Safety Regulation Group

Civil Aviation Authority

CAA PAPER 2011/01

Intelligent Management of Helicopter Vibration Health Monitoring Data

Based on a report prepared for the CAA by GE Aviation Systems Limited

www.caa.co.uk

Safety Regulation Group

Civil Aviation Authority

CAA PAPER 2011/01

Intelligent Management of Helicopter Vibration Health Monitoring Data

Based on a report prepared for the CAA by GE Aviation Systems Limited

May 2012

ISBN 978 0 11792 403 1

Published May 2012

Enquiries regarding the content of this publication should be addressed to:
Flight Operations 2, Safety Regulation Group, Civil Aviation Authority, Aviation House, Gatwick Airport South, West Sussex, RH6 0YR.

The latest version of this document is available in electronic format at www.caa.co.uk/publications, where you may also register for e-mail notification of amendments.

Published by TSO (The Stationery Office) on behalf of the UK Civil Aviation Authority.

Printed copy available from:
TSO, PO Box 29, Norwich NR3 1GN
Telephone orders/General enquiries: 0844 477 7300
Fax orders: 0870 600 5533
www.tsoshop.co.uk
E-mail: caa@tso.co.uk
Textphone: 0870 240 3701

List of Effective Pages

Part	Page	Date	Part	Page	Date
	iii	May 2012	Report	35	May 2012
	iv	May 2012	Report	36	May 2012
	v	May 2012	Report	37	May 2012
	vi	May 2012	Report	38	May 2012
	vii	May 2012	Report	39	May 2012
Contents	1	May 2012	Report	40	May 2012
Contents	2	May 2012	Report	41	May 2012
List of Figures	1	May 2012	Report	42	May 2012
List of Tables	1	May 2012	Report	43	May 2012
Foreword	1	May 2012	Report	44	May 2012
Glossary	1	May 2012	Report	45	May 2012
Executive Summary	1	May 2012	Report	46	May 2012
Report	1	May 2012	Report	47	May 2012
Report	2	May 2012	Annex A Contents	1	May 2012
Report	3	May 2012	Annex A Contents	2	May 2012
Report	4	May 2012	Annex A List of Figures	1	May 2012
Report	5	May 2012	Annex A List of Tables	1	May 2012
Report	6	May 2012	Annex A Glossary	1	May 2012
Report	7	May 2012	Annex A Executive Summary	1	May 2012
Report	8	May 2012	Annex A Report	1	May 2012
Report	9	May 2012	Annex A Report	2	May 2012
Report	10	May 2012	Annex A Report	3	May 2012
Report	11	May 2012	Annex A Report	4	May 2012
Report	12	May 2012	Annex A Report	5	May 2012
Report	13	May 2012	Annex A Report	6	May 2012
Report	14	May 2012	Annex A Report	7	May 2012
Report	15	May 2012	Annex A Report	8	May 2012
Report	16	May 2012	Annex A Report	9	May 2012
Report	17	May 2012	Annex A Report	10	May 2012
Report	18	May 2012	Annex A Report	11	May 2012
Report	19	May 2012	Annex A Report	12	May 2012
Report	20	May 2012	Annex A Report	13	May 2012
Report	21	May 2012	Annex A Report	14	May 2012
Report	22	May 2012	Annex A Report	15	May 2012
Report	23	May 2012	Annex A Report	16	May 2012
Report	24	May 2012	Annex A Report	17	May 2012
Report	25	May 2012	Annex A Report	18	May 2012
Report	26	May 2012	Annex A Report	19	May 2012
Report	27	May 2012	Annex A Report	20	May 2012
Report	28	May 2012	Annex A Report	21	May 2012
Report	29	May 2012	Annex A Report	22	May 2012
Report	30	May 2012	Annex A Report	23	May 2012
Report	31	May 2012	Annex A Report	24	May 2012
Report	32	May 2012	Annex A Report	25	May 2012
Report	33	May 2012	Annex A Report	26	May 2012
Report	34	May 2012	Annex A Report	27	May 2012

Part	Page	Date
Annex A Report	28	May 2012
Annex A Report	29	May 2012
Annex A Report	30	May 2012
Annex A Report	31	May 2012
Annex A Report	32	May 2012
Annex A Report	33	May 2012
Annex A Report	34	May 2012
Annex A Report	35	May 2012
Annex A Report	36	May 2012
Annex A Report	37	May 2012
Annex A Report	38	May 2012
Annex A Report	39	May 2012
Annex A Report	40	May 2012
Annex A Report	41	May 2012
Annex A Report	42	May 2012
Annex A Report	43	May 2012
Annex A Report	44	May 2012
Annex A Report	45	May 2012
Annex A Report	46	May 2012
Annex A Report	47	May 2012
Annex A Report	48	May 2012
Annex A Report	49	May 2012
Annex A Report	50	May 2012
Annex A Report	51	May 2012
Annex A Report	52	May 2012
Annex A Report	53	May 2012
Annex A Report	54	May 2012
Annex A Report	55	May 2012
Annex A Report	56	May 2012
Annex A Report	57	May 2012
Annex A Report	58	May 2012
Annex A Report	59	May 2012
Annex A Report	60	May 2012
Annex A Appendix A	1	May 2012
Annex A Appendix A	2	May 2012
Annex A Appendix A	3	May 2012
Annex A Appendix A	4	May 2012
Annex A Appendix B	1	May 2012
Annex A Appendix B	2	May 2012
Annex A End Notes	1	May 2012
Annex A End Notes	2	May 2012
Annex A End Notes	3	May 2012
Annex A End Notes	4	May 2012
Annex A End Notes	5	May 2012
Annex B Contents	1	May 2012
Annex B List of Figures	1	May 2012
Annex B List of Figures	2	May 2012
Annex B List of Figures	3	May 2012
Annex B List of Figures	4	May 2012
Annex B List of Tables	1	May 2012
Annex B Glossary	1	May 2012
Annex B Executive Summary	1	May 2012
Annex B Report	1	May 2012
Annex B Report	2	May 2012
Annex B Report	3	May 2012
Annex B Report	4	May 2012
Annex B Report	5	May 2012
Annex B Report	6	May 2012
Annex B Report	7	May 2012
Annex B Report	8	May 2012
Annex B Report	9	May 2012
Annex B Report	10	May 2012
Annex B Report	11	May 2012
Annex B Report	12	May 2012
Annex B Report	13	May 2012
Annex B Report	14	May 2012
Annex B Report	15	May 2012
Annex B Report	16	May 2012
Annex B Report	17	May 2012
Annex B Report	18	May 2012
Annex B Report	19	May 2012
Annex B Report	20	May 2012
Annex B Report	21	May 2012
Annex B Report	22	May 2012
Annex B Report	23	May 2012
Annex B Report	24	May 2012
Annex B Report	25	May 2012
Annex B Report	26	May 2012
Annex B Report	27	May 2012
Annex B Report	28	May 2012
Annex B Report	29	May 2012
Annex B Report	30	May 2012
Annex B Report	31	May 2012
Annex B Report	32	May 2012
Annex B Report	33	May 2012
Annex B Report	34	May 2012
Annex B Report	35	May 2012
Annex B Report	36	May 2012
Annex B Report	37	May 2012
Annex B Report	38	May 2012
Annex B Report	39	May 2012
Annex B Report	40	May 2012
Annex B Report	41	May 2012
Annex B Report	42	May 2012
Annex B Report	43	May 2012
Annex B Report	44	May 2012

Part	Page	Date
Annex D Report	5	May 2012
Annex D Report	6	May 2012
Annex D Report	7	May 2012
Annex D Report	8	May 2012
Annex D Report	9	May 2012
Annex D Report	10	May 2012
Annex D Report	11	May 2012
Annex D Report	12	May 2012
Annex D Report	13	May 2012
Annex D Report	14	May 2012
Annex D Report	15	May 2012
Annex D Report	16	May 2012
Annex D Report	17	May 2012
Annex D Report	18	May 2012
Annex D Report	19	May 2012
Annex D Report	20	May 2012
Annex D Report	21	May 2012
Annex D Report	22	May 2012
Annex D Report	23	May 2012
Annex D Report	24	May 2012
Annex D Report	25	May 2012
Annex D Report	26	May 2012
Annex D Report	27	May 2012
Annex D Report	28	May 2012
Annex D Report	29	May 2012
Annex D Report	30	May 2012
Annex D Report	31	May 2012
Annex D Report	32	May 2012
Annex D Report	33	May 2012
Annex D Report	34	May 2012
Annex D Report	35	May 2012
Annex D Report	36	May 2012
Annex D Report	37	May 2012
Annex D Report	38	May 2012
Annex D Report	39	May 2012
Annex D Report	40	May 2012
Annex D Report	41	May 2012
Annex D Report	42	May 2012
Annex D Report	43	May 2012
Annex D Report	44	May 2012
Annex D Report	45	May 2012
Annex D Report	46	May 2012
Annex D Report	47	May 2012
Annex D Report	48	May 2012
Annex D Report	49	May 2012
Annex D Report	50	May 2012
Annex D Report	51	May 2012
Annex D Report	52	May 2012
Annex D Report	53	May 2012
Annex D Report	54	May 2012
Annex D Report	55	May 2012
Annex D Report	56	May 2012
Annex D Report	57	May 2012
Annex D Report	58	May 2012
Annex D Report	59	May 2012
Annex D Report	60	May 2012
Annex D Report	61	May 2012
Annex D Report	62	May 2012
Annex D Report	63	May 2012
Annex D Report	64	May 2012
Annex D Report	65	May 2012
Annex D Report	66	May 2012
Annex D Report	67	May 2012
Annex D Report	68	May 2012
Annex D Report	69	May 2012
Annex D Report	70	May 2012
Annex E Contents	1	May 2012
Annex E List of Figures	1	May 2012
Annex E List of Figures	2	May 2012
Annex E List of Tables	1	May 2012
Annex E Glossary	1	May 2012
Annex E Executive Summary	1	May 2012
Annex E Report	1	May 2012
Annex E Report	2	May 2012
Annex E Report	3	May 2012
Annex E Report	4	May 2012
Annex E Report	5	May 2012
Annex E Report	6	May 2012
Annex E Report	7	May 2012
Annex E Report	8	May 2012
Annex E Report	9	May 2012
Annex E Report	10	May 2012
Annex E Report	11	May 2012
Annex E Report	12	May 2012
Annex E Report	13	May 2012
Annex E Report	14	May 2012
Annex E Report	15	May 2012
Annex E Report	16	May 2012
Annex E Report	17	May 2012
Annex E Report	18	May 2012
Annex E Report	19	May 2012
Annex E Report	20	May 2012
Annex E Report	21	May 2012
Annex E Report	22	May 2012
Annex E Report	23	May 2012
Annex E Report	24	May 2012

Part	Page	Date
Annex E Report	25	May 2012
Annex E Report	26	May 2012
Annex E Report	27	May 2012
Annex E Report	28	May 2012
Annex E Report	29	May 2012
Annex E Report	30	May 2012
Annex E Report	31	May 2012
Annex E Report	32	May 2012
Annex E Report	33	May 2012
Annex E Report	34	May 2012
Annex E Report	35	May 2012
Annex E Report	36	May 2012
Annex E Report	37	May 2012
Annex E Report	38	May 2012
Annex E Report	39	May 2012
Annex E Report	40	May 2012
Annex E Report	41	May 2012
Annex E Report	42	May 2012
Annex E Report	43	May 2012
Annex E Report	44	May 2012
Annex E Report	45	May 2012
Annex E Report	46	May 2012
Annex E Report	47	May 2012
Annex E Report	48	May 2012
Annex E Report	49	May 2012
Annex E Report	50	May 2012
Annex E Report	51	May 2012
Annex E Report	52	May 2012
Annex E Report	53	May 2012
Annex E Report	54	May 2012
Annex E Report	55	May 2012
Annex E Report	56	May 2012
Annex E Report	57	May 2012
Annex E Report	58	May 2012
Annex E Report	59	May 2012
Annex E Report	60	May 2012
Annex E Report	61	May 2012
Annex E Report	62	May 2012
Annex E Report	63	May 2012
Annex E Report	64	May 2012
Annex E Report	65	May 2012
Annex E Report	66	May 2012
Annex E Report	67	May 2012
Annex E Report	68	May 2012
Annex E Report	69	May 2012
Annex E Report	70	May 2012
Annex E Report	71	May 2012
Annex E Report	72	May 2012
Annex E Report	73	May 2012
Annex E Report	74	May 2012
Annex E Report	75	May 2012

INTENTIONALLY LEFT BLANK

Table of Contents

Report

List of Figures

INTENTIONALLY LEFT BLANK

List of Tables

INTENTIONALLY LEFT BLANK

Foreword

The research reported in this paper was funded by the Safety Regulation Group of the UK Civil Aviation Authority (CAA), the US Federal Aviation Administration, Oil & Gas UK, Norwegian CAA and Shell Aircraft Limited, and was performed by GE Aviation (formerly Smiths Aerospace). The work forms part of the UK CAA's helicopter health and usage monitoring system (HUMS) research programme which was instigated in response to the recommendations of the Review of Helicopter Airworthiness (HARP Report - CAP 491), and the subsequent recommendations of the Report of the Working Group on Helicopter Health Monitoring (CAA Paper 85012).

The first HUMS were installed on the North Sea helicopter fleet in the early 1990s which, the CAA believes, contributed significantly to a reduction in the airworthiness accident rate. However, in-service experience and the results of the two helicopter main rotor gearbox seeded defect test programmes, performed under contract to the CAA by AgustaWestland (formerly Westland Helicopters Ltd.) and Eurocopter, indicated that there was scope for improving the effectiveness of HUMS vibration health monitoring (VHM) data analysis.

To address this issue, the application of advanced analysis techniques to HUMS data was investigated by MJA Dynamics Ltd (now part of GE Aviation) under contract to the CAA. The results of that work, published in CAA Paper 99006, clearly demonstrated the potential for significantly improving HUMS VHM data analysis through the use of unsupervised machine learning techniques. However, it was established that further development of the technology would be required to address the practical issues associated with its routine every day use.

Following a competitive tendering exercise, the contract for a programme of work to fulfil this need was awarded to GE Aviation, supported by Bristow Helicopters, and the final project report is presented in this paper. It is the CAA's view that the Advanced Anomaly Detection (AAD) system developed by this project has been clearly demonstrated to be both effective and practical. This view is supported by Oil & Gas UK who have determined to implement the technology for the North Sea helicopter fleets having sufficient utility remaining.

Since the input to the AAD system comprises the output of the existing HUMS and does not affect it in any way, the system could be implemented as an additional stand-alone analysis tool. However, in recognising that the maximum benefit will be obtained from a fully integrated solution, AAD is being introduced via the HUMS supplier for each aircraft which, in most cases, is the helicopter manufacturer. At the time of publication, AgustaWestland have licensed the GE Aviation AAD system developed under this project for the AW139 helicopter fleet, and implementation is progressing. In addition, Eurocopter are developing their own version of AAD, initially for their EC225 and EC175 helicopter fleets, which is scheduled for trials during 2012. Sikorsky are also understood to have an equivalent system although no information is available on plans for trials or demonstrations.

Based on the results of this project the CAA believes that, with the implementation of the GE Aviation or equivalent AAD system, the effectiveness of transmission HUMS will be elevated to a satisfactory level. The next priority is considered to be the extension of HUMS to helicopter rotors. Following the review of the status of helicopter rotor health monitoring reported in CAA Paper 2008/05, GE Aviation were contracted to investigate the application of AAD to helicopter tail rotor VHM data. This work has been completed and will be reported in a separate CAA paper during 2012.

Safety Regulation Group

December 2011

INTENTIONALLY LEFT BLANK

Glossary

5IA/5IT	5-Indicator Absolute / 5-Indicator Trend anomaly model
8IA/8IT	8-Indicator Absolute / 8-Indicator Trend anomaly model
AAD	Advanced Anomaly Detection
AGB	Accessory Gearbox
BIC	Bayesian Information Criterion
CAA	Civil Aviation Authority (UK)
CBR	Case-Based Reasoning
CI	Condition Indicator
DAPU	Data Acquisition and Processing Unit
FS	Fitness Score
GEV	Generalised Extreme Value
GMM	Gaussian Mixture Model/Modelling
HUMS	Health and Usage Monitoring System
IF	Influence Factor
IFS	Industrial Financial Systems
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
LHA	Left Hand Accessory module
M6A/M6T	M6 Absolute / M6 Trend anomaly model
MGB	Main rotor Gearbox
MoD	Ministry of Defence
OEM	Original Equipment Manufacturer
PA	Probability of Anomaly
RHA	Right Hand Accessory module
SOA/SOT	Shaft Order Absolute / Shaft Order Trend anomaly model
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring

INTENTIONALLY LEFT BLANK

Executive Summary

This report documents the results of a five-year research programme commissioned by the Civil Aviation Authority (CAA) to demonstrate the intelligent analysis of helicopter Health and Usage Monitoring System (HUMS) Vibration Health Monitoring (VHM) data. The goal of the programme was to improve the fault detection performance of HUMS.

A HUMS Advanced Anomaly Detection (AAD) capability, and a system to implement it, were developed in Phase 1 of the research programme, which also included an off-line demonstration using a database of historical Bristow Helicopters AS332L HUMS data. In Phase 2 of the programme, the AAD system was subjected to a six-month in-service trial by Bristow Helicopters on their AS332L fleet. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L fleet HUMS data being automatically transferred on a daily basis from Aberdeen.

The CAA contract included an option for a six-month extension to the in-service trial. Building on the success of the initial trial, additional research tasks were carried out to research and develop further advanced HUMS data analysis capabilities prior to the second trial period. These included a data re-modelling task, and the introduction of new Probability of Anomaly (PA) and diagnostic Influence Factor (IF) outputs, with a global anomaly alerting threshold applied based on the PA value. The second in-service trial period evaluated the enhanced AAD system incorporating the updated anomaly models and the new PA and IF outputs. To minimise Bristow's workload associated with the trial, this time GE Aviation also provided an analysis service, sending Bristow daily fleet health status reports.

The new HUMS AAD capability that has been developed is based on density estimation, and incorporates a number of novel features to ensure that the technique is robust, and addresses all the practical issues related to working with in-service HUMS VHM data. The two in-service trial periods provided a large amount of evidence to demonstrate the performance of the AAD system. This clearly highlighted anomalous HUMS VHM Condition Indicator (CI) trends associated with both aircraft defects and HUMS instrumentation faults which are more difficult to detect using traditional HUMS analysis. The two trial periods confirmed that AAD represents a significant advance in HUMS data analysis, resulting in improved fault detection performance, a reduced 'false alert' rate and increased system effectiveness.

The greatest scope for further improvement in the combined performance of the AAD system and its use by maintenance engineers is through the development of a second level of advanced data analysis, based on data mining and automated reasoning technologies. The final element of the research work undertaken comprised a demonstration of new system concepts and capabilities for this second level of advanced data analysis, and the data mining and reasoning technologies that could provide them. The demonstrations showed how secondary analysis of the current AAD system outputs can provide further benefits to operators by reducing workload, automatically highlighting significant anomaly alerts, and providing new information to support maintenance decision making.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Health and Usage Monitoring Systems (HUMS), incorporating comprehensive rotor drive system Vibration Health Monitoring (VHM), have contributed significantly to improving the safety of rotorcraft operations. However, experience has also shown that, while HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected using current threshold setting methods.

This was illustrated by an incident in 2002 involving a Super Puma main rotor gearbox that was removed after a gearbox chip warning. On subsequent inspection, a large crack was found in the bevel pinion. The chip warning occurred purely because a secondary crack fortuitously released a fragment of material. Normally, this type of crack would not generate any debris. Although there were some trends in the vibration Condition Indicators (CIs) for the component, the HUMS did not generate any alerts. This example highlighted a requirement for a better method of detecting anomalous trends in HUMS data, which had also been identified during the Civil Aviation Authority's (CAA's) main rotor gearbox seeded defect test programmes. As a result of the experience gained during this work, the CAA identified the following four areas to be in need of improvement:

- improvement of warning time;
- detection of build defects;
- accommodation of unexpected gear indicator reactions;
- accommodation of reducing gear indicator trends.

Earlier research conducted as part of the CAA's main rotor gearbox seeded defect test programme demonstrated the potential for improving fault detection performance by applying unsupervised machine learning techniques, such as clustering, to seeded fault test data. The CAA therefore commissioned a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841). GE Aviation was contracted to carry out this programme in partnership with Bristow Helicopters, analysing Integrated HUMS (IHUMS) data from Bristow's European AS332L fleet.

GE Aviation conducted research work under Contract No. 841 and four subsequent amendments to this over approximately a five-year period between April 2004 and July 2009. The research work comprised five separate elements, each of which resulted in a task report being produced for the CAA. This Paper presents a summary of the research work performed, with the contents of the five individual task reports being included in Annexes.

The original work under Contract No. 841 was structured as two phases. Phase I started with a literature survey to identify other research work on unsupervised learning, and went on to research and develop an unsupervised analysis technique, referred to as Advanced Anomaly Detection (AAD), to enhance the fault detection performance of HUMS. Anomaly models were built for all the AS332L rotor drive system components. These fused multiple-input HUMS CIs into a single output Fitness Score (FS), indicating how well new data fitted the model's view of normality. Phase 1 also included an 'off-line' demonstration of the advanced anomaly detection

capability on a database of historical AS332L data, and the development of a system implementing this capability for evaluation in an in-service trial. The results of Phase 1 of the original contract are presented in Section 2 and Annex A.

Phase 2 of the contract comprised a six-month in-service trial, conducted by Bristow Helicopters, of the HUMS AAD system developed in Phase I. The trial ran from May to November in 2006. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L IHUMS data being automatically transferred every night from Aberdeen, and Bristow Helicopters having a remote login to the server to check the anomaly detection system outputs. The results of Phase 2 are presented in Section 3 and Annex B.

Building on the success of the initial trial, under amendments to the original contract, four additional tasks were carried out to research and develop further advanced HUMS data analysis capabilities. Bristow continued to use the initial AAD System while these additional research tasks were completed.

A model tuning and re-modelling task benefited from other work ongoing at the time at GE Aviation, re-implementing and improving the modelling technique to provide an enhanced anomaly modelling facility. All the anomaly models were re-built using this improved facility. A probabilistic alerting policy was implemented, with the FS outputs from the anomaly models being converted into normalised Probability of Anomaly (PA) values with a scale of 0-1. On this scale, a value of 0 represents normal data while a value of 1 indicates a clear anomaly. This simplified threshold setting by allowing a single global alerting threshold to be set at a selected PA value and applied to all monitored components. Influence Factors (IFs) were implemented to provide diagnostic information on the causes of an anomaly alert. The IFs are derived from an anomaly model, with one IF for each HUMS CI used to train the model, and provide an indication of how much each CI contributes to an anomaly alert. Again, they are normalised and so can be directly compared with each other. Finally, an investigation was carried out into the use of additional pre- and post-processing techniques to characterise data trends. Although this investigation did not result in the implementation of any new functionality in the AAD system, it did provide a good foundation for the future development of new capabilities for data characterisation. The results of these four additional research tasks are presented in Section 4 and Annex C.

The CAA contract included an option for a six-month extension to the in-service trial of the HUMS AAD system. The enhanced AAD system, incorporating the updated anomaly models and the new PA and IF outputs, was evaluated in this second in-service trial period. To minimise Bristow's workload associated with the trial, this time GE Aviation also provided an analysis service, identifying any detected anomalies which Bristow should follow-up in daily fleet health status reports. Bristow then used the AAD system to investigate the identified anomalies. The second trial commenced at the beginning of January 2008, with the formal trial period completing at the end of June; however the trial was continued informally until 19 December 2008 while two further research tasks were completed. The results of the second trial period are presented in Section 5 and Annex D.

Two final additional research tasks were conducted, demonstrating new system concepts and capabilities for a second level of advanced data analysis, applying data mining and automated reasoning technologies to the outputs from the HUMS AAD system. This is considered to provide the greatest scope for further enhancement in the combined performance of the system and system user. The technology demonstrations clearly showed how secondary analysis of the current AAD system outputs can provide further benefits to users by reducing workload, automatically

highlighting significant anomaly alerts and providing new information to support decision making. The results of these two research tasks are presented in Section 6 and Annex E.

Finally, some overall conclusions and recommendations from the five-year research project are given in Section 7.

2 Phase 1 of Contract No 841 – Development and Off-line Demonstration of Initial HUMS AAD Capability

The task report containing the detailed results of Phase 1 was issued in May 2007 (before Smiths Aerospace became part of GE Aviation), and its contents are reproduced in Annex A. This material also contains some more information on the CAA's requirements for the research.

An anomaly can be defined as something that deviates from normal behaviour. The need for anomaly detection arises in situations where supervised learning of fault related patterns is impractical. Supervised learning requires that every signal acquisition be tagged with a known classification such as healthy, cracked gear, bearing spall, etc. Fortunately, component related faults are rare but this means that there is no large library of tagged fault data. Even if such a library existed, the required size would be indeterminate because symptoms for a single fault type can vary between cases. But normal (healthy) data are commonplace and usually easy to acquire. Recognition of a fault can then be achieved by building a model of normal behaviour against which new data can be compared.

The objective of anomaly detection is to identify abnormal behaviour that might be indicative of a fault. The process is conceptually simple; a model of normal behaviour is built using a training data set, then new data are assessed for their fit against this model. If the fit is not within the model's threshold then it is flagged as anomalous. Nearly all approaches assume that a set of normal data is available to construct a model of normal behaviour. Anomaly detection can be difficult, but HUMS VHM data present significant additional challenges. Individual gearboxes tend to occupy their own space of normality. In addition, due to a lack of feedback from the repair and overhaul process, undetected instrumentation problems and maintenance interventions, etc. it must be assumed that any database of historical HUMS data will contain unknown anomalies.

2.1 Helicopter VHM Data Analysed

The research work was based on IHUMS data from Bristow's European fleet of AS332L helicopters (Figure 2-1). At the start of the programme the fleet comprised 15 AS332Ls – nine located at Aberdeen, five at Scatsta and one at Den Helder. The anomaly modelling included all the following assemblies in the rotor drive system: Main rotor Gearbox (MGB), left and right Accessory Gearboxes (AGBs) and oil cooler fan, Intermediate Gearbox (IGB), and the Tail rotor Gearbox (TGB). The data had already been processed by the IHUMS to extract the CIs for each component. A historical database of these data were collected for development purposes.

Figure 2-1 Bristow AS332L helicopter

The rotor drive system shaft/gear components analysed for this programme are listed in Table 2-1. The table also shows the aircraft sensor used to acquire the component data, and the equivalent IHUMS 'channel' (i.e. analysis number) allocation. The IHUMS performs 32 component analyses. However, three of the components have two gears on a single shaft (the MGB combiner gear and bevel pinion, and the left and right AGB hydraulic drive 47- and 81-tooth gears). For the purposes of anomaly modelling each of these was split into separate components, with separate anomaly models built for each of the two gears on the shaft using the different gear mesh-specific CIs, giving a total of 35 analyses.

Table 2-1 AS332L drive train components analysed

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MGB
2	1	RH high speed input shaft	MGB
1	2	Left torque shaft – fwd end	MGB
2	3	Right torque shaft – fwd end	MGB
3	4	Left torque shaft – aft end	MGB
4	5	Right torque shaft – aft end	MGB
3	6	Combiner gear	MGB
3	6	Bevel pinion	MGB
4	7	Bevel wheel and oil pump drive	MGB
7	8	1st stage sun gear	MGB
7	9	1st stage planet gear	MGB
5	10	1st epicyclic annulus fwd (RH)	MGB

Table 2-1 AS332L drive train components analysed (Continued)

6	11	1st epicyclic annulus left	MGB
7	12	1st epicyclic annulus aft (RH)	MGB
7	13	2nd stage sun gear	MGB
7	14	2nd stage planet gear	MGB
5	15	2nd epicyclic annulus fwd (RH)	MGB
6	16	2nd epicyclic annulus left	MGB
7	17	2nd epicyclic annulus aft (RH)	MGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB
3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB
4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MGB

The CIs calculated by the IHUMS during each component analysis are listed in Table 2-2. Not all of the indicators were used in the anomaly modelling. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data (the actual indicators used in the different anomaly models are listed in Section 2.2.2.3).

Table 2-2 IHUMS Condition Indicators

Condition Indicator	Description
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation

Table 2-2 IHUMS Condition Indicators (Continued)

FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev. (rms)
ESA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA_CVG	Signal average convergence

2.1.1 Data Exploration

A data exploration exercise was carried out to understand the characteristics of the AS332L IHUMS data being analysed. This highlighted the challenges of threshold setting to reliably detect potential defect-related trends in VHM data.

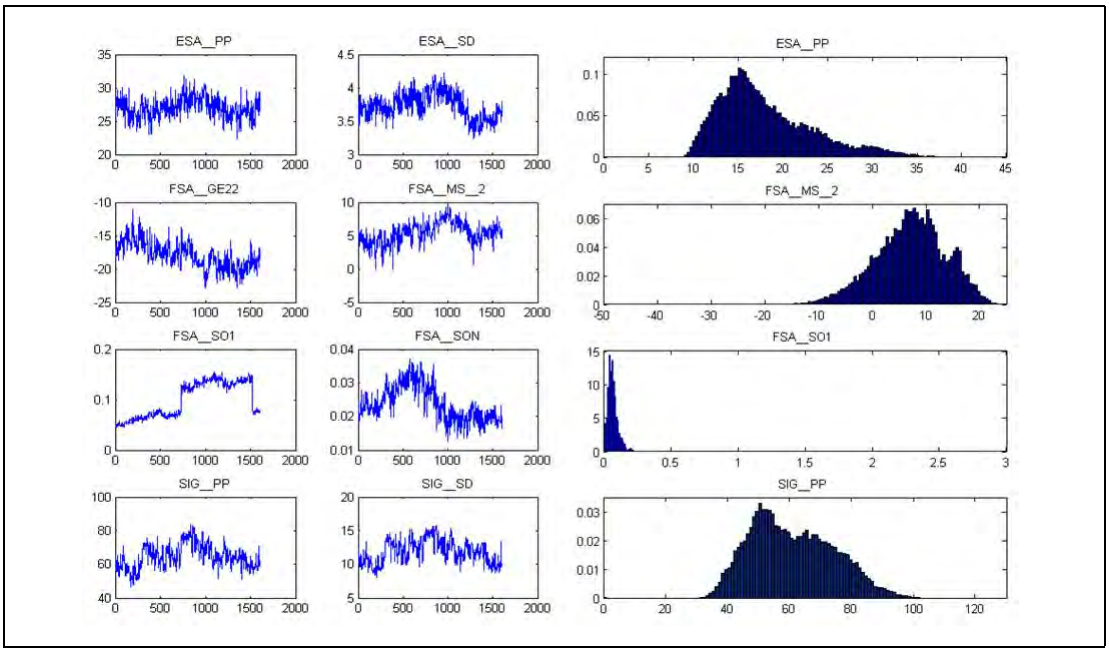

Figure 2-2 Defect-related trends in VHM data: left – typical behaviour of HUMS CIs from a single component; right – fleet histogram for four CIs from one component.

Figure 2-2 (left) shows the typical variability to be found in a set of CIs over the life of a MGB while installed in a specific aircraft. Maintenance often induces a step change in some indicators and this can be seen in FSA_SO1. The signals are typical of a healthy component with no instrumentation defects. Figure 2-2 (right) shows a fleet histogram plot for four indicators taken from the bevel pinion. The indicators have a crude Gaussian shape but they cannot be considered good Gaussian fits.

Figure 2-3 shows a 'box plot' for four CIs from the bevel pinion and a selection of MGBs. The mean value and the range are shown for each gearbox. These plots highlight the variability between different examples of the same component type. The fleet plot for FSA_SO1 looks very different to the other CI fleet plots in that the plot is dominated by two gearboxes. The data for these gearboxes are believed to be polluted by IHUMS instrumentation faults.

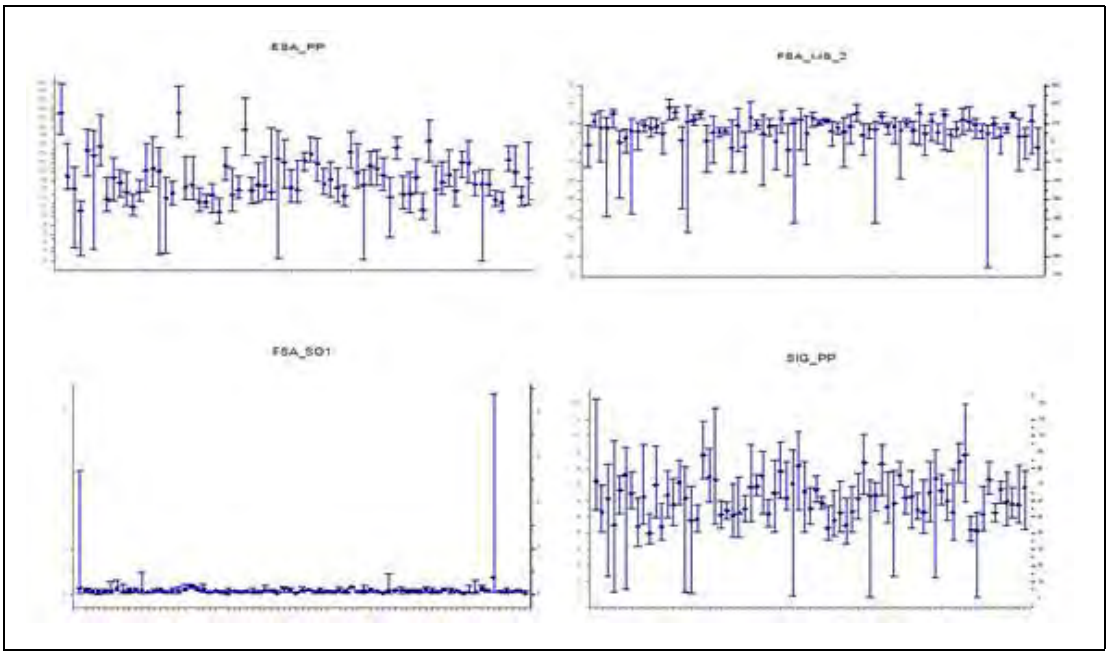

Figure 2-3 Box plot of four CIs, showing variability between aircraft

The data exploration exercise highlighted the highly variable nature of HUMS VHM data. CI histograms appear Gaussian-like, but the variability in the data caused by multiple influences produces a mixing effect that generates complicated distributions. It is inappropriate therefore to fit a single Gaussian distribution to the data. This highlights the weakness of any threshold setting strategy derived from single mode descriptive statistical parameters such as mean and standard deviation.

There is also considerable correlation between some indicator pairs. Redundancy is not implied by this correlation because a de-correlated trend can provide diagnostic information. However, it does mean that care needs to be exercised when performing multivariate analysis because correlated CIs will tend to emphasise an abnormal pattern; this emphasis may be weighted too heavily when CIs measure a similar data characteristic.

2.2 Development of Initial Advanced Anomaly Detection Process

This section describes the initial AAD process that was developed in Phase 1 of the main research contract. This process was refined and enhanced at a later stage of the programme, taking advantage of a GE Aviation internally funded effort to develop a generic data modelling tool. The refinements and enhancements made to the process are described in Section 1.

2.2.1 Data Pre-processing

Different pre-processing was applied to the input IHUMS CI data to enable two types of anomaly model to be built for each monitored component; an 'absolute' and a 'trend' model. The absolute models identify combined CI values that are anomalous in absolute terms, whereas the trend models identify anomalous combined CI trends, irrespective of the absolute values of the indicators.

The pre-processing for the absolute models was minimal, and involved the application of a two-stage filter to each indicator: the first stage removed extreme (physically unbelievable) values and the second stage applied a median filter to remove up to two successive 'noise' spikes in the time-series data. (More than two successive outlying points may indicate a trend shift rather than a noise spike.)

For the trend models, the two-stage filter for the absolute models was used, and then a 'moving median difference' algorithm was applied. Following each new acquisition, the median of the time history was re-calculated and subtracted from the newly acquired value to provide a normalised value. This technique reduced the impact of early post-installation trends, and also small step changes due to maintenance, because the normalised value would gradually recover back to the median base line level. Although there was some distortion of the time-history, this simple approach worked well. The effects of the two stages of pre-processing are illustrated in Figure 2-4.

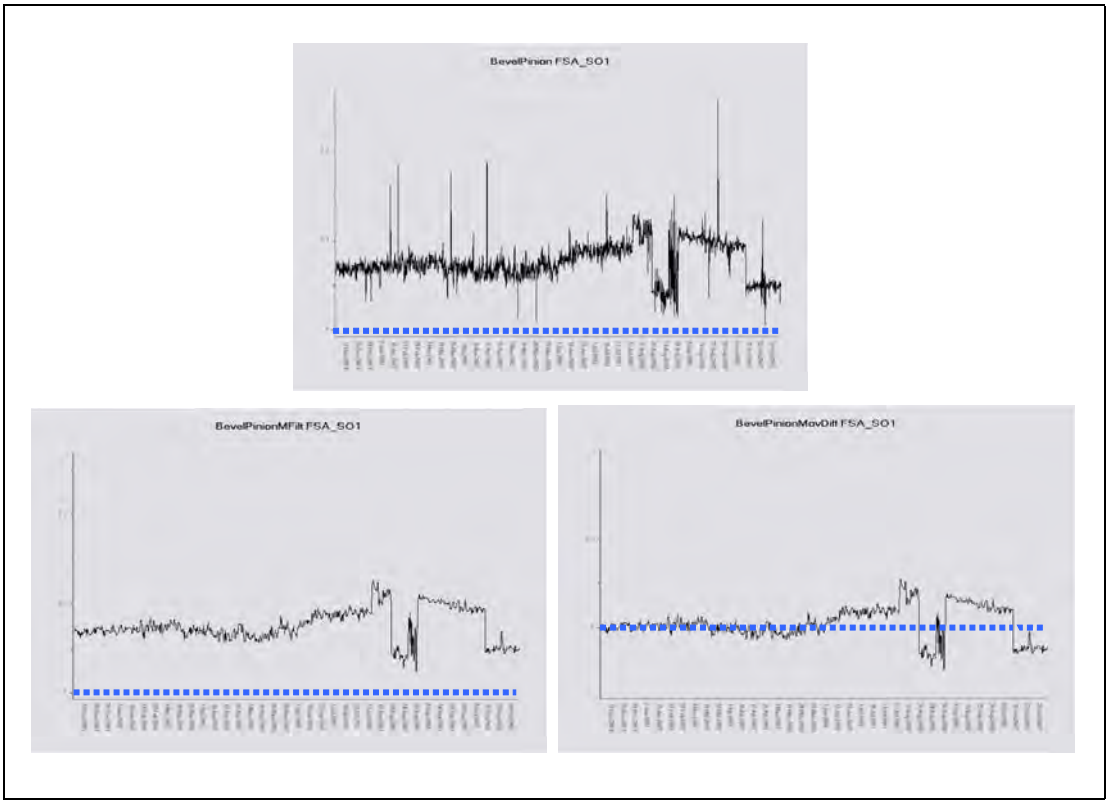

Figure 2-4 FSA_SO1: top – raw CI trace; bottom left – CI trace after first stage pre-processing; bottom right – CI trace after second stage pre-processing (blue dashed line = data origin).

2.2.2 Development of an Anomaly Modelling Capability

2.2.2.1 Gaussian Mixture Modelling

The anomaly detection capability developed is based on density estimation, using Gaussian Mixture Modelling (GMM). The anomaly models are a mixture of multivariate Gaussian distributions and Multinomial distributions (the multinomial aspect of the model is used to provide information for adapting the cluster model as discussed in the next section). A GMM is a probabilistic model and an object can be associated with more than one cluster in the model, unlike a purely distance based approach that performs hard assignment (i.e. an object belongs to a single cluster only).

The anomaly models are like cluster models but they are constructed for a different purpose. Clustering is often used to segment data to discover if the data contain any natural structure. The anomaly models are constructed for density estimation with the aim of identifying regions of space that are likely to represent anomalous behaviour.

In Section 2.1.1 it was shown that the IHUMS data were Gaussian like in shape but a single Gaussian distribution provided a poor model of the data. However, a mixture of Gaussian distributions can be used to model complicated distributions. Figure 2-5 shows the fit of a GMM to data from one HUMS CI. The model contains five clusters – the number being determined using the Bayesian Information Criterion (BIC) measure (see Section 2.2.2.2).

The GMM provides a good representation of the HUMS CI distribution. However, any anomalous CI data will also be included in this model, and therefore it is not suitable for anomaly detection.

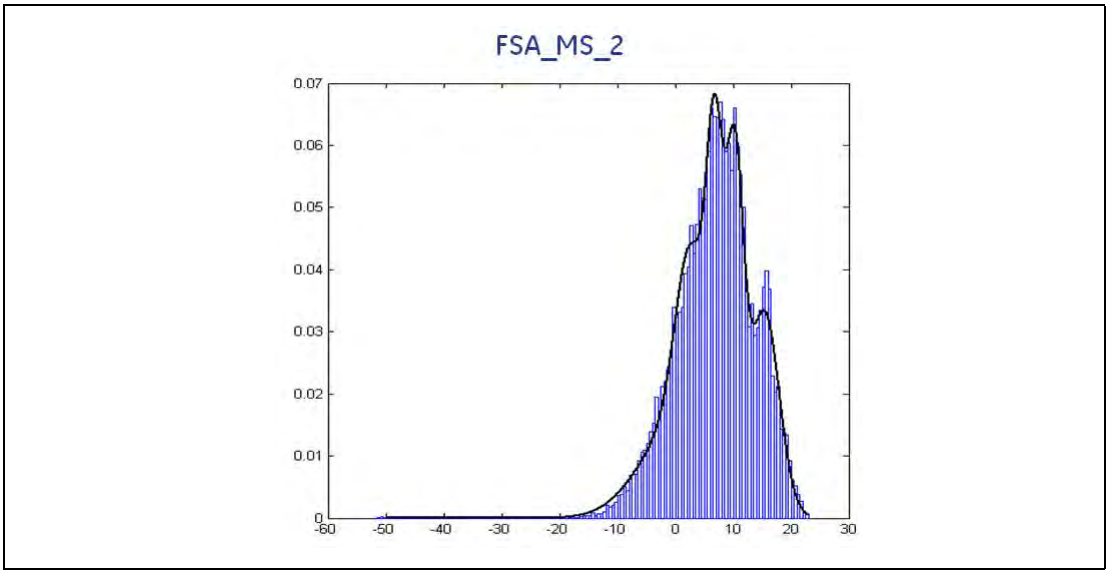

Figure 2-5 Fleet histogram for FSA_MS_2. The solid line is the fit from a GMM.

2.2.2.2 Anomaly Modelling

If it is known that the training data contain outliers, as is the case with HUMS CIs, a GMM has to be adapted so that it does not fit the outlying data. The output from this adaptation process is an anomaly model. The concept of creating an anomaly model is simple. A GMM is constructed, regions in the cluster space (i.e. clusters) that appear not to be representative of normal behaviour are detected, and these clusters are removed so that the final model provides a poor fit to those samples in the training data that are outliers. The model automatically detects the outliers: there is no manual a-priori tagging of outlier cases. However, it is not acceptable to build a standard

GMM and simply remove clusters, as these clusters can contain a mixture of normal and anomalous data. This simple approach can therefore trigger false positive or false negative indications.

The automated model adaptation process that has been developed is complex, but is controlled by a simple tuning parameter. A significant amount of effort was expended developing the adaptation process, as this is key to the successful building of models using in-service data containing various unknown anomalies.

GMM implementations have some method to estimate the optimum number of clusters to apply to a particular data set. There are many techniques for doing this but all of them are heuristic rules which must be carefully applied. One commonly used heuristic is the BIC. When training with data that contain anomalies, optimising the cluster space using the BIC is likely to mix anomalies with normal data. Therefore, the model adaptation process must address two fundamental questions: how is the number of clusters to be determined, and how are clusters identified for removal? The approach taken involves over-specifying the number of clusters. This means that the BIC metric is not used directly but is considered for judging the number of clusters. The number of clusters also links directly to the criteria for cluster removal. A cluster's removal is determined by an information measure based on the cluster's contribution to the 'likelihood' score.

A separate anomaly model is built for each monitored drive system component. Every component is tagged with a 'Component (or Gearbox) Fit' identity (ComponentFitID). A ComponentFitID identifies a specific component fitted to a specific gearbox fitted to a specific aircraft over a period of time. The ComponentFitID is used directly during the anomaly model build process. GE Aviation's cluster algorithm is used to build a GMM that contains information localised to a ComponentFitID. Statistics are then gathered on a generated model to assess the generic capacity of each cluster. A 'generic information measure' is calculated, based on a normalised likelihood measure that is computed using cross-validation (a statistical practice of partitioning data into subsets for training and validation). After the generic capacity of each cluster is scored, the clusters can be ranked as candidates for removal, the number of removals being determined by a threshold called the 'anomaly tuning parameter'. The anomaly tuning parameter is a percentage value that is user specified and indicates the estimated percentage of non-normal (i.e. noisy/anomalous) data in the training set. Data exploration and a review of maintenance history can be used to determine the optimum value of this parameter for a particular data set. Because HUMS VHM data can be relatively noisy, and be affected by undetected instrumentation issues, a relatively high percentage value was used. However, provided there is a good quantity of data to represent normal behaviour, a high value would not be expected to generate false positives.

The output from an anomaly model is a 'likelihood' score, which is referred to by GE Aviation as the FS, as it indicates the goodness of fit of data to a model (the FS decreases as data becomes a poorer fit to the model). Therefore, anomaly models are sophisticated statistical representations of the data generated from in-service experience, fusing sets of HUMS CIs to reduce a complex data picture into a single time history.

2.2.2.3 **Initial Set of Anomaly Models Built for the AS332L IHUMS Data**

Sets of anomaly models were built from a training database containing historical Bristow AS332L IHUMS CI data. With the exception of two drive system components that did not include a gear on the monitored shaft (the oil cooler fan and dual bearing module), every component had anomaly models constructed from two subsets of the CIs listed in Table 2-2.

- '8-Indicator' models, comprising the following eight indicators: {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.
- 'M6' models, comprising the following two indicators: {ESA_M6*, ESA_WEA}, (ESA_WEA is a derived indicator, calculated as the ratio ESA_SD/SIG_SD).

The CIs were split into these two sets to reduce the potential for one indicator to be masked by another; the criterion for this split was derived from engineering knowledge of the indicator design and characteristic behaviour.

Using the different forms of pre-processing, 'absolute' and 'trend' models were built for each set of indicators. Therefore, the majority of drive system components were monitored by four anomaly models – '8-Indicator Absolute' (8IA) and 'Trend' (8IT) models, and 'M6 Absolute' (M6A) and 'Trend' (M6T) models. The exceptions to this were the oil cooler fan and dual bearing module, which were monitored by only two models – absolute and trend models comprising the following '5-Indicator' (5IA and 5IT) set: {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

2.2.3 Generating Anomaly Alerts

Alert thresholds were calculated automatically from the FS data. Each model had an alert threshold and thresholds varied from one model to the next.

The FSs have a one-sided distribution. The FSs for the fleet of training components were first transformed. The extreme tail of the FS distribution is sparsely populated and so the end of the tail was chopped to provide a better fit (the chopped area includes definite anomalies). A Generalised Extreme Value (GEV) distribution was then fitted to the data, after which the FS value corresponding to nth cumulative percentile was determined. This score formed the alert threshold.

An alert was triggered when M out of the last N points were below the threshold (an anomaly is indicated by a negative FS). The default values for M and N were 3 out of 4.

2.2.4 Summary

The complete HUMS AAD process as initially developed is shown in Figure 2-6.

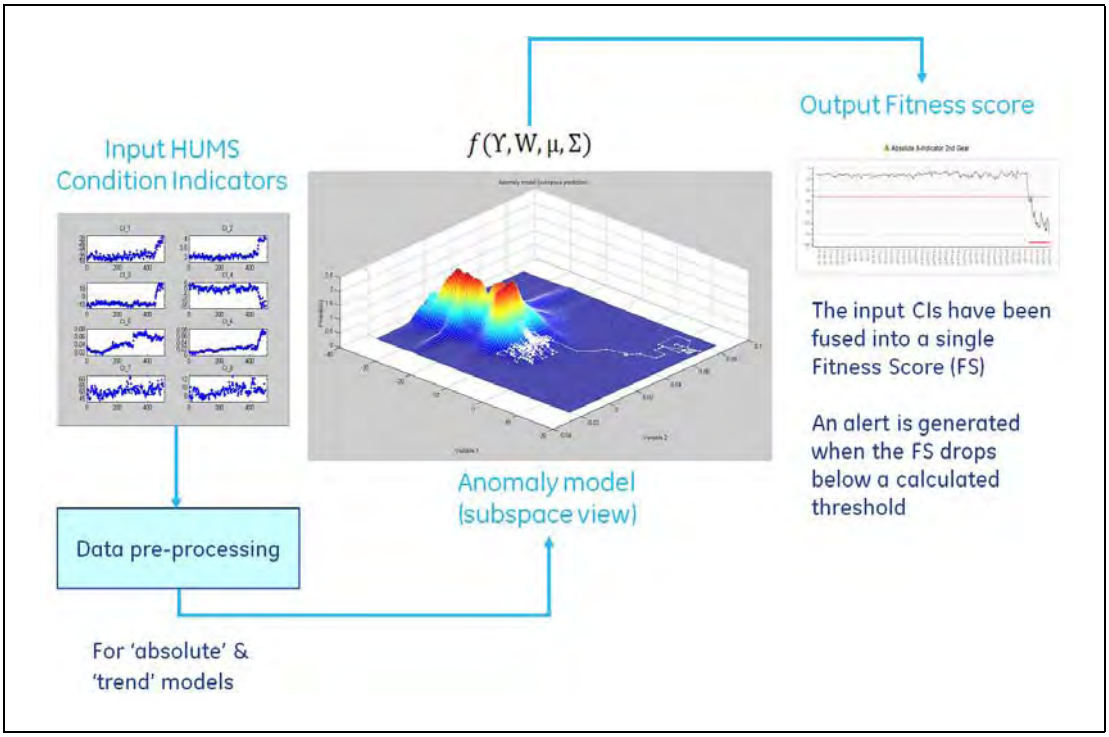

Figure 2-6 Advanced anomaly detection process

2.3 Off-line Analysis and Demonstration

Using the historical database of Bristow AS332L IHUMS data, an off-line analysis was performed to validate that the anomaly models behaved as intended. The key test for the AAD system was its ability to detect the cracked AS332L MGB bevel pinion (Figure 2-7) referred to in Section 1, this being one of the key drivers for initiating the research programme.

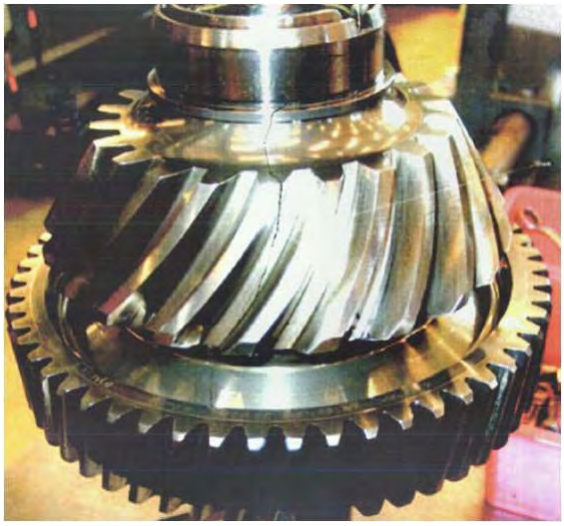

Figure 2-7 Picture of the bevel pinion with a large visible crack

For the bevel pinion case a number of HUMS CIs showed trends, as is evident in Figure 2-8. However, trends are quite common in HUMS data and no indicator on its own reveals any significant outlying behaviour when compared to the fleet. Figure 2-9 presents a 'box plot' of data for eight CIs from over 100 different gearboxes. The blue fleet +3 standard deviation bands shown in Figure 2-9 are computed from the models, from which anomalous data has been removed. The gearbox with the cracked bevel pinion is identified by an arrow. The only CI for this gearbox which appears high or low compared to the fleet is FSA_GE22. However, the mean value of the CI prior to the failure is also high, and the magnitude of the subsequent failure-related trend is not abnormal. It is therefore understandable that the HUMS did not generate any alerts.

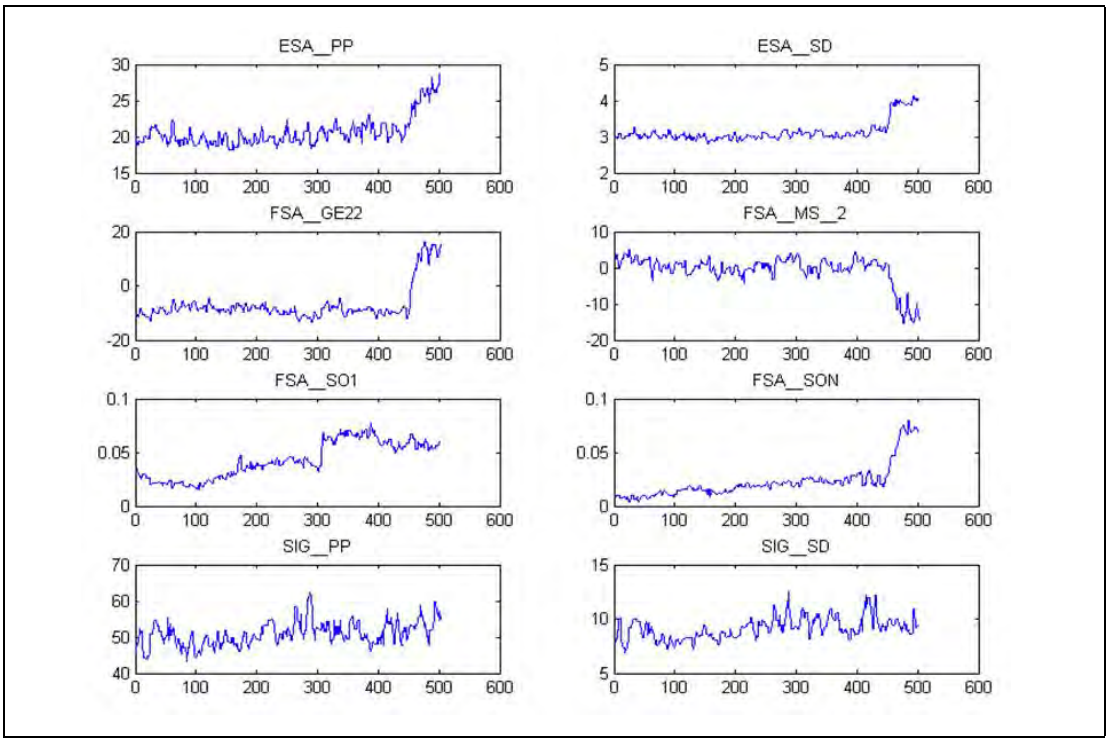

Figure 2-8 HUMS CI trends for the bevel pinion fault case

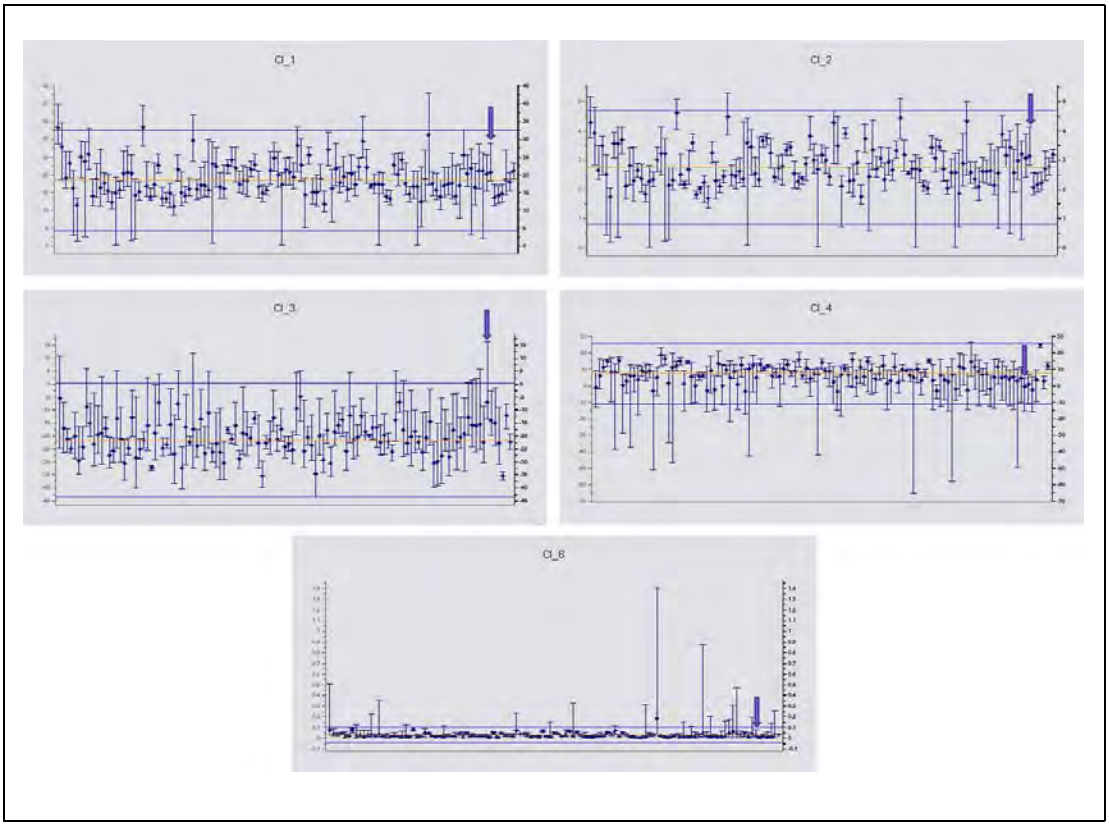

Figure 2-9 Fleet CI plots for the bevel pinion data (each dot is the mean CI value for a particular gearbox, with the vertical lines showing maximum and minimum CI values). The arrow indicates the cracked bevel pinion fault case.

Figure 2-10 shows the '8IT' and '8IA' model FS traces for the cracked bevel pinion. The two key features of the FS trace are the score's position relative to the threshold and the shape of the trace. The FS traces for the cracked bevel pinion would cause concern. FS values are initially high and stable, then clear and fast developing trends appear, with no features suggesting an instrumentation issue. The '8IT' model FS trend is clearly anomalous approximately 40 hours before the gearbox rejection, and the '8IA' model FS trend also becomes anomalous before the rejection. The results demonstrated that the AAD capability developed would have detected the bevel pinion crack.

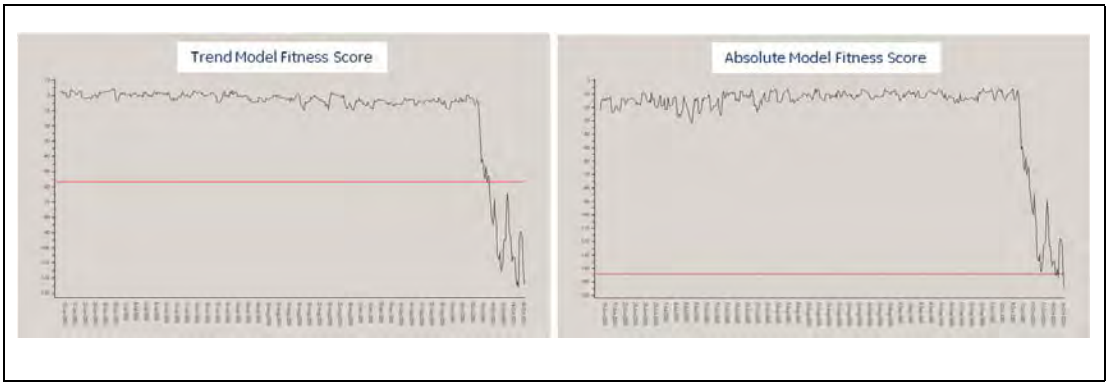

Figure 2-10 FS trends for the bevel pinion: left – '8IT' model; right – '8IA' model.

2.4 Trial HUMS Advanced Anomaly Detection System Implementation

The contractual requirement was simply to provide a system to support Phase 2 of the project, involving an in-service trial of the HUMS AAD capability developed in Phase 1. However, GE Aviation decided, both for technical and logistical reasons, that the best solution would be to implement a web-based system. The work involved exceeded that originally scoped for the programme, so GE Aviation funded the additional development costs internally.

The web-based system implemented for the trial, hosted on a server at GE Aviation, is illustrated in Figure 2-11. HUMS data files were automatically copied every night from the IHUMS ground station to a secure web portal at Bristow Helicopters in Aberdeen. The files were then automatically downloaded to GE Aviation in Southampton, imported into the AAD system's data warehouse, and analysed. Bristow had a remote secure login to the system to view results at any time via a web browser.

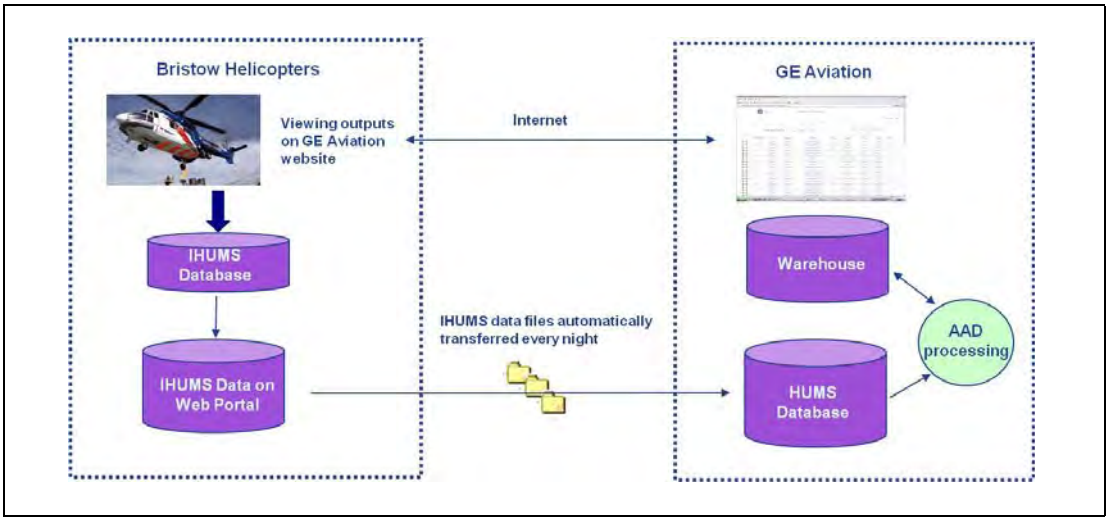

Figure 2-11 Trial AAD system implementation

3 Phase 2 of Contract No 841 – Six-Month In-Service Trial of Initial HUMS AAD Capability

The project report containing the detailed results of Phase 2 of the original contract was issued in July 2007 (again before Smiths Aerospace became part of GE Aviation), and its contents are presented in Annex B.

Phase 2 of the CAA research programme comprised a six-month in-service trial, conducted by Bristow Helicopters between May and November 2006, of the HUMS AAD system developed in Phase I. The system was introduced in Section 2.4, and is described in more detail in Annex B. It allowed the user to easily view new anomaly alerts across the fleet, and then drill down to anomaly model FS traces, plus pre-processed and raw HUMS CI traces, again with an ability to view these on a fleet-wide basis. Some example displays are shown in Figure 3-1.

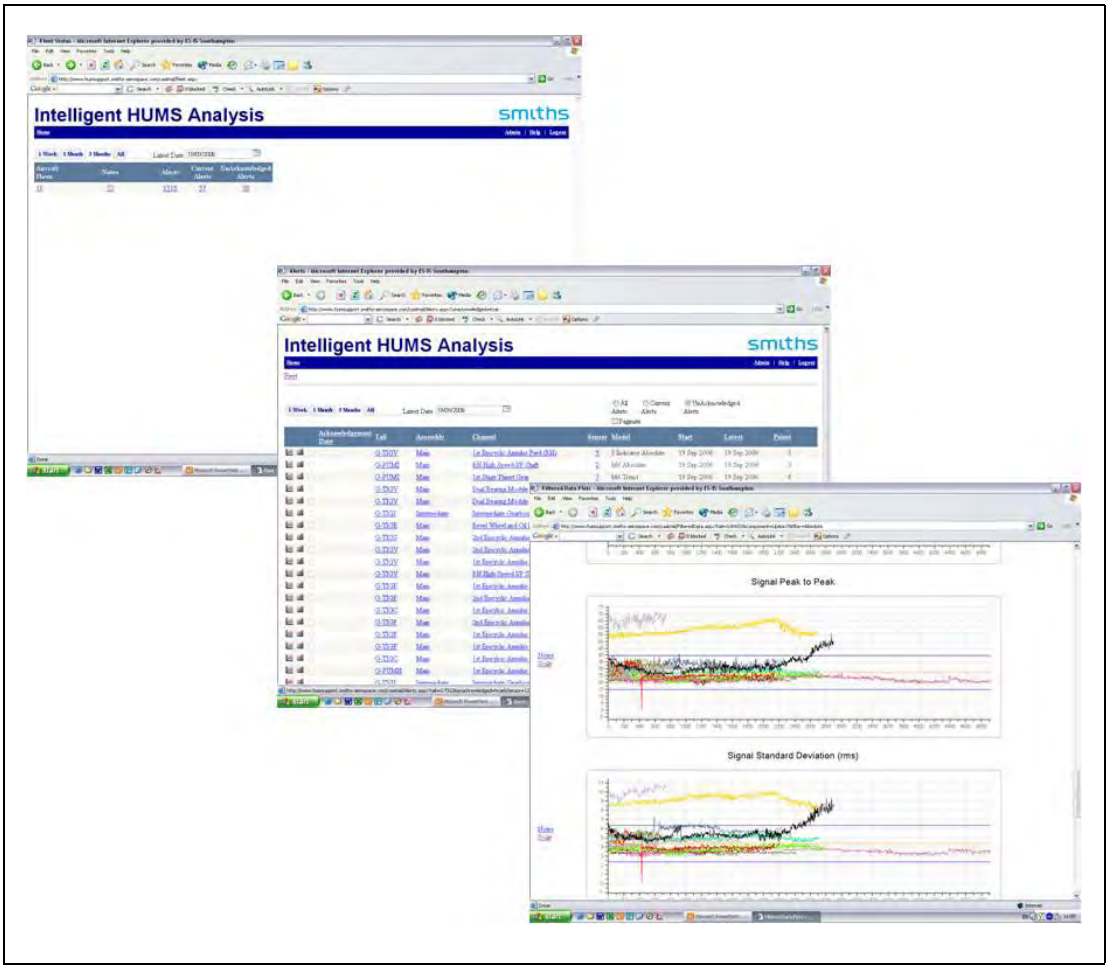

Figure 3-1 Example HUMS AAD trial system displays

In addition to displaying anomaly alerts, for comparison purposes the system could also indicate the occurrence of IHUMS alerts and sensor faults. The user could also enter notes to record any investigation actions and associated findings. This documentary information is shown in Figure 3-2. Facilities were provided for acknowledging alerts, and if a maintenance action caused a step change in the data that triggered a trend model alert a 'reset trend' button enabled the trend pre-processing to be reset, which cleared the alert.

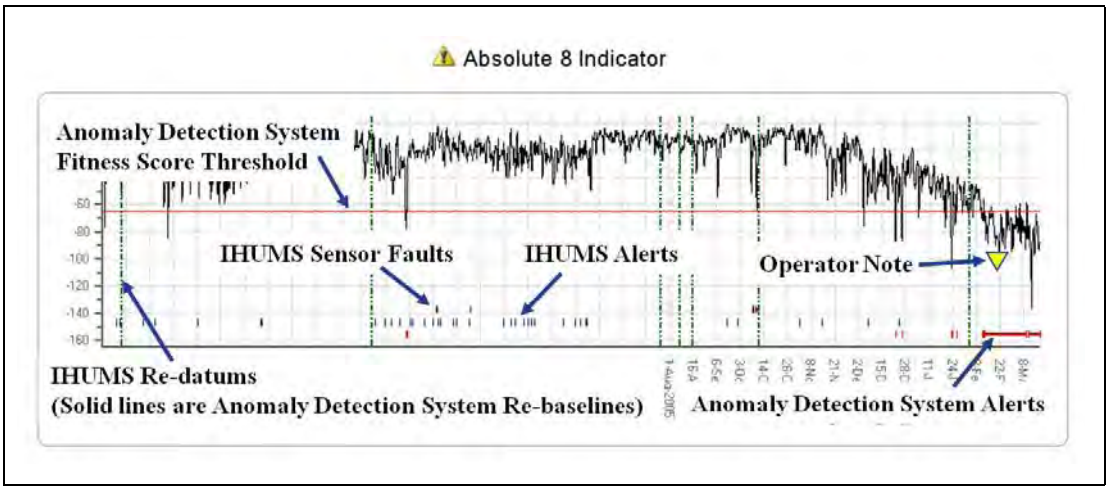

Figure 3-2 Documentary information shown on a Fitness Score trace

3.1 Trial Experience

Trial operational procedures and experience are described in Annex B, which also includes a catalogue of significant findings, and a statistical analysis of the anomaly alerts generated by the system. The reliability of both the nightly automated data download from Aberdeen and the access to the web server was good, demonstrating the practicality and robustness of the web-based implementation of the AAD system.

3.1.1 Catalogue of Significant Findings

Annex B contains a selection of 15 example cases to illustrate the performance of the AAD system. The cases were categorised according to the combined findings of both this system and the traditional IHUMS analysis. In the categorisation, faults were considered to include both aircraft component faults and HUMS instrumentation faults. In the majority of cases it was not possible to positively confirm the presence of a specific fault. Therefore, a correct fault indication was considered to have occurred if:

- the operator determined that maintenance action was required (e.g. to replace an accessory gearbox, or a HUMS accelerometer);
- the maintenance action performed resulted in data returning to normal levels.

1) Cases where AAD identified a fault not seen by the existing HUMS

The six cases in this category related to the detection of one potential MGB fault and five IHUMS instrumentation faults. Based on the CI data, the AAD system proved to be much more effective at detecting instrumentation faults than the IHUMS. The number of sensor and wiring faults detected reflects the fact that the IHUMS is a 'first generation' system, with the installation being performed by the aircraft operator. The AAD system also detected an IHUMS Data Acquisition and Processing Unit (DAPU) fault, which triggered a large number of anomaly alerts. These moved with the DAPU when it was swapped between aircraft.

A potential MGB fault was detected when performing a data review prior to the start of the trial, and was related to the 2nd stage epicyclic annulus gear. The '8IA' model for one annulus accelerometer generated a continuous alert from the first half of February 06, and the '8IT' models for all three annulus accelerometers generated consistent alerts from late 2005. The alerts were primarily caused by rising trends in SIG_SD from all three accelerometer positions. The MGB was rejected for metal contamination on 20 March 2006. The source of the metal contamination could not be identified and, although a gearbox strip report was

requested from Eurocopter, no feedback was received. It was therefore not possible to confirm whether the AAD alerts and rising SIG_SD values were associated with the metal contamination. However, it was noted that such contamination often originates from the mast bearing, and the IHUMS does not detect this. If this was the source of the contamination, it is possible that the anomaly alerts were related as the signal averages for the 2nd epicyclic annulus gear contain main rotor shaft vibration information.

2) Cases where AAD corroborated existing HUMS indications

The cases included in this category were a potential LHA gearbox fault, an oil cooler fault, and two IHUMS instrumentation faults.

The '8IA' and 'M6A' models generated alerts on the left hydraulic idler in a LHA (Left Hand Accessory) gearbox. The CI trends were mirrored in the data for the RHA (Right Hand Accessory) right hydraulic idler, but at lower levels. The IHUMS also generated frequent alerts, resulting in the LHA module being placed on close monitor. The gearbox was subsequently rejected based on the combined IHUMS and AAD indications.

Very high levels of FSA_SO2 and SIG_SD caused continuous '5IA' model alerts on an oil cooler fan. The IHUMS was also generating alerts and the fan was replaced, after which the vibration data returned to normal. No fan or bearing defects were found. However, there was evidence that the fan had been rubbing on the casing, which may have been distorted as a result of the metal securing-band being too tight.

3) Cases where AAD failed to identify a fault that was seen by the existing HUMS

One case was placed in this category. Bristow's HUMS engineer reported that the IHUMS had generated alerts on FSA_SO1 for a TGB (Tail rotor Gearbox) output shaft, and inspections had found damage on three tail rotor flapping hinge bearings. Although the anomaly model FSs responded to the increased SO1 values, neither the '8IA' or '8IT' models generated any alerts. It was reported that bearing damage had also occurred on other aircraft due to a change in grease. An investigation suggested that the reason no anomaly alert had been generated may have been due to the presence of noisy data in the training set that could be related to the existing problems. However, it was also considered that the '8-Indicator' gear models may not be sensitive enough to detect a single FSA_SO1 CI trend on the TGB output shaft. Based on this example it was recommended that, for the monitoring of key input/output shafts (e.g. MGB inputs and TGB output), consideration should be given to implementing new models using only shaft-related parameters (i.e. SO1 and SO2) to increase system sensitivity to shaft faults.

4) Cases where AAD identified an existing HUMS false or premature alert

The IHUMS generated continuous alerts on a RHA gearbox through most of November 2006, and the module was placed on close monitoring. These alerts were triggered by rising trends in ESA_PP and ESA_SD. However, neither the '8IA' or '8IT' models (which have these indicators as inputs) generated any alerts during this period. The reason for this was apparent when viewing fleet displays of the CIs. Trends in these indicators were common, and a number of other aircraft had higher indicator values. The CIs started trending down again in late November 2006, and the RHA gearbox was removed from close monitoring. It was considered that the '8IA/T' models were correct to suppress the indicator trends and not generate alerts.

5) Cases where AAD generated a false or premature alert

After correctly detecting high and variable FSA_SO1 values on a MGB left torque shaft – aft end, the '8IA' model's FS values subsequently remained at or just below the alert threshold, developing a slowly decreasing trend which triggered a continuous alert. This was due to SIG_PP and SIG_SD being at the upper limit of fleet norms, and also having slowly rising trends. The CI levels were not clearly abnormal compared to the rest of the fleet, and the data showed only slowly rising trends. Bristow's HUMS engineer considered that no immediate maintenance action was required, and the anomaly alerts were probably premature.

A second example was the RHA gearbox referenced in item 4 above, where an 'M6T' model produced some alerts that were considered to be premature. While there were trends in ESA_M6 and WEA, these stayed well within the fleet norm and the 'M6A' model did not generate any alerts.

6) Cases of anomaly indications with an unknown outcome

Two examples were included in this category, one relating to an MGB left torque shaft – fwd end, and another to a LHA gearbox hydraulic drive. In both cases, although findings at the end of the trial period were inconclusive, the AAD system was considered to be correct in drawing the operator's attention to anomalous CI data.

3.1.2 Statistical Analysis of AAD System Performance

A statistical analysis was performed of the AAD system and IHUMS alerts in a six-month period of operations from 1 June to 30 November 2006. Table 3-1 shows the number of alerts from the two systems, itemised by IHUMS sensor.

Table 3-1 AAD system alerts and IHUMS alerts (sensor and defect)

Sensor	AAD System Alerts			IHUMS Alerts		
	No. of Analyses	Alerts	Normalised Alerts	No. of Analyses	Alerts	Normalised Alerts
1	2	49	24.5	2	58	29.0
2	2	120	60.0	2	40	20.0
3	9	379	42.1	7	271	38.7
4	6	245	40.8	5	221	44.2
5	2	130	65.0	2	77	38.5
6	2	99	49.5	2	86	43.0
7	6	365	60.8	6	392	65.3
8	1	4	4.0	1	47	47.0
9	1	23	23.0	1	54	54.0
11	2	38	19.0	2	97	48.5
12	2	32	16.0	2	38	19.0
Totals	35	1,484		32	1,381	

As explained in Section 2.1, while the IHUMS performed a total of 32 drive system component analyses, the AAD system performed 35 analyses. The IHUMS alert figures combine both "sensor" and "defect" alerts, as the AAD system alerts did not differentiate between these.

Comparing the number of alerts, during the six-month trial the AAD system triggered a total of 1,484 alerts for the 35 analyses, which equates to an average rate of 42 per analysis. The IHUMS triggered a total of 1,381 alerts for 32 analyses, which equates to an average rate of 43 per analysis. Therefore, using the initial set of anomaly models and thresholds, the AAD system's alert rate was similar to that of the IHUMS. As the comparison was between a system undergoing its first operational trial and a HUMS that has been in service for approximately 15 years, this was viewed as an encouraging result. Furthermore, as the anomaly alerts were intended to draw the operator's attention to any abnormal data behaviour, the AAD system's thresholds had been deliberately set at a relatively sensitive level.

A comparison of the durations of the AAD system and IHUMS alerts showed that, while the IHUMS alerts were cleared quickly, there were a number of long duration AAD system alerts. The IHUMS uses datum thresholds, which are calculated on a limited number of data samples, and the system is normally re-datumed after an alert has been investigated. As a result, the IHUMS may not alert on data that are consistently anomalous compared to the fleet norm. However, the AAD system continues to generate alerts if the data remain abnormal, for example due to persistent IHUMS instrumentation faults.

An analysis was performed of the number of AAD alerts by sensor, analysis, model type and aircraft. It was shown that, in most cases, a majority of the alerts were generated by only a small number of aircraft. This demonstrated that the AAD system alert rate was not driven by random alerting, but by the system's response to particular and persistent aircraft anomalies. A comparison of alerts by model type showed that the '8IA' and '8IT' models generated the greatest number of alerts. The '8IT' models triggered slightly more alerts, but the '8IA' models had significantly more data points in alert. This was believed to be due to the fact that the system identified a significant amount of anomalous IHUMS data due to instrumentation problems.

3.1.3 **Summary**

The initial six-month trial confirmed that the AAD system could highlight anomalous data that were not seen by an existing HUMS. It also provided a clearer picture of anomalous patterns on particular aircraft and drive system components than is possible with the traditional HUMS analysis. It was therefore concluded that the AAD system improved the effectiveness of HUMS, and also provided valuable assistance to an operator in the management of the system.

4 Additional Research Work (Tasks 1 to 4)

Building on the success of the initial trial period, four additional tasks were carried out to research and develop further advanced HUMS data analysis capabilities prior to conducting a second six-month in-service trial. The project report presenting the results of the additional research work was issued in July 2009, and its contents are reproduced in Annex C.

4.1 Re-modelling

After using engineering knowledge to choose the CIs included in a model, the construction of anomaly models is largely an automated task, controlled by a tuning parameter that determines what percentage of clusters to remove from a model to suppress the effects of bad (or fault contaminated) data in the training set. The initial intention of this task was to further enhance AAD system performance by performing model tuning and re-modelling based on experience gained during the first six-month in-service trial.

For example, for critical shafts in the rotor drive system there was a recommendation to include new 'shaft order' models (Shaft Order Absolute – SOA and Shaft Order Trend – SOT) based on FSA_SO1 and FSA_SON.

NOTE: The system architecture allows models to be easily added or removed.

In addition, the trial showed a lack of sensitivity of the IHUMS to many instrumentation problems. Because these problems went undetected, bad data could be acquired over many hours of flying. This raised two concerns for anomaly modelling. First, although the anomaly modelling technique has a unique way of handling bad (anomalous) data in the training set, there comes a point where the quantity of bad data is sufficiently large for it to be classed as normal. Second, the training data were a random selection of 'gearbox fits' (gearbox/aircraft combinations) and there was a risk that the training could be biased by the fact that a large percentage of the bad data might be associated with a small number of gearbox fits.

4.1.1 Enhanced Modelling Capability

Following the completion of the first six-month trial, GE Aviation conducted an internally funded exercise to re-implement the software for the core data mining algorithms used in the anomaly modelling process. This also facilitated a review of the modelling approach from first principles, i.e. from a theoretical viewpoint. The work showed the approach to anomaly modelling to be well founded, and produced ideas to improve the technique. The result of this internally funded exercise was an improvement in the robustness of the technique.

The improved anomaly model construction technique makes greater use of gearbox fit information in the model building process. It now automatically detects when a gearbox fit has insufficient data to contribute to an anomaly model and excludes that particular fit. At the same time it applies equal modelling resource to each of the remaining gearbox fits, and so does not bias the model to fits containing a large number of acquisitions. The improved technique also removes the requirement to categorise data as training, test or validation, which is a common practice in data mining techniques to ensure that built models will generalise to data not used for training. All data can now contribute to building a model. This has the further advantage that online model updates can be performed as new data are acquired. The ability to update models is important, particularly for a new aircraft type. The improved technique also allows the influence of a gearbox fit to be removed from a model without having to repeat the whole training process. This is another valuable feature because the process of model building and review can highlight previously unknown

data problems that can show a gearbox fit to be so abnormal that it should not be used for training.

Another modelling change following the initial trial was to allow clusters to 'rotate' to model correlations between CIs (instead of using fixed axes, the axes of the clusters can rotate to align with the correlations in the data). This did not have a significant impact on the anomalies detected, but it did allow more diagnostic information to be derived from the model, such as identifying a de-correlation between CIs.

4.1.2 Model Tuning and Re-modelling

All the anomaly models were re-built using the improved modelling technique, which provided a significant enhancement to the anomaly modelling capability. The new procedure for building a model was to use all available data apart from cases that were known a-priori to be anomalous. The models for the second six-month trial were built using all Bristow 332L IHUMS data from 1 January 2001.

The anomaly model tuning parameter specifies when cluster removal to eliminate the effects of suspect data should halt. The parameter is set to a value based on a judgement of how much bad data exists in the historical database. The percentage of data (influencing the model) removed would normally be set to a value of 1 or 2 but the default value for the IHUMS data was 8 due to the prevalence of instrumentation issues. If a gearbox fit is known to have developed a fault, or the CI responses are known to be unrepresentative of a normal vibration signature, the data for that gearbox fit are not used for training the model.

The data associated with the catalogue of significant findings from the first six-month trial reported in Annex B was reviewed following the re-modelling. The previously reported anomalies largely remained unchanged following the modelling improvements but there were some minor differences. All six of the cases where the anomaly detection identified a fault not seen by the existing HUMS remained clearly anomalous, and some of the anomalies were further emphasised following re-modelling. The significance of the cracked MGB bevel pinion fault case was further emphasised, with the '8IA' model FS values now exceeding the threshold at a similar point to the previous '8IT' model (i.e. approximately 40 hours before gearbox removal, see Figure 4-1).

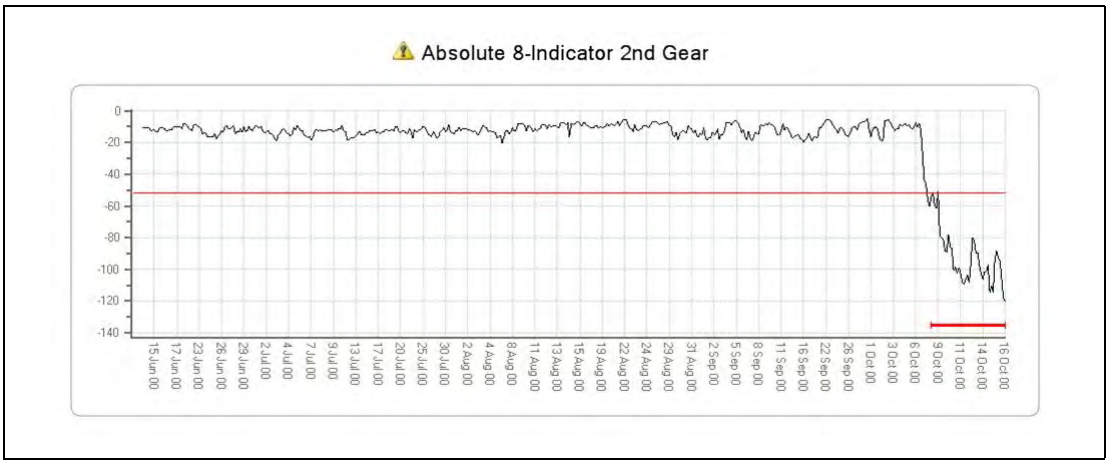

Figure 4-1 The Fitness Scores from the new '8IA' model for the bevel pinion crack case

New 'shaft order' (SO) models were implemented for the MGB input and TGB output shafts using only shaft related parameters (FSA_SO1, FSA_SON). For the TGB case where the original '8-Indicator' models had failed to detect tail rotor flapping hinge bearing damage, the new 'SO' models generated clear alerts for these acquisitions.

4.2 Probabilistic Alerting Policy

During the first six-month trial, alerts were generated using an FS threshold. Each model had its own threshold derived using an Extreme Value Distribution built from training cases with low FS values. Deriving the Extreme Value Distribution was not straightforward, and was made more difficult by the nature of bad data associated with the previously described undetected instrumentation problems. The distribution and scale of the FS values varied between models and therefore, thresholds based on these also varied. It was concluded that the anomaly detection process would be enhanced if the measure of what is anomalous was normalised across models. Normalisation allows model outputs to be compared and could facilitate future enhancements, for example the implementation of a secondary process such as automated reasoning to assess the nature of an anomaly.

The normalised PA measure was therefore derived to create a more robust threshold setting process, and make it easier to interpret the significance of anomalies. The PA is a normalised probability measure that ranges between 0 and 1, and is used for defining the alerting threshold. A global PA threshold can be defined that is common to all models and, for the CAA trial, this was set to 0.9 (the FS equivalent of the 0.9 PA value was different for each model). The default PA threshold for an individual model could be changed if it was deemed that the alert rate was too high or low, however this capability was not needed.

For each model there is a PA distribution which is an Extreme Value Distribution, built using synthesised data. An FS value is passed to the PA distribution and a PA value is returned. Most FS values will return a PA of zero because most acquisitions will be normal. The resolution of the PA distribution was set quite low, so that PA values could often appear to be binary in nature. The resolution can be changed, but a low resolution had the advantage of taking the variability out of the signals to provide a cleaner picture of the output, which helps with visualising results. No information is lost due to the resolution because signal shape and detail is contained in the FS trace, which remains a key output of the anomaly modelling.

Sampling theory was used to derive the PA distribution as this gave better coverage of a model's behaviour, i.e. it was not constrained by anomalies seen to date. The idea of sampling was based on the assumption that a model represents the statistical behaviour of a component type. It is better to build a model of normal behaviour and generate samples from this model that are outliers, as opposed to modelling with a much smaller set of acquired actual outliers. This provides more assured data coverage and is a conservative approach in that, if the model does not accurately represent a component's statistical behaviour, the result will be increased alerts rather than masked faults.

The FS values from the new '8IA' model for the cracked MGB bevel pinion fault case were shown in Figure 4-1. The equivalent PA trace is presented in Figure 4-2. Variability of data within the normal range is now suppressed, and the PA values show a clear response to the developing crack. (The PA distribution was used to set the FS threshold, so the red threshold lines in both figures correspond to the same level of abnormal behaviour.)

Again using internal funds, GE Aviation implemented software for learning the Extreme Value Distribution as part of the anomaly model building process. This implementation included code for generating simulated extreme outliers from a model, fitting an Extreme Value Distribution, and extraction of the FS equivalent to the default PA value of 0.9. When a model is learnt, the PA distribution is automatically generated and its threshold (FS equivalent) stored.

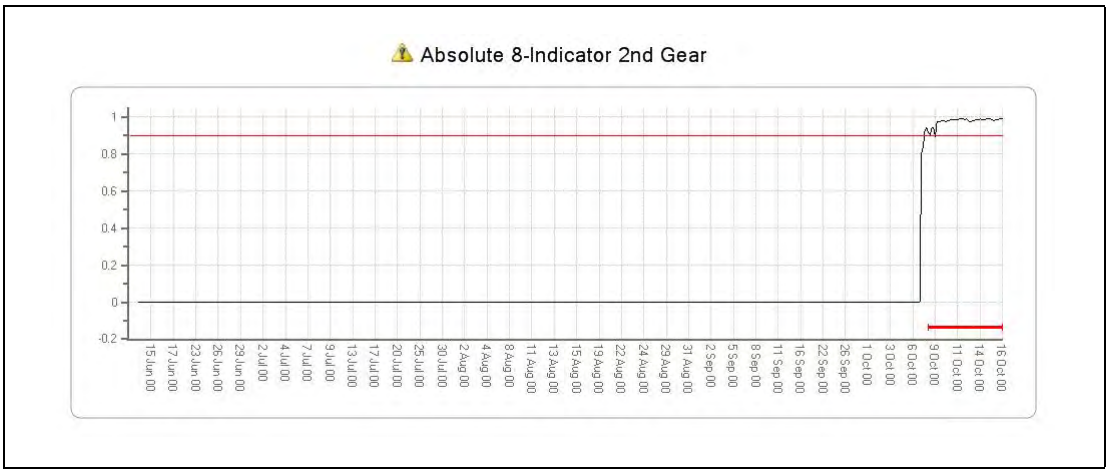

Figure 4-2 The '8IA' model PA output for the cracked MGB bevel pinion fault case

4.3 Influence Factors

In the first six-month trial, an anomaly alert was investigated by a manual inspection of the CIs to see which of them might be responsible for the alert. Charting of CIs is a useful technique but it can lead to incorrect interpretation. The numerical range and scatter of CIs vary considerably, and it can be very difficult to visually assess the significance of a CI trend. Univariate fleet plots of CIs enable viewing of the trend for an individual component in the context of the fleet, but these fleet plots can give a false picture when the data contain a lot of scatter because they hide where the true density of the data lie.

This task was to assess the suitability of IFs, which are part of GE Aviation's advanced data modelling technology, for assisting the interpretation of the significance of individual HUMS CIs, and to implement them for the second six-month trial if found useful. IFs are a type of prediction (i.e. an output from a trained model) that can provide diagnostic information about an anomaly model, and its input variables (e.g. HUMS CIs).

An IF-predicted value is generated for each CI used to train a model. A history of IF values for a model or component (i.e. an IF trace) provides information regarding the influence of different HUMS CIs on the fused FS. These traces are not the same as plots of CIs as they are assessing the contribution of individual CIs to a statistical measure of abnormality. Unlike CIs, IFs are normalised and can be directly compared. There are different types of IF, each type producing a different view on a model or input variable. The interpretation of a variant providing information on correlations between HUMS CIs can be rather complicated. However, this variant has proved very useful on other types of data (e.g. that obtained from multiple sensors). Only one type of IF that allows each CI to be assessed independently of other CIs is presented to an operator, as this type of IF shows trends that are closest to the CI plots and with which the operator is most familiar.

The IFs can be scaled, and the default method of presentation adopted for the HUMS AAD system was to scale the IFs for an anomaly model such that they sum to 1 when the FS crosses the alert threshold. Therefore, it is possible to examine the IF traces for a model and see whether a component is behaving abnormally, and which CIs are contributing to this abnormal response. The IFs are easy to interpret and provide useful diagnostic insight into the cause of abnormal behaviour.

The familiar case of the cracked bevel pinion is used to demonstrate the IFs. The HUMS CI trends for this case were shown in Figure 2-8. The associated '8IA' model

IFs are shown in Figure 4-3, and the '8IT' model IFs in Figure 4-4. Each CI in Figure 2-8 has its own scale and although trends are visible there is no indication within any plot that can suggest a trend is anomalous. The IFs, however, show that several CIs are responding to the developing bevel pinion crack, and that there are some differences between the CIs that are significant in absolute terms, and those that have significant trends. The information contained in these IF plots is extremely difficult to assess in traditional CI plots, and virtually impossible to quantify.

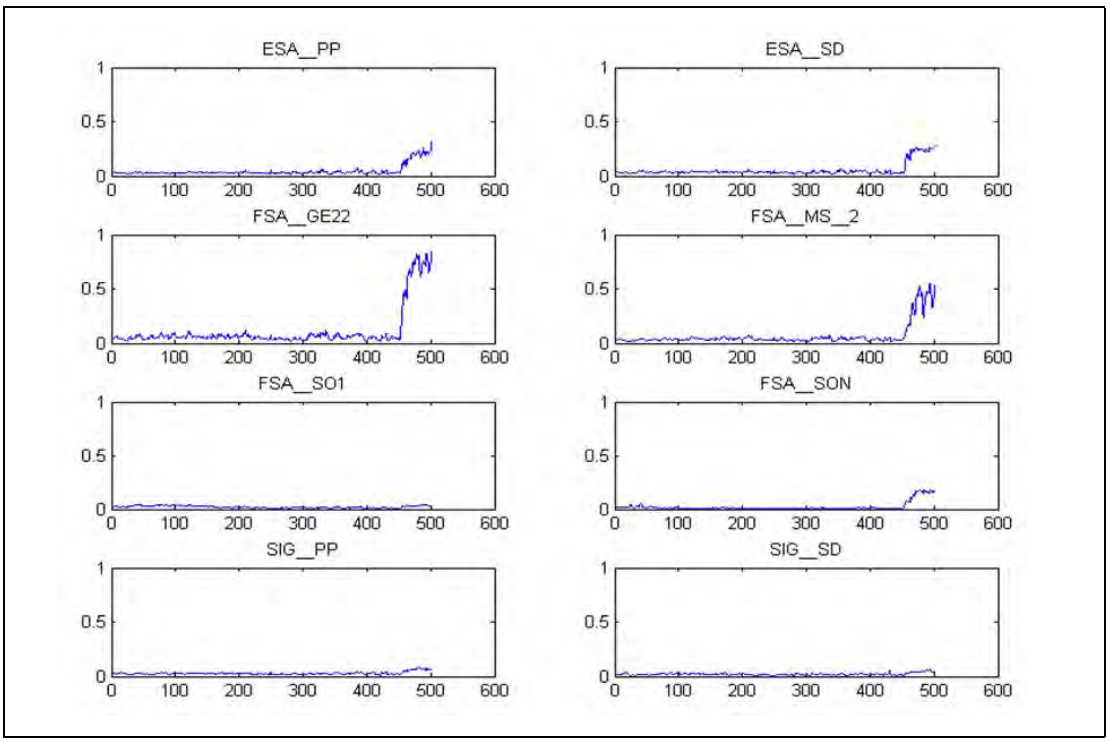

Figure 4-3 IFs for bevel pinion case absolute model

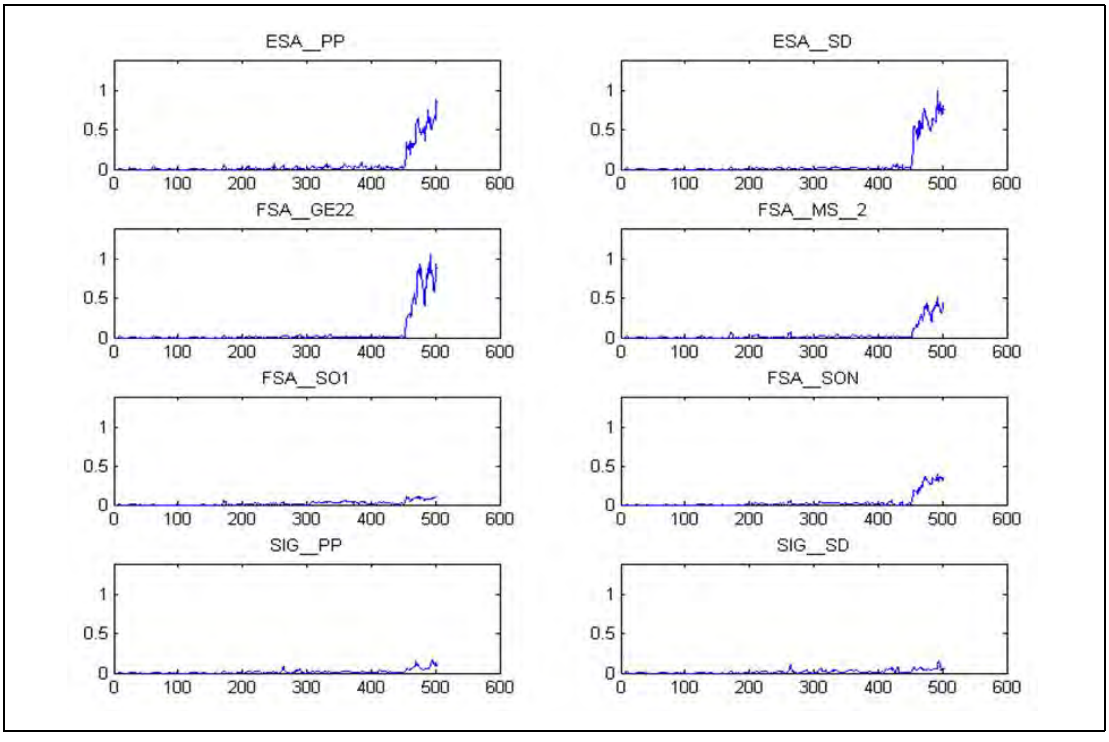

Figure 4-4 IFs for bevel pinion case trend model

4.4 Signal Characterisation

The final additional research task was to investigate the use of additional pre- and post-processing techniques to characterise data trends. Key diagnostic and prognostic information is contained in the shape of a signal. Step changes are often the result of maintenance in which event the cause can be explained but, in other circumstances, it can denote a sudden failure, perhaps in a sensor. A monotonic trend where a signal climbs or drops over successive samples can be indicative of some gradual failure, but it might also be a subtle change due to some unknown environment influence. The onset of erratic 'spikes' in a signal often indicates an instrumentation issue. The ability to detect these signal traits is important when assessing the significance of abnormal data and reasoning about the cause of the anomaly.

Shape characterisation can be very difficult, and the aim of the task was to perform a brief overview of methods that could assist with this characterisation. Because of the uncertainty of the outcome there was no commitment to implement any new techniques, however the topic was considered to warrant continued effort because of the diagnostic information it can provide. Some further work on signal characterisation, including trend detection and severity assessment, was carried out in the final two research tasks, and is described in Section 6.2.2.

Pre-processing involves the transformation of the signal before it is used to train an anomaly model, with the intention of making the model sensitive to key features of a signal such as a trend. The modelling already employs two levels of pre-processing: the first is the median filter to remove single or paired 'spikes', and the second is a moving median difference to detect changes in a signal. The moving median difference is attempting a form of normalisation as opposed to signal characterisation. An investigation was performed into the use of further pre-processing, however it was concluded that transformations on the CIs before input to the model learning should be kept to a minimum. This is because anomaly modelling that is not dependent on a potentially complex pre-processing stage (i.e. fewer assumptions about CIs made prior to modelling) would be more robust. It is also more efficient to act on the single fused signal of the FS than on multiple CIs. Finally, the modelling itself is more stable because it is not subjected to continual rebuilds on different transformations of the CIs. Therefore, the task concentrated on an investigation of options for post-processing to characterise FS trends.

The detection of different signal features such as trends, step changes, etc. would be relatively straightforward if the signals were noise-free and stationary, and the nature of change simple (e.g. a step change or linear trend). It is the continuous variation in HUM CIs without a complete understanding of their cause that makes the task of detecting meaningful changes particularly difficult. A review of the literature revealed no proven reliable methods for characterising signal features in these circumstances.

Although the investigation did not result in the implementation of any new functionality in the AAD system, it did provide a good foundation for the future development of new capabilities for data characterisation. Shape in a signal can be confused or hidden by excessive noise. Smoothing is required to emphasise salient features of a signal and therefore a number of filtering techniques were explored. Segmentation was also seen as an important requirement. Fitting a single model to a complete HUMS CI or FS time history that exhibits different types of behaviour is difficult to achieve reliably and would make the detection of shape characteristics very complicated. Therefore, it was proposed that a signal should be segmented into sections of homogeneous behaviour and a number of candidate segmentation methods were investigated. Within each section, models can then be fitted to

describe the nature of that section of signal and characterise any change, and also indicate the severity of any trend.

4.5 Summary

The additional research tasks resulted in significant enhancements to the HUMS AAD system, which would be evaluated in the second six-month trial period. The final AAD process incorporating these enhancements is shown in Figure 4-5.

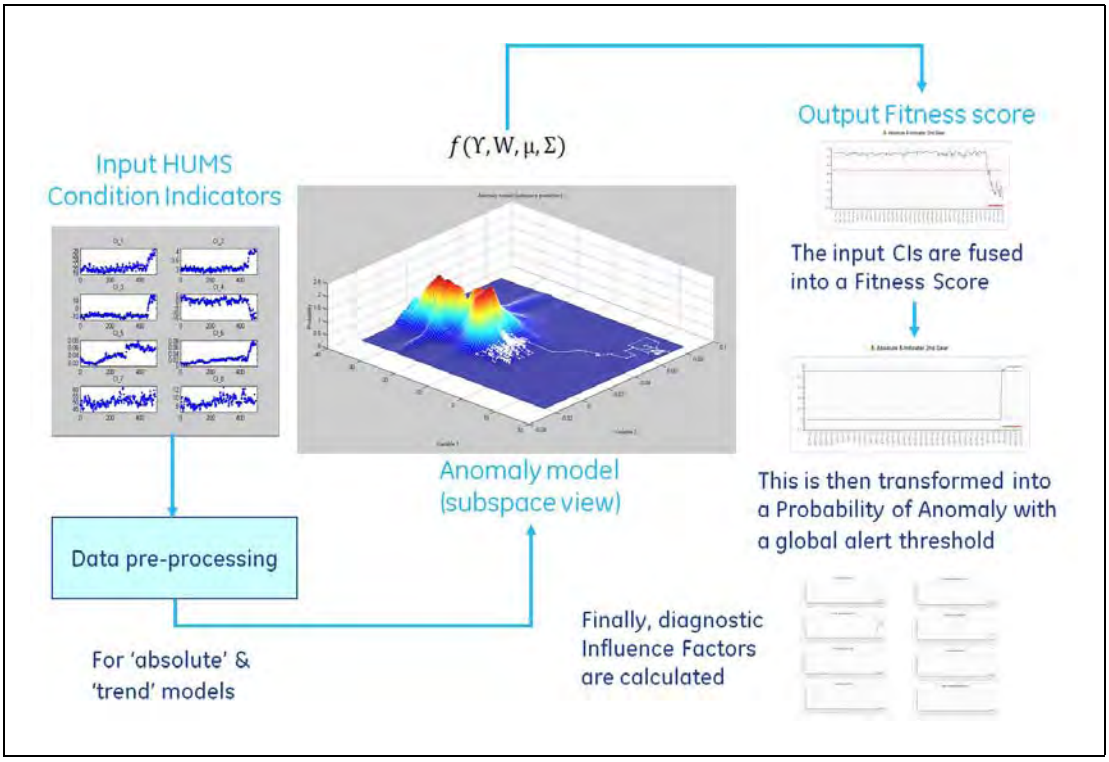

Figure 4-5 Final advanced anomaly detection process

5 Second In-Service Trial Period

The second in-service trial period commenced at the beginning of January 2008, with the formal trial period completing on 30 June. However, the trial was continued informally until 19 December 2008 while two final research tasks were completed. The project report presenting the results of the second trial period was issued in August 2009, and its contents are reproduced in Annex D.

The second trial period evaluated the enhanced HUMS AAD system incorporating the updated anomaly models and the new PA and IF outputs. To minimise Bristow's workload associated with the trial, this time GE Aviation also provided an analysis service, identifying any detected anomalies which Bristow should follow-up in daily fleet health status reports. Bristow then used the AAD system to investigate the identified anomalies. The reason for trialling an analysis and reporting service in addition to the HUMS AAD system was the recognition that some helicopter operators have limited manpower and skills for HUMS support activities. Irrespective of how effective the AAD system is, this still represents another system that must be regularly checked, and many operators are already having to support a number of different HUMS. Some operators might therefore prefer to access the new capability through a service rather than an 'in-house' system.

The only differences made to the system implemented for the second period were the incorporation of the new anomaly models, the addition of the new model PA and IF outputs, and also the implementation of a reporting function that was originally developed for another programme. The system website was updated to include the new functionality. This is described in detail in Annex D, and the new displays are also shown in Figure 5-1. In addition to the PA and IF displays providing new anomaly alerting and diagnostic information, the user can select 'reports' to view fleet health status reports created as part of an analysis service, or for expert diagnostic feedback. In these reports alerts are grouped into tracked items, and new report entries appear in red, whilst entries carried forward from a previous report appear in black.

To maximise the value of GE Aviation's analysis service, Bristow provided access to maintenance records for the Super Puma fleet in its web-based IFS (Industrial Financial Systems) maintenance system. This enabled GE Aviation to view maintenance on tracked parts and identify component replacements, faults and rectification actions, correlating these with anomaly alerts. Any gearbox changes were noted, and applied to the AAD system's database as a component change. A check was also made for other relevant maintenance such as sensor changes, and removals/changes of components attached to gearboxes (e.g. engine, oil cooler fan, rotor head, drive shafts), and trends were reset where necessary. The updated arrangement for the second HUMS AAD trial is illustrated in Figure 5-2.

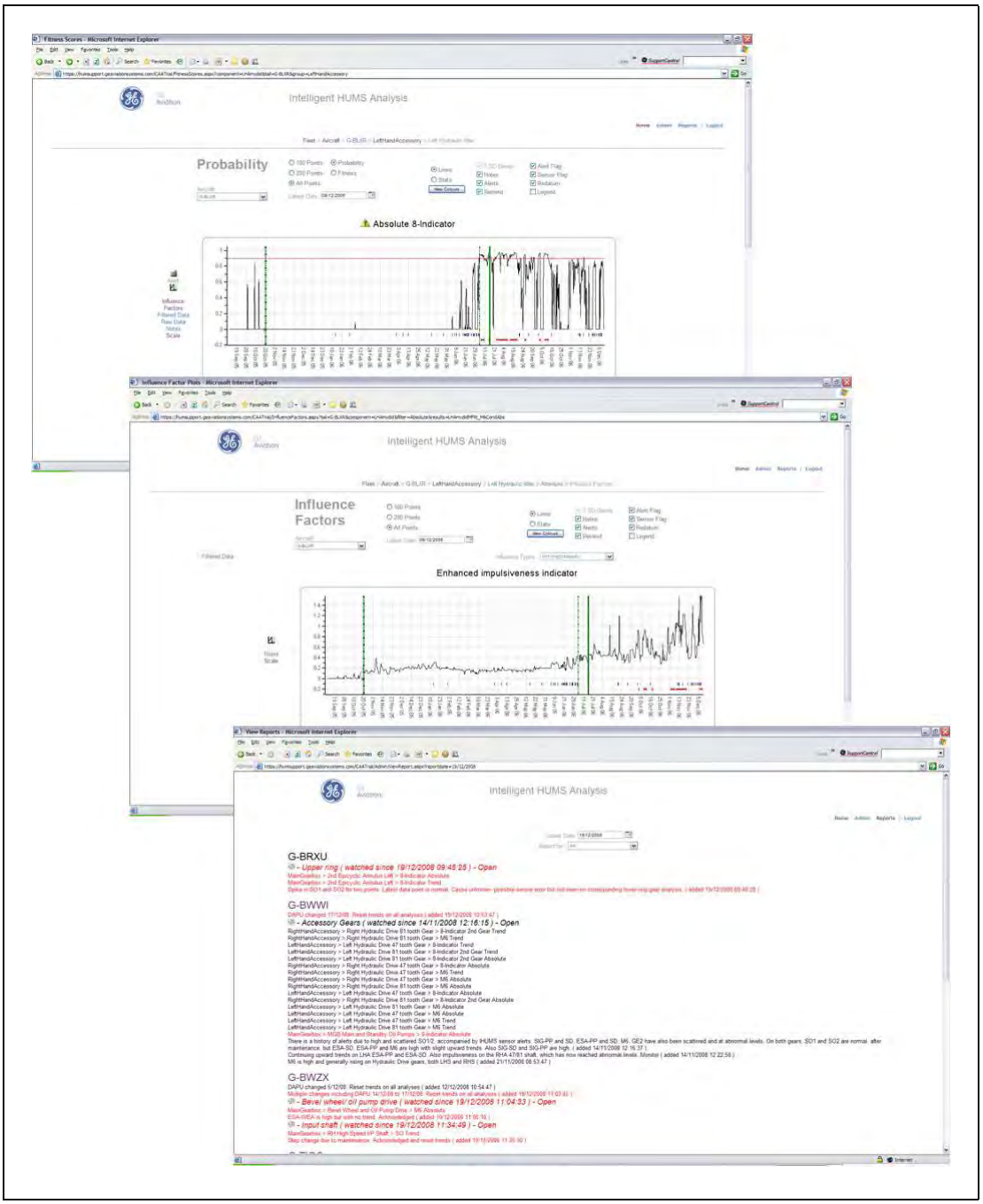

Figure 5-1 New website displays: top – PA display; middle – IF display; bottom – health status report.

