, are R-Squared and RMSE GOF Statistic values for W164167 to W176180 at Turbine-End End-Coil Number 9 (TE09). The series associated with the best performing (as defined in Section 6.10) MLR Model (L4) is highlighted in red, while the series related to the LOWESS Model is highlighted in orange. Performance as measured by the various GOF statistics used throughout is consistently good over the four periods.
Figure A5.5 TE09 R-Squared Values for MLR Models L0, L1, L2, L3 & L4 as well as the LOWESS Model W164167 to W176180 (18/01/2010 to 04/05/2010 AOH 10)

Figure A5.6 TE09 RMSE Values for MLR Models L0, L1, L2, L3 & L4 as well as the LOWESS Model W164167 to W176180 (18/01/2010 to 04/05/2010 AOH 10)
<table>
<thead>
<tr>
<th>2009-2010 MLR and LOWESS Models GOF Statistics (RMSE and R Squared)</th>
<th>W164167</th>
<th>W168171</th>
<th>W172175</th>
<th>W176AOH10</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE01 L2</td>
<td>RMSE</td>
<td>15.58</td>
<td>14.43</td>
<td>15.37</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.83</td>
<td>0.79</td>
<td>0.72</td>
</tr>
<tr>
<td>TE01 LOWESS</td>
<td>RMSE</td>
<td>18.63</td>
<td>14.99</td>
<td>20.24</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.75</td>
<td>0.77</td>
<td>0.46</td>
</tr>
<tr>
<td>TE09 L4</td>
<td>RMSE</td>
<td>9.83</td>
<td>8.95</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.81</td>
<td>0.68</td>
<td>0.84</td>
</tr>
<tr>
<td>TE09 LOWESS</td>
<td>RMSE</td>
<td>10.45</td>
<td>10.04</td>
<td>10.33</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.79</td>
<td>0.60</td>
<td>0.77</td>
</tr>
<tr>
<td>TE17 L3</td>
<td>RMSE</td>
<td>16.97</td>
<td>18.55</td>
<td>18.01</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.50</td>
<td>0.44</td>
<td>0.56</td>
</tr>
<tr>
<td>TE17 LOWESS</td>
<td>RMSE</td>
<td>17.00</td>
<td>18.60</td>
<td>19.68</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.50</td>
<td>0.43</td>
<td>0.50</td>
</tr>
<tr>
<td>TE25 L3</td>
<td>RMSE</td>
<td>8.88</td>
<td>10.05</td>
<td>11.72</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.53</td>
<td>0.27</td>
<td>0.35</td>
</tr>
<tr>
<td>TE25 LOWESS</td>
<td>RMSE</td>
<td>10.08</td>
<td>13.20</td>
<td>13.40</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.36</td>
<td>0.38</td>
<td>0.19</td>
</tr>
<tr>
<td>TE33 L4</td>
<td>RMSE</td>
<td>11.22</td>
<td>11.41</td>
<td>15.55</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.91</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>TE33 LOWESS</td>
<td>RMSE</td>
<td>15.20</td>
<td>12.46</td>
<td>16.66</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.84</td>
<td>0.78</td>
<td>0.83</td>
</tr>
<tr>
<td>TE41 L4</td>
<td>RMSE</td>
<td>9.93</td>
<td>11.10</td>
<td>9.08</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.74</td>
<td>0.63</td>
<td>0.68</td>
</tr>
<tr>
<td>TE41 LOWESS</td>