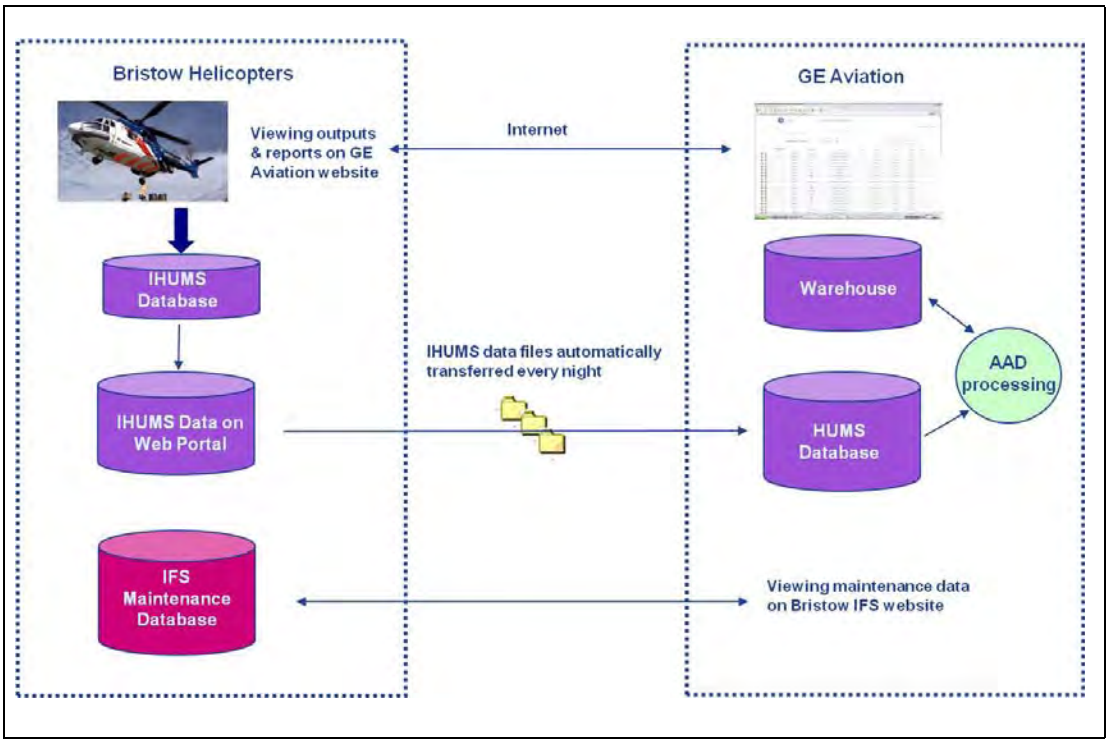

Figure 5-2 Updated HUMS AAD trial arrangement

5.1 Trial Experience

Operational procedures for, and experience from, the second trial period are described in Annex D, which also includes a catalogue of significant findings and a statistical analysis of the anomaly alerts generated by the system.

Again, the reliability of the nightly automated data download from Aberdeen was extremely good, and the updated system website was considered to be easy to use. The analysis service and associated fleet health status reports were effective in reducing Bristow's workload associated with the trial, enabling Bristow's effort to be concentrated on following-up the anomalies that had been flagged by GE Aviation. For a limited period of time during the trial Bristow conducted an evaluation of a new version of the IHUMS ground station software on part of the Super Puma fleet. There were no changes to the files containing VHM data. However, while two versions of the IHUMS software were running concurrently it was necessary to copy two sets of data files to the web portal for downloading by GE Aviation. This required some changes to the data import process. The files were treated as coming from two separate aircraft databases, which were then merged in the data warehouse. GE Aviation also added a capability to automatically detect the movement of aircraft between databases, so that there would be no requirement for further changes as aircraft transitioned between the different versions of the IHUMS ground station software.

5.1.1 Catalogue of Significant Findings

Annex D contains a selection of a further 15 example cases providing additional evidence of the performance of the anomaly detection system to that given in Annex B. Twelve of the cases were from the second trial period, and the remaining three were from the period of operations in 2007 between the two in-service trials when Bristow continued to use the system while GE Aviation carried out the additional research tasks.

To provide consistency with the previous trial the example cases are described under the same category headings as used in Section 3.1.1. This time nine of the 15 cases were placed in Category 2. However, in most of these cases the AAD system gave an earlier and clearer indication of anomalous IHUMS CI trends, often flagging these up on more components than did the IHUMS.

1) Cases where AAD identified a fault not seen by the existing HUMS

The single case in this category related to the detection of a potential fault on the oil cooler fan drive of a LHA gearbox. The '8IA' model was in continuous alert from March 2007, with the '8IT' and 'M6A' models going into almost continuous alert from late January 2008. Anomalies were also detected on other LHA shafts monitored using sensor 3, and on the RHA gearbox (monitored from sensor 4). Although there were no IHUMS alerts, data levels were anomalous for several CIs and there were clear rising trends in ESA_SD and ESA_PP, coupled with an increase in ESA_M6, on the LHA oil cooler fan drive from January 2008. Sensor maintenance was performed on 4 February 2008; however, this had no effect on the rising trends. Although the IHUMS had not alerted, the LHA gearbox was rejected on 13 March 2008 and the data levels returned to normal. Bristow sent the IHUMS data to Eurocopter together with the LHA gearbox and requested a strip report, however by the end of the trial no feedback had been received.

2) Cases where AAD corroborated existing HUMS indications

Three further cases of the detection of potential LHA gearbox faults were included in this category. The left hydraulic idler in a LHA gearbox consistently generated 'M6A' models alerts from June 2007. The IHUMS also generated multiple alerts, and Bristow placed the gearbox on close monitor. The alerts were triggered by a rising trend in ESA_M6 values, followed by a period of increased variability. The gearbox was removed together with other drive system components when aircraft left the European fleet in March 2008.

The left hydraulic drive 47/81T gears in a LHA gearbox consistently triggered '8IA', '8IT' and 'M6A' model alerts from October 2007. These alerts were due to initial clear rising trends in ESA_M6 and ESA_PP, which then stabilised at an increased level. There were no IHUMS alerts at this time. The CI trends started increasing again in October 2008, finally triggering two IHUMS CI alerts. The rising CI trends were also mirrored on the RHA gearbox, causing the '8IA' anomaly model for the right hydraulic drive 47/81T gears to generate continuous alerts from May 2008, with all the remaining models alerting from October 2008. However, no IHUMS CI alerts were triggered on the right hydraulic drive. The MGB (together with both accessories) was removed in December 2008 because of metal contamination. Bristow requested a strip report from Eurocopter to identify the source of the debris, however no report had been received by the end of the trial. All CI levels returned to normal after the MGB change. Although the IHUMS did generate some alerts, the alerts on the left hydraulic drive occurred much later than the anomaly model alerts, and no alerts were triggered on the right hydraulic drive. Therefore, the anomaly models gave an earlier and clearer indication of the abnormal CI trends on these components.

The final LHA example involved multiple gearbox replacements. The anomaly models suppressed initial IHUMS alerts caused by a short duration data spike that resulted in the first rejection of the LHA gearbox in April 2007. As the data returned to normal before the rejection, this gearbox might have continued operating satisfactorily until a scheduled MGB replacement in June 2007. However, the anomaly models corroborated the IHUMS alerts occurring on the new LHA gearbox, resulting in a second gearbox replacement in August 2007. Both the anomaly models and the IHUMS also then flagged a subsequent sensor problem on the gearbox.

The new 'shaft order' models corroborated IHUMS alerts on two different tail rotors. In the first case the alerts cleared after maintenance on the spindles and flapping hinge bearings, and in the second case the alerts cleared after maintenance to replace worn tail rotor pitch link bearings. One of these examples represented a relatively complex maintenance scenario, with both tail rotor faults and an IHUMS instrumentation fault (the shorting-out of a sensor cable) present, and the latter causing some nugatory tail rotor maintenance. The aircraft had transitioned to Nigeria before the appearance of the instrumentation problem. Greater awareness of the abnormal characteristics of the associated FSA_SO1 trend (i.e. highly variable, with many exceptionally high values) might have enabled more rapid diagnosis of this problem.

On an MGB RH input shaft, frequent 'SOA' model alerts were triggered by increased levels and variability of FSA_SO1. A short time later '8IA' and 'M6A' model alerts were triggered, caused by increasing trends in ESA_M6 and ESA_PP. There were also many IHUMS alerts. A left and right sensor swap did not alter the variable behaviour of FSA_SO1. When the engine and Bendix shaft were changed during a period of heavy maintenance the inboard bearing mount was found to be worn, causing some movement of the bearing.

Following the replacement of an oil cooler fan duct due to cracking the FSA_SO1 levels increased, triggering both IHUMS and '5IA' anomaly model alerts. The oil cooler fan was replaced and FSA_SO1 returned to normal, clearing the alerts.

The final two cases in this category involved IHUMS instrumentation and hardware faults. Anomaly model alerts occurred on a number of analyses using sensor 7, caused by a sudden and very large increase in the levels and variability of FSA_SO1 and FSA_SO2. There were no IHUMS CI alerts, however an IHUMS sensor fault was recorded a short time later, after which the data returned to normal (presumably as a result of sensor maintenance). This was a typical example of a number of IHUMS instrumentation faults detected during the trial. Although the IHUMS detected a sensor fault, the anomaly models again gave an earlier and clearer indication of the problem. The final example was an IHUMS DAPU fault. Following a DAPU change there were absolute anomaly model alerts on the TGB input and all three 2nd stage epicyclic gears (sun, annulus and planets). These alerts were caused by step increases in multiple CIs. There were also a few IHUMS alerts. A faulty DAPU was suspected, and this was therefore swapped with the DAPU in another aircraft. All the anomaly model alerts moved with the DAPU, which again caused step increases in multiple CIs. The DAPU swap confirmed the fault, therefore this was removed as unserviceable. Again, although the anomaly model alerts were corroborated by a few IHUMS alerts, the anomaly models gave a clearer and more consistent indication of the anomalous data generated by the faulty DAPU.

3) Cases where AAD failed to identify a fault that was seen by the existing HUMS

In the second trial period no cases were identified of the anomaly modelling failing to identify a fault detected by the IHUMS.

4) Cases where AAD identified an existing HUMS false or premature alert

The only possible case of the anomaly modelling identifying a false or premature IHUMS alert was the first gearbox rejection in the final LHA gearbox example in Category 2.

5) Cases where AAD generated a false or premature alert

One example, involving an MGB RH input shaft, was included in this category. There were consistent '8IA' model alerts on this shaft for a considerable period of time, however there were no IHUMS alerts. Examination of the HUMS CIs showed no levels that were abnormal compared to the rest of the fleet. Similarly, examination of the standard IFs presented to the operator did not identify any CIs that were driving the anomaly indication. GE Aviation then reviewed two other variants of IF that had been implemented primarily to provide additional information on anomaly model behaviour. The IF providing information on the correlation between CIs did show increased levels for SIG_SD and FSA_MS, indicating that an abnormal relationship between these two CIs was the primary driver of the anomaly model alerts. Further investigation confirmed an abnormal correlation between the two indicators, with SIG_SD being higher than normal for a given FSA_MS. The reason for this was unknown, but no action was considered to be required.

6) Cases of anomaly indications with an unknown outcome

Four example cases were included in this category, illustrating the difficulties typically involved in interpreting HUMS VHM data. In all cases the AAD correctly detected an abnormal data trend, but it was not possible to determine the cause. Similar situations often arise with traditional VHM analysis, which is why the concept of 'close monitoring' was introduced.

There was a consistent 'M6A' model alert on an MGB LH input shaft. This was due to a step increase in ESA_M6 to abnormal levels compared to the rest of the fleet. The start of the alert period coincided with an engine removal and replacement. The left and right sensors were swapped, but this made no difference to the data. The alert cleared when the engine was next removed, which resulted in a step decrease in ESA_M6. The cause of the increased signal impulsiveness as measured by ESA_M6 in the period between the two engine removals and replacements could not be identified. There was a similar occurrence on an MGB RH input shaft. Around the time of an engine replacement there was an upward trend in ESA_M6, triggering 'M6A' model alerts. ESA_M6 values were high compared to the rest of the fleet but the trend stabilised and then started reducing, eventually returning to normal levels. Again, the cause of this behaviour could not be identified.

Following the installation of a new MGB, the '8IA' and 'M6A' anomaly models for the bevel pinion and combiner gear went into immediate alert, caused by abnormally high values of ESA_PP, ESA_SD and ESA_M6, although ESA_M6 values gradually declined over the following months. The reason for this could not be determined, but it was speculated that it could possibly be due to a small gear tooth profile or pitching error.

The final example involved an MGB bevel wheel and oil pump drive, where repeated 'M6A' anomaly model alerts were generated from the start of monitoring. The cause of these alerts was abnormally high values of the derived CI WEA (which is the ratio of ESA_SD to SIG_SD), as a result of an abnormally low SIG_SD. However, the most notable feature of the data was the very low amplitude of the bevel wheel mesh tone, FSA_MS. The abnormal data did not trigger any IHUMS alerts. The instrumentation was checked and no problems could be found. Then, over approximately a one-month period, the amplitude of the bevel wheel mesh tone progressively increased, until it was within the bounds of normality for the fleet. The anomaly model alerts cleared at that time. No explanation could be found for the initial abnormally low gear mesh tone, or why it then progressively increased.

5.1.2 **Statistical Analysis of AAD System Performance**

Another statistical analysis was performed of the alerts generated during the second trial period to assess the performance of the AAD system. The statistics presented apply to the period from 7th January to 19 December 2008. For comparison purposes, some of the statistics presented in Section 3.1.2 for the first six-month trial period are also repeated here.

As part of the assessment of AAD system performance, the alerts generated by this system were again compared to those generated by the IHUMS, with the IHUMS alert figures combining both sensor and CI alerts. The IHUMS alert information was obtained from the CI data files extracted by Bristow from the IHUMS ground station. However, it should be noted that a new 'm out of n' alert filtering process had been implemented in the IHUMS ground station between the two trial periods. Therefore, the IHUMS alert figures for the second trial period represent the 'raw' alert data only, not what may be presented to the IHUMS operator. (It is understood that operators have a choice as to whether or not to use the new alert filtering process.)

While the second trial was in progress, GE Aviation undertook another project for the UK Ministry of Defence (MoD), configuring and trialling a HUMS AAD system for the UK fleet of Chinook helicopters. Based on the CAA trial experience, it was considered that an alert threshold set at a PA value of 0.90 may be too sensitive, and that there should be a higher probability that data is anomalous before an alert is triggered. Following an investigation into the effects of varying the PA threshold on the alert rate and fault detection performance, a PA threshold of 0.98 was selected for the Chinook trial. The number of findings during a three-month trial confirmed that the system is still an effective detector of anomalous HUMS CI data using this PA threshold. Therefore, again for comparison purposes, statistics were also calculated for the number of alerts that would have been generated in the second CAA trial period if the PA threshold had been set at 0.98 instead of 0.90.

Table 5-1 presents a summary of the AAD system and IHUMS alerts generated during the two trial periods. AAD system alert statistics are shown for PA thresholds of both 0.90 and 0.98.

Although the first trial period had a duration of six months and the second period a duration of almost one year, there was a decrease in the flying rate of Bristow's AS332L fleet between the two periods as new aircraft were introduced into North Sea operations, and a few of the AS332Ls were transferred overseas. The total flying hours for the AS332L fleet included in the anomaly detection trial were 9,951 hours for the first trial period and 11,096 hours for the second period. Therefore, the most meaningful comparison of the alert rate between the two trial periods is on the basis of alerts per flying hour, rather than alerts per day.

For the first trial period the AAD system alert thresholds were set on the FS values output from the anomaly models, whereas for the second trial the threshold was based on the PA output, with a global threshold at a PA of 0.90 being applied. This PA threshold value was chosen to give a similar sensitivity to the FS thresholds used in the first trial period. The 'IHUMS' data presented in Table 5-1 are the 'raw' alert statistics referred to above. The 'Anomaly FS / PA 0.90' data shows the alerts generated by the AAD system based on FS thresholds in the first trial, and the 0.90 PA threshold in the second trial. The 'Anomaly PA 0.98' data is for the second trial period only, and shows the number of alerts that would be generated if a 0.98 PA threshold had been applied.

Table 5-1 Alert summary

Alert Statistics	First Trial Period	Second Trial Period
Total Alerts		
IHUMS (32 analyses)	1381	1749
Anomaly FS / PA 0.90 (35 analyses)	1484	1509
Anomaly PA 0.98 (35 analyses)		587
Alerts per Analysis		
IHUMS	43.2	54.7
Anomaly FS / PA 0.90	42.4	43.1
Anomaly PA 0.98		16.8
Alerts per Flying Hour		
IHUMS (32 analyses)	0.139	0.158
Anomaly FS / PA 0.90 (35 analyses)	0.149	0.136
Anomaly PA 0.98 (35 analyses)		0.053

As previously discussed, for the first trial period the alert rate of the AAD system was comparable to that of the IHUMS. In both cases the total number of alerts increased in the second trial period, however this was of longer duration, and involved a greater number of flying hours. When the data was factored by the total flying hours, the IHUMS still generated more 'raw' alerts in the second trial period than the first. However, where the new 'm out of n' alert filtering is applied, the IHUMS operator would see fewer alerts than the numbers shown.

The model rebuilding performed between the two trial periods resulted in approximately a 10% reduction in the alert rate (per flying hour) of the AAD system. The figure for the second trial period was also approximately 15% lower than the 'raw' IHUMS alert rate, which actually increased between the trial periods. Table 5-1 shows that increasing the threshold to a PA value of 0.98 would have a significant affect on the alert rate, reducing this by over 60%. With a PA threshold of 0.98, 14 of the 15 example cases described in Section 5.1.1 would still trigger anomaly alerts. The only one that would not now trigger an alert is the oil cooler example in Category 2, where there was a short period of increased FSA_SO1 values following oil cooler maintenance.

Again, an analysis was performed of the number of AAD alerts by sensor, analysis, model type and aircraft. This highlighted differences in the pattern of the alert data between sensors and analyses from the first trial period, suggesting that the relative analysis alert rates were influenced by particular aircraft issues that were present during the period of monitoring. The analysis by model type showed that the 'M6A' models generated the greatest number of alerts, followed by the '8IA' models. This finding was somewhat different than that from the first six-month trial period, where the '8IA' and '8IT' models generated the most alerts. Although the anomaly models were rebuilt between the two trial periods using an enhanced modelling technique, the difference was again believed to be primarily due to the changing issues seen on the aircraft in the two periods of operation.

5.1.3 **Summary**

The second trial period provided further evidence demonstrating the effectiveness of the HUMS AAD system. The system clearly highlighted anomalous HUMS CI trends that are more difficult to detect by traditional HUMS analysis, continuing to out-perform the traditional analysis in successfully highlighting both aircraft problems and HUMS instrumentation faults. The new PA and IF outputs were very useful enhancements, providing clear information to assist the operator in identifying abnormal HUMS CI trends and diagnosing their possible causes.

Bristow was not able to obtain any strip reports on components rejected for anomalous HUMS CI trends or metal contamination during the two trial periods. The lack of feedback information on the condition of drive system components removed from aircraft can prevent any meaningful interpretation of CI trends in terms of component condition. This type of feedback is critical to the on-going development of HUMS data analysis capabilities.

6 Additional Research Work (Tasks 5 and 6)

This section presents a summary of results from two final additional research tasks, demonstrating new system concepts and capabilities for a second level of advanced data analysis, and applying data mining and automated reasoning technologies to the outputs from the AAD system. The project report containing the details of the results of the final research tasks was issued in August 2009, and its contents are reproduced in Annex E.

Because of the relative immaturity of the application of automated data mining and reasoning to VHM data, and the magnitude of the task to research, develop, test and implement new automated data mining and reasoning functions, the outputs of tasks 5 and 6 comprised conceptual demonstrations only. Two key objectives of the work were to demonstrate potential future new capabilities to:

- further enhance the effectiveness of the HUMS AAD system by providing new information to aid the operator's data interpretation and maintenance decision-making processes;
- reduce the operator's workload involved in using the system and responding to anomaly alerts.

6.1 Concepts for New Information Displays

Annex E contains mock-ups of possible new AAD system alert displays to illustrate the new capabilities and output information that the data mining and reasoning technologies are aimed at providing, and how these might be presented to the operator. It was not an objective of this work to design new system displays. The mocked-up displays were purely intended to help communicate the type of new information that could be generated, and to prompt discussion on how this could be presented. For example, some operators may only want a simple output, however a knowledgeable operator may want a comprehensive display bringing together multiple pieces of information. Using the analogy of a cockpit instrument display, this would look bewildering to a passenger, but a pilot can quickly scan the displays and understand the status of the aircraft. The key point is that the anomaly models can provide a lot of information on alerts, and it is clearly beneficial to extract this and make it available to the operator.

A mocked-up new alert information display is shown in Figure 6-1. The intent is to summarise all the information that the system could provide on an alert within one display. This would remove the need for an operator to manually create a picture of an alert by viewing many chart displays, and would also provide more alert information than the operator could obtain from reviewing the existing system data. A key piece of new information is an overall alert significance assessment, intended to ensure that an operator pays close attention to an alert that the system considers has high significance. This assessment fuses the outputs from a ‘data mining significance’ assessment and the automated trend classifier. The ‘data mining significance’ is based on a combination of the rarity of the identified patterns (a significant aircraft component fault is a rare occurrence) and the magnitude of the data change triggering an alert. The automated trend classifier determines the extent to which the data trend fits classifications of 'clear trend', 'noisy' and 'step'. Other information in the display includes 'recognised patterns', indicating the probability that the alert was triggered by a non-benign maintenance action, a HUMS instrumentation problem, or a previously documented fault type. Finally, a 'diagnostic summary' table is shown, summarising the information provided by the sets of IFs calculated for each anomaly model in the form of the significance of six different features of the vibration data.

The figure shows how the display would appear for the cracked MGB bevel pinion fault case. Both the absolute and trend models are in alert, a clear trend is detected, multiple features of the vibration signal are changing, and the data patterns are very unusual. Therefore, the system gives this case a high overall significance assessment. The data patterns do not match any documented fault cases, indicating that these have not been seen before. All this information should prompt an operator to pay very close attention to the alerts, and take appropriate maintenance actions.

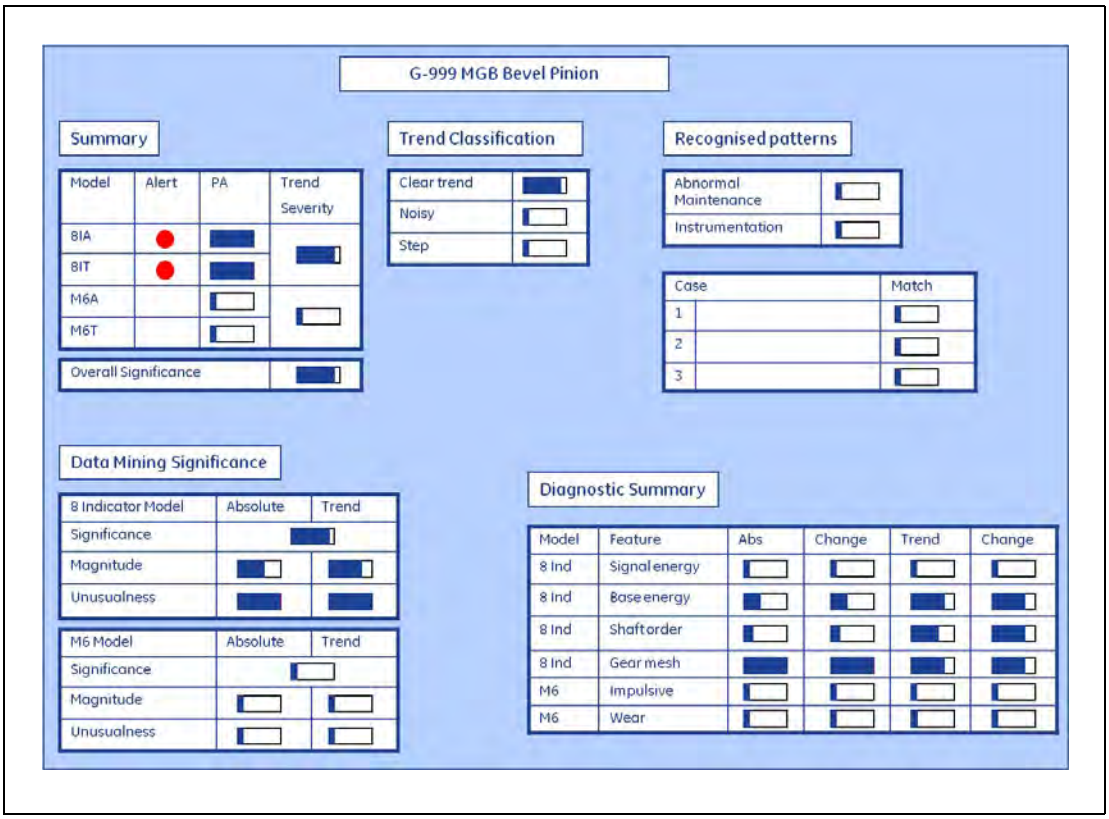

Figure 6-1 Alert information display for cracked MGB bevel pinion

6.2 Data Mining Demonstration

The data mining effort focussed on automated change detection, trend analysis, anomaly significance assessment, and data analysis to support Case-Based Reasoning (CBR).

6.2.1 Change Detection

The purpose of change detection is to reduce workload associated with the AAD system by suppressing repeat alerts triggered by small changes in data that are close to the threshold, while flagging up significant changes in data that are constantly exceeding a threshold (particularly where an alert may have previously been acknowledged) indicating that a fresh review is required. A change detection algorithm was developed for analysing the '8IA' anomaly model IF values obtained from data points in alert. The eight IFs were combined into a normalised 4-dimensional vector with components based on natural pairings of the IFs calculated for the HUMS CIs. The vector angle between each new alert data point and the last acknowledged alert point was calculated. The data for the new alert point was then determined to show no change if both the change in vector angle and the change in magnitude of the FS remained below pre-set thresholds. Using the historical database of in-service trial results, the change detection algorithm was applied to demonstrate both repeat alert suppression and the automated detection of a change in the characteristics of data already in alert.

6.2.2 Trend Analysis

Trend information is critical to determining the nature and significance of an anomaly alert, therefore further investigations were carried out into methods for automated trend detection, classification and severity assessment.

An investigation was carried out into the application of a histogram technique to FS values to extract short-term trend features. The basic approach used was to histogram the data and then detect patterns of transition between different bins of the histogram. Logic-based rules were applied to the histogram to characterise the behaviour of the data as it transitioned between bins. The results obtained showed that the histogram technique could successfully be used to detect and classify short-term trends. However, while the technique was effective over short periods, variability in the data (i.e. signal noise) made it more difficult to implement over longer periods.

Further work was undertaken to develop a trend algorithm more suited to the analysis of longer-term trends. This algorithm first detects the presence of a trend in anomaly model FS or HUMS CI values, and determines whether this is positive or negative. It then assesses the severity and duration of the trend. The algorithm typically identifies multiple trends (occurring at different periods of time) in the complete time history of the FS or CI values for a monitored component. The algorithm design priorities were the quick detection of developing trends (i.e. with a minimum of lag) and good resilience to noisy, non-trending signals to prevent the generation of false alerts.

The trend detection algorithm was applied to FS data from the MGB 2nd stage epicyclic annulus for all 'gearbox fits' in the database, with the gearboxes being ranked according to trend severity. The most significant FS trend identified occurred before the start of any of the in-service trials in September 2003 (Figure 6-2). There was a rapid and very large decreasing trend in the FS values, caused by a rapid increase in SIG_SD. The reason for the trend is unknown, however it is clearly significant. The example highlights the benefits of tools such as automated trend detection and severity assessment in identifying significant trends in large volumes of data.

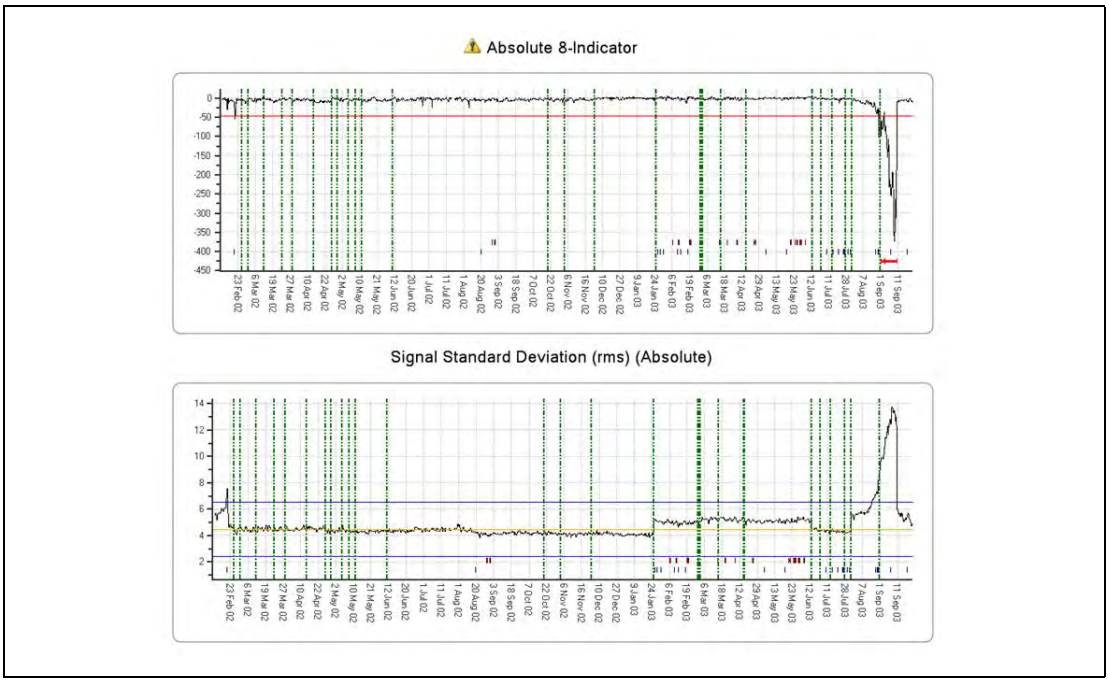

Figure 6-2 MGB 2nd epicyclic annulus aft (RH): top – '8IA' model FS; bottom – SIG_SD.

6.2.3 Anomaly Significance Assessment

A potentially very useful data mining capability would be an automatic assessment of the significance of a detected anomaly on the basis of a combination of a severity assessment and a measure of uniqueness or rarity. The measure of rarity may be of most significance as instrumentation faults can generate highly abnormal data, but are common in comparison to the occurrence of a real aircraft fault. This type of analysis does require a sizeable database of historical data to enable criteria such as rarity to be properly evaluated.

The data modelling performed to demonstrate an example automated anomaly alert significance assessment used the standard IFs (i.e. those presented to the operator) as inputs. The IFs provide information on the HUMS CIs driving an anomaly alert however, unlike the CIs, they are normalised and so enable data from different components to be combined in the modelling process. Different types of data mining model were built and evaluated. Because they had the greatest numbers of CIs as inputs, the work concentrated on the IFs from the '8IA' and '8IT' anomaly models. The cluster statistics used in the automated significance assessment included a distance measure indicating how far a cluster is from the centre of the modelling space, and a distribution measure indicating the mix of 'gearbox fits' associated with the cluster. The distance gives a measure of the severity of an anomaly; the distribution indicates how common a particular pattern in the data is (instrumentation problems are relatively common, whereas a genuine component fault can be a rare event). The results of the anomaly significance assessment showed that this can flag up the most significant anomalies, ensuring that these receive very close attention from the operator.

6.2.4 Data Analysis to Support Case-Based Reasoning

An investigation was performed into data analysis to generate information for CBR. Following an anomaly alert, the objective of CBR is to indicate whether there have been previous cases for which data behaviour was similar to the current case. The objective of this element of the data mining task was to investigate techniques for identifying similar cases in the database. Again, utilising the IFs output by the '8IA' models, different types of data mining model were investigated. Example results for the 2nd epicyclic annulus aft identified 'gearbox fits' with similar low frequency noise characteristics due to sensor faults, and also abnormally high signal levels due to suspect wiring harness problems.

The data analysis for case-based reasoning successfully identified similar cases of anomaly alerts that could be used to assist the diagnosis of different types of HUMS instrumentation problem. However, the work showed some limitations with the approach, where the results obtained were dependent on how the modelling was applied (see Annex E for more information).

6.3 Automated Reasoning Demonstration

The purpose of this task was to demonstrate an automated reasoning capability, performing secondary processing of anomaly detection alerts to help the operator determine the most appropriate action. The reasoning technology was primarily based on 'Bayesian' probabilistic networks, which provide a widely recognised capability for reasoning in conditions of uncertainty (i.e. with limited or possibly conflicting information). Once the anomaly detector has flagged abnormal data, reasoning networks can then fuse multiple features of the anomalous data, together with other information, to automatically indicate the anomaly's significance and possible cause.

A range of potential reasoning inputs are available. These include outputs from the anomaly detection processing such as anomaly model alerts, PA, FS and IF values. They can also include outputs from the data mining tasks, such as measures indicating the significance of alert data, identifying significant trends, and identifying similar cases. Other inputs are aircraft configuration data and existing diagnostic knowledge. The output of the reasoning process should be a summary of key information that can assist an operator to determine the appropriate response to an anomaly alert.

GE Aviation's reasoning tool was used to configure demonstration-reasoning networks, with real data from the CAA in-service trial being used to test them. The task investigated a number of issues, such as: what reasoning capability is required (e.g. what are the desired outputs); how best to address the problem; what inputs to use; how best to design and construct reasoning networks to ensure robustness and supportability; and how would a reasoning capability perform in practice. The reasoning task must carefully consider what type of information should be presented to an operator, as it can be dangerous to provide information that could lead the operator to an incorrect conclusion. To avoid creating large, complex and difficult to manage reasoning networks, the reasoning approach adopted was based on a concept of 'multiple independent experts'. These are independent targeted networks looking for either evidence of a particular potential cause of patterns identified in the HUMS data (e.g. a maintenance action or a HUMS instrumentation problem), or providing a particular piece of information (e.g. the significance of an anomaly alert).

The reasoning demonstration focussed on data fusion and fault diagnosis. The data fusion utilised physical relationships between HUMS sensors and component analyses, generated diagnostic summary information, and provided enhanced anomaly significance assessment capabilities. The fault diagnosis element demonstrated an ability to automatically identify explainable patterns due to maintenance actions and instrumentation faults, and explored the application of CBR. Although not demonstrated on the CAA programme, GE Aviation's probabilistic reasoning models are compatible with the anomaly models and, therefore, anomaly model predictions can be directly used by the reasoning networks. This means that sophisticated reasoning models can be constructed from much simpler and well understood sub-models, relating to different component anomaly models.

6.3.1 Data Fusion by Physical Modelling in Reasoning Networks

Reasoning networks provide an effective way of fusing items of HUMS data from different sensors, or from different components monitored by the same sensor, to increase the confidence in an output. This is a relatively simple form of data fusion, where a reasoning network models the physical relationship between HUMS sensors and monitored components. While of some value, the investigation performed showed that modelling would need to be aircraft and HUM system specific, and that nearly all components are monitored using a single sensor. It was therefore concluded that there is limited scope for the use of physical modelling in a generic reasoning process. However, the in-service trial experience did highlight areas where this type of automated data fusion could be applied, for example in diagnosing the many sensor faults causing anomaly alerts, where the fault affected the results from multiple component analyses performed from that sensor.

6.3.2 Diagnostic Summary Information

The IFs output from the anomaly models provide diagnostic information on the HUMS CIs that are driving an anomaly alert. With two variants of three different types of anomaly model being used in the CAA trial, a total of 24 different IFs were generated, being displayed in 24 different charts. To enhance system usability, there is a desire

to minimise the need for operators to view and interpret multiple charts when investigating an anomaly alert. A diagnostic network was created, using the anomaly model IFs as inputs, to generate diagnostic summary information for an alert in one report table. This simplifies the operator's data interpretation task, provides more meaningful descriptive information, and could be used in the future to drive specific diagnostic outputs. The network converted the IFs / HUMS CIs into more meaningful descriptors and, where two IFs provide related diagnostic information, combined these into a single diagnostic descriptor. Therefore, for the '8-Indicator' model, the eight IFs were converted into the four descriptors 'signal energy', 'base energy, 'shaft order' and 'gear mesh'.

6.3.3 Data Fusion for Anomaly Significance Assessment

A data mining task to determine the significance of anomalies was described in Section 6.2.3, with separate measures being created for IF data from the '8IA' and '8IT' models. A network was created to demonstrate that reasoning networks provide an effective way of fusing different items of information from the two anomaly models, plus additional trend information, to produce an overall significance assessment. The network created fused all the results of the data modelling (i.e. the cluster distance and distribution measures from the absolute and trend models), and then fused these with the results of the trend analysis (trend severity assessment and step change detection). Known fault cases were used to demonstrate the performance of the reasoning network. For example, the cracked MGB bevel pinion generated a high anomaly significance value, with the key factors driving this result being alerts from both the absolute and trend anomaly models, the clear FS trends, and the highly unusual data characteristics.

6.3.4 Identifying Explainable Patterns

The previous reasoning networks are examples of data fusion. The next step in the reasoning process should be automated fault diagnosis. However, aircraft faults can be rare events and even if two similar faults do occur, there can be differences in their specific characteristics. Most anomaly alerts may be caused by maintenance actions or HUMS instrumentation problems. Therefore, the first step in fault diagnosis should be to identify explainable patterns in the HUMS data due to maintenance or an instrumentation fault. Two example reasoning networks were developed to demonstrate a reasoning capability to meet this requirement.

The first network identified anomaly alerts triggered by step changes in the HUMS CI data, and assessed whether these could have been caused by normal maintenance, incorrect (or abnormal) maintenance, or another cause. If an alert is not caused by a step change in the data, it is unlikely to be due to a maintenance action. If the data after a step change is still within normal bounds, then the trend pre-processing should be re-set, but no further action is required. However, if the data after the step change is now abnormal, then incorrect maintenance may have been performed and further rectification maintenance could be required. Fault case examples were used to demonstrate the network, including the installation of a faulty IHUMS DAPU on an aircraft, after which abnormal maintenance was diagnosed. When this was replaced with a functioning DAPU normal maintenance was then diagnosed, requiring no further action other than resetting the trend pre-processing.

The second network detected evidence of a HUMS instrumentation problem. Part of the instrumentation fault network is shown in Figure 6-3.

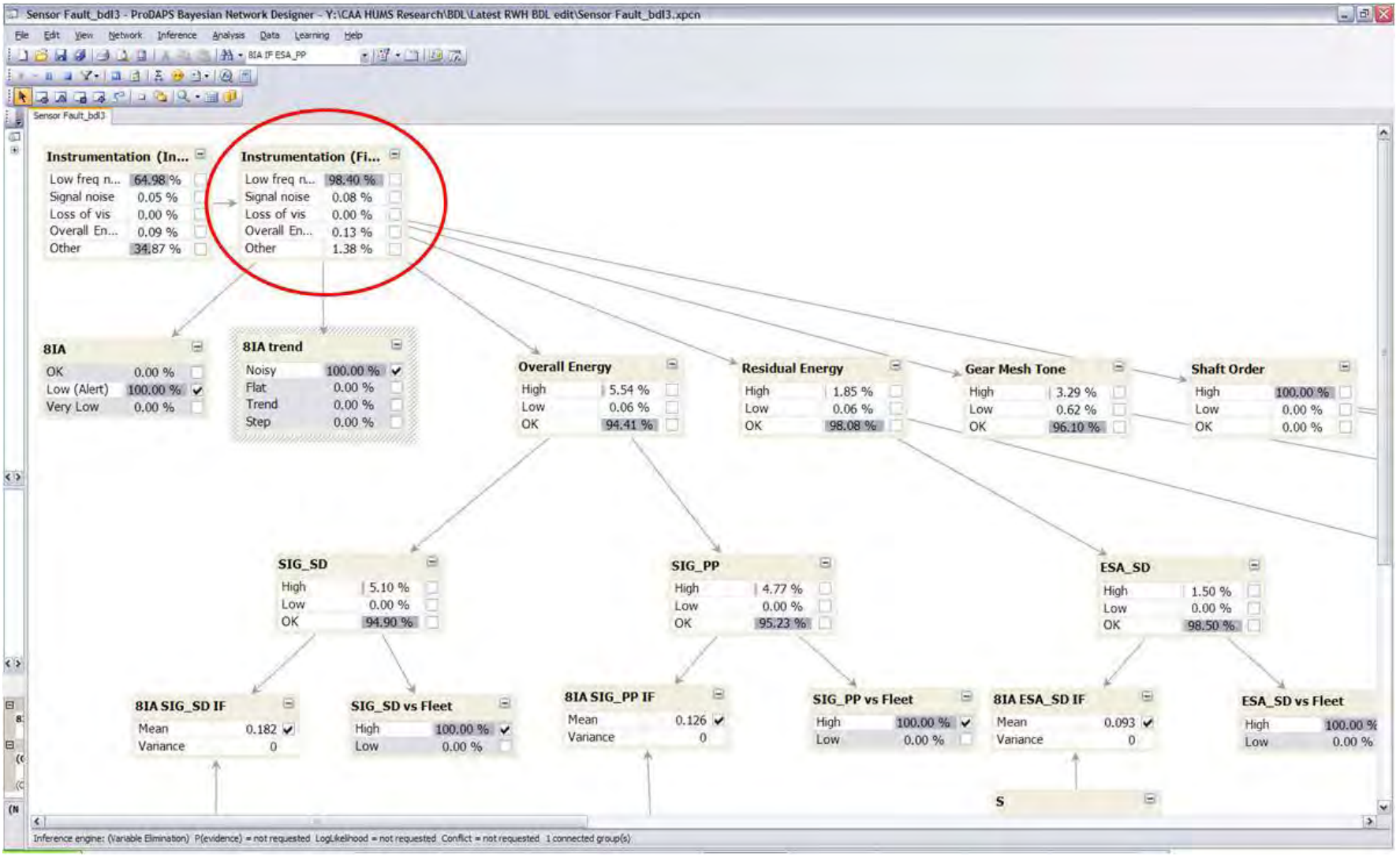

Figure 6-3 Instrumentation fault network

The trial experience showed that different types of instrumentation fault could generate different signal characteristics, such as increases in low frequency noise, high base energy (i.e. general signal noise), high overall signal energy, and loss of visibility of gear mesh frequency vibration. The network therefore identifies the presence of the patterns associated with these four types of instrumentation fault. Figure 6-3 shows the response of the network to data from a left hydraulic drive gear on a LHA gearbox with an accelerometer problem. This caused a noisy FS trend and high FSA_SO1 values. The network has diagnosed an instrumentation problem with the characteristic of low frequency noise.

6.3.5 Case-Based Reasoning

To complete the task, a brief review of CBR was performed. CBR is the concept of using past experience to solve a new problem, and is applicable to many situations such as medical diagnosis, legal proceedings, etc. It does not denote a single algorithm or even class of algorithms to solve a problem but rather a framework for structuring a reasoning system that emphasises the role of previous cases for solving a new problem. Each case describes a problem, its solution if available and outcome. A CBR tool will execute four main processes: retrieve, re-use, revise and retain. Retrieve searches for a stored case that is relevant to the current situation; re-use applies the knowledge of the retrieved case to the new case; revise will test the solution and adapt it if necessary; retain updates the library of cases. Each application domain presents specific challenges and, for this reason, it is difficult to source an off-the-shelf approach that can be directly used to model a specific application.

CBR would be an extension of the reasoning to identify explainable patterns related to maintenance actions and instrumentation faults, and these instrumentation faults could actually be included as cases in a CBR system. If the characteristics of a new anomaly do not match the explainable patterns or any of the CBR cases, the automated anomaly significance assessment capability can still provide useful information to determine the most appropriate response.

6.4 Summary

The AAD system represents the first level of a new advanced HUMS data analysis capability. The normalised PA and IF outputs can be further processed in a second level of advanced data analysis, utilising data mining and automated reasoning tools. This second level of analysis provides the greatest scope for further improving the performance and usability of HUMS, giving valuable additional information to assist the operator in determining the most appropriate response to an anomaly alert. Additional research tasks 5 and 6 involved data mining and reasoning technology demonstrations exploring the approach to, and potential benefits of, a second level of advanced data analysis. The work demonstrated new system concepts and capabilities, and the technologies that could provide them. Further work is required to develop these technologies to a readiness level suitable for implementing in an operational system.

7 Conclusions and Recommendations

This section presents some overall conclusions and recommendations from the Civil Aviation Authority (CAA) Health and Usage Monitoring System (HUMS) research programme, and also provides a definition of the Advanced Anomaly Detection (AAD) capability that has been developed in support of this programme.

7.1 Conclusions

A new HUMS AAD system has been developed, incorporating all of the capabilities defined in Section 7.3. The anomaly detection method is based on density estimation, and incorporates a number of novel features to ensure that the technique is robust, and addresses all the practical issues related to working with in-service HUMS Vibration Health Monitoring (VHM) data. For example, anomaly models are automatically adapted to remove the effect of unknown outliers in the data on which they are built, there is no requirement to split data into training, test and validation sets, and models can be periodically updated as in-service experience accumulates.

The AAD system provides normalised outputs that are easy for an operator to interpret, and that simplify alert threshold setting. Each anomaly model outputs a 'Probability of Anomaly' (PA) value, which is a normalised probability measure that ranges between 0 and 1, and a global alert threshold can be set for all models using this measure. A set of 'Influence Factors' (IFs) is also generated to provide model diagnostic information, with an IF value being calculated for each input HUMS Condition Indicator (CI). Unlike plots of CIs, IF traces assess the contribution of individual CIs to a statistical measure of abnormality derived from the model. IFs are also normalised and can be directly compared.

The AAD system successfully detected anomalous HUMS CI data trends associated with a cracked Main rotor Gearbox (MGB) bevel pinion fault case that was missed by the current HUMS. The results from two in-service trial periods also confirmed that AAD represents a significant advance in HUMS data analysis, resulting in improved fault detection performance and increased system effectiveness. The AAD system out-performed the traditional HUMS analysis in successfully highlighting both aircraft problems and HUMS instrumentation faults. It also gave a clearer picture of anomalous data characteristics on particular aircraft and drive system components than is possible with the traditional HUMS analysis.

A centralised web-based system was implemented for the trials, with a server located at GE Aviation in Southampton, and AS332L HUMS data automatically downloaded every night from Bristow Helicopters in Aberdeen. This was demonstrated to be a reliable and effective solution, and is a proven model for the wider implementation of the HUMS AAD capability.

The normalised PA and IF outputs from the AAD system can be further processed in a second level of advanced data analysis, utilising data mining and automated reasoning tools. In an extension to the research work, a technology demonstration exercise was undertaken to show how this second level of analysis can further improve the performance and usability of HUMS, giving valuable additional information to assist the operator in determining the most appropriate response to an anomaly alert.

During the two trial periods it was not possible to obtain any strip reports on gearboxes rejected for anomalous HUMS CI trends or metal contamination. The lack of feedback information on the condition of drive system components removed from aircraft often prevents any meaningful interpretation of CI trends in terms of component condition. This type of feedback is critical to the on-going development of

HUMS data analysis capabilities, and the absence of such information seriously hinders this development.

7.2 Recommendations

A plan should be developed for the implementation of a HUMS AAD capability on all helicopter fleets involved in UK Continental Shelf offshore support activities. While it is recognised that there may be a few aircraft type and HUMS combinations for which the numbers are too small to make implementation feasible, the plan should cover as many aircraft fleets as possible. Supporting AAD operational requirements should also be defined, addressing issues such as ensuring the effective control of anomaly model development and updating, protecting sensitive data, and the validation of performance through a Controlled Service Introduction.

For the further development of HUMS anomaly detection and VHM capabilities in general, there should be an effective mechanism to enable operators to request feedback on the condition of rejected components. To facilitate this, helicopter Original Equipment Manufacturers (OEMS) and overhaul agencies should have a reliable process for recording component condition information when gearboxes are stripped for overhaul.

Building on the results of the technology demonstrations described in this report, further work should be undertaken to research and develop a second level of advanced data analysis, applying data mining and automated reasoning technologies to the current AAD system outputs. As the primary objective of the second level of advanced data analysis is to provide further assistance to operators, it is recommended that operator feedback is obtained on the contents of this report before additional research and development work is undertaken.

7.3 Definition of AAD for HUMS VHM Data

7.3.1 Full AAD Definition

AAD highlights rotor drive system components whose behaviour is considered abnormal when compared to a population of components. Detection of abnormal behaviour is achieved by measuring a component's behaviour relative to a population model of expected behaviour. This model is learnt from an historical database of training cases. Unknown anomalies may be present in the training data but the model learning method must still be capable of deriving a model that would be sensitive to these training anomalies. Behaviour will usually be monitored by modelling multiple input variables from one or more sensors. The model of normal behaviour will characterise the statistical dependencies between input variables, and model inference should allow predictions where it may be necessary to treat one or more variables as missing in a situation where a variable's value is either unbelievable or the monitoring system has failed to derive a value. Anomaly alerts should be triggered by a model threshold that can be configured via an alert parameter to adjust sensitivity, thus allowing control over the true negative/false positive alert rate. Alert thresholds should be 'data driven' and not biased by a prior belief of a percentage failure rate. The alert threshold should be derived using the full density distribution of the model and not assume a simple distribution such as a uni-modal Gaussian, which often only provides an approximation to the true statistical nature of sensor data. There may be multiple anomaly models for a particular component, and models must be capable of representing different behaviour features, relating to both absolute data values and data trends. Models may also be built from subsets of sensor derived variables where the intention is to emphasise a pattern that might otherwise be masked if modelled using all available variables. While AAD is designed to draw attention to abnormal behaviour, it must also be capable of providing diagnostic information on the input variables driving any anomaly indications.

7.3.2 **Abbreviated AAD Definition**

Rotorcraft HUMS Anomaly Detection is an approach that detects abnormalities in rotor drive system components by comparison of multiple downloaded health monitoring parameters with prepared multi-parameter models of normality for these components. It also provides diagnostic information on the monitoring parameters causing abnormal indications. The multi-parameter models of normality represent the statistical dependencies between monitoring parameters and are based on experience across multiple aircraft within a fleet. The approach incorporates methods to ensure that any unknown abnormalities within this experience do not prevent the detection of similar abnormalities. Models are to be periodically refined based on increasing fleet experience.

INTENTIONALLY LEFT BLANK

ANNEX A

Interim Report on Phase 1 of the Research Project

Based on a report prepared for the CAA by Smiths Aerospace Limited, UK

Table of Contents

List of Figures

INTENTIONALLY LEFT BLANK

List of Tables

INTENTIONALLY LEFT BLANK

Glossary

AGB	Accessory Gearbox
AI	Artificial Intelligence
BHL	Bristow Helicopters Limited
BIC	Bayesian Information Criterion
CAA	Civil Aviation Authority (UK)
CI	Condition Indicator
FS	Fitness Score
GEV	Generalised Extreme Value
GMM	Gaussian Mixture Model
GUI	Graphical User Interface
HUMS	Health and Usage Monitoring System
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
MJAD	MJA Dynamics Limited
MRGB	Main Rotor Gearbox
PCA	Principal Components Analysis
PLATO	Pattern Learning Algorithm Toolkit
PMML	Predictive Model Markup Language
ProDAPS	Probabilistic Diagnostic and Prognostic System
RBF	Radial Basis Function
SOA	Spectrometric Oil Analysis
SQL	Structured Query Language
SVM	Support Vector Machine
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring
WHL	Westland Helicopters Limited

INTENTIONALLY LEFT BLANK

Executive Summary

This is an interim report documenting the results of Phase 1 of a CAA funded research programme to demonstrate the intelligent analysis of helicopter Vibration Health Monitoring (VHM) data. Although existing Health and Usage Monitoring Systems (HUMS) have clearly made a significant contribution to improved safety, there have been examples of defect related trends in VHM data going undetected. This was illustrated by an incident in 2002 involving an AS332L main rotor gearbox that was removed after a gearbox chip warning. On the subsequent inspection, a large crack was found in the bevel pinion. The chip warning occurred purely because a secondary crack fortuitously released a fragment of material. Normally this type of crack would not generate any debris. Although there were some trends in the vibration Condition Indicators (CIs) for the component, the HUMS did not generate any alerts. Existing HUMS struggle to assess the significance of such subtle patterns where no single indicator appears significantly abnormal. Diagnostic inference in existing HUMS tends to be influenced by preconceived ideas of fault related behaviour. The application of a-priori knowledge is important, but this knowledge is limited because faults are rare and physics based models, while useful, are incomplete with regards to in-service systems.

This programme is applying unsupervised machine learning techniques to construct diagnostic models from experiential data. This data driven modelling approach uses historical HUMS data to define models of 'normal behaviour', which can then be used to detect 'abnormal behaviour'. The process is called anomaly detection. Although anomaly detection is not a new discipline, a new approach has been developed here. HUMS data are particularly challenging because the true nature of each case history is not known: it can be assumed that the majority of data represent healthy systems but there will be undocumented instrumentation faults and possibly some undetected component degradation. The anomaly approach adopted constructs density models from historical data for each component and adapts these models so that they reject abnormalities that exist in the training data. Models are built to represent absolute data behaviour (between component variability) and trend behaviour (within component variability). These models are sophisticated statistical representations of the data generated from in-service experience with few assumptions about how the data should behave. They simplify a complex picture by fusing sets of vibration features into a single time history trace called a 'Fitness Score (FS) trace'. The FS measures the degree of abnormality and mirrors the shape of any significant data trends. Unique query facilities allow diagnostic information, such as the CIs with most influence, to be extracted from the models.

The anomaly detector has been implemented as a web-based service and, following a successful off-line demonstration, is currently undergoing an in-service trial in Phase 2 of the CAA programme. The anomaly detector is designed to enhance rather than replace existing HUMS, and there is accumulating evidence that shows the anomaly detector is identifying previously undetected features in the data. For example, the bevel pinion fault case mentioned above shows as a clear anomaly and the model queries reveal significant symptomatic characteristics related to the gear crack. When preparing for the in-service trial, another two gearboxes that were flagged by the anomaly detector were found to have triggered no alerts in the existing HUMS, but were later rejected for metal contamination. A number of previously undetectable instrumentation defects were also revealed, which should help improve the coverage and quality of the current monitoring.

The anomaly detector represents a major advance in HUMS data analysis. It draws attention to those regions in the data that contain significant information, and can be configured to automatically adapt to improve its performance.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Rotorcraft Health and Usage Monitoring Systems (HUMS) were first installed on the North Sea helicopter fleet in the early 1990s. Since that time a large amount of in-service experience has been accumulated, and there is good evidence to show that HUMS have contributed significantly to a reduction in the airworthiness accident rate. A major study conducted in 1999 by the independent research organisation SINTEF (1999) (reference [36]) concluded that: "HUMS was probably the most significant isolated safety improvement of the last decade". In the introductory section to CAP 693 (1999) (reference [10]), the UK CAA state: "It is considered that first generation HUMS, which added comprehensive vibration monitoring to existing health monitoring techniques, has already demonstrated the ability to identify potentially hazardous and catastrophic failure modes, and has already reduced fatal accident statistics." In service experience continues to demonstrate significant safety, operational and maintenance benefits. However experience has also shown that, while HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected and identified using current threshold setting methods. This was illustrated by an incident in 2002 involving an AS332L main rotor gearbox (MRGB) that was removed after a gearbox chip warning. On the subsequent inspection, a large crack was found in the bevel pinion. The chip warning occurred purely because a secondary crack fortuitously released a fragment of material. Normally this type of crack would not generate any debris. Although there were some trends in the vibration Condition Indicators (CIs) for the component, the HUMS did not generate any alerts. But the presence of trends raises the question; could HUMS detection be improved?

Earlier research, conducted as part of the CAA's helicopter MRGB seeded defect test programme, evaluated a number of alternative monitoring techniques including alternative sensors and alternative analysis techniques. The general conclusion at that time was that these alternatives offered no significant improvement over the in-service vibration health monitoring techniques. There was one notable exception, reported in Harrison and Baines (CAA Paper 99006,1999) (reference [17]), which demonstrated the potential benefits of applying machine learning techniques to seeded fault test data, in particular that of unsupervised machine learning such as clustering. The CAA had hoped that the success of this study would motivate industry to build a 'Next Generation HUMS' that may have provided early detection of the bevel pinion crack that occurred three years later. However, the work had been limited to rig test data only, and traditional analysis techniques persisted for in-service HUM systems. The missed bevel pinion crack incident coupled with the belief that a better alerting strategy could be developed motivated the CAA to commission a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841), which is reported here. The CAA contracted GE Aviation to carry out this programme in partnership with Bristow Helicopters Limited (BHL), analysing IHUMS data from BHL's North Sea AS332L fleet. The following extract is taken from the CAA's requirements for this programme of work.

The ongoing review of the results of the two CAA/DfT/HSE/UKOOA funded helicopter MRGB seeded defect test programmes has indicated that there is scope for improving the effectiveness of HUMS data analysis. The main issues identified as requiring attention are:

a) ***improvement of warning time*** *(i.e. the time between warning and component failure) - when conducting retrospective analyses, the presence of defects is nearly always apparent to analysts in the data in advance of any indicator thresholds having being exceeded, and hence any warnings being generated. It should be borne in mind, however, that if a warning and the associated indicator histories are not judged conclusive, it is common practice to fly-on while 'close monitoring' for a defect. Hence improvement in warning time must not be at the expense of the warning's 'quality' for maintenance decision-making.*

b) ***detection of build defects*** *- many warning thresholds are tailored on installation of the component/assembly using a simple 'learning' process. This improves sensitivity without increasing the false alarm rate. However, in the event of a build anomaly or defect these thresholds are set too high, effectively de-sensitising the analysis to the subsequent propagation of defects. Additionally, during the 'learning' period the threshold will, at best, be at a higher fleet average based level, further reducing the protection against defects introduced by maintenance errors. Hence a system that can provide increased sensitivity without increasing the false alarm rate and without requiring a 'learning' period after each maintenance action would represent a significant improvement.*

c) ***accommodation of unexpected gear indicator reactions*** *- the identification of defects in a timely manner can be compromised by the rigid application of preconceived ideas on how defects will manifest themselves in the vibration data. Experience has demonstrated that a wide range of reactions is possible, both in terms of which indicators react and how they respond. A more robust and capable analysis technique is therefore required if effectiveness is to be improved.*

d) ***accommodation of reducing gear indicator trends*** *- certain types of defect can manifest themselves as reducing indicator trends. A technique is required that can detect these.*

Experience has shown that the above issues can be mitigated through the use of well-trained and experienced human analysts. In the in-service environment, however, it is impractical for human analysts to examine all data due to the large quantities generated on a daily basis. Hence a crucial factor in improving the effectiveness of HUMS is the establishment of a more sophisticated means of identifying the sections of data of interest. If this can be achieved it will result in a reduction of the quantity of data requiring detailed analysis, enabling human analysts to focus their efforts where their skills are still essential.

In common with the CAA study reported in Harrison and Baines [17] this current programme of work utilises an unsupervised cluster algorithm but the overall approach is very different. The earlier study, hereafter referred to as Study II, demonstrated the potential of Artificial Intelligence (AI) techniques and was very successful, but the step from laboratory conditions to flight-line operations is significant. Study II is reviewed in Section 2, but some of the more notable features that distinguish the current work from the earlier study include:

a) Study II involved seeded fault tests and the researchers knew which data were considered healthy and which data were post fault initiation. In operational data,

while it would be expected that the majority of acquisitions are sourced from healthy components, there is no a-priori delineation between healthy and non-healthy data. It is not therefore possible to build a 'clean data model' of normality.

b) Study II involved detecting a fault in the knowledge that a fault exists and this reduces the space of uncertainties that will be present in any real-world application of health monitoring. Prior knowledge of a fault in a controlled test means that one can be reasonably sure that trend anomalies in the data are due to the presence of a fault whereas in an operational environment there will be many other factors that can cause an observed anomaly. The baseline for trending in Study II was derived from a cluster model of 'normality' learnt using known healthy data from a single gearbox (there were two tests but each test involved a single gearbox). Operational aircraft histories will contain multiple gearboxes, different gearbox builds, sensor faults, etc.

c) In Study II a single gear had more than one associated sensor. In most HUMS, the majority of gear indicators are generated by a single sensor. There is therefore not the same opportunity to fuse multiple sensors and corroborate findings.

d) The dataset in Study II was small compared with that available from operational aircraft histories. In most data driven analysis, the larger the dataset the better. However these large datasets contain much variability that does not exist in seeded tests. The variability will be due to factors such as the operational regime at the time of acquisition, effects of gearbox builds and instrumentation issues.

The aim of this current programme of work is to develop an anomaly detector for HUMS data. The term 'anomaly' is used throughout this report, and the abnormal HUMS data detection models are described as 'anomaly models'. The particular interpretation of an 'anomaly' in the context of this work is given below after first explaining the notion of 'anomaly'.

The English Oxford Dictionary defines an anomaly as "something that deviates from what is standard or normal". In the research literature the term 'novelty' appears more frequently. The English Oxford Dictionary defines novelty as a "new or unfamiliar thing". The definition of both terms is given because, although it appears that most researchers interpret these terms as meaning the same thing, algorithmic approaches vary widely and some make an implicit assumption that an anomaly is a very rare event.

The need for anomaly detection arises in situations where supervised learning of fault related patterns is impractical. Supervised learning requires that every signal acquisition be tagged with a known classification such as healthy, cracked gear, bearing spall, etc. Fortunately component related faults are rare but this means that there is no large library of tagged fault data. Even if such a library existed, the required size would be indeterminate because symptoms for a single fault type can vary between cases. But normal (healthy) data are usually easy to acquire. Recognition of a fault could then be achieved by building a model of normal behaviour against which a test for the abnormal can be applied.

In the context of this work anomaly detection is considered to be a process that calls attention to events that are of interest. This could result in a higher alerting rate than might be expected from a strict definition of anomaly detection. For example, if instrumentation anomalies have previously been unrecognised for long periods of time, the application of anomaly detection to historical data would, correctly, generate a high number of alerts. However, the triggering of alerts should be a matter of policy and be controllable. This policy determines what events should be flagged and, as with any new system, it needs to evolve with experience. In this data driven

approach, as experience grows the view of what events an anomaly detector should draw attention to may change over time. For example, a set of symptoms viewed today may be considered benign whereas once they caused concern. Next generation HUMS should therefore be able to adapt from experience.

This programme is structured as two phases. Phase I involved the development of unsupervised analysis techniques to enhance existing HUMS, together with their 'off-line' demonstration. It also included a literature survey to identify other research work on unsupervised learning and anomaly detection. Phase 2 involves a six-month in-service trial of the system developed in Phase I, with an option for an additional six-month extension. Phase 2 was dependent on a successful demonstration from Phase I. This demonstration was successful and Phase 2 has now taken place.

This interim report presents the work completed in Phase 1 of the programme, which was the development, off-line analysis and demonstration phase. Section 2 presents a summary of the literature review. Section 3 reviews the existing tools used for this programme. Section 4 presents representative examples of key observations made when conducting exploratory analysis of the IHUMS data. Section 5 introduces the anomaly detection approach and Section 6 describes the software tools developed for this programme. Section 7 provides an interim evaluation of the anomaly detector's performance. Conclusions can be found in Section 8 followed by recommendations in Section 9. A brief review of some fundamental concepts that are pre-requisite to reading this report is provided in Appendix A. Appendix B summarises key equations. Finally, a collection of notes, that are referenced in the main text, provides additional information for the interested reader.

2 Previous Work on Anomaly Detection

This section comprises a brief overview of previous work on anomaly detection. Detailed surveys of the literature have previously been published (for example, Markou and Singh 2003a and 2003b) [23, 24]. A small subset of the literature is described to provide examples of material considered in this programme when planning the approach to the current task. For example, Support Vector Machines [8] are reputed to be good supervised classifiers and they have recently been applied to anomaly detection, so there was good reason to consider their use in this programme of work.

Anomaly detection finds application in any situation that requires the automatic detection of some object, state, or event that is unexpected and where there are insufficient example cases with which to train a supervised classifier. Even when sufficient case data exist, a new class might appear at some later stage that was not available for training or was not foreseen. It is not surprising therefore that anomaly detection finds general application in diverse fields. In addition to machine health monitoring, there has been a great deal of interest in anomaly detection for: network security, video surveillance, image processing and biomedical data analysis. The work on anomaly detection has been very active within certain domains such as intrusion detection, which is concerned with detecting attacks on network systems by differentiating between normal users and intruders. It may be surprising to learn that some of the observations on the practical issues concerning anomaly detection mentioned in this report have much in common with those noted by Eskin et al. (2002) [16] for intrusion detection. For instance, the current work has found that HUMS data that were believed to be largely free of abnormalities do contain anomalies that have gone undetected. While the number of normal (healthy) cases is the majority class, there are a significant number of anomaly cases, which if not handled carefully could lead to a system that would class a fault as normal. Eskin et al. note that network data transactions can contain previously undetected intrusions and the same intrusion can occur multiple times.

Most early applications of anomaly detection made the assumption that all training data were normal. However, it is now more widely acknowledged that some application domains are likely to contain a small percentage of data that are abnormal, i.e. they contain outliers. Methods that explicitly address the existence of outliers usually have a parameter that can be adjusted to tune the trade-off between true positives (correctly classed as abnormal) and false positives (incorrectly classed as abnormal).

The work on anomaly detection has not produced a new class of learning algorithm. Indeed, within the anomaly detection literature, the informed reader will find the familiar machine learning algorithms such as Naive Bayes, Decision Trees, Neural Networks, Nearest Neighbour Clustering, etc.

A general review of the literature on anomaly detection is contained in Markou and Singh (2003a and 2003b) [23, 24]. Their review cites 155 papers. The review is split into two papers, the first covering statistical techniques and the second neural networks. This is a recent review but it is questionable whether neural networks would justify a category of their own were the paper to be written today. Neural networks peaked in popularity in the 1990s but recently there has been growing acknowledgement that many neural network algorithms are not that different to more traditional models or extensions being developed within other disciplines. Neural networks are useful tools but there is no reason to treat them as a special class of model for anomaly detection.

Some machine learning algorithms, such as those associated with supervised classification, would be less suited to the HUMS domain (compared to unsupervised approaches). But they have been used for anomaly detection. For example, Tax and Duin (1998) [38] use the instability of the outputs of simple linear classifiers to detect anomalies. They use a two-class classifier on multi-class problems by training separate classifiers to separate each class from all other classes. For a new object, they compute the variation in the output of the different classifiers and compare it to a 3-standard deviation threshold that is derived from the training data. This type of model has application in situations where the task is to detect a new type of class. But for HUMS data there are no prior multiple classes.

Unsupervised learning algorithms have found wide application in anomaly detection. The most popular are clustering or density estimation[i] methods. Principal Components Analysis (PCA) has also been widely applied. PCA rotates the data onto a new set of axes (principal components) that are weighted composites of the original variables. The first principal component has the highest variance, the second principal component the second highest variance and so on. Often, the first few components contain most of the data variance (or information) in which case it is possible to reduce the dimensional space by mapping (transforming) the original variables onto the first few components. There will be a loss of information but usually the features of most interest will still be present and more visible in the reduced space. Occasionally there is more value in mapping the data onto the minor components. Outliers can be detected within this mapped space by applying a statistical threshold test such as Hotelling's T^2 statistic[ii]. An example of the application of PCA to anomaly detection can be found in Shyu et al. (2003) [35]. Outliers in the data can severely affect the result so training is restricted to normal data.

Clustering is an unsupervised learning technique that is designed to segment objects (data) into clusters (or groups). Assignment of an object to a cluster is supposed to reveal that the object is more similar (according to some measure) to its within cluster neighbours than to objects assigned to other clusters. K-means and hierarchical methods are simple and popular methods for segmenting data. K-means looks at the distance between objects and aims to segment the data such that objects within a cluster are as close to each other as possible but as far from objects in other clusters as possible. A commonly used distance measure is squared Euclidean distance but other measures of similarity exist such as the angle between points, or for binary data, the percentage of bits that differ (Hamming distance). Hierarchical methods construct a tree. The root node is a single cluster that contains all objects, the leaf nodes are individual objects, and the internal nodes are clusters at different scales dependent upon the level they sit within the hierarchy. A range of distance measures exist including those used by K-means. A more sophisticated method is Mixture Modelling. Mixture models are also used for density estimation and regression. They can describe models with mixed data types such as real-valued variables and discrete variables. There is a huge volume of work concerning the theory and application of mixture models and they find application in disciplines that span image analysis to social science studies. Mixture models can utilise a multitude of density distribution types but the Gaussian and Multinomial distributions are popular for modelling continuous and discrete data respectively.

In simple terms, a Gaussian mixture model looks like a cluster model with each cluster parameterised by a mean vector, a covariance matrix, and a weight that denotes the prior probability of the cluster. The covariance matrix can be unrestricted, in which case it will allow a cluster to rotate to model the correlations local to that cluster, or restricted such as diagonal form. An object can be associated with more than one cluster. There are various algorithms for training mixture models but possibly

the most popular is maximum likelihood via Expectation Maximization[iii] first explained by Dempster, Laird and Rubin (1977) [14], (see also Bishop 1995 [4] for a concise introduction, or for more detail McClachlan and Krishnan 1997 [25], and McClachlan and Peel 2000 [26]). Mixture models are also known as generative models. This basically means that the model can be viewed as a machine that could generate the data that was used to train it. A good model will be a good generator. The likelihood score gives a measure of how well the model represents the data, the higher the score the better, hence the term maximum likelihood. The likelihood value for an object gives a measure of how well the model represents that object (or how well the object fits the model). For a model to be useful, it is usually desirable that it provide a generic description of the data which means the number of clusters should be fewer than the number of training objects. The number of clusters is often determined by information metrics such as the Bayesian Information Criterion (BIC)[iv] that seeks to find an optimal trade-off between the likelihood score and the number of clusters. The BIC is a heuristic and only provides guidance as to the number of clusters. Mixture models are discussed further in Section 5.1.

It is easy to see how a mixture model could be used for anomaly detection. A potential anomaly is one that does not fit the model and this fitness could be given by its likelihood score (assuming that the model represents only normal data). Examples of this approach include Bishop (1994) [3] and Tarrasenko et al. (1999) [37].

There have been a few attempts to build a Gaussian mixture model anomaly detector that is robust to training data containing a small percentage of outliers. Lauer (2001) [21] assumes that the percentage of outliers is small (1% or less) and that the outlier distribution along with a rough idea of their number is known. An understanding (assumed) of the outlier distribution is used to constrain the influence of outliers on the estimation of the normal data distribution. Having distributions for both outlying data and normal data permits a simple test for outlier detection: an outlier is flagged when its probability of belonging to the outlier distribution is greater than that for the normal distribution. Eskin (2000) [15] also models the data as two distributions, one for normal data and the other for anomalies. The model likelihood is a product of the likelihood from both distributions. The set of anomalies is initially empty. A model is learnt, and each case is then tested for membership of the anomaly set. This test is done by computing the difference in the model likelihood scores with the case remaining in the normal set and then with the case moved to the anomalous set. If the difference is greater than a pre-specified threshold the case is moved permanently to the anomaly set. The model has to be re-estimated at each step because the set membership changes.

While on the subject of density estimation mention should be given to Extreme Value Distributions and their application to anomaly detection. Density estimation is usually concerned with modelling typical behaviour but Extreme Value Statistics is concerned with characterising untypical behaviour. Extreme value distributions model the extreme deviations from the central location of probability distributions. For many applications values, for modelling, are only available once they exceed a certain threshold (such as payouts for insurance claims). The limiting form of the distribution for observing some extremum is described by one of only three distributions which are all a special case of the generalised extreme value distribution. Extreme values bring to mind the notion of very rare events but extreme value theory has broad application. Roberts (1999) [29] applies extreme value theory to novelty detection in a range of data sets including the detection of epileptic activity in a patient.

Consider now a supervised task where the learnt function is to assign a new case to one of two classes. A popular approach is to learn the probability densities of the two classes first before constructing a decision boundary. A radial basis function network

performs this type of modelling. Defining the class densities can be a difficult task and becomes harder for data described with a high number of dimensions. Vapnik (1995, 2001) [41], [42] discusses the issues with density estimation and introduces a new type of learning machine called the Support Vector Machine (SVM). It was many years earlier that Vapnik suggested a method to find an optimal hyperplane separating two classes. This hyperplane is linear and is constructed in a way so as to maximise the separating margin between the classes. This optimal hyperplane is defined using only a small number of the training points which are called the 'support vectors'. Figure 2-1 shows the margin for two separable classes.

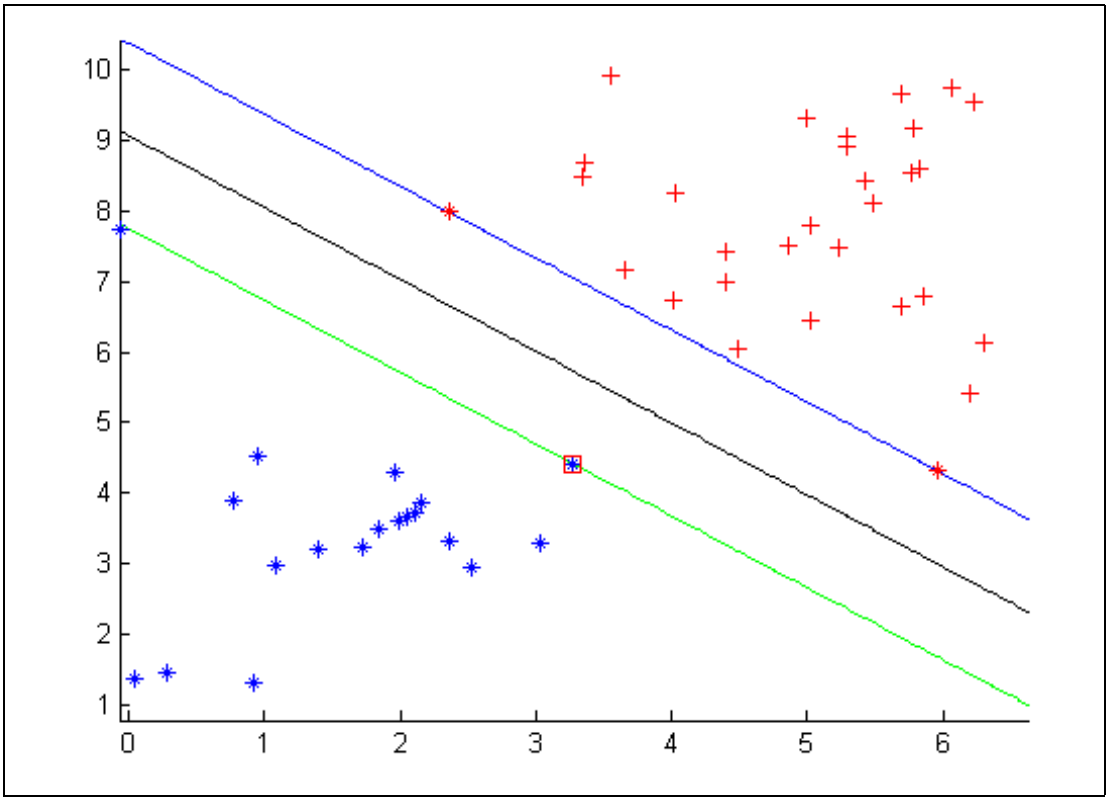

Figure 2-1 Two classes separated by a support vector machine. The dark line is the decision boundary. The green and blue lines are maximum-margin lines and the points on these lines are support vectors.

The original algorithm was a linear classifier. Boser et al. (1992) [5] showed that a non-linear classification problem could be transformed into a linear one by mapping the input space to higher-dimensional feature space. Further advances in the application of SVMs came about when the concept of a soft margin was introduced that permitted some overlap between classes (see Cortes and Vapnik, 1995) [12].

The support vector machine idea has also inspired a type of model that can be applied to anomaly detection. Tax and Duin (1999) [39] give a method that attempts to find a sphere of minimal volume that encloses the training data. If, for simplicity, it is assumed that all training data are representative of the domain (such as healthy data), the sphere will provide a description of the domain and will be capable of rejecting all cases that do not fit the domain (in other words, the outliers). Tax and Duin found that a Gaussian kernel[v] worked well. A similar approach to anomaly detection is given by Schölkopf et al. (1999, 2000) [33, 34]. The data are nonlinearly mapped into a high-dimensional feature space using a kernel such as a Gaussian. The data in feature space are then separated from the origin with maximum margin. Following on from these unsupervised algorithms, Ben-Hur et al. (2001) [2] have defined a method for

Support Vector Clustering. Once again, the Gaussian kernel is the function of choice for the nonlinear mapping. All of these unsupervised methods are tolerant to a small number of outliers in the training data.

The interested reader will find a detailed tutorial on SVMs in Burges (1998) [8].

Aerospace applications

A few applications of anomaly detection within aerospace are described next.

Hayton et al. (2000) [18] have applied the anomaly detection technique described by Schölkopf et al. (1999, 2000) [33, 34] to jet engine vibration pass-off test data. In a pass-off test, an engine is slowly accelerated from idle to full speed and back to idle. The frequency response of the engine to the energy induced by the rotating shafts is tracked. Most energy from a shaft is concentrated in the fundamental tracked order (the amplitude of the vibration at the frequency of the rotating shaft). The tracked energy of each shaft defines a vibration signature. Deviations from the expected vibration signature can reveal diagnostic information such as out-of-balance conditions. The data used for training and testing were pre-classified by Rolls-Royce experts as either normal or anomalous (having at least one abnormal aspect in the vibration signature). The data included 99 Normal engines for training, 40 Normal engines for validation, and 23 engines labelled as abnormal.

Yu et al. (2004) [44] have developed a methodology for detecting anomalies in aircraft engine performance data. For gas path performance parameters, sensor signals are replaced with model based residual values. The residuals are essentially the difference between the measured values and expected values derived from a mathematical performance model. The data contain a lot of variation and noise because the baseline model is for a class of engine and will not model the characteristics of individual engines. The anomaly detector looks for significant shifts in individual parameters. The signal is first segmented to identify the point in time where a shift starts. The segmentation is done by representing the signal as a series of linear segments derived using a recursive linear regression. The start of each segment is denoted by a split point. Not all split points represent significant shifts. Significant shifts are detected using a statistical test (a one-tail two sample T test). All parameters are analysed separately and then a search is performed across all parameters to select a single shift start date. Finally, a normalized shift score over all parameters is computed. Any significant shifts are then passed to a fuzzy diagnostic model to determine the most likely cause. It is noted that at the stage when individual signals are searched for shifts in values, this technique is not multivariate. In a multivariate analysis, subtle shifts across many signals can position a point a long way from its 'normal space' even though each signal in isolation would be close to its nominal response.

Tolani et al. (2006) [40] construct a finite state machine[vi] for anomaly detection in aircraft gas turbine engines. Time series data are represented using a symbolic alphabet. The mapping of a point into a symbol is done via a Wavelet transformation. Another mapping method is to test a point against a set of radial-basis functions and return the identity of the function with maximum response. Each function is associated with a symbol. A training set of symbolic sequences is used to derive the state transitions (from one symbol to the next) and their associated probabilities. Anomalies are detected by computing its likelihood of occurrence. It is noted that the data used in this work is sourced from a generic engine simulated test bench. This data would not contain the noise and variability found in real data and it is therefore difficult to assess how practical this approach is. Describing time series as a sequence of symbols has also been done by others - see, for example, Chan and Mahoney (2005) [11], Daw and Finney (2003) [13], and Salvador et al.(2004) [31].

Brotherton and Johnson (2001) [7] apply a Radial Basis Function (RBF) neural network to hydraulic system data and data from an Auxiliary Power Unit. An RBF is a network consisting of two layers of units. The first layer contains the basis units which can take various forms such as Gaussian functions. The second layer performs some kind of weighting of the outputs from the basis units. Each layer is trained separately. The weights on the second layer can often be obtained directly without the need for training. Brotherton and Johnson use the k-means algorithm to train the first layer. In essence, the anomaly detector is a cluster model. An anomaly is detected when an object falls outside of the boundary of all basis functions (each basis function is actually associated with a specific regime mode). The nearest basis unit is computed using the Mahalanobis distance[vii].

The Signal Invariance Estimator (SIE), Mackey (20001) [22], is a method for anomaly detection in time-correlated signals. It is part of the BEAM system developed by NASA's Jet Propulsion Laboratory for autonomous vehicle health monitoring (Park et. al.) [28]. It computes what is called a Coherence matrix which is basically the cross-correlation between all pairs of signals. So *n* signals would produce an *n* by *n* matrix of cross-correlated values. The coherence matrix will change as a live stream of data is received. A significant change in the coherence matrix is referred to as a mode change (a mode change can be the result of an external shift in state caused by commands). In other words, when the coherence values are in a steady state the coherence matrix is in a specific mode. As the system transitions between modes, the coherence matrix changes and then quickly converges as it settles into the new mode. Most transitions are predictable and nominal (not fault related). Unexpected transitions can be due to faults. Faults that trend will cause drift rather than transitions. This drift will violate the expected convergence law and can indicate an anomaly. In this paper, SIE is demonstrated on eight pressure sensor readings from an aircraft hydraulic system. There are 11 observations (different sets) of which nine are nominal and two are failures. The length of each observation varies between ten and 20 seconds (2000 to 4000 samples). Some nominal data were withheld for false alarm testing. There were some false alarms but the two anomalies were detected.

There is really no discussion or demonstration in most of the above papers regarding anomalies in the training data. This is a little surprising given that anomaly detection is designed to respond to events about which there is little prior knowledge. For large data sets, where it is impossible to manually verify that anomalies do not exist, some contamination of the 'normal training' data could be expected.

CAA Paper 99006, Study II

This study demonstrated the application of machine learning techniques to the analysis of vibration data acquired from defect tests seeded in a S61 MRGB. Westland Helicopters Limited (WHL) conducted the tests using their closed loop back-to-back test rig. MJA Dynamics (MJAD), now part of GE Aviation, conducted the analysis. MJAD's SMART gear diagnostic system was used to derive eight gear indices (GI's). One of the GI's, known as VIS that describes the visibility of a gear in the signal average, could not be computed for some gears. In total there were 14 gears monitored by 11 accelerometers with each gear monitored by three or more sensors and each sensor monitoring more than one gear. Spectrometric Oil Analysis (SOA) data were also acquired. Two tests were conducted. For each test a defect was seeded in the gearbox and MJAD were not informed as to the nature and position of the defect.

This brief overview does not include the oil analysis or the application of supervised learning in this study. SOA analysis was not used to identify the fault but it did provide corroborating evidence and could assist with identifying the nature of the fault. The

supervised learning was used to define alerting boundaries once the faulty component was correctly identified. Unsupervised clustering was the primary analysis technique used to correctly identify the faulty component from each test. MJAD's Pattern Learning Algorithm Toolkit (PLATO, see [17]) was used for the analysis. The vibration and oil data were analysed separately.

The data were supplied in two batches. MJAD were advised by WHL that the first batch could be considered to contain healthy data and the second batch was believed to contain detectable fault characteristics.

The first batch of data from the first test was used to build a normal (healthy) cluster model. Several data points that appeared to be outliers were removed. These outliers were thought to be caused by an erroneous acquisition. There were only five samples for each gear/sensor combination. Two normal cluster models were built: the first for gears that included the VIS indicator, the second for gears without VIS.

The second batch of data that was considered to be fault related contained seven samples for each gear/sensor combination.

The fault related samples were compared to the cluster model of normal behaviour. The default behaviour of the ISODATA cluster algorithm (see Ball and Hall, 1966 [1]), used for Study II, is to assign data points to its nearest neighbour cluster (measure the distance from the point to each cluster centre and choose the cluster whose centre is nearest). Nearest cluster assignment does not take into account the distribution of the data and it is possible for an outlying point to be mislabelled as normal. For instance, outlying data are often clustered as large spheres that enclose other more densely packed spheres (each sphere being a cluster). An outlying point could be much closer in distance to the centre of a densely packed normal cluster than to the centre of the enclosing outlier cluster centre. In this instance, the outlying point gets wrongly classed as normal.

In Study II, a number of cluster assignment approaches were tried. They were all based on approximating a Gaussian distribution from the standard deviation parameters that the ISODATA algorithm computes for each cluster. Association with a cluster could then be expressed as a likelihood measure. It appears that the different assignment techniques were evaluated on the basis of whether any trend information was evident when the distance (now likelihood) from each cluster was plotted for the fault related samples. In other words, if a defect was getting progressively worse, one might expect subsequent samples to move further away from any normal cluster. In this study, the number of clusters was small (typically seven) and so it was possible to plot the distance from each cluster and visually inspect the traces for any evidence of a trend.

Having devised a cluster assignment technique to compute movement of samples away from the normal model, a fault detection criteria was defined that would use the cluster assignment data to select the gear with a defect. A couple of approaches were tested but deemed not practical – they nominated different gears as the faulty component and it was known that these criteria were not practical for implementation in a real system. These first two approaches were based on computing the relative movement of the fault related samples to the normal related samples. It was recognised however that a trend based criteria would provide a practical and more robust method. The chosen algorithm computed the likelihood of a trend being present by comparing the match between the fault related sequence of sample distances to the normal clusters with a sorted version of this sequence – the sorted version represents the perfect trend and the more transitions (swaps) of samples required within the actual sequence to obtain the perfect sequence the less likely it is that a trend is present.

The devised fault detection technique was applied to both seeded defect tests. The gearbox used for the first test was also used for the second test following a re-build. The detection technique successfully identified the gear with a progressive web crack in the first test and the gear crack (that also generated debris due to secondary damage to a thrust washer) in the second test.

Harrison and Baines (1999) [17] make it clear that the blind analysis of the first seeded fault was a difficult task. A number of techniques were tried in an attempt to get the faulty gear to stand out from the other fault free gears. Showing the task to be difficult actually lends more weight to the need for intelligent HUMS analysis techniques because it provides evidence that trend patterns can remain well hidden even when they relate to a clear defect.

Study II was a successful programme of work. While the cluster algorithm can no longer be considered state-of-the-art for analysing this type of data it was supplemented with some innovative ideas that led to the correct identification of the faulty components. Harrison and Baines also provide an honest summation of some of the practical challenges that remained to transition the application of unsupervised techniques from a seeded defect trial to one that works in real-time on a fleet of helicopters. They recommended a further study to explore a number of unknowns to mitigate the risk of this new technology not performing to its potential. But they were clear in their conclusion that unsupervised learning, if well implemented, could alleviate many of the problems experienced with existing HUMS and could greatly increase HUMS performance.

Some observations that clearly distinguish Study II from the current programme of work are listed below.

1 Greater variability in the data. The current programme is concerned with data acquired from a fleet of helicopters. There is much greater variability in the acquired data due to things like: environmental influences during acquisition, variance between different gearbox builds, variance due to helicopter/gearbox fit, and variance due to instrumentation.

2 Sensor/gear monitoring. In study II, each gear was monitored by three or more sensors. In this programme, most components are monitored by a single sensor.

3 Unknown health. In Study II, the gearbox state of health was known. In this programme, although it could be reasonably assumed that most gearboxes will not contain a significant defect, the true state of health will not be known. Of more concern is the unknown quality of acquired data. Given that a large volume of data that have not been assessed by any expert is being processed, it can be expected that there will be components flying with undetected instrumentation or minor maintenance issues. These scenarios will generate data that are not representative of healthy gearboxes (even though the gearbox might be in a good state of health overall).

4 Quantity of data. The small quantity of data acquired for Study II was both a hindrance and a help. A small quantity of data rules out some analysis techniques and constrains the construction of repeatable tests. But limited data makes its management trivial and lends itself more readily to visual assessment. For example, if a cluster model contains very few clusters it is possible to plot the assignment of each sample to each cluster. But for a larger model this is impractical and it becomes impossible to manually assimilate the information about cluster assignments.

5 Model complexity. In this programme several years of flight data were to be analysed covering a total of 35 components. A cluster model would be far too

complex if a single model were to represent all drive train components (Harrison and Baines also suggested that it may be preferable to model gears separately when a sufficient quantity of data is available).

3 Description of Existing Tools Used in the Programme

Section 5 will describe the rationale for the approach to anomaly detection adopted for this programme. The approach is based upon density modelling using a mixture of Gaussian components with multinomial covariates[viii] (discrete variables that are used during modelling).

This section describes the existing tools that were used to implement the anomaly detection. Some ongoing developments that are relevant to anomaly detection and have been utilised on this programme are also introduced. Section 6 describes the software effort that is specific to this programme.

The existing software tools are collectively known as ProDAPS tools, so called because they have been developed under the Probabilistic Diagnostic and Prognostic System (ProDAPS) programme. ProDAPS was a Dual Use Science and Technology (DUS&T) programme jointly funded by GE Aviation and the US Air Force Research Laboratory. ProDAPS represents six years of intensive development to bring techniques from Artificial Intelligence (AI) to the application of aerospace health monitoring. The term 'system' in the ProDAPS acronym is really a misnomer because ProDAPS represents a collection of software components, techniques and knowledge that can be utilised to form part of a generic range of applications.

The ProDAPS tools fall under two broad categories: data mining and reasoning. Data mining is the process of searching for patterns in large historical databases. It finds application in many businesses and all major database vendors, such as Microsoft, IBM and Oracle provide data mining software. Reasoning is the process of automatically deriving conclusions through the application of existing knowledge. AI has developed a range of reasoning techniques but possibly the most promising is Bayesian Network technology. The main tool leveraged for anomaly detection is the data mining tool which is described next.

3.1 ProDAPS Data Mining Tool and Algorithms

The data mining tool is built around an implementation of the Microsoft OLE DB for Data Mining Specification published in 2000. Essentially, it's a protocol based on the SQL language with the purpose of providing an open interface to efficiently integrate mining technology. It also incorporates the Predictive Model Markup Language (PMML) that is a standard for the data mining community proposed by an industry led consortium. The specification was attractive because of the power of its flexibility; it was the sort of framework that had been desired from when PLATO was first interfaced to a SQL database back in 1989. Now Microsoft had delivered a specification and thousands of lines of source code that implemented the framework. This allowed mining algorithms to interface to all types of data source and provided the definitions and language extensions to query information from learnt mining models. This meant that efforts could be concentrated on developing the core learning algorithms.

Why develop a data mining tool when many third party products already exist? This question is even more relevant when it is realised that the algorithms implemented in ProDAPS, such as clustering, decision trees, association rules, etc., are typically found in most major vendor products. The answer lies in the need to develop technology for a specialist application. While most third party products allow development of bespoke solutions through an application programming interface, low level interfaces are not available to provide the required control. Often, the underlying algorithms do not satisfy application specific needs, and the lack of detailed knowledge of the implementation can be a major disadvantage. For example, density

estimation is a difficult task and the nuances of different data sets bring different challenges. So there is a need for a level of algorithm capability and detailed exposition that is not found in any single third party product. The data mining vendors naturally cater for mainstream commercial requirements.

The ProDAPS data mining tool can interface to many types of data source and it has been operated with a range of Graphical User Interfaces (GUIs), including third party products, but its real power is contained in the underlying algorithm engine. For day to day interfacing and quick prototyping of ideas, a GUI that resides in Microsoft's Excel provides a convenient tool. The data sits inside Excel but the algorithm interface to the data is managed via a SQL server engine. The data are not restricted to the 65000 row limit imposed by Excel (Excel's row limit will change with Office 2007). The Excel interface can be used to view any model irrespective of whether or not the model was built from Excel. Some sample screens are shown in Figure 3-1. For large analysis tasks, a bespoke GUI is often implemented to drive the analysis, and this was the approach adopted for this programme of work (see Section 6). SQL Server is the preferred choice of database for large transaction data mining.

The existence of the ProDAPS data mining tool and the detailed knowledge of its core algorithms and capabilities have proved essential to the progress and outcome of this programme.

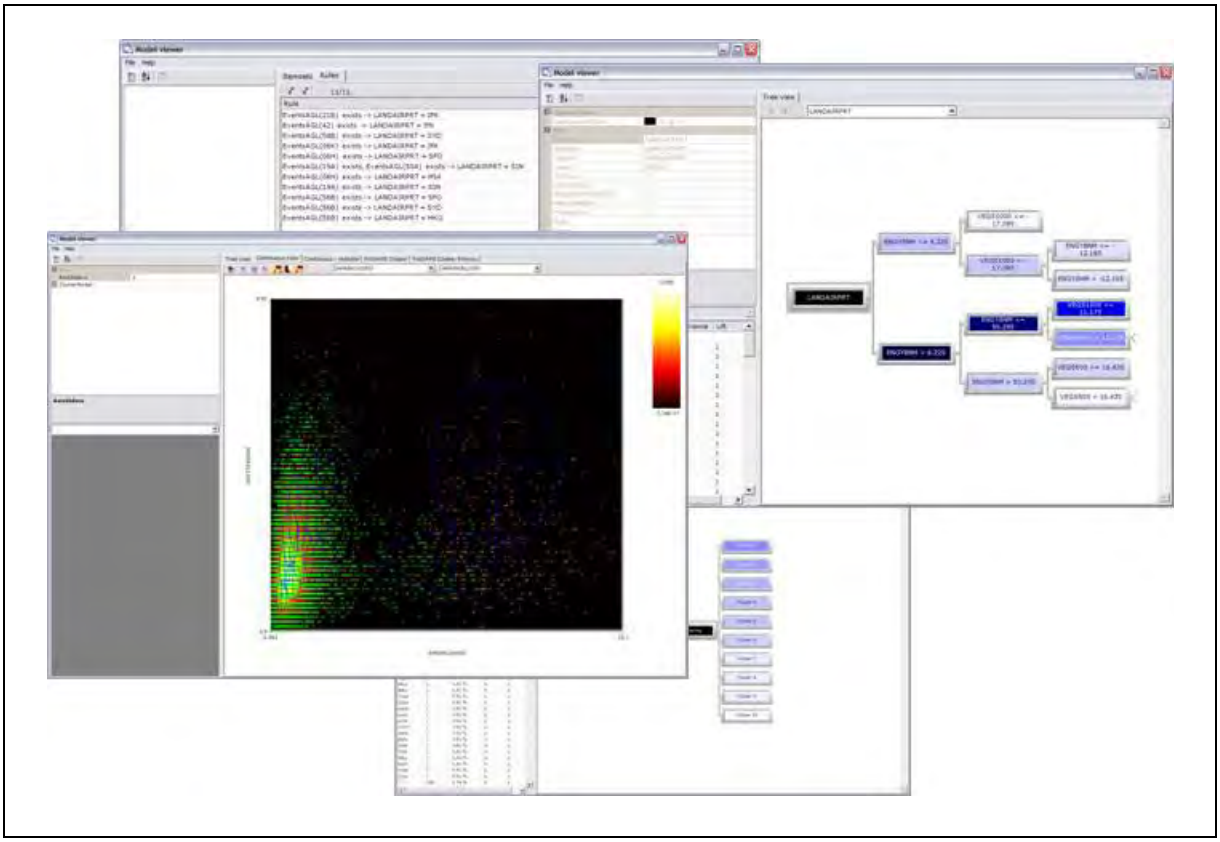

Figure 3-1 Some sample screens of a Microsoft Excel graphical user interface for the ProDAPS data mining tool

3.2 Continued Tool Development Under the ProDAPS Programme

The ProDAPS tool set continues to be expanded as the need arises. These developments are funded by GE Aviation and are designed to have generic application.

3.2.1 **Charting Tool**

This programme spurred the development of an efficient charting tool. It was originally designed to support analysis but its general utility was realised and so a subset of functions were made available for the web server that has been developed to support the live trial (see Section 6).

There is nothing special about the charting tool in terms of functionality; it was simply designed to support efficient analysis. Thus far it supports line charts and a simple type of box plot (a statistical plot also known as a box-and-whisker plot that graphically depicts an object's median, upper and lower quartiles, minimum and maximum values – ProDAPS charts use standard deviation in place of the quartiles). The charting tool allowed time histories to be quickly and conveniently plotted from data held in a database. Fleet comparisons can be made and statistical thresholds viewed – these thresholds are computed against the data held in a current filter (for example, the statistics of a selected group of gearboxes can be viewed). The box plot will show for each component (e.g. shaft/gear) its mean value, 3-standard deviation spread, minimum and maximum values. This provides an easy to assimilate view of fleet statistics.

3.2.2 **Modelling Diagnostics**

The cluster algorithm in the ProDAPS data mining tool supports a range of queries to extract useful information. For example, an expected (predicted) value for an input variable (such as a CI) can be computed along with its variance and support (the sample size, from the training data, that 'support' the prediction – e.g., if the prediction uses a single cluster the support is simply the number of training cases assigned to that cluster). Comparing the expected value to the actual measured value can be useful for diagnostic analysis.

A good implementation of a mixture modelling tool will support variable value predictions. But GE Aviation has been developing some unique capabilities that can assist with diagnosing the quality of a built model and yield additional information about anomalies. For example the concept of an 'Influence Trace' has been developed which is designed to show the variables that have most influence on driving an anomaly. In HUMS, it is customary to drill down and inspect the indicator traces when a component goes into alert. While this is a sensible thing to do it does not emphasise multivariate related information that could be important. An individual trend may look innocent but it could actually be highly unusual given the pattern across the other indicators. Influence traces are designed to capture this multivariate level of detail.

3.2.3 **Reasoning Tool**

ProDAPS has a Bayesian reasoning tool[ix] that contains a collection of features not found in a third party off-the-shelf tool. A lot of emphasis has gone into modelling large networks. Much design effort has also gone into ensuring that the tool has good synergy with the data mining tool to provide an integrated solution for vehicle health management.

Reasoning has a valuable role in health management. For this programme it could be useful in explaining anomalies and providing a secondary level of emphasis to alerts (for example, if an alert is unlikely to have been caused by an instrumentation fault then the alert could receive more emphasis). The application of reasoning technology is beyond the scope of this programme. It is mentioned here because a facility has been developed to embed a mixture model (or an anomaly model) within a Bayesian Network. This facility supports a level of model query not directly provided by the data mining tool. It also facilitates the exploration of different modelling approaches.

4 Helicopter VHM Data Analysed

This section describes the HUMS data analysed.

4.1 Bristow Helicopter AS332L IHUMS Data

The data source was IHUMS data from BHL's European fleet of AS332L helicopters. At the start of the programme there were 15 AS332Ls – nine at Aberdeen, five at Scatsta and one at Den Helder. The analysis included all the AS332L gearbox vibration data: MRGB, left and right hand Accessory Gearboxes (AGBs), Intermediate Gearbox (IGB) and Tail rotor Gearbox (TGB). The data had already been processed by IHUMS to extract its set of CIs.

The data were supplied in a sequence of Westland-format binary files taken from BHL's Aberdeen and Scatsta Windows NT IHUMS servers and the old Aberdeen DOS server, which contained data going back to 1993. The data tables in each file contained a maximum of 2,500 records of processed results. When this limit is reached the oldest records are overwritten.

4.2 Other Data

4.2.1 CHC Scotia AS332L IHUMS Data

The most important fault data set for evaluation of the anomaly detection system was the CHC Scotia IHUMS data for the cracked AS332L bevel pinion. This dataset was from a different fleet but the same model of aircraft.

4.2.2 CAA Seeded Defect Test Data

The CAA's AS332L main rotor gearbox seeded defect test programme was another possible source of fault data for evaluation of the anomaly detection models. GE Aviation had participated in the programme and while the company had access to its own vibration indicator results, the indicators are different to those used by the IHUMS, and therefore it would have been difficult to relate these results to the IHUMS data. The results from Westland's analysis would more closely map onto the IHUMS data but would require additional work to obtain them. It is quite likely that these data would have a different distribution compared to the operational data because they were not subject to the same variable influences. Given the additional effort and the concern that the data might not be directly comparable to the operational data, a decision was made not to pursue this data source.

4.3 Data Extraction and Structuring

The CIs resided in binary flat files which did not provide a good type of data repository for efficient data mining. There was a need therefore to define a database structure and prepare the data for import to this database. For convenience, the data mining database is referred to as the GE Aviation database.

Data from the Scatsta server is merged into the Aberdeen server on a regular basis. Although there is some latency between the two servers a decision was made not to import the Scatsta data directly into the GE Aviation database for fear it could compromise its integrity.

The main requirements identified for the database structure were:

1 Ability to associate acquisitions, flights, aircraft, components and gearbox identities.

2 Ordering of acquisitions so that a component's time history could be identified.

3 A matrix style structure for CI variables so that they could be fetched using a single pass query.

The IHUMS data were accompanied by a flight header table that included for each flight: flight number, aircraft registration number, flight date and airframe hours.

The first difficulty encountered concerned the identification of a gearbox. The IHUMS only stores results by aircraft registration and does not contain any explicit gearbox information. There was no way of reliably tracking gearboxes by serial number but it was possible to use BHL's maintenance record to create a ComponentFitID that denoted a specific gearbox fitted to a specific aircraft over a specific period of time. The task required additional effort because the AGBs have the same serial number as the MRGB when first received from Eurocopter. The AGBs only gain an individual serial number when they are changed independently from the MRGB.

It was initially assumed that the data could be ordered by airframe hours but this proved unreliable so, on the advice of BHL, flight date was used instead. With a few exceptions flight date proved reliable.

The indicators in IHUMS are not stored in a matrix structure and so could not be directly imported into the GE Aviation database. Because there was a desire to keep the data updated as new acquisitions are made, a decision was taken to maintain two databases. The first was a transactional database in a relational form that could be readily updated with new data and would assist with maintaining integrity. The second database was a Warehouse that could be updated regularly but was not designed for online transaction processing. The Warehouse structure breaks the data down into individual component related tables and puts the CIs into a matrix format.

4.4 Data Elements

The shafts/gears analysed for this programme are listed in Table 4-1. Table 4-2 lists the CIs available in the IHUMS data. Not all of the indicators are used for the analysis in this programme. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data. The chosen indicators are described next.

Table 4-1 Components analysed for this programme

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MRGB
2	1	RH high speed input shaft	MRGB
1	2	Left torque shaft – fwd end	MRGB
2	3	Right torque shaft – fwd end	MRGB
3	4	Left torque shaft – aft end	MRGB
4	5	Right torque shaft – aft end	MRGB
3	6	Combiner gear	MRGB
3	6	Bevel pinion	MRGB
4	7	Bevel wheel and oil pump drive	MRGB
7	8	1st stage sun gear	MRGB
7	9	1st stage planet gear	MRGB
5	10	1st epicyclic annulus fwd (RH)	MRGB
6	11	1st epicyclic annulus left	MRGB

Table 4-1 Components analysed for this programme (Continued)

Sensor	Channel	Shaft/Gear	Assembly
7	12	1st epicyclic annulus aft (RH)	MRGB
7	13	2nd stage sun gear	MRGB
7	14	2nd stage planet gear	MRGB
5	15	2nd epicyclic annulus fwd (RH)	MRGB
6	16	2nd epicyclic annulus left	MRGB
7	17	2nd epicyclic annulus aft (RH)	MRGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB
3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB
4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MRGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MRGB

Table 4-2 IHUMS indicators

Condition Indicator	Description
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation (rms)
FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev (rms)
ETA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA__CVG	Signal average convergence

4.4.1 IHUMS Vibration Processing and CIs

A brief description is given of the IHUMS CIs (developed by Westland Helicopters) that have been included in the anomaly modelling process.

Vibration signals acquired from accelerometers mounted on the helicopter transmission system are first processed to produce Signal Averages. The vibration effects of gear, shaft and coupling faults are generally synchronous with the shaft rotation. Therefore Signal Averaging provides the primary method of detecting these faults, as it enhances the signal to noise ratio of synchronous signals. By extracting unique vibration signatures for each gear/shaft rotating at a different speed, Signal Averaging can provide early fault detection, and is able to identify the component on which the fault has occurred.

The production of the Signal Average requires two signals. The first is from an accelerometer and the second is from an azimuth marker, which provides phase locked information on the rotational speed of the system. Sections of vibration data equal to one revolution of the gear/shaft being monitored are ensemble averaged to retain synchronous information, but reduce all asynchronous vibration (originating from other components) to an acceptable noise floor. The process is repeated to generate a Signal Average for each monitored component in the transmission system.

Once acquired, each Signal Average is then analysed to produce a set of CIs. The IHUMS CIs use the following naming convention:

SIG_ These CIs are calculated directly from the Signal Average (i.e. in the time domain).

FSA_ These CIs are calculated from an FFT of the Signal Average (i.e. in the frequency domain).

ESA_ These CIs are calculated from an Enhanced Signal Average (i.e. in the time domain, after specified tones have previously been removed in the frequency domain, to leave the residual content of the signal).

The following ten CIs were included in the anomaly modelling. The explanatory theory statements below are provided by GE Aviation.

SIG_PP The maximum Peak to Peak amplitude of the Signal Average

SIG_SD The Standard Deviation of the Signal Average

Theory: Many faults involving damage can increase the energy of the vibration signal, though for the majority of cases this will be at a relatively late stage of damage progression.

FSA_SO1 The amplitude of the First Shaft Order (i.e. at 1/rev)

FSA_SON The amplitude of Shaft Order N, where N is configurable (this is normally 2 for the Second Shaft Order, i.e. at 2/rev)

Theory: Imbalance of a shaft or gear will result in an increase in 1/rev vibration. Misalignment and cracks generally cause a rise in the 2/rev vibration.

FSA_MS_1 Amplitude at Gear Mesh Frequency (i.e. the mesh tone of gear 1)

Theory: The amplitude of the gear meshing tone can be affected by both general gear tooth damage/wear, or by operational factors.

FSA_GE21 Wideband Modulation (multiple sidebands) (Gear 1)

Theory: Fixed axis gears normally produce a meshing waveform with little or no low frequency amplitude modulation. However, when the gear is affected by damage such as structural failure of the gear web, casing, or support bearings, the meshing waveform begins to modulate at low frequency.

ESA_PP The maximum Peak to Peak of the Enhanced Signal Average

ESA_SD The Standard Deviation of the Enhanced Signal Average

Theory: A range of faults can increase the general residual (or noise) content of the Signal Average, and faults propagating to an advanced stage always increase this.

ESA_M6* Impulsiveness of the Enhanced Signal Average (the 6th statistical moment of the signal)

Theory: The meshing waveform produced by a gear pair is created principally by the transfer of load from tooth to tooth. Tooth bending fatigue cracks, or tooth damage, often result in localised distortion of this waveform, which may appear as an impulsive event in the residual content of the Signal Average.

ESA_WEA A Wear Indicator generated by GE Aviation, calculated as the ratio ESA_SD/SIG_SD

Theory: This is a measure of relative energy content of residual component of the Signal Average, to the total energy.

4.5 Data Exploration

Helicopter manoeuvres effect vibration (see for example Huff et al, 2002) [20] and data therefore are acquired when aircraft are in a stable flight regime. Only automatically triggered acquisitions are considered in this programme of work. Automatic acquisition ensures that the aircraft is within a predefined flight window defined by parameters such as collective pitch, engine torque, bank angle and rate of climb, which helps reduce the variability due to aircraft flight regime.

A review of the general characteristics of IHUMS indicators can be found in Salzer (1994) [32]. Although this is an old paper, in terms of HUMS history, many of the points raised are still relevant today. The paper discusses the data variability influences and the effects of maintenance. Salzer limits his discussion to five CIs (SIG_PP, FSA_SO1, FSA_SO2, FSA_MS_1, and ESA_PP) on the MRGB. A comprehensive statistical analysis across all components has been carried out on this programme but, like Salzer, for this discussion only representative samples of the key data characteristics that have direct bearing on the anomaly modelling approach are presented. It should be noted that Salzer's dataset was limited to that acquired by 1994 whereas an additional 12 years worth of experience is now available.

This discussion concentrates on the following indicators: ESA_PP, ESA_SD, FSA_GE22, FSA_MS_2, FSA_SO1, FSA_SON, SIG_PP and SIG_SD.

All the data presented here have gone through a two stage filtering process. The first filter applies a 'reasonableness limit' to the data to remove extreme outliers whose values look unbelievable. Next, a median filter is applied to remove up to two successive 'spikes' in the time series data.

4.5.1 General Data Characteristics

Figure 4-1 shows the typical variability to be found in a set of CIs over the life of a MRGB while installed in a specific aircraft. Maintenance often induces a step change in some indicators and this can be seen in FSA_SO1. The signals look typical of a healthy component with no instrumentation defects.

Figure 4-2 shows a box plot for a few CIs from the bevel pinion and a selection of MRGBs. The mean value and the range are shown for each gearbox. It is clear from these plots that, not only is there a lot of variability within an individual component, there is a great deal of variability between different examples of the same component type. The fleet plot for FSA_SO1 looks very different to the other CI fleet plots in that the plot is dominated by a couple of gearboxes. Figure 4-3 shows the time histories for FSA_SO1 with each gearbox overlaid. It is clear that the two dominating gearboxes have many acquisitions that are outlying from the distribution of the fleet. So, even after the removal of extreme outliers (from the first stage filter) there is still a considerable quantity of data that look to be different from the fleet. Although not shown, this characteristic is repeated for FSA_SON and it becomes even more pronounced for some components such as the bevel gear in that more gearboxes sustain abnormal values compared to the rest of the fleet for a significant period of time. Due to the significant quantity of data with very high shaft order energy levels, caution needs to be exercised when classifying these as normal or outliers. The question must be asked: Is there a second 'significant normal mode' in the distribution? It is possible to obtain an insight to the answer by looking at the time

histories for FSA_SON for gearboxes that show this characteristic and by visual inspection of the fleet histograms for evidence of a bimodal distribution. Figure 4-4 shows the time history for a single bevel gear with high values for FSA_SON. It is clear that the high values have a high variance. For a significant period of time from around the 900th to the 2,100th sample the data have returned to the main fleet distribution. Figure 4-5 shows a fleet histogram for FSA_SON data from the bevel gear. While this plot is a simple view of the data, there is no clear evidence of a second distinct mode. The majority of the data sit in a tight band but there are a significant number of samples contributing to a very long tail. This plot is distorted by the long tail however, which is further discussed in Section 4.5.3 and in Section 5 where the anomaly modelling process is described.

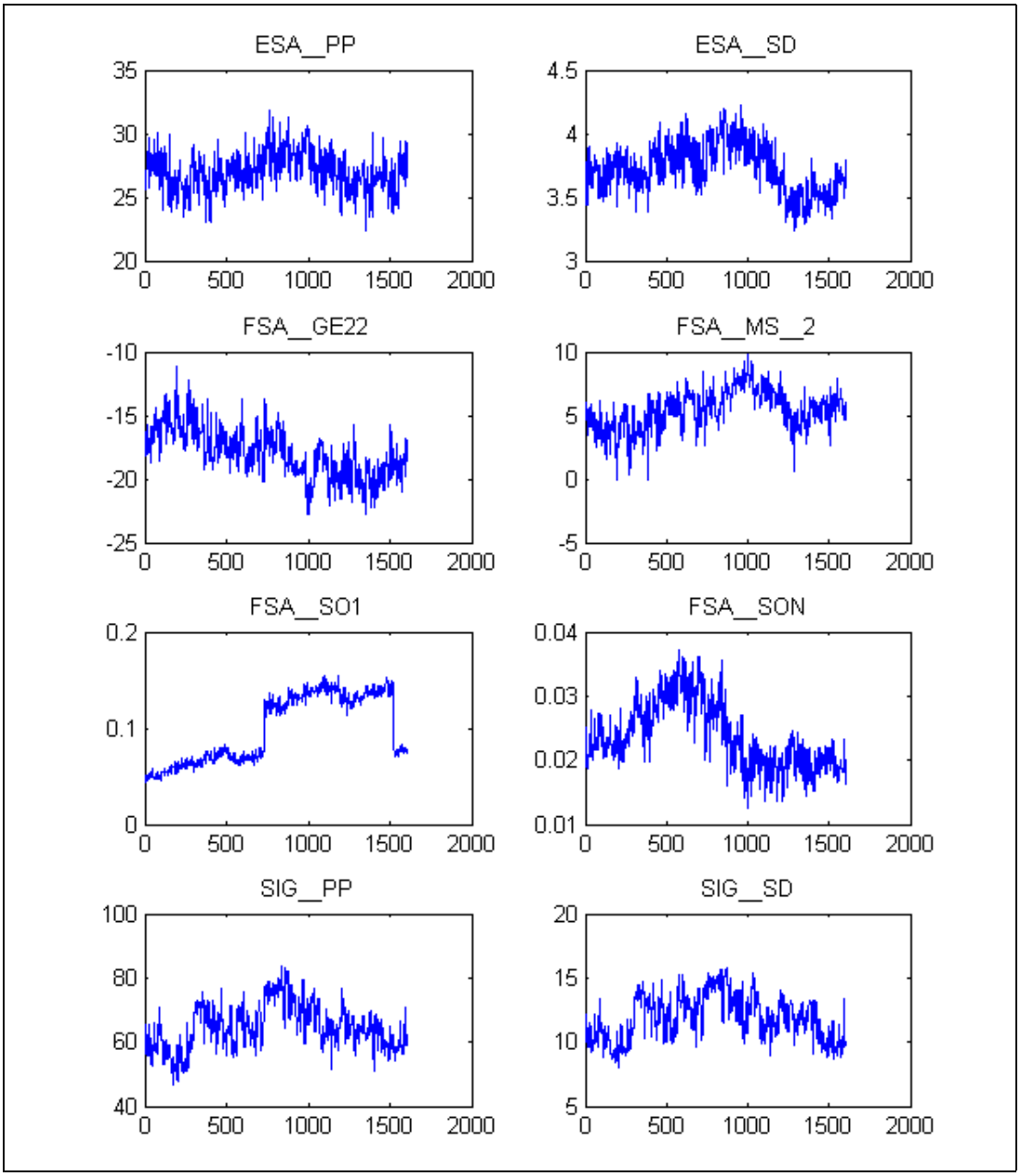

Figure 4-1 Typical behaviour of HUMS Condition Indicators for a healthy component

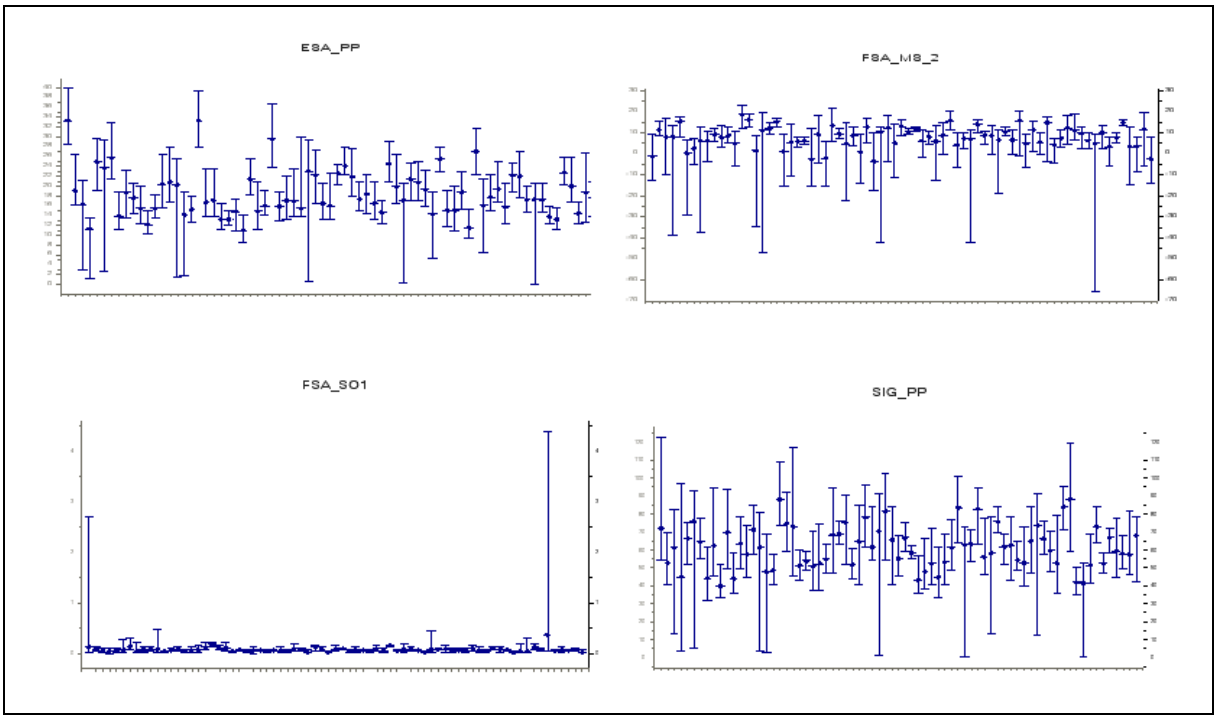

Figure 4-2 Box plot for four indicators (each line is a different example of the same component type). These plots show the variability between aircraft.

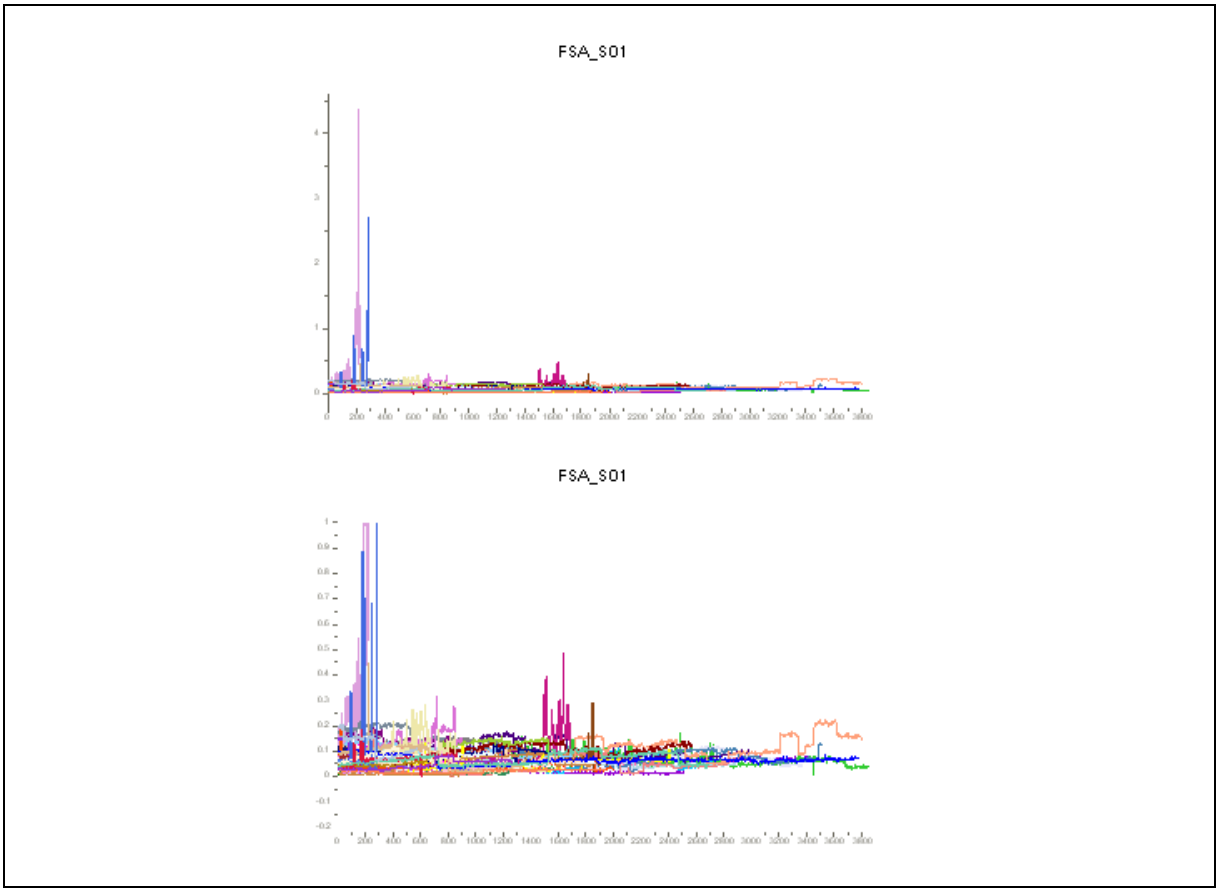

Figure 4-3 Time histories of the bevel pinion FSA_SO1 for a fleet of aircraft. Each colour is a different aircraft fit. The length of each coloured line is dependent on the number of points acquired. The lower plot is the same as the top plot but with the amplitude axis scaled.

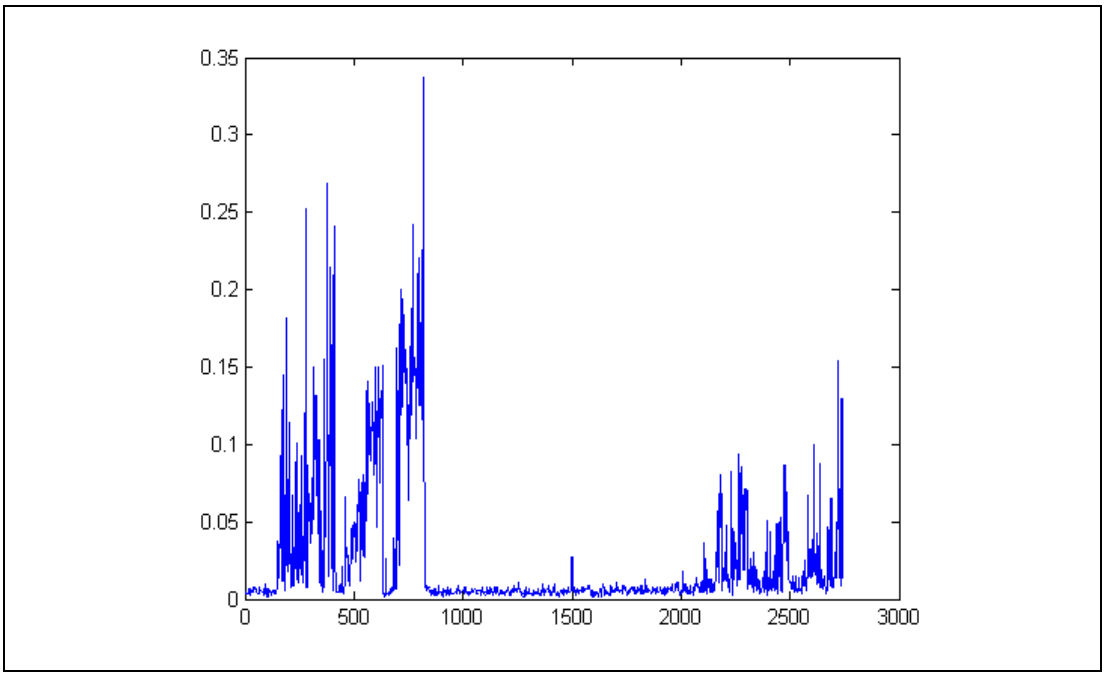

Figure 4-4 Time history for FSA_SON for a single bevel gear component

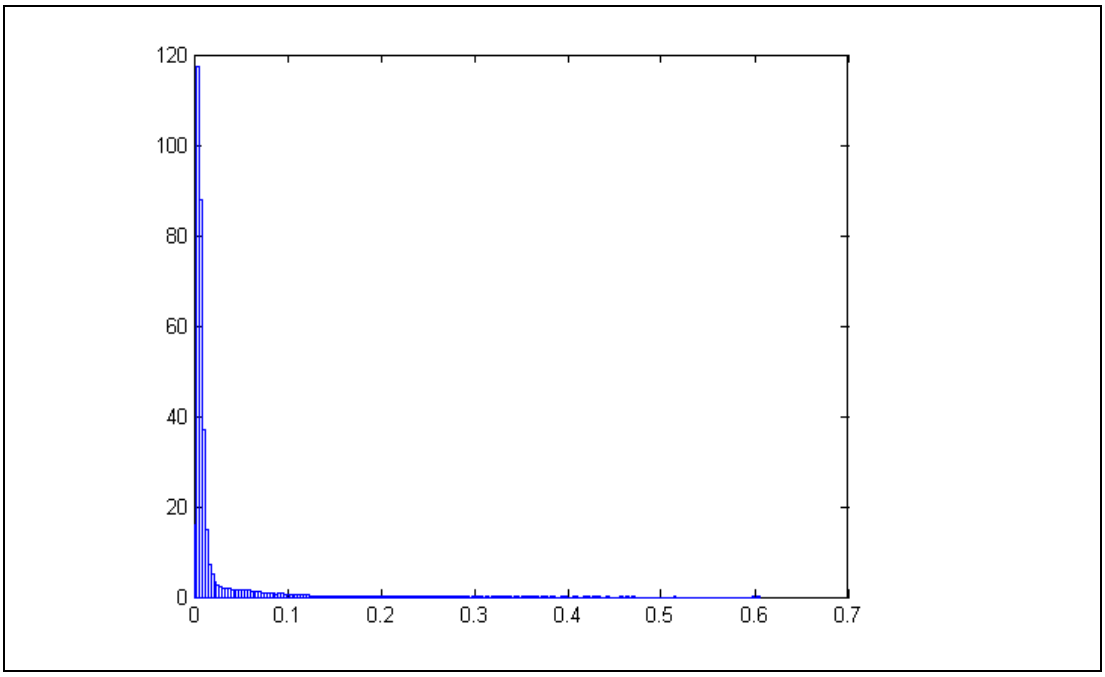

Figure 4-5 Fleet histogram for the bevel gear FSA_SON data

4.5.2 Correlation

Correlation between indicators can increase the spatial separation between gearboxes when analysed in multivariate space.

Knowledge of the CIs shows that some should be correlated. This section presents measures of this correlation for the bevel pinion. Caution needs to be exercised in the analysis because outlying data can affect the true correlation.

Table 4-3 shows high correlations between the following pairings {(ESA_PP, ESA_SD), (FSA_GE22, FSA _MS2), (SIG_PP, SIG_SD)}. These correlations are to be expected.

Table 4-3 Correlations between each indicator pair

	ESA_PP	ESA_SD	FSA_GE22	FSA_MS_2	FSA_SO1	FSA_SON	SIG_PP	SIG_SD
ESA_PP	1	0.9692	0.2946	0.0642	0.2795	0.2307	0.5211	0.4386
ESA_SD	0.9692	1	0.2784	0.076	0.263	0.1915	0.5686	0.4896
FSA_GE22	0.2946	0.2784	1	-0.8382	0.194	0.0688	0.0397	-0.0124
FSA_MS_2	0.0642	0.076	-0.8382	1	-0.0331	0.0786	0.1669	0.1808
FSA_SO1	0.2795	0.263	0.194	-0.0331	1	0.4181	0.2404	0.226
FSA_SON	0.2307	0.1915	0.0688	0.0786	0.4181	1	0.102	0.099
SIG_PP	0.5211	0.5686	0.0397	0.1669	0.2404	0.102	1	0.9488
SIG_SD	0.4386	0.4896	-0.0124	0.1808	0.226	0.099	0.9488	1

A Principal Component Analysis (PCA) can also be useful for testing the inherent dimensionality of the data. Table 4-4 has ranked the variance of each component as a percentage of total variance. The last three components contribute less than 1% of the variance and over 90% of the variance is contained in the first four principal components. While most of the variance is contained in the first few principal components, all indicators contribute some information and it is possible that some of the more interesting anomalies could be more visible in the minor components. The information regarding correlation does mean that when a point is on the edge of a distribution any correlated indicators will further emphasise its outlying characteristics. Some care must therefore be exercised when judging anomalies if the data contain a degree of duplication.

Table 4-4 Principal components ordered according to variance

Principal Component	Percent of variance by component	Cumulative
1	40.37	40.37
2	24.14	64.51
3	15.42	79.93
4	11.18	91.11
5	6.84	97.95
6	1.13	99.08
7	0.57	99.65
8	0.35	100

4.5.3 Distribution

Knowledge regarding the distributions of CIs is important for modelling because many types of model assume a certain theoretical distribution. Salzer (1994) [32] found that no other distribution offered advantages over the Gaussian distribution even though the Gaussian was, for many indicators, a rather poor fit of the data.

Figure 4-6 shows a histogram plot for four indicators taken from the bevel pinion. The indicators have a crude Gaussian shape but they cannot be considered good Gaussian fits. Figure 4-7 plots a histogram for FSA_GE22 which is the most Gaussian like but a Gaussian is still not a good fit according to the probability plot in Figure 4-7 and chi-square goodness-of-fit test at the 5% level.

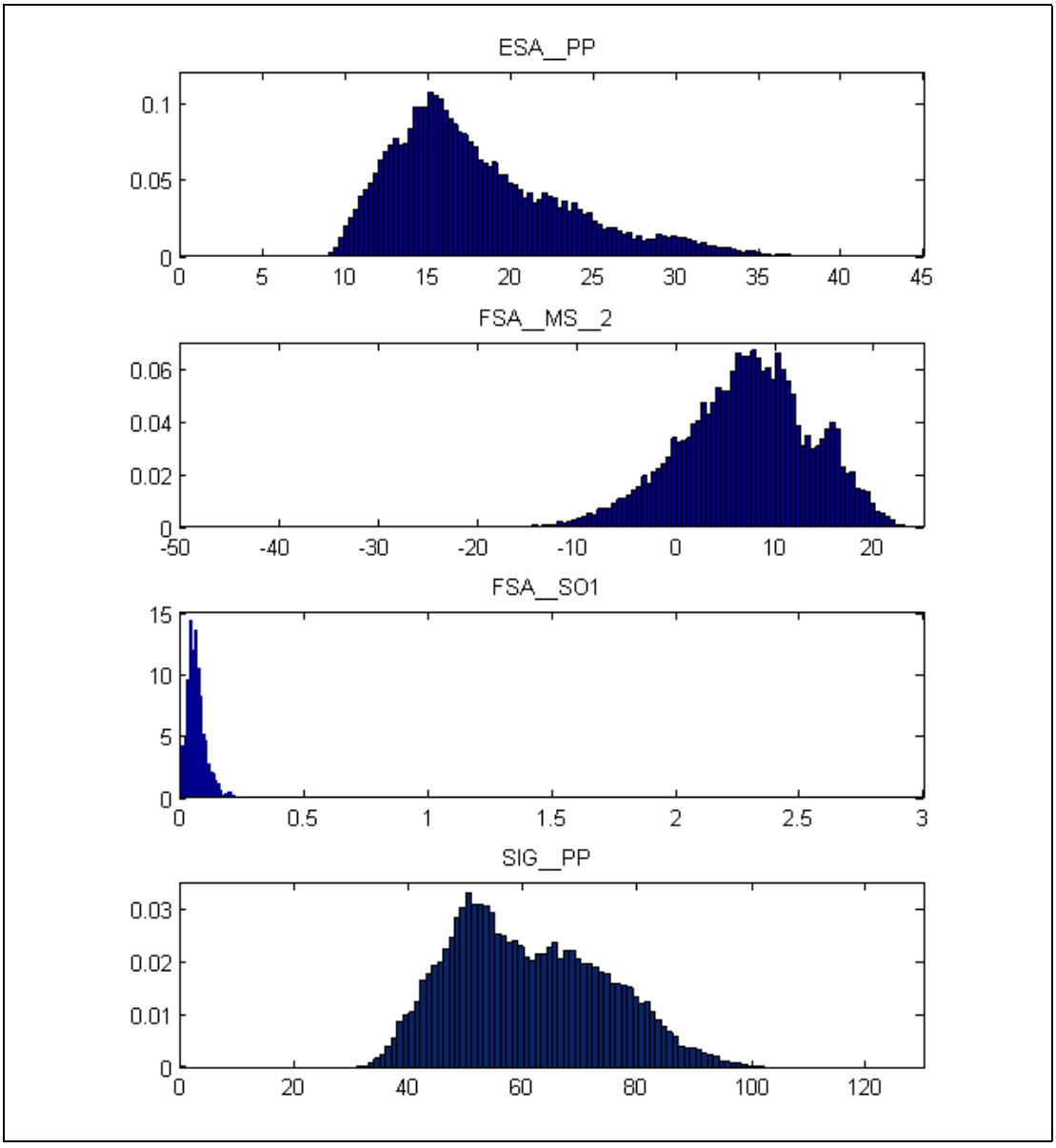

Figure 4-6 Fleet histograms for four Condition Indicators from the bevel pinion

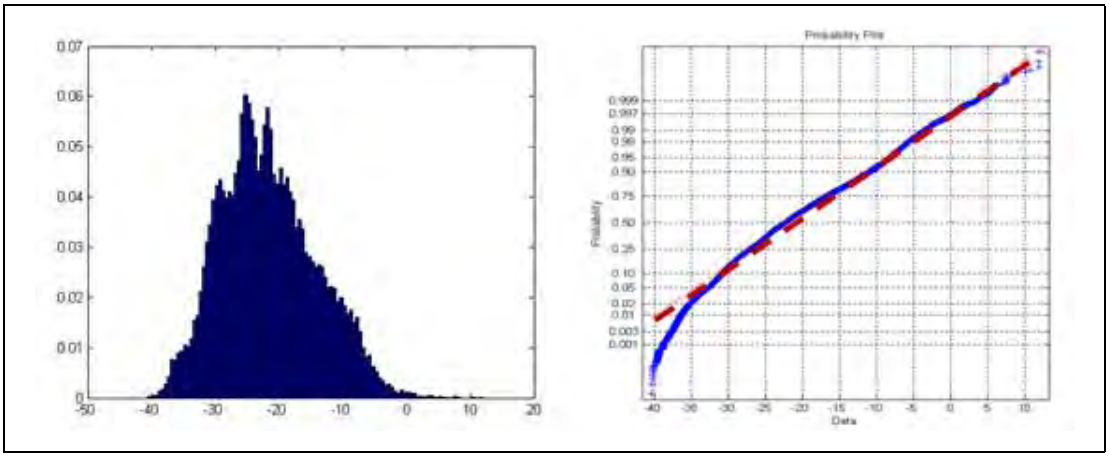

Figure 4-7 Histogram of FSA_GE22 for the bevel pinion and its probability plot showing that the Gaussian distribution does not provide a good fit (the sample data should fall near the red line for a good fit)

The histogram for the shaft order energy, FSA_SO1, looks to be one-sided with a very long tail. The distribution is similar to that for FSA_SON from the bevel gear shown in Figure 4-5. The histogram views are however distorted by the very long tail. Figure 4-8 shows the distribution of FSA_SON from the bevel gear data after it has been transformed with the log function. The bulk of the data now looks to be Gaussian in nature. On first inspection there appears to be a second mode in the right hand tail but closer examination suggests a uniform distribution that falls rapidly away for large values. A more sophisticated type of density model did not reveal any clear second mode.

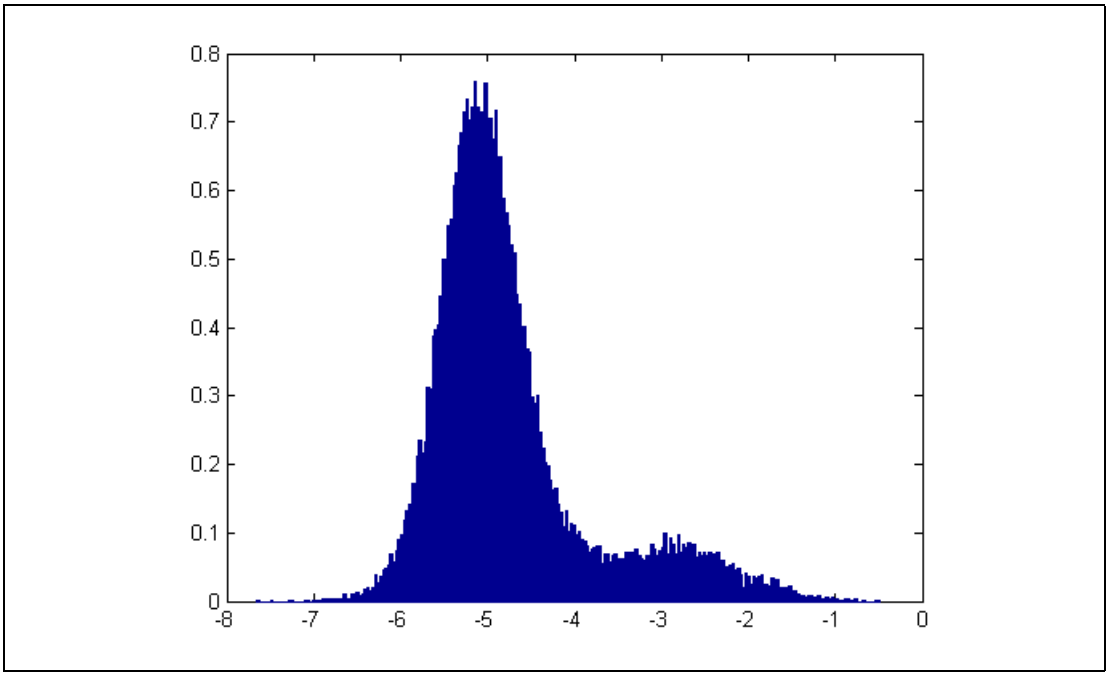

Figure 4-8 Histogram for FSA_SON for the bevel gear after transforming with the log function

4.5.4 **Summary of Data Characteristics**

HUMS data are highly variable. There is variability within the CI time histories for single components, variability in CIs between components of the same type, and different CI trend characteristics across component types in that some types are more prone to trend and suffer a higher percentage of instrumentation issues.

Instrumentation defects can distort the data distribution and will certainly challenge the threshold setting strategies used by existing HUMS. Some instrumentation defects may only be visible in a subset of CIs with the other indicators appearing not to be affected.

CIs appear Gaussian like. But the variability in the data caused by multiple influences produces a mixing effect that generates complicated distributions. It is inappropriate therefore to fit a single Gaussian distribution to CIs. This highlights the weakness of any threshold setting strategy derived from descriptive statistics such as mean and standard deviation.

There is considerable correlation between some indicator pairs. Redundancy is not implied by this correlation because a de-correlated trend can provide diagnostic information. However, it does mean that care needs to be exercised when performing multivariate analysis because correlated parameters will tend to emphasise an abnormal pattern: this emphasis may be weighted too heavily when there are parameters that measure a similar data characteristic.

5 Development of Anomaly Detection Process

There is growing understanding within the AI community that there exists no algorithm that can be championed over all others for all tasks within a problem domain. This is certainly true of anomaly detection. The choice of algorithm depends on the algorithm's applicability to the type of representation deemed most appropriate for the problem at hand. Clustering was always a prime candidate for the anomaly detector but there are many types of cluster algorithm including some based on Support Vector Machines.

There was a particular interest in the Support Vector Machine (SVM) approach when performing the literature survey because it is a relatively new algorithm in terms of its broad adoption and it might have alleviated some of the difficult problems encountered with other techniques. For example, density estimation, which is a popular approach to learning and the basis of the ProDAPS cluster algorithm, can be difficult in that the core algorithms have to be supplemented with additional techniques to avoid poor models. While there is general recognition that SVM's provide a powerful learning paradigm for many supervised tasks, the same cannot be said for the unsupervised techniques. There is no evidence in the literature to show that these machines provide any clear advantage over other approaches to anomaly detection (see for example Heller et al. [19]). Furthermore, for single class[x] machines the fit of the enclosing boundary separating the normal data from outlying data is dependent on the choice of kernel function. Similar issues exist for density estimation but therein lays the point: the performance of both unsupervised SVM's and density estimation are dependent upon the wise choice of model options and parameters. Unsupervised SVM's are not restricted to single class machines and Ben-Hur et al. (2001) [2] have demonstrated a useful approach to support vector clustering. As stated earlier, there are many different types of cluster algorithm. Some are a restricted form of a more general class of model. Some algorithms have wider application than others, but the choice is dependent on the problem to be solved. It is interesting to note that the ISODATA algorithm, that has been around for many years and was implemented in PLATO in 1989, is still widely used (see for example Memarsadeghi et al., 2005) [27].

In this programme, density estimation via mixture modelling was chosen to underpin the approach to anomaly detection. The choice was clear given what was said above and the following:

- Mixture modelling is a very flexible form of data modelling with many types of application.
- It can be used for clustering, domain description, regression, imputation[xi] and more.
- It is widely used by many researchers in many domains from image analysis to social science studies.
- A very good implementation exists within ProDAPS that includes some unique functionality.
- A great deal of experience exists with this algorithm on a range of data and it is the most widely used ProDAPS algorithm.
- Mixture models can be represented as a graphical model. They are a subset of a more general unifying probabilistic graphical model theory[xii] that includes Bayesian networks, conditional Gaussians[xiii], and Markov models[xiv] etc. This has practical importance because a health management system is typically layered as a hierarchical structure with modules processing data at different levels of

information. Communication between levels is more robust when modules conform to a unifying theory. Finally, within the ProDAPS programme, work has started on the fusion of individual mixture models to provide a sophisticated higher-level type of model. This approach is directly relevant to anomaly detection.

5.1 Type of Mixture Model

The anomaly models are a mixture of multivariate Gaussian distributions and Multinomial distributions[xv]. For the time being the multinomial aspect of the model, which is used to provide information for adapting the cluster model, can be ignored. The clusters in a mixture model are usually referred to as components, but the term cluster shall be used here to avoid confusion with HUMS drive-train components.

The anomaly models are like cluster models but they are constructed for a different purpose. Clustering is often used to segment data to discover if the data contain any natural structure. In commercial applications, clustering can identify customers that share traits and behave in similar ways (such as spending habits and product purchase preferences). The anomaly models are constructed for density estimation with the aim of identifying regions of space that are likely to represent anomalous behaviour.

In Section 4.5.3 it was shown that the IHUMS data were Gaussian like in shape but the Gaussian distribution provided a poor model of the data. That section was referring to a single Gaussian, but a mixture of Gaussians can model more complicated distributions. Figure 5-1 shows the Gaussian distribution for FSA_MS_2 from the bevel pinion. The solid line is the fit from a Gaussian mixture model (GMM). The model contains five clusters; the number determined using the BIC measure. The mixture model provides a good representation of the distribution.

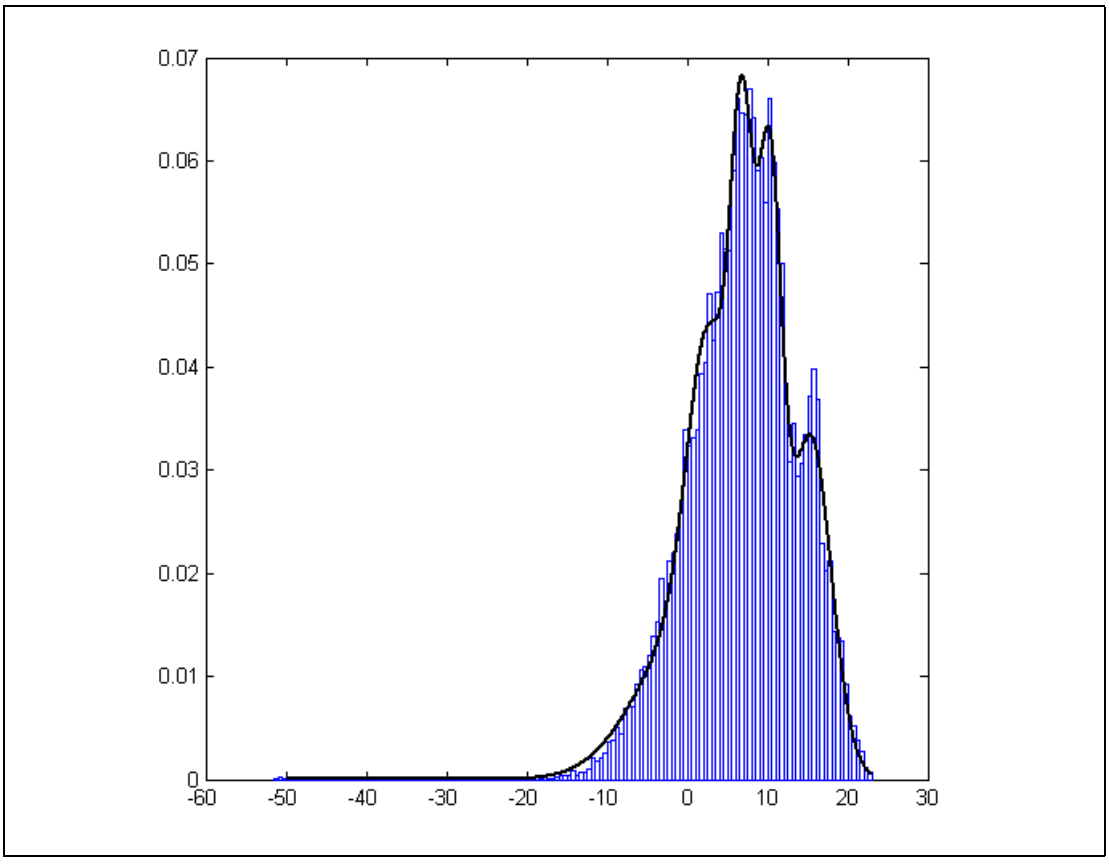

Figure 5-1 Fleet histogram for FSA_MS_2 from the bevel pinion. The solid line is the fit from a mixture model.

Constructing a GMM is not always straightforward. The ProDAPS implementation uses Expectation Maximization for training and the objective is to maximise the likelihood score. The models generated by many types of cluster algorithm are subject to the starting conditions in that the initial positioning of the clusters is randomly set and different executions of the same modelling task can yield different results. Often the results from different runs on the same model building task are similar but sometimes they are markedly different. The problem with the maximum likelihood algorithm is that the maxima can be localized, in which case the main contribution to the likelihood score is contained within a small region of space, the result being a poor representation of the data. This could occur for example where one part of the space has many clusters but another part of the space has too few. There have been several proposed methods to solve this difficulty based on sub-sampling (a random selection from the training data), or splitting and merging clusters, but there is still no universally accepted method for solving it. Therefore models need to be carefully inspected once built. Another problem is the occurrence of singularities which happen when a cluster collapses onto a data point. The likelihood for singularities increases in higher dimensional space (e.g. using more indicators). The calculation of the likelihood requires a matrix inversion and this can prove impossible when the number of points assigned to a cluster is too few. This is because the inherent dimension can be lower than the input dimension; for example, if there is 100% correlation between two input variables the inherent dimension is 1 whereas the input dimension is 2. Singularities can be avoided through the use of prior distributions to initialise the covariance matrix. The above points illustrate the need for an in-depth knowledge of the underlying algorithms used to define an anomaly detection method.

A mixture model is a probabilistic model and an object can be associated with more than one cluster, unlike a purely distance based approach that does hard assignment (an object belongs to a single cluster). A mixture model will handle missing data. For example, a model could be built with eight indicators but predictions can be done with one or more indicator values missing (treated as null). Prediction with a reduced set of indicators is different to building a model with fewer indicators in the first place and then doing predictions because the full model is conditioned on all of the indicators used in training. This flexibility is useful for anomaly modelling because it is possible to test the fitness of an object against a model using a subset of indicators if it is suspected that one or more acquired values are unreliable.

5.2 Overview of the Anomaly Modelling

The concept of the approach is simple. A mixture model is constructed, regions in the cluster space (clusters) that appear not to be representative of normal behaviour are detected, and these clusters are removed so that the final model provides a poor fit to those samples in the training data that are outliers. The model automatically detects the outliers: there is no manual a-priori tagging of outlier cases. The removal of clusters seems a foolhardy thing to do because the model construction phase has tried to optimise the data fit and this fit includes the outlying data. But ProDAPS permits the construction of models in a way that minimizes the risk of removing clusters that represent normal behaviour. It does this by using a normalised likelihood metric that is sensitive to outlying regions (see Section 5.4).

To explain the concept of the approach, use shall be made of a dataset that is very familiar to the machine learning community, called "Iris". The Iris dataset contains 150 sets of measurements from three classes of Iris. The measurements are sepal length, sepal width, petal length and petal width. There are 50 cases from each class. The three classes are readily separated by a supervised classifier. A cluster algorithm will also do a good job of segmenting the data into class related clusters, although

there can be some mixing of classes. Figure 5-2 shows a scatter plot of the data. The matrix of plots is symmetrical and all information therefore is contained in either the upper or lower off-diagonal plots. Each row and each column represents a measured attribute. The matrix includes plots for every combination of attribute pairs. It is easy to see from these plots that the class Setosa is well separated from the other two classes and there is some overlap between Versicolor and Virginica. For this illustration there is no interest in outlier detection. Instead, the objective is to extract the class that could be considered the most distinct (or different to the others). In the context of the four attributes presented here, examination of the 4th row and 3rd column of the plot matrix shows that Setosa is tightly clustered and clearly separated from the other two classes. In all plots there is some overlap between Versicolor and Virginica. Setosa looks to be the most distinct class.

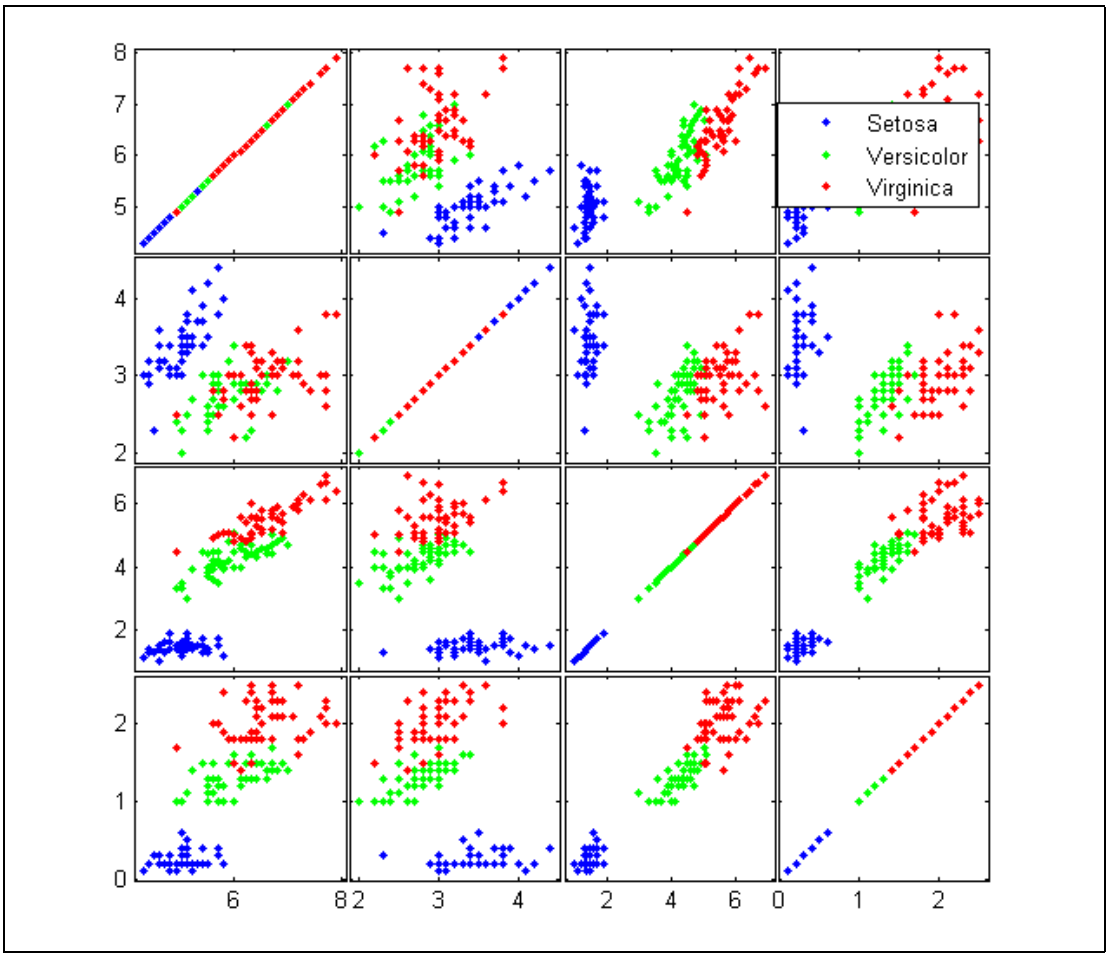

Figure 5-2 Plot matrix for the Iris data. Each row and column represents a different measurement attribute.

A standard mixture model was constructed for these data. Figure 5-3 shows the log likelihood scores for each case. From hereon the term Fitness shall be used instead of log likelihood. An object with a poor fit gets a low Fitness Score (FS). The FSs for all classes are high and they overlap. It is hardly surprising that all cases fit the model since the model has been trained to represent all of them.

Next, a model was constructed using the anomaly modelling approach. Figure 5-4 shows the FSs from this model. It is clear that Setosa has the lowest fit. The anomaly model has identified Setosa as being different to the other two classes. This is not saying that Setosa samples are anomalous, just that the other two classes have more features in common. There is an expectation of difference within these data because

they have a tagged identity (their classification). But there is no a-priori measure of this difference: this 'difference' information is contained in the measurements and it is fortunate in this example, that the data are few and have a simple structure that permit visualisation of different characteristics. But the patterns of interest in real world problems are often well hidden and charting data proves inefficient, unreliable and impractical for large quantities of data. There is a need therefore for technology that can automatically highlight regions of interest in the data.

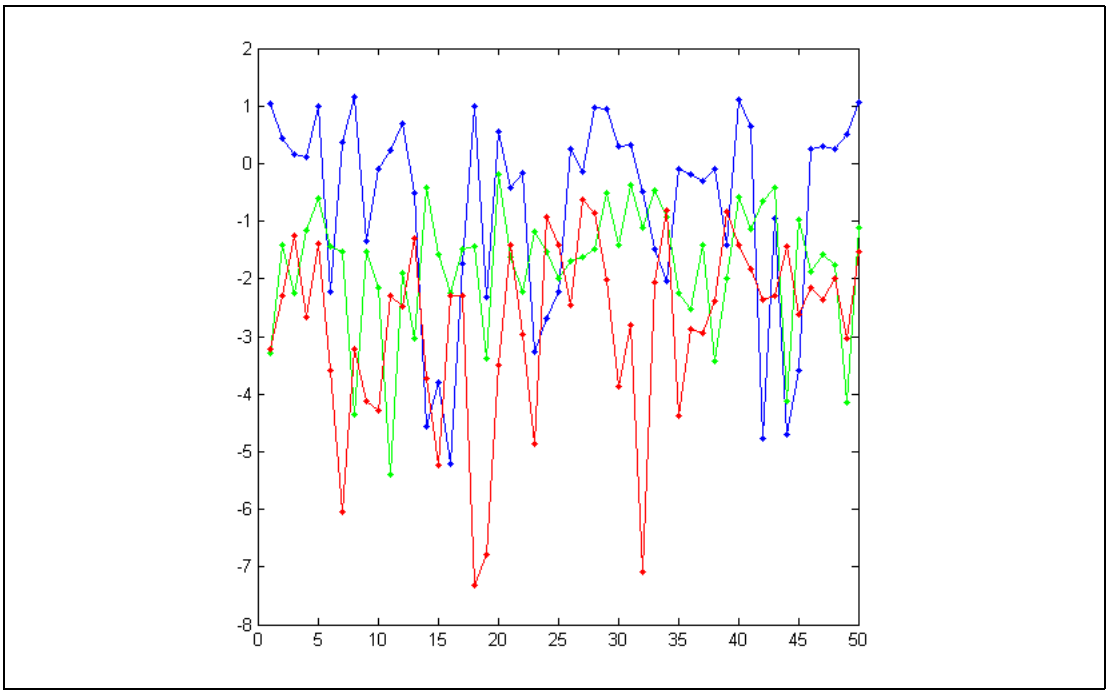

Figure 5-3 Fitness Scores for each class of Iris predicted by a mixture model

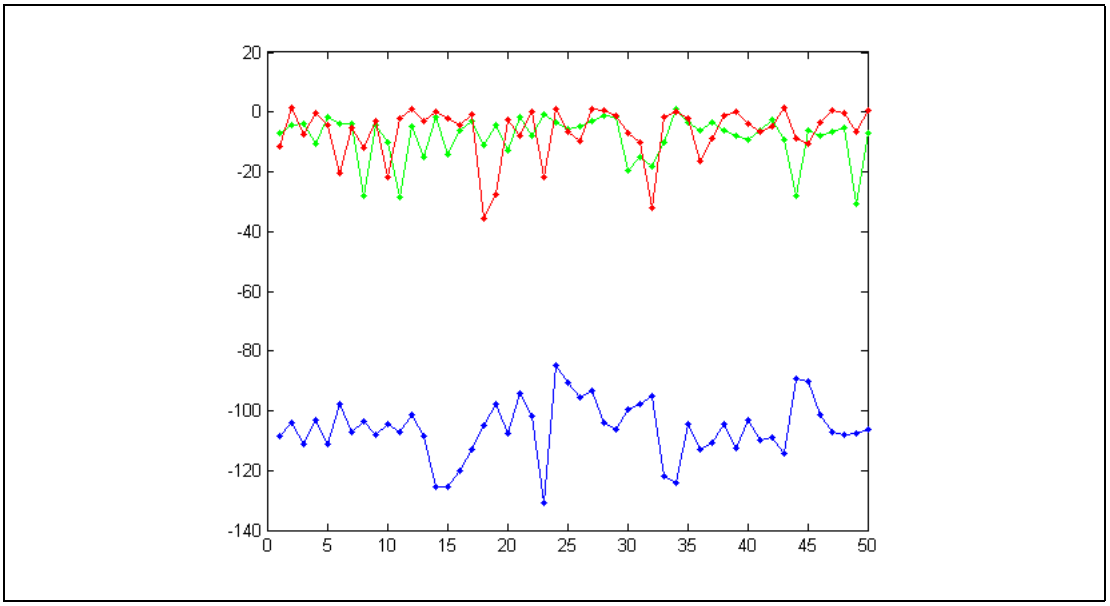

Figure 5-4 Fitness Scores for each class of Iris predicted by an adapted mixture model

5.3 Properties of the FS

The FS is the basic measure used to flag anomalies. It is a standard measure used in mixture modelling, but here it has a more general use. When analysing time histories with multiple variables, the FS can act as a fusion of the variables, reducing the multiple time histories to a single view. Whether or not such a fusion is appropriate depends on the type of information sought from the model. For anomaly modelling of time history data, interpreting the FS as a fused parameter is intuitively useful. However, there is a need to consider some of its properties to enable the correct interpretation of an FS time history.

Consider a univariate Gaussian centred at 0 with a variance of 1, and imagine that it represents the fleet statistics for a single indicator exhibiting normal behaviour. Suppose now there is a single gear that is trending away from normal space in a positive direction such that successive acquisitions are producing larger values. A linear trend plotted against time would appear as a straight line with the slope indicating the rate at which the trend is moving. The FS will start off high when the gear is within normal space. As the gear trends away the FS will fall. If another gear was to trend at the same rate but negatively, so that successive acquisitions are producing lower values, the FS would look exactly the same as for the first trend. Both trends differ in their direction but they are both moving away from normal space represented by the Gaussian. The rate of change of the FS will vary with distance from the centre of the Gaussian. So a linear trend in indicator space appears as a curve in FS space. This non-linear response of the FS is the result of the Gaussian model whose distribution determines that an object is increasingly looking less normal the further it moves away. If it moves too far, the FS may reach a limit after which it will stay flat even though the object may still be trending. An object that reaches the FS limit is outside of the model's response and clearly outlying. Such objects would at the very least be on 'close monitor' (i.e. being regularly checked) and it is still possible to track the trend using a more localised model.

Suppose now there is a multivariate Gaussian model built with three indicators. To keep things simple the mean and variance parameters for each indicator are the same. Now consider that a single component is trending in the same manner (same rate and starting state) across all three indicators. If the indicators are statistically independent, the likelihood score is the product of the likelihood score for each indicator (each individual score being the same in this example). The FS is the log of the likelihood score. This means that the FS for the three indicators combined is the same as a single indicator score weighted by 3. In other words, the combined FS trends three times the rate of any single indicator trend. The end result is that a gear with a trend across multiple indicators will be seen as more anomalous than a gear trending across fewer indicators. This response is desirable but care needs to be taken. It is not necessarily correct to emphasise an anomaly when indicators are highly correlated if this correlation is a result of them responding to the same signal characteristic.

5.4 Adapting the Cluster Space

If there are known outliers in the training data, as is the case with HUMS, a mixture model has to be adapted so that it does not fit the outlying data. The adaptation works by removing clusters. Building a standard mixture model and simply removing clusters is fraught with risk. To understand this it is necessary to consider the nature of the HUMS data being analysed.

In Section 4.5 it was seen that the shaft order energies have extremely long tails that contain a significant quantity of data. An opinion was not given at that time as to whether these data should be treated as outlying. Salzer (1994) [32] gives an example

of high variability in FSA_SO1 from a high speed input shaft fitted to an S61 and comments that this type of behaviour is seen occasionally on the S61 and less frequently on the Super Puma.

Knowledge of the existence of an unusual behaviour does not make the data normal and there is evidence to suggest these acquisitions should be treated as outliers. Many of these cases have very high energies and high variability that is random in nature. These patterns can exist for long periods of time and then disappear following some maintenance action (if they are due to instrumentation problems for example). The signals in such outlying states appear to carry no information other than there is something wrong in the acquisition. If these data are considered normal they could easily mask fault related trends. The decision has been to treat such data as outlying that require removal from the model. These data are so large in value that they can be seen as 'globally outlying': on the edge of the cluster space.

Global outliers are easy to spot. But there exist another type of outlier that is known as 'locally outlying' (see for example, Breunig et al., 2000 [6]). Local outliers are more hidden in the data and sit on the edge of 'local dense regions'. These local outliers will appear in HUMS training data whenever there are cases of unlabelled data that are associated with component degradation. The data may be unlabelled as fault related because either symptoms went undetected or components were rejected for some reason but their physical integrity remains unknown (because there was no feedback from strip reports etc).

Implementations of mixture modelling software have some method to estimate the optimum number of clusters. There are many techniques for doing this but all of them are heuristic rules that will fail under certain conditions (such as localized maxima for example). One commonly used heuristic is the BIC. The likelihood score for a model will increase as the number of clusters increases and so the BIC penalises this score on the basis of model complexity where complexity is judged by the number of clusters. Optimising the cluster space, when training with data that contain anomalies, is likely to mix anomalies with normal data and removal of such clusters would impact on the number of false positives/negatives. So there are two fundamental questions: first how is the number of clusters determined? And secondly how are clusters identified for removal?

The approach taken involves over specifying the number of clusters. This means that the BIC metric is not used directly but is considered for judging the number of clusters. The number of clusters also links directly to the criteria for removing clusters. A cluster's removal is determined by some information measure based on the cluster's contribution to the likelihood score.

Consider a single component type within the drive train such as the bevel gear. The time history of an indicator from an individual bevel gear can vary considerably; this variability is referred to as 'within component' variation. For a given indicator, there will be a great deal of variability between bevel gears fitted to different gearboxes; this is referred to as 'between component' variation. Every component is tagged with a component fit identity (ComponentFitID). A ComponentFitID identifies a specific component fitted to a specific gearbox fitted to a specific aircraft over a period of time. The ComponentFitID is used during the construction of anomaly models. Anomaly models are built using only one type of component: in other words, there will be a model for the bevel pinion, another model for the bevel gear and so on. The ComponentFitID is used either indirectly or directly in the build process.

The indirect use of the ComponentFitID is fairly straightforward in that it does not directly impact the structure of the mixture model during training. Once the mixture model is built statistics about the ComponentFitID's are gathered for each cluster.

These statistics measure the association between ComponentFitID's and clusters. The result is information about a cluster's generic capacity: its representation of the fleet of components (i.e., if a cluster represents a number of different components then it has high generic capacity whereas a cluster relating to a single component has low generic capacity). This generic information measure varies with the number of clusters but it is possible to define a heuristic measure, much like the BIC, that strikes a balance between the number of clusters and the generic information content. Clusters with low generic capacity are removed. One measure of generic capacity is based on normalised Entropy[xvi]. Entropy provides a measure of uncertainty. To understand how entropy can be applied to a mixture model it is first necessary to be reminded of how a mixture model can be used to simulate HUMS data. A cluster can be selected at random according to the cluster prior distribution and a set of HUMS indicators generated according to the selected cluster's probability distribution. A single generated set of indicators represents a single acquisition. Information relating to the ComponentFitID contained in the generating cluster can be used to tag the simulated acquisition with a component identity. This tag will be a multinomial distribution in that there will be a probability of association with each ComponentFitID (the simulated acquisition could have come from a number of different components). If a cluster has zero entropy, then it is possible to be certain about the ComponentFitID of the simulated acquisition. Higher entropy values mean more uncertainty. Intuitively, high uncertainty means a higher generic capacity which means the cluster is more likely to represent normal behaviour.

The direct use of the ComponentFitID is more complicated in that the mixture model is conditioned on the ComponentFitID. Basically, the ProDAPS cluster algorithm is used to build a mixture model that contains information localised to a ComponentFitID. ProDAPS will then gather statistics on a generated model to assess the generic capacity of each cluster. The direct approach does not use Entropy. Instead the measure is based on a normalised likelihood measure that is computed using cross-validation (a statistical practice of partitioning data into subsets for training and validation). After the generic capacity for each cluster is scored, the clusters can be ranked as candidates for removal, the number of removals being determined by a threshold called the 'anomaly tuning parameter'. The anomaly tuning parameter is a continuous value that specifies the maximum difference in the normalised likelihood score between the original model (no clusters removed) and the new model (with clusters removed). This form of modelling can suffer local maxima but these are easy to detect via the scoring process.

While the indirect modelling approach can provide useful results, the adaptation tends to be swamped by global outliers and it can fail to detect local outlying regions. All of the models, therefore, are built using the direct approach.

The general form of the modelling equations can be found in Appendix B.

5.5 Setting Alert Thresholds

Alert thresholds are determined from the FS. Each model has an alert threshold and thresholds vary from one model to the next. The thresholds are determined automatically and the number of alerts generated will vary by component type. This is a desirable effect because some components may be more difficult to reliably monitor and some are more susceptible to defects.

There is a sense of 'overkill' in the current method for setting thresholds in that the FS is a measure from a probability distribution, but a secondary probability distribution is then derived based on the FSs for the fleet of training components. The mathematical property of the FS results in a one-sided distribution. The approach is to fit a probability distribution to the FSs and set a threshold on this secondary

distribution. It turns out that a Generalized Extreme Value (GEV)[xvii] distribution provides a good fit. The FSs are first transformed. The extreme tail of the FS distribution is sparsely populated and so the end of the tail is chopped to provide a better fit (the chopped area includes definite anomalies). A GEV distribution is then fitted after which the FS corresponding to n^{th} cumulative percentile (typically 0.995) is determined. This score is the alert threshold.

5.6 Data Pre-processing

The anomaly detector is not some magical black box at which data can be thrown and out pops sensible anomalies. The anomaly models will detect behaviour that is unusual but they know nothing about the nature of the particular behaviour that the HUM system should detect. The data have to be pre-processed to reveal the characteristic behaviour that requires modelling.

There are two types of behaviour that have been modelled. The first type is addressed by modelling the between component variation and such models are referred to as 'absolute models'. Each component tends to occupy its own level of vibration energy. Even though a large variation in vibration between components is expected, there will be a level above which a component's vibration appears so unusual that it should be flagged as anomalous. The absolute model is designed to detect such cases. The second type of behaviour concerns within component variation and such models are referred to as 'trend models'. As the name implies, these models are designed to detect significant trends. The notion of a significant trend is important because trends are not uncommon in HUMS.

The pre-processing for the absolute models is minimal and only requires the application of a two stage filter to each indicator: the first stage removes unreasonable values and the second stage applies a median filter to remove up to two successive spikes in the time series data.

There were various candidate methods considered for the trend pre-processing involving things like segmentation followed by linear regression fits: basically the time history for an indicator is segmented into sections and a line is fitted to each section. The idea of the trend pre-processing is to detect the relative change in an indicator and provide a measure that normalises out the between component variation. The ideal scenario would use an independent signal source against which data could be normalised but this does not exist. Others have also explored techniques for describing trends in HUMS indicators. Wigg et al. (2006) [43] have explored Sigmoid fits and Wavelet transformations. However, determining a satisfactory solution would require much empirical investigation and so currently a simple approach is used. For this, the two stage filter used for the absolute model is applied, and then a 'moving median difference' filter is used. Following each new acquisition, the median of the time history is re-calculated and subtracted from the newly acquired value to provide a normalised value. The objective of this technique is to reduce the impact of early post installation trends, and also small step changes due to maintenance, because the normalised value will gradually recover back to the median base line level. There is some distortion therefore of the time-history but in practice this simple approach appears to work quite well. Its weakness is that, while it achieves the normalization objective, it acts as a difference calculator and cannot distinguish between step changes, short duration changes, and slow or fast developing trends. There is therefore a potential level of information missing and it is planned to re-visit the trend pre-processing should the opportunity for further development exist. The pre-processing is illustrated in Figure 5-5.

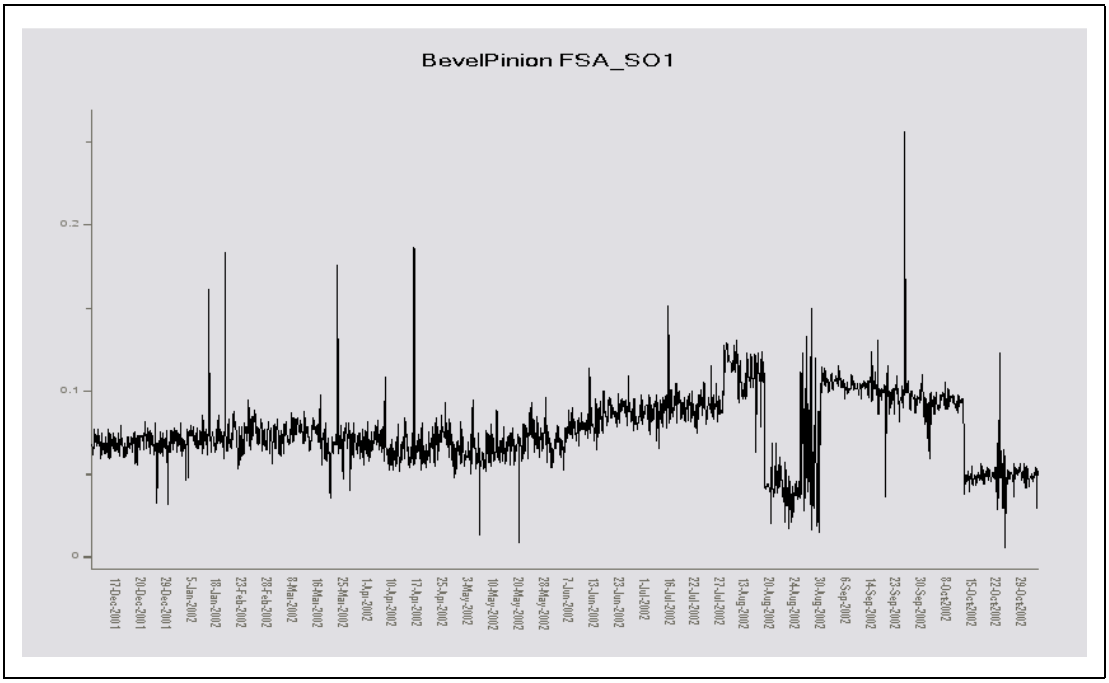

Figure 5-5a FSA_SO1 Condition Indicator for a specific gearbox before pre-processing

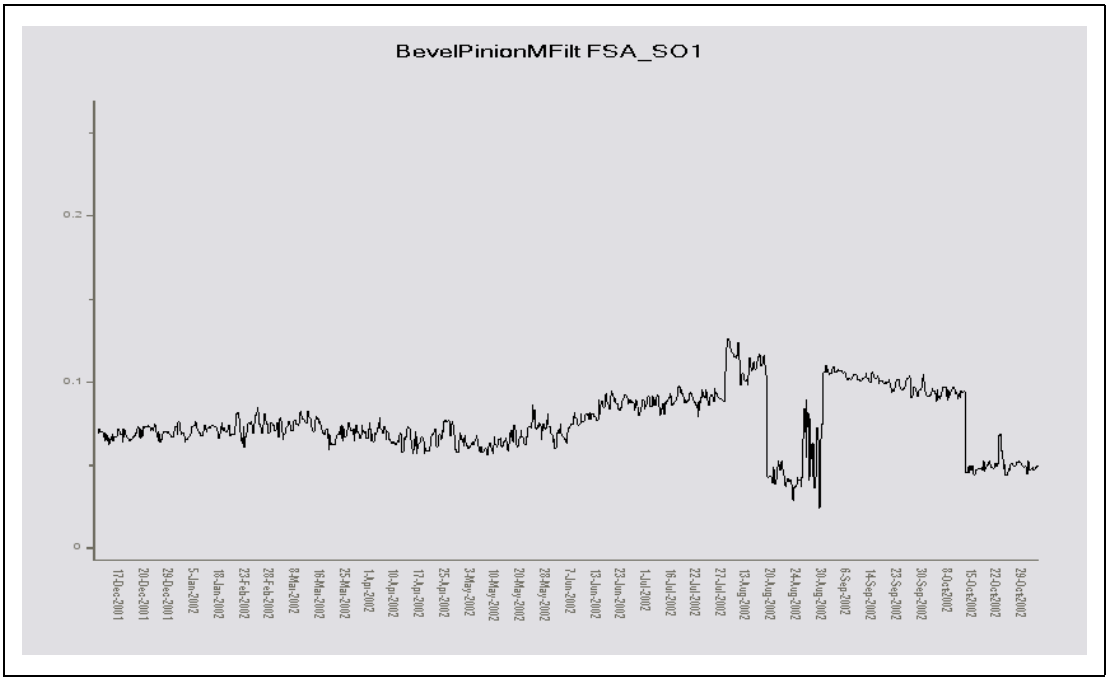

Figure 5-5b FSA_SO1 Condition Indicator after first stage pre-processing. Spikes have been removed

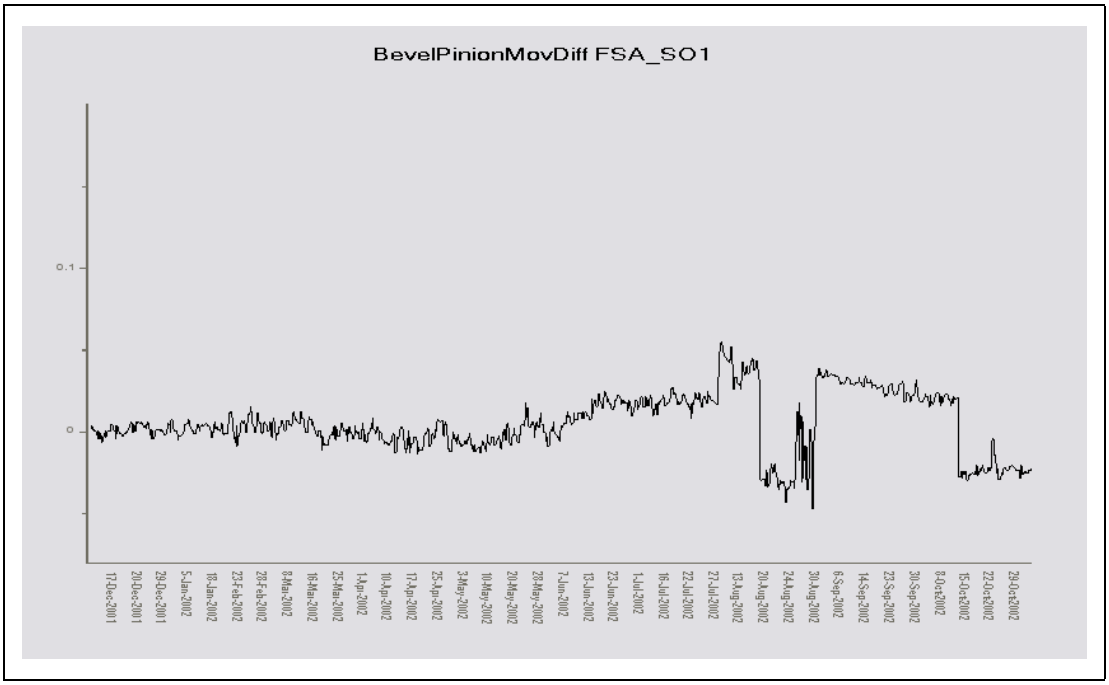

Figure 5-5c FSA_SO1 Condition Indicator after second stage pre-processing (for trend modelling). The signal shape is close to that in Figure 5-5b but the signal has been translated to give an approximate median value of 0.

5.7 Definition of Anomaly Models

Every component is represented by four anomaly models. These models differ according to the indicators used and the pre-processing. The indicators are split into two sets: set 1 is referred to as the '8-indicator' models and comprises {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD} and set 2 is referred to as the M6 models and comprises {ESA_M6*, ESA_WEA}. The indicators are split to reduce the potential for one indicator to be masked by another; the criterion for this split is from engineering knowledge of the indicators. Absolute and trend models are built for each set of indicators. The oil cooler fan and number 1 bearing are exceptions in that they contain single indicator sets comprising {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}. In summary, for each component there will be an: '8-indicator' absolute model, '8-indicator' trend model, M6* absolute model and M6* trend model.

6 Development of Anomaly Detection Software

This section describes the software built to perform the off-line analysis and the software built to execute the live trial.

6.1 Data Analysis Software

The analysis software is split into two components. The first component defines the anomaly model building and anomaly detection methods. The second component implements batch task processing to make the modelling process efficient. A task defines a how a model is built.

6.1.1 Anomaly Modelling Tool

The anomaly modelling tool is a component that is dependent upon the ProDAPS data mining tool. This component is completely separate from the data mining tool in that the data mining tool has no knowledge of the anomaly modelling component. The anomaly modelling tool sends requests to the data mining tool. These requests include the construction of a mixture model, model predictions to score the clusters for model adapting, and model predictions to process flights and produce FSs.

The model construction request contains data that specifies the data source, choice of indicators, mixture model build parameters, model naming and model storage location. For scoring the cluster space, the interaction with the data mining tool is at a low-level in that the communication takes place case by case.

The general schema for the anomaly modelling tool is shown in Figure 6-1.

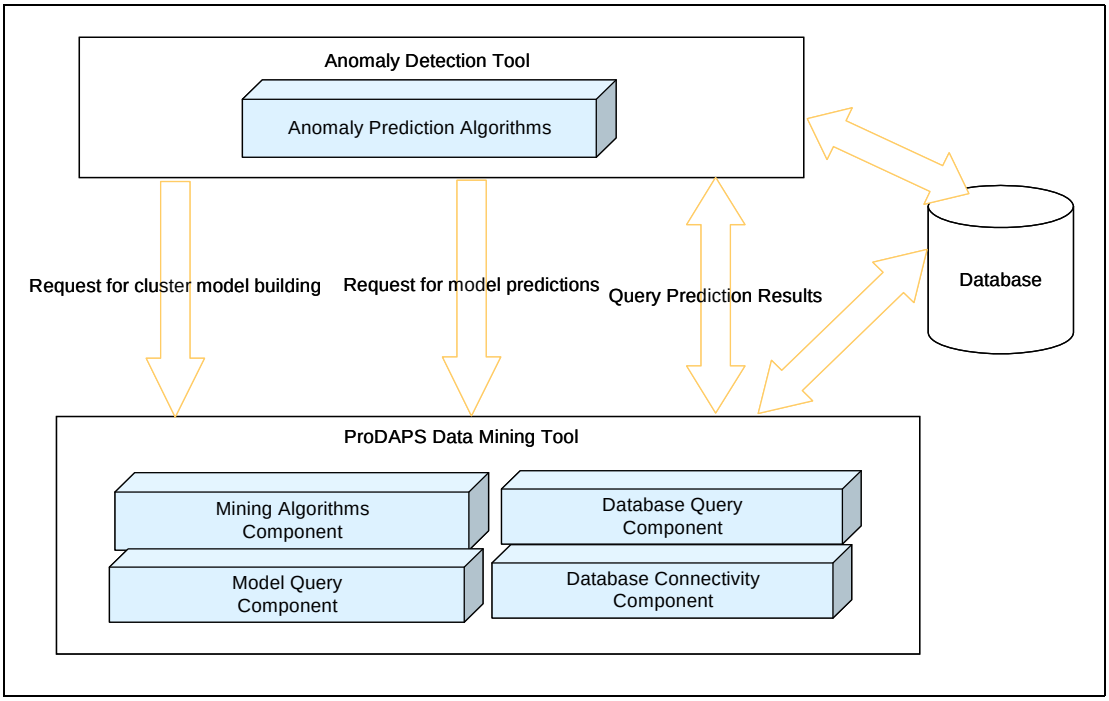

Figure 6-1 Anomaly detection tool's interface to the ProDAPS data mining tool

6.1.2 Model Building

There are 136 anomaly detection models for this programme. Model building is split into three tasks: build a mixture model, score the clusters, and adapt the model. A task is a script that states how the task is to be executed. The scoring process is the most complex and computationally intensive task. Manual control of this whole process is tedious, very time consuming to repeat and is prone to error and so a tool

was implemented to automate the process. This tool is used to configure and execute tasks. The tasks are batch processed. Batches of tasks can be distributed to different machines for processing. Both the off-line analysis and in-service trial would have been restricted to a subset of drive train components if this level of automation did not exist.

A picture of the graphical user interface (GUI) for this tool is shown in Figure 6-2. For a class of model, such as the trend models for the MRGB, many of the parameters in a task script are repeated. The GUI is split into a number of window panes. Some panes allow the definition of parameter values common to a class of model; other panes define model specific parameter values. This provides a way to simultaneously generate multiple task scripts.

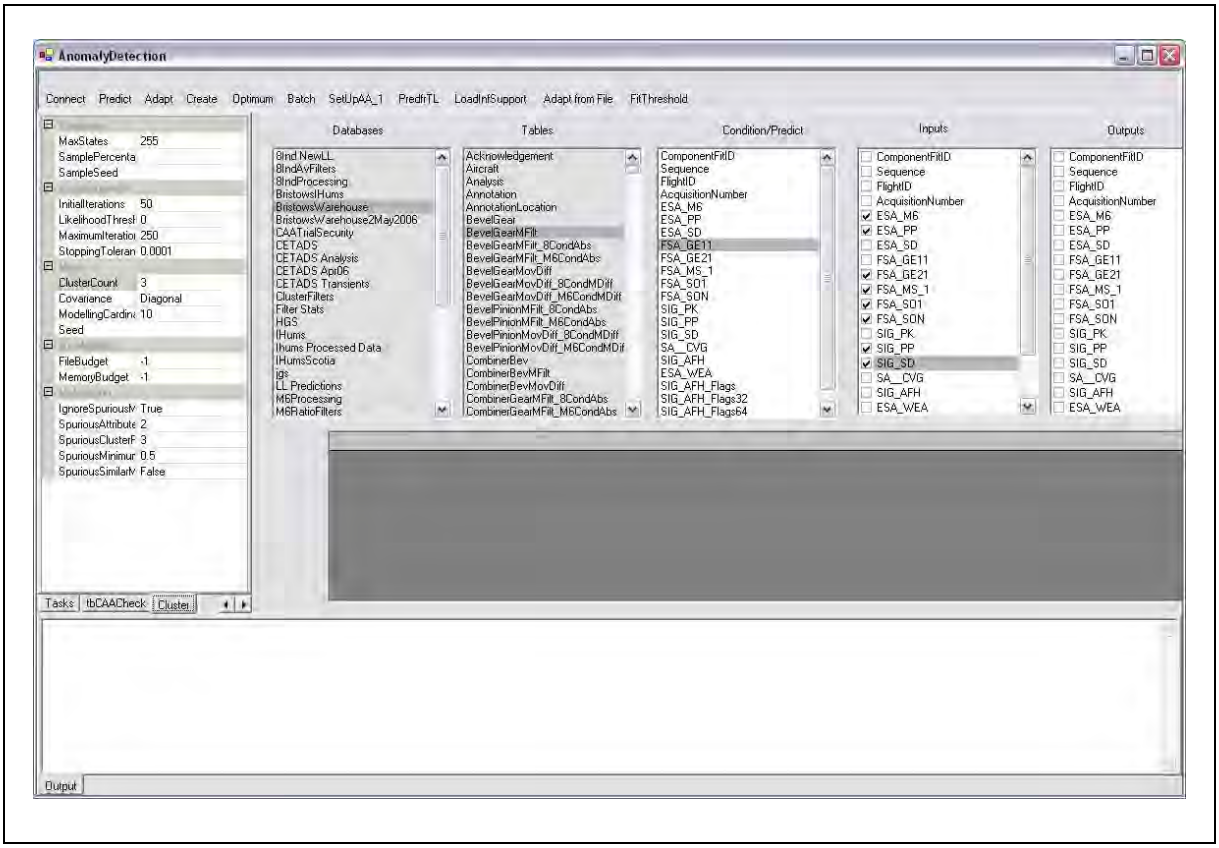

Figure 6-2 User interface for model task construction and execution

6.2 Anomaly Detection System Software

The original intention for the in-service trial was to provide client software at the operator's (BHL) base. There was concern though regarding maintenance of trial software located with the customer. Because the output from this programme is not a production system there were bound to be teething problems and it was possible to envisage a need to provide model updates after the in-service trial had started. It was deemed preferable to have the anomaly detector sited at GE Aviation in Southampton and to provide the operator with some form of networked access. Emulation software was considered, where the operator would remotely drive a conventional windows application that resided on a machine in Southampton. This option was easy from a software implementation perspective but was considered a second choice to the more elegant solution of a web based server. The final choice was for a web based solution for a number of reasons:

- Software could easily be maintained at Southampton and would save cost in the long term.
- Engineers in Southampton could see what the operator was seeing. This meant that the process of evaluating the trial could be more robust as it would enable regular dialog between the operator and developers in Southampton.
- The database at Southampton would continue to be populated with new acquisitions. This would aid future development and would provide long term persistence of historical data (IHUMS starts to overwrite its binary files data once the acquisitions exceed 2,500).

Although a web based server was the best solution for technical and support reasons the work involved would exceed that originally scoped for the programme so GE Aviation took a decision to fund additional development costs. The two main risks with this solution were reliability of the data download from Aberdeen, and the speed of response to the user when plotting charts. The only way to check the data download from Aberdeen was to run a sequence of tests that would last for one week. These tests were to check unknowns such as available bandwidth and procedures to pass through internet firewalls. The main concern with the user's experience was the speed of response when plotting charts and interacting with the chart. A decision was made to do all intensive processing on the server and to upload charts to the client (at Aberdeen) as images. A few tests with some dummy data and prototyped screens confirmed this approach provided acceptable user interaction.

Figure 6-3 shows the data flow between Aberdeen and Southampton. HUMS data are automatically transferred overnight from Bristow's secure Web Portal, imported into the anomaly detection system's data warehouse and analysed. Bristow have a remote secure login to the system to view results at any time via a web browser.

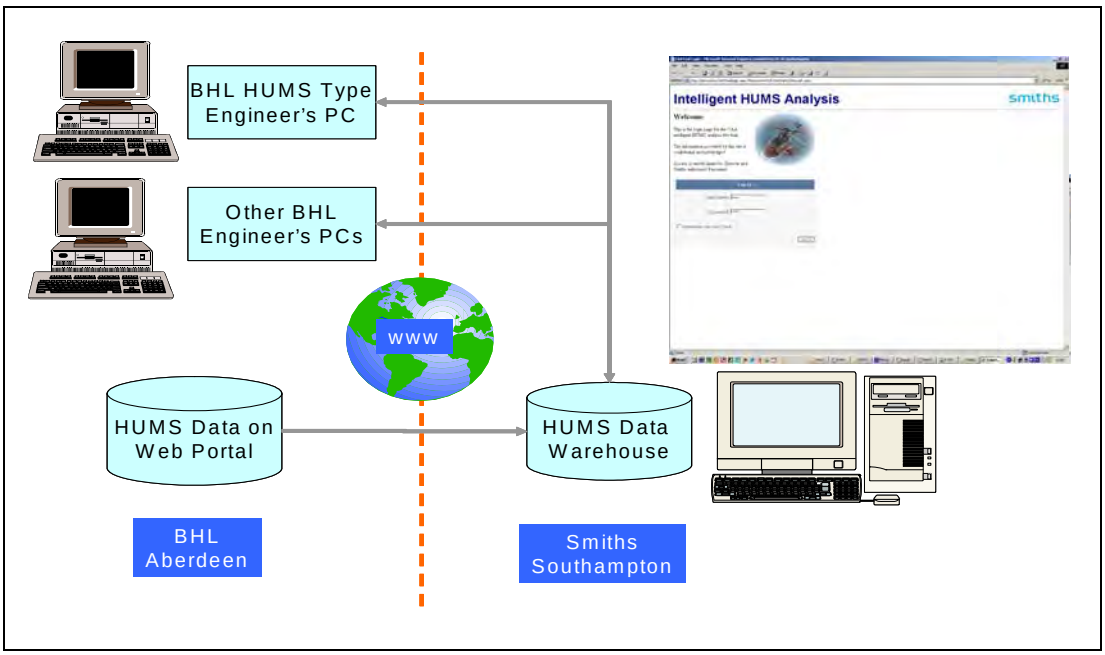

Figure 6-3 Data flow between Aberdeen and Southampton

After logging in, the user is presented with a fleet view showing aircraft, the number of alerts (totals, current, and unacknowledged), and the number of notes. The user has the facility to enter notes against an acquisition. The notes are useful for tracking findings during the trial and are used by the operator to record such things as maintenance actions or observations with their operational HUMS. The user can limit

their view of the data to one week, one month, three months or the complete history. There is also a calendar that allows the data to be viewed as it was at some previous date. This option is necessary to access data relating to components that have been replaced because the web server's 'latest date' view, which is the default, focuses on those components that are currently fitted to the fleet. The user can drill down via an aircraft tail number, by notes or type of alert. At the component level, such as bevel gear, the FS time histories can be viewed. These traces are annotated with alerts and any notes the operator has made. The user can then drill down to either the filtered or pre-filtered CIs. The indicator plots show, for those indicators used to build models, the mean, and upper/lower three standard deviation bands that have been derived from the anomaly model. These bands are different to those that would be computed from the fleet data because the anomaly models are adapted to reduce the impact of outlying data. Some screen plots of the web system are shown in Figure 6-4.

Figure 6-4 A selection of screen shots from the web based anomaly detection trial system

7 Results of the Off-line Analysis and Demonstration

Four models were built for each of the 35 gearbox components monitored (the exceptions being the oil cooler fan and number 1 bearing which have two models). These include the '8-indicator' absolute and trend models, and the M6* absolute and trend models as detailed in Section 5.7. Table 7-1 lists the number of gearboxes and acquisitions used for training the anomaly models. The number of gearboxes and acquisitions differ because the life of an assembly in an aircraft varies and the choice of training data is restricted to cases whose gearbox/aircraft association can be verified and the gearbox/aircraft fit has 20 or more acquisitions. Table 7-2 lists the number of validation gearboxes. The number of test gearboxes continually grows. For example the number of MRGBs is currently 119.

Table 7-1 Number of gearboxes used for training

Assembly	Number of gearboxes	Number of acquisitions
MRGB	60	70343
IGB	59	100256
TGB	58	89908
LHAGB	118	52804
RHAGB	118	52246

Table 7-2 Number of gearboxes used for validation

Assembly	Number of gearboxes
MRGB	28
IGB	23
TGB	49
LHAGB	45
RHAGB	16

7.1 Anomaly Model Assessment

The anomaly modelling technique has been verified on synthetic data where there is complete knowledge of the data distribution and anomalous behaviour.

For the IHUMS data a laborious process of cross checking FS traces with IHUMS indicator traces was undertaken. This confirmed that any response (lack of fit) in the FS was due to one of more indicators changing behaviour. The reverse procedure was also carried out to confirm that significant indicator trends were reflected in the FS traces. The charting tool, developed because of this program, assisted enormously in this task because it provides efficient charting of large quantities of data and it is easy to compare a component with the fleet.

Global descriptive fleet statistics were also compared with those generated by the anomaly models. The anomaly models produce different statistics because they are

adapted to reduce the impact of outlying data. These statistics reveal the support (number of cases) for the normal model of behaviour (as defined by an anomaly model) and give an indication of the number of training cases that are considered to be outliers. The statistics also reveal whether the adaptation is driven by multiple indicators.

Tests with synthetic datasets validate that the anomaly modelling behaves as intended. The charting and descriptive statistics give assurance that the anomaly modelling is doing something sensible with HUMS data. Validating the anomaly approach for HUMS requires testing with new data. The bevel pinion fault case is a rare opportunity to test against a known in-service fault. This case is discussed next. Other cases will be documented as the in-service trial progresses. Statistics regarding alert frequencies will be collated. This case portfolio has started and three new cases are listed in Section 7.4.

7.2 Results of the Analysis of the AS332L Bevel Pinion Fault HUMS Data

7.2.1 Fault Detection

The significance of the bevel pinion fault case was stated in the introduction to this report. For this case, which had a large crack in the bevel pinion (see Figure 7-1), a number of indicators showed trends as evident in Figure 7-2. But trends are quite common in HUMS data and no indicator on its own reveals any significant outlying behaviour when compared to the fleet. Figure 7-3 plots the trend CI time histories for both training and validation data. Each gearbox is coloured differently and its number of points depends on the length of time the component was fitted to the aircraft. The three standard deviation bands shown are computed direct from the database (not the anomaly model). These bands do not represent HUMS alert thresholds. The fault case is labelled 999 and is annotated on those plots where it tends to occupy the tail end of the distribution and is visible on the plot. It should be remembered that many outliers have already been removed in the two stage filtering process. This fault case did not trigger any IHUMS alerts.

Figure 7-1 Picture of the bevel pinion with a large visible crack

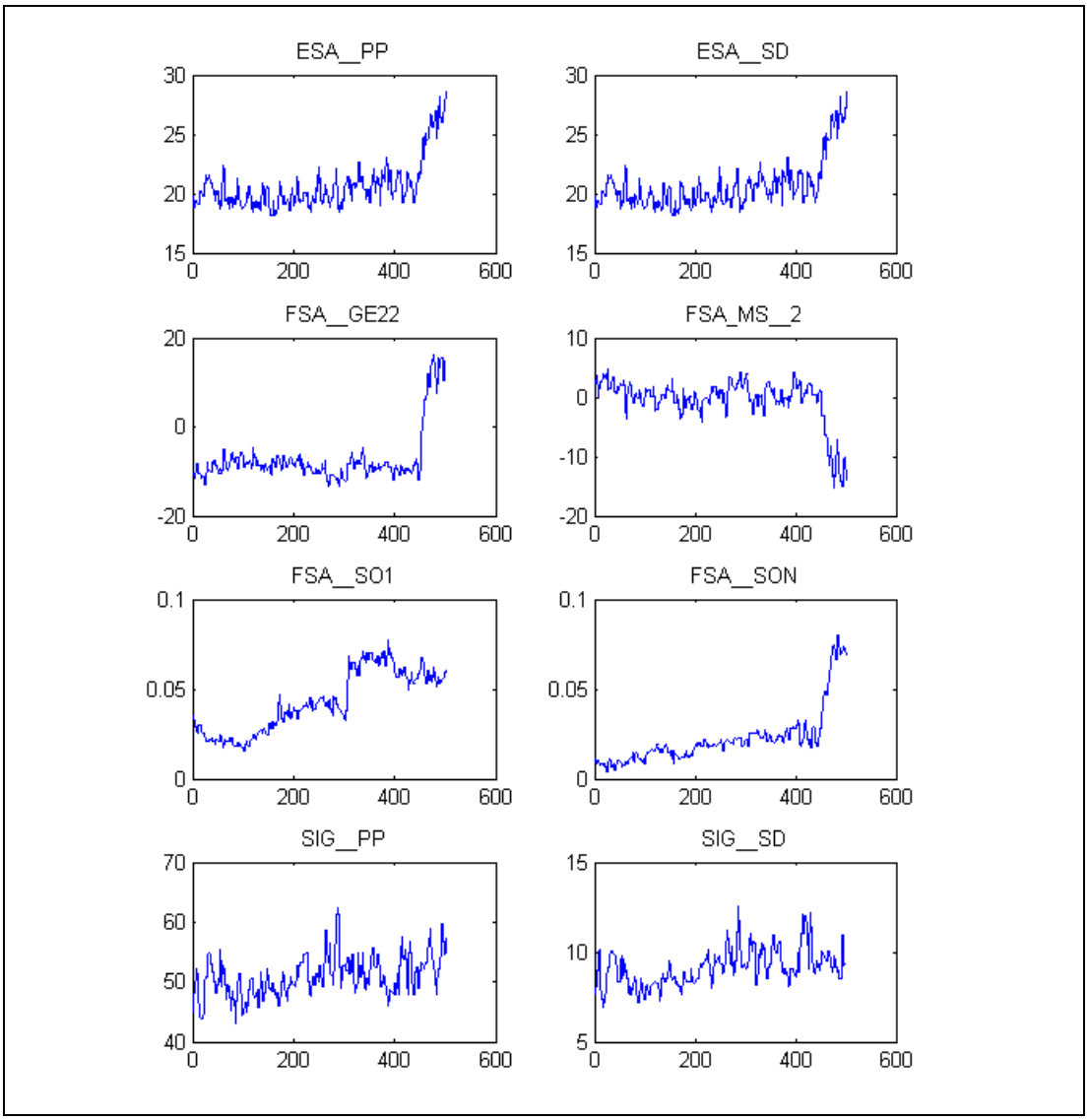

Figure 7-2 Absolute filtered indicator trends for the bevel pinion fault case

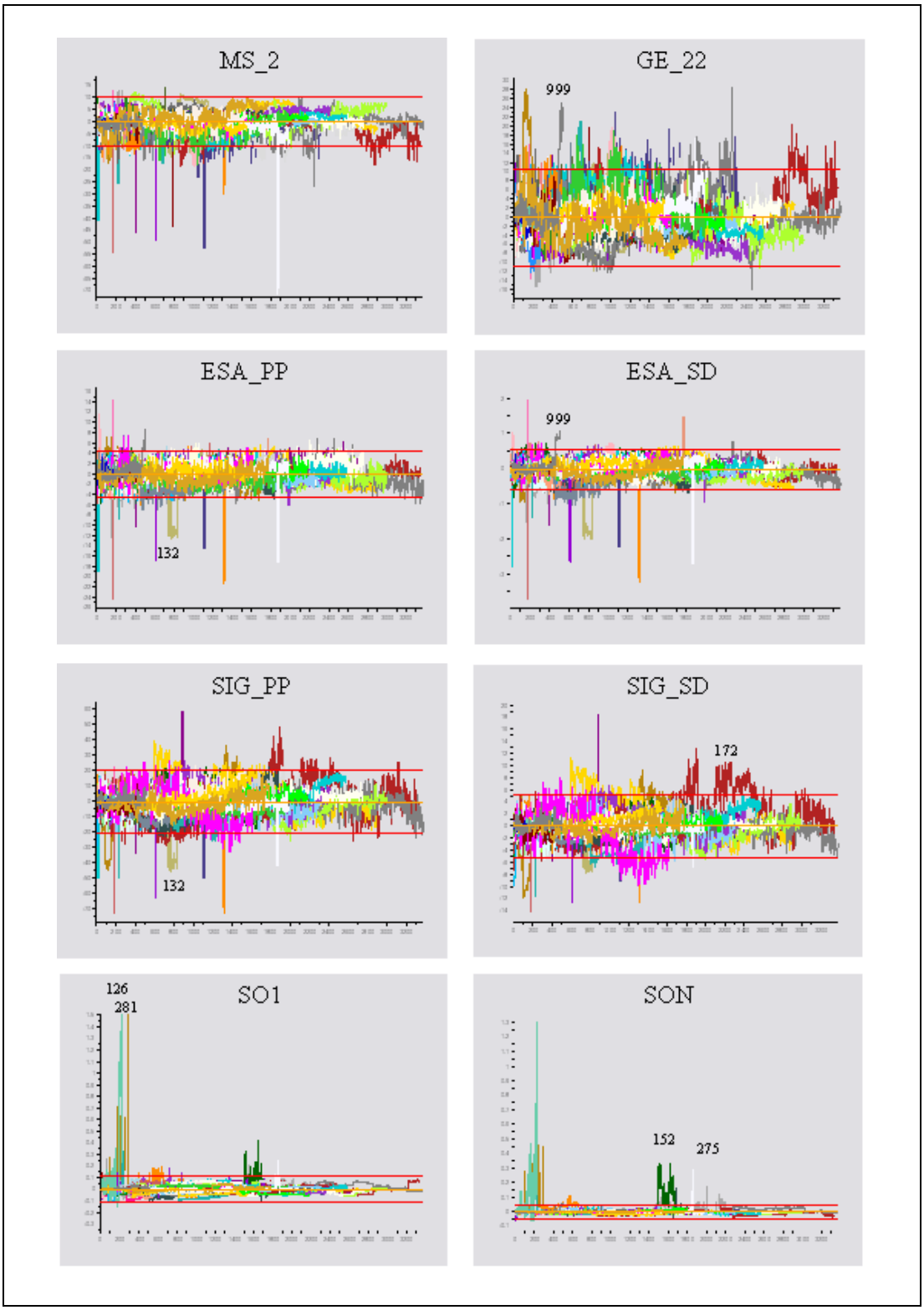

Figure 7-3 Fleet trend indicator plots for the bevel pinion from the complete training and validation data

Figure 7-4 shows the trend FS traces for the training and validation data. The fault case was not used for training but was included as part of the validation set. There are some gearboxes with very low FSs and these are annotated with their component fit identities as 126, 281, 132, 152 and 275. These data have unbelievable values and are indicative of sensor problems. One of the gearboxes was rejected because the

resident HUMS expert was not available to consult on the cause of the high vibration indication. The gearbox only had 100 hours remaining life and while in this instance the rejection was not costly, the anomaly detector would have confirmed a sensor fault. Sensor (or instrumentation) issues tend to show as global outliers with high variability. This global outlying behaviour is evident in the shaft-order indicator plots shown in Figure 7-3. Gearbox 172 has FSs in alert with a variable behaviour pattern. It was discovered however that these were caused by a series of maintenance actions.

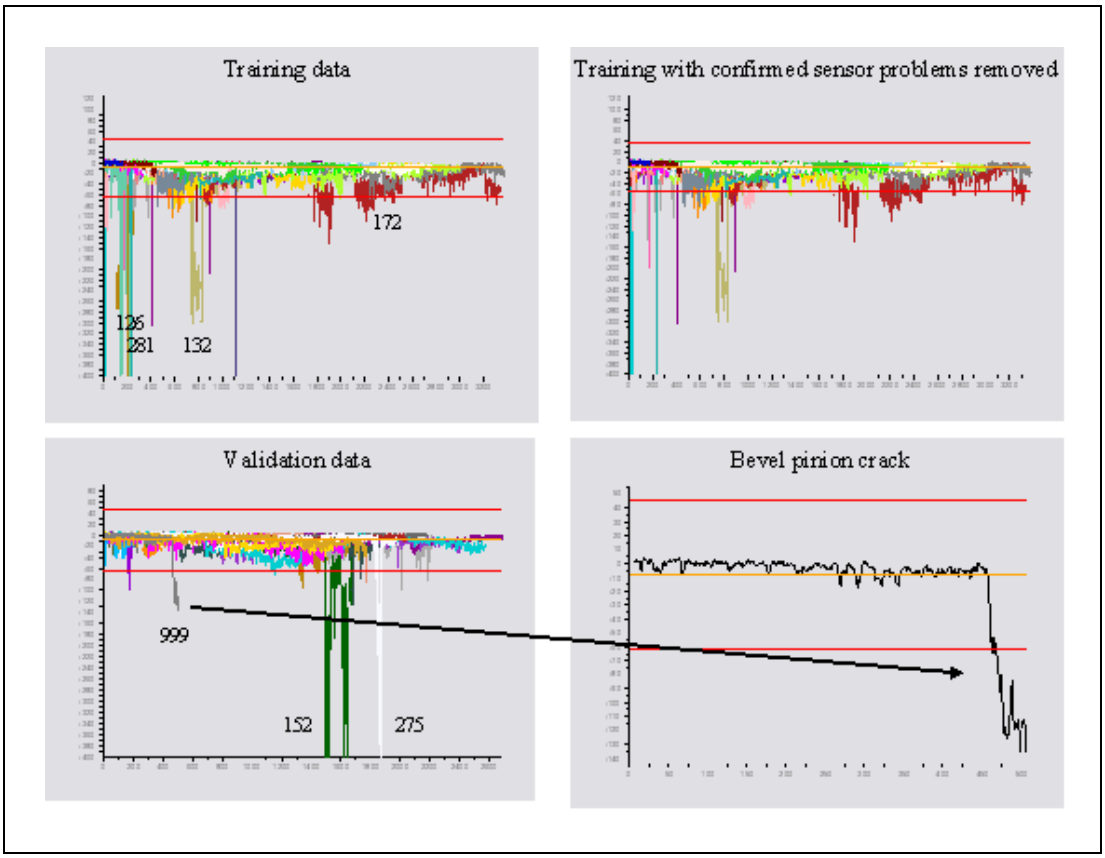

Figure 7-4 Trend Fitness Scores for the bevel pinion. The fault case is tagged 999 and its Fitness Score is also shown separately (bottom right). The trend clearly mimics the indicators and it exceeds the threshold many hours before being rejected for metal contamination.

The FS traces in Figure 7-4 are summarising a large quantity of data. Each FS is a fusion of eight indicators representing almost half a million points (if each indicator were plotted separately). Care needs to be taken when viewing fleet data because the distribution tends to look uniform when in reality it is very sparse below the alert threshold as shown in Figure 7-5. It is necessary to remember that fleet data plots only provide a crude overview of the data and are unreliable for assessing the significance of outlying data. The two key features of the FS trace are the score's position relative to the threshold and the shape of the trace. As a rule of thumb, those traces with values below the threshold that have been trending are of concern because these are more likely to be component related.

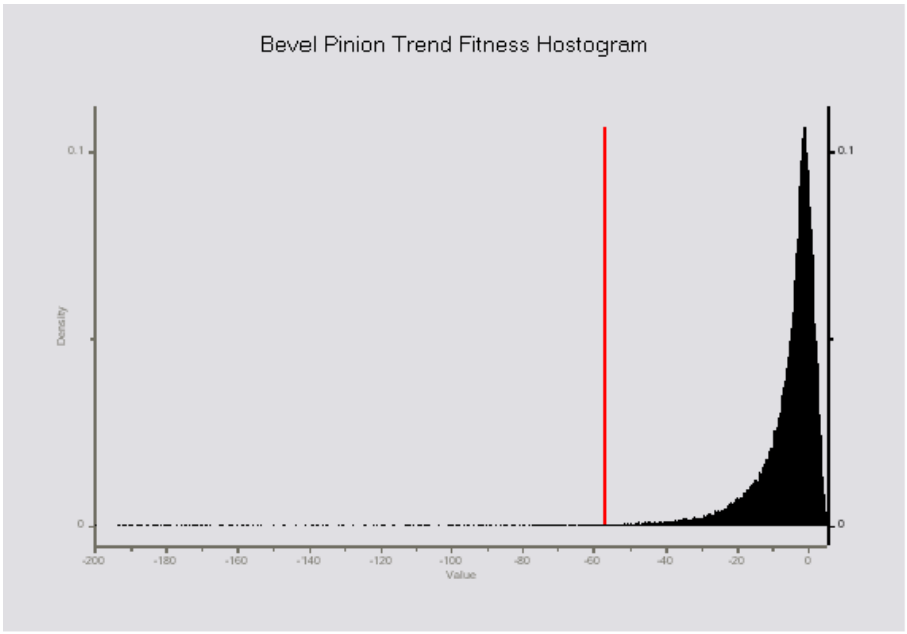

Figure 7-5 Fleet trend Fitness Score histogram for the bevel pinion. Note that the data are sparse around the threshold and below.

Examination of the fault case, flagged as 999, reveals anomalous behaviour that would cause concern. There is a clear and fast developing trend. The trend is clearly outlying and shows no features suggesting an instrumentation issue. Concern over this case is further emphasised when examining the FS trace before the trend develops. Before the start of the trend the FS is very high and stable. So a very normal looking component quickly develops a highly abnormal behaviour. It is not uncommon for some component types to exhibit trends but these often appear on installation. Because this component has been installed for some time, and these components do not generally exhibit fast developing trends, the anomaly alerts associated with 999 are highly significant and suggest a fault is present.

7.2.2 Information Extraction

ProDAPS provides a number of unique prediction capabilities and one type of prediction produces what are called 'Influence Traces'. These traces show which indicators are contributing most to the FS. There are a number of candidate techniques for producing influence traces but the outputs from a method that produces easy to interpret results are presented here. The influence traces are normalised so that they sum to 1 at the point corresponding to the minimum value of the FS. There are a number of informative statistics associated with these traces but for now the only item of interest is the overall picture they describe relating to the bevel pinion fault case.

Figure 7-6 shows the FS trace for the fault case from the absolute '8-indicator' model. Only a handful of acquisitions exceed the alert threshold but it is interesting to note that the trend is clearly visible in the plot. This shows the data are trending towards a low density region. Figure 7-7 shows the influence traces corresponding to the FS trace in Figure 7-6. In this type of plot, it is not unusual for most values to sit around zero because the whole plot is normalised relative to the lowest FS. The traces show that FSA_GE22 and FSA_MS_2 are globally outlying at the point where the FS exceeds the threshold. The other indicators are not significant in a global outlying sense.

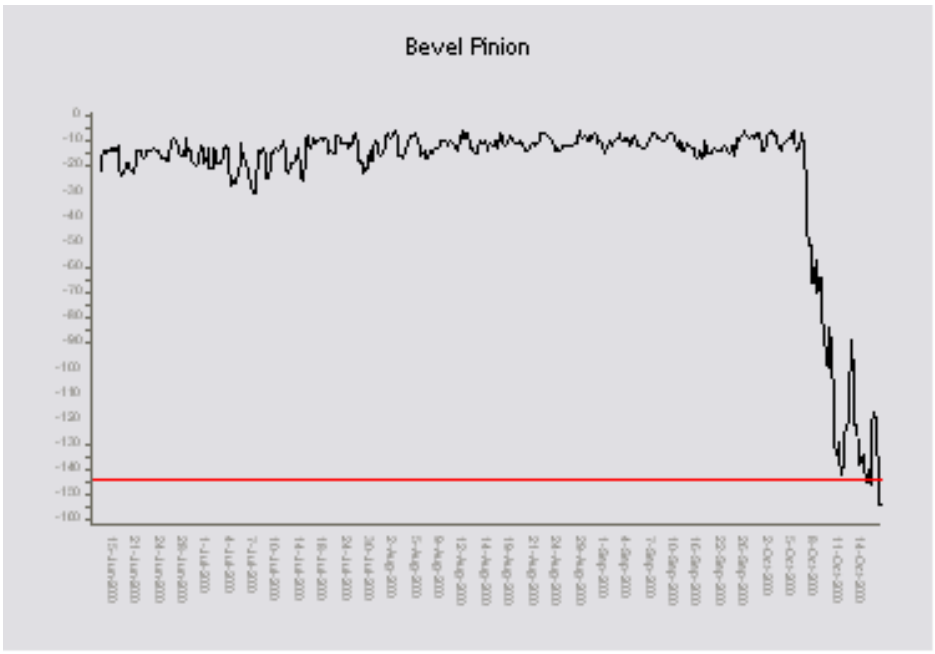

Figure 7-6 Absolute model Fitness Score trace for the bevel pinion fault case

Figure 7-8 shows the influence traces corresponding to the FS trace from the trend model included in Figure 7-4. Additional indicators now have an influence and in particular FSA_SON. The significance of FSA_SON is not visible in the CI trend. Although this indicator clearly trends it looks very benign in the context of the fleet. It is also possible to see the risk with existing HUMS of misinterpreting a trend's significance when bad data are allowed to pollute alert setting thresholds. The indicator traces for FSA_SON in Figure 7-3 show a significant quantity of outlying data that would mask this trend if the anomaly modelling did not remove its influence. But existing HUMS alerts will be affected by these data since these data are not flagged as unusual. Discounting any indicator's influence due to masking is a concern because it throws away potential diagnostic information and it de-emphasises the significance of a trend. It is known from the analysis presented in Section 5 that ESA_PP and ESA_SD are correlated and so are FSA_GE22 and FSA_MS_2, thus emphasising the need to retain all significant features of the data.

The influence of FSA_MS_2 and FSA_SON suggest that there is damage local to the gear (as opposed to responding to damage remote from the gear). This analysis of the fault provides evidence that corroborates with the discovery of a crack in this component.