<td>RMSE</td>
<td>11.17</td>
<td>10.27</td>
<td>11.30</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.68</td>
<td>0.69</td>
<td>0.51</td>
</tr>
<tr>
<td>EE03 L3</td>
<td>RMSE</td>
<td>4.25</td>
<td>3.87</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.63</td>
<td>0.38</td>
<td>0.68</td>
</tr>
<tr>
<td>EE03 LOWESS</td>
<td>RMSE</td>
<td>4.97</td>
<td>3.91</td>
<td>4.01</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.49</td>
<td>0.37</td>
<td>0.58</td>
</tr>
<tr>
<td>EE11 L4</td>
<td>RMSE</td>
<td>8.23</td>
<td>6.52</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.65</td>
<td>0.76</td>
<td>0.42</td>
</tr>
<tr>
<td>EE11 LOWESS</td>
<td>RMSE</td>
<td>9.28</td>
<td>7.95</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.56</td>
<td>0.64</td>
<td>0.14</td>
</tr>
<tr>
<td>EE19 L2</td>
<td>RMSE</td>
<td>6.14</td>
<td>6.37</td>
<td>6.40</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.80</td>
<td>0.65</td>
<td>0.74</td>
</tr>
<tr>
<td>EE19 LOWESS</td>
<td>RMSE</td>
<td>7.52</td>
<td>7.65</td>
<td>7.53</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.70</td>
<td>0.49</td>
<td>0.65</td>
</tr>
<tr>
<td>EE27 L4</td>
<td>RMSE</td>
<td>5.67</td>
<td>5.06</td>
<td>4.51</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.93</td>
<td>0.92</td>
<td>0.87</td>
</tr>
<tr>
<td>EE27 LOWESS</td>
<td>RMSE</td>
<td>7.83</td>
<td>5.54</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.87</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>EE35 L4</td>
<td>RMSE</td>
<td>9.34</td>
<td>8.86</td>
<td>7.98</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.65</td>
<td>0.51</td>
<td>0.74</td>
</tr>
<tr>
<td>EE35 LOWESS</td>
<td>RMSE</td>
<td>8.77</td>
<td>9.69</td>
<td>8.12</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.69</td>
<td>0.42</td>
<td>0.73</td>
</tr>
<tr>
<td>EE43 L4</td>
<td>RMSE</td>
<td>8.33</td>
<td>9.49</td>
<td>8.93</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.60</td>
<td>0.37</td>
<td>0.64</td>
</tr>
<tr>
<td>EE43 LOWESS</td>
<td>RMSE</td>
<td>9.99</td>
<td>9.90</td>
<td>9.43</td>
</tr>
<tr>
<td></td>
<td>R Squared</td>
<td>0.43</td>
<td>0.32</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table A5.11 Goodness of Fit Statistics for all End-Coils from W164167 to W176AOH10 for the Best Performing MLR Models and the LOWESS Model.
Table A5.11 displays R-Squared and RMSE values for the best performing (as defined in Section 6.10) MLR Model and the LOWESS models at all 12 End-Coils. It can be seen that the performance of the models across all end coils is broadly similar, with the notable exception of TE25, where the R-squared value is negative at W176180 and to a lesser extent EE11 where the R-squared value is higher for W168171 than for the other three periods, which contradicts the trend set by most of the other end-coils.