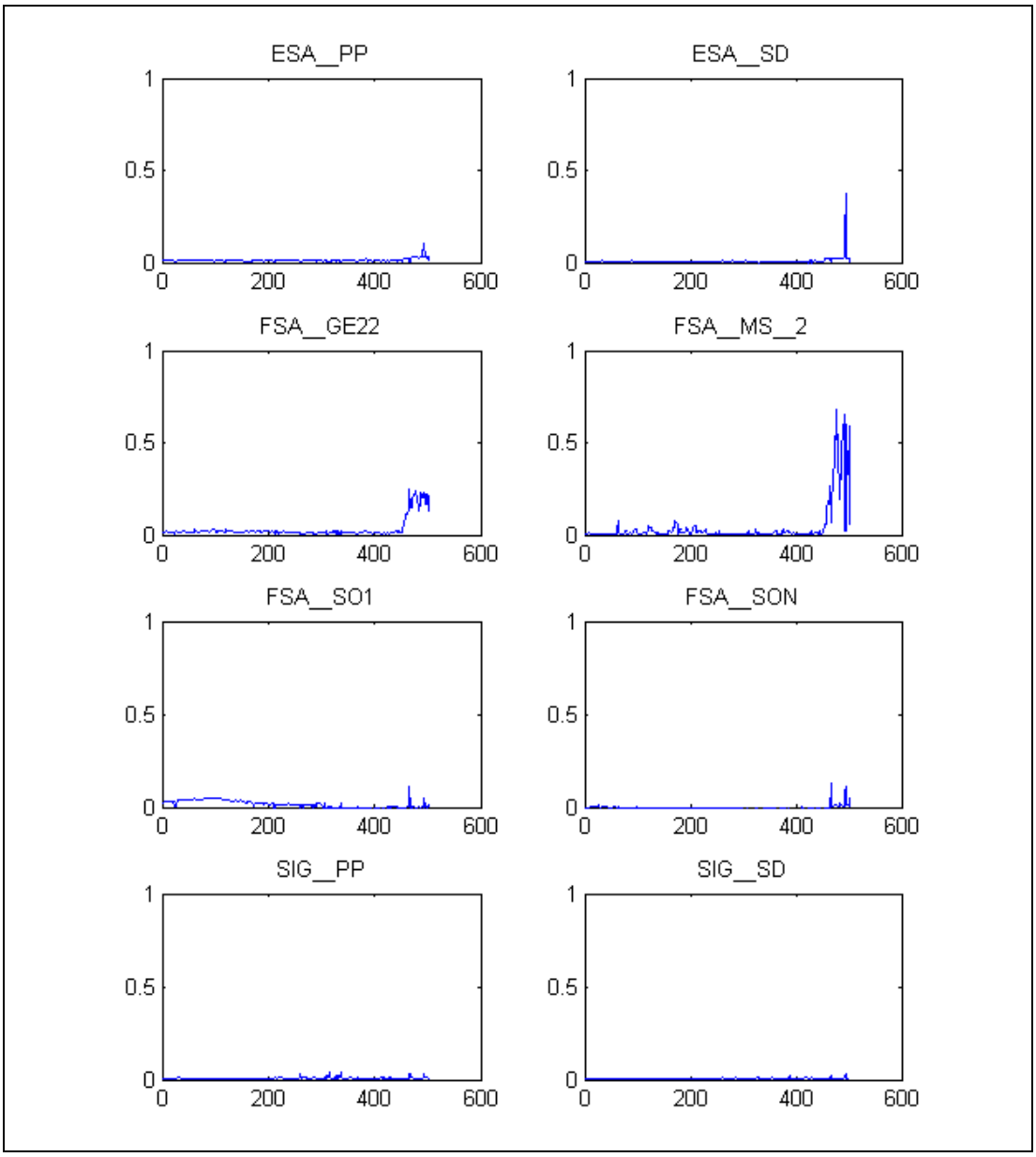

Figure 7-7 Influence traces for the bevel pinion fault derived from the absolute model. The significant features are FSA_GE22 and FSA_MS_2.

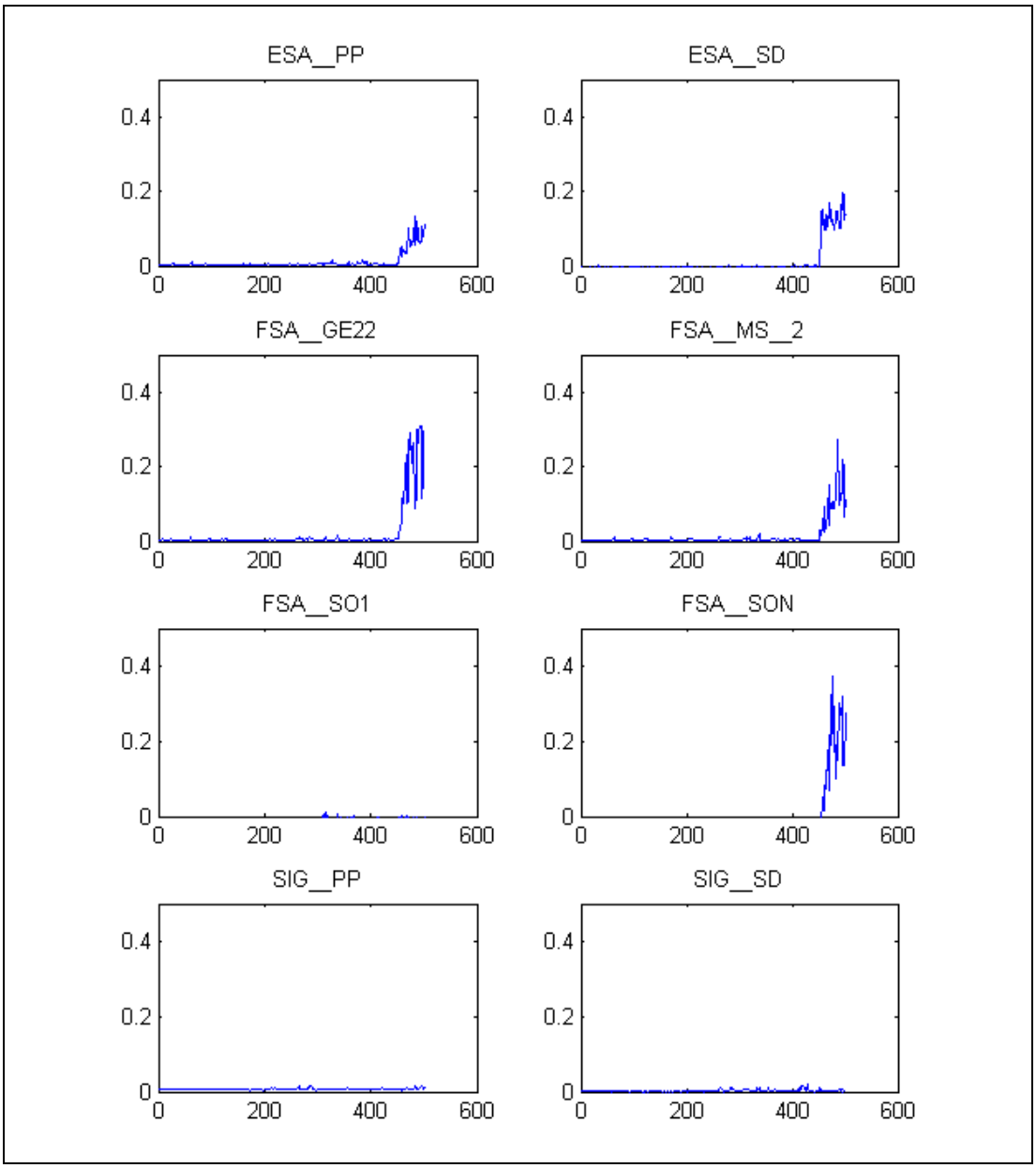

Figure 7-8 Influence traces for the bevel pinion fault derived from the trend model. There are more features evident in this plot compared to the absolute model shown in Figure 7-7. The significance of FSA_SON is also evident.

7.3 Example Results of the Analysis of BHL Current Fleet HUMS Data

Following the off-line analysis, the trial system was configured and applied to BHL's current fleet in preparation for the in-service trial. Three prominent anomalies were highlighted before the in-service trial commenced, none of which had produced any IHUMS warnings. One involved a suspected wiring issue and the other two concerned gearboxes rejected for metal contamination. These cases are listed below.

1 G-TIGC. The lower sun and other components in the MRGB, analysed from accelerometer 7, showed very high and variable trends in ESA-SD / ESA-PP and SIG-SD / SIG-PP. BHL reported that there had been a known problem on some epicyclic accelerometer channels that had been addressed by a modification involving the fitting of Endevco wiring with welded connectors. Not all aircraft have been modified. BHL are to find out whether this aircraft has the old or new wiring.

2 G-BWZX. The upper annulus in the MRGB is one of the few components monitored by multiple sensors. All three accelerometers (5, 6 and 7) showed significant trends in the FSs. These trends are unusual and emphasise the significance of the anomaly. The anomalies were driven by clear rising trends in SIG-SD / SIG-PP. No IHUMS warnings were generated but BHL stated that these appeared to be genuine gearbox trends as they were seen on three channels. A few days later the gearbox was rejected for metal contamination. A strip report has been requested from Eurocopter.

3 G-TIGT. This component was rejected five months before the trial but was not included in the training or validation data and only came to light when configuring the trial system. There were clear anomalies on the absolute and trend models of the TGB input caused by high values of ESA-SD / ESA-PP and high and rising values of SIG-SD / SIG-PP. BHL reported that the TGB had been replaced for metal contamination on the 31/12/2005.

8 Conclusions

The CAA's motivation to commission this programme of work was a belief that there is scope to improve existing HUMS. The findings at the interim stage of this programme confirm this belief. Existing HUMS essentially perform univariate analysis in that they treat each indicator independently when computing its position relative to a threshold. The thresholds are defined indicator by indicator. Component related faults can appear as local outliers meaning that they occupy the edge of space in a low density region that is likely to be contained by the global (fleet) space. It is possible for a single indicator that is trending because of a fault not to exceed its threshold. A repeat of such trends across multiple indicators can position acquisitions a long way from normal space even though none exceed their individual threshold. An example of this type of behaviour was seen with the bevel pinion fault case.

Existing HUMS fleet thresholds have reduced fault detection sensitivity because the between component variance is large and these thresholds are only breeched by a component that is trending significantly away from a baseline that is positioned well within the fleet space of normal behaviour. Learnt thresholds, where a newly installed component has its own thresholds defined after acquiring an initial set of samples, are designed to be more sensitive to behaviour local to a component. But learnt thresholds have limitations. There is no sense of 'self-awareness' when calculating learnt thresholds and it is possible for a threshold to be defined using acquisitions from an abnormal state. Learnt thresholds should be defined with a notion of statistical support. Even when a component remains in the same state (i.e. does not trend) the statistics regarding outliers change as more data are acquired. This means that a component's threshold should be conservative in its early life and less so as more data are acquired. This approach acknowledges that confidence regarding a threshold's setting increases with more data. If confidence is not modelled, there is a risk that initial thresholds will be set too high. Ideally, these thresholds should be set with knowledge learnt from a fleet of historical data, which is the type of approach used in this programme.

Existing HUMS effectiveness is undermined because they do not learn from the considerable amount of information hidden in historical data. Some might argue that human experts continually learn about HUMS, and while there is some truth in this statement, the learning is highly selective in that a human's attention is usually only drawn to the data following a HUMS alert. Humans are excellent at recognizing shapes in charts but struggle to quantify the significance of subtle patterns, and also cannot view data in multidimensional space.

This programme uses a data driven approach to analysing HUMS data. The idea is to let the data inform as to what is normal and abnormal. This type of analysis assists diagnosis when knowledge is incomplete, and for many health monitoring applications, the amount of undiscovered knowledge is considerable.

This programme has developed an anomaly detection method built on existing data mining algorithms. The core algorithm is density estimation by mixture modelling. The anomaly model is constructed by adapting the mixture model to remove the effect of outliers in the training data. The modelling technique automatically suggests outlying regions for removal. This is a new approach to anomaly detection for applications that contain a considerable quantity of outlying data (in excess of 1%).

At the start of the live trial in Phase 2 of the programme the alerting thresholds have been set cautiously. There will be a tendency to generate false alerts but at this stage it is important to minimise the risk of missing things that could be significant. Many alerts can be expected to be instrumentation related. For now the policy suggested

in the introduction to Section 2 has been adopted, which is that the anomaly detector should bring to attention anything that is unusual. The anomaly detector is not a diagnostician. It is believed that there is a secondary level of processing required to diagnose the significance of an alert. This is currently done manually but it should be possible to automate much of it and the anomaly model can provide useful diagnostic inputs to this reasoning process. It will take the best part of the initial Phase 2 trial to optimise the alerting policy. It would be wrong at this stage to overly worry about the frequency of alerts because incorrect suppression, due to regular instrumentation issues for example, could hide fault related incidences.

The anomaly detector is uncovering previously unseen anomalies. It has shown that the bevel pinion fault case, that went undetected by the HUMS, did contain significant trends that were anomalous. Also the features of the anomaly should have caused concern. But the anomaly detector is also highlighting previously undetected instrumentation issues and there is growing evidence to suggest that these instrumentation issues have a disruptive impact on the effectiveness of health monitoring. For example, the shaft order energies often contain instrumentation anomalies that can mask important trends. The anomaly detector rejected these instrumentation anomalies when building the anomaly models.

The anomaly detector addresses the CAA's requirements given in the introduction. The anomaly detector:

1 Improves warning time;

2 Will not be influenced by build defects;

3 Will accommodate unexpected gear indicator reactions because it is data driven and does not utilise any preconceived ideas about fault related symptoms;

4 Accommodates reducing gear indicator trends; the Fitness Score (FS) responds the same for increasing and reducing trends.

5 Calls attention to those regions of data that contain useful information.

With existing HUMS it is not possible to quantify their true effectiveness because there is no tool to monitor their performance. Generally awareness is limited to information pertaining to that portion of data that relate to confirmed correct HUMS rejects or HUMS misses. It is evident, that for significant periods of time, there is no awareness of a HUMS' visibility of the components it is monitoring. It could be argued that this issue may be specific to the HUMS analysed on this programme, but this is unlikely. HUMS data contain a lot of hidden information that should be extracted. The anomaly detector can isolate the regions of data that are information rich. As a minimum, this capability should be viewed as a valuable addition to existing HUMS analysis that should be utilised to increase knowledge and assist with on-going improvements to HUMS effectiveness.

9 Recommendations

Anomaly detection for HUMS should be viewed as a two stage process. The first stage concerns the identification of behaviour requiring investigation. The second stage performs an investigation to identify the likely cause. This programme has been concerned with the first stage and relies on human interpretation for the second stage. However care needs to be exercised during the second stage because there will be a tendency to do a univariate search for the cause of a multivariate anomaly. In other words there is a danger of underestimating the significance of anomalies. The alerts from the first stage should also err on the side of caution. The anomaly detector is data driven and, compared to existing HUMS, operates in a much broader domain (because it does not make assumptions) meaning that it naturally widens the space of events that can trigger an alert. There is therefore a requirement to maximise the information that can be obtained from the first stage to assist the second stage. Ideally, the second stage should be automated as much as possible.

The following is a list of recommendations on future research areas arising from the work that has been completed to date. These tasks will enhance the information flow from stage 1 analysis and demonstrate supporting technology for stage 2.

1 Pre-processing: There is a great deal of variability in HUMS data (e.g. due to instrumentation problems and maintenance actions) that could mask important trends. The current trend pre-processing performs a form of differencing and does not distinguish between a developing trend, step change, or occurrence of noise in the data. It is recommended that other trend pre-processing options are explored.

2 Model tuning: The application of this technology is very new and trial experience is needed to refine the models.

3 Probabilistic alerting policy: Setting a hard threshold on FSs to demarcate interesting from non-interesting data is not the best approach and limits system capability. There could be a tendency to interpret FSs in a manner similar to reading a linear temperature scale but the distribution is not linear. A probabilistic measure of an 'anomaly index' is more appropriate, this would also normalise the anomalies and assist with reasoning across shafts.

4 Influence traces: The in-service trial system does not include the facility to generate influence traces and these traces have not been widely utilised in analysing the FSs derived from the IHUMS data. Influence traces provide valuable diagnostic information and reduce the need for 'drill down' to raw IHUMS data and comparisons with fleet views. Influence traces would also provide a useful input to any reasoning component such as that recommended next.

5 Automated reasoning: More directed information could be provided by reasoning with anomaly outputs. FS exceedances bring to attention unusual behaviour but the likely cause of this behaviour has to be inferred manually. A reasoning component could automate this process, provide consistent interpretation, and quantify features of the anomaly FS trace. Information across shafts could be fused to identify instrumentation issues. Reasoning could be performed on the nature of the anomaly – trends vs. high variability vs. step changes etc. Reasoning could be performed on the indicators driving a trend to provide more detail on the significance of anomalies. Finally case-based reasoning could be performed to search for any similar previous cases.

6 Data mining of anomalous trend features to test theoretical models: Once implemented, the influence traces can point to the indicators that are driving trends. It would be very informative to mine the identified features to test established diagnostic knowledge and further develop this.

10 References

1 Ball, G. H. and Hall., D. J. (1966) ISODATA, an iterative method of multivariate data analysis and pattern classification. In *IEEE International Communications Conference*, Philadelphia, June 1966.

2 Ben-Hur, A., Horn, D., Siegelmann, H. T., and Vapnik, V. (2001). Support Vector Clustering. *Journal of Machine Learning Research*, 2, 125-137.

3 Bishop, C. M. (1994) Novelty Detection and Neural Network Validation. *Proc. IEEE Conference on Vision and Image Signal Processing*, 217-222.

4 Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.

5 Boser, B. E., Guyon, I. M., and Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler, editor, *5th Annual ACM Workshop on COLT*, pages 144-152, Pittsburgh, PA, 1992. ACM Press.

6 Breunig, M. M., Kriegel, H., Ng, R. T. and Sander, J. (12000). LOF: Identifying Density-Based Local Outliers. *Proc. ACM SIGMOD, Conf. On Management of data*.

7 Brotherton, T. and Johnson, T. (2001). Anomaly Detection for Advance Military Aircraft Using Neural Networks. In *Proceedings of the 2001 IEEE Aerospace Conference, Big Sky Montana, March 2001*.

8 Burges, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121-167.

9 Callan, R. E. (2003). *Artificial Intelligence*. Palgrave Macmillan.

10 CAP 693 (1999) "Acceptable means of compliance, Helicopter Health Monitoring, CAA AAD 001-05-99". Civil Aviation Authority, London, May 1999.

11 Chan, P. and Mahoney, M. (2005). Modeling Multiple Time Series for Anomaly Detection. In *Proc. IEEE Intl. Conf. on Data Mining,* 90-97.

12 Corinna Cortes, C., and Vapnik, V. N. (1995), Support-Vector Networks*, Machine Learning, 20.*

13 Daw, C. S., and Finney, C.E.A. (2003). A review of symbolic analysis of experimental data. *Review of Scientific Instruments*, Vol. 74, no. 2, 915-930.

14 Dempster, A. P., Laird, N. M., and Rubin, D. B. (1997). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal statistical society B*, 45, 51-59.

15 Eskin, E. (2000) Anomaly Detection over Noisy Data using Learned Probability Distributions. In *Proc. 17th International Conf. on Machine Learning* 255-262

16 Eskin, E., Arnold, A., Prerau, M., Portnoy, L and Stolfo, S. (2002) A Geometric Framework for Unsupervised anomaly Detection: Detecting Intrusions in Unlabeled Data. In D. Barbara and S. Jajodia (editors), *Applications of Data Mining in Computer Security*, Kluwer

17 Harrison, N., Baines, N. C. (1999). Intelligent Management of HUMS Data: The use of Artificial Intelligence Techniques to Detect Main Rotor Gearbox Faults. Study II, CAA Paper 99006.

18 Hayton, P., Schölkopf, B., Tarassenko, L., and Anuzis, P. (2000). Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra. NIPS. 946-952.

19 Heller, K. A., Svore, K.M., Keromytis, A.D., Stolfo, S.J. One Class Support Vector Machines for Detecting Anomalous Windows Registry Accesses. url = "citeseer.ist.psu.edu/644377.html"

20 Huff, E. M., Tumer, I. Y., Barszcz, E., Dzwonczyk, M. and McNames, J. Analysis of Maneuvering Effects on Transmission Vibrations in an AH-1 Cobra Helicopter *Journal of the American Helicopter Society*, 47(1):42–49, January 2002.

21 Lauer, M. (2001). A Mixture Approach to Novelty Detection Using Training Data with Outliers. In *Proceedings of the 12th European Conference on Machine Learning*. Lecture Notes In Computer Science; Vol. 2167, 300-311.

22 Mackey, R. (2001) Generalized cross-signal anomaly detection on aircraft hydraulic system. In *Proceedings of the 2001 IEEE Aerospace Conference, Big Sky Montana, 2001*.

23 Markou, M. and Singh, S (2003a) Novelty Detection: A Review - Part 1: Statistical Approaches. *Signal Processing* 83(12), 2481-2497.

24 Markou, M. and Singh, S (2003b) Novelty Detection: A Review - Part 2: Neural Network based Approaches. *Signal Processing* 83(12), 2499 - 2521.

25 McClachlan, G, and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley.

26 McClachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

27 Memarsadeghi, N., Mount, D. M., Netanyahu, N. S., and Le Moigne, J. (2005). A Fast Implementation of the ISODATA Clustering Algorithm. Submitted to the International Journal of Computational Geometry and Applications.

28 Park, H., Mackey, R., James, M., Zak, M., amd Baroth, E. BEAM: Technology for Autonomous Vehicle Health Monitoring. http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/12068/1/02-0897.pdf

29 Roberts, S.J. (1999). Novelty Detection using Extreme Value Statistics. In *IEE Proceedings on Vision, Image & Signal Processing,* 146(3):124—129.

30 Rumelhart, D. E., Hinton, G.E. and Williams, R.J. (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, Vol. 1 (Rumelhart, D. E., McClelland, J. L. and the PDP Research Group eds), pp 318-362. Cambridge, MA: MIT Press.

31 Salvador, S., Chan, P. and Brodie, J. Learning States and Rules for Time Series Anomaly Detection (2004). In *Proc. 17th Intl. FLAIRS Conf.*, 300-305.

32 Salzer, M. L. W. (1994) Health & Usage Monitoring of Rotorcraft Transmission Systems – Review of Service Experience. *Aerospace Industries Seminar S247.* Institution of Mechanical Engineers.

33 Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J., and Williamson, R.C (1999). Estimating the support of a high-dimensional distribution. TR MSR 99-87, Microsoft Research, Redmond, WA, 1999.

34 Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylory, J., and Platt, J. (2000). Support Vector Method for Novelty Detection. In *Neural Information Processing Systems*.

35 Shyu, M.L., Chen, S.C., Sarinnapakorn, K. & Chang, L.W. (2003), A novel anomaly detection scheme based on principal component classifier. In *IEEE Foundations and New Directions of Data Mining Workshop, IEEE Int. Conf. on Data Mining (ICDM'03)* (Melbourne, FL), 172.

36 SINTEF (1999) Helicopter Safety Study 2, SINTEF, 15 December 1999.

37 Tarassenko, L., Nairac, A., Townsend, N. and Cowley. P. (1999). Novelty detection in jet engines", *IEE Colloquium on Condition Monitoring, Imagery, External structures and Health, 41-45*.

38 Tax, D. M. J., and Duin, R. P. W. (1998) Outlier Detection using Classifier Instability. In *SSPR '98/SPR '98: Proceedings of the Joint IAPR International Workshops on Advances in Pattern Recognition,* 593-601.

39 Tax, D. M. J., and Duin, R. P. W. (1999). Support vector domain description. *Pattern Recognition Letters*, 20, 1191-1199.

40 Tolani, D. , Yasar, M. and Shin Chin Ray, A. (2006). Journal of Aerospace Computing, Information and Communication. 3, 44-51.

41 Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, Berlin.

42 Vapnik, V. N. (2001). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

43 Wiig, J., Noura, H., Brun-Picard, D., and Derain, J. P. (2006b). Trend-Based Transmission System Diagnosis. *45th IEEE Conference on Decision and Control*. San Diego, CA. USA.

44 Yu, L. J., Cleary, D. J., and Cuddihy, P. E. (2004) A Novel Approach to Aircraft Engine Anomaly Detection and Diagnostics. In *Proceedings of IEEE Aerospace Conference, Big Sky, Montana*.

Appendix A

This appendix provides a basic brief review of some topics that are pre-requisite to understanding this report.

1 Machine Learning

A learning machine is a program that automatically learns a function from a set of training data. A simple example is learning the linear function of a line that separates two classes of data. Training data usually reside in a table made up of rows representing training cases and columns describing variables that act as inputs (and outputs where applicable) to the learning machine. A case is also called an object or training instance or exemplar or sample. A variable is also called a feature or attribute. A collection of variables for a single case is referred to as a vector or point in multidimensional space. HUMS data reside in tables where rows represent acquisitions and columns describe processed features such as CIs. There are two fundamental types of learning machine when the objective is to learn something about the class structure of the data. Supervised learning is utilised when each training case has one or more attributes denoting its class. The class information is used directly in training to inform the machine as to its success in separating the data into their correct classification. Unsupervised learning is used when there is no prior classification and the purpose is to search for the existence of any meaningful class structure. Both types of machine learning are sometimes used together.

A 2-dimensional problem is a task that entails two training attributes. It is possible to chart (plot) real-valued data in one or two dimensions. An *n*-dimensional problem is a task involving *n* attributes, where *n* is an integer. The data used for learning are usually divided into three sets: a training set, a validation set and a test set. The validation set is used to assess the performance of the learnt machine. Re-training may be required when the performance does not satisfy requirements. Because the need for re-training is informed by the validation data, a test data set is used to evaluate the final performance of the learnt machine. Generalization is a term used to describe how well a learnt machine represents the test data. For example, if a supervised machine is over trained its performance on the training data is generally good but its performance on the test data will suffer because it over-fits the training data and cannot generalize to new cases.

Consider the data in the scatter plot in Figure A-1. Each case is a point in the 2-dimensional plot. It is easy to see that the data separate nicely into three groups. If each point is labelled with its group identity a supervised learning machine could be used to find a function that separates the three classes (a multi layer perceptron for example, see Rumelhart et al., 1986 [30]). The objective is for the learnt machine to know how to assign a new point to a class when that point has no prior classification. In this example it is easy to visualize the classification but for most real world tasks such trivial assignment is not possible. The learnt function could be in the form of two lines where each line separates a pair of classes. It is possible to imagine many such functions that differ only in terms of their parameter values (such as slope and intercept, more commonly referred to as weights) describing the lines. Some learnt functions will be better than others in that they will make fewer mistakes when classifying new cases: they are good generalized functions. If the points in the scatter plot have no prior classification, an unsupervised machine could be used to discover the three groups. In this instance, a two-dimensional plot reveals the structure (the

three groups). But real world problems rarely yield their hidden structure so easily. A cluster algorithm is a machine that is capable of learning a function that describes this structure. In this scatter plot there is no overlap between the groups and the data are called separable. In the context of clustering, a group is a cluster.

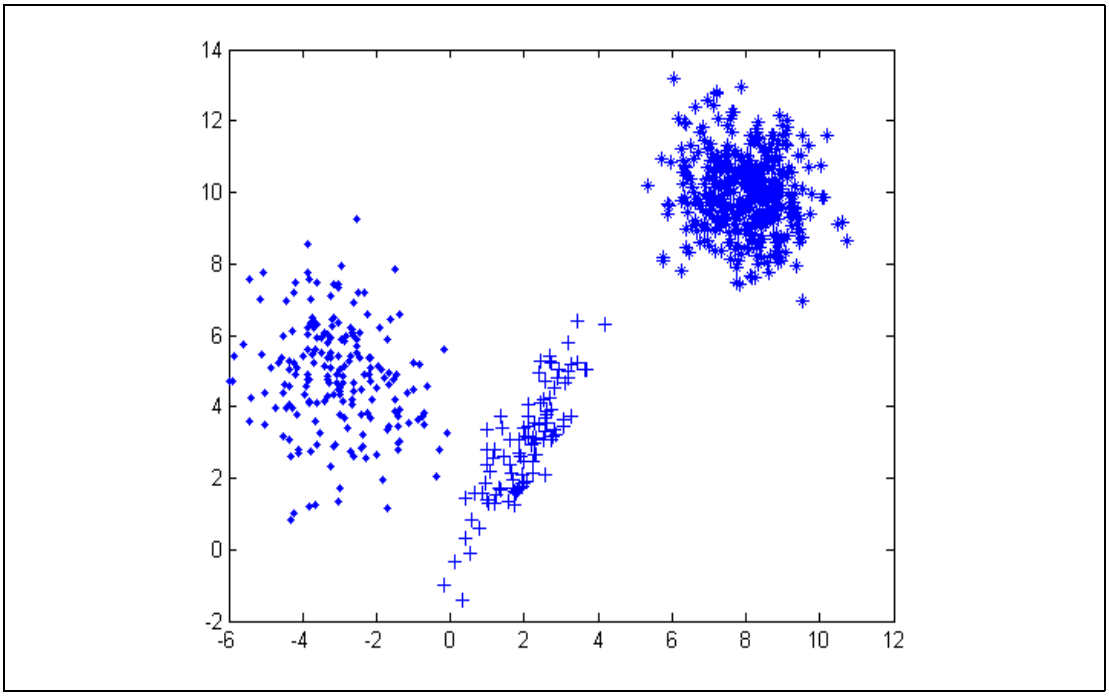

Figure A-1 Three classes of data. The class marked with '*' is linearly separable from the other two because it is possible to draw a straight line to separate it. To separate the class '+' from the other two classes would take two lines and would represent a non-linear classification task.

A learnt function is also called a model. For example, the simplest type of cluster model for Figure A-1 would contain six parameters consisting of three pairs with a single pair denoting the centroid (mean values) of a cluster. The learnt function's capacity to reveal the true structure depends on the type of algorithm used. For example, if all the points were plotted using the same marker (such as '*') a level of ambiguity creeps in regarding those points that sit between the first ('.') and second ('+') clusters.

A cluster algorithm such as k-means uses a distance measure to indicate the degree of closeness between pairs of points (the cluster centroid is also a point). A commonly used measure is the Square Euclidean distance. In two dimensional space, this distance is:

$$d^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

where x and y denote attribute values and the subscripts index a point.

k-means performs what is called hard assignment which means that a point is assigned to a single cluster. A statistical model can assign a point to multiple clusters and its degree of association with a cluster is expressed as a probability (all associations will sum to 1).

A more detailed introduction to machine learning can be found in Callan (2003) [9].

2 Normal Distribution

A density function describes how the values of a random variable are distributed. A normally distributed variable has a density function similar to that shown in Figure A-2. The function is described by two parameters: μ (its mean) and σ (its variance being σ^2). The function is defined by the equation

$$f(x) = f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

This density function tells us how likely or how probable a value of *x* is. Strictly speaking likelihood and probability differ in that likelihood refers to a past event with a known outcome whereas a probability is a prediction of a future event.

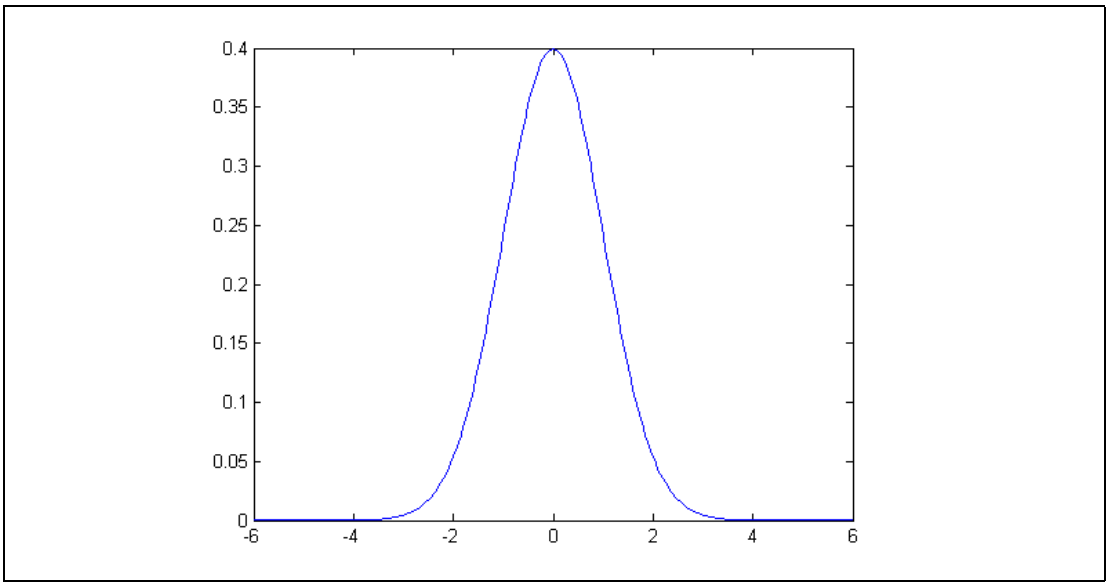

Figure A-2 Plot of a normal distribution with a mean of 0 and variance of 1

Figure A-3 shows the cumulative distribution corresponding to the density in Figure A-2. The cumulative function gives the probability of being equal to or less than *x*.

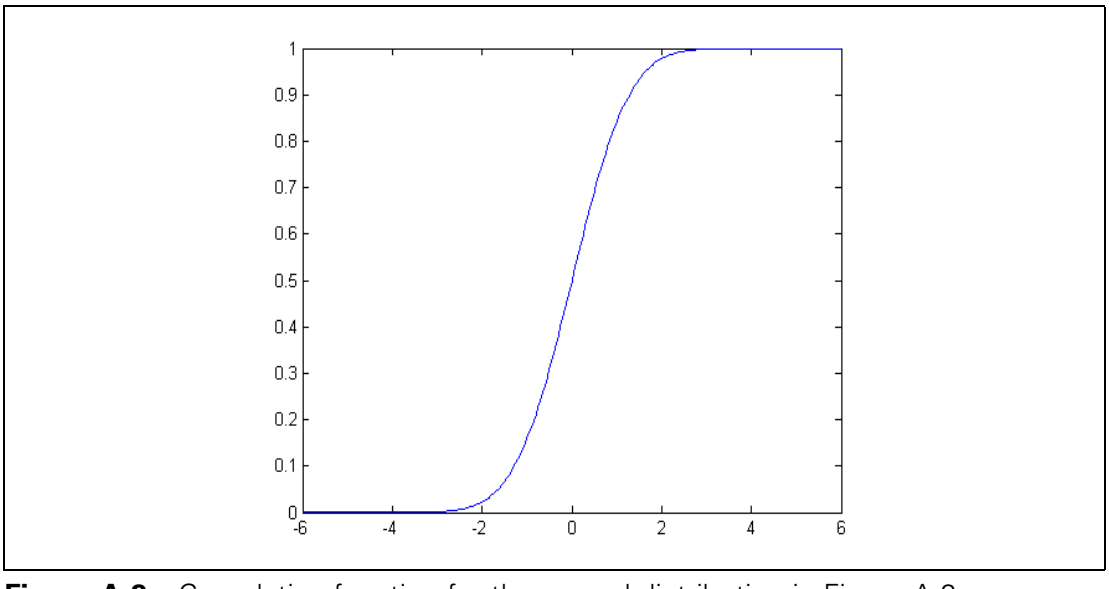

Figure A-3 Cumulative function for the normal distribution in Figure A-2

In this report the term Gaussian distribution is used instead of Normal distribution to avoid confusion when talking about normal data (as opposed to anomalous data).

A Gaussian distribution can be multivariate. A 2-dimensional Gaussian is shown in Figure A-4. In this example, the Gaussian is spherical in that the variance is the same along each dimension.

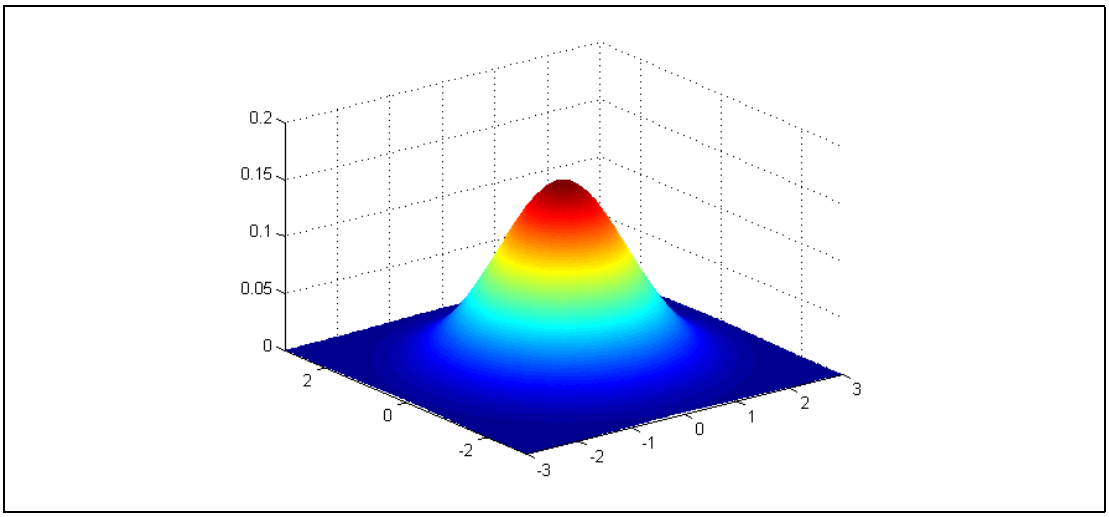

Figure A-4 A multivariate Gaussian in two dimensions. The likelihood score is the vertical axis and reaches its maxima at the centre. The variance along each dimension (the base) is 1 and the distribution is centred at (0, 0).

A cluster model can be described using Gaussians. These models are called Gaussian Mixture Models. Each cluster is a multidimensional Gaussian (unless the data are univariate). Each cluster is parameterised by its mean and variance along each dimension. There is also a parameter denoting the cluster's prior probability. Basically this is a value that indicates the proportion of training cases associated with a cluster. For example, suppose a three-cluster model has been trained with 100 cases. If 50 cases are associated with cluster 1, 30 with cluster 2 and 20 with cluster 3, the probabilities are 0.5, 0.3 and 0.2 respectively. The number of cases associated with a cluster is also referred to as support. So cluster 1 has a support of 50 (which can also be expressed as a proportion, i.e. 0.5).

3 Conditional Probability

Conditional probability is the probability of an event B given that event A has occurred and is written as P(B|A). This is read as "the probability of B given A". Conditioning a probability usually results in a reduced sample space. For example, the sample space of human subjects is restricted if any probability measure is conditioned on a human being a sports person. It is possible to speculate that the probability of a subject being overweight would be lower if the samples were restricted to horse jockeys.

The joint probability of event A and event B is written as P(A, B) and refers to the probability of both events occurring together.

The probability of B given A is

$$P(B \mid A) = P(A, B) / P(A)$$

If A and B are independent their joint probability is

$$P(A, B) = P(A)P(B)$$

Appendix B

1 Equations for a Gaussian Mixture Model

For a dataset X, and clusters K the likelihood of X is

$$p(X \mid K) = \prod_{i=1}^{N} p(x_i \mid K)$$

Summing over all clusters

$$p(X \mid K) = \prod_{i=1}^{N} \left[\sum_{k=1}^{K} p(x_i \mid K) p(k) \right]$$

Taking logs the product over each case becomes a sum

$$p(X \mid K) = \sum_{i=1}^{N} \log p(x_i \mid K)$$

The sum of cluster priors is 1

$$\sum_{k=1}^{K} p(k) = 1$$

For a Gaussian mixture model the likelihood score for a case is

$$p(x_i \mid k, \theta_k) = \frac{1}{(2\pi)^{d/2} \left|\Sigma\right|^{1/2}} \exp\left[-\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right]$$

The prior and posteriors are calculated using Bayes rule

$$p(K \mid X) = \frac{p(X \mid K) p(K)}{p(X)}$$

2 Indirect Model

For the indirect model, and ComponentFitID set Z, the entropy H is

$$H(k) = \sum_{j=1}^{Z} \frac{p(z_j \mid k)}{z_j} \log_2 \frac{p(z_j \mid k)}{z_j}$$

3 Direct Model

The general structure for the direct modelling approach is

$$p(\mathrm{M}_m) = p(X \mid K) p(K \mid Z) p(Z \mid R)$$

Where M is a model and R is a set of constraints on Z. R defines the associations between members of Z.

Each member of M is an individual model. The model can be indexed by a component c, and sensor s. Model Fitness Scores can then be fused according to:

$$p(\mathrm{M}_{c,s}) = \prod_{m}^{M} p(M_m), \quad p(M_m) = \begin{cases} p(M_m) & m \in c,s \\ 1 & m \notin c,s \end{cases}$$

The Fitness Scores can also be fused, and often more meaningfully, by defining a discrete distribution that combines the Fitness Scores (or alerts) from multiple models. For example, the model to provide a fused alert, A, over multiple models could take the form

$$p(A) = P(A \mid A_{m \in c,s}) \prod_{m}^{M} P(A_m \mid M_m) \quad p(M_m) = \begin{cases} p(M_m) & m \in c,s \\ =1 & m \notin c,s \end{cases}$$

This structure can then be extended to provide an alert policy.

End Notes

i Density estimation is a process for constructing an estimate of an unobservable probability density function from observed data. The density function provides a model for how data are distributed. For a discrete variable the distribution can be visualised with a histogram. For a continuous variable the visualization is a curve (or smoothed histogram). The observed data are considered to be a random sample from the estimated density function. Density estimation is said to be ill-posed because there are many estimates that could generate the observed data.

ii Control limits for process variables are often used to detect abnormal deviations. The limit for a single variable might be set 3 sigmas above the median. Hotelling's T^2 statistic is often used for process control monitoring where a single limit is defined for multiple variables. This statistic is the multivariate generalization of the t-distribution that is used to test the statistical significance of the difference between two sample means. In a process control context, the t-distribution is used to test if data are consistent with the process mean. The T^2 statistic is often used to compare a point (new case) with a sample distribution; the distance measure, between the point and the sample, is essentially the squared Mahalanobis distance (see vii). For control monitoring, Hotelling's T^2 statistic can be seen as a way to test for outliers by computing a normalized multivariate distance and comparing it to a T^2 control limit.

iii Expectation Maximization (EM) is a general technique that has broad application to many different types of model. It is widely used in the context of mixture models. In a Gaussian mixture model each cluster is a multivariate Gaussian distribution with parameters: cluster probability, mean vector and covariance matrix. The cluster probability is the probability that the cluster is responsible for generating a data point. The mean vector gives the location of the cluster and the covariance its dispersion and rotation. During learning, the model parameters (parameters for all clusters) are adjusted to maximise the likelihood of the data (training data). The EM algorithm is a two step iterative solution for learning a mixture model. The algorithm starts by assuming values for the parameters. The expectation (E) step, uses the current parameters to estimate the posterior probabilities (evaluates the cluster probabilities from the training data). The maximization (M) step, uses the posterior probabilities to re-estimate the parameters.

iv The definition of Bayesian Information Criterion (BIC) is:

$$BIC = -2\ln L + k\ln n$$

n - the number of observations (cases)
L - the maximised likelihood
k – number of parameters

BIC is a heuristic for model selection, the preferred model having the lowest value of BIC. For a Gaussian cluster model the parameters are the means, support and covariance's for each cluster. A cluster learning algorithm seeks values for the model parameters that maximise the likelihood (a measure of how well the model fits the data). The likelihood value increases with increasing number of parameters. The BIC makes a trade-off between the likelihood and the number of parameters. Given two models with identical likelihood, the model with fewer parameters (clusters) is the preferred model.

v The term Kernel has a number of mathematical meanings but in statistics it refers to a weighting function used in non-parametric density estimation. A common choice is a Gaussian kernel

$$k(u) = \frac{1}{(2\pi h^2)^{1/2}} e^{-\frac{u^2}{2h^2}},\ u = \|x - x_n\|$$

where *h* is the standard deviation of the Gaussian component. Density estimation is done by placing a Gaussian over each training point and summing the contribution over the whole data set. So the likelihood for a specified point is a weighted sum of all the Gaussian training points.

Kernel functions are also used in pattern recognition tasks to transform data so that non-linear classification problems become linearly separable. This idea is illustrated with a contrived example in Figure N-1. The two classes defined by the single variable are not linearly separable. However, the data can be transformed to a linearly separable problem by mapping the data into a higher dimensional space (two dimensions in this case). Each point in the transformed space is described by the outputs from two Gaussian kernel's.

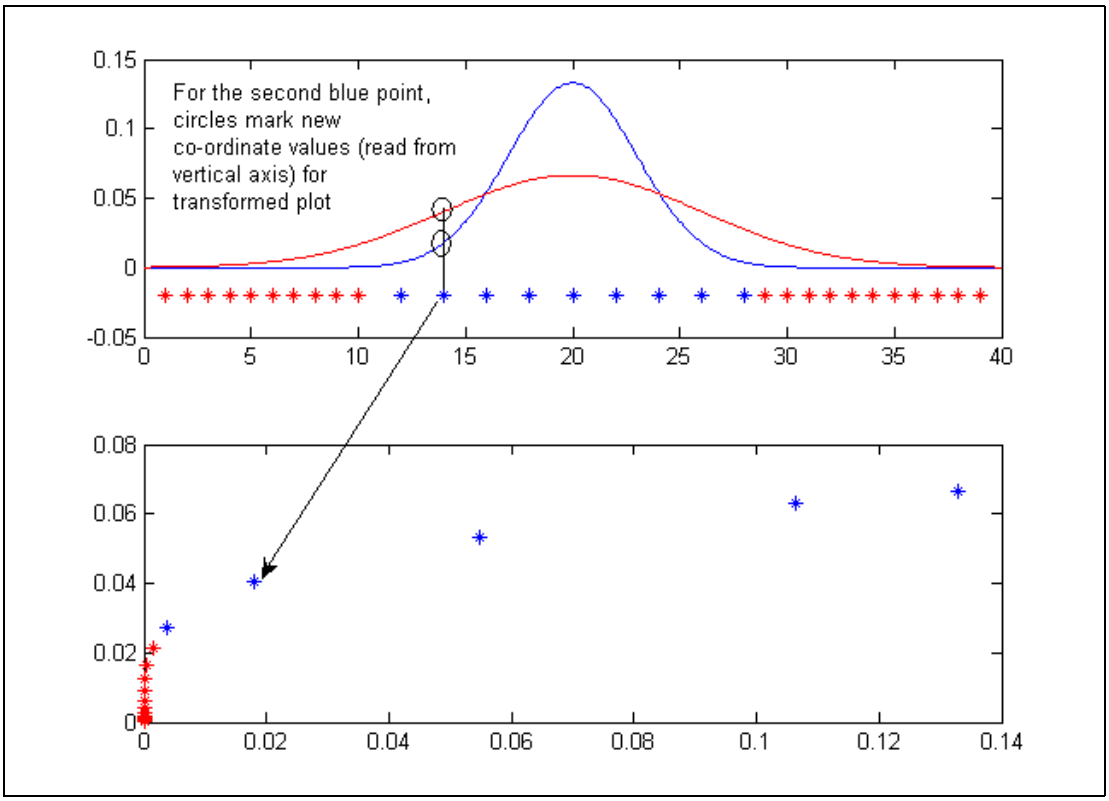

Figure N-1 Kernel functions used in pattern recognition. The top plot shows the distribution of points from a single variable (they all have the same value on the vertical axis). There are two classes distinguished by colour but these classes cannot be separated using a linear function. Two Gaussian functions have been drawn and these act as the kernel functions to transform the data into two dimensions. The Gaussians return two outputs for each point which are plotted in the lower chart. There appear to be fewer points because the problem's symmetry maps some data (from the same class) to the same location. The data can now be separated with a line (it is now a linear problem).

vi There are various types of finite state machine (FSM) but the basic form consists of a set of states, a start state, a set of inputs and a transition function that is a rule for mapping the current state and input symbol to the next state. A transition may also be associated with an output. An FSM is a model of computation. For example, consider the task of determining whether a sequence of inputs consisting of 0s and 1s contains an odd number of 1s. A machine can be built that outputs 0 if the input so far contains an even number of 1s and outputs a 1 if the input sequence contains an odd number of 1s. The table below shows some example inputs and outputs. The output lags the input by one time step.

Input	Output
001010	001100
11100111	10111010
01010	01100

This FSM can be viewed as a graph as shown in Figure N-2. The machine starts in state s1. If the input in state s1 is 0 the machine remains in s1 but will transition to s2 if the input is 1. When in s2, a 0 input will keep the machine in s2 but an input of 1 will transition the machine to s1. The machine outputs a 0 when it has taken a transition and either remains in s1 or moves to s1 from s2. The machine outputs a 1 when it has taken a transition and either remains in s2 or moves to s2 from s1.

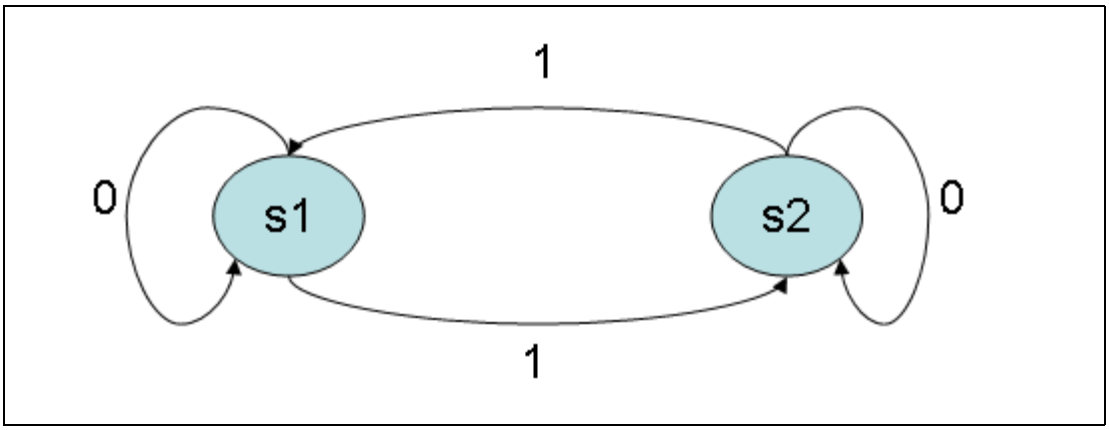

Figure N-2 A finite state machine to detect an odd numbers of 1s in a binary sequence. The outputs are not shown, only the conditions for transiting states.

vii Mahalanobis distance. This distance is defined as:

$$d(x) = [(x-\mu)^T \Sigma^{-1} (x-\mu)]^{1/2}$$

where $\mathbf{x}$ is a vector, $\boldsymbol{\mu}$ is a vector of mean values and Σ is a covariance matrix. The distance measure is normalized to be scale invariant and is sensitive to correlations within the data. Consider for example a set of data that forms a single cluster with the y-axis variance much smaller than the x-axis variance. The density of data is more concentrated about the mean on the y-axis compared to the x-axis. Two new observations (points) can have the same spatial (Euclidean) distance from the cluster centre but different statistical distance because the distance is variable (*x* or *y*) dependent. If there is correlation between *x* and *y*, the direction of maximum variance will lie on a vector which is a weighted composite of *x* and *y*.

viii A covariate is a variable that does not necessarily affect the model during learning but information regarding its statistical significance is desired. For example, consider a cluster model that uses continuous parameters such as bank angle, speed, etc to describe the approach of an aircraft type into different airports. The airport location is a discrete variable (treated as being multinomial) that is not used to train the model but is used as a covariate so that the association between a cluster and airport location can be determined.

ix The Bayesian reasoning tool is a probabilistic graphical modelling tool that describes statistical dependencies between variables. The structure for representing knowledge is often called a network. For example, there is an association between smoking and lung cancer but the symptoms for lung cancer, such as breathing difficulty and coughing, are also associated with other diseases. Causes, symptoms and diagnostic tests are all variables and are represented as nodes in the network. Related nodes are connected and probability distributions are given to define the dependency of relationships. For instance, the likelihood of an abnormal X-ray result given someone has cancer can be quantified.

x A single class machine attempts to learn a separation between a class of interest and all other classes. For example, all known healthy cases of vibration acquisitions could be labelled as positive (healthy) and all other cases left unlabelled (these could include faults and other anomalies). In the context of unsupervised learning, the positive labelled data are the only cases used for training and the machine learns a function that encloses most of the data. There is only one class, which in this example represents healthy data.

xi Imputation is the processes of estimating a missing value. A vector (e.g., the set of CIs for a component) may have one or more missing values. Sometimes cases with missing data values can be removed from the analysis but for some applications this is not practical. It can also be dangerous if the cases with missing values are not 'missing at random' but follow a systematic pattern – though many imputation methods assume missing at random. A simple technique is to substitute a missing value with the mean or median. This approach is often naïve and can lead to misleading results. If a model for the data exists (such as a mixture model) substitution values can be found that maximise the likelihood estimate – in other words substitutes are found that maximises the fit of the vector with the model. Substitutes can also be found using a technique called Multiple Imputation (MI). MI generates *m* (typically 3 – 10) versions of the missing values by sampling from a probability distribution. These *m* versions are then analysed using methods appropriate to the given task (for example, estimating regression coefficients) and finally combined to produce final estimates and confidence intervals.

xii Unifying probabilistic graphical model theory refers to the ability to represent different types of probabilistic models as graphical models. A Bayesian network is a graphical model. ProDAPS, for example, can use the same graphical tool to represent and inference with different types of model such as traditional Bayesian networks, mixture models, conditional Gaussians and Principal Component Analysis. This ability brings a number of advantages. It provides a way to visualise and structure a model, complex models can be constructed from simpler models and it allows the integration of outputs from different levels of computation (e.g., from feature extraction to fusion and diagnostic reasoning).

xiii Suppose that variables *x* and *y* are both Gaussian and *y* is conditionally dependent upon *x*. Without any observation on *x*, the expected value of *y* (i.e., *y*'s prior) is the mean for variable *y* and the variance for *y* expresses the uncertainty in the prior. An observation for x changes the expected value for *y* and its variance. Knowing *x* will increase the confidence in the value of *y*. Such a model can represent the following type of relationship:

$$y = \beta x + \epsilon$$

where β is a regression coefficient and ϵ is Gaussian noise. The models naturally extend to many variables.

xiv There are many examples where data sequences exhibit short term trends. These can be found in historical trends of the daily closing prices of company share prices or the weather. A Markov model is a probabilistic model that represents the dependency between the current observation and preceding observations. Consider a simple discrete Markov model for the weather where an observation, taken at 9am in the morning, is recorded as either sunny, overcast or rain. In this example a first order model is being considered which means that the current observation is dependent only on the previous observation. Each observation is associated with a point in time (day). The model is visualised as a graph structure (a set of nodes and links) where each node is a state (corresponding to an observation) and a link is a transition between states. The weather example will have three states and each state links to itself and the other two states with a value that denotes the probability of transiting to the next state from the current state (it can remain in the same state). Such a model can be used to predict the next day's weather or calculate the probability of a sequence of observations such as {sunny, sunny, rainy, cloudy, sunny}. There are different types of Markov model and they have many applications. Hidden Markov models have wide application to speech recognition and biological sequence analysis. In a hidden Markov model the state is not directly observable. To assist understanding consider a scenario where somebody is laid up in bed for a long period in a room with no windows so they can't directly observe the weather and there is only ever one visitor who will not say what the weather is doing. The visitor often carries an umbrella or wears a hat. There is a high probability it is raining if the umbrella is carried but it could also be cloudy. The hat is usually worn on sunny days but occasionally is worn on a cloudy day if the weather looks less certain and on rare occasions will be worn on a rainy day. The true state of the weather is not directly observable but has to be inferred from observations regarding the visitor. In speech processing a state will correspond to a unit of sound called a phoneme. Phonemes collectively make up the sounds for words and there are about 40-50 phonemes for English. There will be uncertainty in matching a phoneme to a state because different words

share phonemes. A hidden Markov model conveys stochastic structure for the language being modelled and the context defined by surrounding sounds, in a spoken phrase, helps reduce ambiguity.

xv The Multinomial distribution is a generalization of the Binomial distribution. The Binomial distribution is the discrete distribution of the number of successes in a sequence of *n* independent binary trials. The Binomial distribution is defined by two parameters, *n* and *p*, where *n* is the number of trials and *p* is the probability of success. For example, suppose 30% of the population are smokers. If 1000 people were selected at random, the number of smokers picked would be a random variable that follows a Binomial distribution with $p = 0.3$ and $n = 1000$. Intuitively, the most likely number of selected smokers would be 300 and the likelihood of picking 500 smokers would be very small. The Binomial distribution is approximately Gaussian for large values of *n*. The multinomial distribution extends the binomial distribution to the situation where a trial result can be one from a number of possible states. For a K-state variable and *n* trials, a histogram could be plotted which shows the counts for each observed state. The counts sum to *n*. The multinomial distribution gives the likelihood of observing the histogram. For example, the histogram for observing each face of a fair die in six trials has a count of 1 for each state (a state representing a face from 1 to 6). The likelihood for this result is 0.0154 (an improbable result).

xvi For a random variable *X* with probability $p(x)$, entropy is defined as

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

It is a lower bound on the average number of bits required to describe the random variable. Consider a message that is a sequence made up of the characters {U, S, W, P, R, O, A, E} with respective frequencies {2, 4, 8, 8, 64, 64, 128, 256}. A uniform binary code requires 3 bits to represent eight different objects and the total message length would be 1602 bits (sum of frequencies multiplied by 3). The frequencies are expressed as probabilities to compute the entropy; the value returned is 2 bits which means that the average number of bits required to send each character is 2. The uniform coding scheme uses an average of 3 bits which is somewhat higher than that suggested by the measure of entropy. Is there a more efficient coding scheme? The answer is to use a non-uniform coding scheme where the most frequent characters use the shorter codes. For example, the characters could be given the following codes {1111111, 1111110, 111110, 11110, 1110, 110, 10, 0}. Notice that no binary string is a prefix to another; this property ensures there is no ambiguity in decoding the long string of bits representing the message. The total message length is now 1090 bits and the average number of bits is 2.04. The entropy measure gave an average value of at least 2; the actual value is 2.04 which represents a significant compression with no loss of information.

The above example is used to help explain that entropy is a measure of uncertainty or information. If you are not told that a certain event has happened then you are missing no information (because you knew what the outcome would be). In the message example, predicting the next character in the sequence to be E would yield an expectation of success on almost half of all guesses whereas a uniform distribution would reduce this expectation to 1 in 8. An unexpected event yields more information; intuitively it appears that a formal modelling for how we might receive information begins to emerge. For example, the prediction of a failing component is considered highly valuable information.

xvii The Generalized Extreme Value distribution is a probability density function that provides a single representation of three simpler distributions. These simpler distributions differ in the way their tails decrease. The distribution is used to model extreme events. It can be used to model the largest (or smallest) values from a large set of repeated and identically distributed random measurements (or observations). For example, a biscuit manufacturer could record the heaviest biscuit (of a type) out of every batch of 10000 and model these maxima using an extreme value distribution.

INTENTIONALLY LEFT BLANK

ANNEX B

Report on Phase 2 of the Research Project: Six-Month Operational Trial

Based on a report prepared for the CAA by Smiths Aerospace Limited, UK

Table of Contents

Report

INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK

Glossary

AGB	Accessory Gearbox
CAA	Civil Aviation Authority (UK)
CI	Condition Indicator
DAPU	Data Acquisition and Processing Unit
FFT	Fast Fourier Transform
FS	Fitness Score
HUMS	Health and Usage Monitoring System
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
LHA	Left Hand Accessory module
MJAD	MJA Dynamics Limited
MGB	Main rotor Gearbox
OEM	Original Equipment Manufacturer
RHA	Right Hand Accessory module
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring

INTENTIONALLY LEFT BLANK

Executive Summary

This report documents the results of Phase 2 of a CAA funded research programme to demonstrate the intelligent analysis of helicopter HUMS Vibration Health Monitoring (VHM) data. This programme is applying unsupervised learning techniques to HUMS data to construct an anomaly detection system with the goal of improving HUMS fault detection performance. This data driven modeling approach uses historical HUMS data to define models of 'normal behaviour' which can then be used to detect 'abnormal behaviour'.

An unsupervised learning technique for anomaly detection, and system to implement it, were developed in Phase 1 of the CAA research programme which also included an off-line demonstration using a database of historical AS332L HUMS data. The results of Phase 1 are reported separately in reference [1]. In Phase 2 of the programme, which is the subject of this report, the anomaly detection system was subjected to a six-month in-service trial by Bristow Helicopters. The system was implemented as a web server at GE Aviation in Southampton, with AS332L HUMS data being automatically transferred every night from Aberdeen, and Bristow having a remote login to the server to check the anomaly detection system outputs.

The initial six-month trial has demonstrated that the anomaly detection system represents a major advance in HUMS data analysis, resulting in improved fault detection performance. The system has highlighted anomalous data that were not seen by the existing HUMS, and has given a much clearer picture of anomalous data characteristics on particular aircraft and drive train components than is possible with the traditional HUMS analysis. The system can therefore improve the effectiveness of HUMS, and also provide valuable assistance to an operator in the management of the system.

A number of aircraft instrumentation faults were detected that were not seen by the existing HUMS. In at least one case, significant trends on a drive train component were detected that may have been fault related (Section 4.1). Unfortunately, although a gearbox strip report had been requested to confirm the finding, no strip report was produced. The anomaly detection system has also corroborated aircraft drive train component and instrument faults that were detected by the existing HUMS (Section 4.2). Conversely, in one case it has correctly suppressed the significance of trends that were triggering HUMS alerts; the gearbox was placed on 'close monitor' as a result of the HUMS alerts, but was later 'stood down' without any maintenance actions being performed (Section 4.4). A few anomaly models were found to be somewhat over-sensitive, and require 'tuning' (Section 4.5). However, there was also a case in which the anomaly models failed to alert on an out of balance indication, detected by the IHUMS, on a tail rotor gearbox output shaft due to tail rotor flapping hinge bearing damage (Section 4.3). Some re-modelling with a small set of shaft-related indicators may be required for the monitoring of key gearbox input and output shafts.

The anomaly detection system has highlighted the fact that the IHUMS data on Bristow's AS332Ls is relatively noisy due to instrumentation issues on the aircraft. There were a significant number of HUMS accelerometer, insulation washer, and cable or connector problems, which may be due to the fact that these aircraft received some of the first HUMS installations. Not only do these trigger anomaly model alerts, but their prevalence pollutes the model training data, which can result in a loss of fault detection sensitivity.

The trial verified that the usability of the web-based anomaly detection system was good, and the operator was able to use it effectively. The responsiveness of the website to remote commands accessing various chart displays over the internet was found to be acceptable. The reliability of the automated overnight data transfer process and the anomaly detection system website was good.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Health and Usage Monitoring Systems (HUMS), incorporating comprehensive drive train Vibration Health Monitoring (VHM), have contributed significantly to improving the safety of rotorcraft operations. However, experience has also shown that, whilst HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected using current threshold setting methods. Earlier research, conducted as part of the CAA's helicopter main rotor gearbox seeded defect test programme, demonstrated the potential for improving fault detection performance by applying unsupervised machine learning techniques such as clustering to seeded fault test data. The CAA has now commissioned a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841), which is the subject of this report. GE Aviation was contracted to carry out this programme in partnership with Bristow Helicopters, analysing IHUMS data from Bristow's North Sea AS332L fleet.

The work under Contract No. 841 was structured as two phases. Phase I started with a literature survey to identify other research work on unsupervised learning, and went on to develop an unsupervised analysis technique to enhance the fault detection performance of HUMS which was subjected to an 'off-line' demonstration on a database of historical AS332L data. It also included the development of a system implementing the technique that could be evaluated in an in-service trial. The results of Phase 1 are reported separately in reference [1].

The unsupervised learning approach is known as anomaly detection. The need for anomaly detection arises in situations where supervised learning of fault related patterns is impractical. Supervised learning requires that every signal acquisition be tagged with a known healthy or particular fault classification. Fortunately component related faults are rare, but this means that there is no large library of tagged fault data. Even if such a library existed, the required size would be indeterminate because symptoms for a single fault type can vary between cases. A large majority of the acquired HUMS data can be assumed to be associated with a normal (healthy) component state. Recognition of a fault could therefore be achieved by building a model of normal behaviour against which a test for the abnormal can be applied. In the context of this work, anomaly detection is considered to be a process that calls attention to abnormal data that requires investigation.

Phase 2 of the CAA research programme comprised a six-month in-service trial, conducted by Bristow, of the anomaly detection system developed in Phase I. The system was implemented as a web server at GE Aviation in Southampton, with AS332L IHUMS data being automatically transferred every night from Aberdeen, and Bristow having a remote login to the server to check the anomaly detection system outputs. The results of the six-month trial are presented in this report. Section 2 gives an overview of the IHUMS data analysed, the anomaly processing performed, and the web-based anomaly detection system used in the trial. Section 3 describes the trial operating procedures implemented by Bristow, and Bristow's operational experience with the anomaly detection system. A catalogue of significant trial findings, with 15 example cases, is presented in Section 4, and a statistical analysis of the anomaly detection system performance is contained in Section 5. Finally, some conclusions and recommendations are given in Section 6.

2 HUMS Anomaly Detection Processing and System

2.1 IHUMS Data Analysed

IHUMS VHM data was acquired and analysed from the following assemblies in the AS332L drive train: Main rotor Gearbox (MGB), left and right Accessory Gearboxes (AGBs) and oil cooler fan, Intermediate Gearbox (IGB), and the Tail rotor Gearbox (TGB). Thirty-five drive train component analyses were defined, as listed in Table 2-1. The table also shows the aircraft sensor used to acquire the component data, and the equivalent IHUMS "channel" (i.e. analysis number) allocation. The IHUMS performs thirty-two component analyses. However, three of the components have two gears on a single shaft (the MGB combiner gear and bevel pinion, and the left and right AGB hydraulic drive 47- and 81-tooth gears). For the purposes of anomaly modelling, each of these was split into separate components, with separate anomaly models built for each of the two gears on the shaft using the different gear mesh-specific Condition Indicators (CIs).

Table 2-1 AS332L drive train components analysed

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MGB
2	1	RH high speed input shaft	MGB
1	2	Left torque shaft - fwd end	MGB
2	3	Right torque shaft - fwd end	MGB
3	4	Left torque shaft - aft end	MGB
4	5	Right torque shaft - aft end	MGB
3	6	Combiner gear	MGB
3	6	Bevel pinion	MGB
4	7	Bevel wheel and oil pump drive	MGB
7	8	1st stage sun gear	MGB
7	9	1st stage planet gear	MGB
5	10	1st epicyclic annulus fwd (RH)	MGB
6	11	1st epicyclic annulus left	MGB
7	12	1st epicyclic annulus aft (RH)	MGB
7	13	2nd stage sun gear	MGB
7	14	2nd stage planet gear	MGB
5	15	2nd epicyclic annulus fwd (RH)	MGB
6	16	2nd epicyclic annulus left	MGB
7	17	2nd epicyclic annulus aft (RH)	MGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB
3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB

Table 2-1 AS332L drive train components analysed (Continued)

4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MGB

The CIs calculated by the IHUMS during each component analysis are listed in Table 2-2 (a more detailed description is presented in reference [1]). Not all of the indicators have been used in the anomaly modelling. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data.

Table 2-2 IHUMS Condition Indicators

Condition Indicator	Description
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation (rms)
FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev (rms)
ETA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA_CVG	Signal average convergence

The IHUMS data provided for the anomaly processing was extracted from Bristow's Aberdeen IHUMS ground station. This held the master database, and was regularly updated with data from IHUMS ground stations at satellite bases such as Scatsta. During the six-month in-service trial, IHUMS data from a fleet of 15 Bristow AS332L aircraft was analysed on a daily basis. The data from each aircraft was associated with the serial numbers of the individual gearboxes installed in that aircraft using a "ComponentFitID". When a gearbox was changed the related ComponentFitID was updated in the anomaly detection system so that data could always be associated with a particular gearbox or accessory module serial number.

2.2 Data Processing for Anomaly Detection

A detailed description of the data processing performed is presented in reference [1]. A brief summary of the three stages of the anomaly detection processing is given below.

2.2.1 Data Pre-Processing

Two types of characteristic data behaviour are modelled in the anomaly detection system. The first type is addressed by modelling between-component variation, and such models are referred to as 'absolute models'. Each component tends to have its own level of vibration energy. Even though a variation in vibration between components is expected, there will be a level above or below which a component's vibration appears to be so unusual that it should be flagged as anomalous. The 'absolute model' is designed to detect such cases. The second type of behaviour concerns within-component variation and such models are referred to as 'trend models'. As the name implies, these models are designed to detect anomalous trends in component data. The notion of an anomalous trend is important because trends are not uncommon in HUMS vibration data.

The pre-processing for the absolute models is minimal, and involves the application of a two stage filter to each indicator: the first stage removes unreasonable values and the second stage applies a median filter to remove up to two successive spikes in the time series data.

The objective of the trend pre-processing is to detect the relative change in an indicator by providing a measure that normalises out the between-component variation. For this, the two stage filter for the absolute model is used, and then a 'moving median difference' algorithm is applied. Following each new acquisition, the median of the time history is re-calculated and subtracted from the newly acquired value to provide a normalised value. This technique reduces the impact of early post-installation trends, and also small step changes due to maintenance, because the normalised value will gradually recover back to the median base line level. There is some distortion of the time-history, but in practice this simple approach appears to work well. Its weakness is that, whilst it achieves the normalisation objective, it acts as a difference calculator and cannot distinguish between step changes, short duration changes, and slow or fast developing trends.

2.2.2 Anomaly Models

Using the pre-processed data, anomaly models were constructed for each drive train component. The models were constructed from a training database of historical data, and adapted so that they rejected any abnormalities existing in the training data. Models were built to represent absolute data behaviour (between-component variability) and trend behaviour (within-component variability). These models are sophisticated statistical representations of the data generated from in-service experience, with few assumptions about how the data should behave. They fuse sets of CIs (i.e. vibration features) to reduce a complex data picture into a single time history called a 'Fitness Score' (FS) trace. The FS measures the degree of abnormality in the input data and mirrors the shape of any significant data trends. It represents a 'goodness of fit' criterion, indicating how well data fits a model of normality. Therefore the FS has a decreasing trend as data becomes increasingly abnormal.

With the exception of two drive train components that do not include a gear on the monitored shaft, every component has anomaly models constructed from two sub-sets of the CIs listed in Table 2.2:

- '8-indicator' models, comprising the following eight indicators: {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

- 'M6' models, comprising the following two indicators: {ESA_M6*, ESA_WEA}, (ESA_WEA is a derived indicator, calculated as the ratio ESA_SD/SIG_SD).

The indicators were split to reduce the potential for one indicator to be masked by another; the criterion for this split was derived from engineering knowledge of the indicator design and characteristic behaviour.

Using the different forms of pre-processing, 'absolute' and 'trend' models were built for each set of indicators. Therefore the drive train components are monitored by four anomaly models – '8-indicator' absolute and trend models, and 'M6' absolute and trend models. The oil cooler fan and dual bearing module are exceptions in that they are monitored by only two models – absolute and trend models comprising the following '5-indicator' set: {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

2.2.3 Alert Thresholds

The FS output from each anomaly model has a threshold to alert the operator when a FS has decreased to a level at which it can be considered to be indicating the presence of anomalous data. The thresholds are determined automatically, based on a probability distribution fitted to the FSs derived from the training database, and vary from one model to the next. The number of alerts generated will also vary by component type. This is acceptable because some components are more susceptible to defects, and other components may be more difficult to monitor reliably. An alert is generated on a component whenever one or more of the individual component anomaly models goes into an alert state.

2.3 Anomaly Detection System

For both technical and support reasons, it was determined that the best implementation solution for the anomaly detection system would be as a web-based system, hosted on a server at GE Aviation in Southampton. Figure 2-1 shows the data flow between Aberdeen and Southampton. HUMS data are automatically transferred overnight from Bristow's secure web portal, imported into the anomaly detection system's data warehouse, and analysed. Bristow have a remote secure login to the system to view results at any time via a web browser.

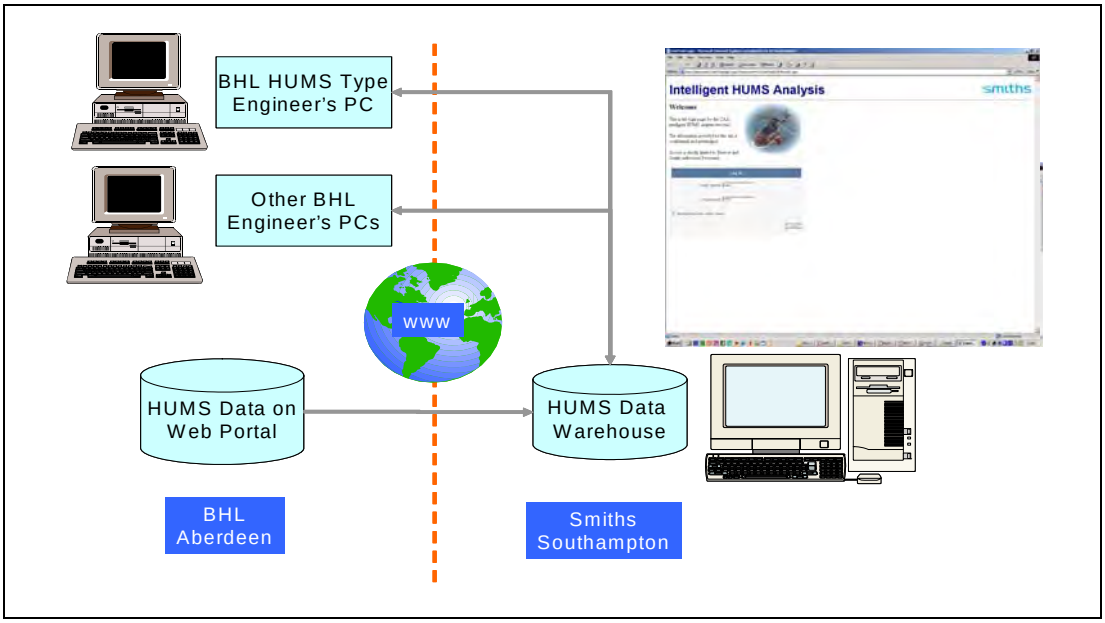

Figure 2-1 Data flow between Aberdeen and Southampton

After logging into the system via the screen in Figure 2-1, the user is presented with a fleet view showing the number of aircraft, alerts (total, current, and unacknowledged), and user-entered notes in the system (Figure 2-2). The user can limit their view of the data to one week, one month, three months, or can view the complete history. There is also a calendar that allows the data to be viewed as it was at some previous date. This option is necessary to access historical data relating to components that have been replaced because the web server's 'latest date' view, which is the default, includes only those components that are currently fitted to the fleet.

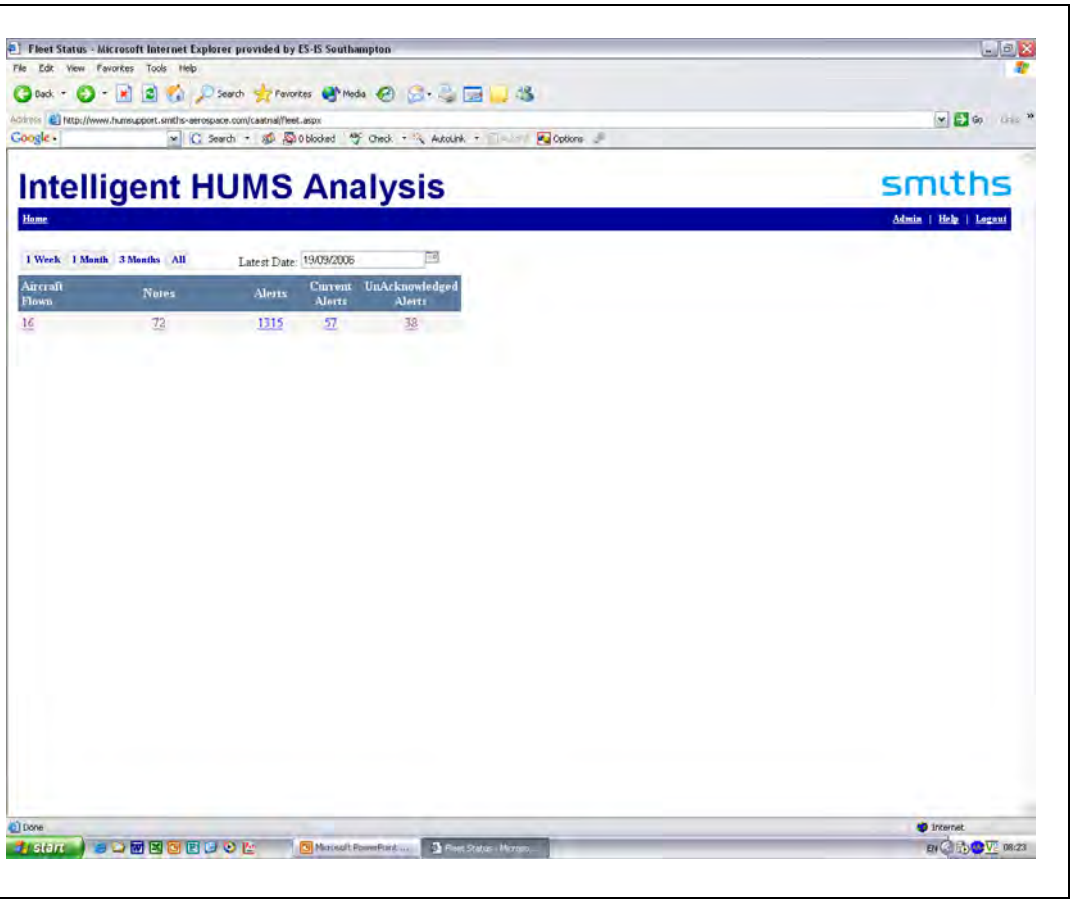

Figure 2-2 Top level summary

The user can drill down via an aircraft tail number, by user notes, or by type of alert. For routine daily checks the user would normally drill down by unacknowledged alerts to view new alerts generated from the anomaly processing of data downloaded from the previous day's flying (Figure 2-3).

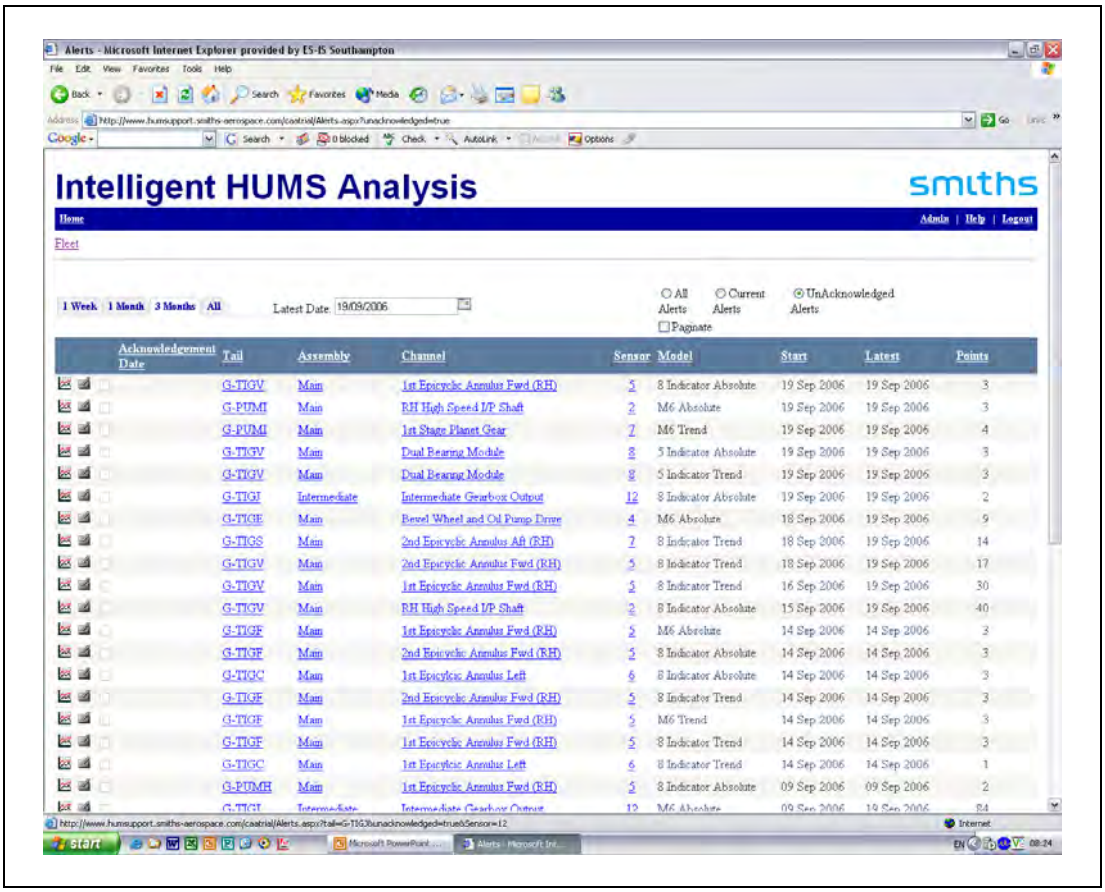

Figure 2-3 Unacknowledged alerts

By clicking on the chart icon at the left hand side of an alert record, the user can then drill down to the anomaly model FS traces for the component in alert (Figure 2-4). These traces are annotated with alerts and any notes the operator has entered to document items such as the findings from any follow-up investigation or maintenance actions. (NB: A note appears as a yellow triangle on the trace display). A maintenance action causing a step change in some of the IHUMS CIs could trigger a trend model alert. Therefore a facility is provided to reset the trend pre-processing, which re-baselines the trend model data and clears the alert.

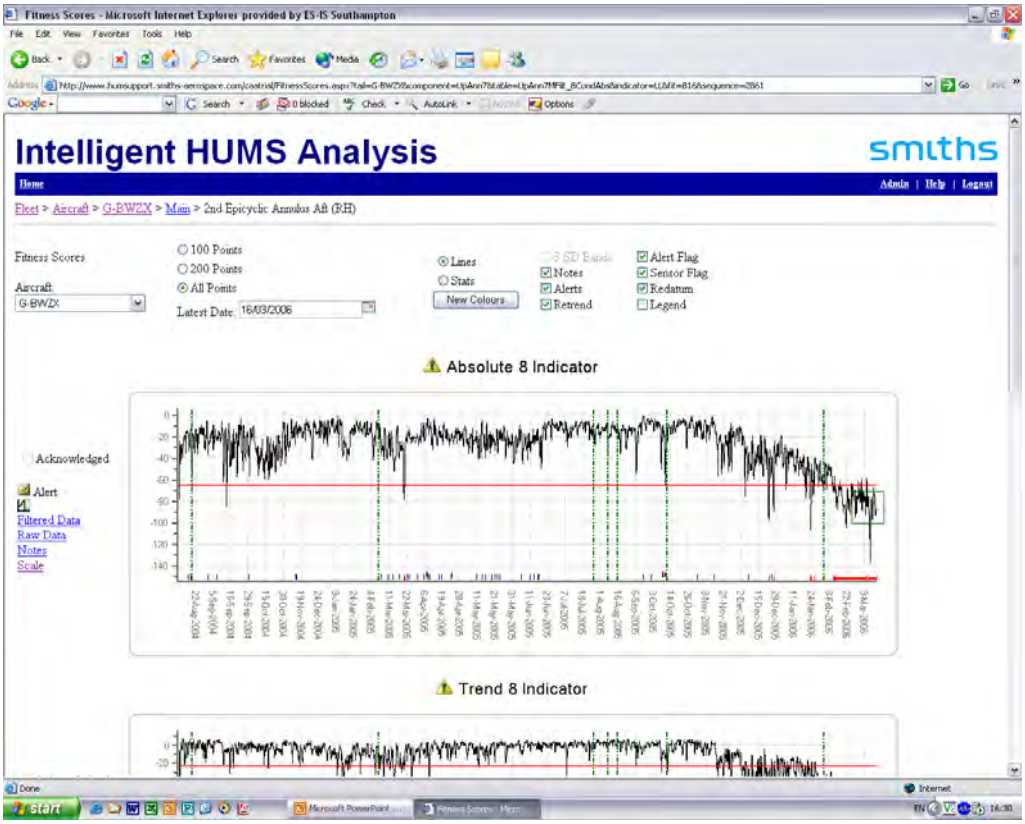

Figure 2-4 Fitness Score trace

For further investigation, the user can drill down to either the filtered or pre-filtered IHUMS CIs (Figure 2-5). For those indicators used to build models, the indicator plots show the upper and lower three standard deviation bands that have been derived from the anomaly model. These bands are different to those that would be computed directly from the fleet data because the anomaly models have been adapted to reduce the impact of outlying data.

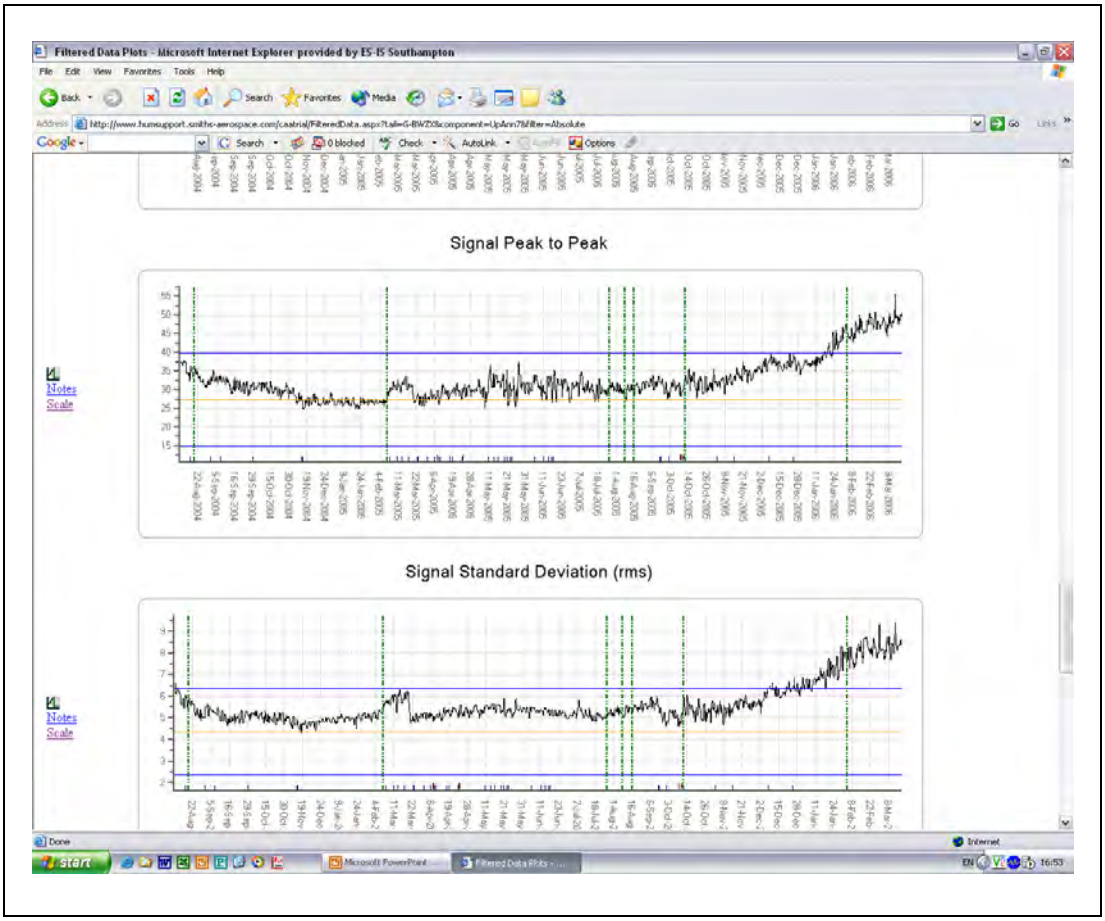

Figure 2-5 Condition Indicator display

It is also possible to view IHUMS CI data for a fleet of aircraft either as multiple trend plots or as statistical plots (Figure 2-6). Once an investigation and any associated follow-up action has been completed the user can acknowledge an alert and, if appropriate, enter a note.

Figure 2-6 Fleet data displays

Figure 2-7 shows how the user can view alerts and notes by aircraft, assembly, and component, and also drill down to the FS and IHUMS CI traces via this component hierarchy. A display of user entered notes is presented in Figure 2-8. Again, by clicking on the chart icon in a note record, the user can drill down to the FS traces to which the note applies.

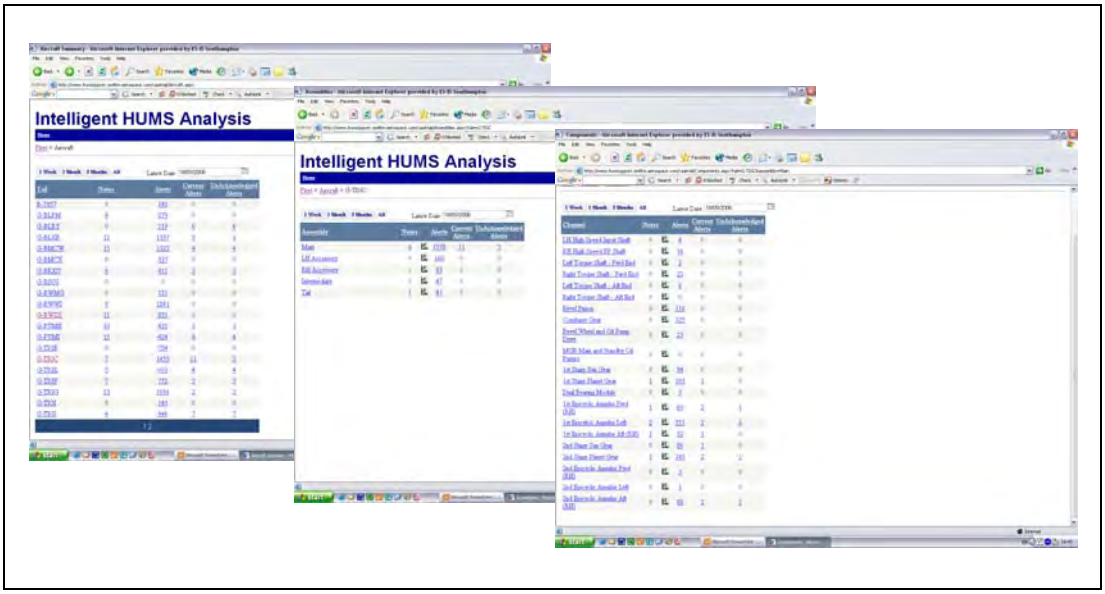

Figure 2-7 Drill down by aircraft, assembly and component

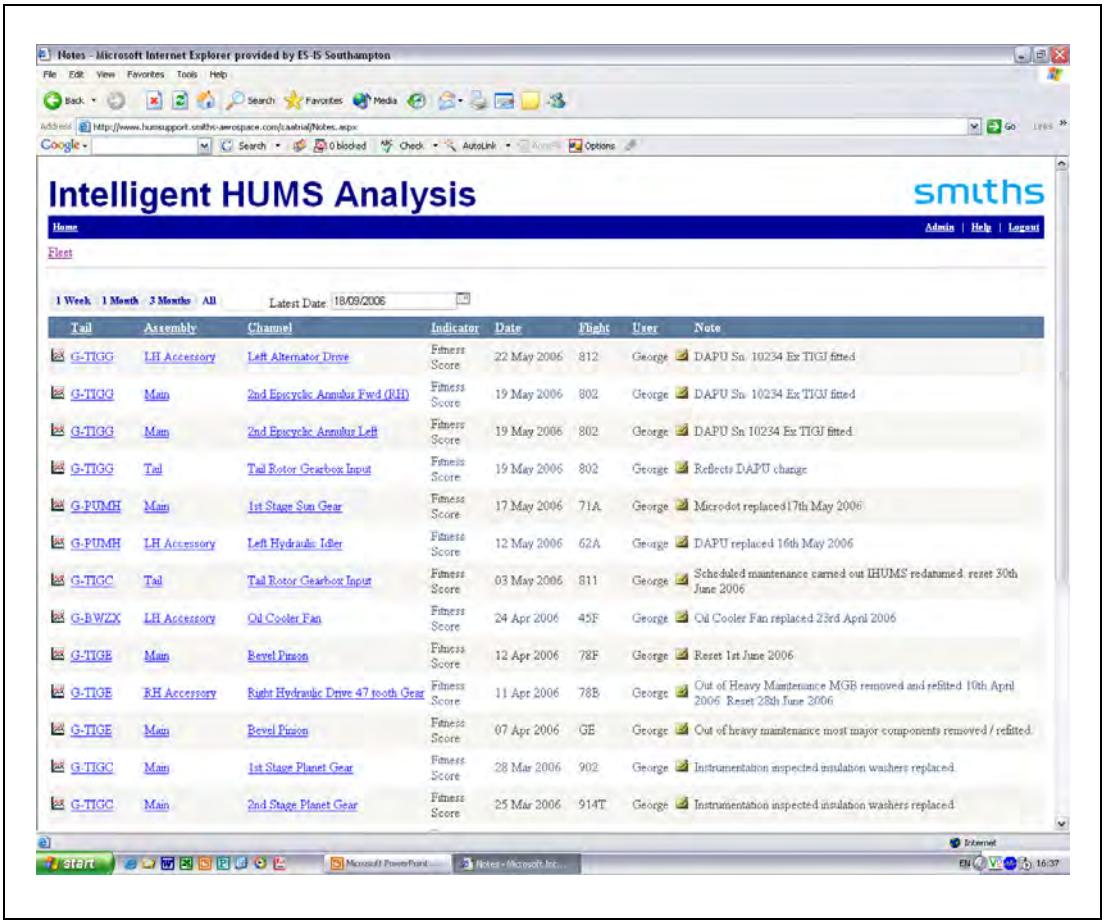

Figure 2-8 Notes display

Figure 2-9 provides a key to the documentary information shown on the anomaly detection system FS traces that are included in Section 4 of the report.

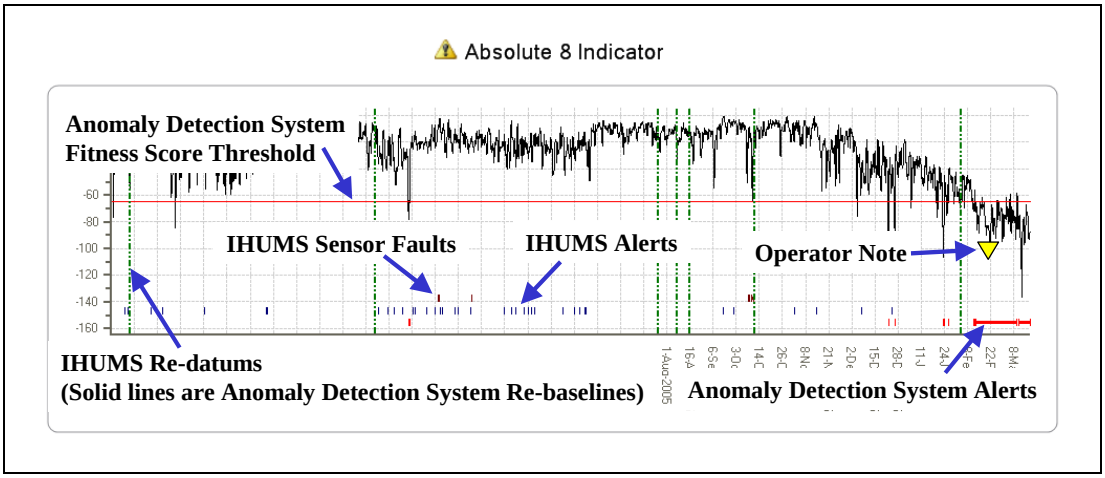

Figure 2-9 Key to documentary information shown on a Fitness Score trace

3 Trial Operational Procedures and Experience

Bristow's HUMS engineer was provided access to the live web-based system on 10 May 2006 via a dedicated login and password, and the official start date for the six-month trial was 22 May.

3.1 Trial Operational Procedures

Apart from occasions when other commitments did not permit it, the Bristow HUMS engineer logged into the anomaly detection system on a daily basis, and performed the following data investigation tasks:

1 Review new component anomaly alerts, and try to substantiate which IHUMS CIs had triggered the warning.

2 View the data associated with the component in the IHUMS ground station, and look for trends or abnormal indications in the IHUMS graphical displays.

3 Check the IHUMS re-datum history for any actions that could correlate with the alert occurrence.

4 Plot IHUMS raw and enhanced vibration signal data to look for suspect instrumentation faults, and plot IHUMS linear FFT data to check for abnormal peaks at certain frequencies or shaft orders.

5 Search the aircraft maintenance database for any maintenance actions or component replacements that could be associated with the alert.

6 Input a note on the relevant anomaly detection system FS chart with reported findings, acknowledging the warning and, if required, resetting the trend pre-processing.

If the data investigation suggested a problem on the aircraft (most commonly related to the IHUMS instrumentation), a maintenance task would be raised for investigation and/or rectification of the problem.

Once a month an aircraft gearbox replacement report was compiled and sent to GE Aviation for updating of the ComponentFitIDs. The maintenance and gearbox removal information was originally extracted manually from Bristow's AMIS (Aircraft Maintenance Information System). Later in the trial the maintenance system changed to IFS (Industrial Financial Systems), which allowed information to be accessed directly from the database for a selected period of time.

3.2 Trial Operational Experience

The reliability of the nightly automated data download from Aberdeen was extremely good, with no failures occurring. Although the process was fully automated, this did require Bristow's HUMS engineer's PC to be running for the internal data transfer from the IHUMS ground station database to the externally accessible web portal. This PC was switched off during periods when the engineer was absent from the office, and the data transfer process was therefore interrupted on these occasions. The importing of IHUMS data into the master ground station database at Aberdeen from satellite bases could also be interrupted during these periods. In addition, aircraft occasionally transferred between bases. This sometimes resulted in IHUMS data files being received that contained some records that were not in the correct date sequence for a particular aircraft. The out of sequence records had to be either ignored or manually corrected during the import into the anomaly detection system.

The reliability of access to the web server at GE Aviation was generally good. There were only a small number of occasions when the site was down for more than

30 minutes (i.e. an operationally insignificant period of time). The reliability of access must be very good for any production system implementation. A key concern was the speed of response when plotting charts and interacting with a chart. Therefore all intensive processing was performed on the server (at Southampton) and charts were uploaded to the client (at Aberdeen) as images. Experience showed that the system response was quite acceptable with the 0.8 Mbits/second upload bandwidth available during the trial.

The web based anomaly detection system was found to be easy to use, with an uncomplicated display that was straightforward to navigate. The graphs displayed were tidy and generally uncluttered, except possibly when data was plotted against the fleet. However, colours and a legend have been provided to help identify each aircraft registration and location on the graph. The help document provided on the website was very comprehensive, and contained all the information needed to use the system.

Based on the initial period of operating experience, a few system modifications were recommended at a project review meeting on 8 June, including:

- Inclusion of the date next to the alert acknowledgement check box.
- If the data remains in alert following initial acknowledgement, enabling the alert to be re-acknowledged at a later time, with a date update.
- Adding a notes field to the alert records to document trial findings.
- Auto-scaling the fitness charts.
- Improving the colour selection for the fleet displays of data, always showing the currently selected aircraft in black for consistency and conspicuity.
- Adding IHUMS alert and sensor fault flags, and also IHUMS re-datum points, to the fitness traces.

The above recommendations were incorporated into the software on 26 June 2006 and 23 August 2006. The addition of the notes field and the IHUMS flags were the most significant modifications. It was originally proposed that a spreadsheet would be used to correlate anomaly detection alerts with IHUMS warnings and document the trial findings. However, managing this had created too high a workload.

At the 8 June project meeting, it was also requested that a login for the Bristow line engineers be added to give them the ability to view data in the anomaly detection system. The line engineers already had access to an IHUMS ground station, however it was found that the anomaly detection system could give better visibility of anomalous data associated with aircraft instrumentation problems.

A couple of further minor modifications were implemented following another project review meeting on 11 October. These included displaying the name of the person acknowledging any alert in the system to provide traceability.

Checking the anomaly detection system on a daily basis was a new work task for the operator. However, the primary workload issue was found to be the time needed to investigate anomaly alerts by checking the associated data in the IHUMS ground station. It could also be difficult to assess the significance of data trends that had been flagged as anomalous when viewing these in the IHUMS. The default IHUMS display shows the 2,000 results that are retained in the system, and this can include data from multiple gearboxes. When manual zoom functions are used to focus on recent data, the apparent visual significance of the trends observed in the IHUMS depended on factors such as the number of data points plotted, and the scaling of the charts.

4 Catalogue of Significant Findings

This section presents a selection of 15 example cases to illustrate the performance of the anomaly detection system. These have been categorised as follows:

1 Cases where anomaly detection has identified a fault not seen by the existing HUMS.

2 Cases where anomaly detection has corroborated existing HUMS indications.

3 Cases where anomaly detection has failed to identify a fault that was seen by the existing HUMS.

4 Cases where anomaly detection has identified an existing HUMS false or premature alert.

5 Cases where anomaly detection has generated a false or premature alert.

6 Cases of anomaly indications with a currently unknown outcome.

The 15 examples were selected on the basis of the notes entered into the anomaly detection system by Bristow's HUMS engineer, and a joint review of the trial findings by GE Aviation and Bristow. The numbers of examples presented for each of the above categories is broadly proportionate to the overall trial experience.

In this analysis, faults are considered to include both aircraft component faults and HUMS instrumentation faults. In the majority of cases it is not possible to positively confirm the presence of a specific fault. For the purposes of categorisation, a correct indication is considered to have occurred if:

- The operator has determined that maintenance action is required (e.g. to replace an accessory module, or a HUMS accelerometer).
- The maintenance action performed has resulted in data returning to normal levels.

The cases are illustrated with a series of charts of FSs, HUMS CIs, and fleet views of CIs, that are taken directly from the anomaly detection system. Unless otherwise stated, all charts of HUMS CIs show the median filtered data that forms the input to the anomaly models.

4.1 Cases where Anomaly Detection has Identified a Fault not seen by the Existing HUMS

Example 1: MGB 2nd Epicyclic Annulus

This example occurred before the official start of the first six-month trial period. The '8-indicator' absolute model for G-BWZX's MGB 2nd epicyclic annulus aft (RH) generated a continuous alert from the first half of February 2006 (Figure 4-1). The '8-indicator' trend models for all three MGB 2nd epicyclic annulus accelerometers (aft (RH), left and fwd (RH)) generated consistent alerts from late 2005 (Figure 4-2 - Figure 4-4). The 'M6' trend models for the aft (RH) and left accelerometers generated consistent alerts from the beginning of February 2006 (Figure 4-5 - Figure 4-6). The IHUMS did not generate any alerts during these periods.

All the alerts were primarily caused by rising trends in the Signal Standard Deviation CIs from all three accelerometer positions (Figure 4-7 - Figure 4-9). In all cases indicator values trended outside of fleet norms, however a couple of other aircraft were generating extremely high values for this indicator (Figure 4-10 - Figure 4-12). It is believed that these very high values may have been caused by instrumentation problems (see examples 3 and 4 later in this section). Signal Peak-Peak values also increased, however the Enhanced Standard Deviation CIs did not change, resulting in

a decreasing trend in the Wear indicators (Figure 4-13 - Figure 4-14) that caused the 'M6' trend model alerts. The MGB was rejected for metal contamination on 20 March 2006. The source of the metal contamination could not be identified and, although a gearbox strip report had been requested from Eurocopter, no feedback was received.

It is not possible to confirm whether the anomaly detection system alerts and rising Signal Standard Deviation values were associated with the metal contamination. However, it is noted that such contamination often originates from the mast bearing, and the IHUMS does not detect this. If this was the source of the contamination, it is possible that the anomaly alerts were related as the signal averages for the 2nd epicyclic annulus gear contain main rotor shaft vibration information.