The poor performance of TE25 is due largely to behaviour changes related to generator temperature which are dealt with in some detail in Chapter 9.
Monitoring and Regression Based Modelling of End Winding Vibration in Large Synchronous Generators

Publications

Monitoring of Generator End Turn Vibration

Mr. David Shally
Dublin Institute of Technology
daveshally@gmail.com

Mr. Michael Farrell
Dublin Institute of Technology
Michael.Farrell@dit.ie

Mr. Kevin Sullivan
Dublin Institute of Technology
Kevin.Sullivan@dit.ie

Abstract: The principal focus of this paper is the development of regression based models for the prediction of vibration levels in the end windings of a 288MVA Gas Turbine Generator. These models are part of a wider end-winding monitoring system. For the purpose of this study, a regression approach was applied. This approach was selected, as it did not require any additional major investment in hardware or software once the original system had been installed. Equipment already available at the plant was utilised to enhance the capabilities of the monitoring system. These models etc. should prove a valuable and straightforward tool in assessing the condition of the machine end-windings.

I. INTRODUCTION

End winding vibration has long been a concern for the design and maintenance engineer alike. Current flowing in the rotor and stator give rise to magnetic fields. The resulting forces lead to vibration within the stator core, but more seriously at the stator end windings and their support structures. The end winding structures have a high susceptibility to vibration damage on account of the complexity of their structure, the number of materials used in their construction and the difficulty in supporting them properly.

This project was initiated to monitor the end winding behaviour of a 288MVA Synchronous Generator with a known end winding vibration (EWV) problem. The generator was put into service in late 2002.

During inspections carried out at two consecutive scheduled annual outages, damage to the end winding structures was discovered. The damage included cracked end winding support brackets. Large quantities of dusting due to severe fretting were also discovered.

It was determined that the natural frequencies of some coils/coil sections of the generator and the end winding structure as a whole were at, or very close to, the 100Hz magnetic forcing frequency. As a result the coil sections with natural frequencies at or close to this level were being excited, resulting in raised vibration levels and the aforementioned damage to the machine.

On the first occasion the machine was repaired on site and put back into service. However, after the second discovery, which included the cracked support brackets, the manufacturer imposed some operating restrictions in order to minimise the risk of a major failure. The main stipulations were:

1. The machine must operate as a base load unit in so far as possible,
2. The machine must be shut down for a visual inspection and ‘bump-test’ at quarterly intervals in order to assess its condition
3. The number of ‘shut downs’/’run ups’ per quarter was limited.

At this juncture, the operator considered the purchase of an end-winding vibration monitoring system. It was agreed with the manufacturer that if the system was purchased and the data made available them, the requirement for quarterly inspections could subsequently be dropped. The dropping of such quarterly inspections enabled a valuable increase in plant availability. In addition, the number of scheduled generator ‘run ups’ and ‘shut downs’ was significantly reduced. Once the monitoring system was installed, the problem could be measured on an ongoing basis and the machine could be taken out of service should vibration levels become too high.

A major repair was carried out to the end winding support structure in April/May 2007. Following this repair, the static natural frequencies of both structures had been reduced from around 111Hz and 99.7Hz, to around 92Hz at both ends.

The work documented in this paper focuses on the development of a methodology, which models the behaviour of the end winding structure of a relatively healthy machine. The models run on OSIsoft’s PI Advanced Computing Engine (ACE) and through RT-WEBPARTS. These are available to all interested staff through the company intranet.

II. VIBRATION MONITORING SYSTEM

The vibration monitoring system was provided by VibrosysM Inc, who specialise in the field of motor and generator condition monitoring. The system currently monitors vibration levels on both the Gas Turbine (GT) and Steam Turbine (ST) Generators.

Each generator is fitted with 12 fibre optic accelerometers, 6 per end. Vibration is measured in the radial plane only. Data is stored to a low resolution database (average vibration data stored at 5-15 minute intervals). The cost of installing sensors to record vibration in all three measurement planes (Axial, Radial and Tangential) was deemed prohibitively high. The decision to monitor data in the radial plane was made based...
on the results of a ‘bump-test’ carried out two months prior to the installation of the system. The results of the test showed that vibration levels were highest in the radial plane. Vibration levels were lowest on the tangential/circumferential plane. The distribution of the sensors around the end winding structure and where they are located along the windings is depicted in Figure 1.

In order to increase the capabilities of the monitoring system, the sensors were also connected to a high resolution data acquisition system already used at the plant. This system is used mainly to monitor plant data during ‘run-ups’ and shut-downs’. This system is linked with OSIsoft’s PI data acquisition and storage system.

The regression models are calculated in PI Advanced Computing Engine (ACE) which uses Microsoft Visual Basic.NET 2003 as its program development environment. The calculated values are stored to the PI database.

RT-WEBPARTS is used to develop web pages on the company intranet to allow all plant staff access to plant information from any PC connected to the LAN.

III. DEVELOPMENT OF MODELS

The machine returned to service in May 2007. The development of models to predict machine EWV behaviour commenced once there was sufficient data available to generate a reasonably good model.