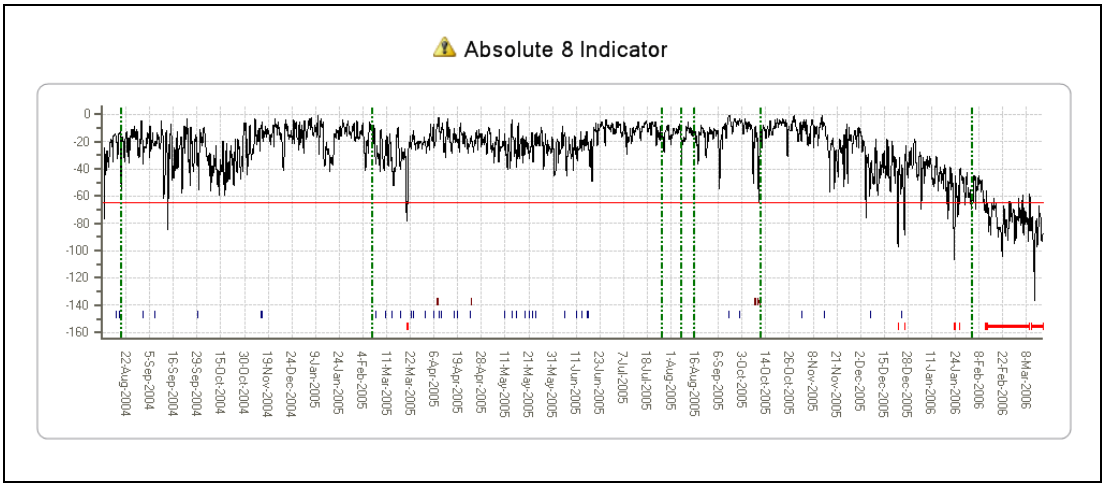

Figure 4-1 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Fitness Score

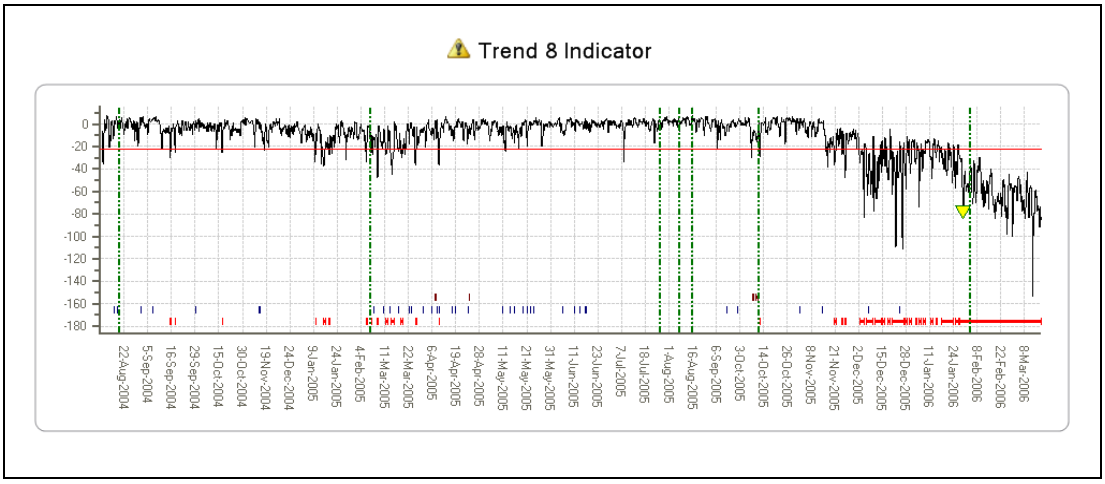

Figure 4-2 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Fitness Score

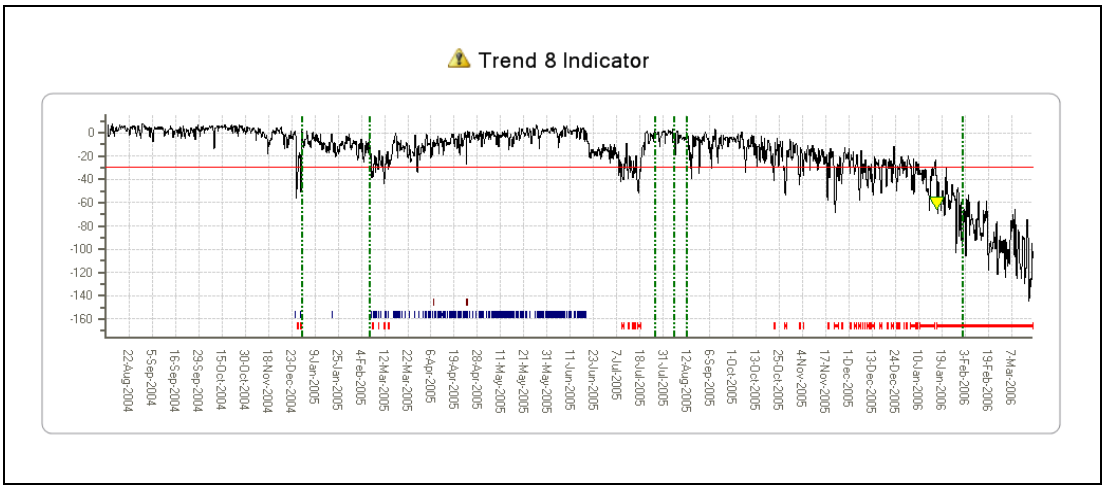

Figure 4-3 G-BWZX MGB 2nd epicyclic annulus left – Fitness Score

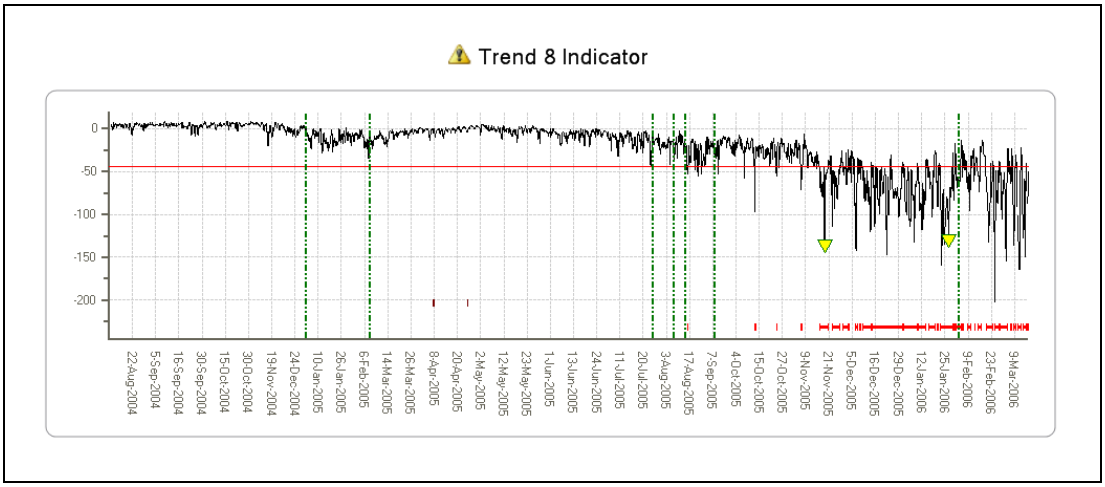

Figure 4-4 G-BWZX MGB 2nd epicyclic annulus forward (RH) – Fitness Score

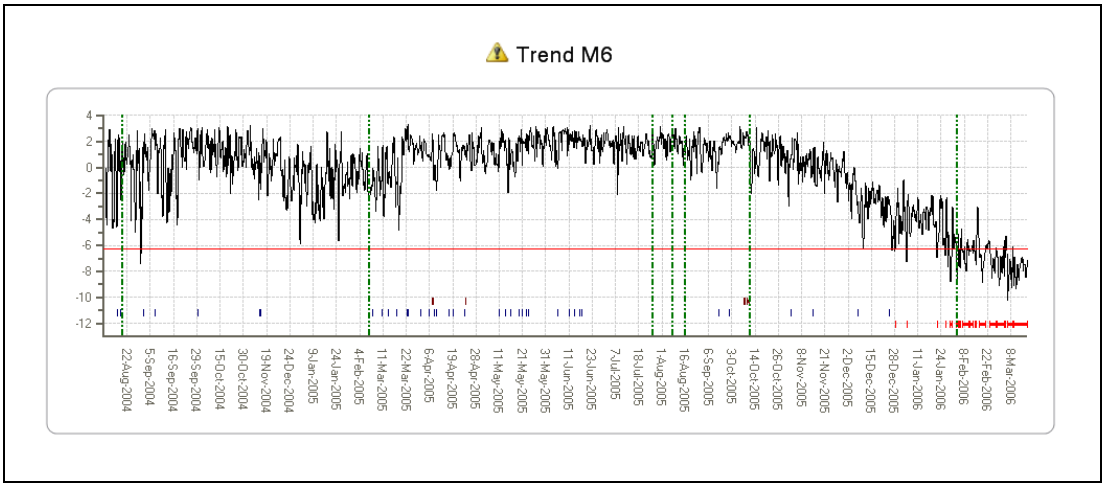

Figure 4-5 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Fitness Score

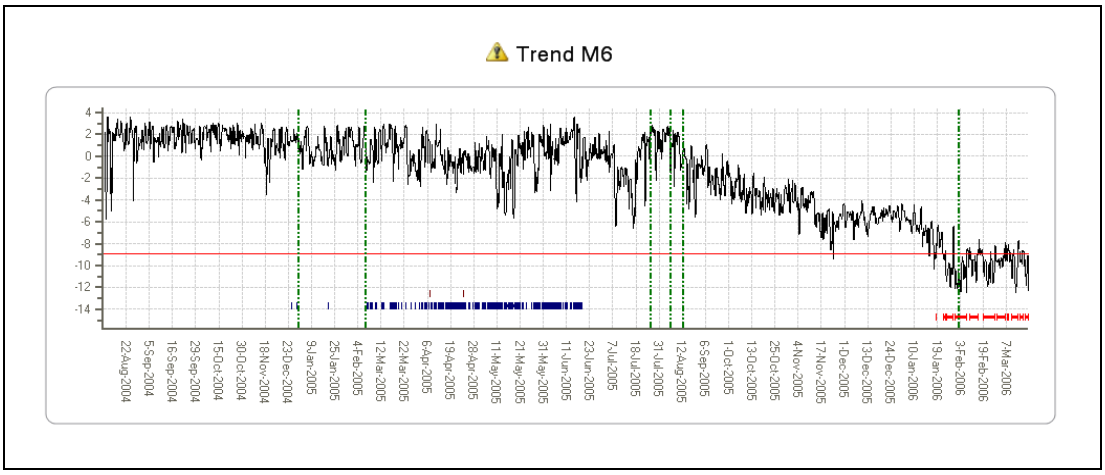

Figure 4-6 G-BWZX MGB 2nd epicyclic annulus left – Fitness Score

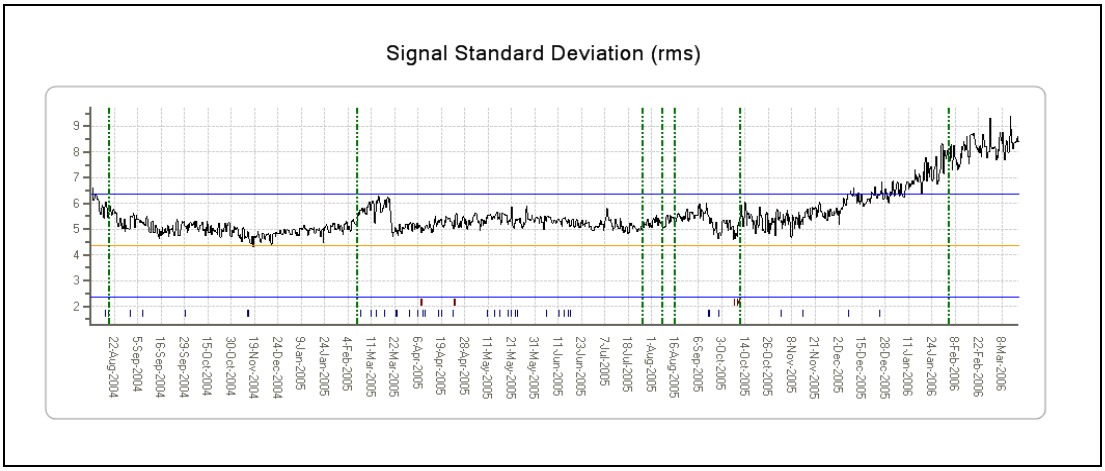

Figure 4-7 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Condition Indicator

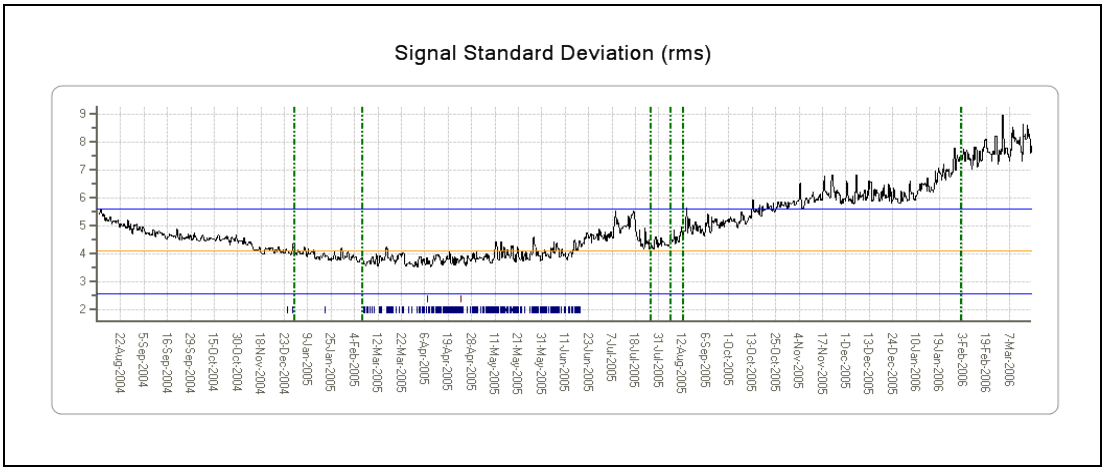

Figure 4-8 G-BWZX MGB 2nd epicyclic annulus left – Condition Indicator

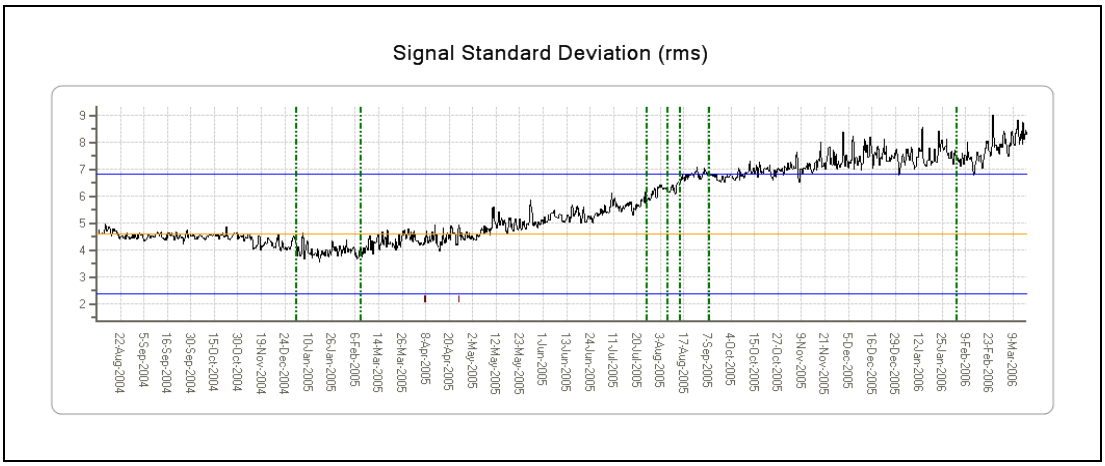

Figure 4-9 G-BWZX MGB 2nd epicyclic annulus forward (RH) – Condition Indicator

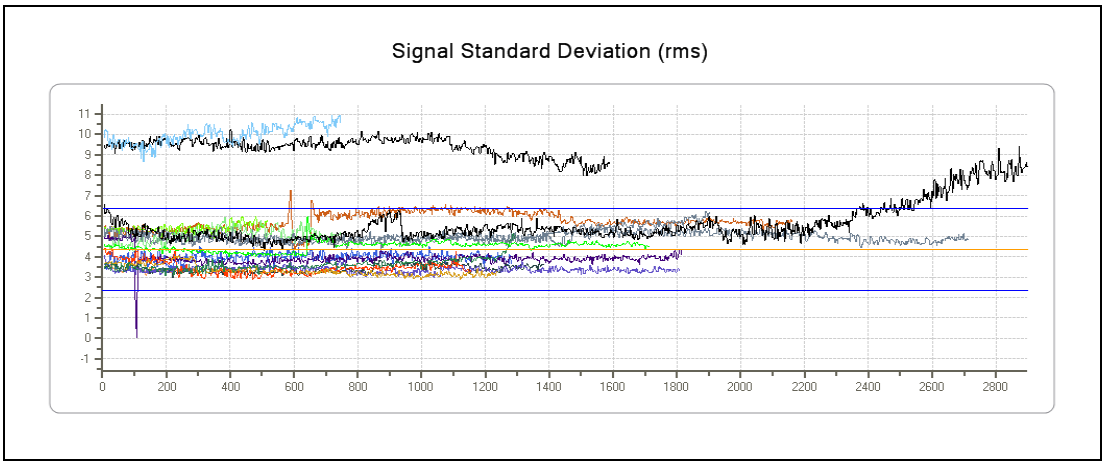

Figure 4-10 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Condition Indicator – fleet view

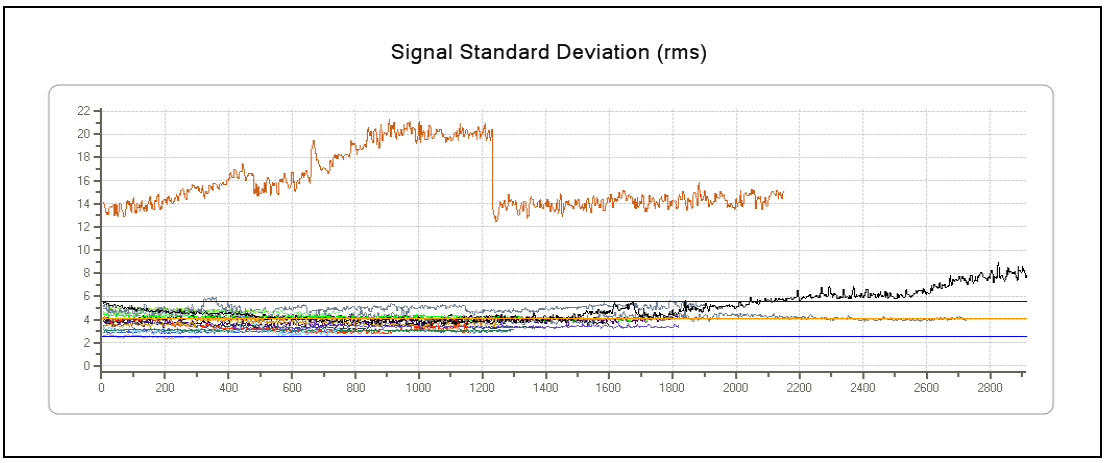

Figure 4-11 G-BWZX MGB 2nd epicyclic annulus left – Condition Indicator – fleet view

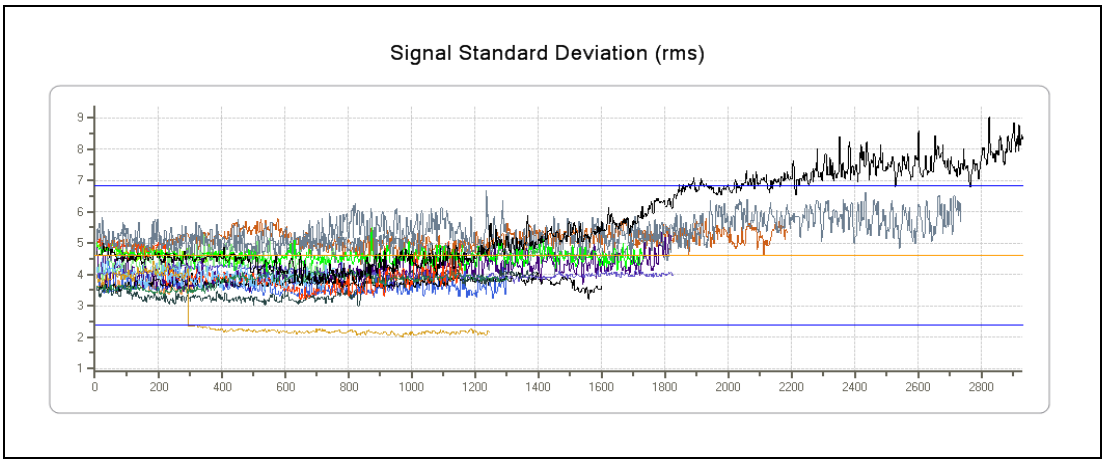

Figure 4-12 G-BWZX MGB 2nd epicyclic annulus forward (RH) – Condition Indicator – fleet view

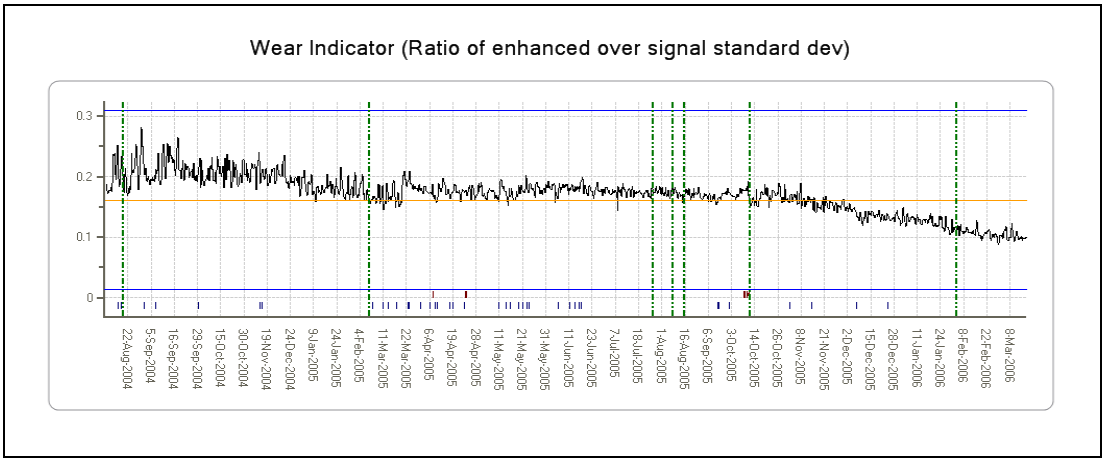

Figure 4-13 G-BWZX MGB 2nd epicyclic annulus aft (RH) – Condition Indicator

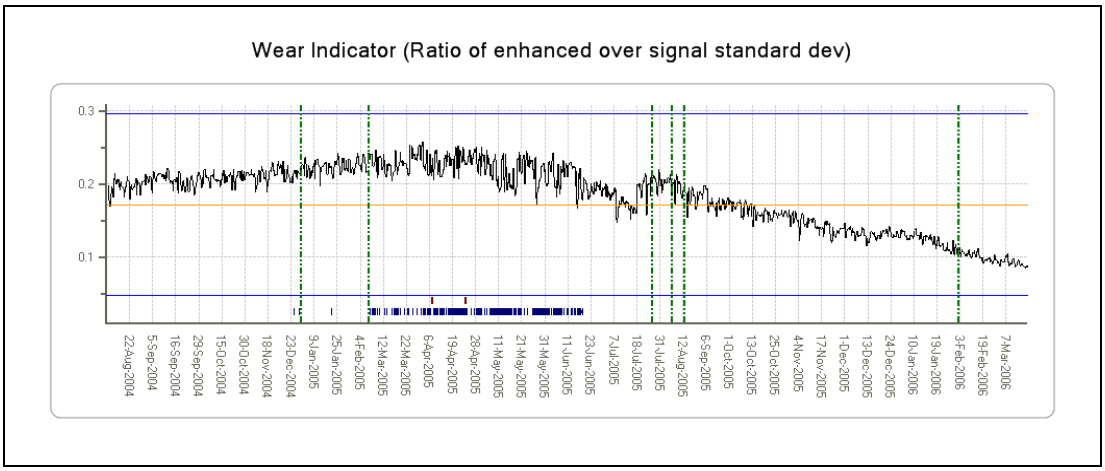

Figure 4-14 G-BWZX MGB 2nd epicyclic annulus left – Condition Indicator

Example 2: IHUMS DAPU (Instrumentation)

The IHUMS DAPU was fitted into G-TIGJ on 24 April 2006, and then removed on 18 May 2006 and fitted into G-TIGG. It was fitted back into G-TIGJ on 16 June 2006, and finally removed as unserviceable on 19 June 2006. There were no IHUMS alerts indicating the DAPU problem.

Figure 4-15 shows that the large number of anomaly alerts that were present on G-TIGJ on 18 May 2006 move to G-TIGG on 19 May 2006. Illustrating the effect of the DAPU change on one example MGB shaft, Figure 4-16 shows the FS trace from the '8-indicator' trend model for the MGB 1st stage planet gear in G-TIGJ, and Figure 4-17 shows the FS trace from the '8-indicator' absolute model for the same component in G-TIGG. The two periods of very low FSs on G-TIGJ and the one period on G-TIGG correlate with the dates on which the DAPU was fitted and removed from the two aircraft.

The anomaly detection system has clearly identified the DAPU problem. Returning a DAPU for repair can be expensive, and in some previous cases it has been difficult to conclusively demonstrate that a problem is related to the DAPU.

1 Week | 1 Month | 3 Months | All Latest Date: 18/05/2006

Tail	Notes	Alerts	Current Alerts	UnAcknowledged Alerts
G-BLPM	0	34	0	0
G-BLRY	0	44	6	6
G-BLXR	0	145	9	9
G-BMCW	0	48	1	1
G-BWWI	0	23	0	0
G-BWZX	1	69	1	1
G-PUMH	2	54	4	4
G-PUMI	0	11	5	5
G-TIGC	3	173	14	14
G-TIGE	3	53	11	11
G-TIGF	0	63	1	1
G-TIGG	0	128	3	3
G-TIGJ	0	212	25	24
G-TIGO	0	21	0	0
G-TIGS	0	92	6	6
G-TIGT	1	50	2	2
G-TIGV	0	147	1	0

1 Week | 1 Month | 3 Months | All Latest Date: 19/05/2006

Tail	Notes	Alerts	Current Alerts	UnAcknowledged Alerts
G-BLPM	0	34	0	0
G-BLRY	0	44	6	6
G-BLXR	0	148	2	2
G-BMCW	0	48	1	1
G-BWWI	0	23	0	0
G-BWZX	1	69	1	1
G-PUMH	2	53	2	2
G-PUMI	0	11	5	5
G-TIGC	3	171	14	14
G-TIGE	3	55	12	12
G-TIGF	0	64	2	2
G-TIGG	3	147	21	21
G-TIGJ	0	214	3	2
G-TIGO	0	21	0	0
G-TIGS	0	92	6	6
G-TIGT	1	49	2	2
G-TIGV	0	147	1	0

Figure 4-15 Aircraft summary display

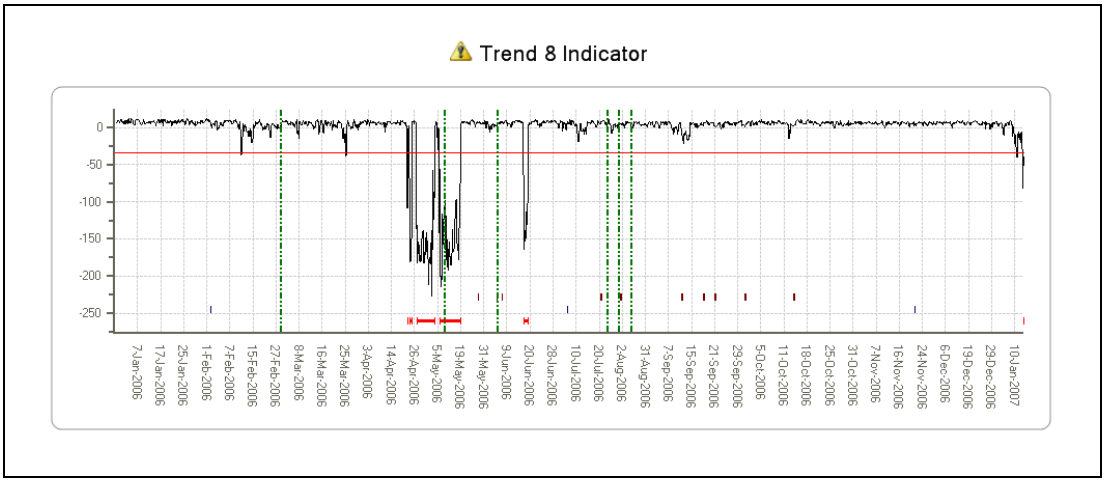

Figure 4-16 G-TIGJ MGB 1st stage planet gear – Fitness Score

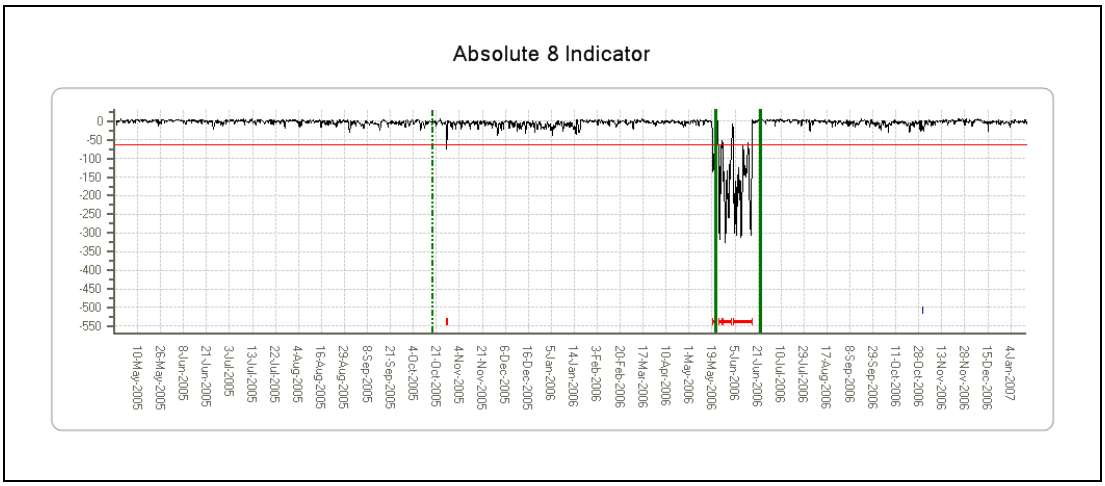

Figure 4-17 G-TIGG MGB 1st stage planet gear – Fitness Score

Example 3: IHUMS MGB Epicyclic Instrumentation

The '8-indicator' absolute models for all G-TIGC's MGB 1st and 2nd stage epicyclic gear analyses performed using data from accelerometer 7 (i.e. six different analyses) were in continuous alert from April 2005. The same models were also in continuous alert for the 1st stage annulus analyses performed using data from accelerometers 5 and 6. The IHUMS did not generate any alerts during this period.

Figure 4-18 shows the FS trace for the 1st epicyclic annulus aft (RH) data, which is flat-lining at its minimum allowed value. Figure 4-19 and Figure 4-20 show the Signal Standard Deviation indicator values for G-TIGC – there is no clear trend in the data, however values are many times higher than on the rest of the fleet. The MGB was removed on 13 December 2006 due to an aircraft rollover. When the new MGB was fitted all data returned to normal levels. It is considered that the anomaly alerts were identifying an MGB epicyclic stage wiring harness problem.

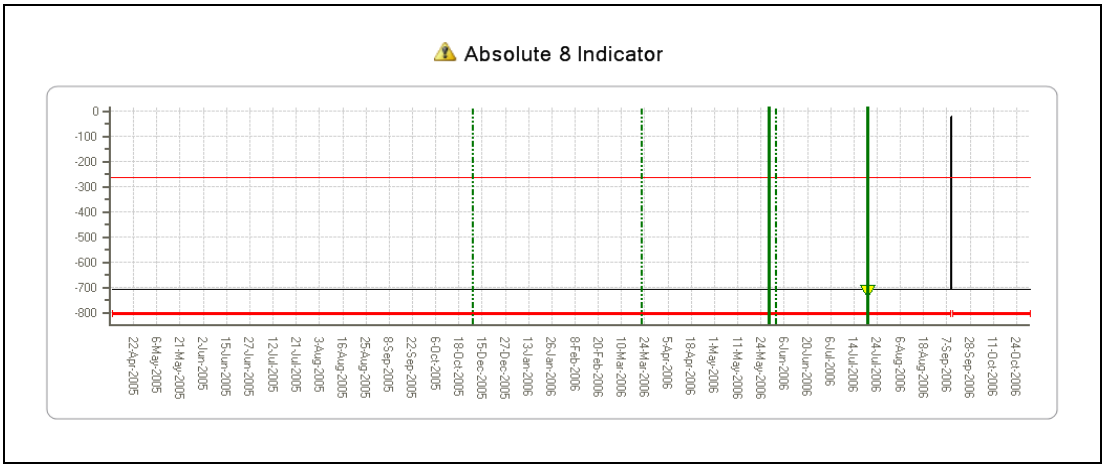

Figure 4-18 G-TIGC MGB 1st epicyclic annulus aft (RH) – Fitness Score

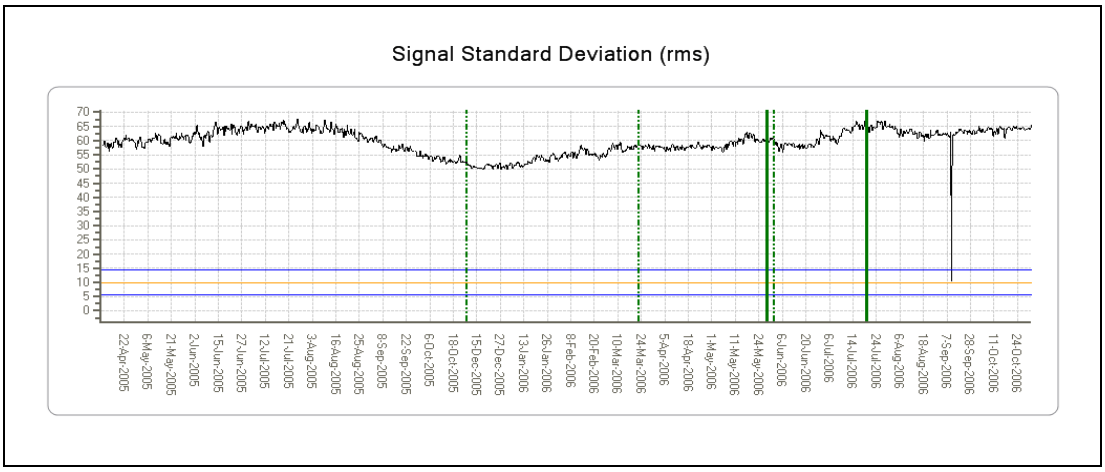

Figure 4-19 G-TIGC MGB 1st epicyclic annulus aft (RH) – Condition Indicator

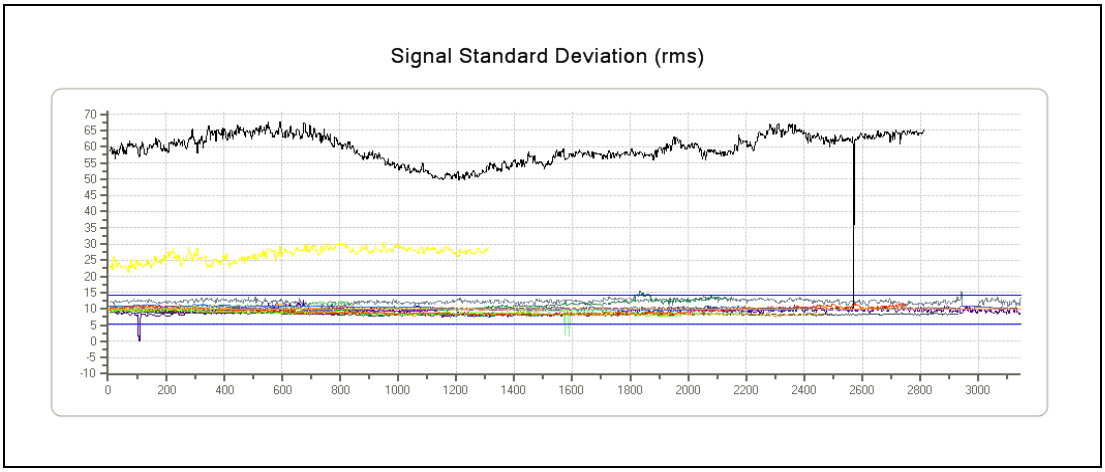

Figure 4-20 G-TIGC MGB 1st epicyclic annulus aft (RH) – Condition Indicator – fleet view

Example 4: IHUMS MGB Epicyclic Instrumentation

This example is very similar to the previous case. Multiple '8-indicator' absolute models for G-TIGS's MGB epicyclic gear analyses performed using data from accelerometer 7 were in continuous alert from May 2006 to 8 January 2007 when the gearbox reached the end of its service life and was removed. Again, there were no IHUMS alerts.

Figure 4-21 shows the FS trace for the 2nd epicyclic annulus aft (RH) data, which is at a very low value until the gearbox change at which point it returns to normal. Figure 4-22 - Figure 4-25 show the Enhanced and Signal Standard Deviation indicator values for G-TIGS, which are much higher than on the rest of the fleet. Accelerometer 7 was swapped with accelerometer 3 on 31 August 2007, however this had no effect on the data. It is therefore again considered that the anomaly alerts were identifying an MGB epicyclic stage wiring harness problem associated with accelerometer 7.

The problem was eliminated when the gearbox was changed. At a gearbox change the accelerometers and harness are normally moved across from the old to the new gearbox. However, the cleanliness of the wiring on the replacement gearbox indicated that a new wiring harness had been fitted. A new standard of accelerometer wiring had been previously trialled on a limited number of aircraft, and it was believed that this had reduced the instrumentation problems associated with the MGB

epicyclic stage. However, no decision had been made to use this on a fleet wide basis, and the old standard of cable is still being used for the accelerometer wiring harnesses.

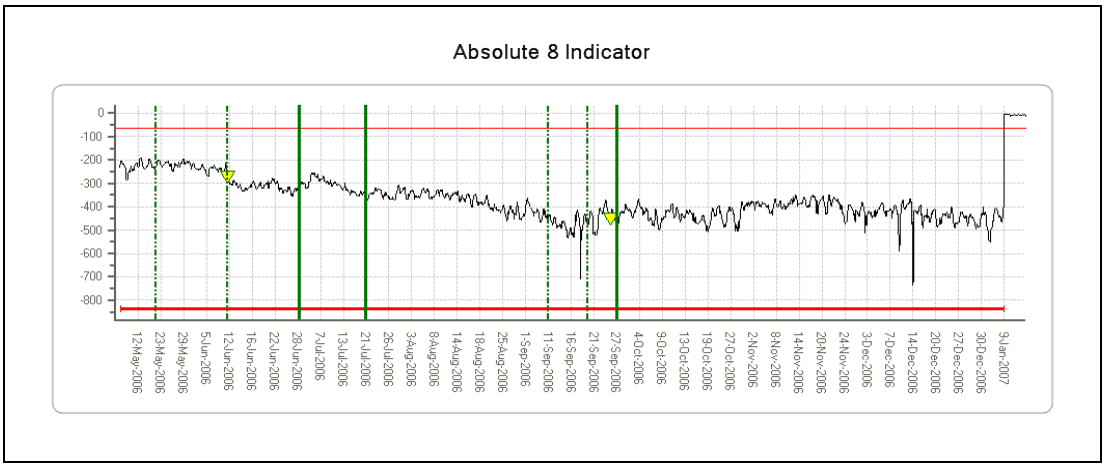

Figure 4-21 G-TIGS MGB 2nd epicyclic annulus aft (RH) – Fitness Score

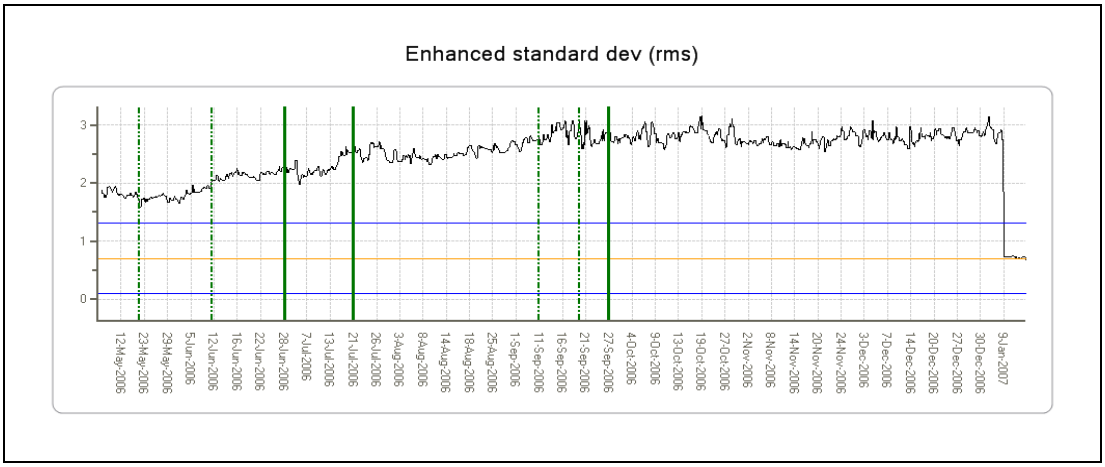

Figure 4-22 G-TIGS MGB 2nd epicyclic annulus aft (RH) – Condition Indicator

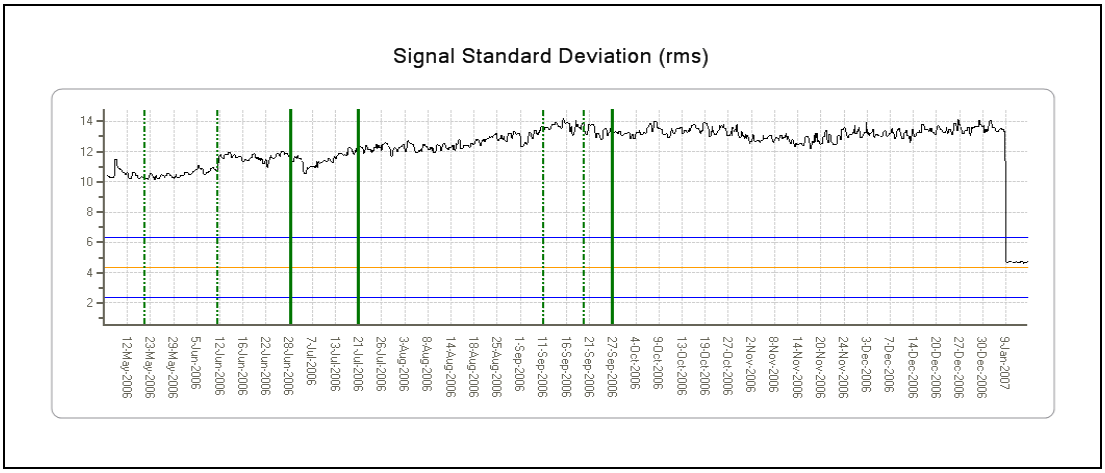

Figure 4-23 G-TIGS MGB 2nd epicyclic annulus aft (RH) – Condition Indicator

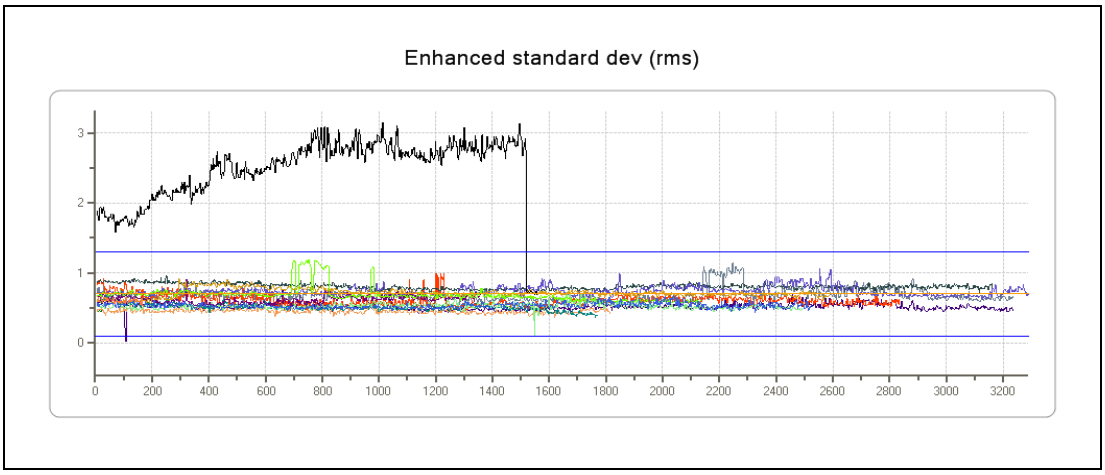

Figure 4-24 G-TIGS MGB 2nd epicyclic annulus aft (RH) – Condition Indicator – fleet view

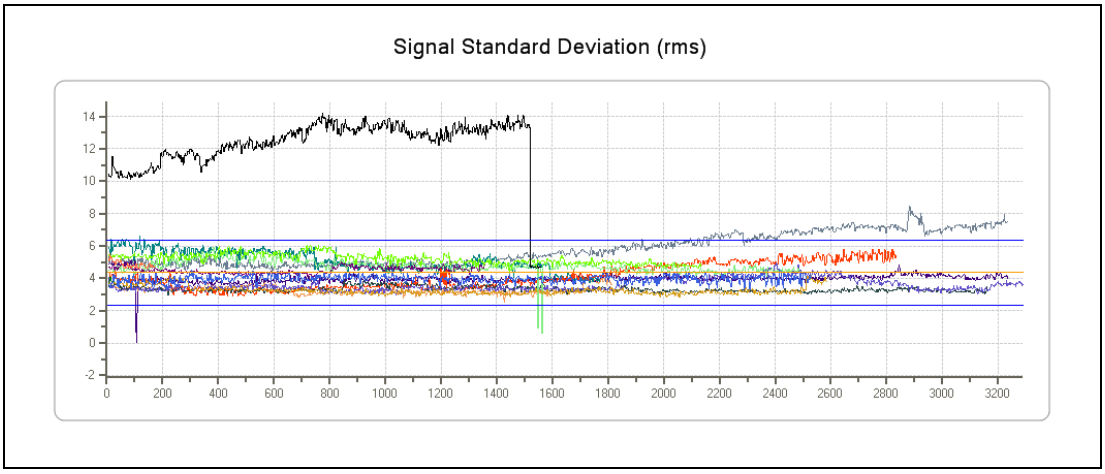

Figure 4-25 G-TIGS MGB 2nd epicyclic annulus aft (RH) – Condition Indicator – fleet view

Example 5: Accelerometers 3 and 4 Wiring Found Swapped

The '8-indicator' absolute model for G-TIGE's MGB bevel wheel and oil pump drive was in continuous alert from July 2006 to October 2006, and the FS did not return to normal values until the RHA Module was replaced on 5 November 2006 (Figure 4-26). The key CIs that were driving the low FSs were the low amplitude of the bevel wheel mesh magnitude, and high and erratic SO1 values (Figure 4-27 – Figure 4-30).

The cause of the high and erratic SO1 values is unknown, however it is the low amplitude bevel wheel mesh magnitude that is of most interest. There is no reason why changing the RHA Module should affect this. In fact, when the RHA Module was replaced following IHUMS warnings on the RH alternator drive the wiring for accelerometers 3 and 4 was found to be swapped. The bevel wheel mesh magnitude returned to its normal level when the accelerometer wiring was corrected. There is no information on when the wiring error first occurred. The bevel wheel data may indicate that the problem existed since the time that the MGB was installed in July 2006, but it is not possible to confirm this.

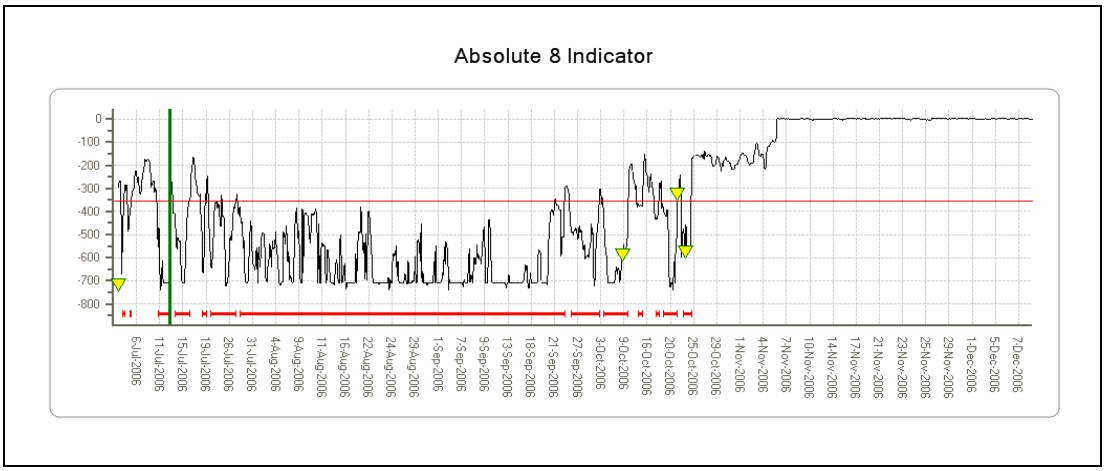

Figure 4-26 G-TIGE - MGB - bevel wheel and oil pump drive – Fitness Score

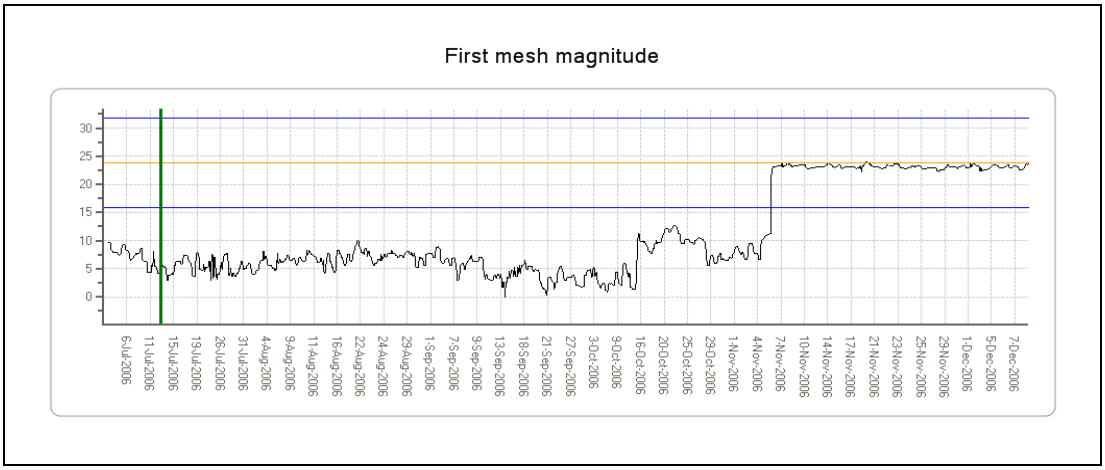

Figure 4-27 G-TIGE - MGB - bevel wheel and oil pump drive – Condition Indicator

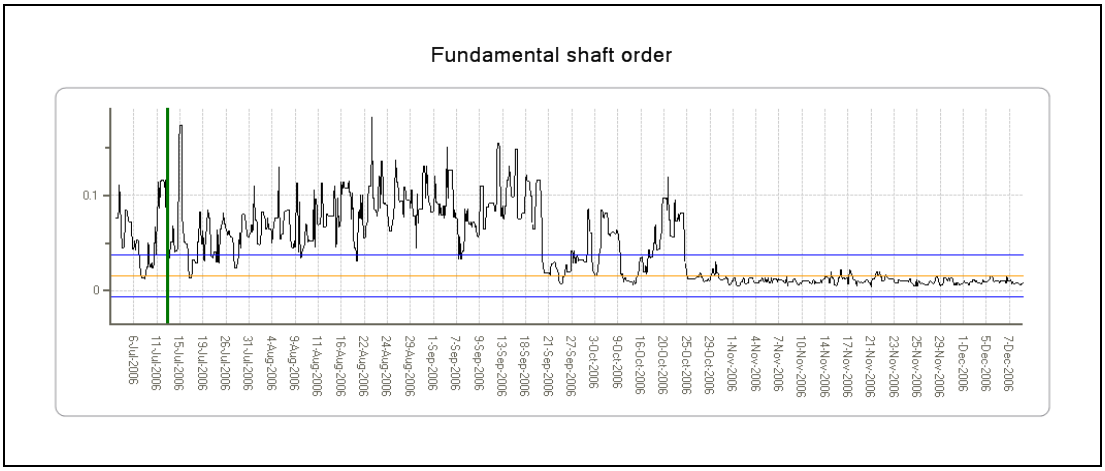

Figure 4-28 G-TIGE - MGB - bevel wheel and oil pump drive – Condition Indicator

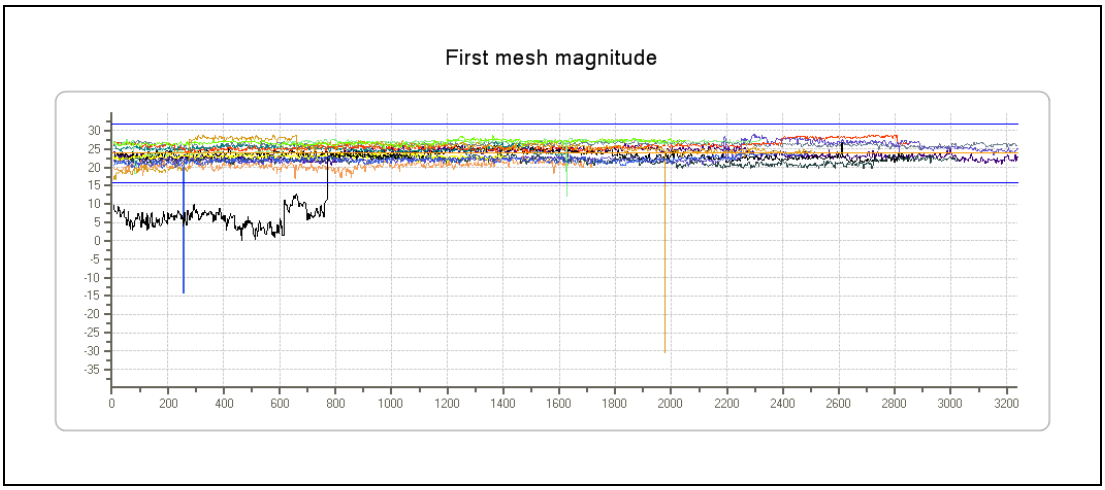

Figure 4-29 G-TIGE - MGB - bevel wheel and oil pump drive – Condition Indicator – fleet view

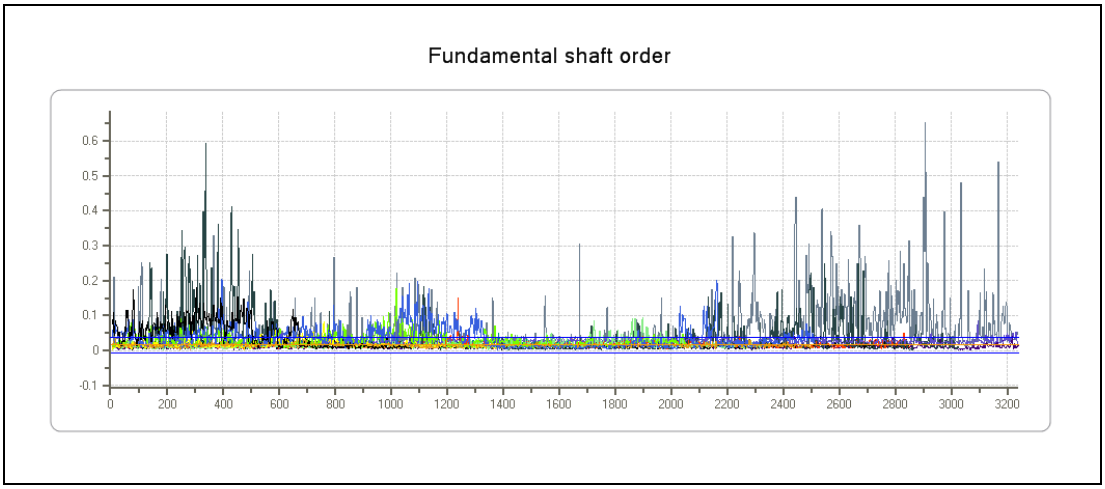

Figure 4-30 G-TIGE - MGB - bevel wheel and oil pump drive – Condition Indicator – fleet view

Example 6: Accelerometer 3 (Instrumentation)

The '8-indicator' absolute models for G-TIGG's LHA left hydraulic idler and left hydraulic drive 47-tooth gear produced low and variable FSs from late 2005, which generated a large number of alerts (Figure 4-31 and Figure 4-36). The CIs that were driving the low FSs were the very high and erratic SO1 and SO2 (1/rev and 2/rev) values (Figure 4-32 - Figure 4-35 and Figure 4-37 - Figure 4-40). A review of the IHUMS data confirmed that the majority of aircraft have low SO1 and SO2 values, but a few aircraft have high values.

The LH hydraulic pump seized on 2 December 2006, but there were no changes in values after pump replacement. No. 3 and 4 accelerometers were swapped on 9 November 2006, but this had no effect on the data. Finally No. 3 accelerometer was swapped with No. 6 on 20 January 2006 and the SO1 and SO2 values returned to normal. It is concluded that the abnormal SO1 and SO2 values were caused by an IHUMS instrumentation problem (either with No. 3 accelerometer or its cable and connector). It is not known exactly what maintenance was performed when accelerometer 3 was first swapped; for example whether the insulation sleeve and washer were changed at the same time. Detailed differences in the maintenance actions may explain why the problem was not rectified at the first attempt.

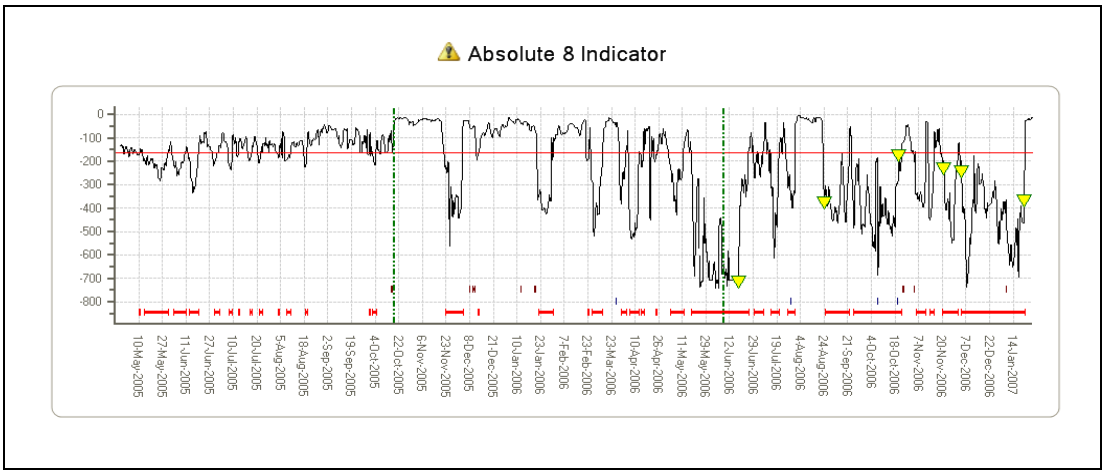

Figure 4-31 G-TIGG - LHA - left hydraulic idler – Fitness Score

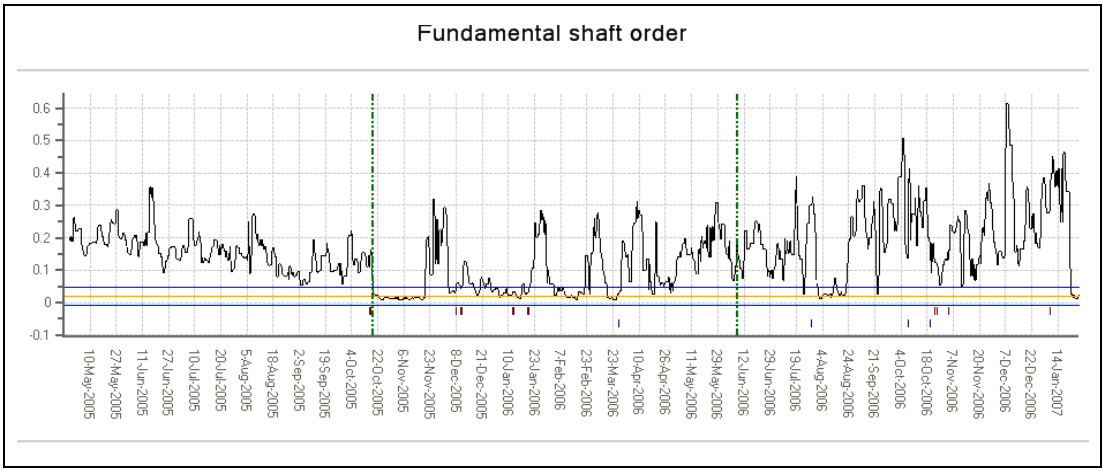

Figure 4-32 G-TIGG - LHA - left hydraulic idler – Condition Indicator

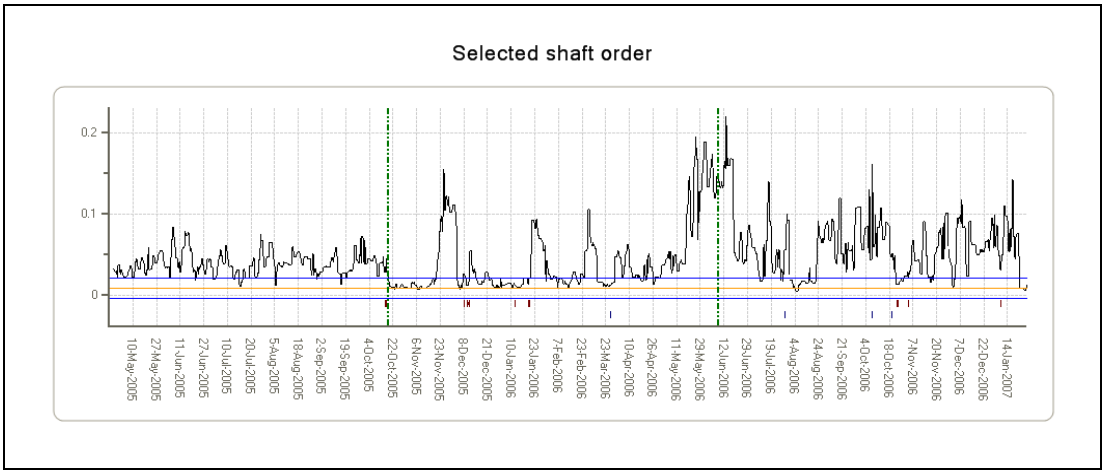

Figure 4-33 G-TIGG - LHA - left hydraulic idler – Condition Indicator

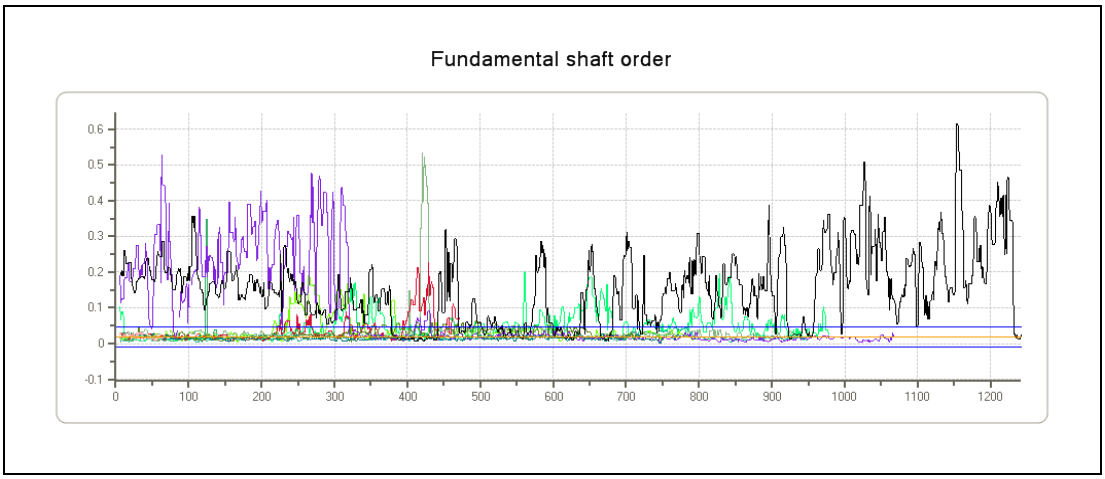

Figure 4-34 G-TIGG - LHA - left hydraulic idler – Condition Indicator – fleet view

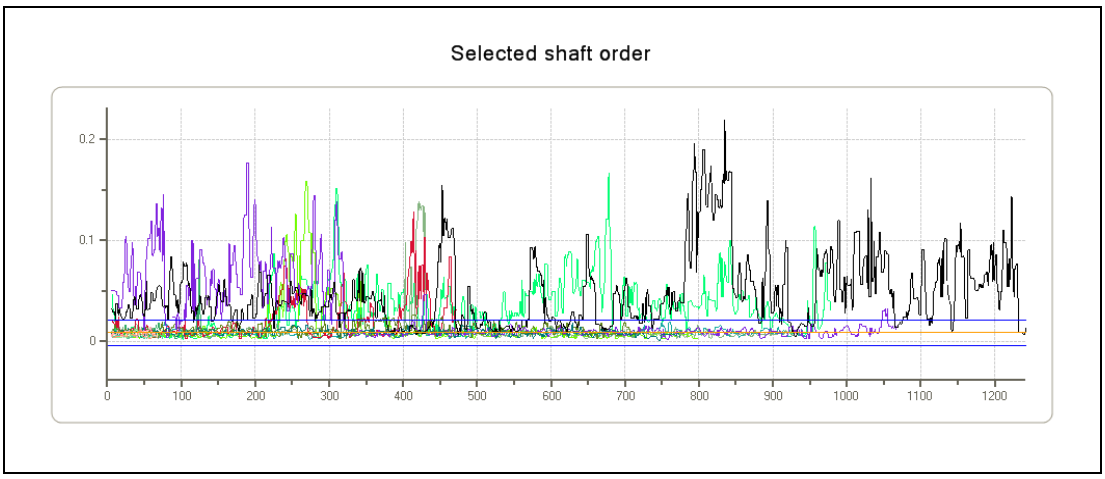

Figure 4-35 G-TIGG - LHA - left hydraulic idler – Condition Indicator – fleet view

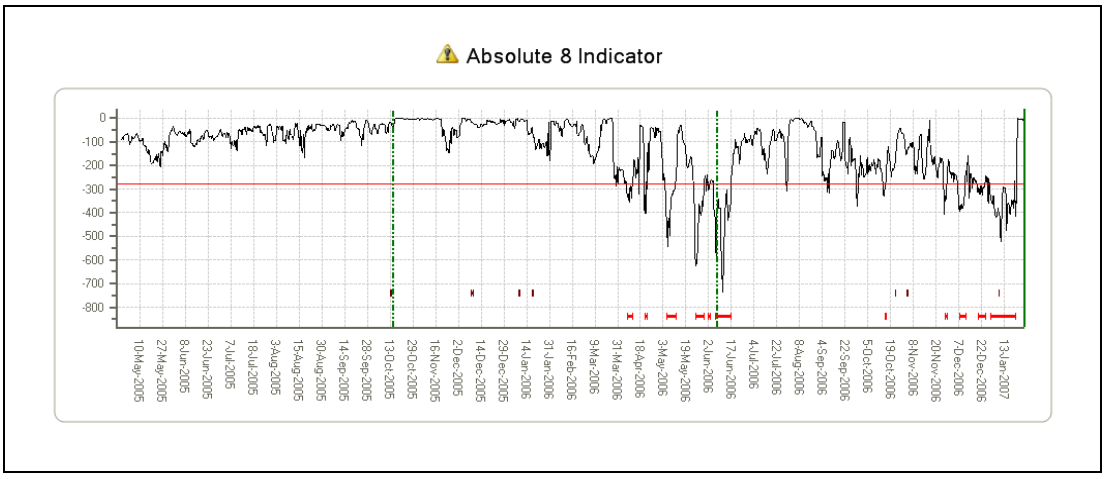

Figure 4-36 G-TIGG - LHA - left hydraulic drive 47-tooth gear – Fitness Score

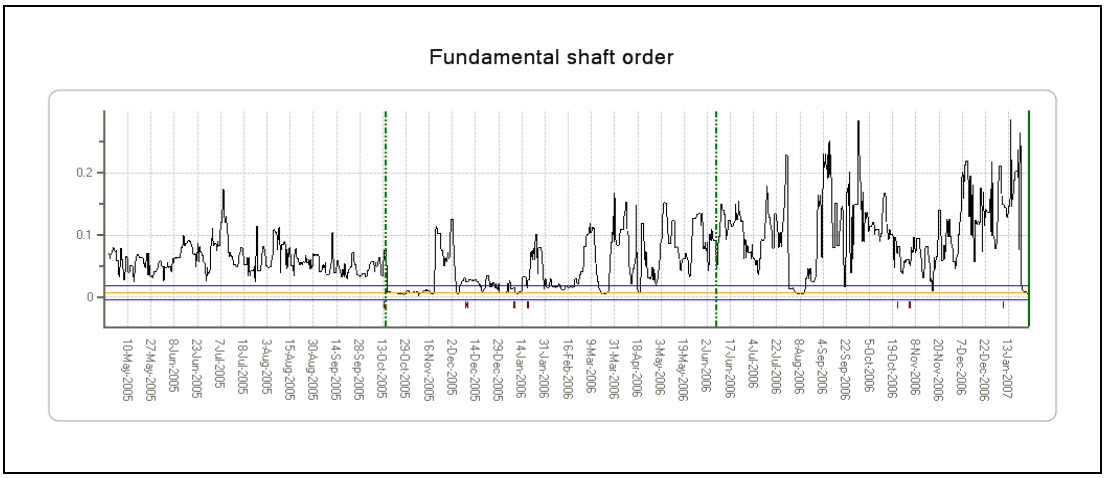

Figure 4-37 G-TIGG - LHA - left hydraulic drive 47-tooth gear – Condition Indicator

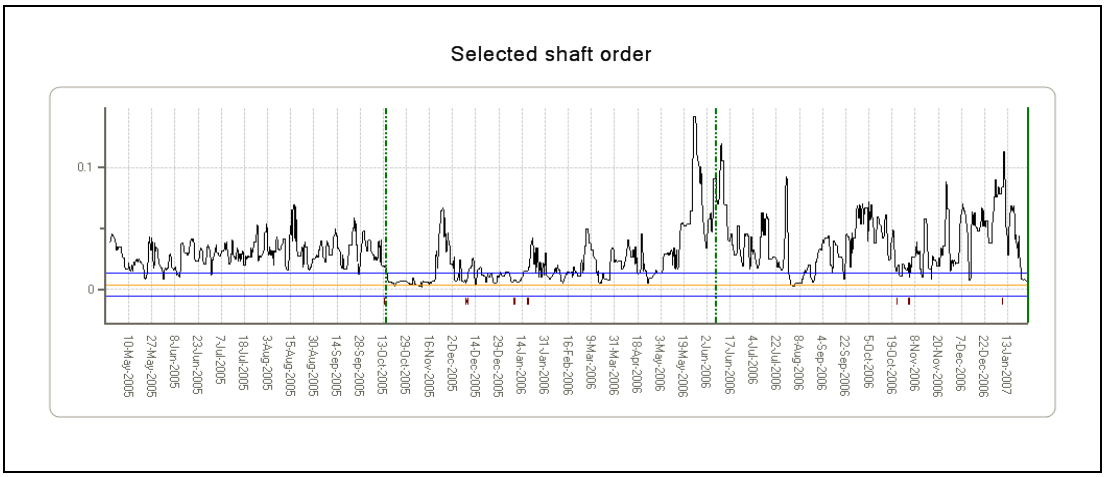

Figure 4-38 G-TIGG - LHA - left hydraulic drive 47-tooth gear – Condition Indicator

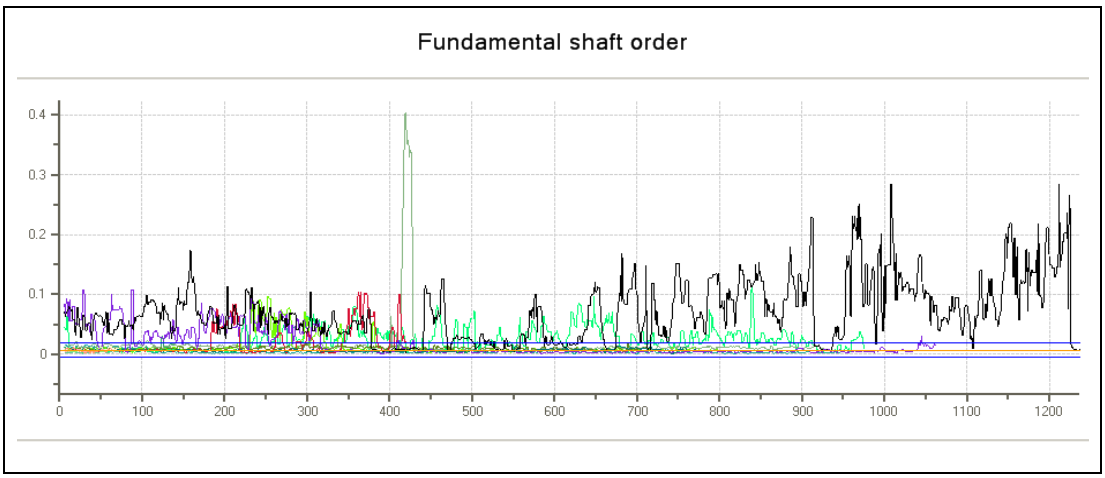

Figure 4-39 G-TIGG - LHA - left hydraulic drive 47-tooth gear – Condition Indicator – fleet view

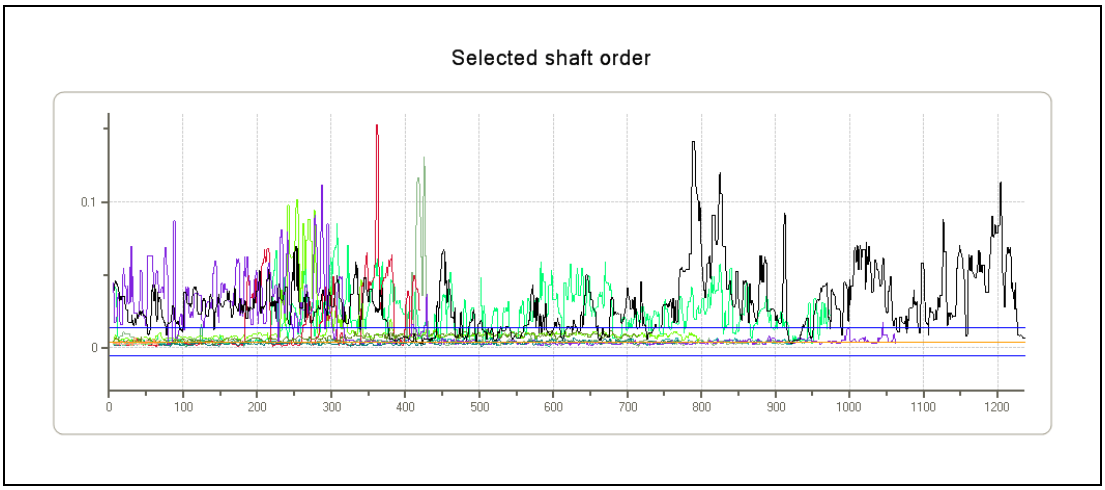

Figure 4-40 G-TIGG - LHA - left hydraulic drive 47-tooth gear – Condition Indicator – fleet view

4.2 Cases where Anomaly Detection has Corroborated Existing HUMS Indications

Example 1: LHA Module Rejected

The '8-indicator' absolute model for G-BLXR's LHA left hydraulic idler generated periodic alerts in March 2006, and then a continuous alert from May 2006 (Figure 4-41). The 'M6' absolute model generated consistent alerts from the beginning of November 2006 (Figure 4-42). Although occasional alerts occurred before then, the IHUMS generated frequent alerts from June 06, resulting in the module being placed on close monitor. The accelerometer insulation washer was replaced on 4 July 2006 and the IHUMS re-datumed. These actions cleared the IHUMS alerts, but had no effect on the data. Frequent IHUMS alerts commenced again in November 2006, and the module was again on close monitor. The module was finally rejected in December 2006 when the enhanced impulsiveness (M6*) indicator values reached 200 (Figure 4-43 and Figure 4-44). The indicator trend was mirrored in the data for the RHA right hydraulic idler, but at lower levels. The Enhanced Peak-Peak/Standard Deviation and Signal Peak-Peak/Standard Deviation indicators also had increasing trends, reaching high but stable levels from July 2006 (Figure 4-45 - Figure 4-48).

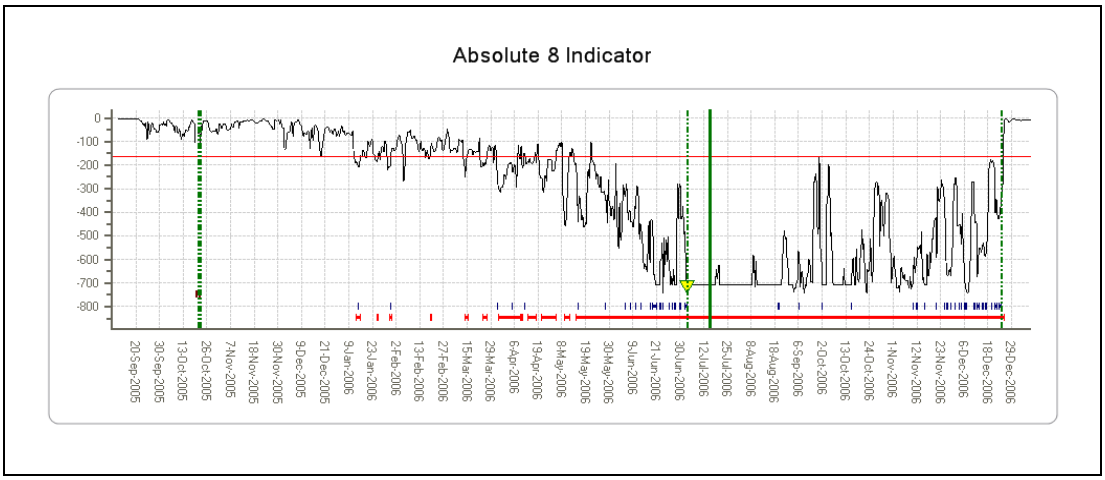

Figure 4-41 G-BLXR LHA left hydraulic idler – Fitness Score

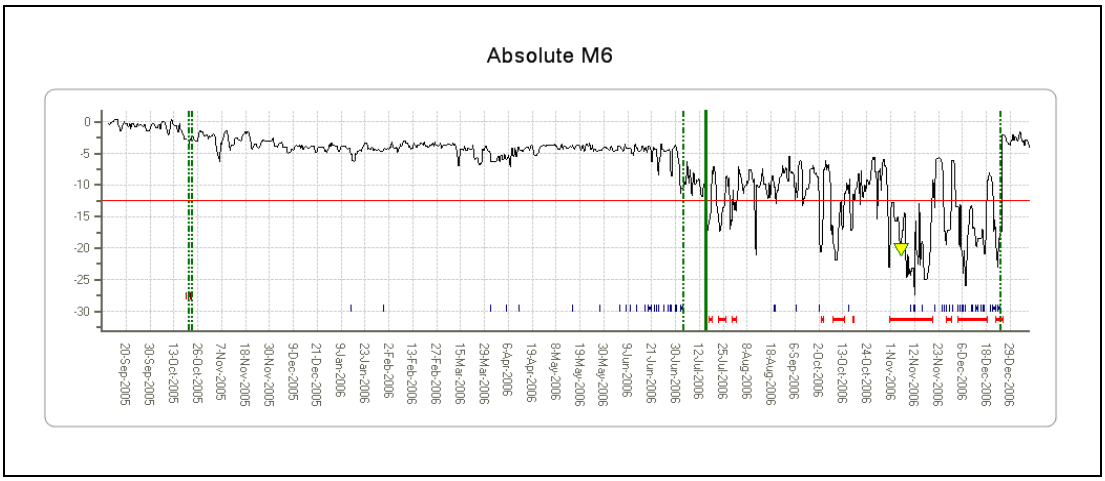

Figure 4-42 G-BLXR LHA left hydraulic idler – Fitness Score

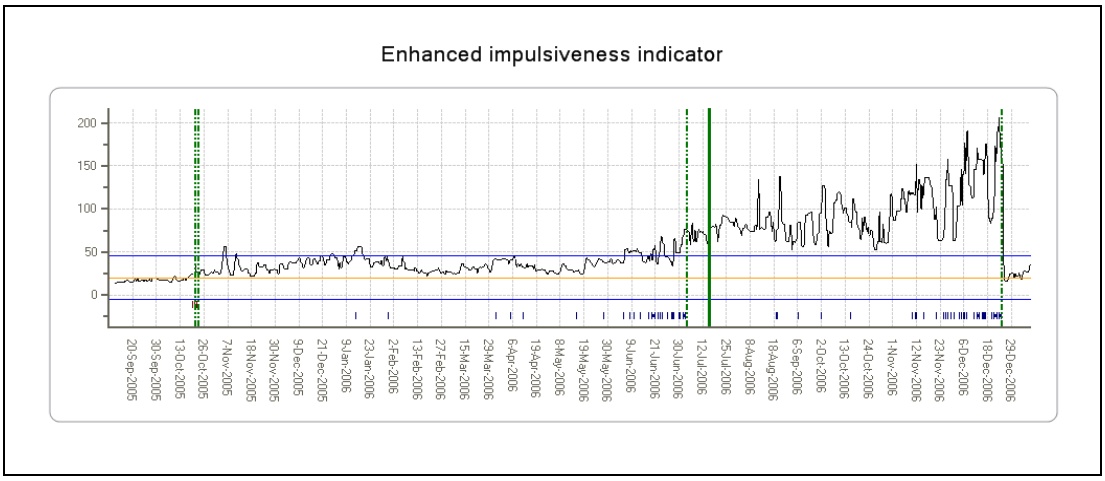

Figure 4-43 G-BLXR LHA left hydraulic idler – Condition Indicator

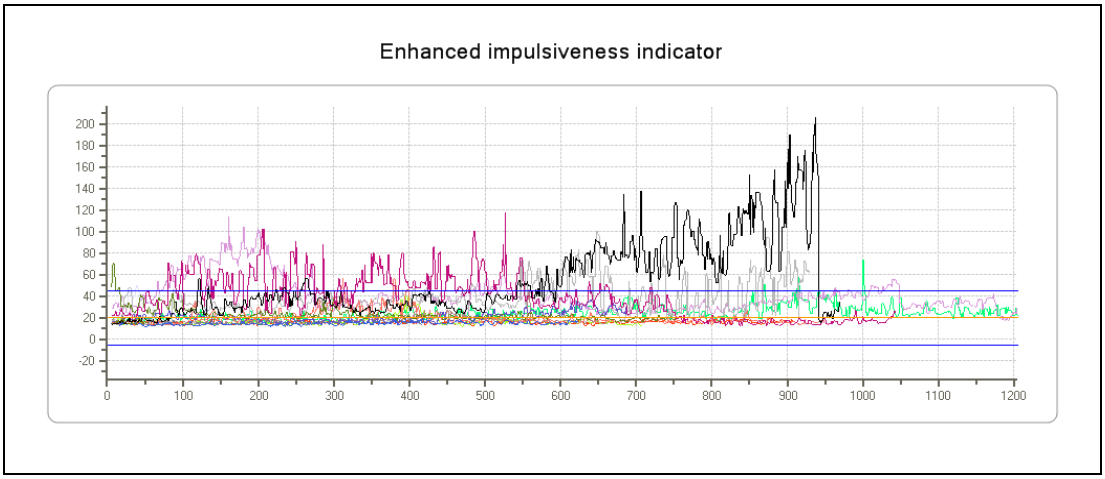

Figure 4-44 G-BLXR LHA left hydraulic idler – Condition Indicator – fleet view

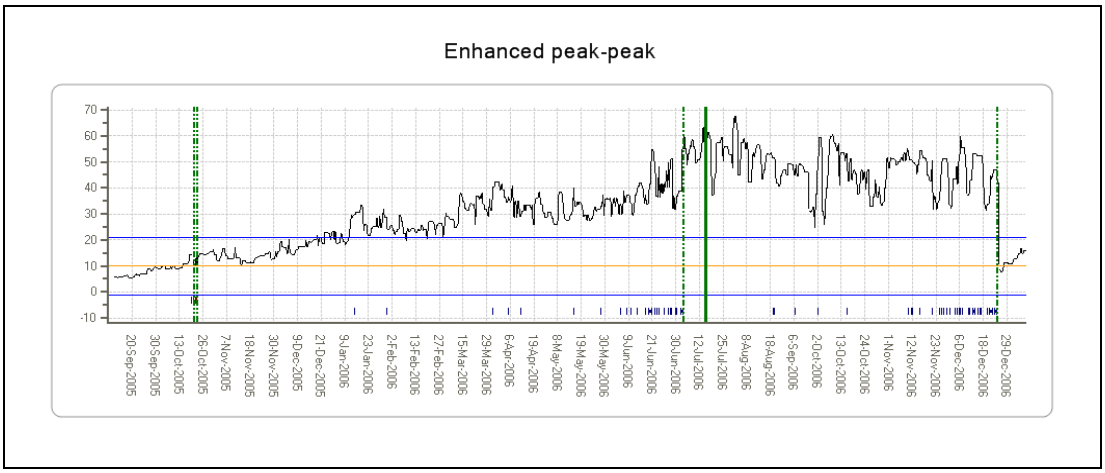

Figure 4-45 G-BLXR LHA left hydraulic idler – Condition Indicator

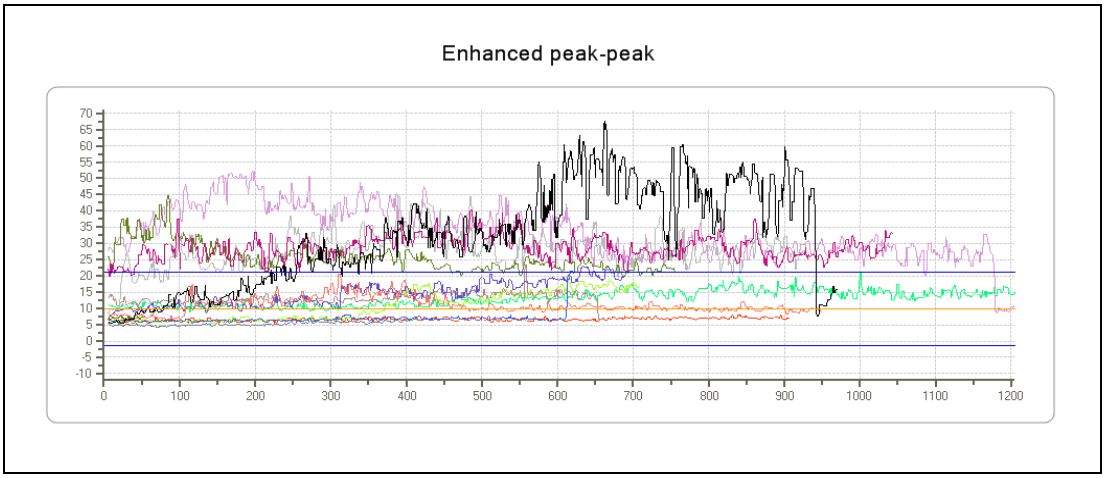

Figure 4-46 G-BLXR LHA left hydraulic idler – Condition Indicator – fleet view

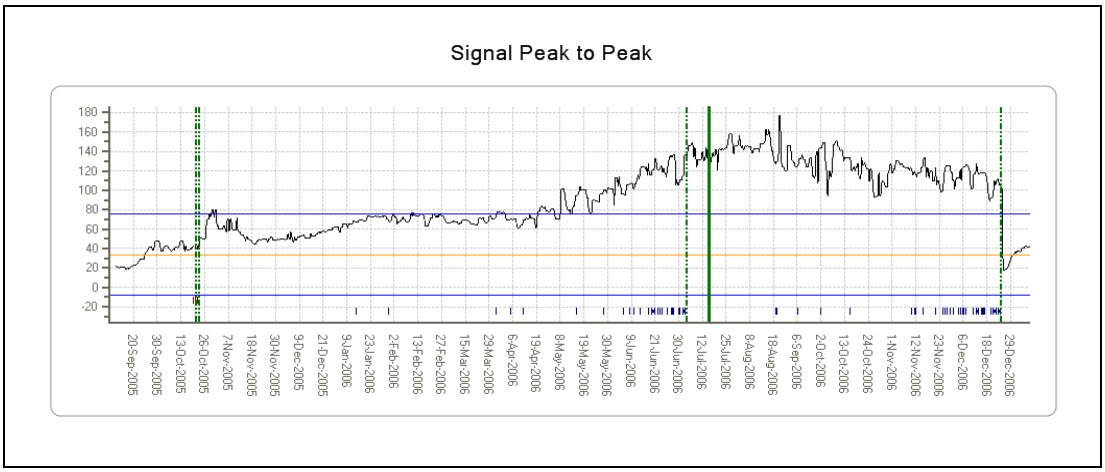

Figure 4-47 G-BLXR LHA left hydraulic idler – Condition Indicator

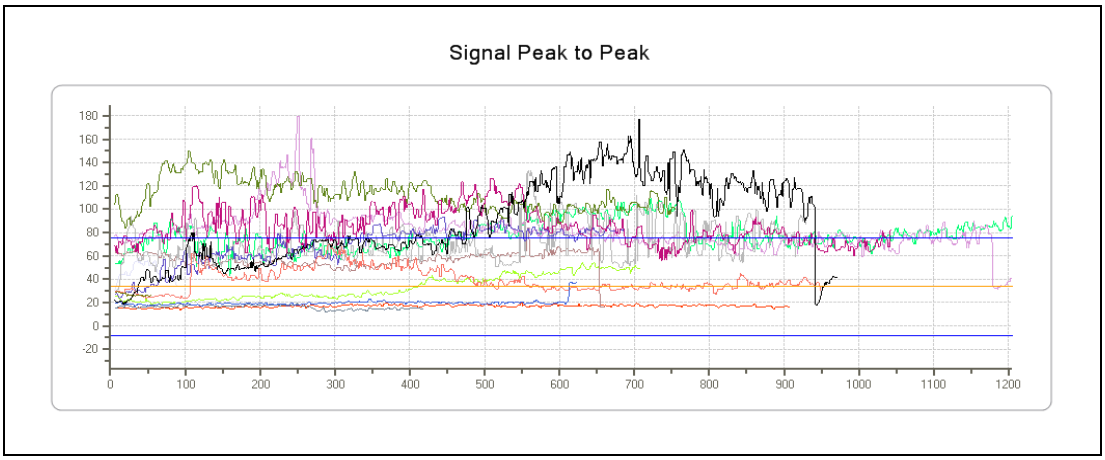

Figure 4-48 G-BLXR LHA left hydraulic idler – Condition Indicator – fleet view

Example 2: Oil Cooler

The '5-indicator' absolute model for G-BWZX's oil cooler fan generated a continuous alert from March 2006 (Figure 4-49). The cause was very high levels for the SO2 (2/rev) vibration and Signal Standard Deviation indicators (Figure 4-50 - Figure 4-53). The IHUMS was also generating alerts. The oil cooler fan was replaced on 23 April 2006 and the vibration data returned to normal. No fan or bearing defects were found, however there was evidence that the fan had been rubbing on the casing, which may have been distorted as a result of the metal securing band being too tight.

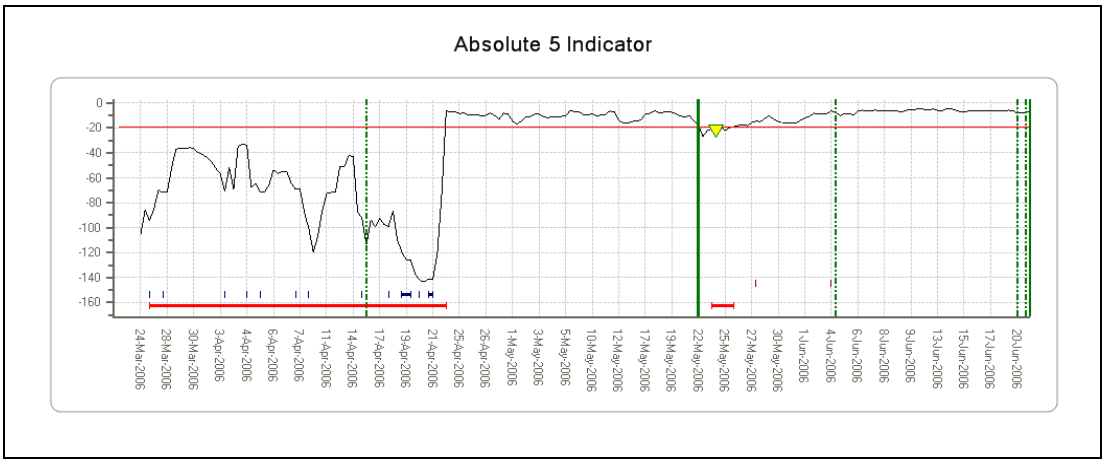

Figure 4-49 G-BWZX oil cooler fan – Fitness Score

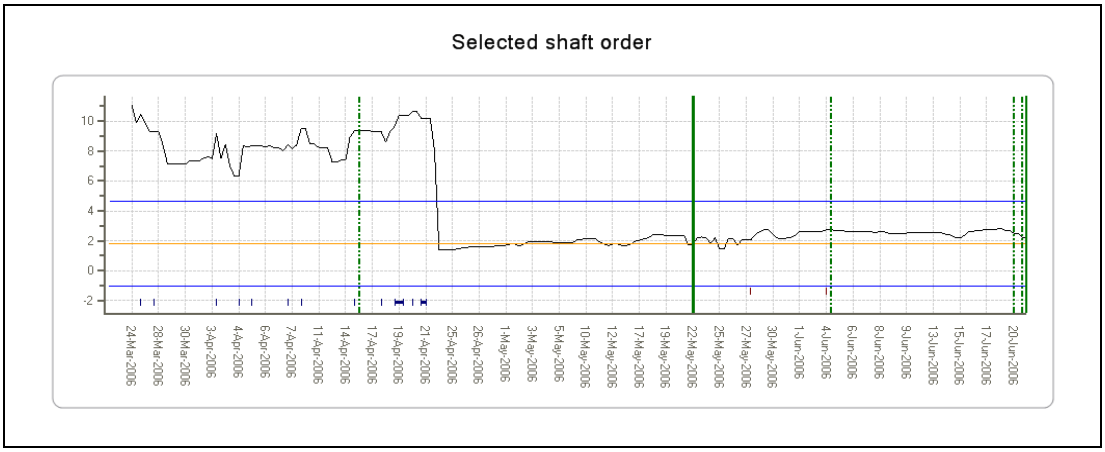

Figure 4-50 G-BWZX oil cooler fan – Condition Indicator

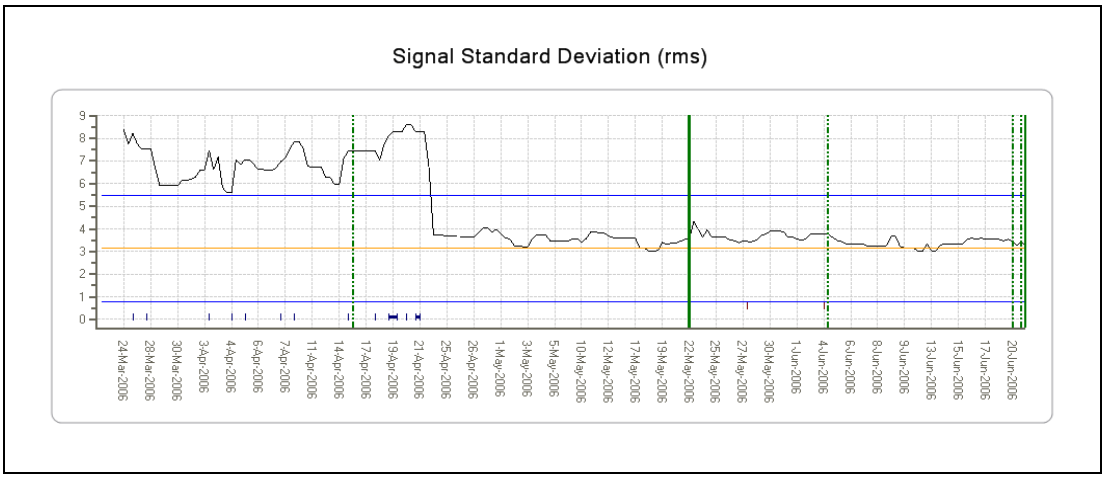

Figure 4-51 G-BWZX oil cooler fan – Condition Indicator

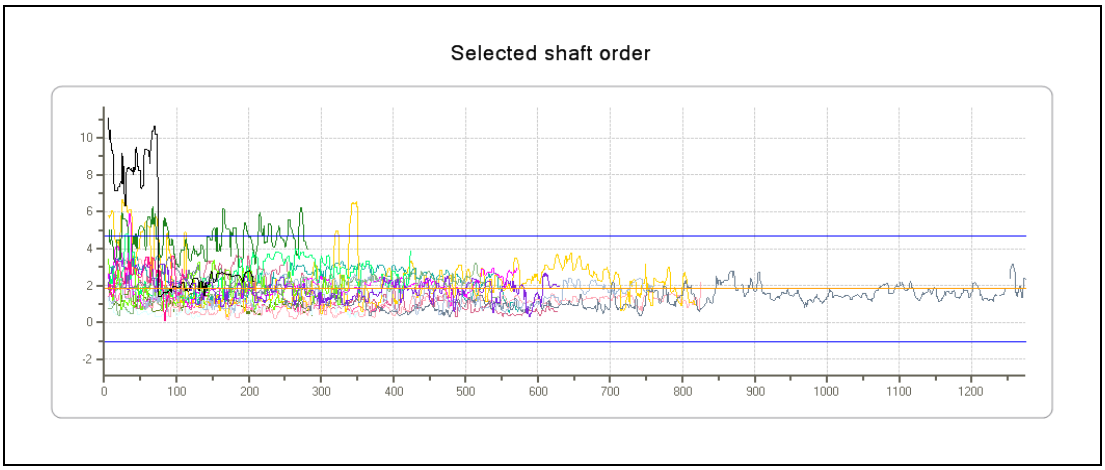

Figure 4-52 G-BWZX oil cooler fan – Condition Indicator – fleet view

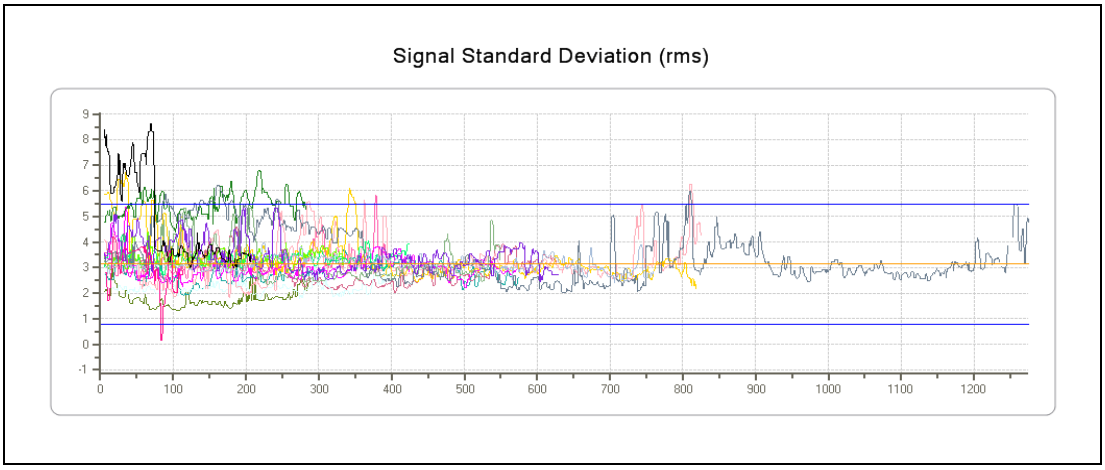

Figure 4-53 G-BWZX LHA oil cooler fan – Condition Indicator – fleet view

Example 3: Aircraft Instrumentation

The '8-indicator' absolute models were generating continual alerts from May or July 2005 on three of G-PUMI's MGB shafts: RH high speed input shaft, right torque shaft - fwd end, and 2nd epicyclic annulus fwd (RH) (Figure 4-54 - Figure 4-56). The IGB output shaft '8-indicator' and 'M6' absolute models were also generating alerts from May 2005 (Figure 4-57 - Figure 4-58). The MGB had been replaced on 6 July 2005 during a period of heavy maintenance, at which time the IGB was also removed and refitted. The cause of the anomaly alerts was very low accelerometer signal levels, which created the very low First Mesh Magnitude and Signal Standard Deviation indicator values seen in Figure 4-59 - Figure 4-62 (the indicators trends for the right torque shaft - fwd end mirror those for the RH high speed input shaft, and are therefore not shown). The high Wear indicator values for the IGB output shaft indicate that the low amplitude signal has a large noise content (Figure 4-63).

The IHUMS generated a significant number of convergence faults on the IGB, and also a number of alerts and sensor faults, most notably on the MGB 2nd epicyclic annulus fwd (RH). The No. 5 accelerometer installation was inspected on 23 May 2006 and the insulation washers were replaced, however this had no effect on the data. The No. 2 accelerometer installation was inspected and the No. 1 and 2 accelerometers were swapped on 30 May 2006, however this again had no effect on the data. The IGB accelerometer installation was inspected on 23 June 2006, but no findings were noted. The IHUMS DAPU was replaced on 23 September 2006, but again this did not have any effect.

It was suspected that there was an undetected IHUMS instrumentation issue on the aircraft, and the wiring harnesses were checked during a period of heavy maintenance in January 2007. It was noted that the earthing block for the DAPU did not appear to be in good condition, and there may possibly have been an earthing problem. It is considered that the anomaly detection system provided good confirming evidence for the IHUMS indications.

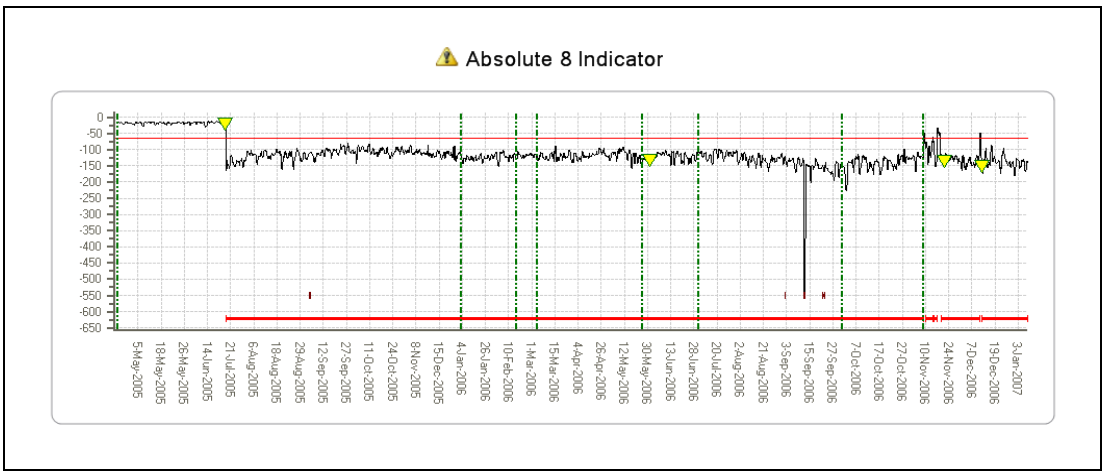

Figure 4-54 G-PUMI MGB RH high speed input shaft – Fitness Score

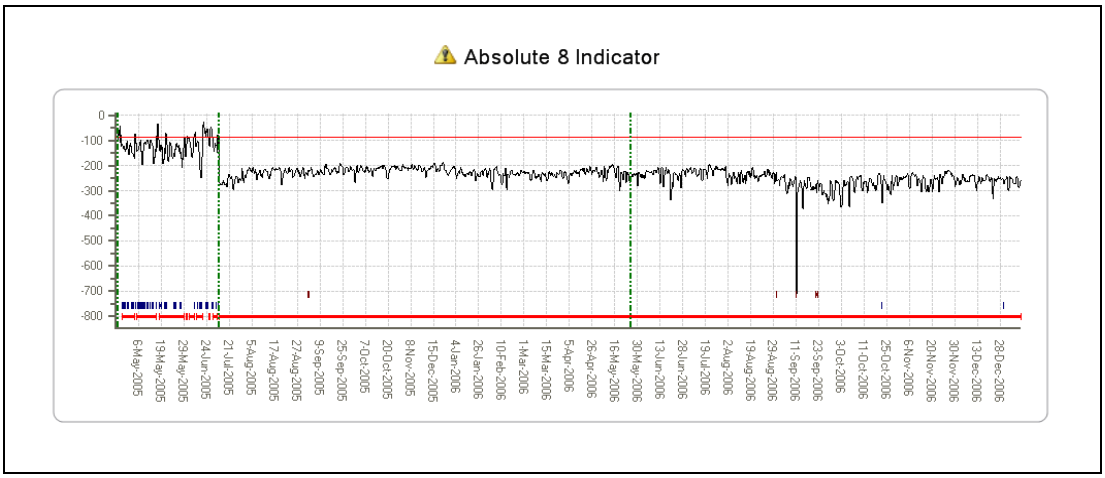

Figure 4-55 G-PUMI MGB right torque shaft - forward end – Fitness Score

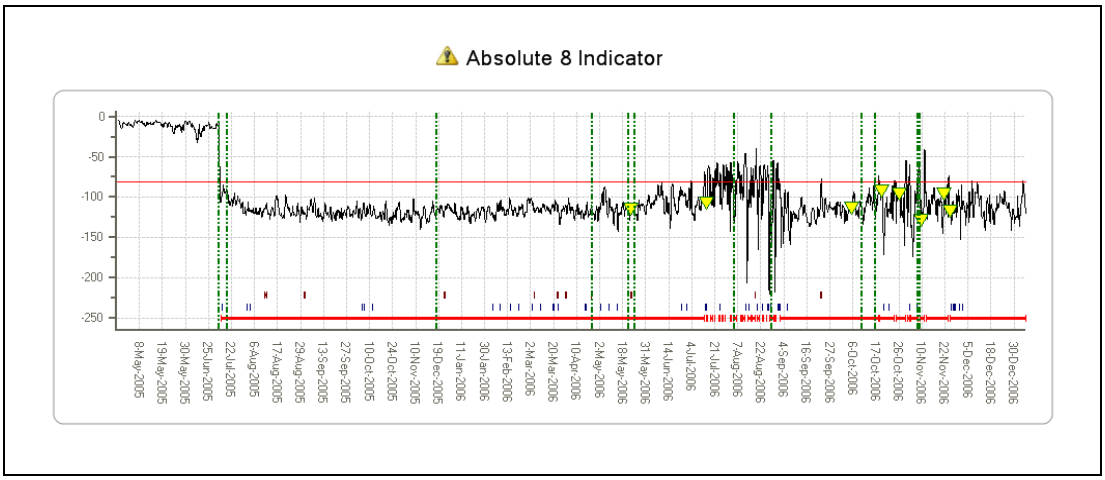

Figure 4-56 G-PUMI MGB 2nd epicyclic annulus forward (RH) – Fitness Score

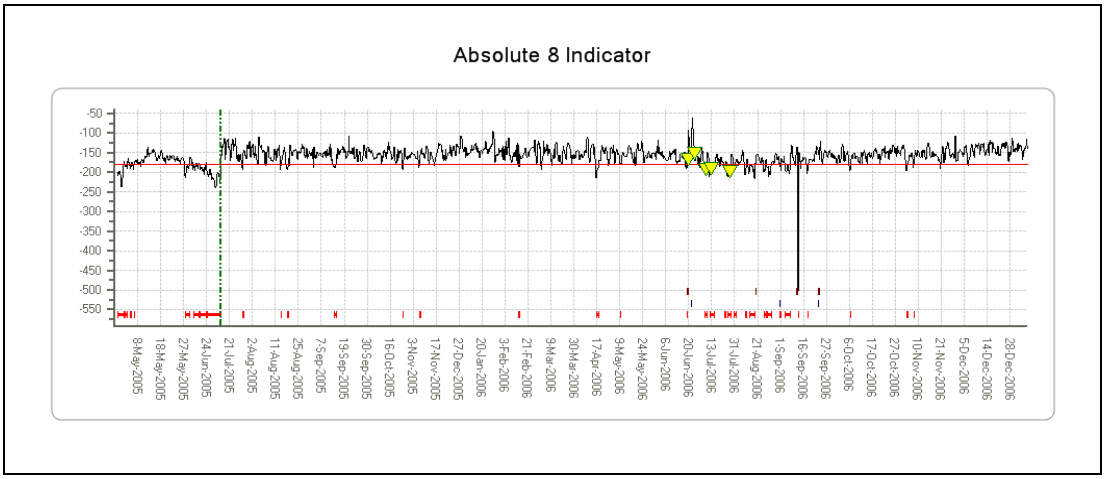

Figure 4-57 G-PUMI IGB output – Fitness Score

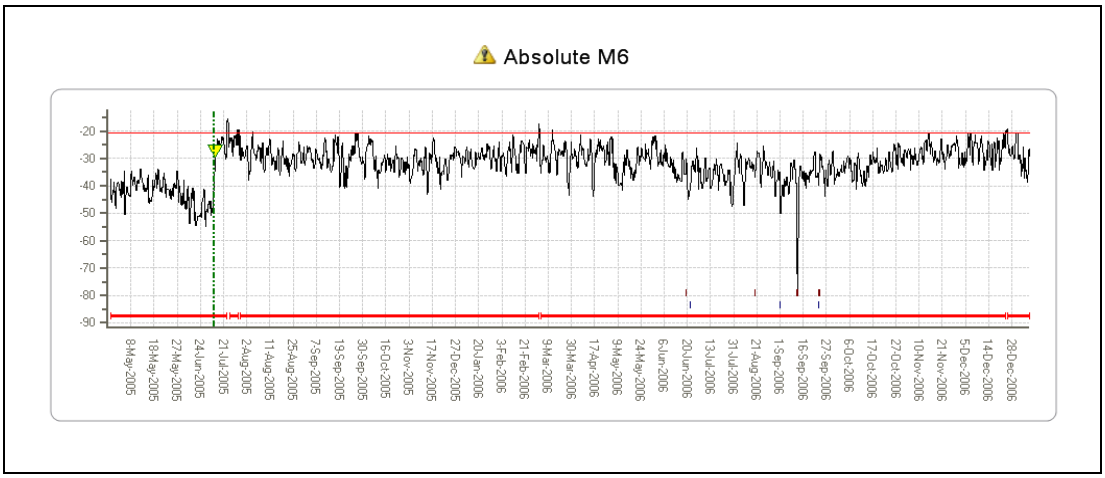

Figure 4-58 G-PUMI IGB output – Fitness Score

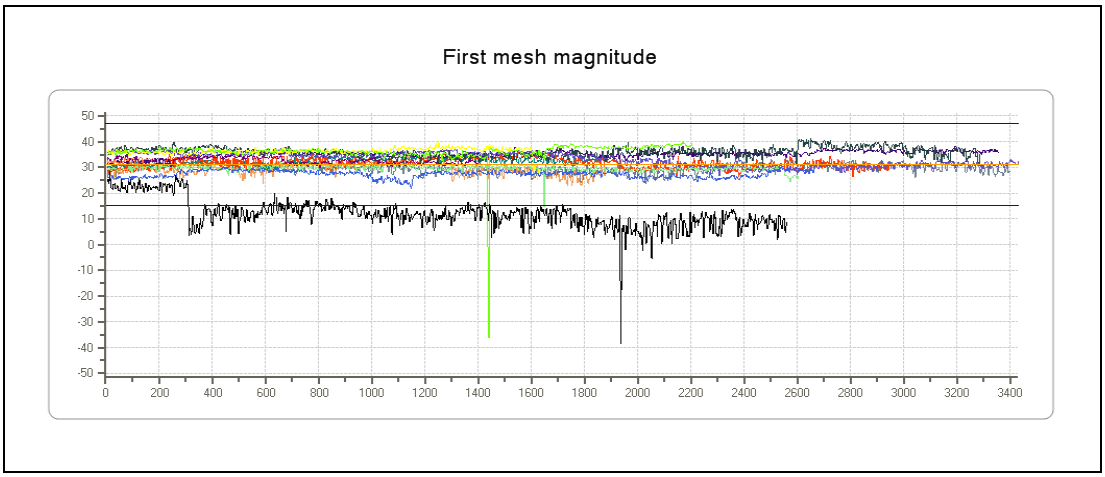

Figure 4-59 G-PUMI MGB RH high speed input shaft – Condition Indicator – fleet view

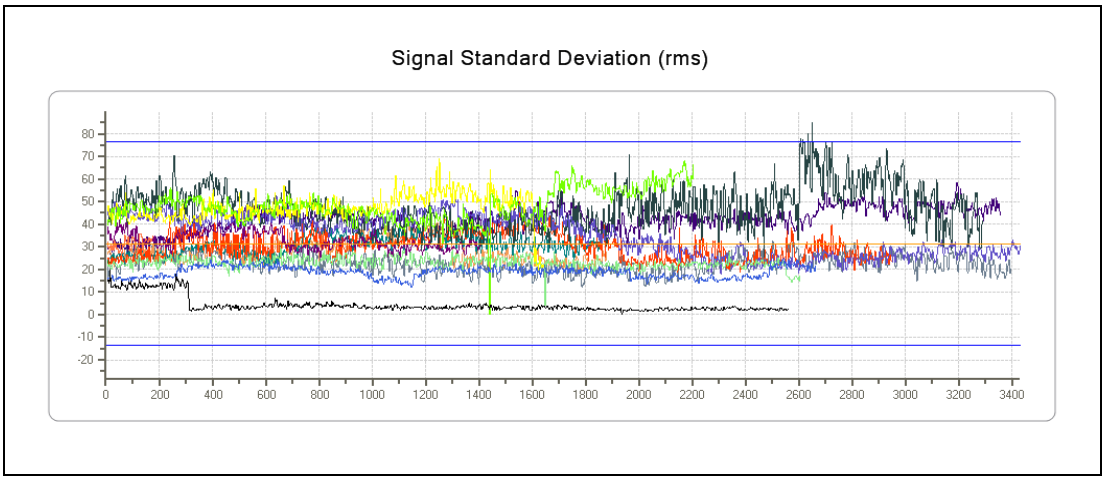

Figure 4-60 G-PUMI MGB RH high speed input shaft – Condition Indicator – fleet view

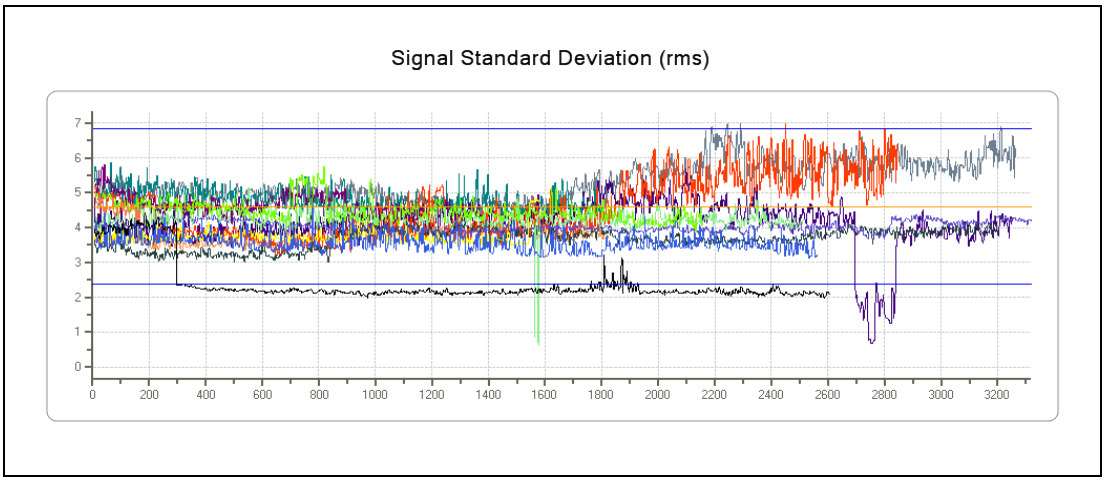

Figure 4-61 G-PUMI MGB 2nd epicyclic annulus forward (RH) – Condition Indicator – fleet view

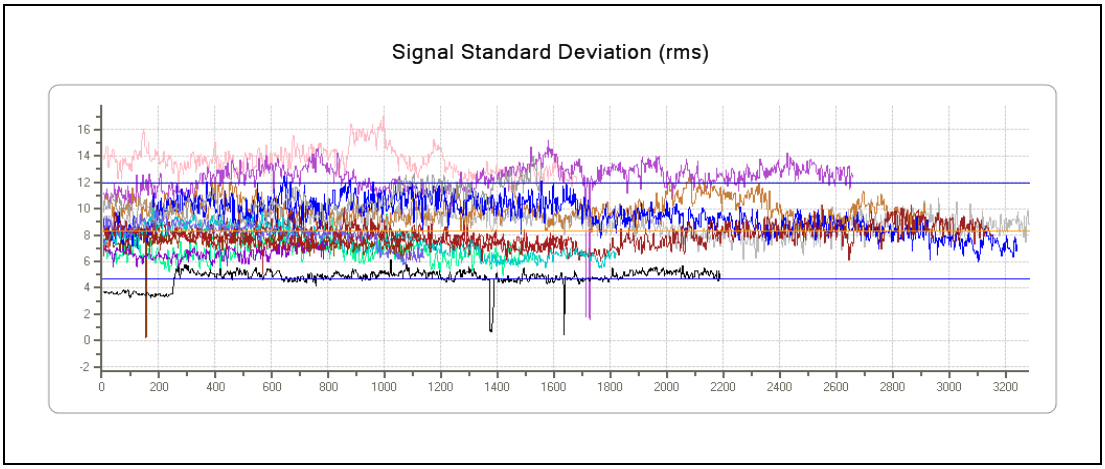

Figure 4-62 G-PUMI IGB output – Condition Indicator – fleet view

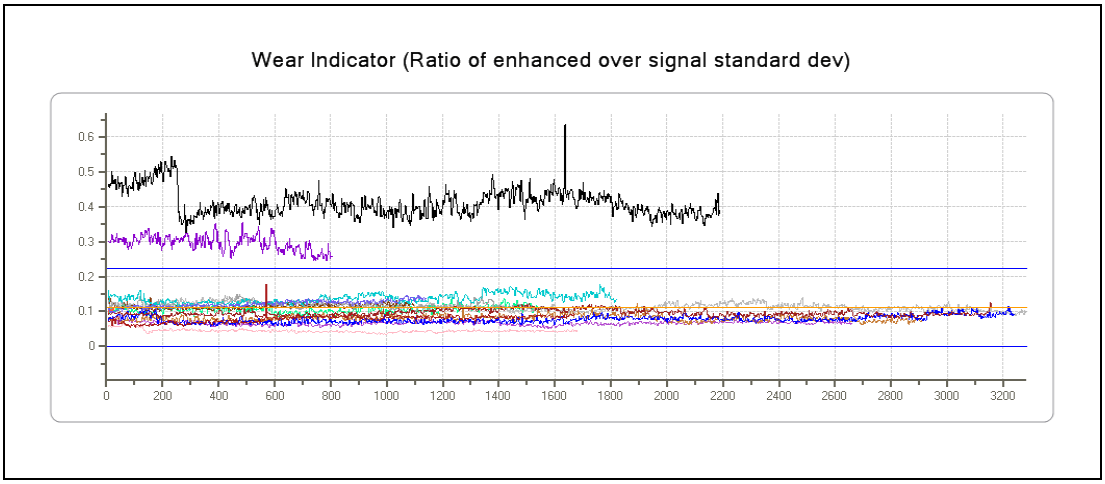

Figure 4-63 G-PUMI IGB output – Condition Indicator – fleet view

Example 4: Accelerometer 7 (Instrumentation)

The '8-indicator' absolute model for G-BLXR's MGB 1st epicyclic annulus aft (RH) had discrete periods of low and variable FSs which resulted in alerts, most notably in the period July 2006 to September 2006 (Figure 4-64). These were caused by periods of abnormally high SO1 values (Figure 4-65 - Figure 4-66).