The outcome of the repair carried out in April of 2007 was the reduction of the static natural frequencies of both end winding structures to 92Hz (from 111 Hz Exciter End, 99.7 Hz Turbine End). Dynamic natural frequency values tend to be 3 – 10 Hz below static values, due mainly to the effects of heating.

The effect on vibration levels was that, at the Turbine End (TE)/Connected End (CE), the behaviour of the overall structure has become much more uniform. Variance has reduced, and peak vibration levels are significantly lower (30 - 40% approx).

At the Exciter End (EE)/Non-Connected End (NCE) however, peak vibration levels post repair are largely unchanged. The behaviour of the individual end bars are much less uniform than at the Turbine End. Certain sensors record large vibration level variances, associated mainly with load, power factor and temperature (e.g. at bars 03, 19 and 43). Sensors at other ‘end bars’ record minimal variation in vibration levels with load etc. (e.g. bar 27).

From Fig. 2 it can also be seen that for sensor positions there is a rise in vibration for a given change of condition while at other locations vibration levels drop.

\[ \text{Graphical representation of vibration data over time} \]

The voltages at the Connected End, and the Non-Connected End are also plotted at both the Turbine end and the Exciter end. The voltages are shown in the figure for 04-February-2008.

From the trend data it was apparent that while finding and applying suitable models to data taken from some sensors
Many different forms of regression were employed in order to model the behaviour of each point of measurement on both ends of the machine. Relevant plant data such as Active Power, Reactive Power, Ambient Air Temperature, Average Stator Slot Temperature etc. were used as input/independent variables. Ten plant variables were used in all.

A number of models calculated using various programs and varying greatly in complexity were applied to the data.

Model types tested included:

- MLR with 4-10 plant input/independent variables
- Weighted Least Squares Regression
- ARX, ARMA, ARMAX etc. using Matlab 7
- Kernel Based Regressions
- Stepwise Regression using 10 plant input/independent variables
- MLR using plant inputs, as well as, interactions and polynomials etc.
- Stepwise Regression using plant inputs/independent variables, as well as, interactions and polynomials etc.

All of the various model types were compared in order to choose the type to be implemented on the system. It was found, as predicted, that vibration behaviour at certain positions was much easier to model accurately that at others. It was found that no matter which regression method was used, that the same measurement points performed well or poorly.

Models calculated using the Stepwise function in Matlab 7 and the kernel based regressions proved to be the most accurate overall. The quality of the fit was determined based mainly on $R^2$ and RMSE figures. All model types were tested with new data for a number of months before the final models were chosen for each end winding end bar.

During various tests it was discovered that although the model predictions for some sensor positions were not as accurate as for others, they were still useful, and in conjunction with other data could be used reasonably reliably to assess machine condition.

In arriving at an appropriate model for end use, a number of criteria had to be taken into account namely:

- Model complexity
  (Number of independent variables included, model type etc.)

- Ease of application of model on plant system
  (Number of independent variables included, model type etc.)

- Accuracy of the model ($R^2$, RMSE etc.)

The models chosen for application on PI ACE were calculated using Multiple Regression, with variables for inclusion in the model selected using the Stepwise function of MATLAB 7. The resulting models are multiple regression models calculated using relevant plant data as well as various calculated variables created from the original variables in order to optimise the models i.e. polynomials, interactions etc.

The Stepwise function in MATLAB was used to calculate the best possible model and to eliminate statistically insignificant independent variables. The Stepwise function of MATLAB uses Backward Elimination and Forward Selection procedures in order to select independent variables for inclusion and exclusion. (The significance level for inclusion and exclusion are based on p-values for that variable)

Using the GUI one can choose between competing models in order to select the most appropriate model.

The result of the regression is a constant/intercept plus a coefficient of regression related to each independent variable (in Matlab the coefficient is zero if the variable is omitted). The regression equation is therefore of the standard MLR form

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n$$

(1)

Where: $y$ is the dependent variable
- $\alpha$ is the constant/y-axis intercept
- $\beta_n$ is the coefficient of regression for independent variable $n$
- $x_n$ is the independent variable $n$

These types of model can be programmed into PI ACE easily, are in relative terms not computationally intensive (when compared to kernel based regression methods etc.), and are among the most accurate models to have been run as part of this project.