The IHUMS generated accelerometer Nos. 5, 6 and 7 instrumentation defects on 28 May 2006. All three accelerometers were checked, and the insulation washers and sleeves were replaced. After a period of low SO1 values these suddenly increased again around 1 July 2006, which may have been caused by various maintenance actions carried out between 25 and 28 June 2006. On 19 September 2006 the IHUMS generated shaft imbalance and misalignment defects (a late indication of the increase in SO1/SO2), and also Nos. 3 and 7 accelerometer defects. No. 7 accelerometer was swapped with No. 5, and new washers were fitted. On 19 December 2006 No. 7 accelerometer was swapped with that from another aircraft.

This example clearly illustrates the sensitivity of the IHUMS SO1 and SO2 CI data to instrumentation issues, which can be difficult to rectify. Whilst the IHUMS did generate alerts, the timing of these did not correlate well with the significant increases in SO1. This illustrates the fact that the IHUMS alert generation is very dependent on the actual CI levels at the time of previous re-datums. This can result in a lack of consistency in the IHUMS alerting.

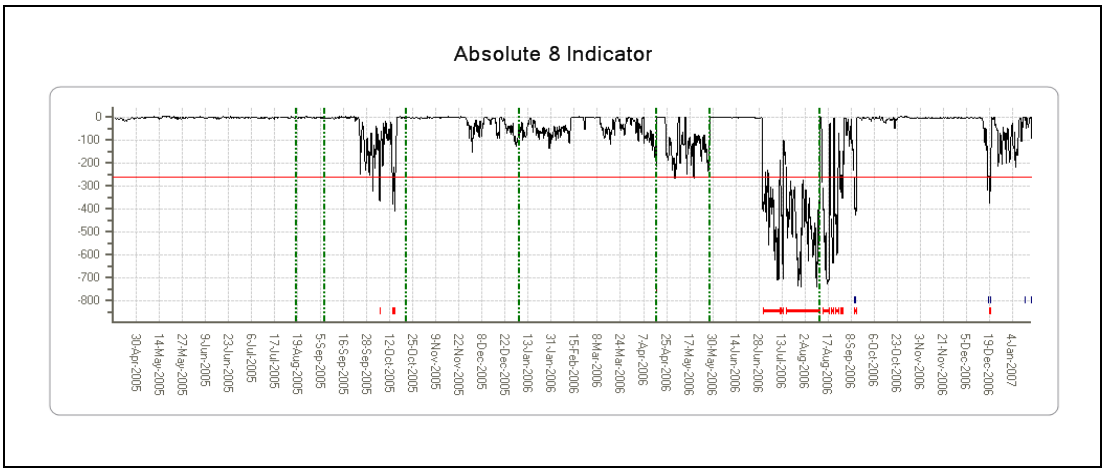

Figure 4-64 G-BLXR MGB 1st epicyclic annulus aft (RH) – Fitness Score

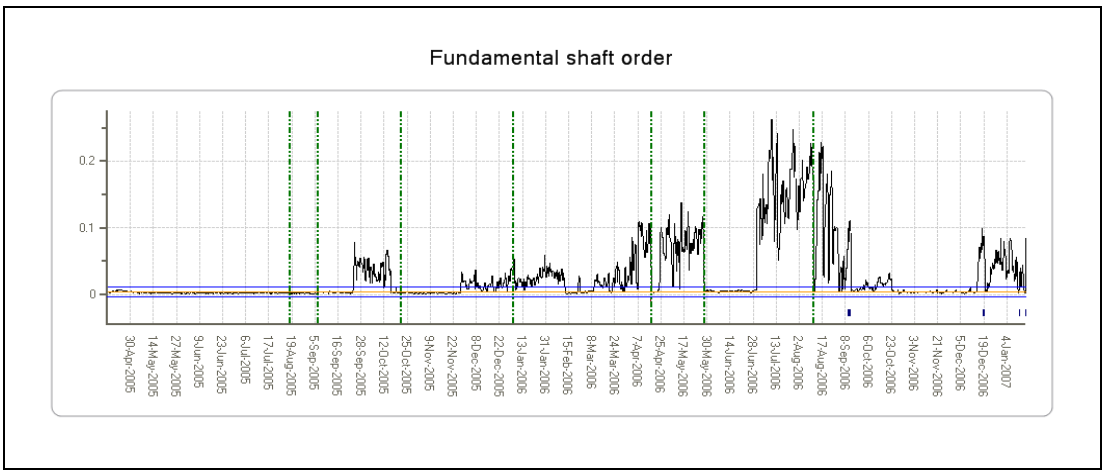

Figure 4-65 G-BLXR MGB 1st epicyclic annulus aft (RH) – Condition Indicator

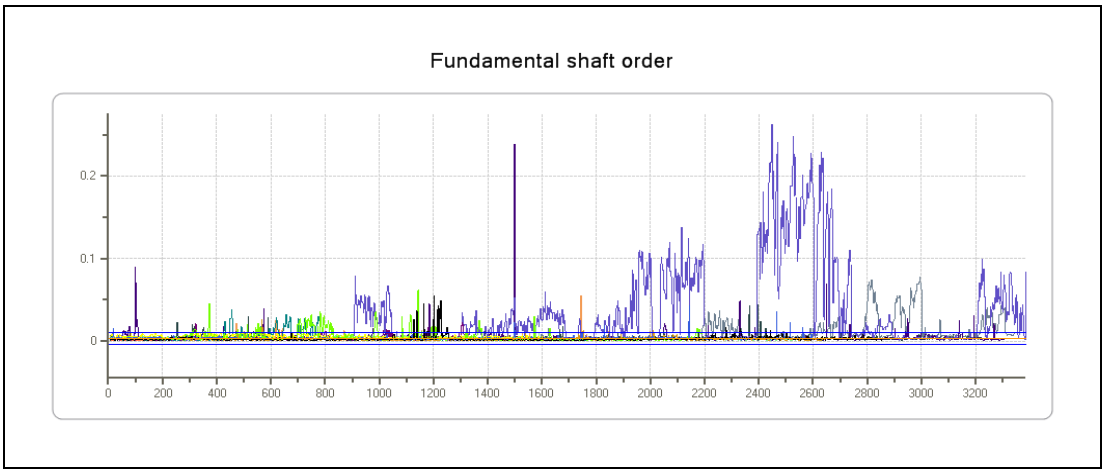

Figure 4-66 G-BLXR MGB 1st epicyclic annulus aft (RH) – Condition Indicator – fleet view

4.3 Cases Where Anomaly Detection has Failed to Identify a Fault that was Seen by the Existing HUMS

Example 1: TGB Output Shaft

At the 8 June project review meeting the Bristow HUMS engineer reported that the IHUMS had generated an alert on SO1 for the TGB output shaft on G-BWWI, and an inspection had found damage on the TGB Output bearings. No anomaly alert had been generated for this shaft. SO1 increased to a higher level on 22/23 June 2006, triggering further IHUMS alerts (Figure 4-67 and Figure 4-68 show the raw and median filtered SO1 data respectively). Following an inspection, three tail rotor flapping hinge bearings were replaced on 23 June 2006. Although the anomaly model FSs responded to the increased SO1 values, in neither case did they generate any alerts (see Figure 4-69 and Figure 4-70 for the '8-indicator' absolute and trend model FSs).

A fleet view of the SO1 indicator is presented in Figure 4-71. One aircraft has produced higher SO1 values than G-BWWI, this is G-TIGF (the data is coloured light grey). The '8-indicator' absolute and trend anomaly models did generate alerts in this case, however other IHUMS CIs were also high.

The Bristow engineer reported that TGB Output bearing damage had also occurred on other aircraft due to a change in grease, and that different aircraft were believed to be using different types of grease. An investigation suggested that the reason no anomaly had been generated may be due to the presence of noisy data in the training set that could be related to the existing problems. However, for the monitoring of key input/output shafts (e.g. MGB inputs and TGB output), consideration should be given to implementing new models using only shaft related parameters (i.e. SO1 and SO2) to increase system sensitivity to shaft faults.

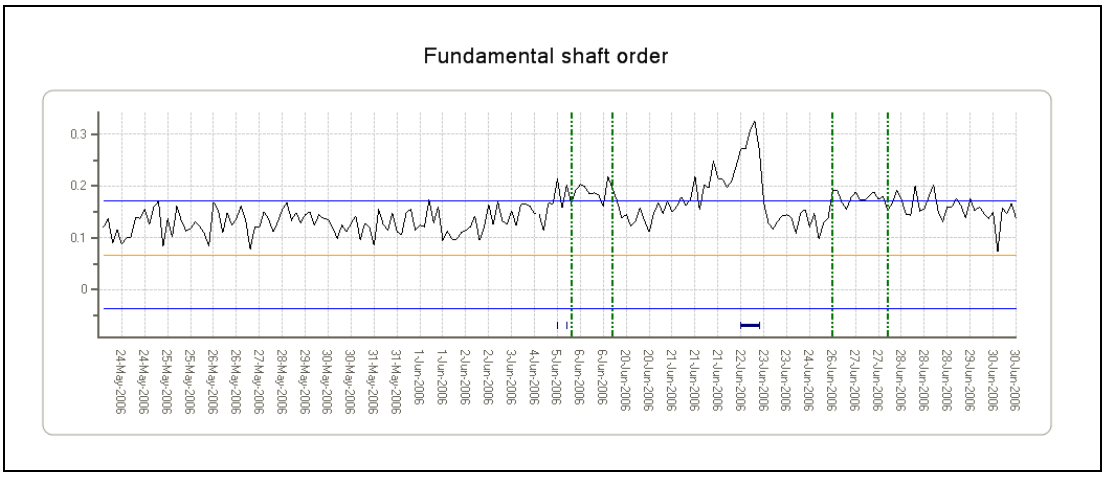

Figure 4-67 G-BWWI TGB output – Condition Indicator (raw data, 200 points)

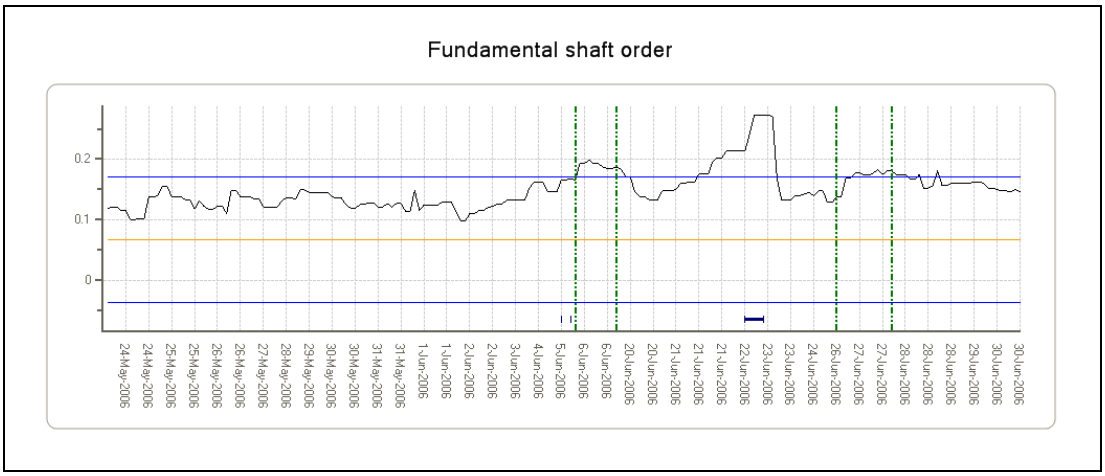

Figure 4-68 G-BWWI TGB output – Condition Indicator (median filtered data, 200 points)

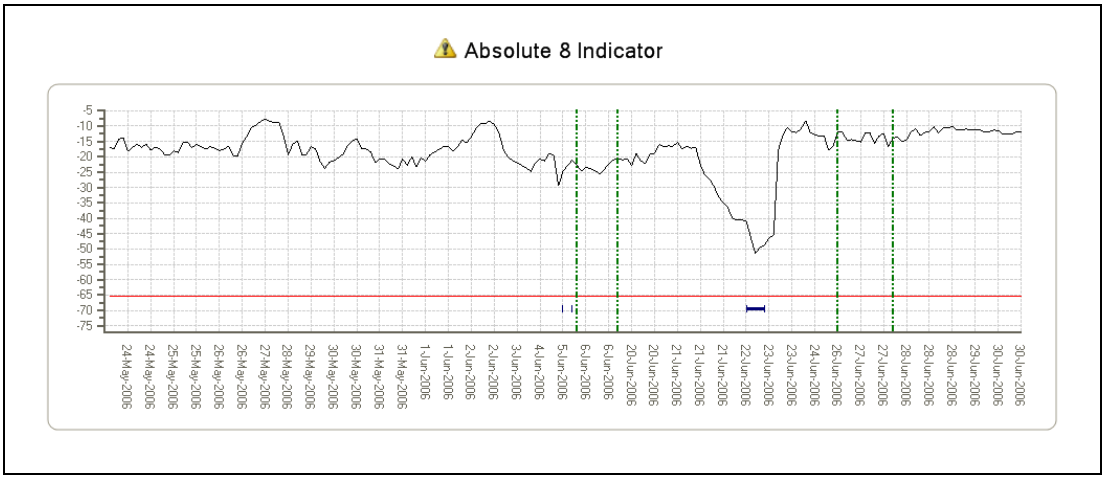

Figure 4-69 G-BWWI TGB output – Fitness Score (200 points)

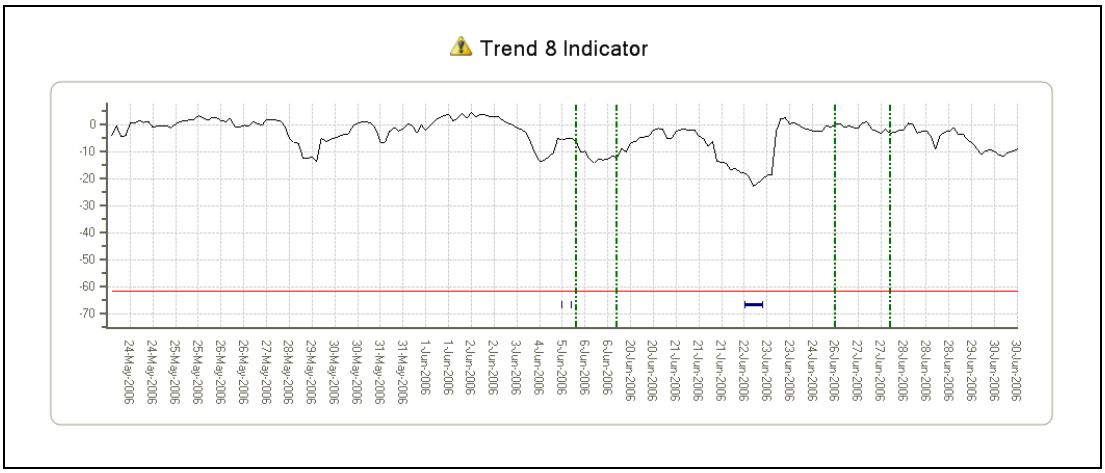

Figure 4-70 G-BWWI TGB output – Fitness Score (200 points)

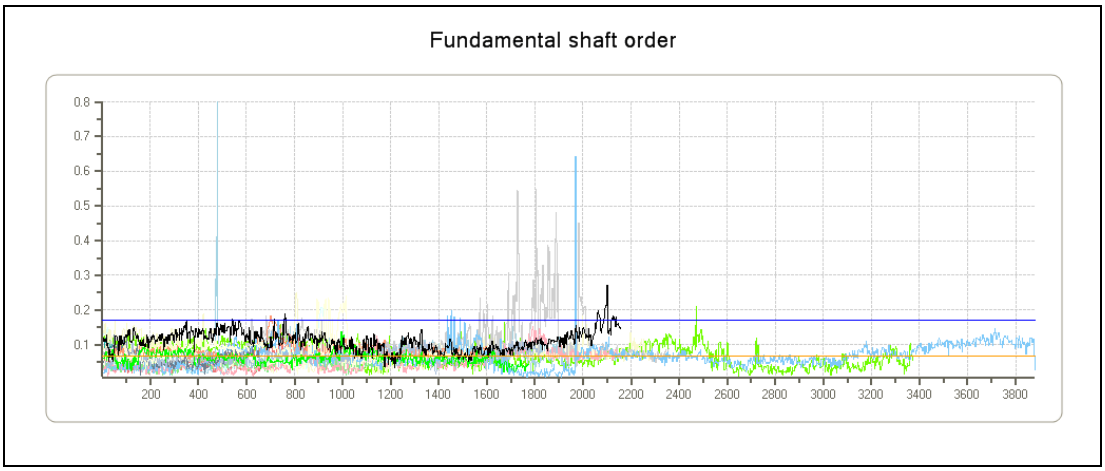

Figure 4-71 G-BWWI TGB output – Condition Indicator – fleet view

4.4 Cases Where Anomaly Detection has Identified an Existing HUMS False or Premature Alert

Example 1: RHA Module

The IHUMS was generating continuous alerts on G-BWZX's RHA Module through most of November 2006 (the dark blue line at the bottom of Figure 4-72 and Figure 4-73), and the module was placed on close monitoring. These alerts were triggered by rising trends in the Enhanced Peak-Peak and Enhanced Standard Deviation CIs. However, neither the '8-indicator' absolute or trend models (which have these indicators as inputs) generated any alerts during this period (Figure 4-74 and Figure 4-75). The reason for this is apparent when viewing fleet displays of the indicators (Figure 4-76 and Figure 4-77). Trends in these indicators are common, and a number of other aircraft have higher indicator values.

The indictors started trending down again in late November 2006, and the RHA Module was removed from close monitoring. It is considered that the '8-indicator' models were correct to suppress the indicator trends and not generate alerts.

Although the 'M6' absolute model did not generate any alerts (Figure 4-78), the trend model did produce some alerts in November 2006 (Figure 4-79). Whilst there are trends in the Enhanced Impulsiveness and Wear indicator values, these stayed well within the fleet norm (Figure 4-80 and Figure 4-81). Therefore, like the IHUMS, the 'M6' trend model can be considered to have generated premature alerts, and may be too sensitive on this component.

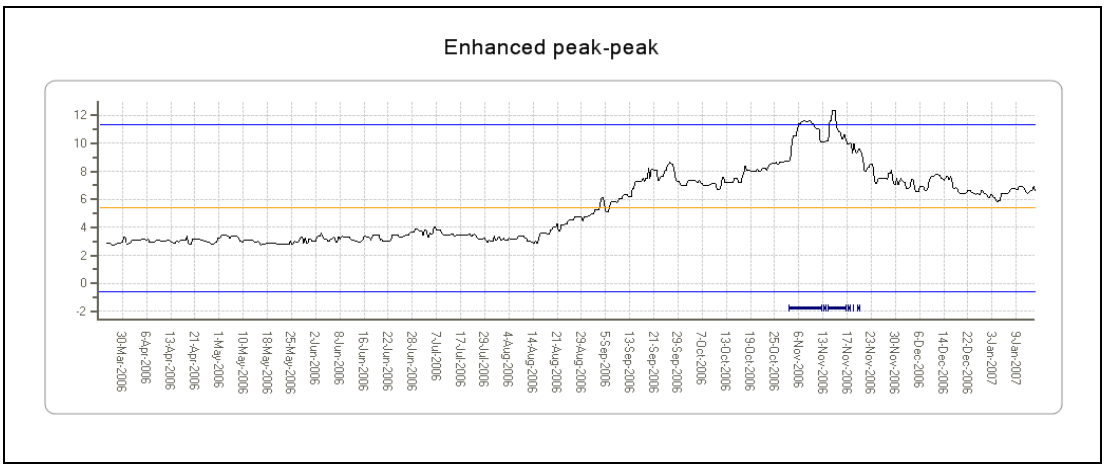

Figure 4-72 G-BWZX RHA right hydraulic idler – Condition Indicator

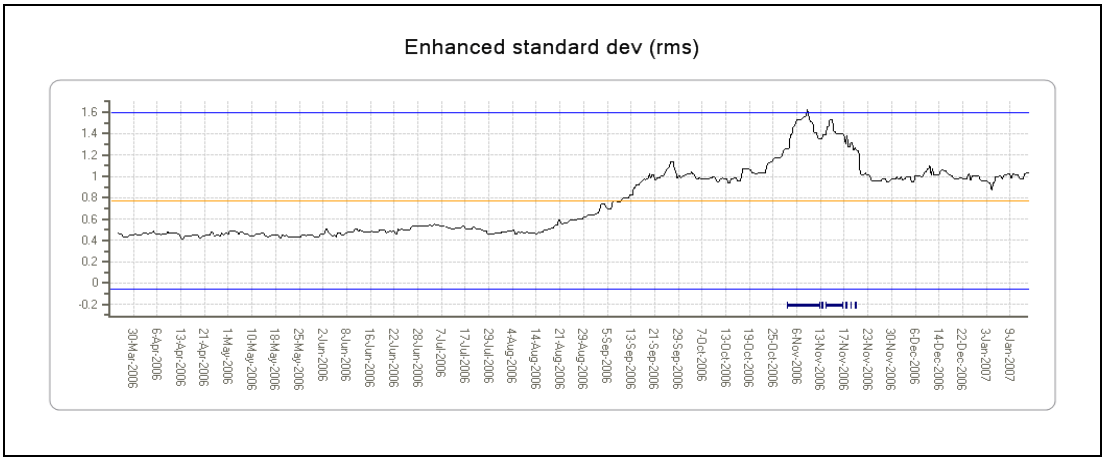

Figure 4-73 G-BWZX RHA right hydraulic idler – Condition Indicator

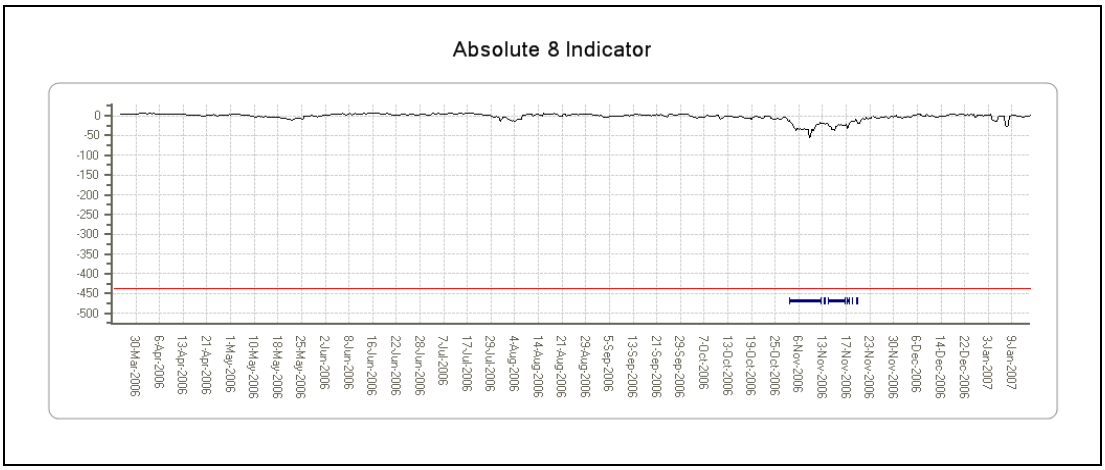

Figure 4-74 G-BWZX RHA right hydraulic idler – Fitness Score

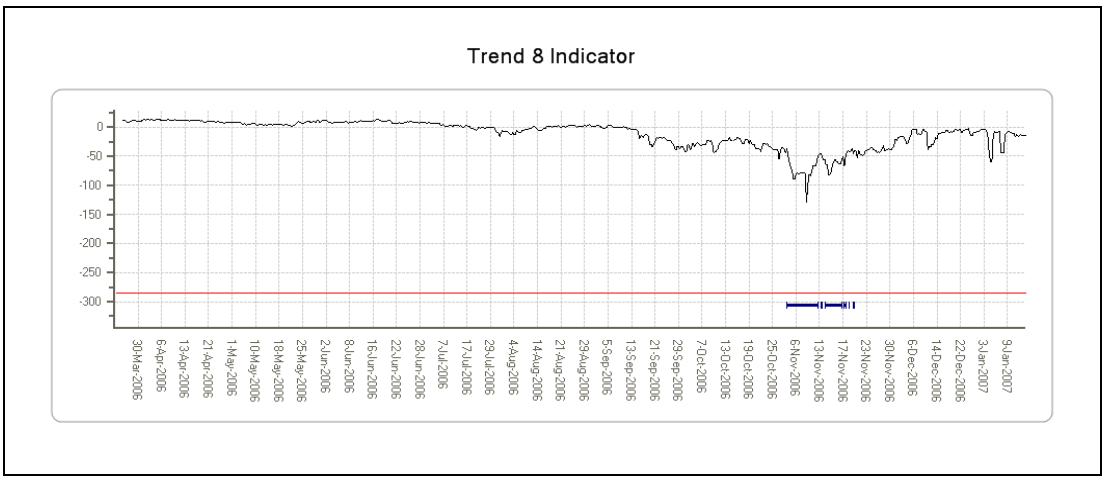

Figure 4-75 G-BWZX RHA right hydraulic idler – Fitness Score

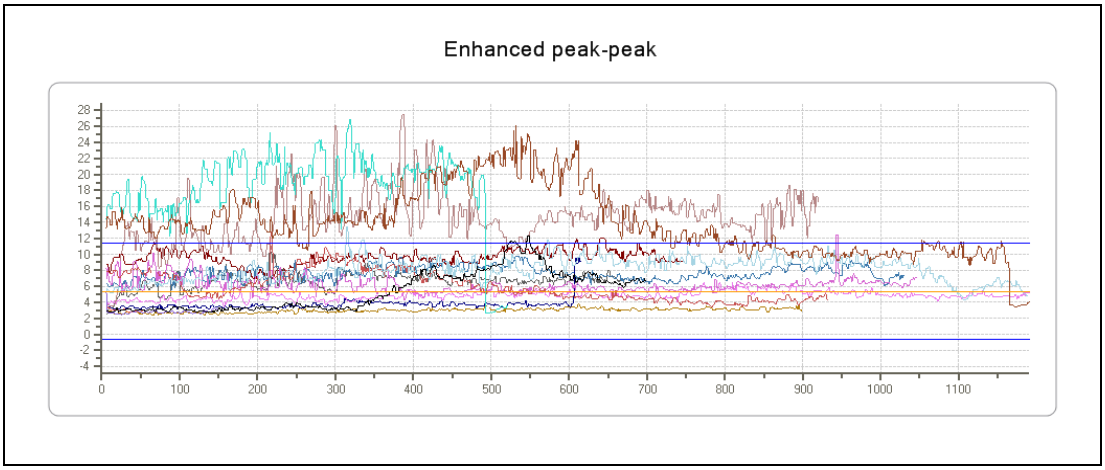

Figure 4-76 G-BWZX RHA right hydraulic idler – Condition Indicator – fleet view

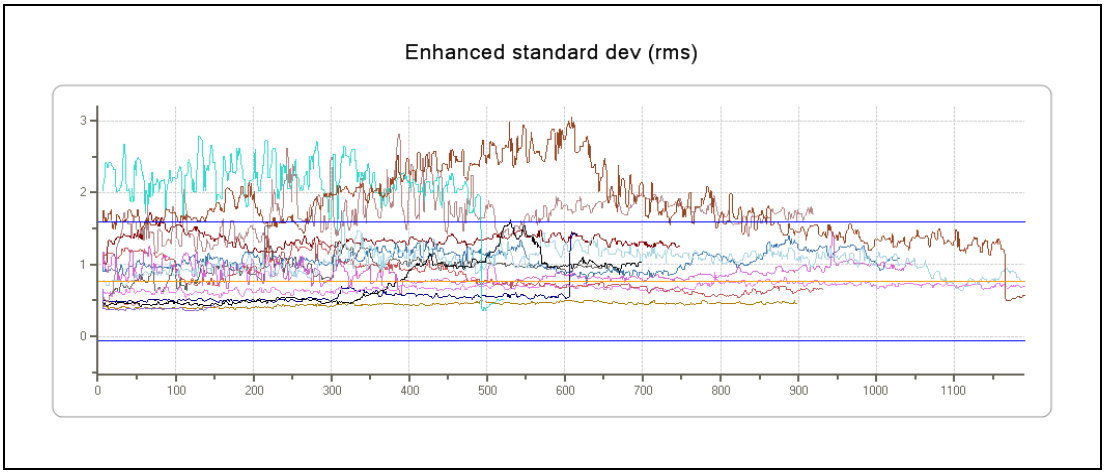

Figure 4-77 G-BWZX RHA right hydraulic idler – Condition Indicator – fleet view

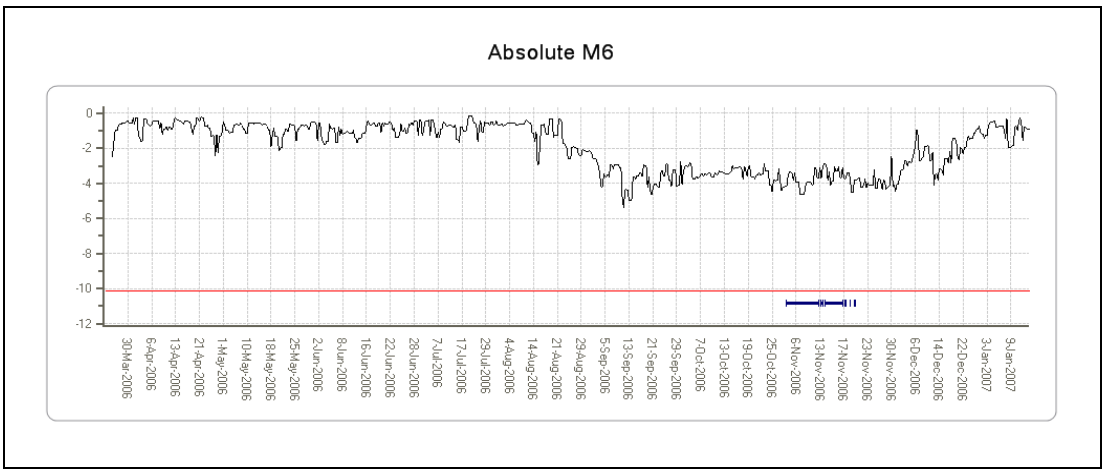

Figure 4-78 G-BWZX RHA right hydraulic idler – Fitness Score

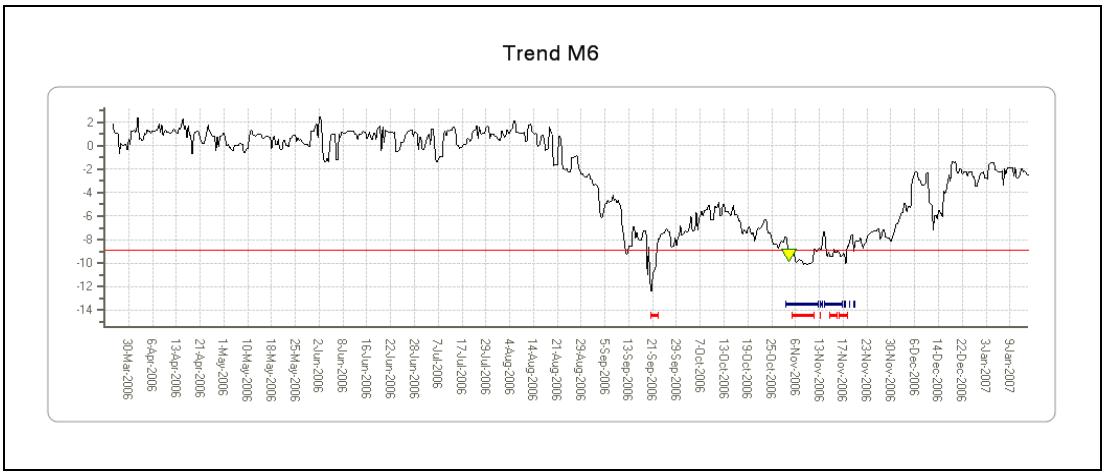

Figure 4-79 G-BWZX RHA right hydraulic idler – Fitness Score

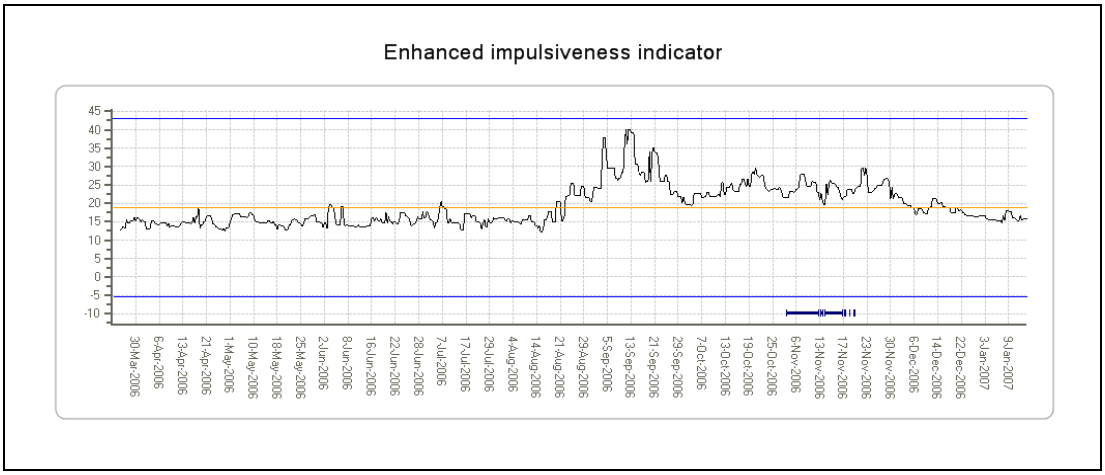

Figure 4-80 G-BWZX RHA right hydraulic idler – Fitness Score

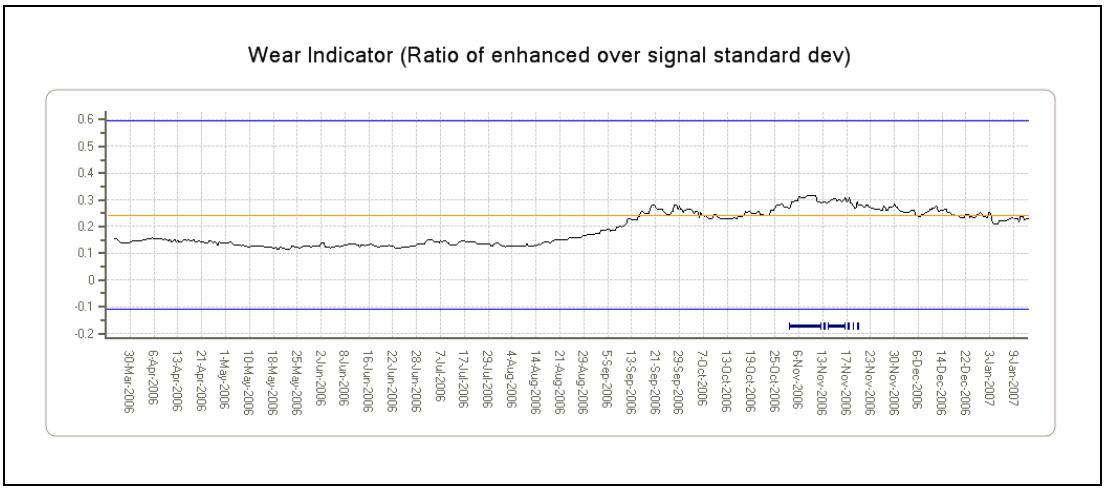

Figure 4-81 G-BWZX RHA right hydraulic idler – Condition Indicator

4.5 Cases Where Anomaly Detection has Generated a False or Premature Alert

Example 1: MGB – Left Torque Shaft – Aft End

The '8-indicator' absolute model for G-TIGT's left torque shaft - aft end produced low and variable FSs up to November 2005 (Figure 4-82). These were caused by high and variable SO1 values, which returned to normal levels at that time (Figure 4-83). However, the FSs subsequently remained at or just below the alert threshold, developing a slowly decreasing trend and continuous alert from August 2006. These were due to the Signal Peak-Peak and Standard Deviation indicators being close to the upper limit of fleet norms, and also having slowly rising trends (Figure 4-84 - Figure 4-85).

The CI levels were not clearly abnormal compared to the rest of the fleet (Figure 4-86 - Figure 4-87), and the data showed only very slowly rising trends. Bristow's HUMS engineer therefore concluded that the anomaly models are too sensitive on the left torque shaft - aft end.

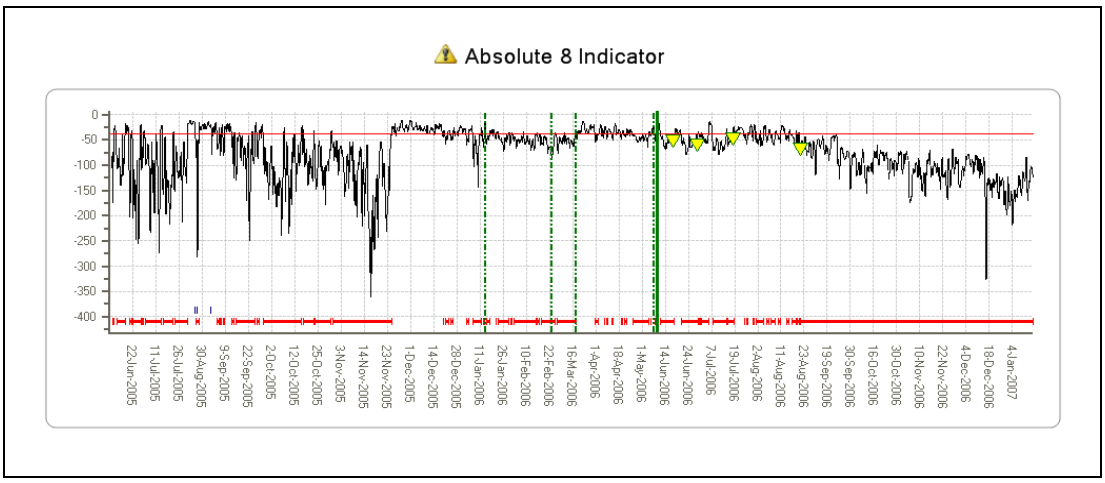

Figure 4-82 G-TIGT - MGB - left torque shaft - aft end – Fitness Score

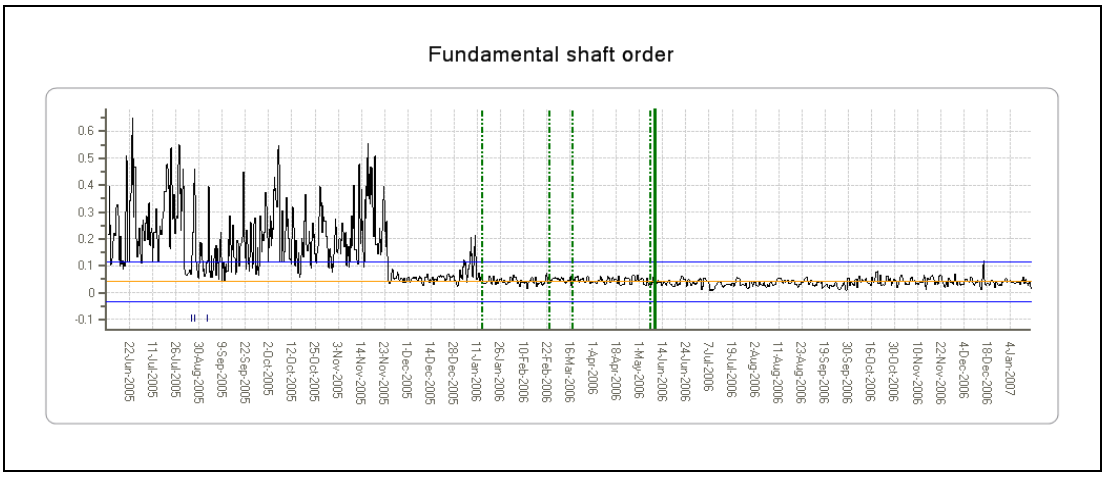

Figure 4-83 G-TIGT - MGB - left torque shaft - aft end – Condition Indicator

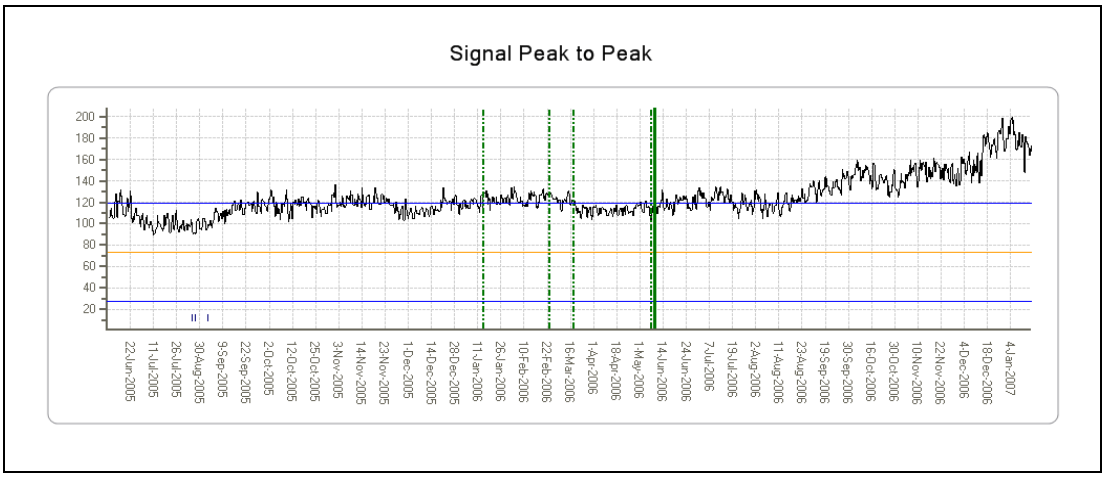

Figure 4-84 G-TIGT - MGB - left torque shaft - aft end – Condition Indicator

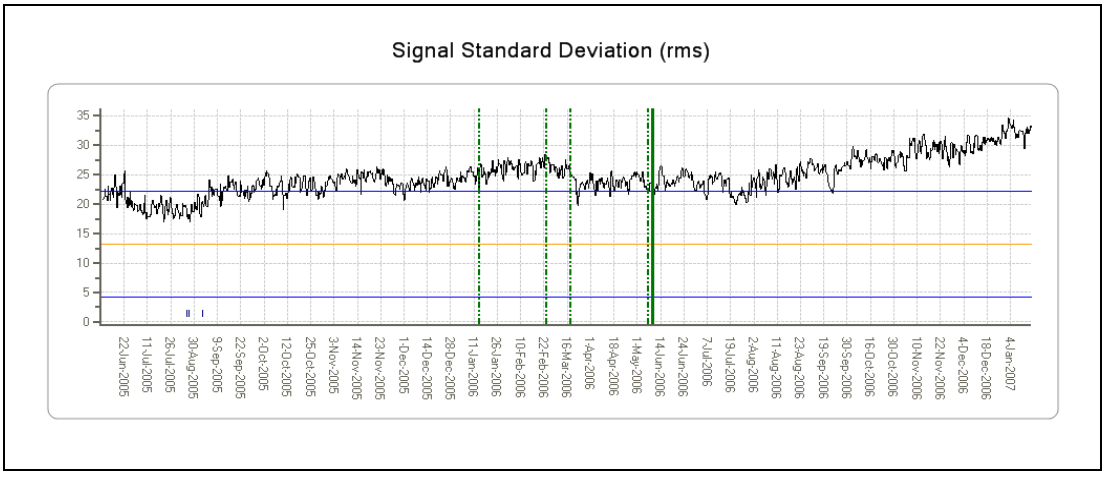

Figure 4-85 G-TIGT - MGB - left torque shaft - aft end – Condition Indicator

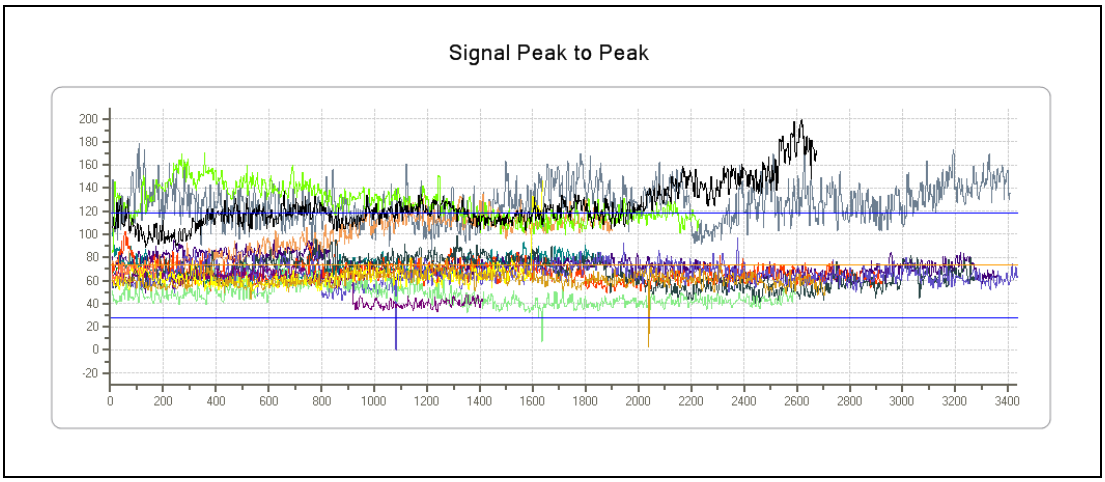

Figure 4-86 G-TIGT - MGB - left torque shaft - aft end – Condition Indicator – fleet view

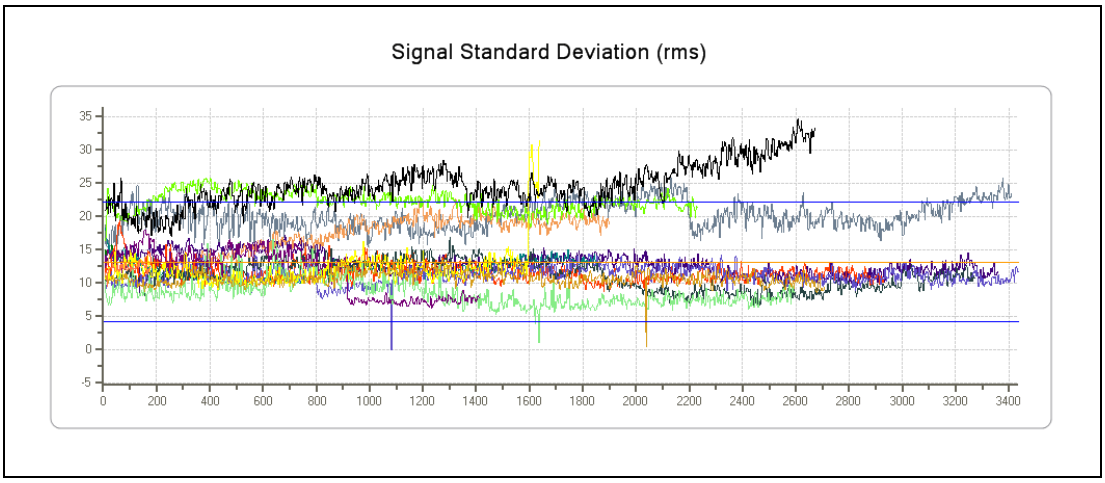

Figure 4-87 G-TIGT - MGB - left torque shaft - aft end – Condition Indicator – fleet view

Example 2: RHA Module (Example 1 in Section 4.4)

Referring to example 1 in Section 4.4, whilst the '8-indicator' models were considered to have correctly suppressed indicator trends triggering IHUMS alerts, the 'M6' trend model did produce some alerts that were considered to be premature. Therefore the example is also referenced in this section.

4.6 Cases With a Currently Unknown Outcome

Example 1: MGB Left Torque Shaft – Fwd End

The '8-indicator' absolute model for G-BMCW's MGB left torque shaft - fwd end had a clear decreasing trend for a period of time, generating a continuous alert from July 2006 (Figure 4-88). The trend model was in alert from November 2006 (Figure 4-89). However, no IHUMS alerts were generated during this period. Figure 4-90 - Figure 4-92 show the rising trends in the Enhanced Peak-Peak, Enhanced Standard Deviation, and Signal Standard Deviation indicators. The final indicator levels were significantly higher than on any of the other aircraft in the fleet. The '8-indicator' absolute model for the mating LH high speed input shaft had also been in alert from November 2006 (Figure 4-93) due to high values of the Signal Standard Deviation indicator (Figure 4-94).

The only identified relevant maintenance action was removal of the No. 1 and 2 engines for inspection on 3 December 2006. This was followed by an IHUMS re-datum on the LH high speed input shaft, however it had little effect on the data (it is noted that there have been ten IHUMS re-datums on this shaft during the period shown). Although at the time of writing the findings for this case are inconclusive, owing to the high levels of some CIs, the anomaly detection system is considered to be correct in drawing the operator's attention to this feature in the data.

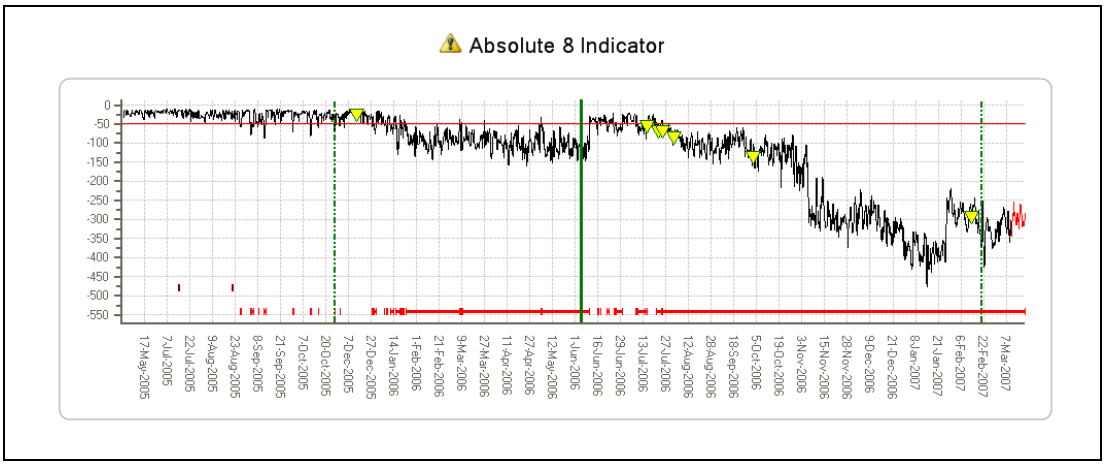

Figure 4-88 G-BMCW MGB left torque shaft - forward end – Fitness Score

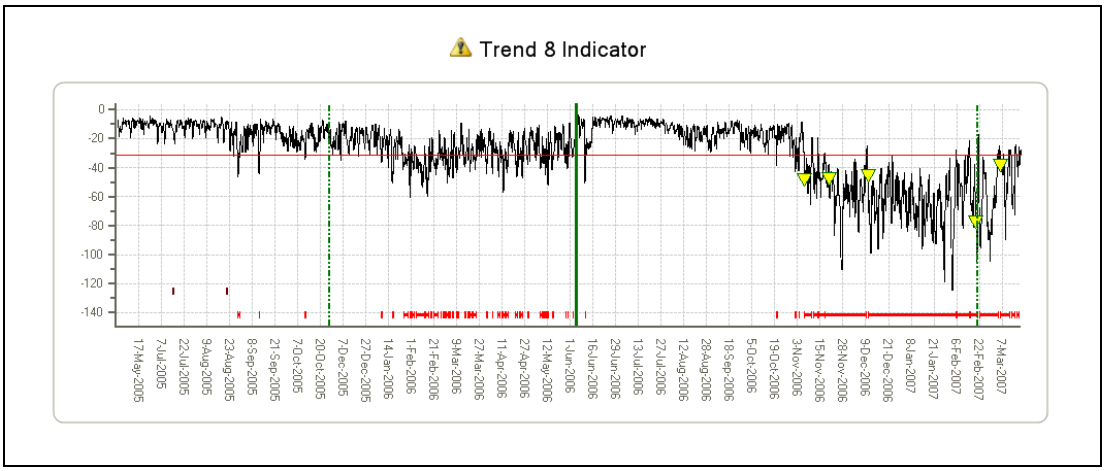

Figure 4-89 G-BMCW MGB left torque shaft - forward end – Fitness Score

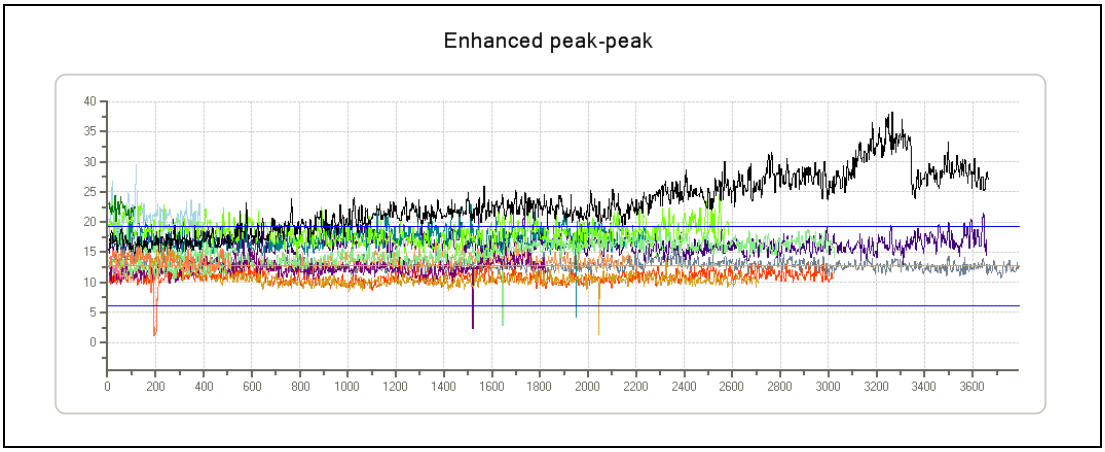

Figure 4-90 G-BMCW MGB left torque shaft - forward end – Condition Indicator – fleet view

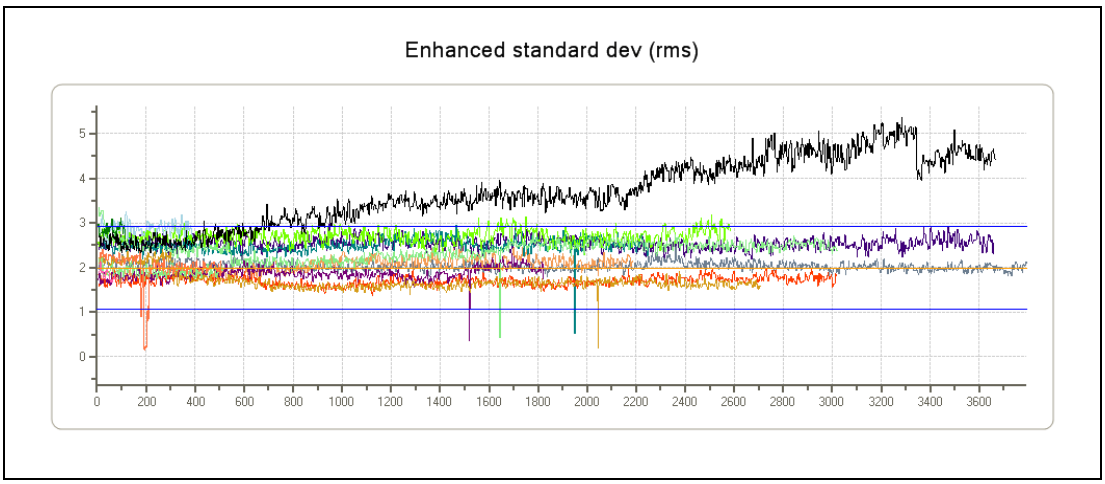

Figure 4-91 G-BMCW MGB left torque shaft - forward end – Condition Indicator – fleet view

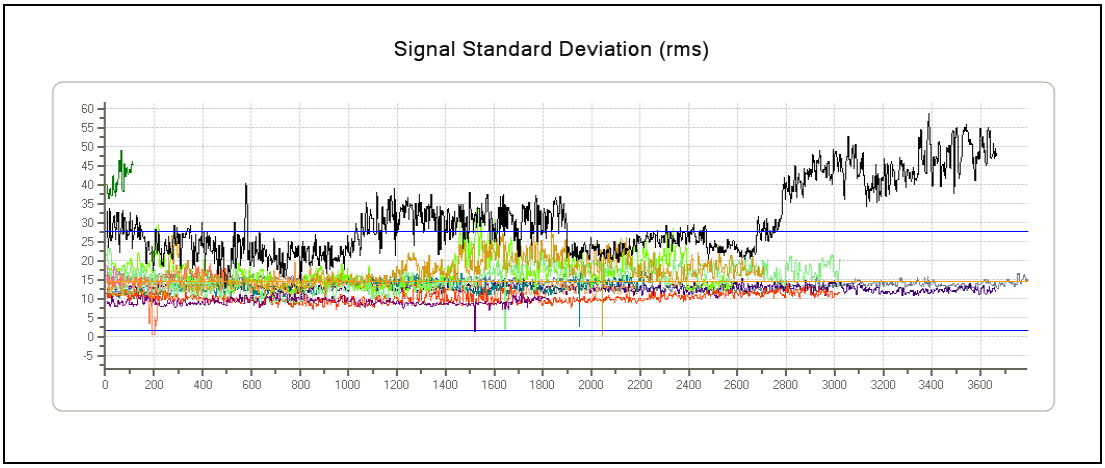

Figure 4-92 G-BMCW MGB left torque shaft - forward end – Condition Indicator – fleet view

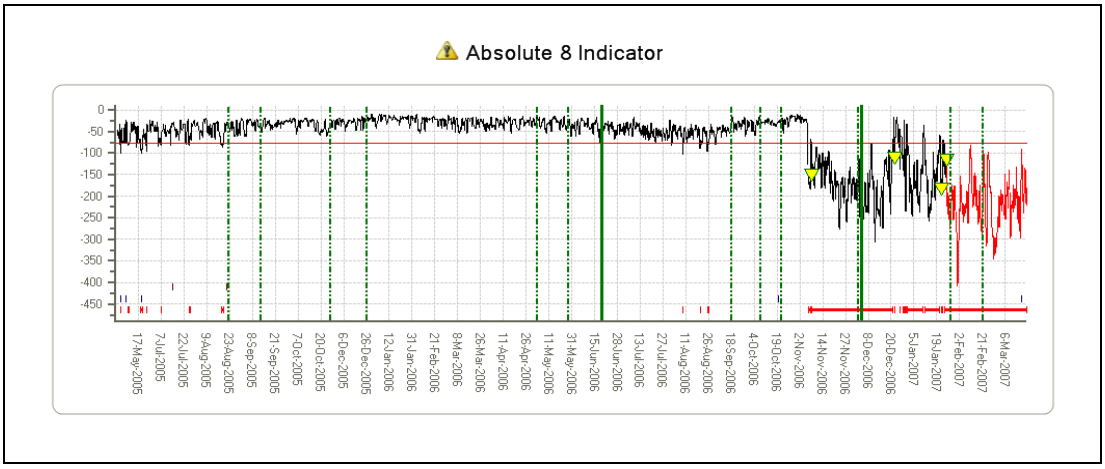

Figure 4-93 G-BMCW MGB LH high speed input shaft – Fitness Score

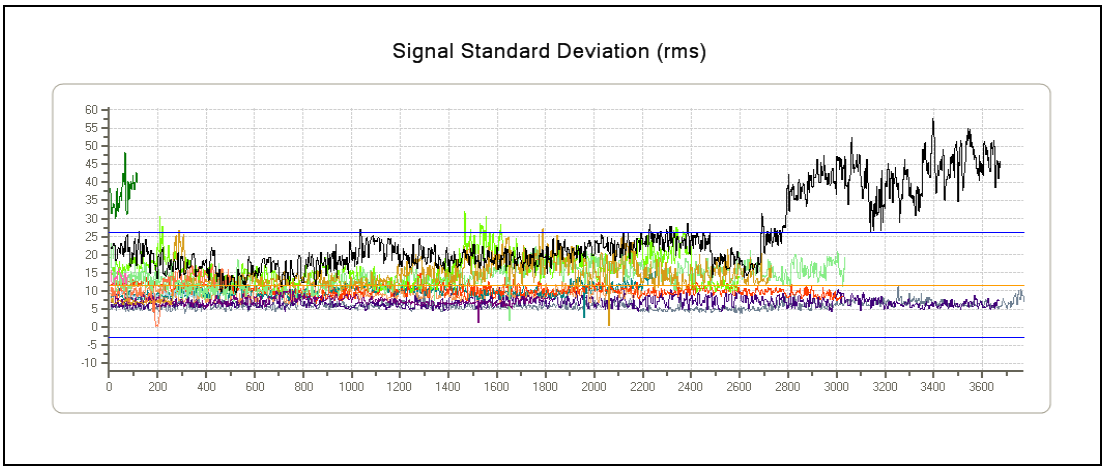

Figure 4-94 G-BMCW MGB LH high speed input shaft – Condition Indicator – fleet view

Example 2: LHA - Left Hydraulic Drive 47-Tooth Gear

The '8-indicator' trend model for G-TIGT's LHA left hydraulic drive 47-tooth gear generated alerts in the period August 2006 to November 2006 (Figure 4-95 - Figure 4-96). These were caused by clear rising trends in multiple indicators, including Enhanced Standard Deviation and Signal Peak-Peak (Figure 4-97 - Figure 4-101). There were no IHUMS alerts, but the trends were confirmed in the IHUMS.

The No. 3 accelerometer was replaced on 27 September 2006, however this had no effect on the data. The LH hydraulic pump was refitted and replaced on 6 October 2006, at which time the pump quill drive was found to be worn. The rising CI trends stopped, but indicator levels remained high. It is therefore possible that there may have been other wear within the module. Although not conclusively resolved, this was considered to be a valid anomaly detection system indication.

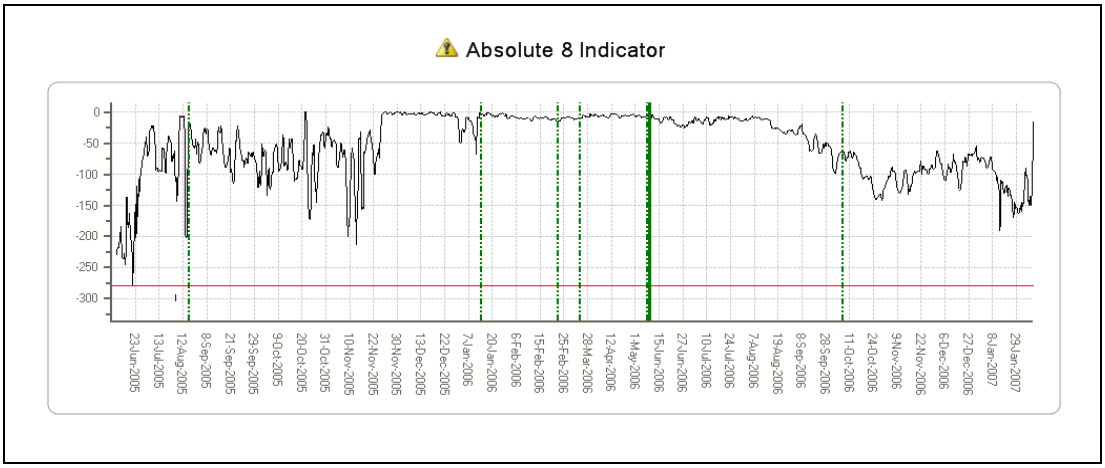

Figure 4-95 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Fitness Score

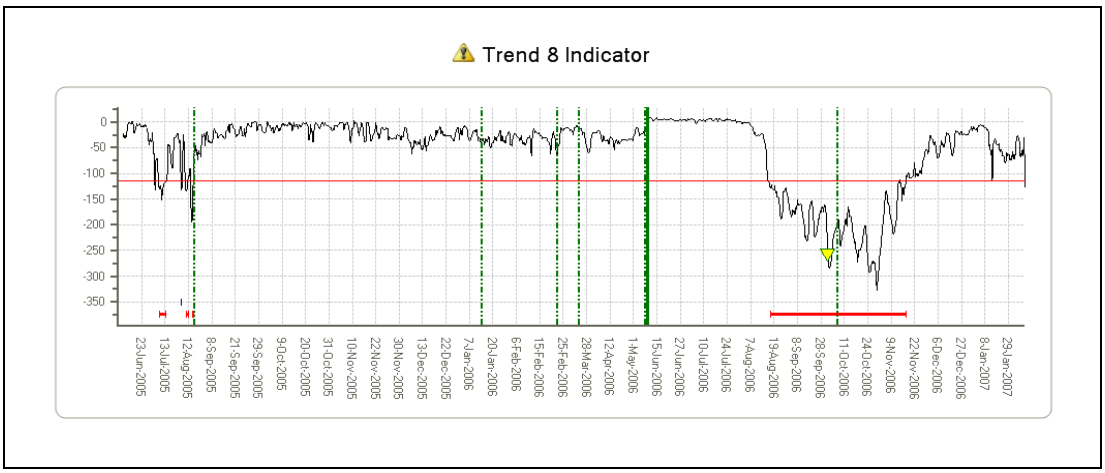

Figure 4-96 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Fitness Score

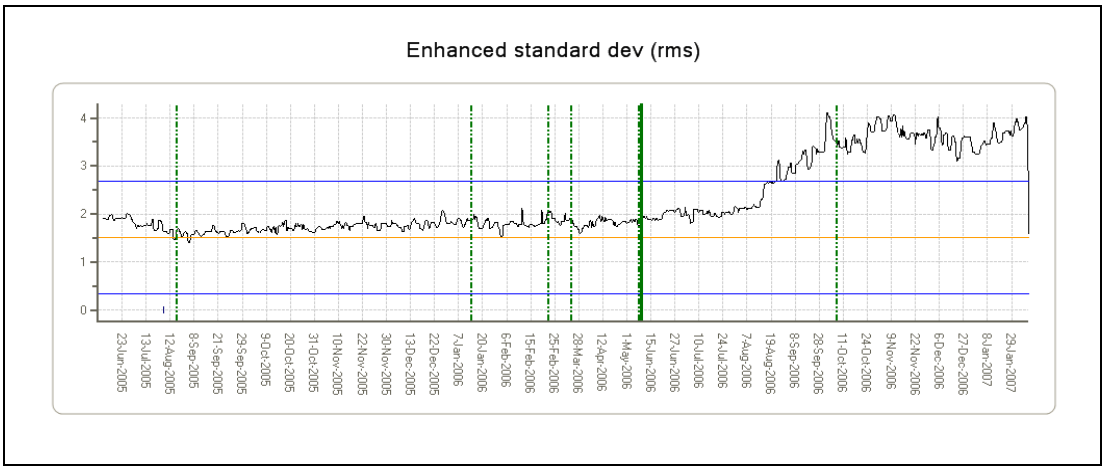

Figure 4-97 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Condition Indicator

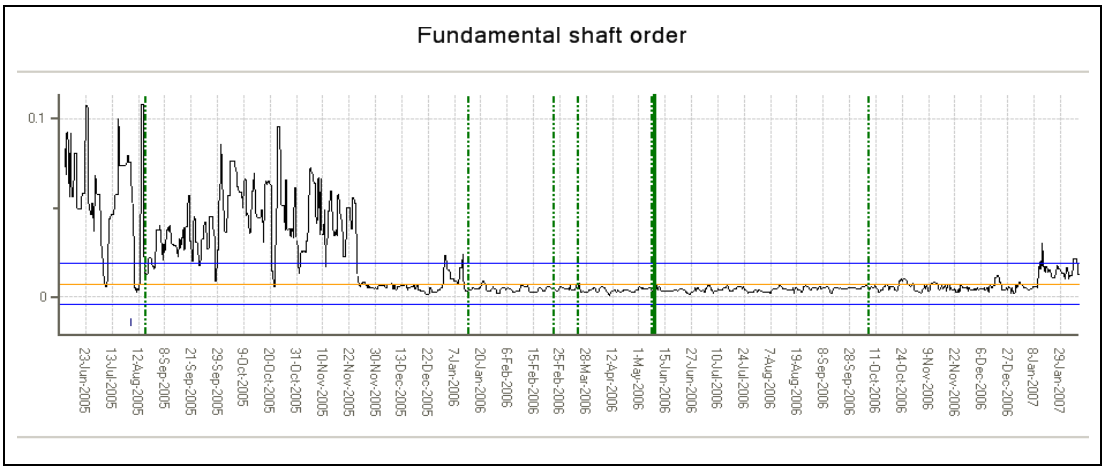

Figure 4-98 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Condition Indicator

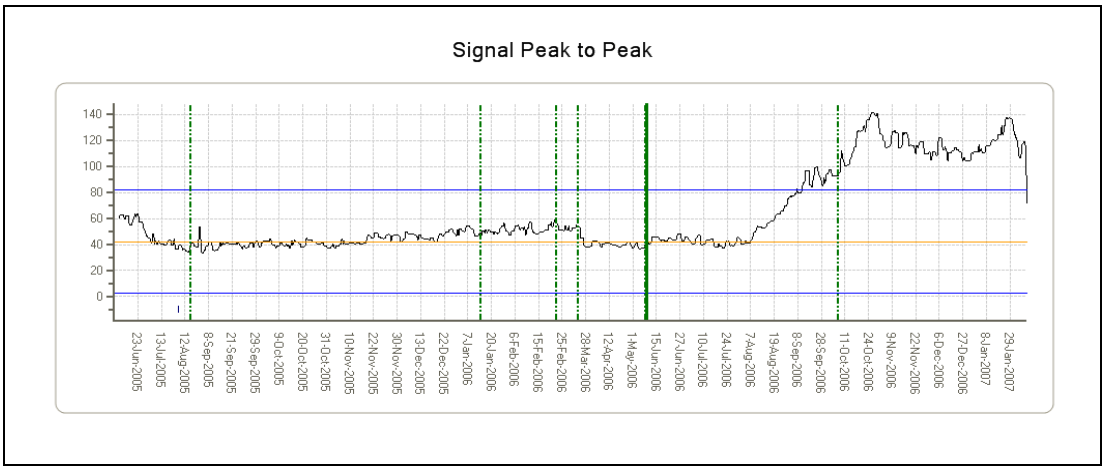

Figure 4-99 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Condition Indicator

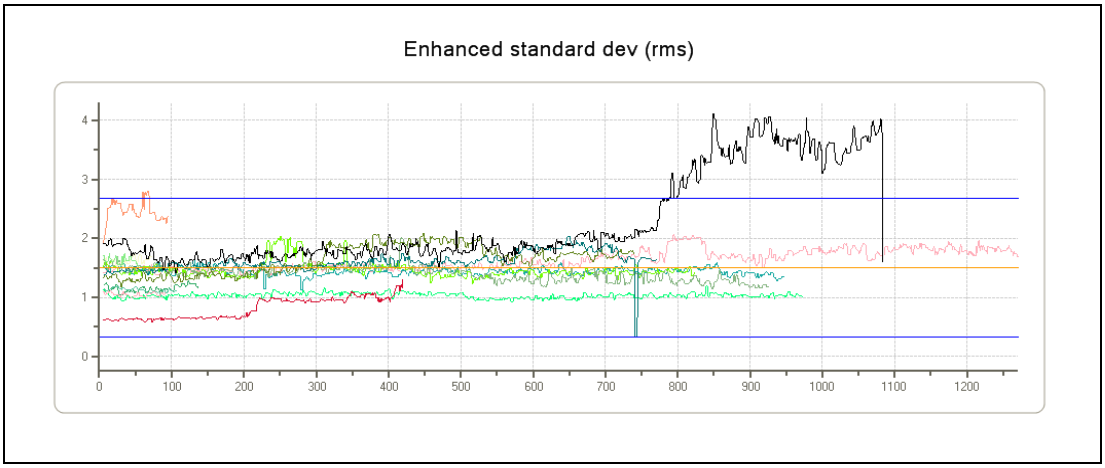

Figure 4-100 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Condition Indicator – fleet view

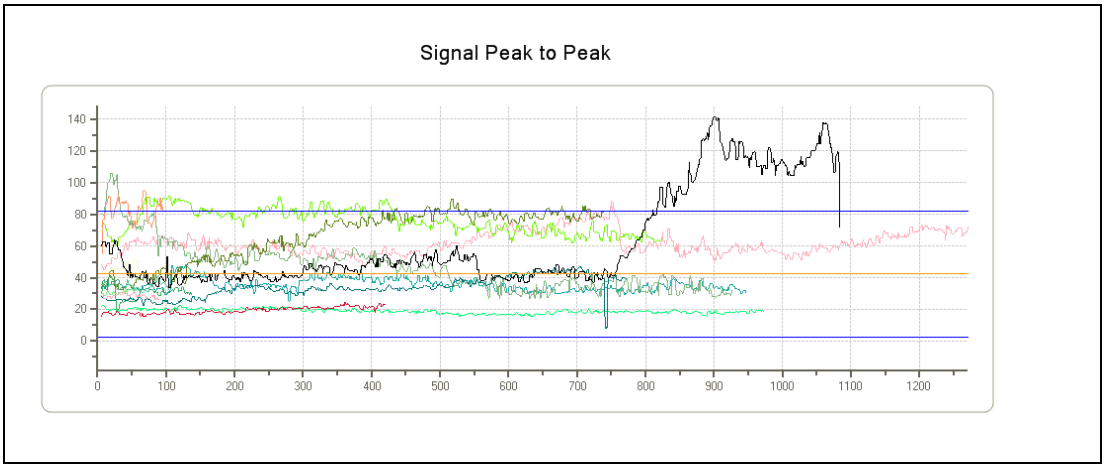

Figure 4-101 G-TIGT - LHA - left hydraulic drive 47-tooth gear – Condition Indicator – fleet view

5 Statistical Analysis of System Performance

All the statistics presented in this section apply to a six-month period of operations from 1 June to 30 November 2006.

5.1 Anomaly Detection System and IHUMS Alerts by Sensor

Table 5-1 and Table 5-2 show the number of anomaly detection system and IHUMS alerts respectively, broken down by IHUMS sensor.

The IHUMS performs a total of 32 drive train component analyses. There are three drive train components that have two gears on the monitored shaft. For each of these, separate anomaly models have been created for the two gears, using different gear mesh related CIs in each of the models. The anomaly detection system therefore performs 35 drive train component analyses. The IHUMS alert figures combine both "sensor" and "defect" alerts, as the anomaly detection system alerts currently do not differentiate between these.

Comparing the number of alerts, during the six-month trial the anomaly detection system triggered a total of 1,484 alerts for the 35 analyses, which equates to an average rate of 42 per analysis. The IHUMS triggered a total of 1,381 alerts for 32 analyses, which equates to an average rate of 43 per analysis. Therefore, using the initial set of anomaly models and thresholds, the anomaly detection system's alert rate was similar to that of the IHUMS. Considering that the comparison is between a system undergoing its first operational trial, without any opportunity for refinement, and a HUMS that has been in service for approximately 15 years, this is an encouraging result. Furthermore, as the current purpose of the anomaly alerts is to draw the operator's attention to any abnormal data behaviour, the anomaly detection system's thresholds were deliberately set at a relatively sensitive level. The system's alert rate should decrease as a result of the model tuning exercise that will be performed before the planned second six-month trial period.

Table 5-1 Anomaly detection system alerts

Sensor	No. of Analyses	Alerts	Normalised Alerts
1	2	49	24.5
2	2	120	60.0
3	9	379	42.1
4	6	245	40.8
5	2	130	65.0
6	2	99	49.5
7	6	365	60.8
8	1	4	4.0
9	1	23	23.0
11	2	38	19.0
12	2	32	16.0
Totals	**35**	**1,484**	

Table 5-2 IHUMS alerts (sensor and defect)

Sensor	No. of Analyses	Alerts	Normalised Alerts
1	2	58	29.0
2	2	40	20.0
3	7	271	38.7
4	5	221	44.2
5	2	77	38.5
6	2	86	43.0
7	6	392	65.3
8	1	47	47.0
9	1	54	54.0
11	2	97	48.5
12	2	38	19.0
Totals	**32**	**1,381**	

The comparative durations of the anomaly detection system and IHUMS alerts are shown in Figure 5-1 and Figure 5-2 (for clarity, the duration is truncated in the first figure and the alert count is truncated in the second). It can be seen that there are a number of long duration anomaly detection system alerts, with a few alerts being present throughout the six-month trial period. In contrast, the vast majority of the IHUMS alerts have a duration of only one day. There are several reasons for this difference. The IHUMS is the "executive system", and therefore IHUMS alerts must be responded to. Also the IHUMS uses datum thresholds, which are calculated based on a limited number of data samples. As a result, the IHUMS may not alert on data that is consistently anomalous compared to the fleet norm. The system is also normally re-datumed after an alert has been investigated, which would re-set the threshold and clear the alert. However, the anomaly detection system will continue to generate alerts if the data remains abnormal. This is consistent with the persistent IHUMS instrumentation faults that were detected.

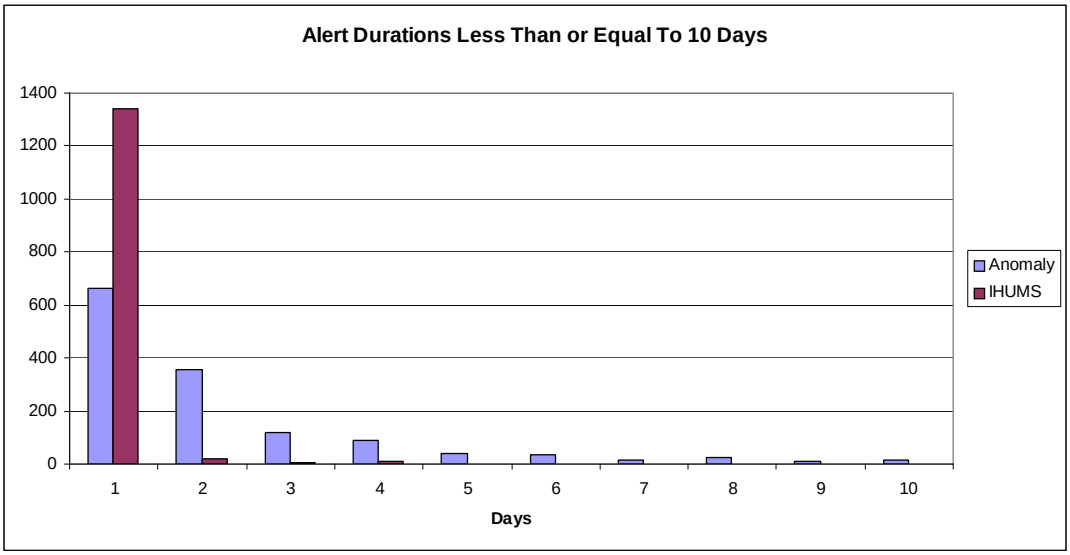

Figure 5-1 Anomaly detection system and IHUMS alert durations (duration truncated at ten days)

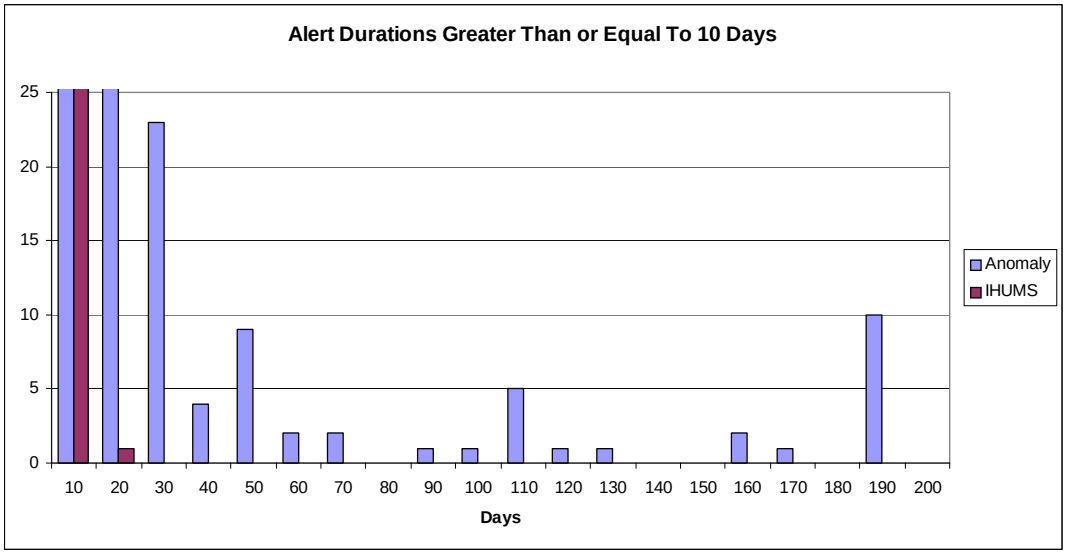

Figure 5-2 Anomaly detection system and IHUMS alert durations (alert count truncated at 25)

Table 5-1 and Table 5-2 also show the number of alerts that have been triggered on data from each IHUMS sensor. As the number of component analyses performed using data from each sensor varies, the alert figures per sensor have been normalised by the number of analyses using that sensor to give a more meaningful comparison between sensors. This comparison is presented graphically for both the anomaly detection system and the IHUMS in Figure 5-3. Figure 5-4 compares the number of individual data points in alert for each sensor, with the data again being normalised by the number of analyses performed per sensor. The fact that there are more data points in alert for the anomaly detection system than the IHUMS is due to the longer duration of the alerts generated by this system, as illustrated in Figures 5-1 and 5-2. The data points in alert are also presented in a pie chart format in Figure 5-5.

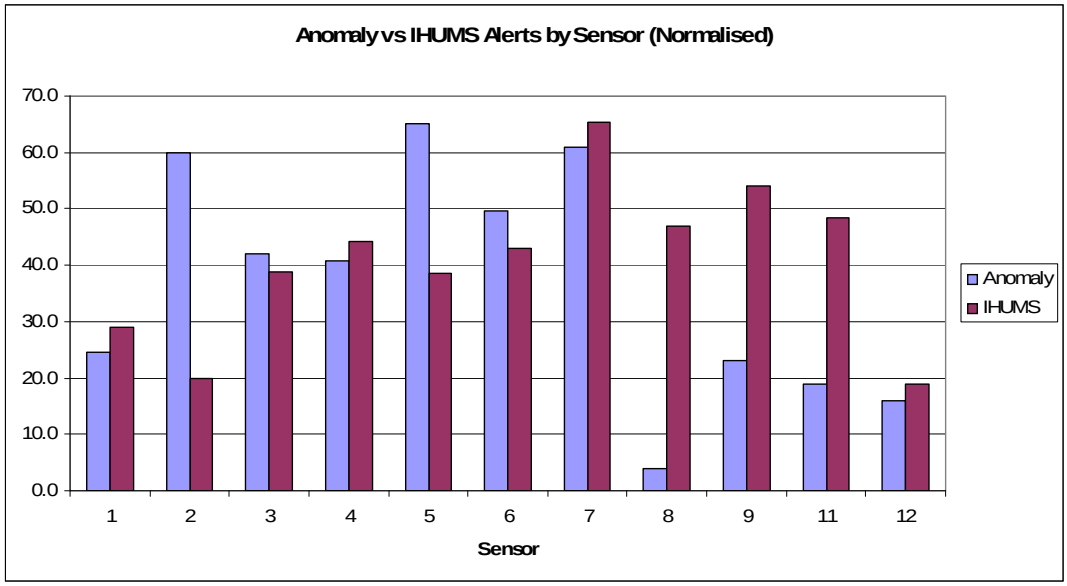

Figure 5-3 Anomaly detection system and IHUMS alerts by sensor (normalised)

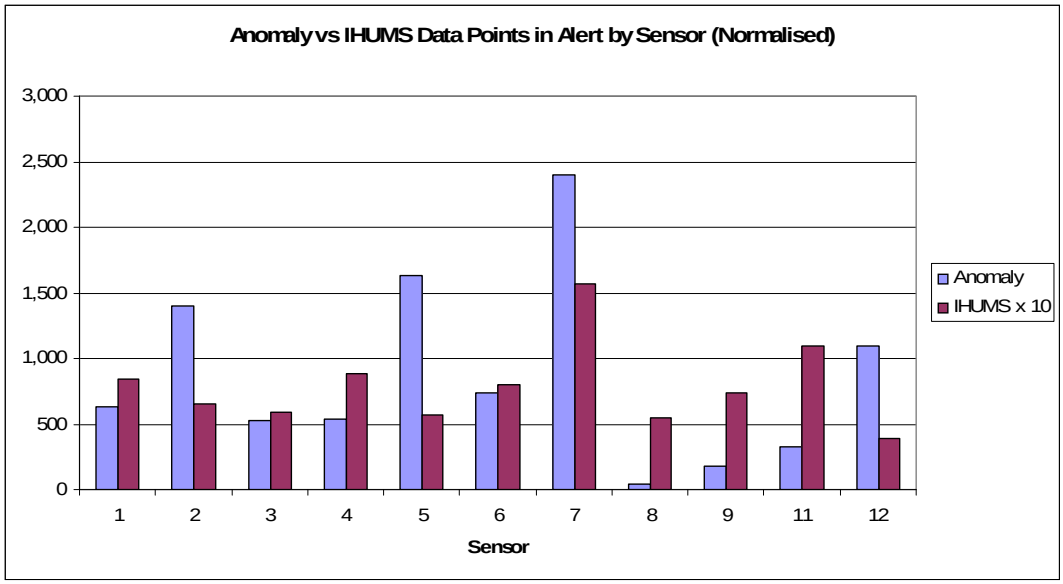

Figure 5-4 Anomaly detection system and IHUMS data points in alert by sensor (normalised)

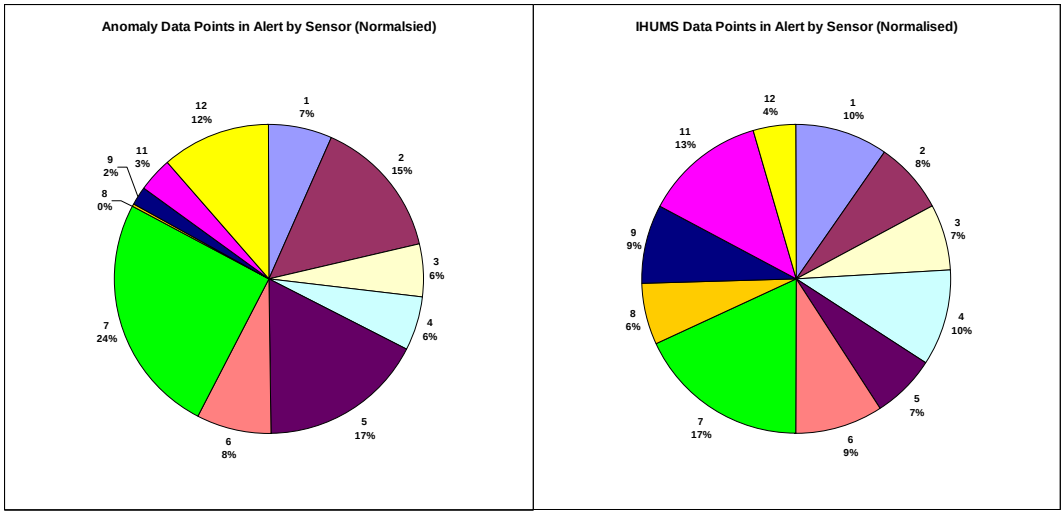

Figure 5-5 Anomaly detection system (left) and IHUMS (right) data points in alert by sensor (normalised)

There is some difference in the balance of alerts across the different sensors between the anomaly detection system and the IHUMS (Figure 5-3). Sensor 7 has the greatest number of data points in alert in both systems (Figure 5-4 and Figure 5-5), however the actual percentage is greater in the anomaly detection system. This reflects the issues with sensor 7 that were discussed in Sections 4.1 and 4.2.

5.2 Anomaly Detection System and IHUMS Alerts by Analysis

Figure 5-6 shows the number of anomaly detection system and IHUMS alerts that have been triggered on the results from the analysis of each drive train component. Figure 5-7 shows the number of individual data points in alert by component analysis.

The anomaly detection system has triggered more alerts on the 'MGB left torque shaft – aft end' than on any other component (Figure 5-6). This numerical evidence supports the operator's perception described in example 1 of Section 4.5 that the anomaly models are too sensitive on this component. One of the most notable differences in alerts between the anomaly detection system and the IHUMS is on the 'Tail rotor Gearbox output', where the IHUMS has generated significantly more alerts. This may suggest a need to increase the sensitivity of the relevant anomaly model, particularly to shaft order vibration related to tail rotor problems (see also example 1 in Section 4.3).

Both systems have a high number of data points in alert on the six 1st and 2nd stage epicyclic gear analyses that use sensor 7 (Table 2-1), but the numbers are relatively higher for the anomaly detection system (Figure 5-7). This correlates with the data in Figure 5-4.

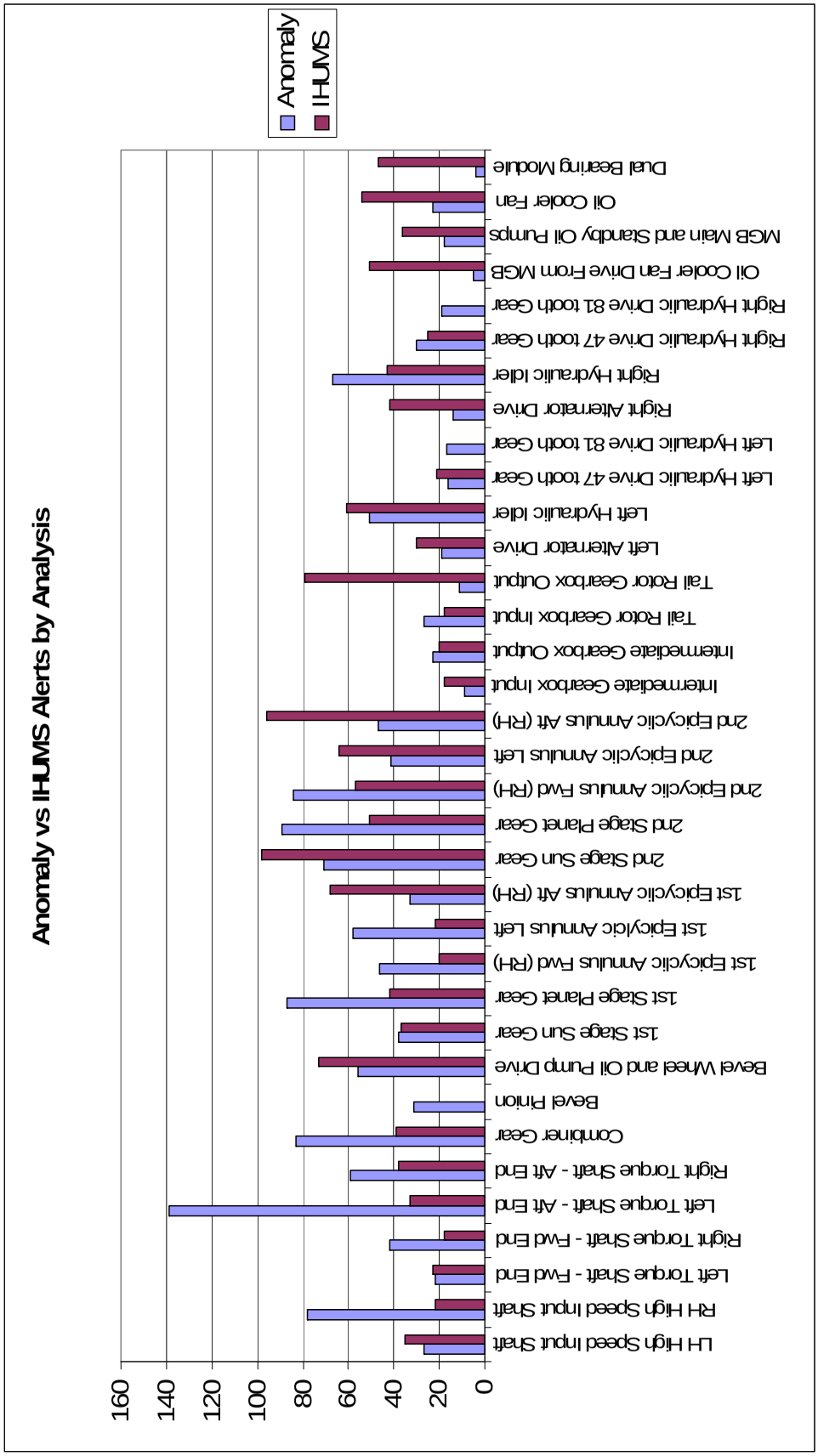

Figure 5-6 Anomaly detection system and IHUMS alerts by component analysis

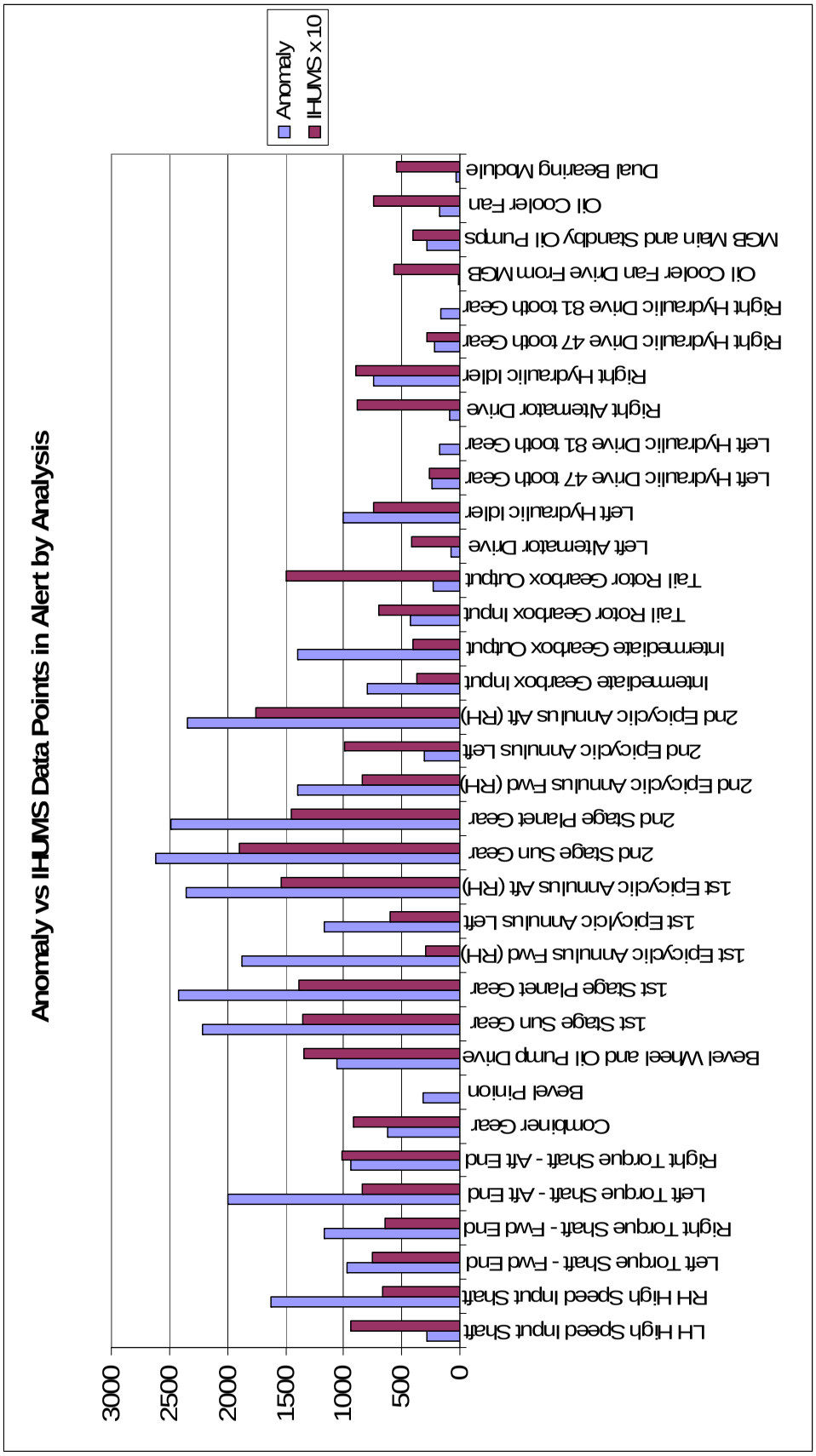

Figure 5-7 Anomaly detection system and IHUMS data points in alert by component analysis

5.3 Anomaly Detection System Alerts by Sensor and Aircraft

Figure 5-8 presents a series of pie charts showing, for each sensor, the proportions of anomaly detection system data points in alert for each aircraft included in the six-month trial. The first chart (top left of Figure 5-8) shows the proportions of the total data points in alert for each sensor (in this case the data has not been normalised by the number of analyses performed per sensor). By far the greatest proportion of data points in alert are from sensor 7 (42%). Conversely, very few alerts have been triggered on data from sensor 8.

The data points in alert from sensor 1 are dominated by G-BMCW (89% of total), which provides a statistical confirmation of the significance of the anomaly indications from the 'MGB left torque shaft - fwd end' described in example 1 in Section 4.6.

The data points in alert from sensors 2 and 12 are dominated by G-PUMI, with 63% and 70% of the total respectively. The aircraft also generates the greatest proportion of the data points in alert on sensor 5 (29%). Again, this confirms the significance of the anomaly indications on multiple components (including the 'MGB RH high speed input shaft', 'right torque shaft – fwd end', and '2nd epicyclic annulus fwd (RH)', plus the 'Intermediate Gearbox output') described in example 3 in Section 4.2. These were believed to have been caused by a DAPU earthing problem.

The MGB epicyclic instrumentation problem on G-TIGC described in example 3 in Section 4.1 is clearly reflected in the pie charts for sensors 5, 6 and 7. The aircraft generates the largest proportion of data points in alert on sensors 6 and 7 (44% and 41% respectively), and also a significant proportion of data points on sensor 5 (27%). A similar instrumentation problem on G-TIGS (example 4 in Section 4.1), but affecting data from sensor 7 only, can also be seen, with the aircraft generating 40% of the data points in alert on this sensor. The two aircraft therefore account for 81% of all the data points from sensor 7.

On several other sensors, the majority of the data points in alert are only generated by two or three aircraft. For example, G-TIGE generated 31% of the data points from sensor 4, which was largely due to the anomalous bevel wheel and oil pump drive data described in example 5 in Section 4.1. All these examples illustrate how the statistical information highlights particular aircraft anomalies. It also demonstrates that the anomaly detection system alert rate is not driven by random alerting, but the system's response to these particular anomalies.

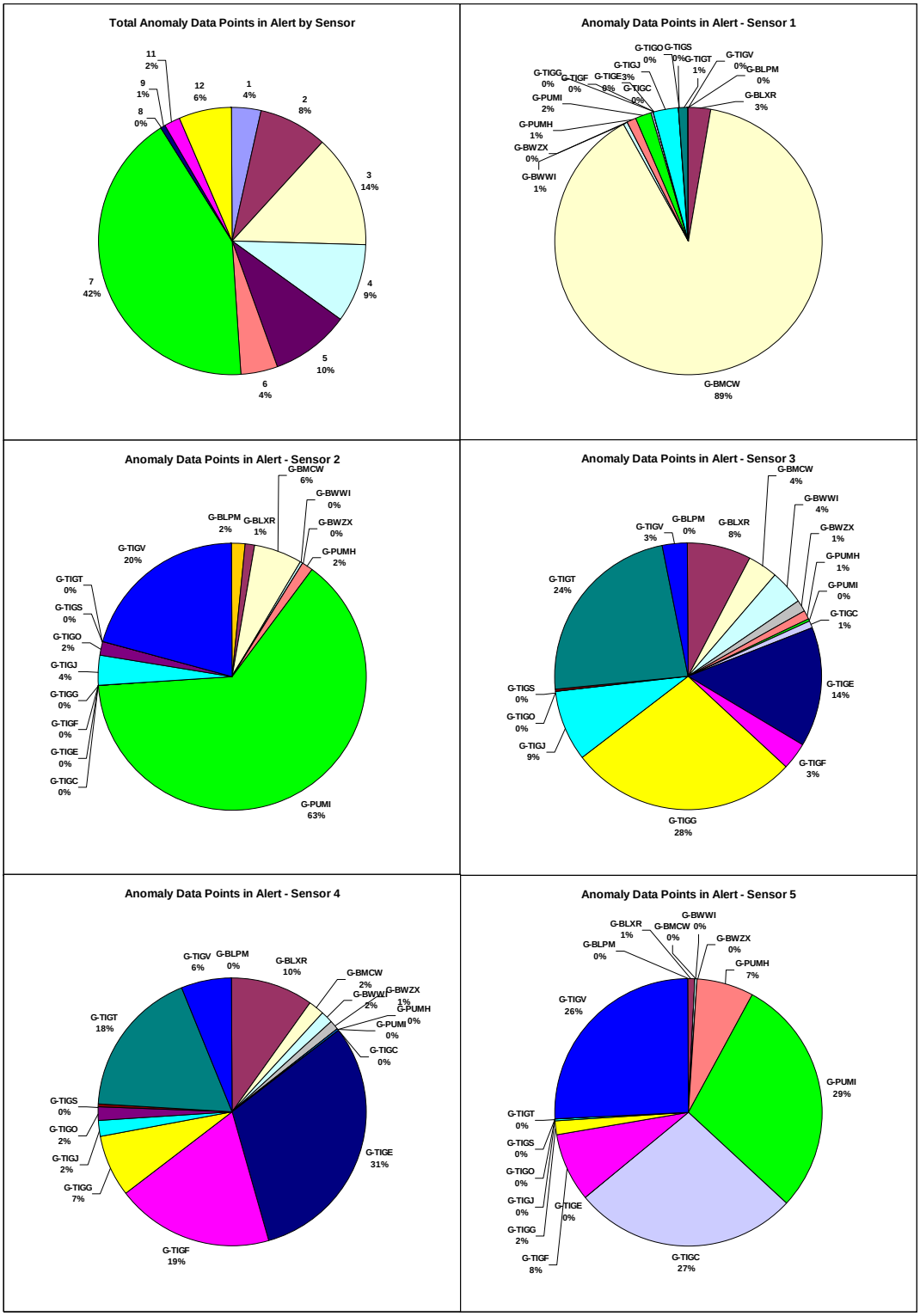

Figure 5-8 Anomaly detection system data points in alert by sensor and aircraft – Part 1

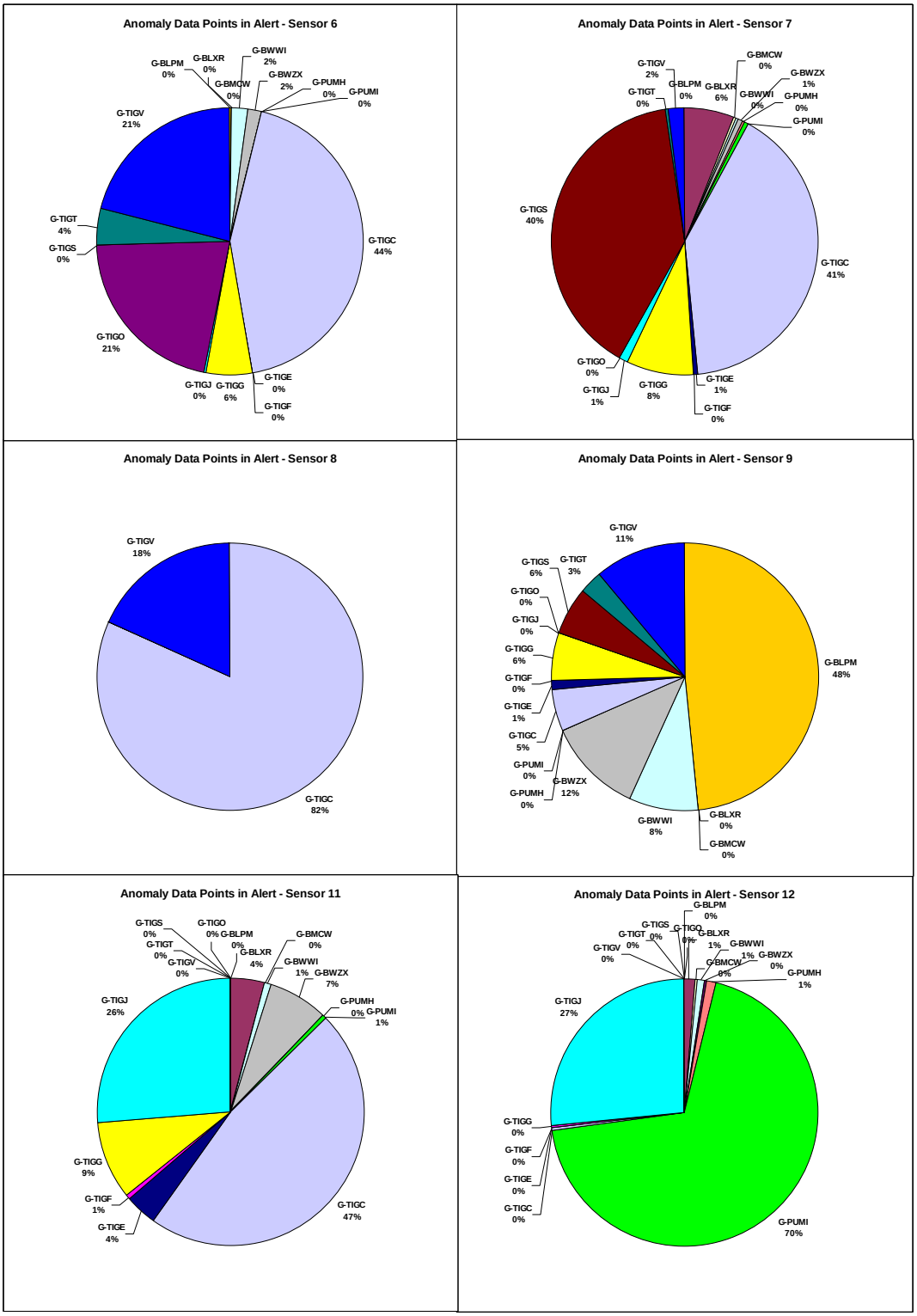

Figure 5-8 Anomaly detection system data points in alert by sensor and aircraft – Part 2

5.4 Anomaly Model Performance

A large majority of the drive train components were monitored by four anomaly models – '8-indicator' absolute and trend models, and 'M6' absolute and trend models. There were two components without gears that were monitored by two

'5-indicator' models. An anomaly detection system alert was triggered whenever one or more of the models went into an alert state. Counting the individual model outputs separately, these triggered a total of 2,213 alerts, which relates to the 1,484 system alerts. The difference of 729 is a result of multiple models going into alert at the same time.

Figure 5-9 shows the number of alerts triggered by each of the six different types of anomaly model. Figure 5-10 shows the number of data points in alert by model type. Both the greatest number of alerts and data points in alert were generated by the '8-indicator' models. There was, however, a marked difference in the ratio between the '8-indicator' absolute and trend models. The trend models triggered slightly more alerts, but the absolute models had significantly more data points in alert. The explanation for this is believed to be that the system has identified a significant amount of anomalous IHUMS data due to instrumentation problems and, in particular, problems associated with sensor 7.

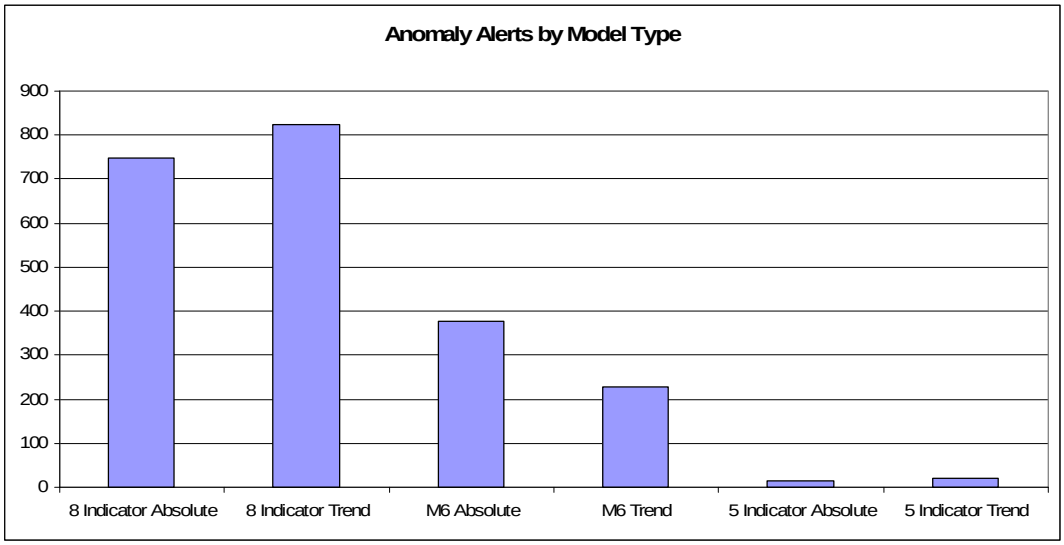

Figure 5-9 Anomaly alerts by model type

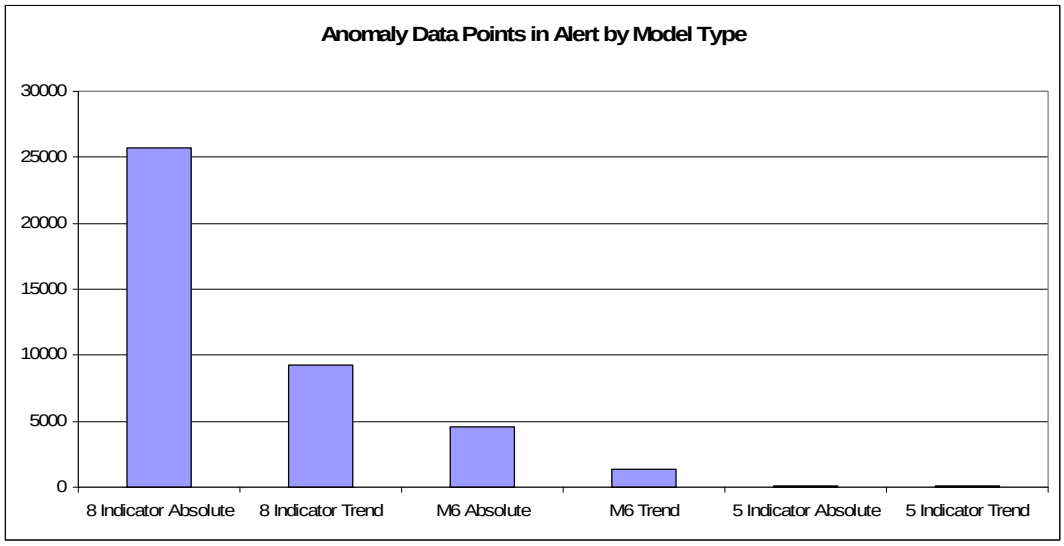

Figure 5-10 Anomaly data points in alert by model type

To further assess the performance of the individual models, breakdowns of alerts and data points in alert by component analysis and sensor are presented for each of the four main model types (i.e. '8-indicator' absolute and trend, and 'M6' absolute and trend). Figure 5-11 and Figure 5-12 show the alerts and data points in alert generated by the '8-indicator' absolute models built for each component (there is one model per component).

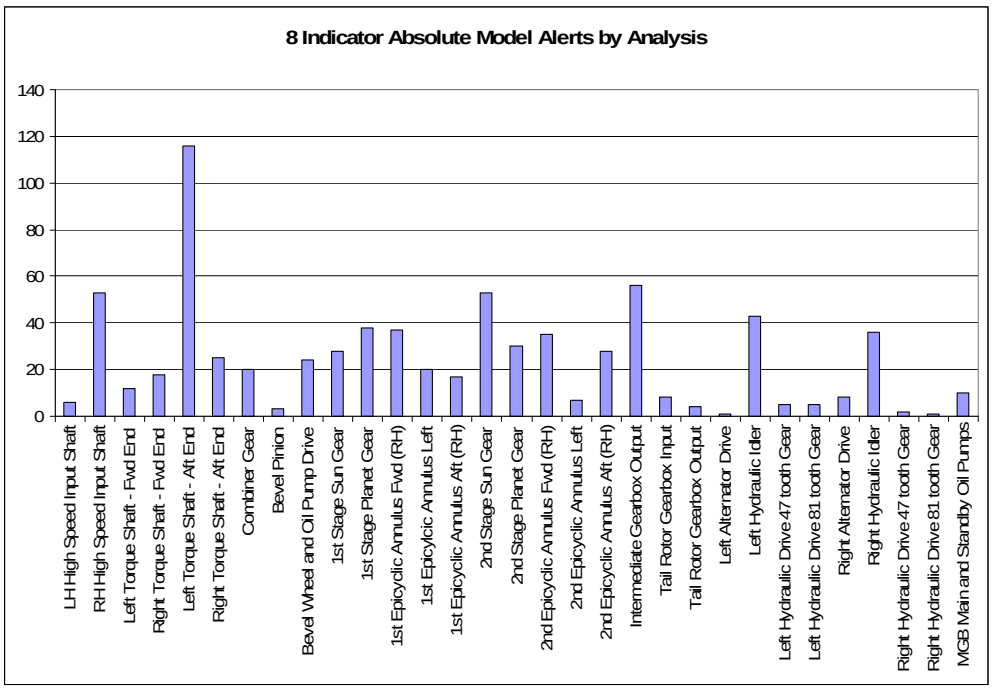

Figure 5-11 '8-Indicator' absolute model alerts by component analysis

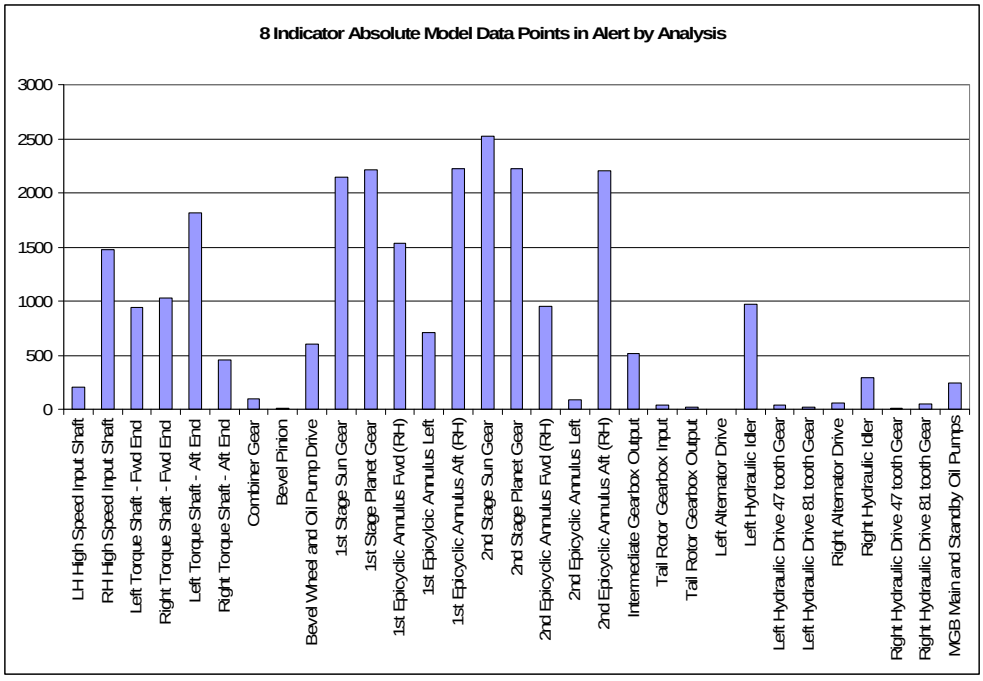

Figure 5-12 '8-Indicator' absolute model data points in alert by component analysis

There were over twice as many alerts on the left torque shaft – aft end than on any other component (Figure 5-11) which, as commented previously, suggests that some tuning of this model is required. The greatest number of data points in alert were from the six epicyclic component analyses that use sensor 7 (Figure 5-12). The fact that this sensor caused the most '8-indicator' absolute model data points in alert is illustrated in Figure 5-13. The right hand chart showing normalised data (where the number of data points for a sensor is divided by the number of analyses using that sensor), gives the most meaningful picture of the relative prominence of sensor 7. It should be noted that 81% of the data points in alert on this sensor were generated by only two aircraft (Figure 5-8).

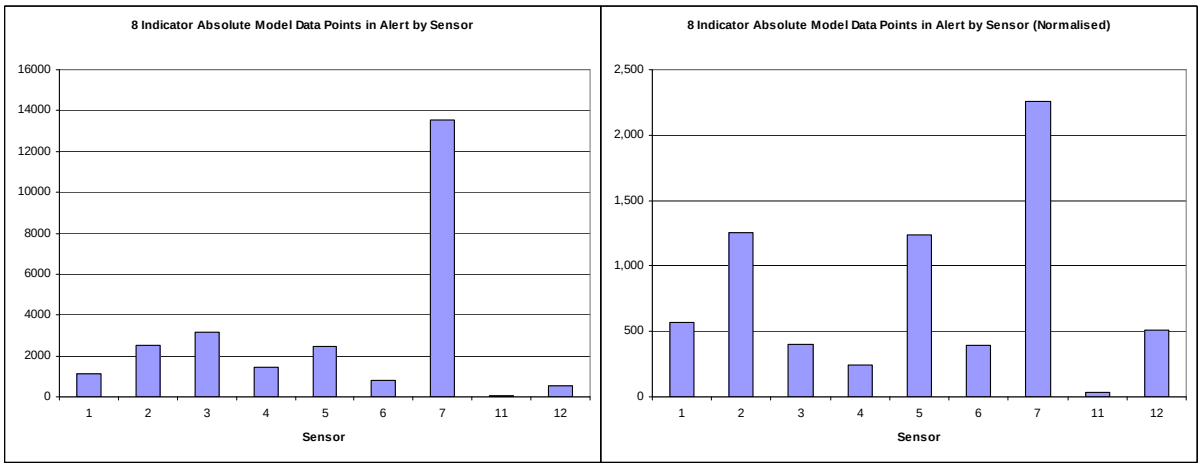

Figure 5-13 '8-Indicator' absolute model data points in alert by sensor (the chart on the right shows normalised data)

Figure 5-14 and Figure 5-15 show the alerts and data points in alert generated by the '8-indicator' trend models for each component.

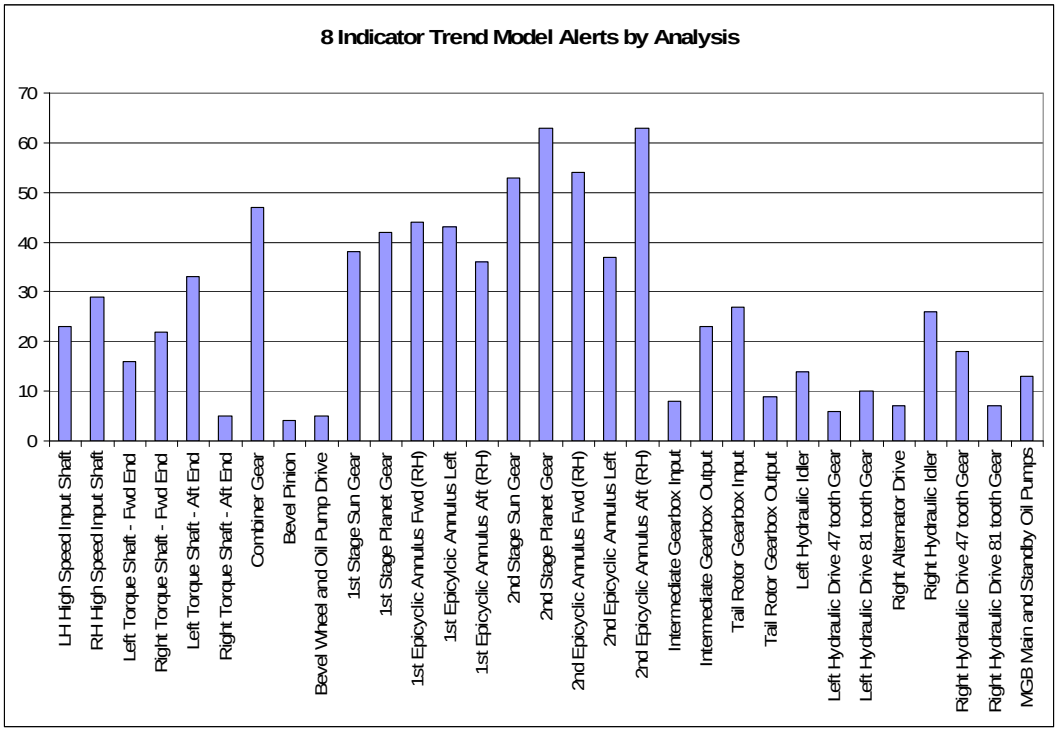

Figure 5-14 '8-Indicator' trend model alerts by component analysis

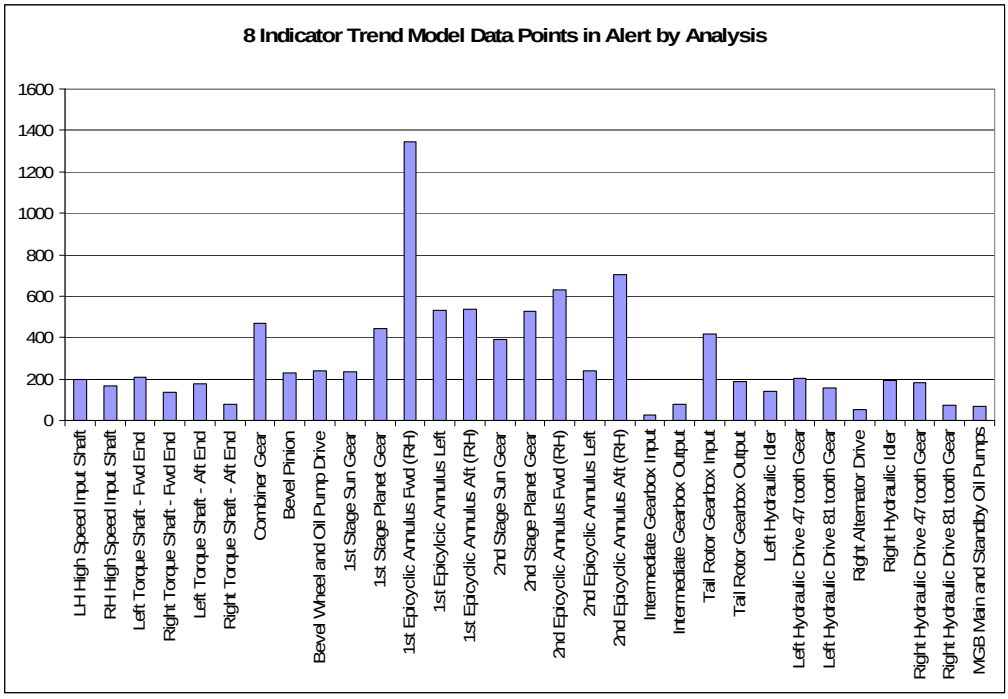

Figure 5-15 '8-Indicator' trend model data points in alert by component analysis

Although not prominent in the alert data (Figure 5-14), the 1st epicyclic annulus fwd (RH) '8-indicator' trend model has approximately twice as many data points in alert as any other model (Figure 5-15). Displaying the data points by sensor illustrates the prominence of sensor 5 (Figure 5-16), which is the sensor used for analysing the 1st epicyclic annulus fwd (RH) component. Figure 5-8 showed that a high proportion (82%) of the total number of data points in alert (i.e. from all models) on sensor 5 were generated by three aircraft.

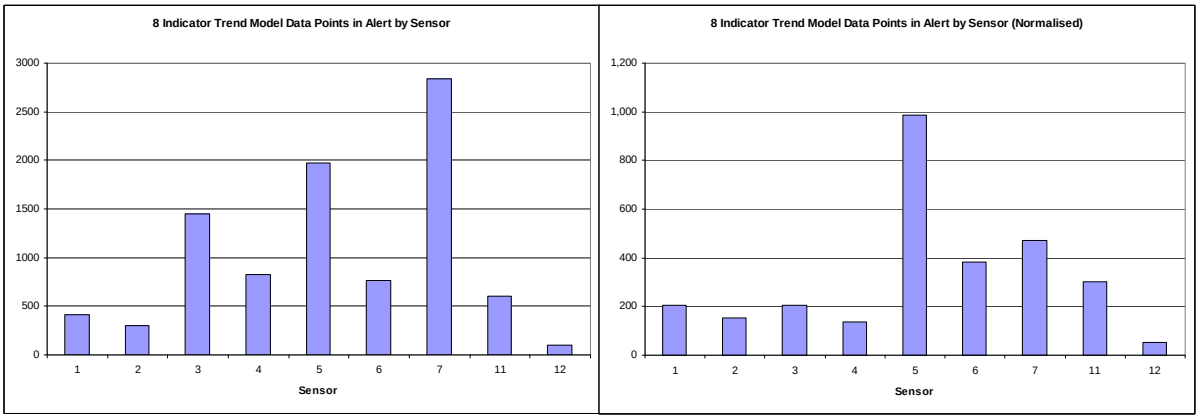

Figure 5-16 '8-Indicator' trend model data points in alert by sensor (the chart on the right shows normalised data)

The alerts and data points in alert generated by the 'M6' absolute models for each component are shown in Figure 5-17 and Figure 5-18. The high number of alerts on the bevel wheel and oil pump drive is primarily due to instrumentation problems on G-TIGE described in example 5 in Section 4.1. By far the largest number of data points in alert are from the Intermediate Gearbox. The fact that these are associated with only a small number of alerts indicates that these are connected with a problem on a particular aircraft. The aircraft in this case is G-PUMI, as described in example 3 in Section 4.2. This aircraft is also responsible for the predominance of sensor 12 in Figure 5-19.

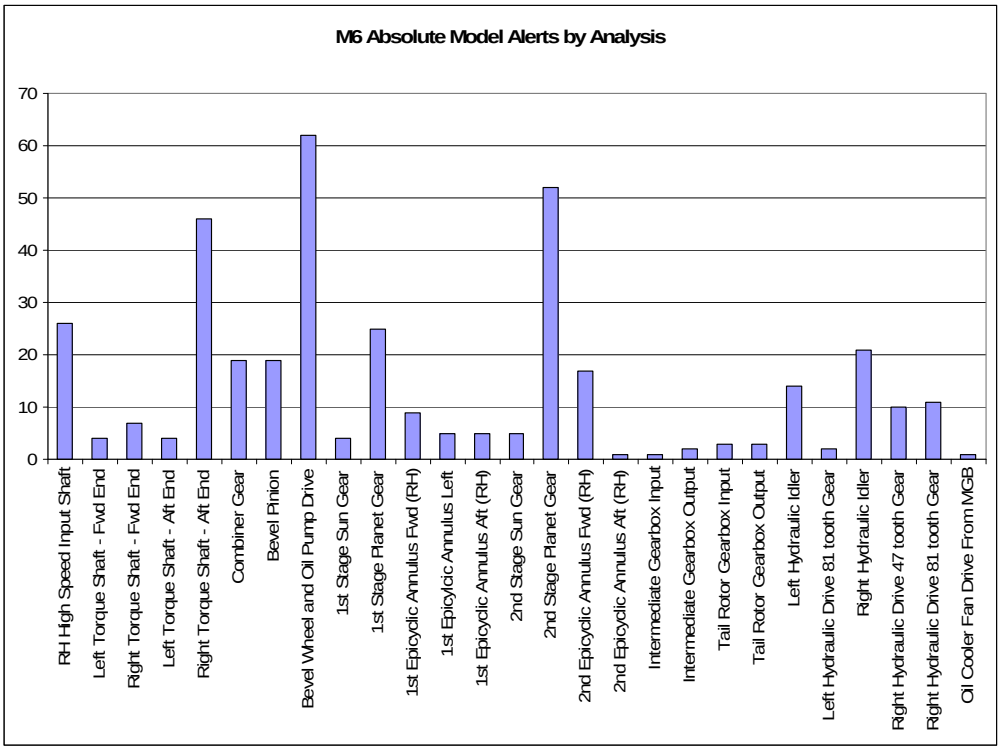

Figure 5-17 'M6' absolute model alerts by component analysis

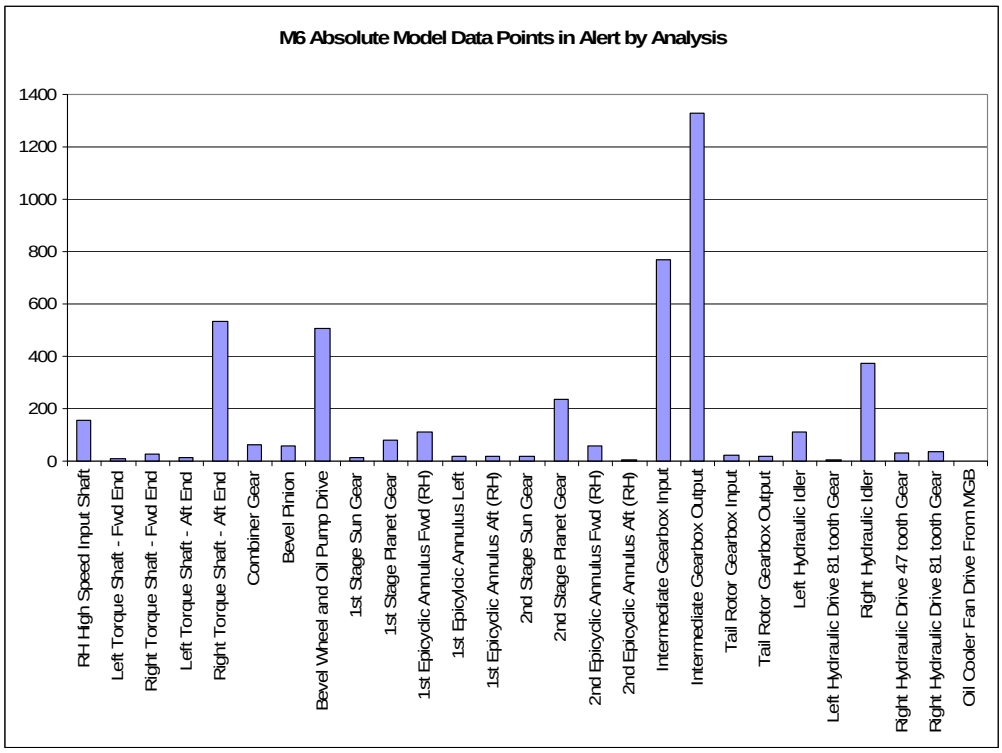

Figure 5-18 'M6' absolute model data points in alert by component analysis

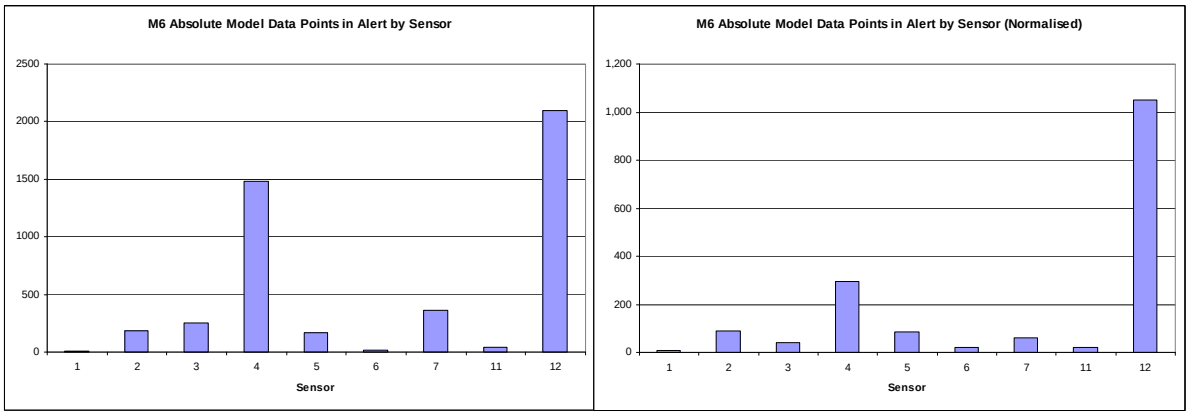

Figure 5-19 'M6' absolute model data points in alert by sensor (the chart on the right shows normalised data)

Finally, the M6 trend model data is shown in Figure 5-20 to Figure 5-22. The high number of data points in alert generated by the 1st epicyclic annulus fwd (RH) component and the associated sensor 5, is due to specific aircraft issues, including the previously discussed instrumentation issues on G-PUMI.

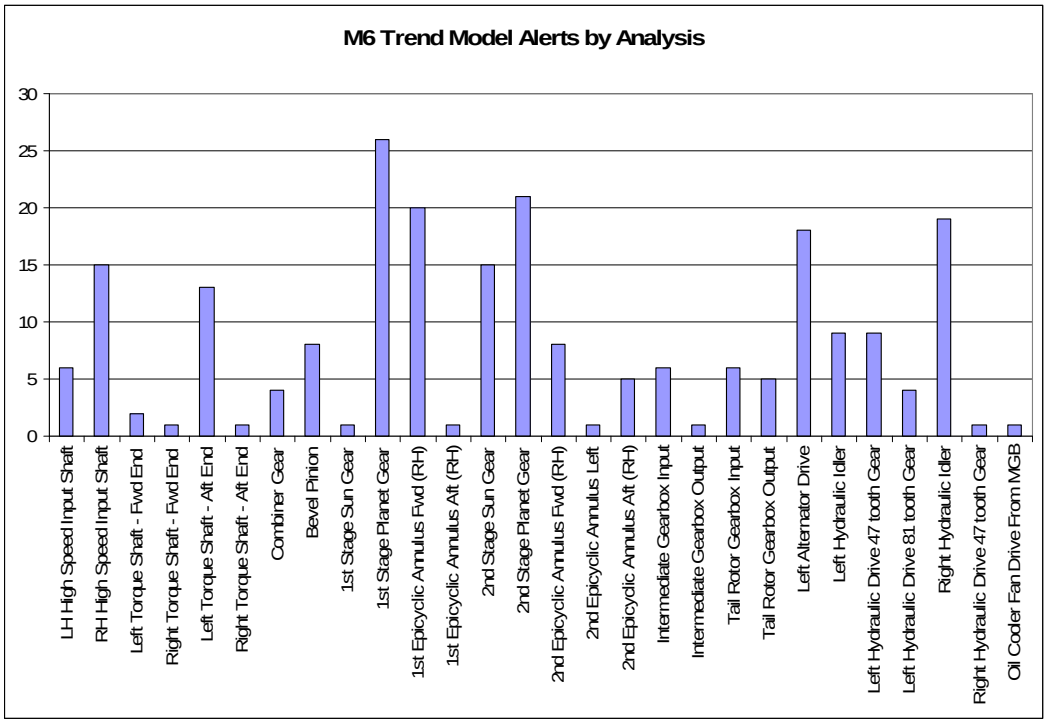

Figure 5-20 'M6' trend model alerts by component analysis

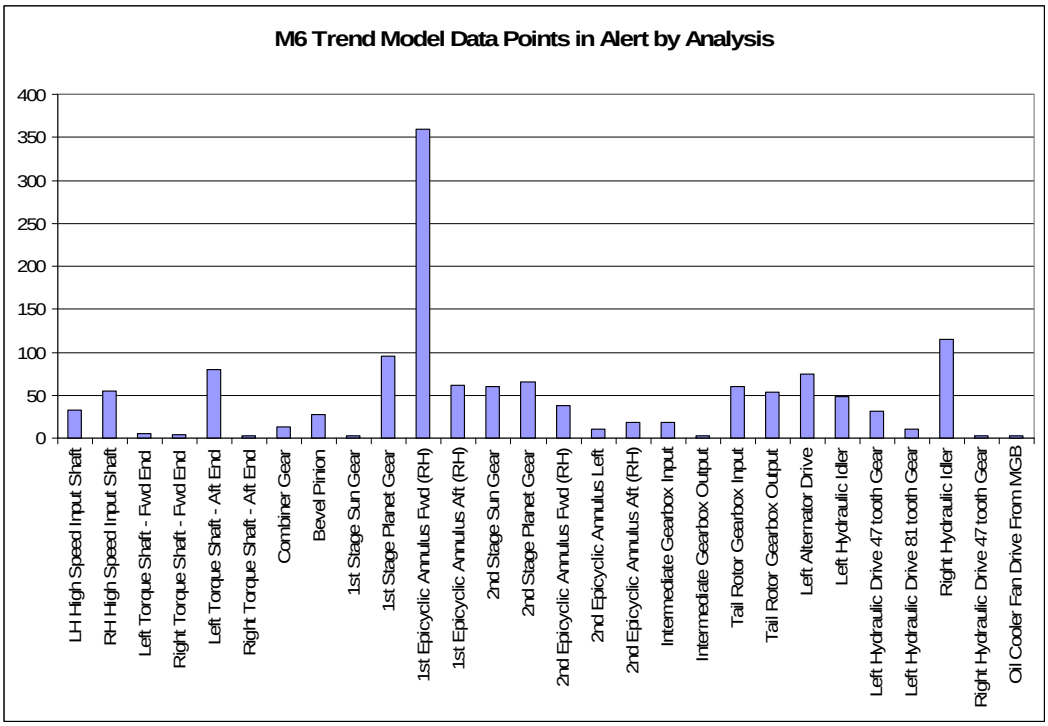

Figure 5-21 'M6' trend model data points in alert by component analysis

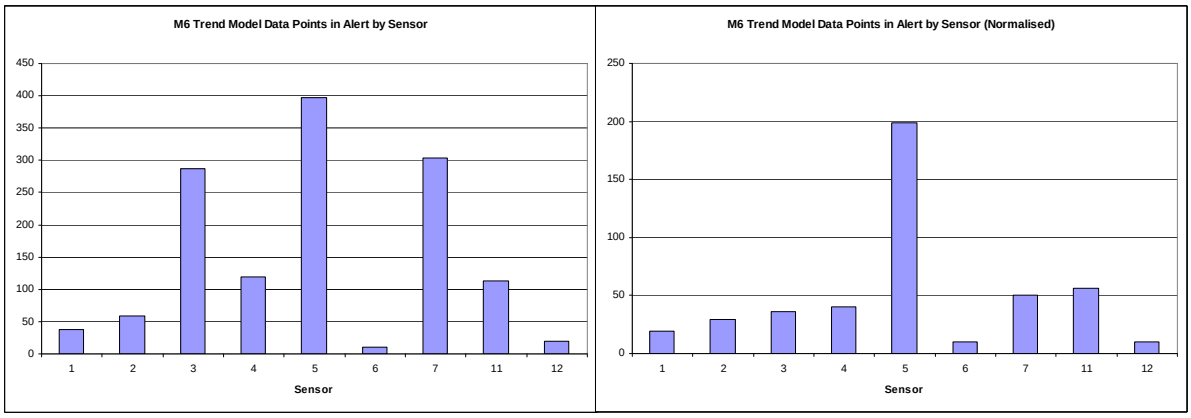

Figure 5-22 'M6' absolute model data points in alert by sensor (the chart on the right shows normalised data)

6 Conclusions and Recommendations

6.1 Conclusions

The initial six-month trial has confirmed that the anomaly detection system represents a major advance in HUMS data analysis. The system has highlighted anomalous data that were not seen by the existing HUMS. It has also provided a much clearer picture of anomalous patterns on particular aircraft and drive train components than is possible with the traditional HUMS analysis. The system can therefore improve the effectiveness of HUMS, and also provide valuable assistance to an operator in the management of the system.

The trial verified that the usability of the web-based implementation of the anomaly detection system was good. The operator found this to be user friendly, with a simple interface that enabled straightforward navigation to different items of information. The responsiveness of the website to remote commands accessing various chart displays over the internet was found to be acceptable. The reliability of the automated overnight data transfer process was very good, and the reliability of access to the anomaly detection system website was also good. Very high reliability would be required for both of these aspects of system performance in any production implementation.

The anomaly detection system detected a number of aircraft instrumentation faults that were not seen by the existing HUMS. In at least one case, it also detected significant trends on a drive train component that may have been fault related. Unfortunately, although a gearbox strip report had been requested to confirm the finding, no feedback could be obtained. The system has also corroborated aircraft drive train component and instrument faults that were detected by the existing HUMS. Due to the prevalence of trends in the accessory gearbox data, in one case the anomaly detection system correctly suppressed the significance of trends that were triggering HUMS alerts. The gearbox was placed on 'close monitor' as a result of the HUMS alerts, but was later 'stood down' without any maintenance actions being performed.

The anomaly detection system has highlighted the fact that the IHUMS data on Bristow's AS332Ls can be significantly affected by noise due to instrumentation issues on the aircraft. There have been a number of HUMS accelerometer, insulation washer, and cable or connector problems, which may be due to the fact that these aircraft had some of the first HUMS installations. The instrumentation problems primarily affect the low frequency narrow band indicators SO1 and SO2, causing frequent periods of very high and erratic SO1 and SO2 values. Not only do these trigger anomaly model alerts, but their prevalence pollutes the model training data, which can result in a loss of fault detection sensitivity. Wiring harness problems on the main gearbox epicyclic gears have been shown to generate abnormally high signal levels which could mask fault related trends. Accelerometer 7 on the MGB epicyclic stage created the largest number of instrumentation-related anomaly alerts, but 81% of all the data points in alert were associated with only two aircraft.

Using the initially defined set of alert thresholds, the anomaly detection system generated approximately the same number of alerts as the current HUMS. Given that the comparison is between a system undergoing its first operational trial, without any opportunity for refinement, and a HUMS that has been in service for approximately 15 years, this is considered to be a good result. Furthermore, the initial anomaly thresholds were defined with the intent of drawing the operator's attention to 'interesting' items of data, which can result in a higher alert rate than would be the case if these were intended to indicate with confidence the potential presence of a

component fault. Although the '8-indicator' trend model generated slightly more alerts than the absolute model, the absolute model generated by far the greatest number of data points in alert. This is believed to be due to the noisy nature of the HUMS data described above.

The trial results have indicated that some anomaly models may be too sensitive, in particular the models for the MGB left torque shaft – aft end. Conversely, one example case suggested that the sensitivity of models to faults on the TGB Output shaft may be suppressed by the fact that the model also includes gear indicators. Therefore some model tuning and/or re-modelling would be beneficial.

It has proved to be relatively difficult to investigate anomaly detection system alerts by checking the data in the current HUMS where the HUMS has not also triggered an alert. In this case it is necessary to rely on a visual interpretation of data trends in the HUMS, and the visual impression of these is highly dependent on the amount of historical data viewed and the vertical scaling of the plot. The fleet (Condition Indicator) CI displays in the anomaly detection system are considered to give the clearest indication of the significance of any data trends.

The lack of feedback information on the condition of drive train components removed from aircraft, such as from gearbox strip reports, often prevents any meaningful interpretation of HUMS CI trends in terms of component condition. This type of feedback is critical to the on-going development of HUMS data analysis capabilities, and the absence of such information is seriously hindering this development.

6.2 Recommendations

OEMS and overhaul agencies should be strongly encouraged to routinely record component condition information when gearboxes are stripped down for overhaul. In addition, there should be a reliable mechanism to enable operators to request information on the condition of a particular component, or group of components, that have associated HUMS CI trends, or have been generating metallic debris.

Based on the results of the six-month trial period, some model tuning should be performed to adjust the sensitivity of models on certain shafts. In addition, for key input and output shafts (e.g. the MGB inputs and TGB output), consideration should be given to the implementation of additional shaft models (i.e. with a smaller set of shaft related indicators) to improve sensitivity to shaft related problems. (NB: A recommendation on model tuning was included in reference [1]).

Where feasible, efforts should be made to improve the standard of the HUMS instrumentation on Bristow's AS332L fleet to address the noisy data issues that have been identified during the trial.

For any web-based production system implementation, the reliability of access to the website, and of the routine data transfer process, must be assured. Therefore appropriate infrastructure and procedures must be implemented to guarantee this.

7 References

1 GE Aviation report REP1697(2): "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Interim Report on Phase 1 of the Research Project". May 2007.

INTENTIONALLY LEFT BLANK

ANNEX C

Report on Additional Research Work (Tasks 1 to 4)

Based on a report prepared for the CAA by GE Aviation Systems Limited, UK

Table of Contents

Report

INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK

Glossary

AGB	Accessory Gearbox
BIC	Bayesian Information Criterion
CAA	Civil Aviation Authority (UK)
CI	Condition Indicator
DAPU	Data Acquisition and Processing Unit
FS	Fitness Score
HUMS	Health and Usage Monitoring System
IF	Influence Factor
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
LHA	Left Hand Accessory module
MGB	Main rotor Gearbox
PA	Probability of Anomaly
ProDAPS	Probabilistic Diagnostic and Prognostic System
RHA	Right Hand Accessory module
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring

INTENTIONALLY LEFT BLANK

Executive Summary

This report documents the results of additional research work conducted under a CAA funded research programme to demonstrate the intelligent analysis of helicopter HUMS Vibration Health Monitoring (VHM) data. The programme is applying unsupervised learning techniques to HUMS data to construct an anomaly detection system with the goal of improving HUMS fault detection performance. This data driven modelling approach utilises existing HUMS data to define models of 'normal' behaviour, which can then be used to detect 'abnormal' or 'anomalous' behaviour that may be associated with a defect.

An unsupervised learning technique for anomaly detection, and system to implement it, were developed in Phase 1 of the CAA research programme, which also included an off-line demonstration using a database of historical AS332L HUMS data. In Phase 2 of the programme, the anomaly detection system was subjected to a six-month in-service trial by Bristow Helicopters. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L fleet HUMS data being automatically transferred every night from Aberdeen, and Bristow having a remote login to the server to check system anomaly indications. The results of Phases 1 and 2 are reported in references [2] and [3] respectively.

Building on the successful in-service trial of the HUMS anomaly detection system, four additional research tasks have been performed to further develop its advanced HUMS data analysis capabilities. These have resulted in significant enhancements to the system, which will be evaluated in a second six-month in-service trial.

A number of improvements have been made to the modelling technique created during the initial system development phase. All anomaly models have been re-built using the improved technique, and a detailed review of the outputs has confirmed enhanced modelling performance. In addition, new shaft models have been built for the critical MGB input and TGB output shafts, resulting in the detection of a tail rotor flapping hinge bearing problem that was previously missed. A new "Probability of Anomaly" (PA) output has been derived for each anomaly model, which is a normalised probability measure that ranges between 0 and 1. A more robust method for setting thresholds on anomaly models has been implemented, based on the PA values. A global default threshold of 0.9 PA is now set across all models and components. New "Influence Factors" (IFs) have also been implemented to provide model diagnostic information. IF traces assess the contribution of individual HUMS CIs to a statistical measure of normality derived from the model. IFs are also normalised and can be directly compared. Finally, the research has identified an opportunity to perform shape characterisation of anomaly model "Fitness Score" (FS) outputs to detect the nature of a change (i.e. identifying significant trends, step changes, and noise spikes in the HUMS data).

In addition to further enhancing both the performance and usability of the HUMS anomaly detection system, the new capabilities will also facilitate future developments to further increase the effectiveness of HUMS such as the application of automated diagnostic reasoning, and data mining to search for new knowledge.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Health and Usage Monitoring Systems (HUMS), incorporating comprehensive drive train Vibration Health Monitoring (VHM), have contributed significantly to improving the safety of rotorcraft operations. However, experience has also shown that, whilst HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected using current threshold setting methods. Earlier research (see [1]), conducted as part of the CAA's helicopter main rotor gearbox seeded defect test programme, demonstrated the potential for improving fault detection performance by applying unsupervised machine learning techniques such as clustering to seeded fault test data. The CAA therefore commissioned a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841). Smiths Aerospace (now GE Aviation) was contracted to carry out this programme in partnership with Bristow Helicopters, analysing IHUMS data from Bristow's North Sea AS332L fleet.

The work under Contract No. 841 was structured as two phases. Phase I started with a literature survey to identify other research work on unsupervised learning, and went on to develop an unsupervised analysis technique to enhance the fault detection performance of HUMS which was subjected to an 'off-line' demonstration on a database of historical AS332L data. It also included the development of a system implementing the technique for evaluation in an in-service trial. The results of Phase 1 are reported in reference [2].

The unsupervised learning approach is known as anomaly detection. The need for anomaly detection arises in situations where supervised learning of fault related patterns is impractical. Supervised learning requires that every signal acquisition be tagged with a known healthy or particular fault classification. Fortunately component related faults are rare, but this means that there is no large library of tagged fault data. Even if such a library existed, the required size would be indeterminate because symptoms for a single fault type can vary between cases. A large majority of the acquired HUMS data can be assumed to be associated with a normal (healthy) component state. Recognition of a fault could therefore be achieved by building a model of normal behaviour against which a test for the abnormal can be applied. In the context of this work, anomaly detection is considered to be a process that calls attention to abnormal data that requires investigation.

Phase 2 of the CAA research programme comprised a six-month in-service trial, conducted by Bristow Helicopters, of the anomaly detection system developed in Phase I. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L IHUMS data being automatically transferred every night from Aberdeen, and Bristow Helicopters having a remote login to the server to check the anomaly detection system outputs. The results of Phase 2 are reported in reference [3].

The CAA contract included an option for a six-month extension to the in-service trial of the HUMS anomaly detection system. Building on the success of the initial trial period, a number of additional work packages were identified and recommended to research and develop further advanced HUMS data analysis capabilities prior to the second six-month trial. The work covered the following four topics:

1 Pre-processing: Pre-processing is an important step in identifying significant features of the data upon which the anomaly detection is based. Pre-processing is designed to represent data in a way that characterises behaviour to be modelled during the anomaly learning process. For example, there is diagnostic information contained in the manner in which a component exhibits a significant increase in vibration (such as a step change or trend). Behavioural traits have to be described to models through pre-processing transformations.

2 Model tuning and re-modelling: The new modelling technology has been implemented as a fully automated process. However, models can only be as good as the data used to train them. For instance, if the training data contain a large number of similar anomalous cases then the model will consider these cases to represent normal behaviour. A-priori knowledge can provide insight into the importance of individual input features. It is possible for subtle behaviour on important features to be masked by other features. Reviewing model performance can highlight any deficiencies caused by bad data or masking effects. The review might suggest a need for model tuning (e.g. to remove the effect of detected bad data), or for re-modelling using a different sub-set of input features.

3 Probabilistic alerting policy: The primary output from an anomaly model is a 'Fitness Score' (FS), which provides a measure of abnormality. Alerts are triggered by setting a threshold on the FS, but using this directly to assess anomalies is not ideal. Each anomaly model has its own FS threshold and the FSs from different models are not directly comparable. Also, the anomaly models are statistical models and any future additional processing would be limited if anomalies are treated in a binary fashion (i.e. in or out of alert). The objective of this task is to describe anomalies on a normalised probability scale. On this scale, a value of 0 represents normal data whilst a value of 1 would indicate a clear anomaly.

4 Influence factors: Following the detection of an anomaly, further diagnostic analysis needs to be carried out to determine its nature (e.g. an instrumentation or component defect). Diagnostic information is contained in the HUMS Condition Indicators (CIs) (input features), but interpreting the significance of the levels of, or changes in, these indicators in their raw (or pre-processed) form can be difficult. The scale and characteristic behaviour can vary markedly between individual indicators and gearbox components. Indicators are not viewed with reference to what is normal; although standard deviation bands can be drawn on fleet plots, these standard parameter bands will be biased because they are derived from a fleet that usually contain a proportion of bad data. An influence factor is derived from an anomaly model and can give an indication of how much an anomaly might be apportioned to a specific CI. For a model, there will be one influence factor per CI used to train the model. A history of influence factor values computed from a sequence of acquisitions constitutes a set of influence traces. In other words, for a specific component and model, influence traces are time histories in the same sense that CI plots are component time histories. Influence factors provide measures that are suited to further automated analysis such as diagnostic reasoning. The objective of this task was to explore the application of influence factors to HUMS data and, if deemed suitable, to implement them in the anomaly detection system.

This report presents the results of the further work performed on each of the above topics under an amendment to the original contract. Section 2 first provides some background information by giving a brief overview of the status of the HUMS anomaly detection system prior to the start of the additional work. Section 3 presents the results of the investigation into further data pre/post-processing but for reasons explained therein, this section has been titled Data Characterisation, Section 4 describes the model tuning and re-modelling that has been performed. The implementation of a probabilistic alerting policy and influence factors for anomaly model diagnostics are discussed in Sections 5 and 6 respectively. Finally, some conclusions and recommendations are presented in Section 7.

2 Background Information on the Status of the HUMS Anomaly Detection System at the Start of the Further Work

2.1 IHUMS Data Analysed

The anomaly detection system analysed IHUMS VHM data from the following assemblies in the AS332L drive train: Main rotor Gearbox (MGB), left and right Accessory Gearboxes (AGBs) and oil cooler fan, Intermediate Gearbox (IGB), and the Tail rotor Gearbox (TGB). Thirty-five drive train component analyses were defined, as listed in Table 2-1. The table also shows the aircraft sensor used to acquire the component data, and the equivalent IHUMS 'channel' (i.e. analysis number) allocation. The IHUMS performs thirty-two component analyses. However, three of the components have two gears on a single shaft (the MGB combiner gear and bevel pinion, and the left and right AGB hydraulic drive 47- and 81-tooth gears). For the purposes of anomaly modelling, each of these was split into separate components, with separate anomaly models built for each of the two gears on the shaft using the different gear mesh-specific CIs.

Table 2-1 AS332L drive train components analysed

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MGB
2	1	RH high speed input shaft	MGB
1	2	Left torque shaft - fwd end	MGB
2	3	Right torque shaft - fwd end	MGB
3	4	Left torque shaft - aft end	MGB
4	5	Right torque shaft - aft end	MGB
3	6	Combiner gear	MGB
3	6	Bevel pinion	MGB
4	7	Bevel wheel and oil pump drive	MGB
7	8	1st stage sun gear	MGB
7	9	1st stage planet gear	MGB
5	10	1st epicyclic annulus fwd (RH)	MGB
6	11	1st epicyclic annulus left	MGB
7	12	1st epicyclic annulus aft (RH)	MGB
7	13	2nd stage sun gear	MGB
7	14	2nd stage planet gear	MGB
5	15	2nd epicyclic annulus fwd (RH)	MGB
6	16	2nd epicyclic annulus left	MGB
7	17	2nd epicyclic annulus aft (RH)	MGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB

Table 2-1 AS332L drive train components analysed (continued)

3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB
4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MGB

The CIs calculated by the IHUMS during each component analysis are listed in Table 2-2 (a more detailed description is presented in reference [2]). Not all of the indicators were used in the anomaly modelling. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data (the actual indicators used in the different anomaly models are listed in section 2.2.2).

Table 2-2 IHUMS Condition Indicators

Condition Indicator	Description
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation (rms)
FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev (rms)
ETA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA_CVG	Signal average convergence

2.2 Data Processing for Anomaly Detection

A detailed description of the data processing performed is presented in reference [2]. A brief summary of the three stages of the anomaly detection processing is given below.

2.2.1 Data Pre-Processing

Two types of characteristic data behaviour were modelled in the anomaly detection system. The first type was addressed by modelling between-component (of the same type) variation, and such models were referred to as 'absolute models'. Each component tends to have its own level of vibration energy. Even though a variation in vibration between components is expected, there will be a level above or below which a component's vibration appears to be so unusual that it should be flagged as anomalous. The 'absolute model' was designed to detect such cases. The second type of behaviour concerns within-component variation and such models were referred to as 'trend models'. As the name implies, these models were designed to detect anomalous trends in component data. The notion of an anomalous trend is important because trends are not uncommon in HUMS vibration data.

The pre-processing for the absolute models was minimal, and involved the application of a two stage filter to each indicator: the first stage removed unreasonable values and the second stage applied a median filter to remove up to two successive spikes in the time series data.

The objective of the trend pre-processing was to detect the relative change in an indicator by providing a measure that normalised out the between-component variation. For this, the two stage filter for the absolute model was used, and then a 'moving median difference' algorithm was applied. Following each new acquisition, the median of the time history is re-calculated and subtracted from the newly acquired value to provide a normalised value. This technique reduced the impact of early post-installation trends, and also small step changes due to maintenance, because the normalised value would gradually recover back to the median base line level. Although there was some distortion of the time-history, this simple approach worked well. Its weakness was that, while it achieved the normalisation objective, it acted as a difference calculator and could not distinguish between step changes, short duration changes, and developing trends.

2.2.2 Anomaly Models

Using the pre-processed data, anomaly models were constructed for each drive train component. The models were constructed from a training database of historical data, and adapted so that they rejected any abnormalities existing in the training data. Models were built to represent absolute data behaviour (between-component variability) and trend behaviour (within-component variability). These models were sophisticated statistical representations of the data generated from in-service experience, fusing sets of CIs (i.e. vibration features) to reduce a complex data picture into a single time history called a 'FS' trace. The FS measured the degree of abnormality in the input data and mirrored the shape of any significant data trends. It represented a 'goodness of fit' criterion, indicating how well data fitted a model of normality. Therefore the FS had a decreasing trend as data became increasingly abnormal.

With the exception of two drive train components that did not include a gear on the monitored shaft, every component had anomaly models constructed from two sub-sets of the CIs listed in Table 2.2:

- '8-indicator' models, comprising the following eight indicators: {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

- 'M6' models, comprising the following two indicators: {ESA_M6*, ESA_WEA}, (ESA_WEA is a derived indicator, calculated as the ratio ESA_SD/SIG_SD).

The indicators were split to reduce the potential for one indicator to be masked by another; the criterion for this split was derived from engineering knowledge of the indicator design and characteristic behaviour.

Using the different forms of pre-processing, 'absolute' and 'trend' models were built for each set of indicators. Therefore the drive train components were monitored by four anomaly models – '8-indicator' absolute and trend models, and 'M6' absolute and trend models. The oil cooler fan and dual bearing module were exceptions in that they were monitored by only two models – absolute and trend models comprising the following '5-indicator' set: {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

2.2.3 Alert Thresholds

A threshold was applied to the FS output from each anomaly model to alert the operator when a FS had decreased to a level at which it could be considered to be indicating the presence of anomalous data. The thresholds were determined automatically, based on a probability distribution fitted to the FSs derived from the training database, and varied from one model to the next. An alert was generated on a component whenever one or more of the individual component anomaly models went into an alert state.

2.3 Anomaly Detection System

The anomaly detection system was implemented as a web-based system, hosted on a server at GE Aviation in Southampton. The system is described in reference [3].

The web server had a simple-to-use interface, enabling the user to quickly access anomaly alerts, FS traces, HUMS CI trends, and notes documenting follow-up actions. After clicking on a chart icon in an alert record, the user was presented with the anomaly model FS traces for the component in alert (Figure 2-1). This represented the primary system output display associated with an anomaly alert.

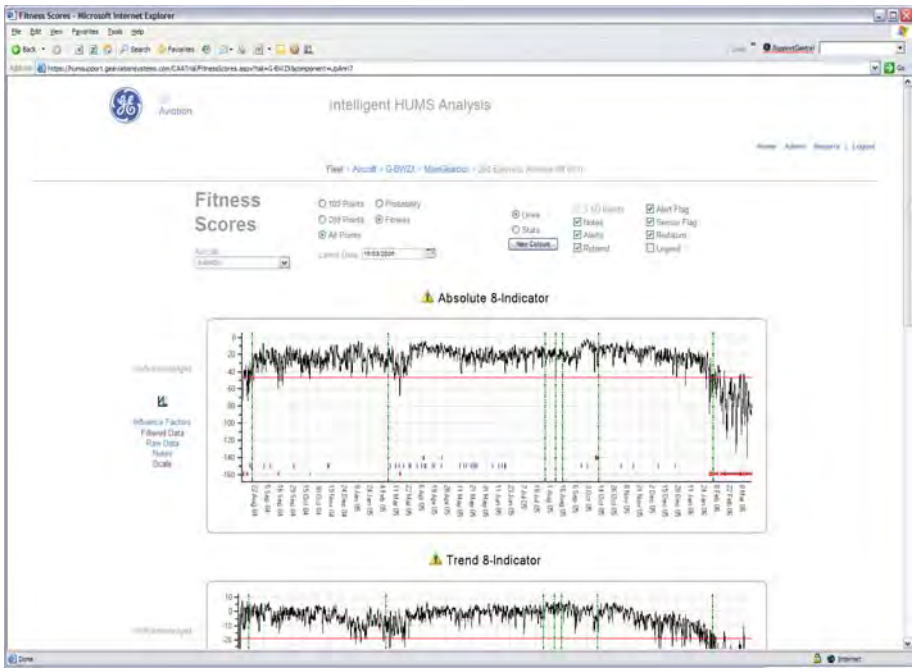

Figure 2-1 Fitness Score trace

For further investigation, the user could 'drill down' to either the filtered or pre-filtered IHUMS CIs (Figure 2-2). For those indicators used to build models, the indicator plots showed the upper and lower three standard deviation bands that had been derived from the anomaly model. These bands were different to those that would be computed directly from the fleet data because the anomaly models had been adapted to reduce the impact of outlying data.

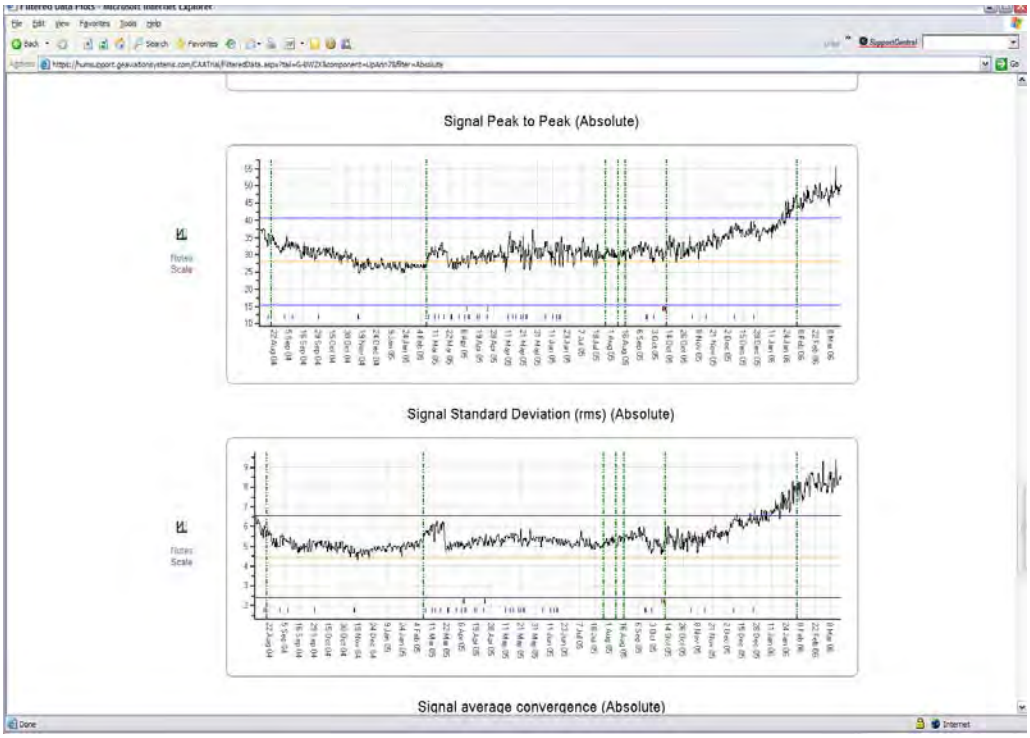

Figure 2-2 Condition Indicator display

It was also possible to view IHUMS CI data for a fleet of aircraft either as multiple trend plots or as statistical plots (Figure 2-3). Once an investigation and any associated follow-up action had been completed the user could acknowledge an alert and, if appropriate, enter a note.

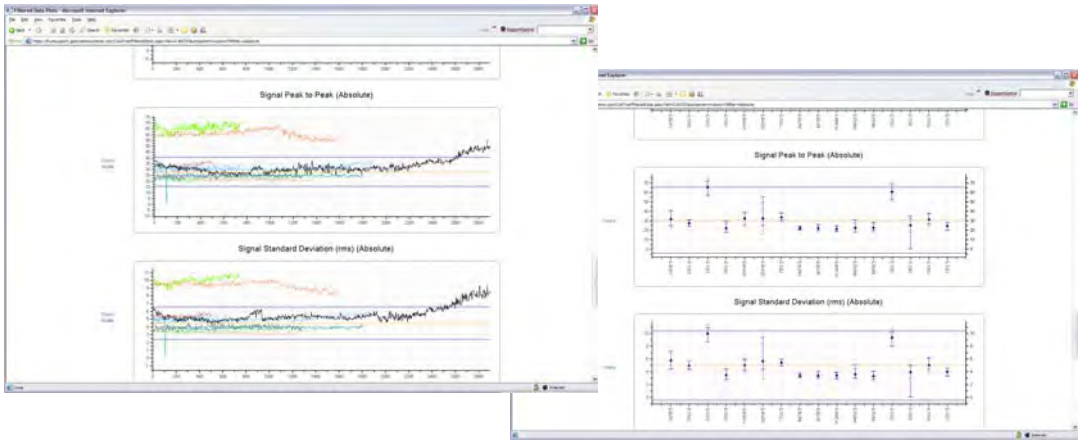

Figure 2-3 Fleet data displays

The anomaly detection system FS traces included various items of documentary information; these are shown in Figure 2-4.

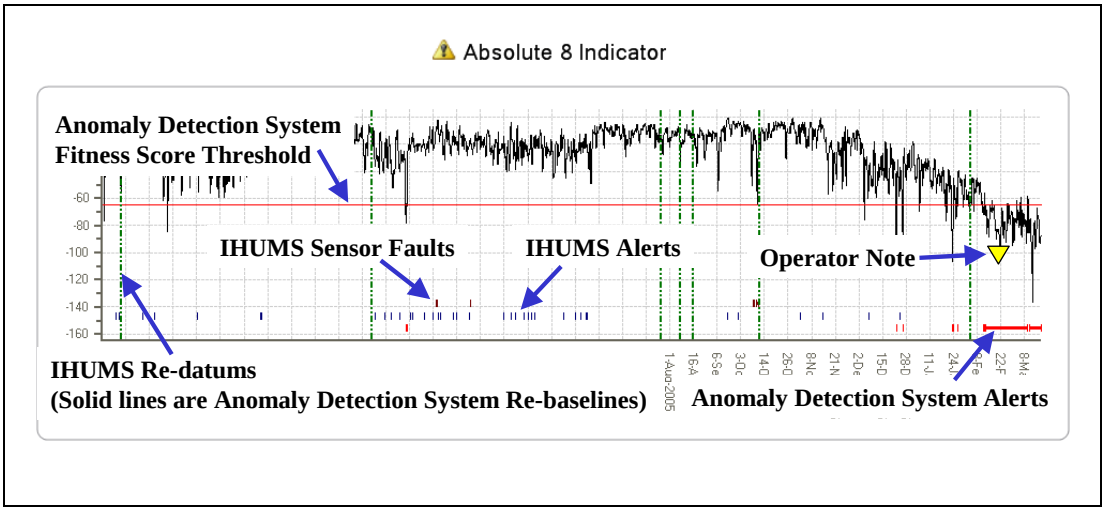

Figure 2-4 Documentary information shown on a Fitness Score trace

3 Data Characterisation

Key diagnostic and prognostic information is contained in the shape of a signal. Step changes in a signal are often the result of maintenance in which event the cause can be explained but, in other circumstances, it can denote a sudden failure, perhaps in a sensor. Monotonic trends where a signal climbs or drops over successive samples can be indicative of some gradual failure but it might also be a subtle change due to some unknown environment influence. The onset of erratic 'spikes' in a signal often indicates an instrumentation issue. The ability to detect these signal traits is important when assessing the significance of abnormal data and reasoning the cause of the anomaly.

Pre-processing implies a transformation of the signal before it is used to train a model (anomaly model in this case). The idea is to make the model sensitive to key features of a signal such as a trend. The modelling already employs two levels of pre-processing: the first is the median filter to remove singleton or paired 'spikes', and the second is a moving median difference to detect shifts in a signal. This latter pre-processing stage is basically a signal differencing calculation that represents the magnitude of changes in a signal's level as opposed to how the signal is changing (e.g. a slow developing trend or step change). The moving median difference is attempting a form of normalization as opposed to signal characterisation.

This task was originally referred to as pre-processing but, with hindsight, it is better described as signal characterisation because there is a requirement to detect shape in the FSs (i.e. processed data) as well as the CIs. During the course of this work it has been concluded that transformations on the CIs before input to the model learning should be kept to a minimum. This is because anomaly alerting that is not dependent on a potentially complex pre-processing stage (i.e. fewer assumptions about CIs made prior to modelling), would be intuitively safer. It is also more efficient to act on the single fused signal of the FS than on multiple CIs. Finally, the modelling itself is more stable because it is not subjected to continual rebuilds on different transformations of the CIs.

This assumption is sound provided that the FS mirrors any significant shape that shows in an input signal (CI). Experience from trial results shows this to be the case, but the assumption should nevertheless be treated with some caution.

Shape characterisation can be very difficult. The Phase I work to develop the anomaly modelling technique presented many new challenges, and the resource to develop comprehensive shape detection techniques did not exist. The aim of the current task was to perform a brief overview of methods that might assist with this task. There was no commitment to implement new techniques because there was no confidence that useful methods could be derived within a reasonable time and budget. In other words, other tasks described in this report took budget priority because development in these areas was more advanced and therefore more likely to be successful. Nonetheless, this task warrants continued effort because it is required for better informed diagnosis.

3.1 The Nature of CI Time Histories

The types of behaviour that would be most useful to model are shown in Figure 3-1. They include developing trends, step changes and periods of excessive noise. Detection of some shapes can initially appear simple. Detecting step changes might only require a statistical test for a change in a signal's average value. However things are rarely that simple with real world data and step changes in HUMS data are often accompanied by changes in signal variance and, while the human eye is still good at identifying these step changes, it can be difficult to derive an algorithm that recognises all of the variations that can occur.

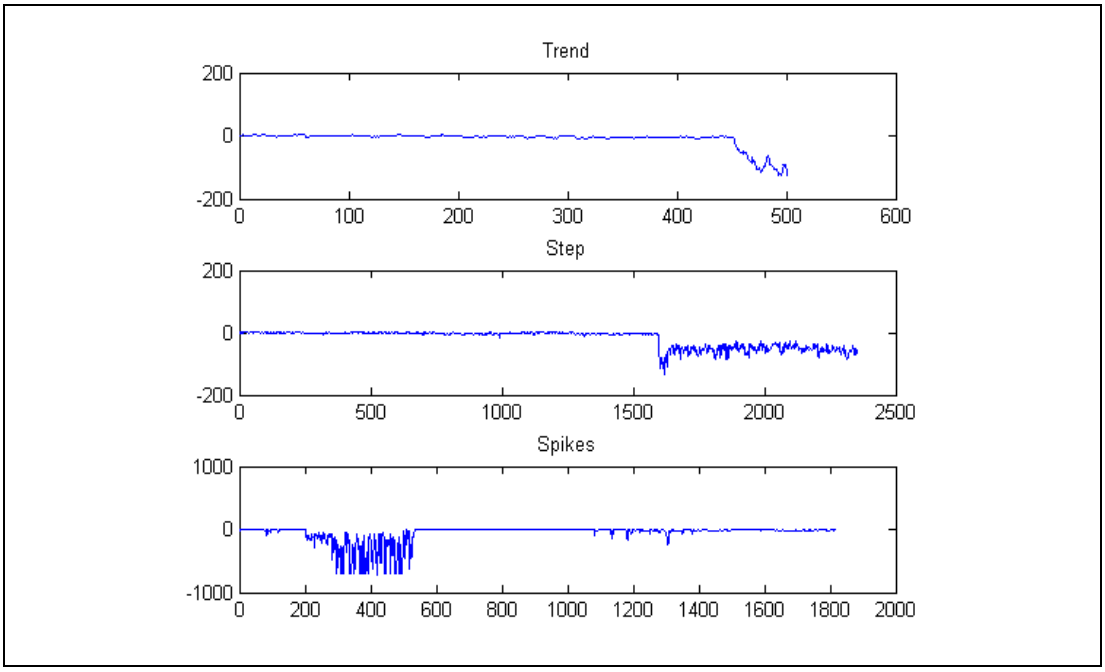

Figure 3-1 Three types of behavioural trait. The top plot shows a long period of stable normal behaviour followed by a fast developing trend. The second plot shows a step change with the introduction of increased noise. The third plot shows a signal with a period of noisy erratic behaviour.

Noise and non-stationary behaviour are two characteristics of HUMS CIs that complicate the recognition and tracking of salient features. Noise in this context refers collectively to signal variance caused by instrumentation problems and environment influences. CIs tend to drift over time meaning that their average and variance tend to vary (the non-stationary nature) throughout the life of a component fit. A CI's time history can also appear to cycle which may, in part, be due to environment influences.

A trend in a CI refers to a monotonic shift in a CI's level over a contiguous time window. The trend may be non linear but, for diagnostic or prognostic analysis, it will be sufficient in the majority of cases to represent this trend using a linear model. The challenge is detecting the trend against the noisy and non-stationary nature of a signal.

Extreme levels of noise, seen as regular large spikes, are often due to an instrumentation fault. The third plot in Figure 3-1 illustrates a signal showing a period of extreme noise. The detection and correction of such behaviour is important if the HUM system is to function correctly. Therefore any noise removal technique must smooth out small fluctuations while emphasising extreme shifts.

3.2 Modelling Different Signal Traits

A review of the literature has revealed no proven reliable methods for detecting different signal features such as trends, step changes etc. These tasks would be relatively straightforward if the signals were noise free, stationary and the nature of change simple such as clean discrete shifts (e.g. step change) or linear trends. It is the continuous variation in HUM CIs without a complete understanding of their cause that makes the task of detecting meaningful changes particularly difficult. For example, a time window on a CI will sometimes reveal cyclic type behaviour but there is often no knowledge of what is causing it. Cyclic behaviour is a common feature in many time series but it is usually understood, e.g. seasonal fluctuations in retail sales. Such trends are known as periodic because the cycles repeat and extend over long periods, but a HUMS CI might cycle over several hundred flights after which it shifts to another statistical mode. Figure 3-2 shows a HUMS CI for a bevel gear that is cycling within the normal range of the data. The non-stationary nature of the signal is clearly evident. Many time series data exhibit periodic patterns and understanding the cause of the periodic influence helps in modelling the series and making forecasts. The initial period of behaviour of the signal in Figure 3-2 shows a rising trend and, with no knowledge to suggest otherwise, the assumption would be that the trend will continue whereas it reverses at around 200 hours. Although this signal is atypical of HUMS CIs it is not uncommon to see short term trends that either plateau or reverse. Trend detection therefore is of little value unless its statistical significance can be assessed. Judging significance is the job of anomaly detection.

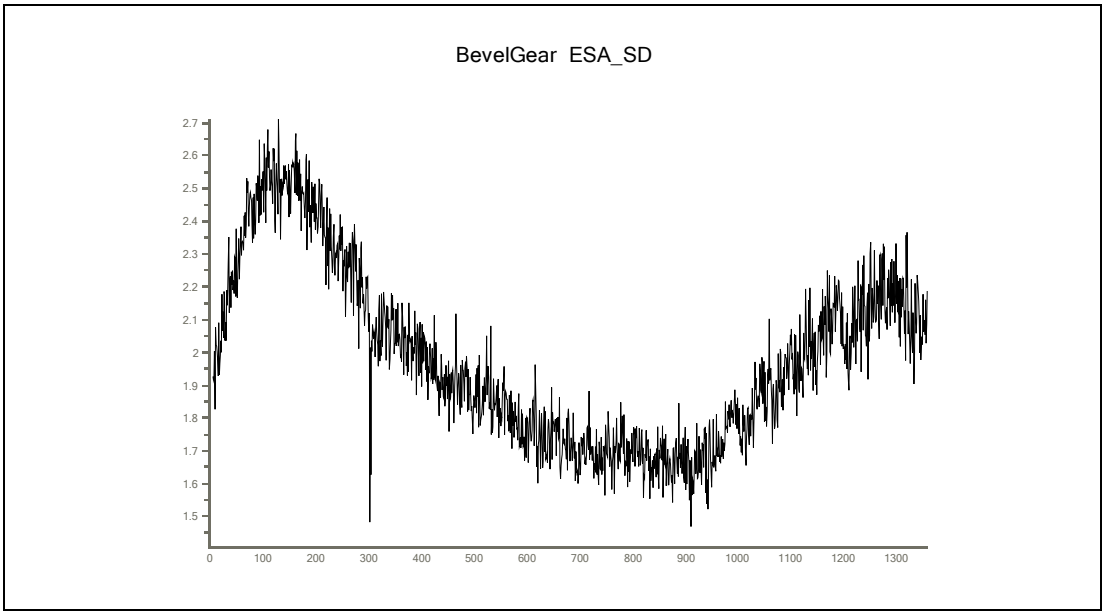

Figure 3-2 Time history for a single gearbox component. The non-stationary nature of the signal is evident but it cycles within the normal range of the fleet.

Given that there are many potential influences on the behaviour of CIs that give rise to windows (periods) of different behaviour, it is difficult (if not impossible) to assume a single model (representation) for a signal over its complete time history. Rather, it seems more pragmatic to segment a signal into what are homogeneous periods of behaviour (rising or falling trends, stable signal, etc).

Segmenting a signal requires the ability to detect the onset of a meaningful change in the signal level but changes can be hard to detect due to the high level of variance found in typical CIs. There are two initial processing stages that are seen as holding the key to successful shape recognition: the first is a smoothing function to remove low level noise to emphasise salient features, and the second is a function to

segment a signal into homogeneous periods of behaviour. The nature of a signal within a segment can then be represented using a simple linear model. The remainder of this section explores some potential candidate methods to implement these functions.

Many signal processing and pattern matching techniques are candidates for some aspect of signal characterisation. Techniques can be drawn from a broad spectrum of disciplines that include process control, time series analysis and signal processing such as wavelets and filtering. Techniques from these disciplines can often be applied to the same problem. For example, data smoothing can involve simple exponential weighting or the application of wavelet transforms (see paragraph 3.2.1 for more information).

3.2.1 **Removing Noise**

Care needs to be taken when applying some transformation function to change the nature of a signal. A clear understanding of a function's effect is required to ensure that the result has only desirable properties. For instance, smoothing may be inappropriate if fitting a model to data (such as a mixture model, see reference [2]) but, in the context discussed here, smoothing is to be applied in order to make salient features more visible.

There are many functions for smoothing a signal. A simple technique is to apply a moving average where a combination of previous, current and following samples are averaged. A moving average is simply a summation over a window of samples (e.g. eight points) where each sample is first multiplied by a constant (1 over the number of samples). This type of transformation is captured by a more general class of function known as the digital filter ([4]). An online digital filter only has knowledge of the current and previous samples and transforms the current sample as a weighted sum of the current sample, previous samples and outputs from earlier transformed samples (feedback). If the filter uses feedback it is called an infinite impulse response filter (IIR) and without feedback it is called a finite response filter (FIR). The online requirement can be relaxed if the task requires a retrospective analysis of the signal (e.g. analysing the FS), in which case filtering can use following samples (to the right of the current sample). In any event, filtering is often a simple process involving a summation of weighted samples. Coefficients determine the weighting of each term in the summation and therefore determine the response of the filter. Defining or learning the coefficients is not trivial but, once designed, the application of a filter is straightforward.

Useful application of a filter requires an understanding of its properties. For example, a moving average filter tends to reduce peaks in the data and squashes them out, so careful consideration has to be given before applying this filter to HUMS CIs because the significance of outlying data can be suppressed. Another type of filter that is generally well suited to data smoothing is the Savitzky-Golay filter. This filter was described in a 1964 paper [5] published in the journal of Analytical Chemistry, and the paper's wide citation is testament to its perceived value. This filter applies a moving window in a manner similar to a moving average but it uses a polynomial to approximate the data. For each point, a least squares polynomial fit is performed using all data points within the window and the point is substituted with the value of the polynomial at that point. The computation involved sounds horrendous but in fact it is very quick because the polynomial coefficients can be computed independently of the actual data. The fitting function needs as input the number of previous points, the number of following points and the order of the polynomial.

Figure 3-3 shows a CI and the effect of smoothing with a moving average filter, FIR filter and Savitsky-Golay filter. It should be emphasised that this application is for illustration and there has been no attempt to optimise the filter designs for application to CI data. The moving average has squashed the step change that occurs near to sample 140. This squashing extends the duration of the step change and reduces its amplitude. The FIR filter provides much improvement over the moving average and the Savitsky-Golay filter is even closer to the original signal.

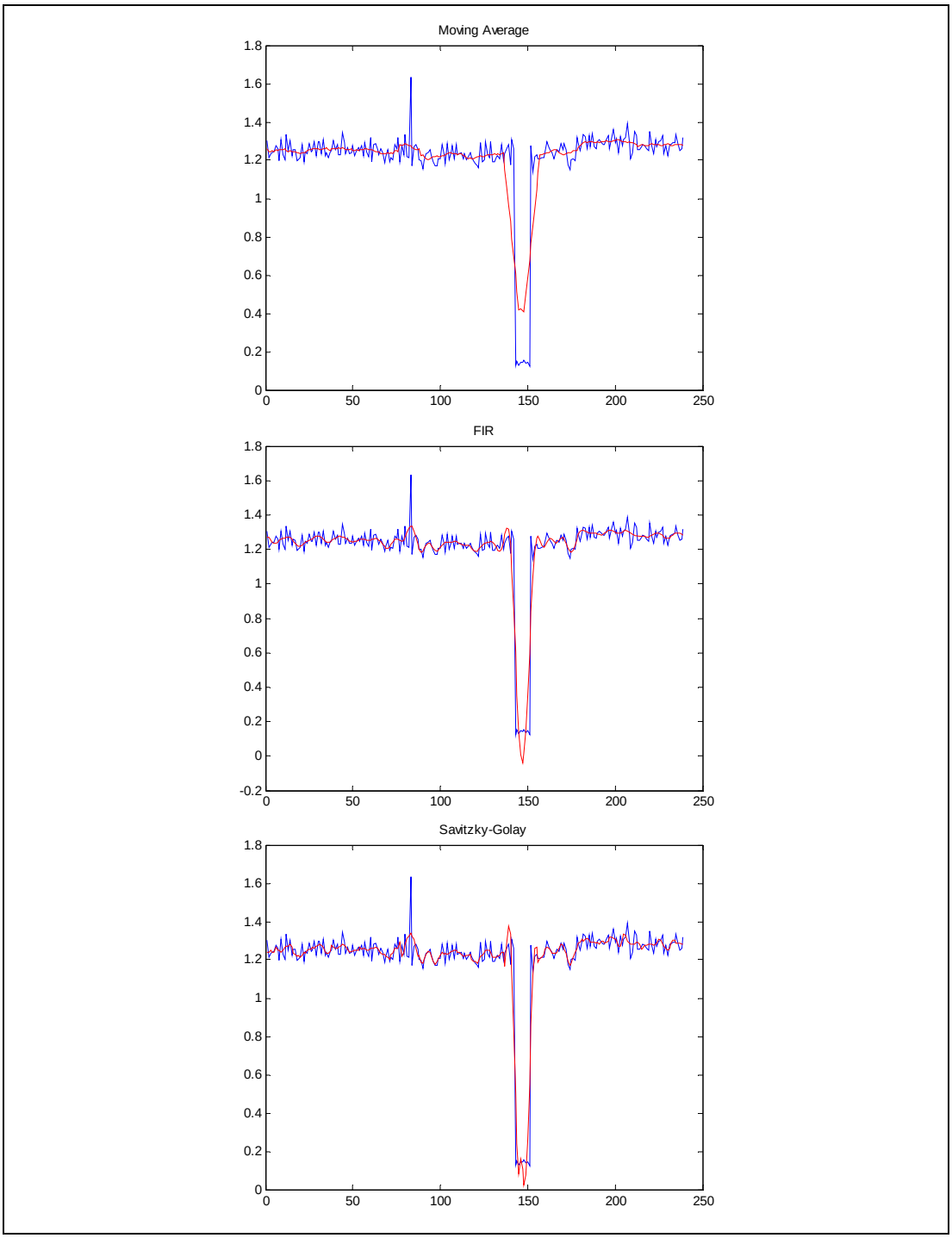

Figure 3-3 The effect of applying different filters to a signal that has a step change lasting for several samples. It can be seen that the moving average filter provides a poor representation of the step reducing its amplitude and extending its duration. The FIR filter does a reasonable job and there is some further improvement with the Savitzky-Golay filter.

Another technique for smoothing is the Discrete Wavelet Transform (DWT). This transform is similar to the Discrete Fourier Transform (DFT) in that both transforms are linear operations that utilise a set of basis functions. Like the DFT, the DWT represents the original signal by a sum of weighted basis functions. Coefficients denote the weights. The basis functions in the DFT are sinusoids of different frequencies. Unlike the DFT the basis functions in the DWT are not restricted to a single set of basis functions – these basis functions are generically referred to as wavelets. Two variables are used to change the scale and location of the chosen (analysing) wavelet. Scale determines resolution and location determines its position. So the original signal is represented by a set of functions that vary in scale and position. This representation provides localization in space in addition to frequency and can be usefully exploited for compression, feature detection and noise reduction.

The DWT is computed by simultaneously passing the signal through low pass and high pass filters – these filters are half band filters and have to be related to each other (known as a quadrature mirror). This filtering is applied recursively to the low pass outputs until the desired level of decomposition is reached. Since the low pass filtered output has a highest frequency which is half the highest frequency of the input signal it can be down sampled by a factor of 2 (according to the Nyquist rule). The signal now has half the number of points and the scale is now doubled. The end result is a decomposition of the signal into a smoothed approximation and detailed representation. Suppose the original signal has 1024 data points and a frequency band of 0 to 2π rad/s. The output of the high pass filter contains 512 points but only spans π to 2π rad/s. So it has half the time resolution but double the frequency resolution. These samples make up the first level of decomposition. The low pass filter also has 512 samples but spans 0 to π rad/s. The successive filtering of the low pass signal provides the second level of decomposition. This procedure iterates until two samples remain. The DWT of the original signal is obtained by concatenating all coefficients starting from the last level of decomposition. The DWT has the same number of coefficients as points in the original signal. The more dominant frequencies in the original signal have high coefficients in the region of the DWT signal that includes those frequencies (hence the time localization). The time localization depends on the frequencies where most information resides, and is better when this information resides in the high frequencies (since these frequencies are represented by more samples). The DWT will give good time resolution for high frequencies and good frequency resolution for low frequencies.

An inverse transform can translate these DWT coefficients back to the original signal. A signal can be 'cleaned' or smoothed by setting those coefficients that fall below a specified threshold to zero before performing the inverse transform. Figure 3-4 illustrates the application of smoothing using the DWT. In this example the DWT offers no advantage over the FIR filter and is less optimal than the Savitzky-Golay filter. However there has been no attempt to optimise any of these filters for the signal being analysed. Further information on the application of wavelets to noise reduction can be found in [6] and [7].

No attempt has been made to evaluate an optimal filter for application to CIs. It is possible that different filters are required to emphasise different features and optimum selection would require a detailed evaluation over many CI time histories. Because smoothing is only one part in a process to identify significant features, candidate methods need to be considered in the context of segmentation.

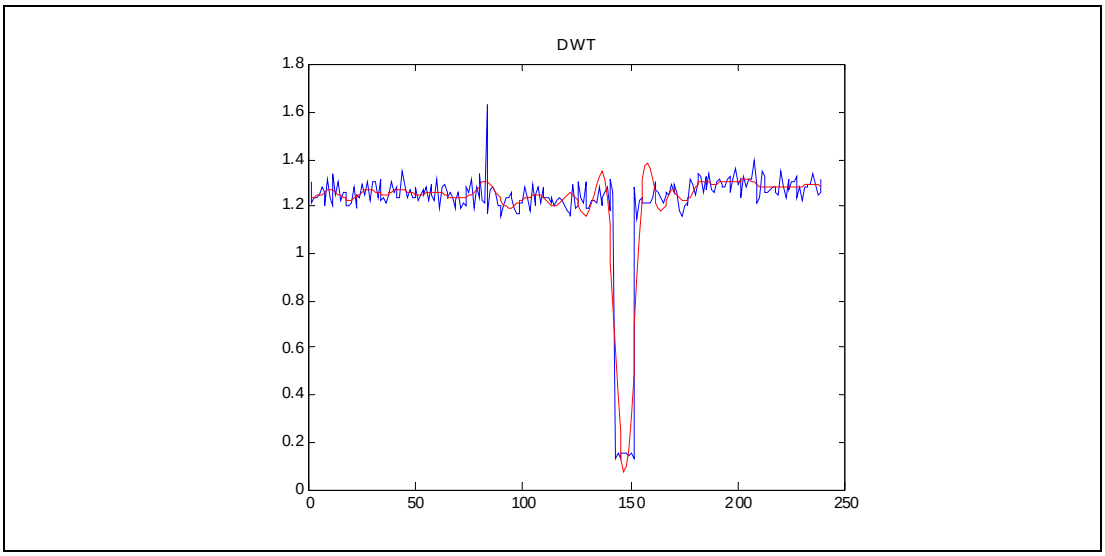

Figure 3-4 Smoothing using a discrete wavelet transform

3.2.2 Segmenting

The ability to segment a signal (such as a CI time history) into homogeneous regions of behaviour is seen as a fundamental requirement. Once segmented, the underlying nature of a segment (e.g. stable, trending, excessive noise) can be detected more easily. Segmentation however is a non-trivial task and is heavily influenced by scale and time. Plotting a collection of independent signals on the same chart and varying the scale (y-axis) and time (x-axis) illustrates how perception of a signal can change. For example, a signal with 100 points that displays a clear trend over a ten-point window could appear more like a step change when re-scaled and viewed alongside signals containing two- or three-thousand samples. Similarly, a short section of an individual signal might exhibit a trend that appears to be contained within the noise when viewed against the whole time history of the signal. Detection therefore requires statistical context (to give a sense of scale) and rules to define characteristic features (such as a step change).

Several potential approaches to segmenting a signal will be explored in this section. No attempt is made to characterise a signal, only to detect when the nature of the signal has changed.

There are a number of statistical techniques employed in process control monitoring to detect when there is a significant shift in a monitored feature such as the mean or variance. Shewart charts provide a well known and simple approach to process control monitoring. Shewart was working at Bell Labs in the 1920s when he reasoned that there was a variation in a process (e.g. the manufacture of piston rings) that could be considered natural or within control. An output of the process, such as the diameter of a piston ring, is plotted on a chart which is overlaid with mean, upper and lower control limits. Any measurement above and below the lower and upper control limits respectively is considered to be within control. Points outside of the limits are considered statistically unlikely and such points are indicative of an out of control process.

Control charts are designed to detect significant shifts in data and, as such, have potential application to tasks discussed here. Indeed such ideas form the basis of threshold setting in traditional HUM systems. However, for general change detection there is a requirement to detect shifts when data are still within control limits.

CUSUM charting, introduced by Page [8], is one type of process control technique that is capable of detecting more subtle changes in something like the process mean. CUSUM charts plot the cumulative sum of the deviation of each sample from some reference value. Limits are not parallel to the x axis as with Shewart charts but defined via what is called the V-mask which is a V rotated anticlockwise by 90 degrees with its origin placed over the last point. However the feature of the CUSUM technique that is of interest here is not the actual value but the shape of the chart. If a signal spends a contiguous period of time below its overall mean (reference value) the CUSM chart will show a negative slope and positive slope if there is a consistent period above the mean. This is illustrated in Figure 3-5 and shows that a contiguous period above the mean gets reflected in a monotonic positive slope in the CUSUM. The CUSUM has been applied to a typical HUMS CI in Figure 3-6. The reference value is the mean over all the data. The tendency of the data to drift over time is evident in the CUSUM plot and there are essentially two modes split by the mean value with the first mode having a positive slope and the second a negative slope.

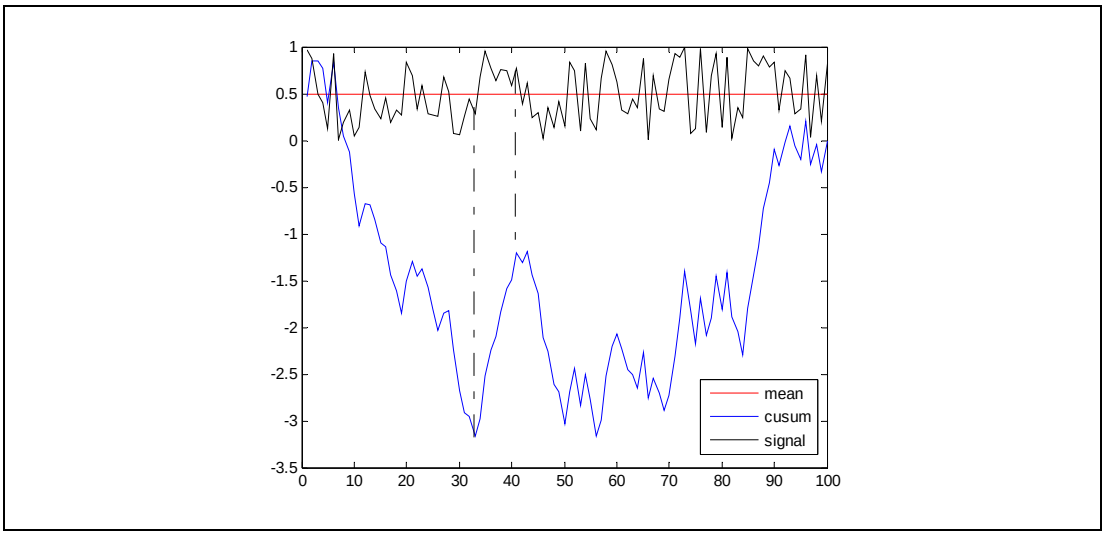

Figure 3-5 Effect of CUSUM on a uniform random signal. The original signal is a uniform random signal and is shown in black. The mean value, which is the reference value, for the complete time history is shown in red. The CUSUM is plotted in blue. The short section of signal between the dashed lines illustrates what the CUSUM does – the data points in this section are all above the mean and the CUSUM responds by having a monotonic positive slope.

The CUSUM technique does have potential application to the signal segmentation sought here. The segmentation needs to be automated and so the CUSUM would not be employed in a traditional charting manner. However it would be relatively straightforward to define rules that will detect long duration shifts in the data. For example, a bimodal shift should display a CUSUM with a maxima and monotonic trends with opposite signs either side. This is illustrated in Figures 3-6 and 3-7 and the peak in the CUSUM corresponds to the mode change in the CI.

Time series modelling (see [9]) refers to a collection of statistical techniques that are used to understand the nature of data and to make predictions. Data are assumed stationary or made stationary (for example by differencing adjacent samples in the series or taking logs). An autoregressive (AR) process assumes that a sample is a linear combination of previous samples plus a random error. Independent of the AR process, it could also be assumed that a sample might be affected by errors in previous samples and this describes a moving average (MA) process. Both the AR and MA processes are modelled as weighted sums of the current and previous

samples. The order of the AR or MA model determines how many previous samples are used in modelling the data. An ARMA model is a generalization that combines AR and MA models. Autoregressive Integrated Moving Average (ARIMA) models are a further generalization that includes transformations to make the data stationary. Differencing can be used for de-trending the data to remove non-stationary effects, due to seasonal influences for example. Variance can be made stationary by taking logs.

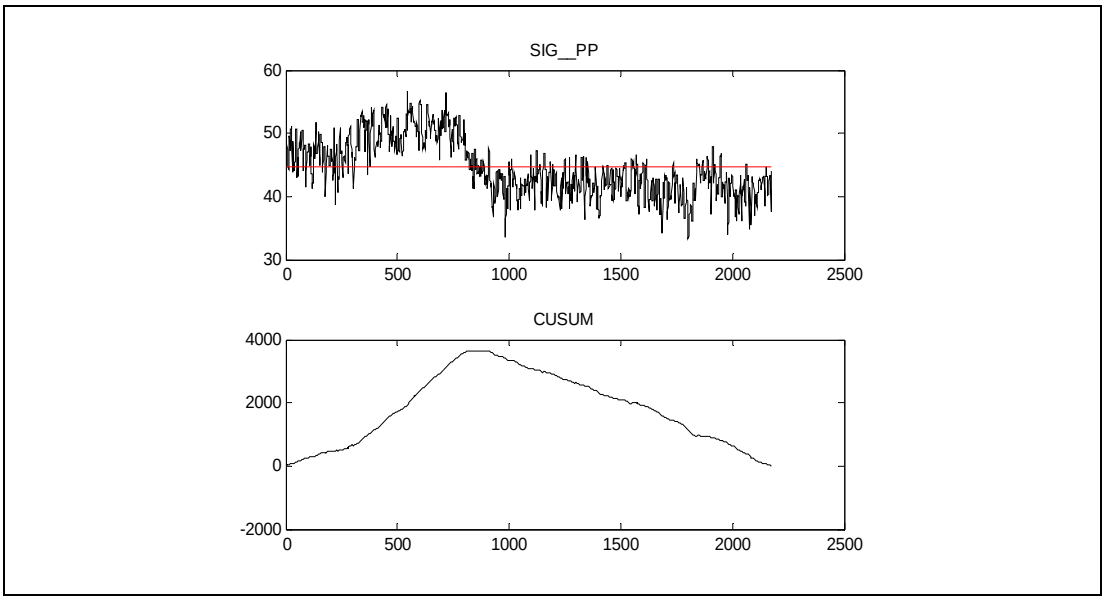

Figure 3-6 Effect of a bimodal signal on CUSUM

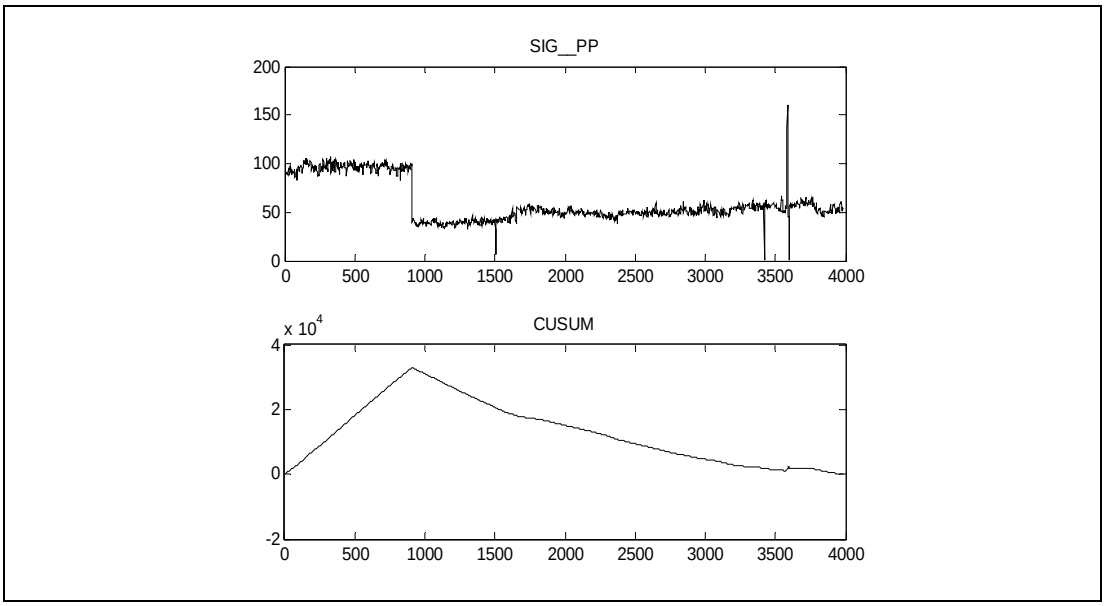

Figure 3-7 Effect of a step change on CUSUM. The step change in the CI results in a CUSUM with a single peak and almost clean monotonic trends either side of the peak.

In Figure 3-8 the first 15 points were used to fit an autoregressive (AR) model by removing the trend then using AR to model the signal's stochastic nature. The dotted line shows the predicted trend over successive points. The following 12 points provide a good forecast of the trend after which the nature of the trend changes and the forecast breaks down. This example illustrates that a short duration pattern can be reasonably represented using a very simple technique.

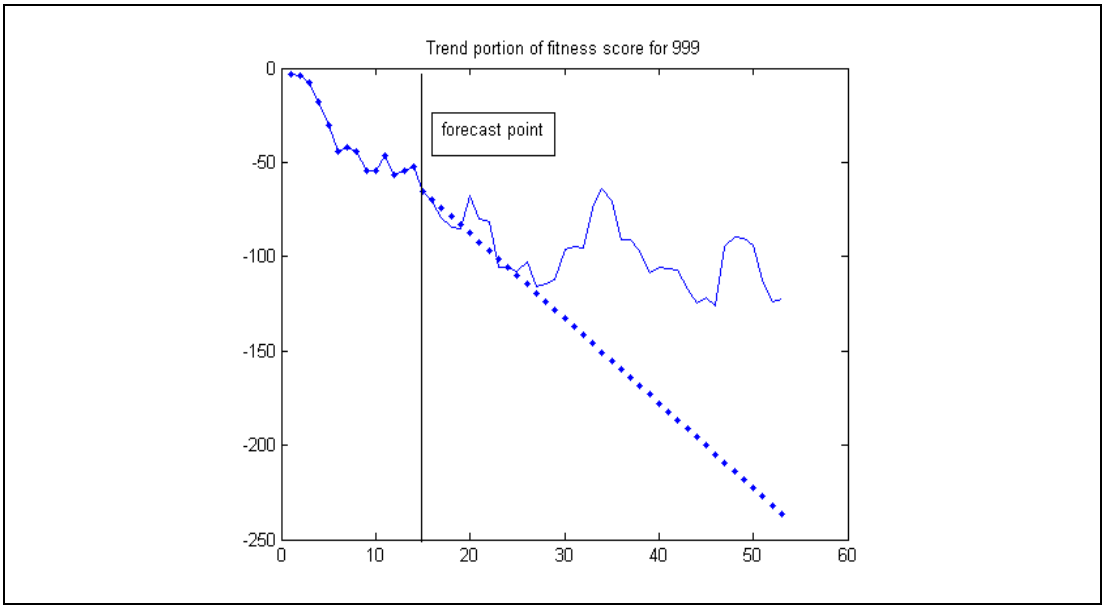

Figure 3-8 Example application of an AR model. A simple time series model fitted to the first 15 points in the time history produces a model for predicting the trend in the following 12 acquisitions.

As seen above time series models can be used to predict future values. If applied over a short time window, the assumption is that samples within the window will look similar. If there is a sudden shift in the data there should be a larger error in the predicted value which suggests a potential method for detecting meaningful shifts in the data. The method would simply plot the residuals (difference between actual and predicted values) and any shift will show itself as a spike. This approach has been tried using something similar to an autoregressive model where the previous *k* points are used to predict the current sample. Its application to a non-stationary signal that has no sudden shift is shown in Figure 3-9. The residuals plot is quite spiky but no residual stands out above the background noise. The signal is the one used in Figure 3-6 where the CUSUM technique reacted to the two models. The regression technique has not identified the shift between the two modes because the shift occurs slowly over several points. Contrast this with the example in Figure 3-10. The sudden step change produces a clear spike. It can also be seen that real spikes in the actual signal produce a similar response in the residual plot but also introduce some additional spikes. The additional (or spurious) spikes in the residuals are due to increased prediction error caused by successive spikes in the original signal – for example, due to the scale of Figure 3-10 it is not easy to see that there are in fact two drop out spikes at around 1500. These two dropouts will lead to a low value being predicted for the next sample and this introduces a large positive residual because the next sample has returned to the normal mode for the data (the normal mode following the step change).

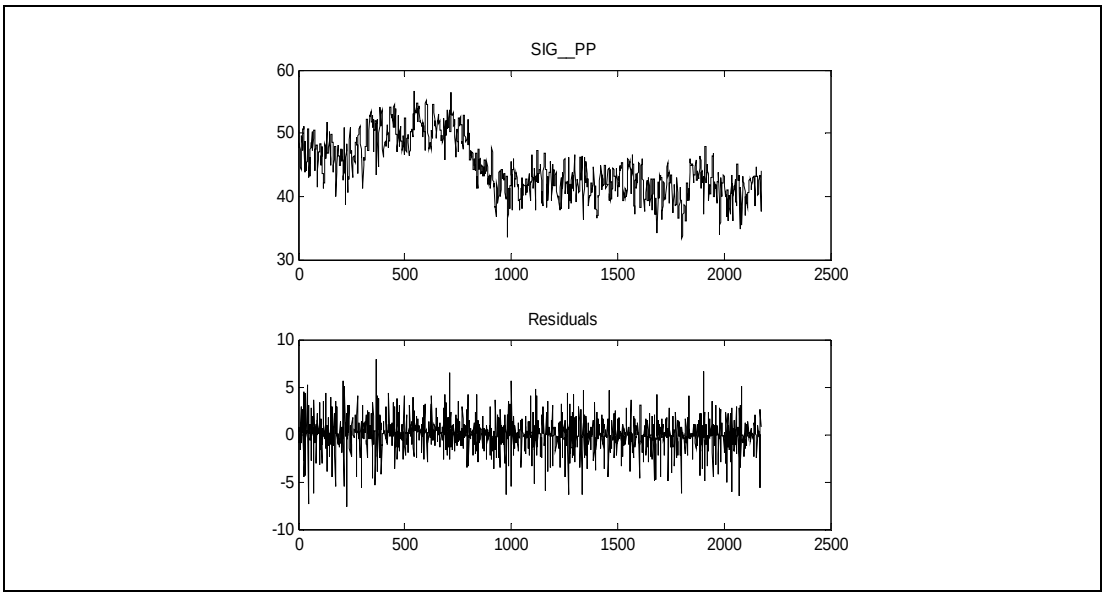

Figure 3-9 Application of AR to a non-stationary signal with no sudden shift. A simple AR type of model has been used to predict the next sample from the preceding four samples and the residuals calculated by subtracted the predicted from the actual values. The residuals show no evidence of a sudden change in the data.