IV. APPLICATION OF MODELS TO NEW DATA

For a number of months after the application of the selected models no problems were encountered with them. However, the plant moved from a predominantly loaded regime to a cycling or ‘two-shifting’ regime for a number of weeks. Changing to this mode of operation led to conditions being outside of normal, as encountered to that point, post repair i.e. non-cycling duty. No data pertaining to this mode of operation (i.e. cycling/‘two-shifting’) was available when the models were created. As a result the models proved unreliable in predicting end bar vibration during cycling duty.

As a consequence of this, the regression models were recalculated using data representative of all possible operating conditions, including two-shifting mode.
In generating these models there was some loss in accuracy for the more normal operating conditions (i.e. non-cycling), however improvement in terms of greater model robustness outweighed any minor losses in overall accuracy.

The resulting models can be used reliably for all operating conditions between the Minimum Stable Generation (MSG) condition and Maximum Generation. For conditions outside these normal operating boundaries i.e., ‘start-ups’ and ‘shut downs’ etc, where output power is below MIN GEN, the accuracy of the models cannot be guaranteed.

Fig. 3. Turbine End (TE)/Connected End (CE) Bar 33: Measured data (Blue) -v- Model Predicted Data (Red). Learning Data

To date, the results have been encouraging with the quality of the fit of the predicted vibration data sets to the actual measured data sets remaining largely constant since the application of the newest models. The figures below display data seven months apart for the same vibration measurement location (Exciter End/Connected End Bar No.33)

Fig. 4. Turbine End (TE)/Connected End (CE) Bar 33: Measured data (Blue) and Model Predicted Data (Red) -v- Time. 27-August-2007 to 23-September-2007

It can be seen that the measured behaviour of the coil is well predicted by the model in both cases. The quality of the prediction can also be seen to be at a comparable level for both periods. The Goodness of Fit Statistics used to assess the quality of the predictions are, $R^2$, $R^2_{PRED}$ and RMSE. $R^2$ is calculated as:

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

Where: $\text{SSE}$ is the Sum of Squares due to Error

$\text{SST}$ is the Total Sum of Squares (or sum of squares about the mean)

$R^2$ is used only for the learning data and values can range from 0 to 1.

$R^2_{PRED}$ is used for all data outside of the learning data

$$R^2_{PRED} = 1 - \frac{\text{PRESS}}{\text{SST}}$$

Where: $\text{PRESS}$ is the Predicted Error Sum of Squares

$\text{SST}$ is the Total Sum of Squares

The $R^2$, $R^2_{PRED}$, and RMSE (Root Mean Squared Error) values (for data sets 4 weeks in length) from May/June ‘07 (learning data) to May ’08 are shown below in figures 6-9. The $R^2$ value applies only to the learning data (the leftmost point on the trends). For all data sets outside of the learning data the $R^2_{PRED}$ ‘quality of fit’ statistic is applied. The RMSE statistic is used universally.

It can be seen that for models on the TE/CE, values have remained relatively steady throughout. On the EE/NCE the values can vary quite widely, particularly during periods of ‘two-shifting’/cycling operation. On the EE particularly, the RMSE figure is a more reliable measure of accuracy of prediction than $R^2$. 
Since the development and application of the models there has been no noticeable deterioration in the condition of the machine end-windings (verified by an inspection and ‘bump test’ carried out at a scheduled outage in April 2008). The work of Fortin and Duffeau [4] also reinforces confidence in the viability of the method. The accuracy of the predicted vibration values has remained acceptable over a period of 12 months. When used along with other statistical and graphical methods, as well as periodic inspections and ‘bump tests’, these models are a useful tool for operations staff in assessing end winding condition.

REFERENCES