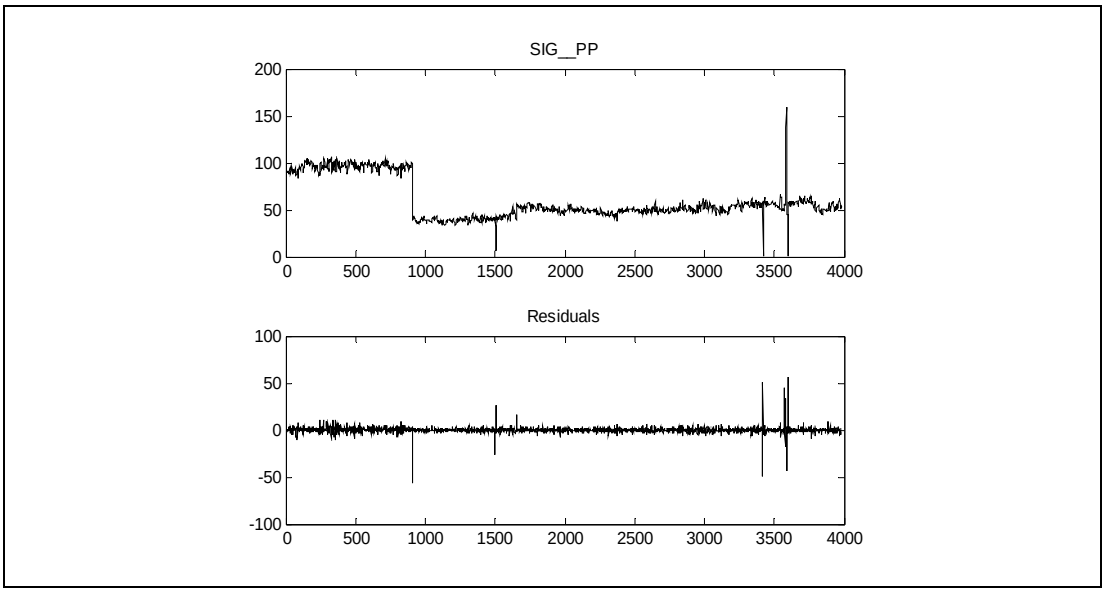

Figure 3-10 Application of AR to a signal with a step change. The residual plot is derived in the same manner as that in Figure 3-9. Sudden changes in the signal are evident in the residuals.

A further example is shown in Figure 3-11 where in this case the data actually trends down over seven samples. This example is the familiar trend FS for the bevel pinion crack case. The trend looks similar to a step change but the plot is deceptive due to the scale. The trend develops quickly and most of the drop is contained within four samples. In this example the residuals have been calculated slightly differently to the earlier examples in that the predictions are based on six samples but an offset (lag) has been introduced to delay the predictor samples by five points – in other words, the immediate preceding five samples are not used but the six samples prior to that.

The residuals respond over seven samples similar to the original signal. However the residual plot allows a simple threshold to be applied that, when exceeded, will indicate a significant shift in the data. Once these shifts have been detected their nature can be investigated by applying rules to detect the nature of the change.

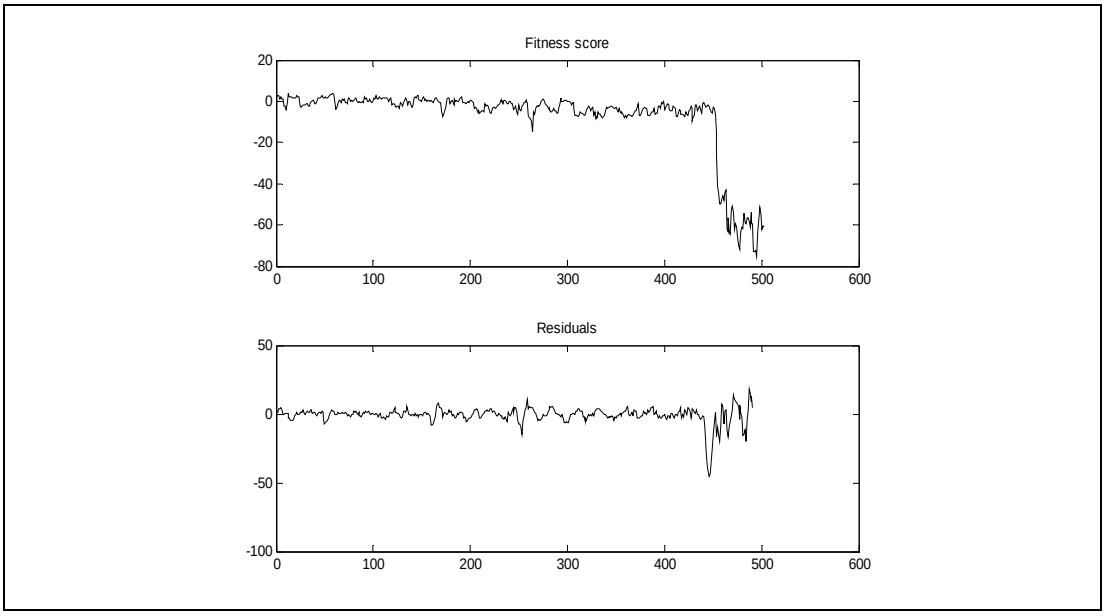

Figure 3-11 Application of AR to a signal with a fast trend. The residual plot uses six previous samples for predictions but an offset of five is introduced so that predictor samples start 11 samples before the current sample. The Fitness Score develops a fast dropping trend over seven samples. This shift in the data could be detected from the residual plot using a simple thresholding technique.

The concept of calculating residuals and applying a threshold to detect significant shifts shows real promise. It is a simple idea based on a simple expectation that neighbouring samples will look similar most of the time. The method presented here would need to be extended to remove the introduction of spurious spikes but this should not prove difficult.

The application of mixture modelling (upon which the anomaly modelling is built, see reference [2]) has been used to explore segmenting time histories. Mixture modelling is designed to identify different statistical modes in the data. However, pattern recognition techniques are rarely perfect and usually make some use of heuristic rule(s) to make decisions. Such a decision for mixture modelling is how many clusters to use. There are various heuristics designed to help answer this question, and one such metric is the BIC described in reference [2].

Figure 3-12 shows the effect of applying a mixture model to segment a FS. In this example the model chose two clusters and the data were cleanly segmented between the steady behaviour and the negative developing trend. Segmentation was relatively easy in this case.

The FS threshold for a model could also play a role in segmentation. For example, the threshold for the model in Figure 3-12 would separate most of the trend from the rest of the signal.

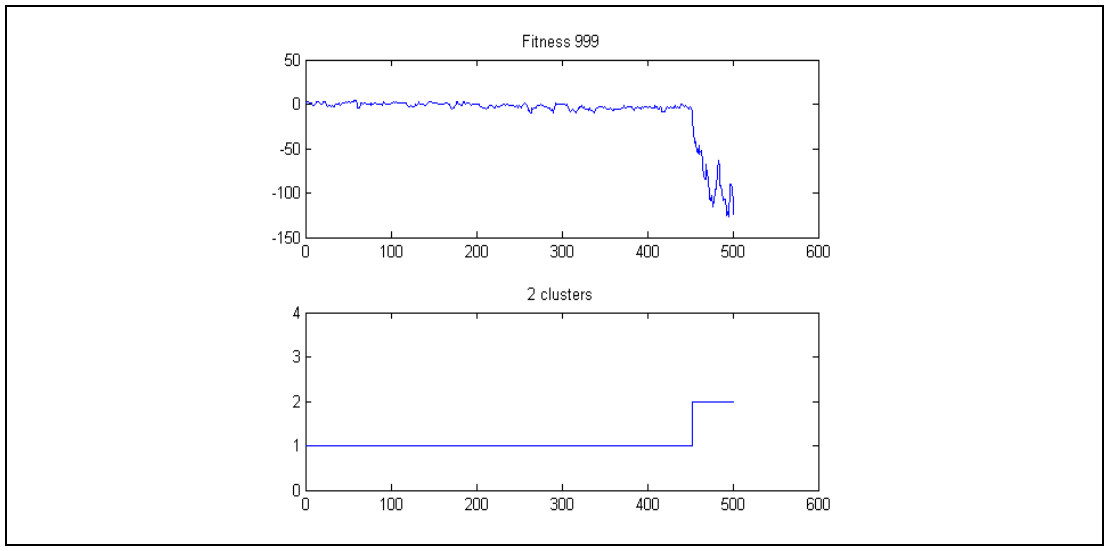

Figure 3-12 Application of a mixture model to segment a signal. A mixture model is used to successfully segment the signal in the upper plot into its two periods of steady behaviour and rapidly decreasing trend.

Instrumentation faults will often produce a highly erratic signal where CIs display a high level of variance oscillating between very different energy levels. The variation usually exceeds the level of noise in the original signal. Figure 3-13 shows the FSs for a component generated by a model of the bevel gear. Visually, this signal appears to have three periods of distinct behaviour – a steady, albeit random, background signal with a period of intense noise starting around 200 hours and lasting some 300 hours. The highly variable nature of the signal masks the two fundamental types of behaviour when a mixture model is applied. The lower left hand plot in Figure 3-13 shows the cluster assignment for a two cluster model. It has failed to cleanly identify the two types of behaviour. The original FS has been filtered in the upper right hand plot and the two cluster model re-generated in the lower plot. The segmentation is still not clean but it is an improvement on the original segmentation.

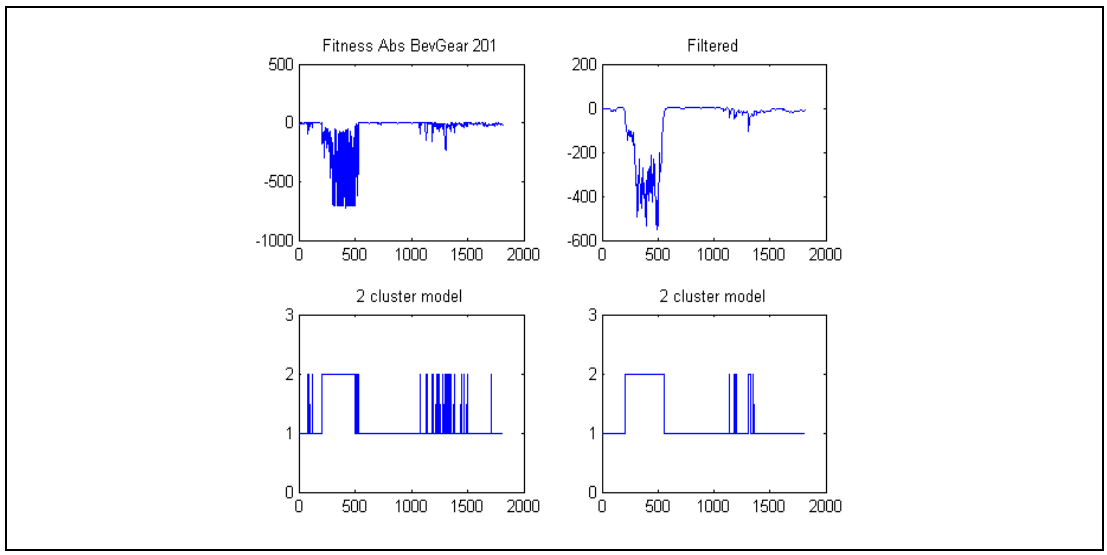

Figure 3-13 Application of a mixture model to segment a highly noisy signal. The two plots on the left show the result of applying a two cluster mixture model to segment a highly noisy signal. The plots on the right show the result of clustering after applying a filter to suppress some of the more erratic behaviour. The result is not perfect but the output is much cleaner.

Initially, mixture modelling seemed a good candidate for segmentation because it is designed to find the natural density modes in data. However, it can be computationally quite expensive and can blur different segments because there is no notion of time localization. It is possible to overcome this drawback but the effort required makes it a less attractive option.

Consideration was given to describing a signal using symbols. These symbols would be concatenated into strings to describe the nature of any trend. But this soon becomes a complicated task. First the signal needs to be discretized (or binned) and this in itself is a challenge. There are different discretization methods, each method suited to different signal traits. To optimise an alphabet for the symbols, there would need to be a learning procedure that could analyse a signal to recognise associations between signal patterns and symbol descriptions.

3.3 The Way Forward

This research task has resulted in no defined method(s) for characterising a signal. It was appreciated from the start that this was the likely outcome. However the study has proved useful for deciding on a way forward.

Shape in a signal can be confused or hidden by excessive noise. Smoothing is required to emphasise salient features of a signal. A number of filtering techniques have been explored. The selection of an appropriate filter requires more exhaustive evaluation and there has not been the resource on the current contract to do this. It may prove necessary to employ more than one type of filter, perhaps targeted at detecting different signal traits.

Segmentation is also seen as a fundamental requirement. Throughout its life a component can exhibit different levels and types of behaviour. Fitting a single model to a complete time history that exhibits different types of behaviour is difficult to achieve reliably and would make the detection of shape characteristics very complicated. The proposal is to segment a signal, whether a CI or a FS, into sections of homogeneous behaviour. There are a number of candidate segmentation methods and, as with smoothing, it is likely to require more than one method.

The type of techniques explored in Section 3.2.2 will mark sections of a signal where there has been a change. At this point the significance of the change or its nature is not determined. Rules are then applied to detect if the change is spurious and to give an indication of its nature (e.g. step, erratic spiking). End points of a section are then defined and a model is fitted to describe the nature of that section of signal. It is envisaged that these models can be assumed linear regression fits with a level of noise – a linear Gaussian model. The model will indicate the severity of any trend.

Segmentation is seen as the most challenging task. A combination (and maybe variation) of the techniques explored here is expected to provide a reasonable first attempt but, along with methods for smoothing, there needs to be an opportunity to perform a detailed evaluation.

3.4 Summary

The task results fell short of any implementation due to the effort required to research and decide a sensible way forward. There are no off-the-shelf solutions for characterising signal shape. Recognising step changes and spikes would seem a somewhat trivial task, but the nature of HUMS data means that even simple patterns can be masked by random erratic behaviour due to instrumentation issues. For example, maintenance can result in a step change in a signal but if a sensor is also disturbed the step change can be accompanied by spikes.

There is confidence that, with additional effort, a first generation implementation for characterising signal shape could be achieved. It may not be perfect but would provide useful diagnostic information and is an essential stage in achieving further diagnostic automation.

4 Model Tuning and Re-modelling

The construction of anomaly models is largely an automated task. The input CIs are initially chosen using engineering knowledge, some modelling parameters are specified, the data source is defined (i.e. which gearboxes to use in training), then the automated build process is executed. Models are built for 35 component types covering some three million acquisition points.

In most legacy health monitoring systems there will be a quantity of data that are anomalous because problems (either instrumentation or component faults) have gone unrecognised due to incomplete knowledge (features in the data that were unknown and therefore not detectable and only become apparent through more advanced analysis). The common procedure is for data to be scrutinised only when an alert has been flagged. There is no regular data mining to look for inconsistencies or unusual patterns.

The first six-month trial, reported in reference [3], highlighted a significant number of HUMS instrumentation issues which had previously gone unrecognised. The proportion of acquisitions polluted by instrumentation problems varied between components. The failure to recognise instrumentation problems meant that they could persist for long periods of time. While the anomaly model learning handles a percentage of bad data, there comes a point at which the quantity of anomalous data is so large that the model considers the data to be normal.

It is clear therefore that the performance of the anomaly modelling is dependent not only on the underlying technology but also on the quality of data used for training. If, for example, there are data relating to known faults or instrumentation problems, these data should not be used for training otherwise the model learning process has to commit resource to identifying data that are already known to be anomalous. Also prior anomalies can lend support to other anomalies in the data, thus making their behaviour appear normal.

A quantity of bad data is to be expected in the historical database. The distribution of this bad data can vary by component because some components can be harder to monitor than others, and there can also be issues with particular elements of the HUMS instrumentation. The anomaly modelling technique targets anomalies to be removed on a purely statistical basis. The technique will allow the number of anomalies to be removed to vary by component, but it has no knowledge of the criticality of components or the relative importance of CIs for different components (e.g. FSA_SO1 for shafts). So there needs to be a facility to tune models based on experience. The tuning can affect model sensitivity and/or the alert rate. Model sensitivity refers to the type of pattern that the model will consider to be anomalous. Occasionally there will be a need to re-model data due to a CI masking effect where the level of, or changes in, some CIs are not seen as significant when analysed with other more noisy CIs. For example, FSA_SO1 and FSA_SON are important for monitoring shafts, but anomalies in these CIs can be masked if other CIs are generating more anomalies. In this type of situation it is sometimes necessary to build additional models using subsets of the primary modelling CIs.

4.1 Enhanced Modelling

A detailed statistical comparison of alert rates between the IHUMS and anomaly detection system, categorised by model type and component type, was given in reference [3]. The overall alert rates were comparable but, as expected, the alerts were not necessarily synchronised because the two systems can respond to different features. One finding was the lack of sensitivity of the IHUMS to many

instrumentation problems. Because these problems went undetected, bad data could be acquired over many hours of flying. This raised two major concerns for anomaly modelling. First, although the anomaly modelling technique has a unique way of handling bad (anomalous) data in the training set, there comes a point where the quantity of bad data is sufficiently large for it to be classed as normal. Second, the training data was a random selection of component fits (gearbox / aircraft combinations) and there was a risk that the training could be biased by the fact that a large percentage of the bad data might be associated with a small number of component fits.

The effect of this can be seen in Figure 4-1. A bar chart shows the number of FS threshold exceedances categorised by; training, current (currently flying), and all (training and current and some additional historical data not selected for training), for components in the MGB. It is clear that the current fleet of components (which were not used for training) are generating many more exceedances. However, the results are biased by problems that have gone unrecognised and are not being addressed. The bias in this example is showing that the current fleet contains more anomalies compared to the component fits used for training. For each of the two of the component types shown in the lower plot of Figure 4-1, a single component fit (i.e. aircraft / gearbox component combination) has been removed from the data set. This has had a dramatic effect on the relative threshold exceedances between the training, current and 'all' data sets, showing that much of the bias can be contained in a single component fit.

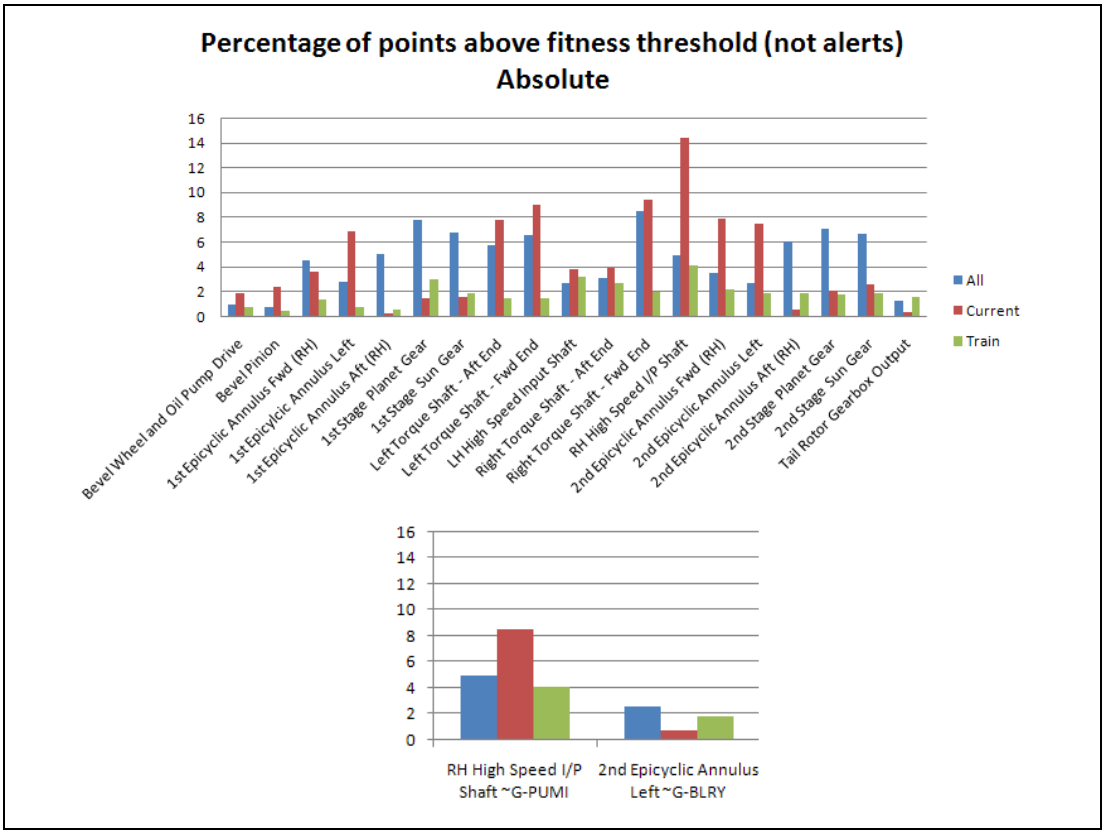

Figure 4-1 Distribution of Fitness Score threshold exceedance for absolute models. The top plot shows the number of component fits exceeding the anomaly model's threshold for each component type. The data are split according to 'training', 'current' (was flying when counted) and 'all' (training and current and some additional historical data not selected for training). The lower plot shows, for two component types, the impact of removing a single component fit.

Figure 4-1 shows absolute models. Figure 4-2 shows the effect of single component fits on the trend models. The bias in the trend models is not as large as that in the absolute models because the pre-processing performs a type of normalisation which reduces a secondary inherent bias due to the tendency of different component fits to have different 'normal' vibration energy levels.

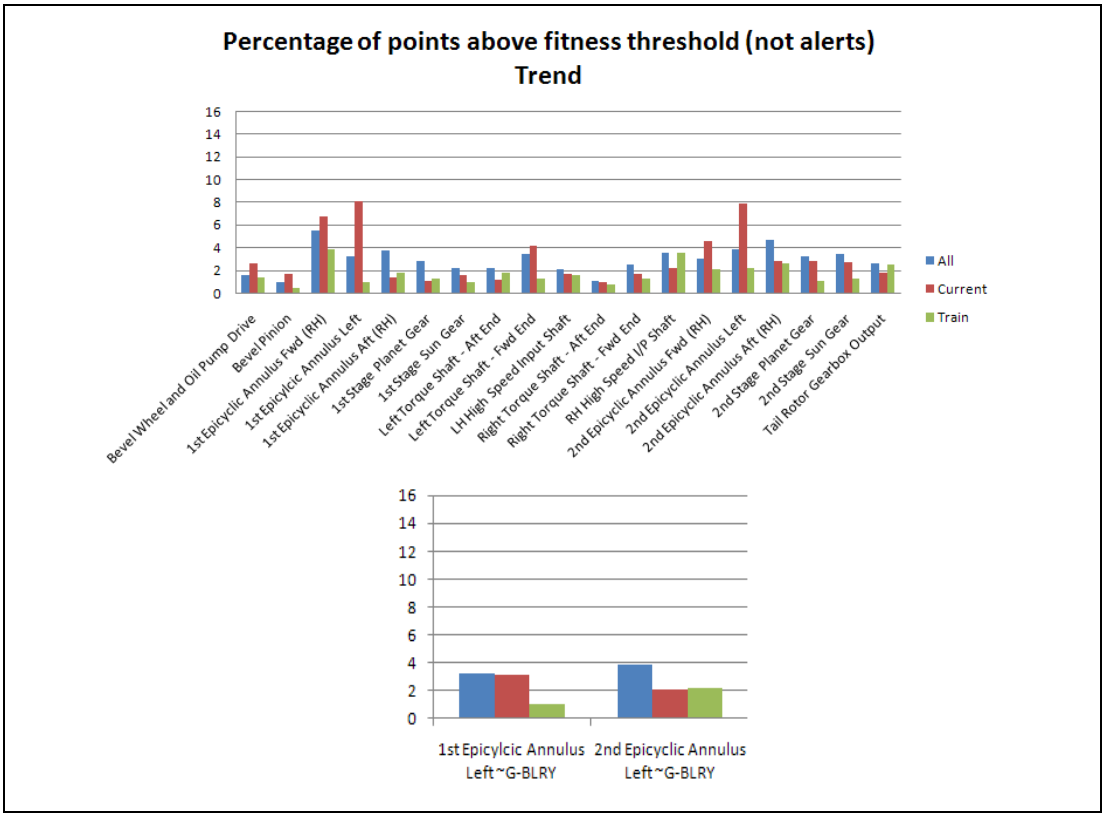

Figure 4-2 Distribution of Fitness Score threshold exceedance for trend models. The top plot shows the number of component fits exceeding the anomaly model's threshold for each component type. The data are split according to 'training', 'current' (was flying when counted) and 'all' (training and current and some additional historical data not selected for training). The lower plot shows, for two component types, the impact of removing a single component fit.

Following the completion of the first six-month trial there was an internally funded effort to re-implement the software for the core data mining algorithms used in the anomaly modelling process. This also facilitated a review of the modelling approach from first principles (i.e. from a theoretical viewpoint). The work showed the approach to anomaly modelling to be well founded, and produced ideas to improve the technique. As a result of the internally funded effort the technique is now more robust.

The improved anomaly model construction technique automatically detects when a component fit has insufficient data to contribute to an anomaly model, and will not bias the model to component fits containing a large number of acquisitions. These improvements also mean that there is no need to categorise data as training, test or validation which is a common practice in data mining techniques to ensure that built models will generalise to data not used for training. In other words, all data can contribute to building a model. This also has the advantage that online model updates can be performed as new data are acquired. The ability to update models is important, particularly for a new aircraft type. The improved technique also allows the influence of a component fit to be removed from a model without having to repeat the whole

training process. This is another valuable feature because the process of model building and review can highlight previously unknown data problems that can show a component fit to be so abnormal that it should not be used for training.

Another modelling change following the initial trial has been to allow clusters to rotate to model correlations between CIs (see Figure 4-3). This would not be expected to have a significant impact on the anomalies detected, but it does allow more diagnostic information to be derived from the model, such as a de-correlation between CIs.

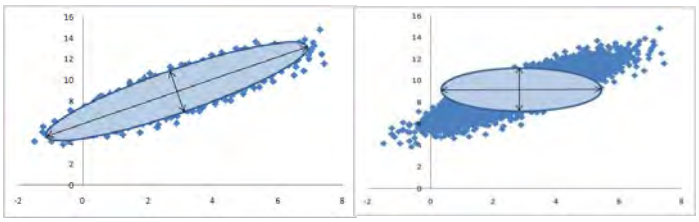

Figure 4-3 Cluster rotation. The data points shown in the right plot are clearly correlated but the cluster (ellipse) has a diagonal covariance matrix and will not model the correlation whereas the cluster in the plot to the left is described by a full covariance matrix and therefore models the correlation.

4.2 Model Tuning and Re-modelling

A model can be tuned by changing a single parameter, called the "percentage removed", that specifies when cluster removal to eliminate suspect data should halt. The percentage removed parameter is set to a value based on a judgement of how much bad data exists in the historical database. This judgment is based on data exploration, initial model builds and knowledge of how often faults occur. For example, data exploration will sometimes show that data are split into two general distributions with the second containing a much lower percentage of samples and characterised by a very large variance. The second distribution is clearly different to normal behaviour and its support gives an indication about the quantity of bad data. The percentage removed parameter would normally be set to a value of 1 or 2 but the default value for the IHUMS data is 8 due to the prevalence of instrumentation issues. If a component fit is known to have developed a fault, or the CI responses are known to be unrepresentative of a normal vibration signature, the data for that component fit are not used for training the model. Occasionally, there will be no a-priori knowledge that a component fit's CIs are unrepresentative of the expected vibration signature, but this becomes apparent later during a model review. In this situation, the model can be instructed to re-tune itself ignoring that component.

The new standard procedure for building a model is to use all available data apart from cases that are known a-priori to be anomalous. The models for the second six-month trial have been built using data since 1 January 2001. All models have been re-built using the enhanced technique. A catalogue of significant findings was reported in reference [3] and these findings have been reviewed following the re-modelling.

The anomalies reported in reference [3] have largely remained unchanged following the modelling improvements but there are some differences. Six cases were cited where the anomaly detection identified a fault not seen by the existing HUMS. All of these cases remained clearly anomalous and for some the anomalies were further emphasised following re-modelling. These cases included epicyclic anomalies on three sensors fitted to a gearbox that was eventually rejected for metal contamination, two gearboxes affected (at different times) by a faulty DAPU, and four other instrumentation related fault cases.

The significance of the CHC Scotia bevel pinion fault case is further emphasised, with the FSs becoming less variable across the fleet while at the same time maintaining a clear developing anomalous trend for the fault case. The absolute model shows this case exceeding the model fitness threshold many hours before gearbox removal, whereas the original models triggered alerts only at the time of removal.

Section 4.2 in reference [3] reported cases where the anomaly modelling corroborated existing HUMS alerts. Example 3 from this section cited a suspected instrumentation problem with G-PUMI. Figure 4-4 shows the FSs from the new model and includes additional acquisitions since the end of the first trial period. This figure shows a step change in the FSs for the 2nd epicyclic annulus fwd (RH) when the MGB had been replaced on 6 July 2005. The signal becomes more variable following maintenance, and then highly erratic during the second half of 2006. The HUMS generated intermittent alerts throughout this period while the anomaly detection generated almost continuous alerts for a number of components. An instrumentation issue looked to be the likely cause and the wiring harnesses were checked during a period of heavy maintenance in January 2007. There was a suspected earthing problem with the DAPU. The anomaly model shows that the component's vibration data returned to normal when the aircraft resumed flying in July 2007. This is an example of a general problem with the trial fleet alluded to earlier, where instrumentation problems have previously gone unrecognised and lead to a significant quantity of bad data being recorded.

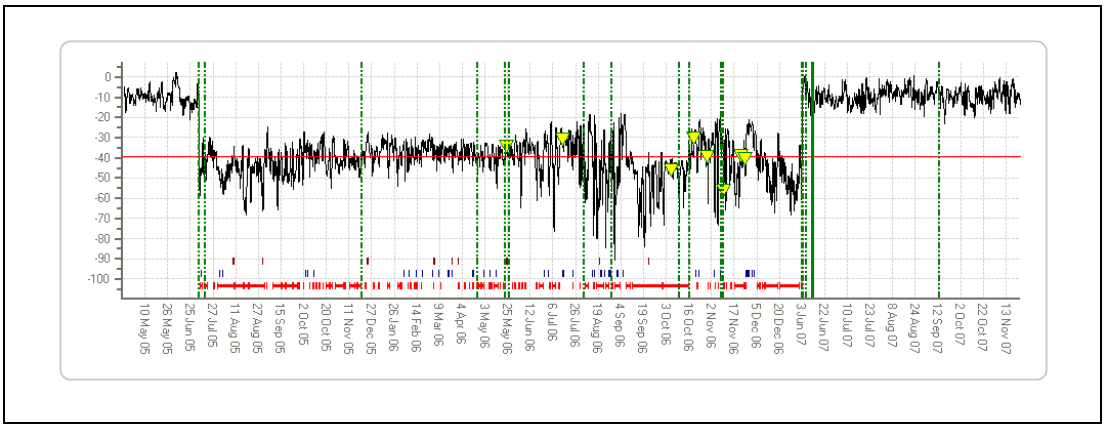

Figure 4-4 Fitness Scores for the 2nd epicyclic annulus forward (RH) for a single component fit generated by the improved modelling technique

Section 4.5 in reference [3] reported a case (Example 1) where the operator believed that the anomaly detection system was alerting prematurely. Figure 4-5 shows this case on the left torque shaft – aft end is still considered to be a clear anomaly. The early alerts are due to shaft order energy CIs behaving erratically and are probably due to an instrumentation problem. When this problem disappears the component is still generating alerts but the signal is steady, hovering around the alert threshold until eventually its trend becomes increasingly anomalous. This FS trend is due to the signal average energy trending higher. Figure 4-6 shows a fleet plot (110 gearboxes) for the Signal Standard Deviation and Signal Peak to Peak. The component is highlighted in black. A significant trend away from the fleet is clearly evident. Statistically it is correct for the anomaly system to flag this case as anomalous. Engineering judgement based on experience has to be used in the interpretation of anomalies to decide if action should be taken. But it is prudent to draw attention to significant patterns. The gearbox was replaced on the 11 February 2007 and as of the 27 November 2007 there have been no anomalies generated for this aircraft/ component. There are likely to be some alerts for the absolute models that, if persistent, could be viewed as nuisance alerts. An example is seen in Figure 4-5

where the FS hovers around the threshold for a long period before it eventually becomes progressively more abnormal. Nuisance alerts can be controlled to some extent by changing the alert threshold. A more complete solution would be to provide a secondary level of processing that recognises the nature of the alert (e.g. a stable FS), and can reason that nothing significant has changed since the alerts were last reviewed. Secondary processing is further discussed in Section 6.

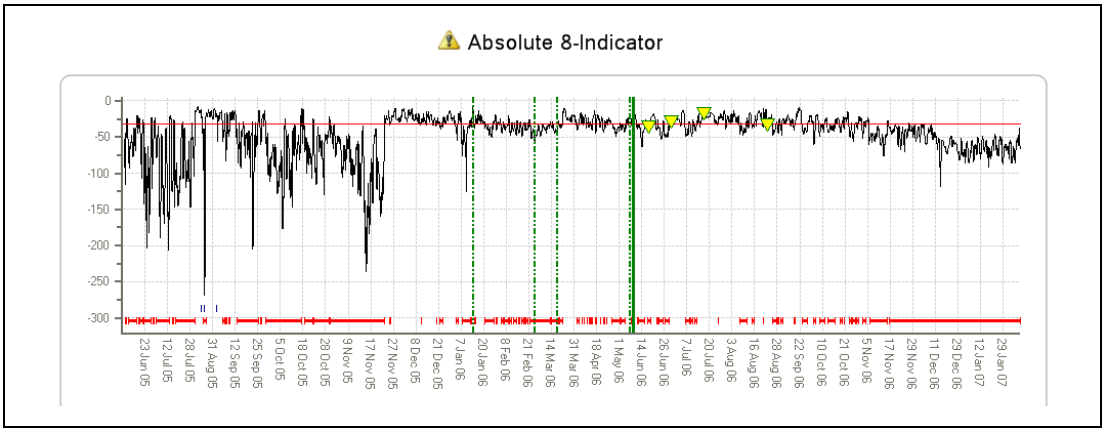

Figure 4-5 Fitness Scores for the left torque shaft – aft end for a single component fit generated by the improved modelling technique

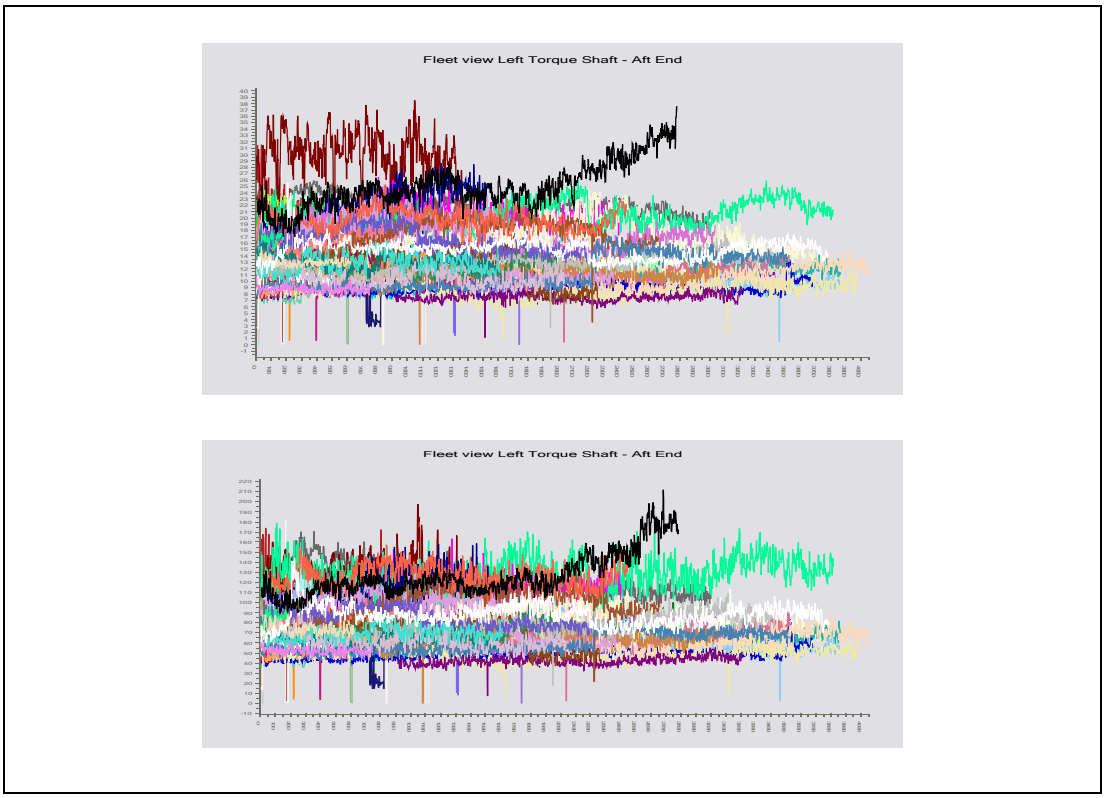

Figure 4-6 Fleet plots (110 gearboxes) of SIG_SD (top) and SIG_PP (bottom) for the left torque shaft – aft end

There was a case cited in reference [3] (Section 4.3 Example 1) where the anomaly modelling failed to identify a fault seen by the existing HUMS. The IHUMS generated an alert on FSA_SO1 for the TGB output shaft and inspection showed damage to the tail rotor flapping hinge bearings. This case is an example of how knowledge about the components being monitored can feed into the anomaly model building process.

The models for the TGB Output were built using eight CIs and it was suspected that noise in other CIs (i.e. not FSA_SO1) could be masking more subtle changes in FSA_SO1. A decision was made to implement new models for key Input/Output shafts (i.e. MGB Inputs and TGB Output) using only shaft related parameters (FSA_SO1, FSA_SON) so that the modelling would be more sensitive to shaft faults. The output from the new model for this case is shown in Figure 4-7. The missed fault related incident occurred on 22/23 of June 2006, but Figure 4-7 shows that the new model generates clear alerts for these acquisitions.

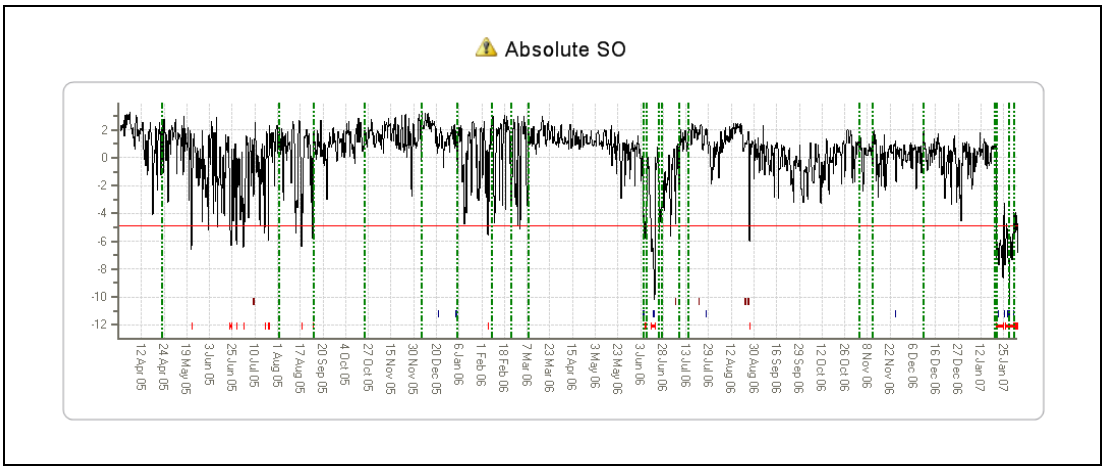

Figure 4-7 Fitness Score outputs from a model of the tail output shaft

4.3 Summary

The improved modelling technique has provided a significant enhancement to the anomaly modelling capability. This improved technique is more robust, no longer requires data to be split into training, test and validation sets, and provides a method to automatically adapt the models as new data are acquired.

It was originally envisaged that this task would result in a few of the existing trial models being re-tuned and new shaft models built for the MGB inputs and TGB output. Instead, all models have been re-built using the improved modelling capability. A detailed review of the models has confirmed enhanced performance. The anomaly in the cracked bevel pinion case is further emphasised, the documented anomalies in reference [3] have been confirmed, and the new shaft models alert to the previously missed problem with a TGB output shaft.

5 Probabilistic Alerting Policy

During the first six-month trial, alerts were defined using a FS threshold. Each model had its own threshold derived using an extreme value distribution built from training cases with low FS values. Deriving the extreme value distribution was not straightforward. The principal difficulty was caused by the nature of bad data mentioned in Section 4 that could lead to the training data being biased. The bias affects the FS distributions, and hence the extreme value distribution and consequently the threshold. Also, the extreme value distribution was built using training cases with the lowest FSs. Selecting these samples was based on a simple and rather arbitrary rule, and these samples could largely be generated by three or four components (due to the bias).

Setting a FS threshold can be a difficult task because the distribution and scale of the FSs vary between models. The problem is analogous to defining univariate thresholds for HUMS CIs. Different CIs have different scales and behaviour, dependant on the component being monitored. The anomaly modelling provides a significant enhancement to existing HUMS because it detects previously unseen faults, and it fuses many CIs into a single anomaly measure. But the anomaly detection could be enhanced if the measure of what is anomalous is normalised across models. Normalisation allows model outputs to be compared, and such a measure could be fed into a secondary process such as automated reasoning to assess the nature of an anomaly.

A normalised measure called the "Probability of Anomaly" (PA) has been derived. This has proved to be a successful enhancement. Alert threshold settings are now more robust, and it is easier to interpret the significance of anomalies. The measure will also provide a valuable input to future enhancements such as reasoning. In addition, it will facilitate data mining of anomalous patterns in the search for new knowledge.

5.1 Development

The PA is a normalised probability measure that ranges between 0 and 1. The alerting threshold is defined for a model using the PA. A default PA threshold is defined that is common to all models, and is currently set to 0.9. The FS equivalent of the 0.9 PA will be different for each model. The default PA threshold for an individual model can be changed if it is deemed that the alert rate is too high or low.

For each model there is a PA distribution which is an extreme value distribution (see reference [2]). An FS value is passed to the PA distribution and a PA value (probability) is returned. Most FS values will return a PA of 0 because most acquisitions will be normal. The PA distribution is built using outlying data. There is quite a narrow margin between a FS value that returns a PA of 0 and FS value that returns a PA of 0.9 or greater. So the resolution of the PA distribution is quite low and PA values can often appear to be binary in nature. The resolution can easily be changed but a low resolution has the advantage of taking the variability out of the signals to provide a cleaner picture of the output, which helps with visualising results and future enhancements such as data mining. No information is lost due to the resolution because signal shape and detail is contained in the FSs which remain a key output of the anomaly modelling.

Sampling theory is used to overcome the bias mentioned earlier. The improved modelling technique discussed in Section 4 overcomes modelling bias due to the strong influence of some gearboxes. Sampling to derive the PA distribution gives better coverage of a model's behaviour, i.e. it is not constrained by anomalies seen

to date. (The reason that sampling is used is that it is not possible to derive the extreme value distribution analytically.) The idea of sampling is based on the assumption that a model represents the statistical behaviour of a component type. It is better to build a model of normal behaviour and generate samples from this model that are outliers, as opposed to modelling with a much smaller set of acquired outliers. This provides more assured data coverage and is a conservative approach in that, if the model does not accurately represent a component's statistical behaviour, the result will be increased alerts rather than masked faults.

The basic procedure will be explained with a simple synthetic dataset. Figure 5-1 shows a scatter plot for data with three features (x, y, z). Only plots of x against the other two variables are shown, but there is correlation between x and z and some correlation between y and z. The red points have the lowest FSs. Some red points are outliers only on a single feature and so they can look normal when plotted using other features. The red points are used to fit a PA distribution that is shown in Figure 5-2. To show the output of the process some additional random data was generated. The features for these data points are presented in Figure 5-3, which shows a developing trend on y. The FSs for the new samples are plotted in Figure 5-4. The red points are those with a PA value of 0.9 or greater.

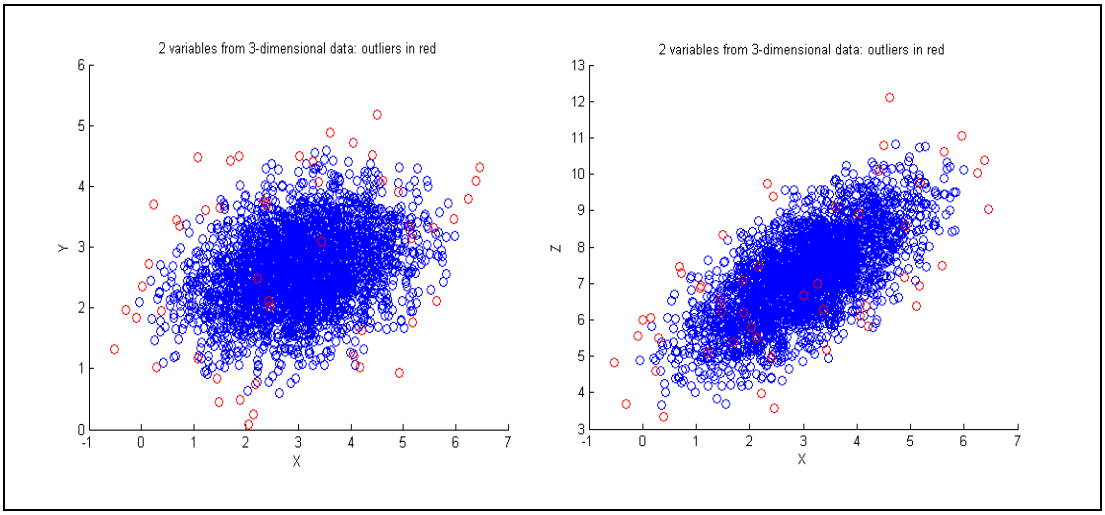

Figure 5-1 Synthetic data set with outliers shown in red

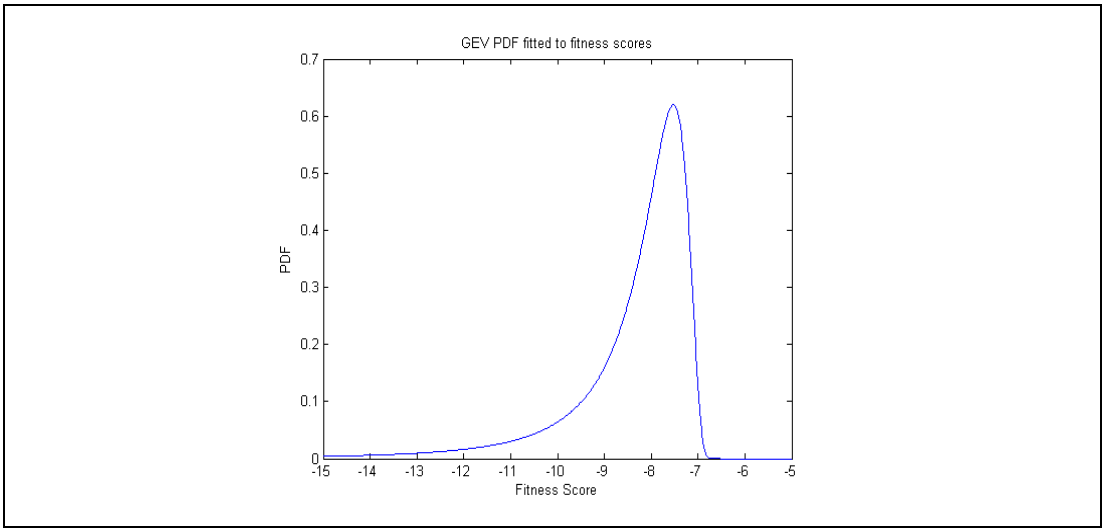

Figure 5-2 An extreme value distribution generated using the red points in Figure 5-1

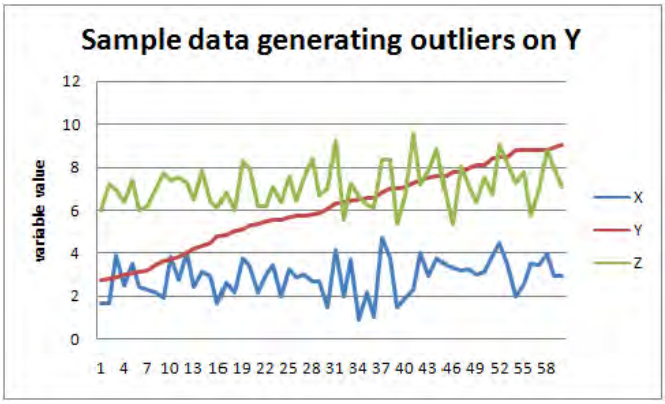

Figure 5-3 Additional random data (not used to train the model) with developing trend in y

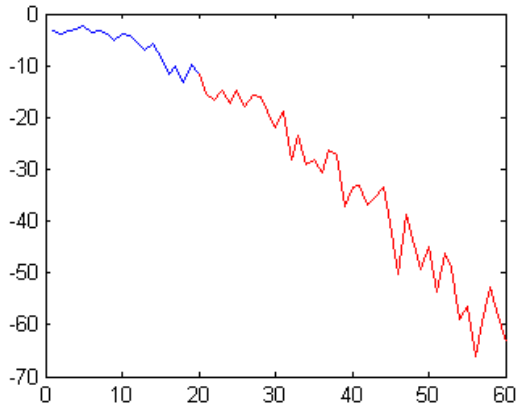

Figure 5-4 Fitness Scores for the additional random data. The red points have exceeded the threshold defined using the extreme value distribution in Figure 5-2.

5.2 Example Outputs Using the PA

Figure 5-5 shows a fleet plot (110 gearboxes) of eight CIs for the first stage plant gear. It is clear that different gearbox / aircraft fits have their own vibration characteristics. They can exhibit considerable variation throughout their life, and trends are not uncommon. Two of the gearboxes have an association in that a faulty DAPU fitted to one aircraft was removed and fitted to a second aircraft, and later returned to the original aircraft. The faulty DAPU generated no IHUMS alerts (this case was reported in reference [3], Example 2 Section 4.1)

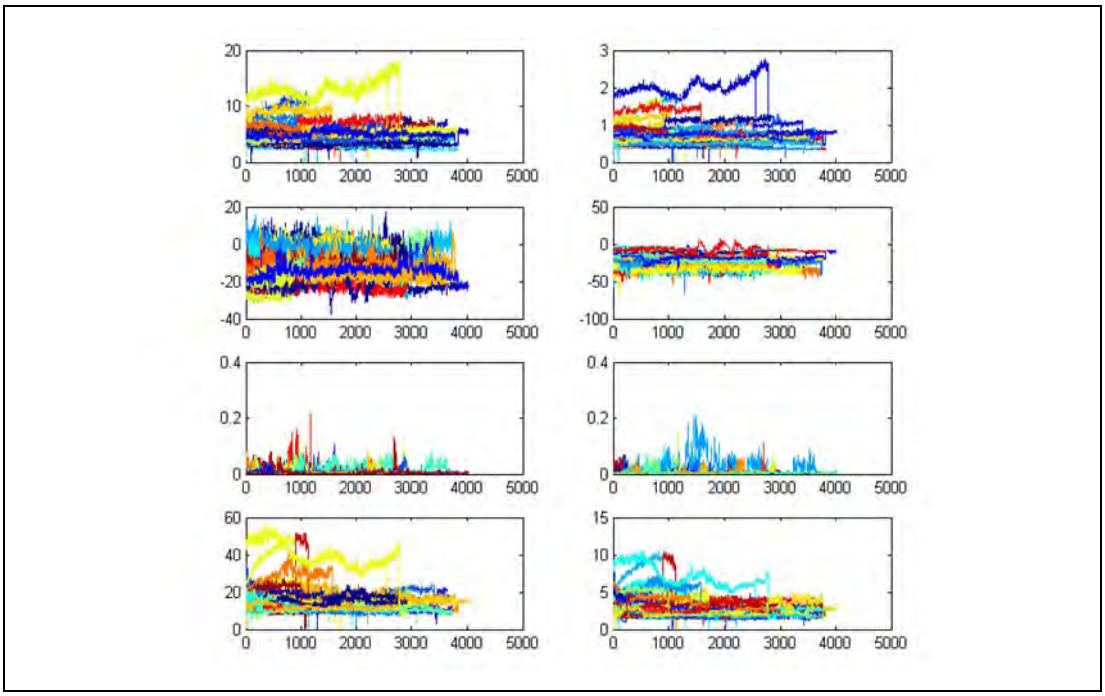

Figure 5-5 Fleet plot of eight CIs from the first stage planet gear

Trying to spot a fault-related association in the data shown in Figure 5-5 is almost impossible to do manually, and it is certainly impractical. The two gearboxes in question have been extracted from Figure 5-5 and presented in Figure 5-6. Having highlighted the two components and given information that there is an association, the reader might guess where in the signals the association lies. But it should be clear that to spot associations in a large quantity of noisy data is a very challenging task.

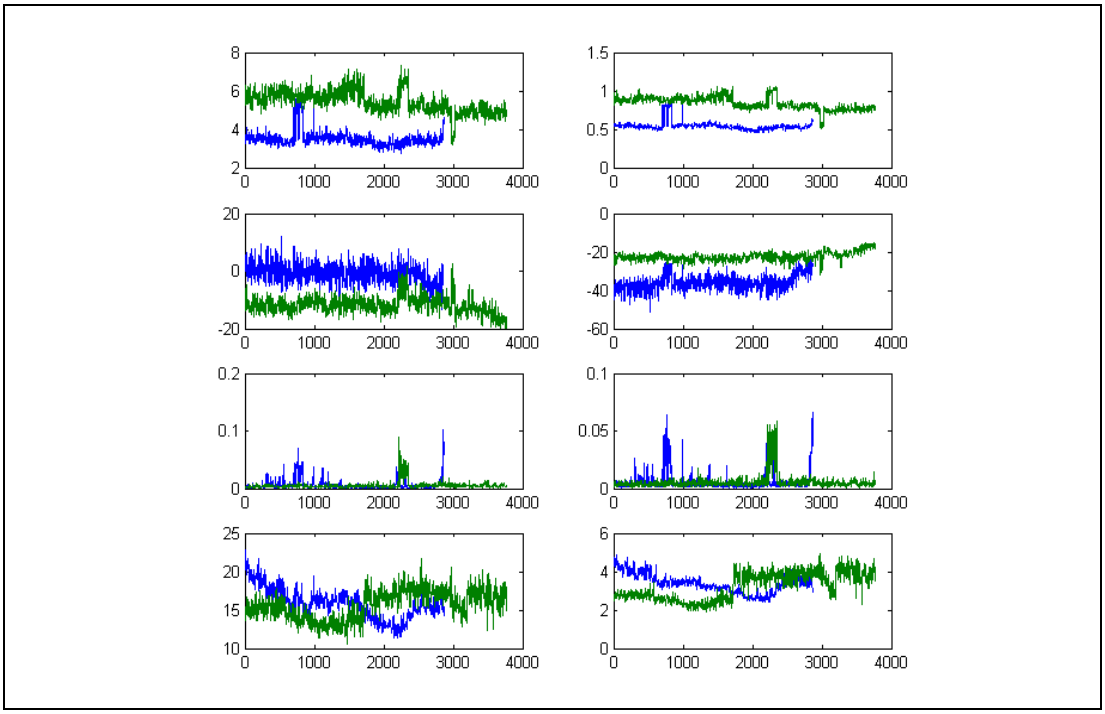

Figure 5-6 Two components, having an instrument fault related association, extracted from Figure 5-5

Figure 5-7 shows the PA outputs from anomaly models for the two components. The PAs are fused outputs that replace the eight CIs with a single trace that removes the noise and highlights the periods exhibiting abnormal behaviour. The IHUMS DAPU was fitted into G-TIGJ on 24 April 2006, and then removed on 18 May 2006 and fitted into G-TIGG. It was fitted back into G-TIGJ on 16 June 2006, and finally removed as unserviceable on 19 June 2006.

Figure 5-7 PA traces for G-TIGJ (top) and G-TIGG (bottom). Two alert periods are clearly visible on the top plot, and a single alert period in the bottom plot. The end of the first alert period for G-TIGJ coincides with the start of the alert for G-TIGG, and the end of the alert for G-TIGG corresponds to the start of the second alert period for G-TIGJ.

The PA traces not only clearly identify abnormal behaviour, they transform the data into an output that allows secondary processing to be performed such as reasoning or data mining. For example, data mining for associations between components becomes straightforward once the periods of abnormal behaviour can be clearly described (as with the PA).

The anomaly models for the initial trial period clearly highlighted the abnormal behaviour in the CHC Scotia bevel pinion crack case. The models generated a consistent period of alerts on the trend model, and an alert on the last data point on the absolute model. The FSs and the PA trace for this case, generated from the new absolute model, are shown in Figure 5-8. This model is now generating clear alerts some 45 hours before the end of the component's life. This case now generates consistent absolute and trend alerts, further emphasising its clear abnormal behaviour. The PA distribution was used to set the FS threshold, so the red threshold lines in both plots correspond to the same level of abnormal behaviour.

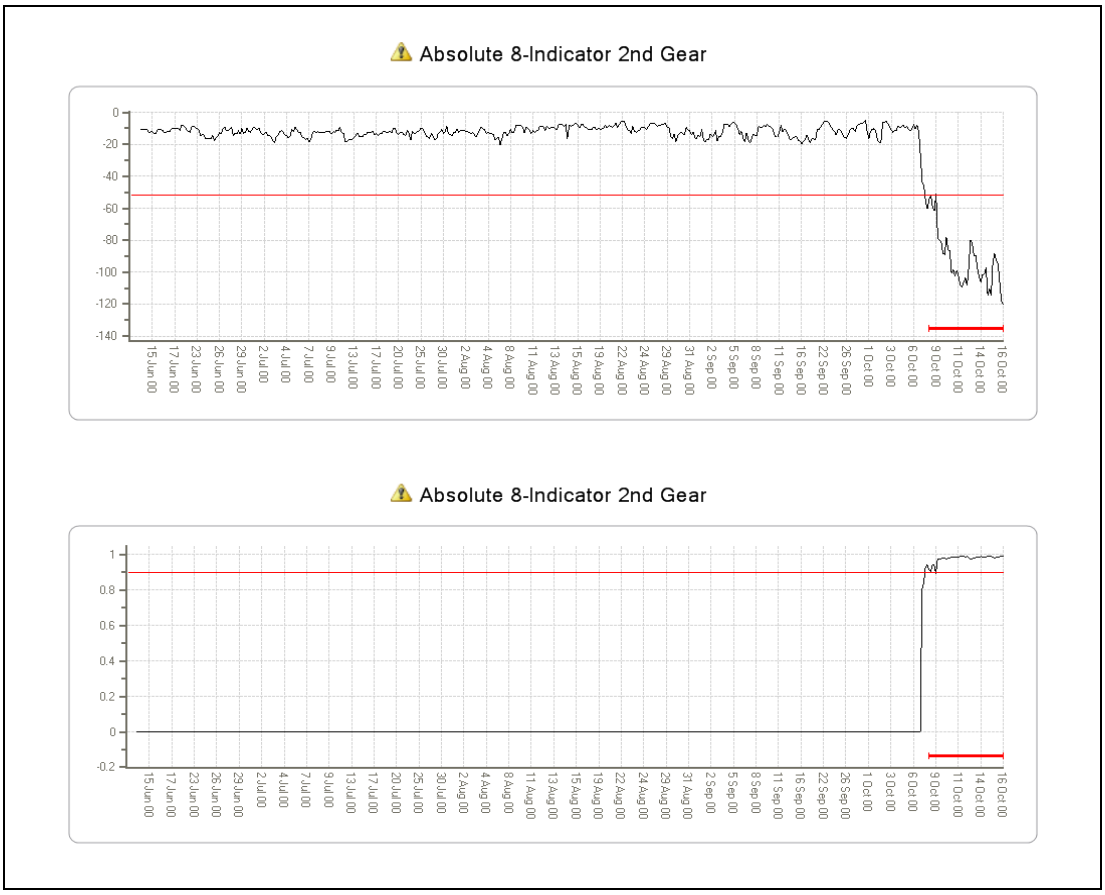

Figure 5-8 New absolute model results for bevel pinion crack case. The top plot shows the Fitness Score, the bottom plot the PA output.

The fitness and PA traces for the TGB output shaft case presented in Section 4.2 are shown in Figure 5-9. The red threshold lines in both plots correspond to the same level of abnormal behaviour. The FS trace shows the component exhibits highly variable behaviour, but only a relatively small number of points exceed the threshold before the component eventually enters into a period of sustained alerts until the gearbox is removed. The PA output shows more clearly the periods of erratic behaviour and cleans up the output by suppressing normal behaviour.

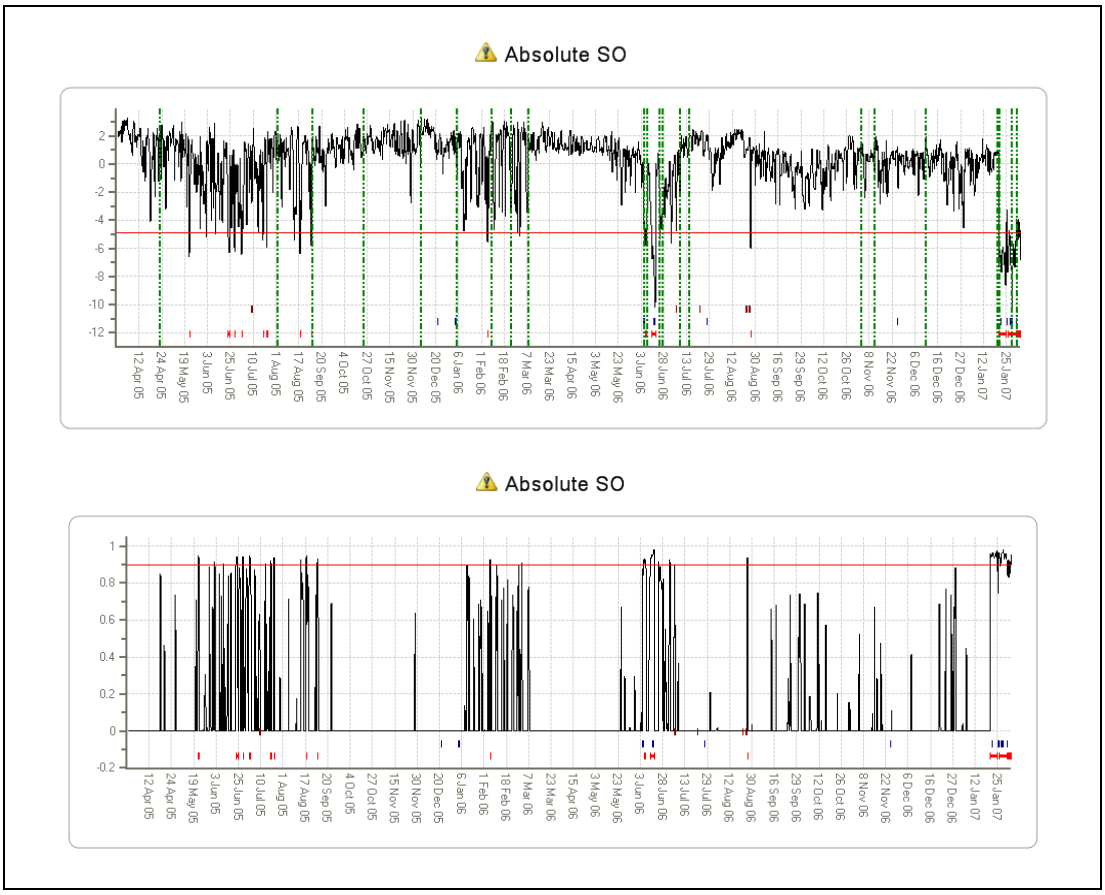

Figure 5-9 Fitness Score and PA traces for a single component fit from a shaft model of the TGB output

5.3 Implementation of Probabilistic Alerting

Software for learning the extreme value distribution has been implemented as part of the anomaly model building process. This implementation includes code for generating simulated extreme outliers from a model, fitting an extreme value distribution, and extraction of the FS equivalent to the default PA value of 0.9. When a model is learnt, the PA distribution is automatically generated and its threshold (FS equivalent) stored.

A new plot type has been added to the CAA trial web site to display the PA traces. The PA trace is now the default view of anomalous behaviour (Figure 5-10) but the FS traces can be selected and, as stated earlier, these remain a key output because they retain shape information and highlight any abnormal trend (e.g. is the anomaly getting worse?).

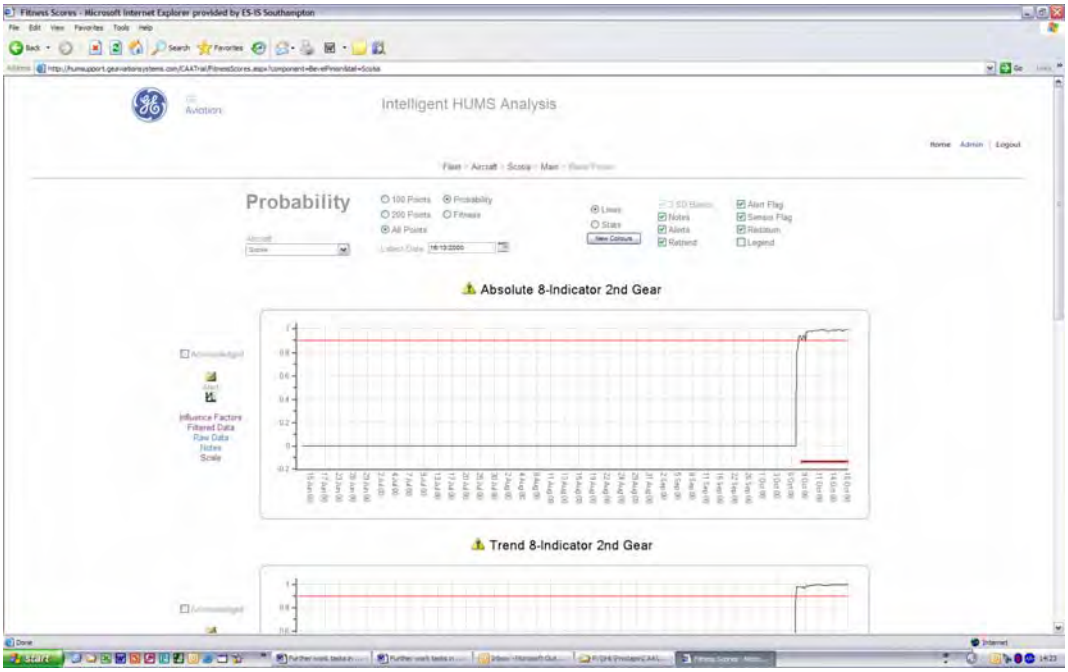

Figure 5-10 Default PA traces for an anomaly alert

5.4 Summary

A robust method for setting thresholds on anomaly models has been presented. The procedure is fully automated and has the advantage that a global default threshold is set across all models and components. The default value can be adjusted for an individual model if it is deemed to generate either too few, or too many alerts.

The method generates a PA distribution which is an extreme value distribution derived from extreme outliers simulated from a model. PA traces provide a simplified view of abnormal behaviour and have utility for future secondary processing such as reasoning and data mining.

6 Influence Factors

Influence Factors (IFs) are a type of prediction generated by GE Aviation's ProDAPS (Probabilistic and Diagnostic System) that can provide diagnostic information about a model (e.g. an anomaly model), and its input training variables (e.g. HUMS CIs). A prediction is an output from a trained model when a case is presented, for example the FS for a HUMS acquisition. There are different types of IF, each type producing a different view on a model or input variable.

Currently, an anomaly alert is usually followed up with a manual inspection of the CIs to see which of them might be responsible for the alert. Charting of CIs is a useful technique but it can lead to incorrect interpretation. The numerical range and scatter of CIs vary considerably, and it can be very difficult to visually assess the significance of a CI trend. Univariate fleet plots of CIs help view the trend of an individual component in the context of the fleet, but these fleet plots can give a false picture when the data contain a lot of scatter because they hide where the true density of the data lie.

This task was to assess the suitability of IFs for assisting the interpretation of the significance of individual HUMS CIs, and to implement them for the second six-month trial if found useful.

6.1 Properties of Influence Factors

An IF is a single prediction, in that an acquisition will generate a single IF predicted value for each CI used to train a model. A history of IFs for a single component fit produces an IF trace. IF traces provide information regarding the influence of different HUMS CIs on the fused FS. These traces are not the same as plots of CIs as they are assessing the contribution of individual CIs to a statistical measure of abnormality. Unlike CIs, IFs are normalised and can be directly compared.

IFs are unique to ProDAPS, but there is another method proposed by researchers within the chemical industry to assess the contribution of variables to a process that is outside of normal operating conditions. Statistical process control was introduced in Section 3 but the application was univariate (monitoring a single variable). More complicated process monitoring takes account of multiple process variables and is known as multivariate statistical process control (MSPC). There are two popular approaches to MSPC, one based on principal component analysis (PCA), the other on projection to latent structures (PLS). These have also been extended to batch processes ([10] and [11]) where a model for nominal process control can be constructed from past successful production batches.

For MPCA, data from different batches are concatenated into one large measurement matrix. The columns of this matrix are mean centred and scaled to unit variance. PCA is then performed on the correlation matrix of the measurement matrix. PCA is a linear transformation of the measurement variable space to a new set of axes (principal axes). Each principal axis is a weighted combination of the original variables. The principal axes are ordered according to the amount of variance in the measurement variables they represent. The first axis will describe most variance, the second axis the most variance remaining (after discounting the first axis variance) and so on. A number of minor (least variance represented) principal axes will be dropped. A projection of the measurement matrix onto the remaining principal axes results in a compressed representation that captures most of the information contained in the measurements while reducing small variations (noise).

Outliers are detected within the transformed measurement space. Two measurements for outlier detection are calculated: Hotelling's T^2 and the squared

prediction error (SPE). The T^2 is a measurement within the transformed principal space and the SPE is a measurement within the residual space (error between the measurement and predicted value from the reduced principal transformation). Thresholds are applied to T^2 and SPE calculations to detect outliers.

Once an outlier is detected the next question is "what is the cause?" and to assist this analysis contribution plots are calculated that give a visual measure of the contribution of each measurement variable to the outlier. The calculations are relatively straightforward.

The role of MPCA, T^2, SPE and contribution plots for monitoring chemical processes is similar to anomaly detection and influence factors presented in this report. The key difference between ProDAPS anomaly models and MPCA models is that ProDAPS anomaly models are more complicated in that they represent normal behaviour with a mixture of Gaussian distributions. MPCA uses a single distribution. The contribution plots used with MPCA could be used with ProDAPS anomaly models in place of influence factors. However the usefulness of the result would depend on the association a sample has with a single cluster because MPCA and contribution plots work with a single distribution (cluster). It could also prove difficult to perform a meaningful PCA on a cluster with little support (training cases assigned).

There are several types of IF, each type designed to provide different information about a model and/or an input feature (e.g. a CI). The type of IF is determined by three attributes, each attribute being a true/false option. The first attribute is called 'modal' and if set to true the calculations are based around a single cluster which is the one most associated with the case (e.g. acquisition). The calculations are simplified when modal is selected, but the result loses fidelity for cases that are associated with multiple clusters. The second attribute is what is called a 'posterior weighting'. This option determines the contribution the other CIs can make to the calculation of the IF for the selected CI. The third attribute is called 'Forced Diagonal' which if set to true removes the influence of correlations in the calculations of IFs.

All types of IF have been generated for the IHUMS data. The interpretation of the correlation variant ('Force Diagonal' set to false) is rather complicated and it needs to be fully evaluated to see in which contexts it has most value (quite often this type of IF produces more information about the model rather than the CIs). There is no suggestion of presenting more than one type of IF to an operator since this would complicate their diagnostic task as opposed to simplifying it. The type presented to the operator is 'Forced Diagonal' set to true, modal and weighting set to false. This combination of options allows each CI to be assessed independently of other CIs, and is therefore the type of IF that will be closest to the CI plots that the operator is most familiar with.

An IF can be scaled, and this is the default method for user presentation. Scaling an IF means that an IF trace is generated relative to the alert threshold for the anomaly model. When an IF has a value of 1 or greater it means that the individual IF could have driven the FS into alert. Equivalently if, for a given point in time, the individual IFs sum to 1 or greater, the FS is in alert. Therefore it is possible to examine the IF traces for a model and see whether a component is behaving abnormally and which CIs are contributing to this abnormal response. IFs can appear to be flat and have a value of zero if the data are considered by the model to be normal. An IF trace therefore may not appear to have the same shape as its related CI trace. It is also possible for the scaling to change the shape of a CI in a way that on first inspection seems counter intuitive. For instance, suppose a single CI (e.g. FSA_SO1) has high values that are abnormal and, after a period of time, it further increases in value (i.e. becomes even more abnormal). If the initial period of abnormality is due only to the

individual CI but the later period of increased abnormality is due to multiple CIs, it is possible for the IF (e.g. FSA_SO1) to drop because its relative contribution has dropped (other CIs are now responding). Therefore a decrease in an IF does not necessarily mean that the corresponding CI has decreased. IF traces are a new view on the data and they cannot be expected to behave in exactly the same way as CIs. Once the user understands how to read an IF trace, they are easy to interpret and provide useful insight into the cause of abnormal behaviour.

6.2 Example Model Outputs of Influence Factors

The familiar case of the cracked bevel pinion will be used to illustrate that the IF type chosen for the six-month trial extension is easy to interpret and does provide useful information. Figure 6-1 shows the HUMS CIs for the bevel pinion case, Figure 6-2 the absolute IFs, and Figure 6-3 the trend IFs. Each CI in Figure 6-1 has its own scale and although trends are visible there is no indication within any plot that can suggest a trend is anomalous. Statistical bands such as mean plus three standard deviations can be overlaid on these CI plots, but this view on the data is restricted to a more traditional HUMS approach and fails to exploit the benefits of the anomaly modelling. Also, as explained in reference [2], data can look normal when individual CIs are examined, but collectively they may be abnormal.

The IFs presented in Figures 6-2 and 6-3 show that several CIs are responding in the bevel pinion case. In the absolute model, FSA_GE22 (wideband modulation) is clearly dominant, followed by FSA_MS_2 (mesh magnitude), and there is response in the enhanced signal measurements ESA_PP and ESA_SD. It is very evident that these four CIs sum to greater than 1 over the last 40+ hours, and the model has therefore exceeded its threshold. The trend model shows the statistical change in the enhanced signals to have similar significance to that for the wideband modulation. FSA_SON also receives slightly more emphasis in the trend model. The information contained in these IF plots such as the statistical significance of individual CIs in both absolute and trend terms is extremely difficult to assess in traditional CI plots, and virtually impossible to quantify.

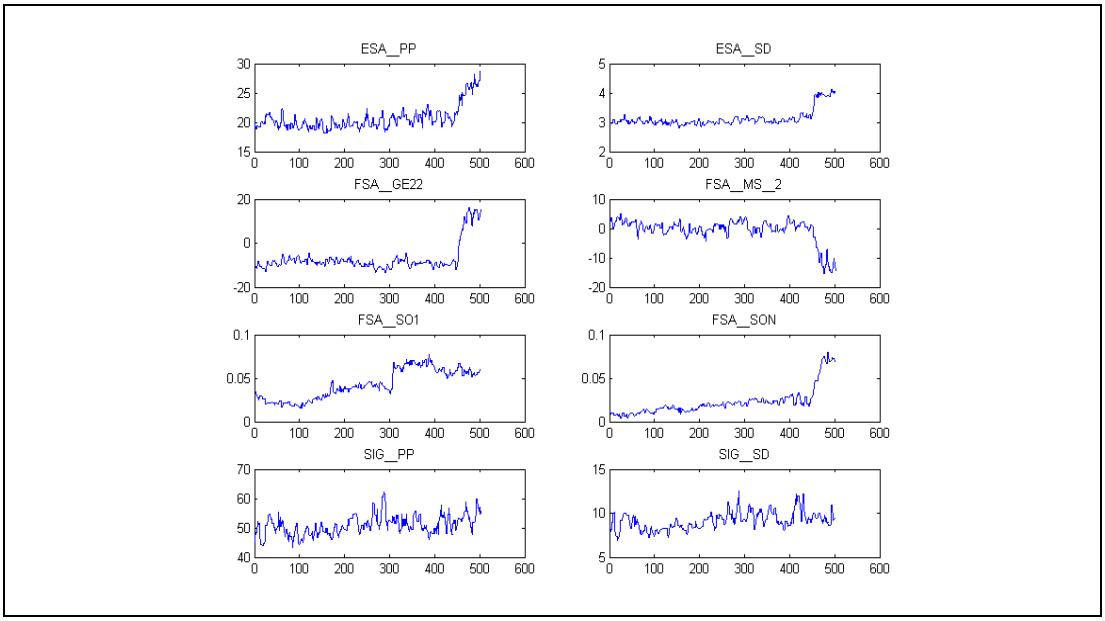

Figure 6-1 HUMS CIs for the bevel pinion fault case

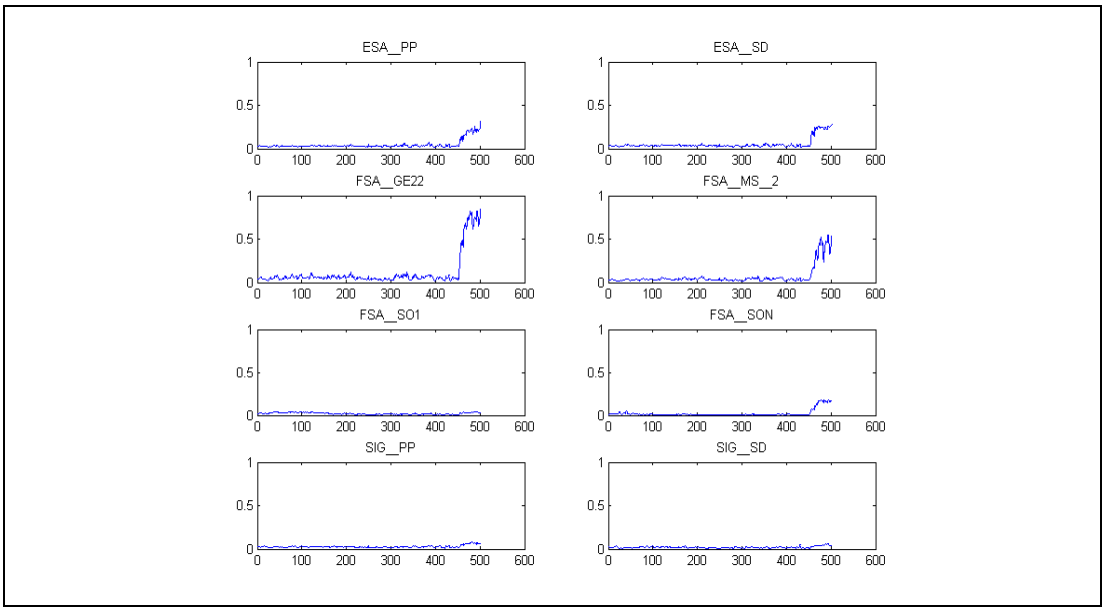

Figure 6-2 IFs for bevel pinion case absolute model

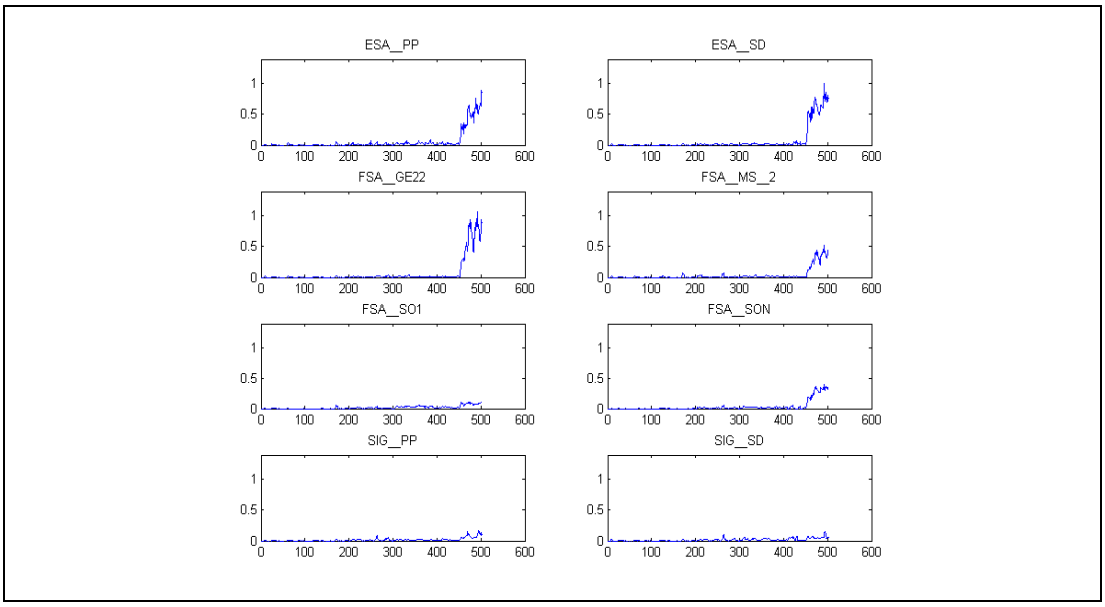

Figure 6-3 IFs for bevel pinion case trend model

It was explained in reference [2] that the HUMS anomaly detection has been configured as a two stage process. The first stage relies on the anomaly modelling to flag attention to anything unusual in the data. The second stage is to interpret the nature of the anomaly and decide what action is necessary. The second stage is at present a manual process but the intention is for this stage to be largely automated in the future. The two stage concept implies that the first stage (detection of abnormal behaviour) should err on the cautious side, and generate false positives in preference to false negatives. Additional information (such as the nature of the anomaly and what is contributing to it) becomes available at the second stage, and difficult borderline decisions are easier to resolve with additional information. IFs, FSs and PAs help make an automated second stage process a reality. PAs provide a very clean view of where attention should be focussed – this was illustrated in the first example in Section 5.2. A PA also enables a global single threshold to be used, because the default value is 0.9 for all models, but 0.9 equates to a different FS

threshold for each model. It is possible therefore to simplify a reasoning architecture because it does not need specialised constructs for representing the associations between individual component types and abnormal behaviour. The same logic applies to the IFs because they are normalised to the same scale. The FS still has a role because it retains shape information (e.g., is a component's abnormal behaviour getting worse).

Figure 6-4 shows the PA and FSs from Example 1 cited in Section 4.1 of reference [3] where the second stage epicyclic was generating trends in the data from three sensors. This case generated no IHUMS alerts and the component was eventually rejected for metal contamination. The PA trace shows that a number of acquisitions generated alerts, but the alert periods were all of short duration apart from the last one that spanned several weeks. The FSs show the final alert period has a degrading trend. Figure 6-5 displays the IFs and these show that the earlier alerts are due to the enhanced energy measures but these are not trending. The final alert period is clearly driven by an increasing trend in the energy of the signal average (SIG_PP and SIG_SD). The CIs for this case are shown in Figure 6-6 along with conventional statistical three standard deviation bands that are derived from the anomaly model (as opposed to being calculated from the database of samples). It can be seen that the IFs respond only to those time periods when vibration signals exhibit outlying behaviour.

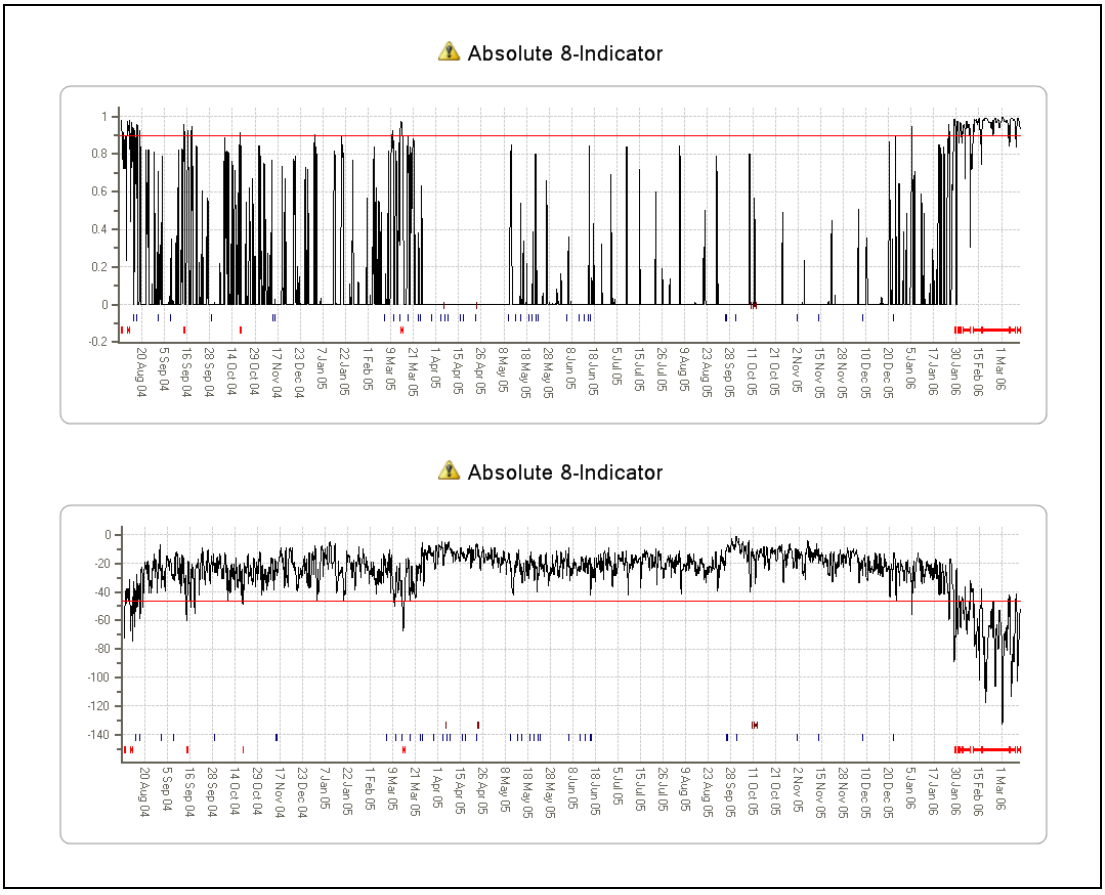

Figure 6-4 Fitness and PA Scores for a second stage epicyclic model

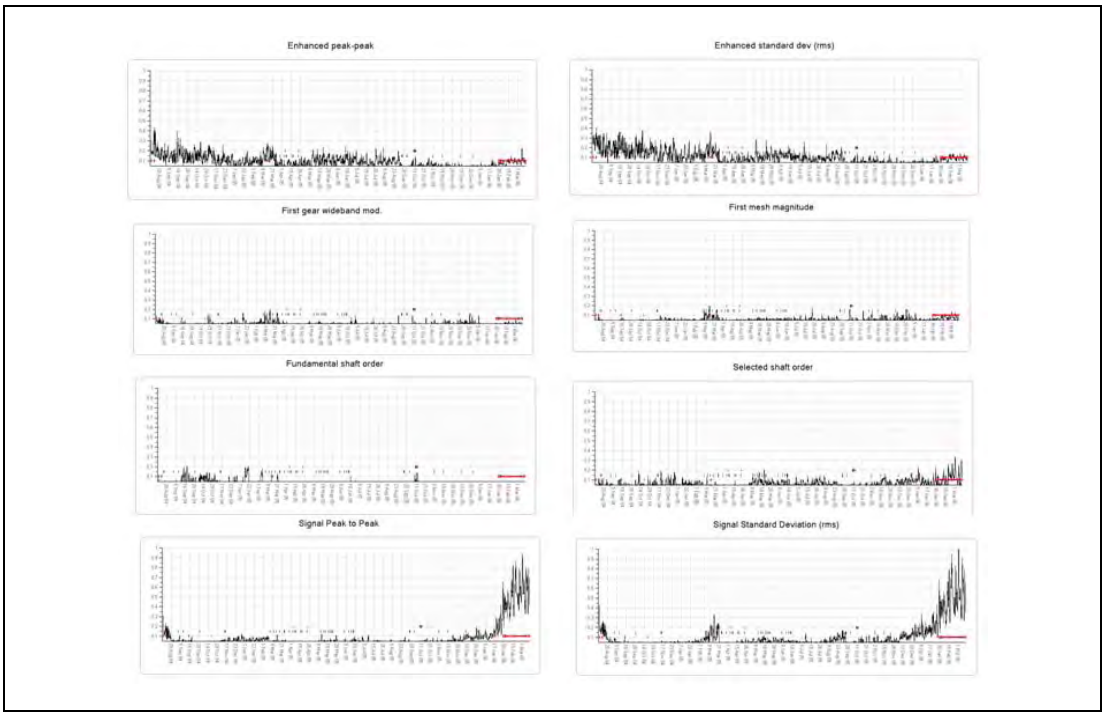

Figure 6-5 IFs generated by the model for the component shown in Figure 6-4

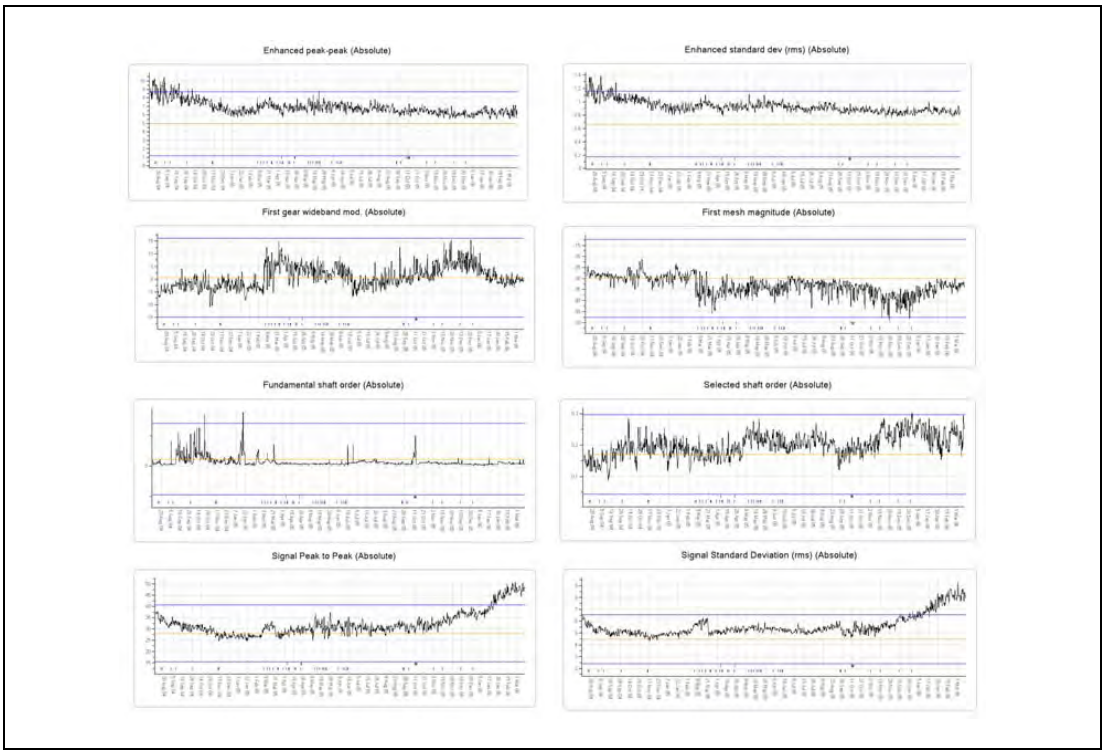

Figure 6-6 The CIs corresponding to the IFs shown in Figure 6-5

The next example shows a case where a component is in continuous alert over a long period, but the nature of the anomaly suggests no action is required other than close monitoring. Figure 6-7 confirms the continuous alert spanning several months that originally started with an increasing trend. Figure 6-8 shows that the anomaly is due mainly to the enhanced signal average energies. These energy CIs are also responsible for the initial trend. An alert generated by an anomaly model means that the data should be reviewed it does not mean a fault exists. The nature of the alert

(e.g. trending fast, CIs responsible, etc.) is then analysed to see if any action should be recommended. The alerts presented here are from absolute models and the nature of absolute thresholds means that alerts will be generated from time to time, that require no immediate action. The summed enhanced signal average energy IFs in Figure 6-8 do not exceed a value of 1 and this means that the alerts are close to the FS threshold. This along with knowing that the anomalies are due to background noise (feature measured by enhanced energies) means there is no immediate concern and the action is to keep monitoring.

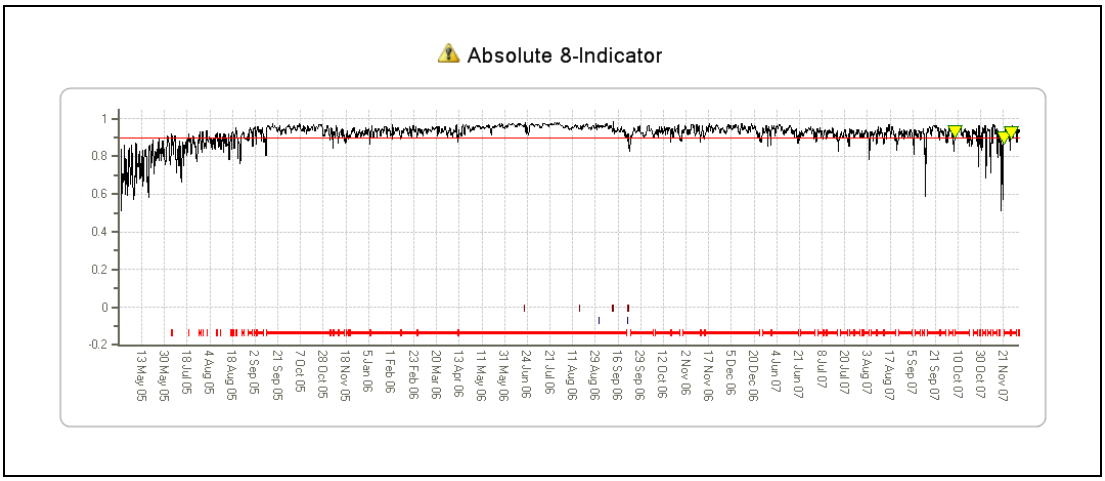

Figure 6-7 PA trace from an IGB input

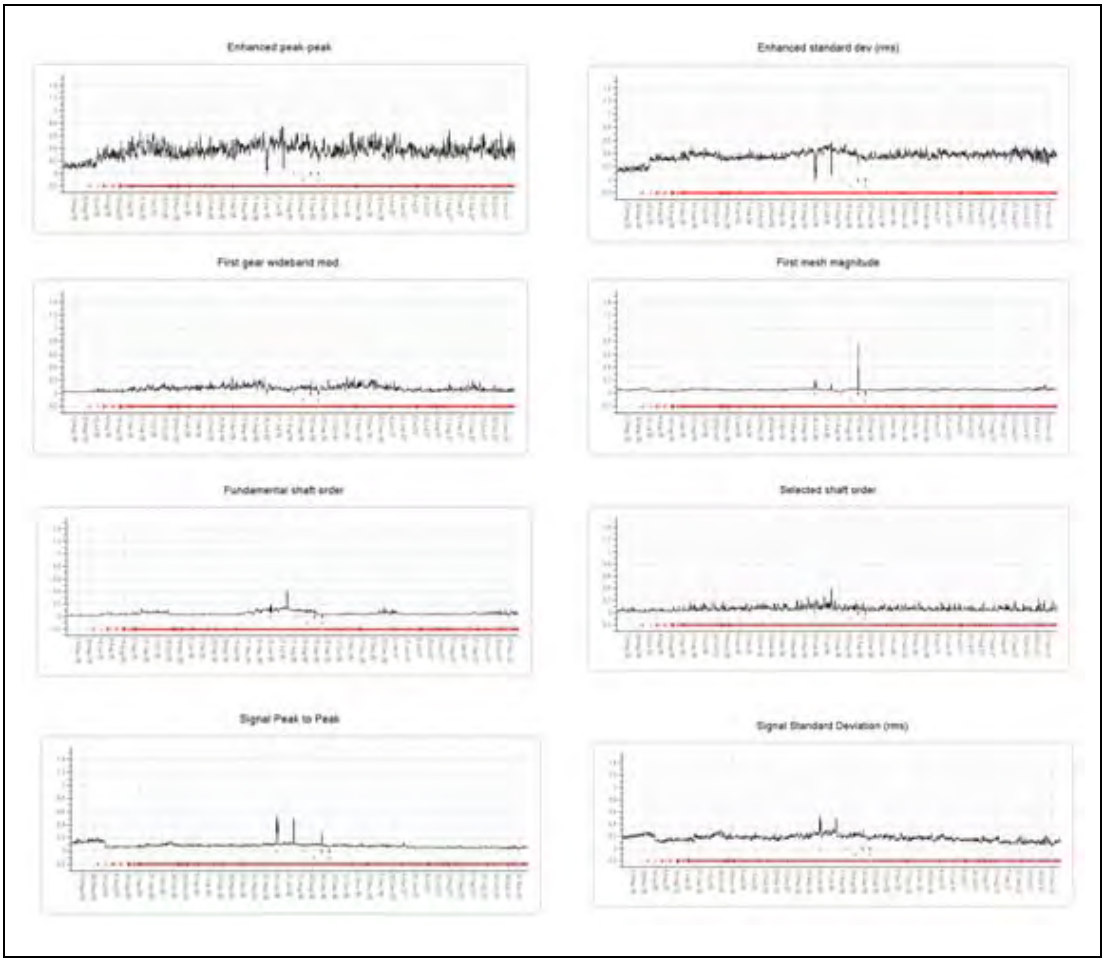

Figure 6-8 The IFs generated by the IGB input model from Figure 6-7

6.3 Web Site Implementation of IFs

A new plot type has been added to the CAA trial web site to display IFs as the first level of supporting information for a PA or FS trace, as shown in Figure 6-9. (Figures 6-5 and 6-8 are also example plots from the web site). Only the 'Forced Diagonal' type is available to external users, but the weighted and modal variants are available to internal GE Aviation users for continued evaluation.

Presentation of IFs created a slight challenge in that the IFs for each indicator will be different, but they are normalised in a way so that they can be directly compared (i.e. the IF with the highest value makes the largest contribution to the anomaly score). For the operator to be able to quickly compare IFs, they need to be drawn on the same scale. There was no time or budget remaining to devise a sophisticated strategy that would work in all cases. So a simple solution was adopted. It was explained earlier that a total IF value of 1 or greater (either from a single IF trace or a sum of traces) means that the model threshold has been exceeded. A decision was made therefore to default the scale of IF plots between -0.2 and 1.4. The minimum value of -0.2 is chosen so that plots can be annotated with other information such as alert flags. Occasionally, extreme anomalous behaviour means that an IF will exceed the default plot scale. When this situation occurs, the user has the ability to rescale the axes. A convenient option allows the user to scale all of the IFs to the same scale as a selected IF. Using this option, the user can identify the most abnormal IF and scale the other IFs relative to it.

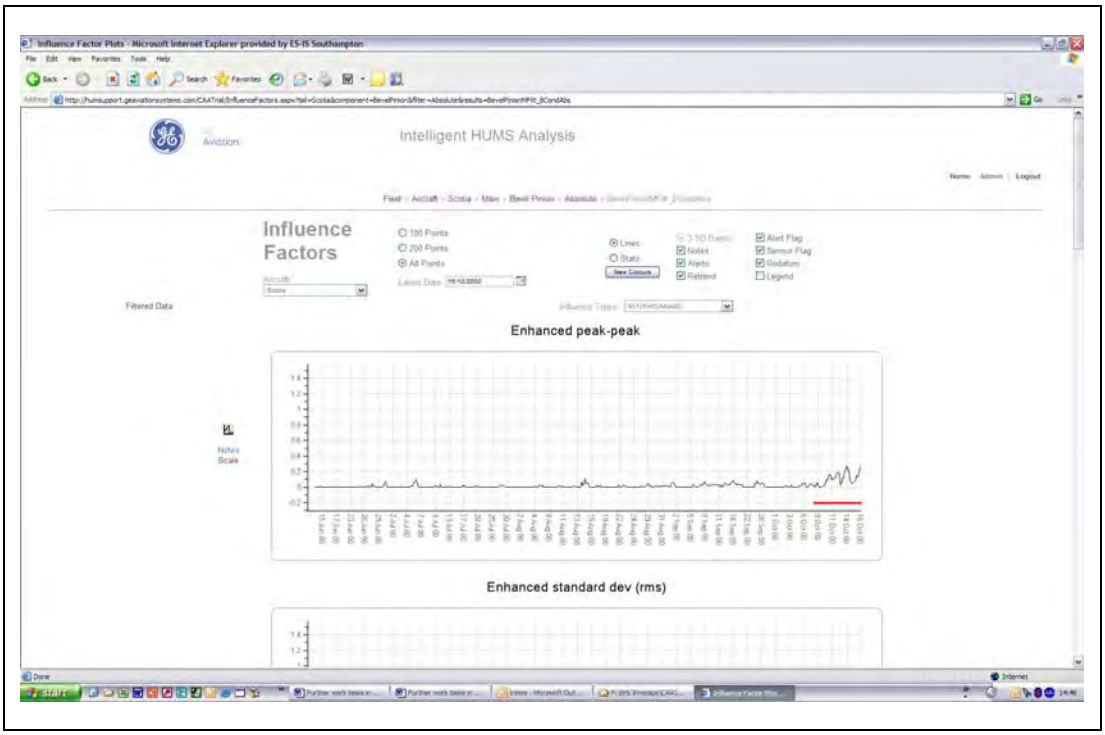

Figure 6-9 IFs displayed in the CAA trial web site

6.4 Summary

A type of influence factor known as 'Forced Diagonal' has been shown to be a valuable model output in assessing individual HUMS CI contributions to the FS. Unlike CIs, IFs are normalised and can be compared to each other. Furthermore, the scaling of IFs provides a measure of severity in that a summed (across all IFs) value of 1 or greater means that the model threshold has been exceeded. IFs therefore provide useful diagnostic information and are represented in a way that can be directly utilised by an automated diagnostic reasoner.

7 Conclusions and Recommendations

7.1 Conclusions

Building on the successful development and six-month in-service trial of the HUMS anomaly detection system, four additional research tasks have been performed to further develop its advanced HUMS data analysis capabilities. These have resulted in significant enhancements to the anomaly detection system, which will be evaluated in a second six-month trial period.

Research has been performed to detect shifts in a signal and identify shapes such as trends, step changes, and noise spikes. The original proposal was to focus on methods to transform CIs to representations that would emphasise these shapes and therefore make the anomaly models sensitive to these features. However it was concluded that pre-processing (transformations before modelling) should be minimised provided key diagnostic information, such as significant trends, do not remain hidden from the anomaly modelling. This strategy makes fewer assumptions before anomaly modelling and is intuitively safer because anomaly alerting is not dependent on a potentially sophisticated pre-processing stage. The moving median difference calculation for building trend models is designed to capture within component variations and should make shifts in the data visible to models. If this strategy works correctly, the shape of any signal feature that leads to a significant shift in the data should be carried through to the Fitness Score (FS). Any FSs that generate alerts can then be analysed to characterise their shape.

It was concluded that two key elements to the identification of shape were signal smoothing and segmenting. Segments could then be characterised using a linear Gaussian model which would indicate any rate of change and the amount of noise present. Segmentation is the most difficult part. A number of techniques were reviewed and it was concluded that with further effort a first generation automated technique for shape detection could be implemented.

An internally funded re-implementation of the anomaly modelling software provided the opportunity to make a number of improvements to the modelling technique created during the initial system development phase. The improved technique is more robust, no longer requires data to be split into training, test and validation sets, and provides a method to automatically adapt the models as new data are acquired. The latter feature has particular significance in the context of applying the technique to new aircraft types or variants. Although not expected to have a significant impact on results, it is also now possible to allow clusters to rotate to model correlations between HUMS CIs.

All anomaly models have been re-built using the improved anomaly modelling capability. A detailed review of the outputs has confirmed enhanced modelling performance. The anomalous data trends associated with the cracked bevel pinion case are further emphasised, and the documented anomalies in reference [3] have been confirmed. In addition, new shaft models have been built for the critical MGB input and TGB output shafts. The TGB output shaft model now alerts to the increase in 1/rev vibration associated with a tail rotor flapping hinge bearing problem that was previously missed.

A new "Probability of Anomaly" (PA) output has been derived for each anomaly model, which is a normalised probability measure that ranges between 0 and 1. PA traces provide a simplified view of anomalies to facilitate easier interpretation of their significance, with the model FS traces being retained to provide detailed anomaly trend information. A more robust method for setting thresholds on anomaly models

has been implemented, based on the PA values. The method generates a PA distribution which is an extreme value distribution derived from extreme outliers simulated from a model. The procedure is fully automated and has the advantage that a global default threshold of 0.9 PA is now set across all models and components. The FS equivalent of the 0.9 PA will be different for each model. The default PA threshold for an individual model can be adjusted if necessary to control the model's alert rate. The PA traces, together with the global default PA threshold, have proved to be valuable enhancements of the anomaly modelling output. They provide a simplified view of abnormal behaviour, and will also facilitate future developments such as automated reasoning, and data mining of anomalous patterns to search for new knowledge.

The anomaly modelling output has been further enhanced by the implementation of 'Influence Factors' (IFs) to provide model diagnostic information. An IF value is calculated for each input HUMS CI value. However, unlike plots of CIs, IF traces assess the contribution of individual CIs to a statistical measure of abnormality derived from the model. IFs are also normalised and can be directly compared. Three variants have been implemented for evaluation by GE Aviation. However only one type of IF, known as 'Forced Diagonal', will be presented to the operator as this provides a simple to interpret output for assessing individual HUMS CI contributions to a model's FS, and hence to an anomaly indication. Furthermore, scaling of the IFs gives a measure of severity in that a summed value of 1 (across all IFs) or greater means that the anomaly model threshold has been exceeded. IFs therefore provide useful diagnostic information, and can be directly utilised by a future automated diagnostic reasoner.

7.2 Recommendations

The work described in this report has provided an enhanced anomaly detection capability, and generated new information in the form of PA and IF traces that can be used for further information processing and analysis. It is therefore recommended that the following additional tasks are performed:

1 Demonstrate the application of an automated reasoning capability to the anomaly model outputs, adding further intelligence to the advanced HUMS analysis, and guiding the operator in the correct interpretation of, and response to, the analysis results. (Research Task 5 in reference [12]).

2 Demonstrate the application of data mining to the advanced HUMS analysis outputs, testing existing diagnostic knowledge and discovering new knowledge that can be used to further enhance HUMS effectiveness. (Research Task 6 in reference [12]).

3 Implement and evaluate techniques for characterising shape of a signal. This information will assist diagnostic reasoning and automatic detection of behavioural features such as trends will allow their significance to be quantified.

8 References

1 Harrison, N., Baines, N. C. (1999). *Intelligent Management of HUMS Data: The use of Artificial Intelligence Techniques to Detect Main Rotor Gearbox Faults*. Study II, CAA Paper 99006.

2 Smiths Aerospace report REP1697(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Interim Report on Phase 1 of the Research Project*. May 2007.

3 Smiths Aerospace report REP1712(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Phase 2 of the Research Project: Six-Month Operational Trial*. July 2007.

4 Antoniou, A. *Digital Filters: Analysis, Design, and Applications*, New York, NY: McGraw-Hill, 1993.

5 Savitzky, A. and Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36: 1627–1639.

6 Donoho, D. Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data, Different Perspectives on Wavelets, *Proceeding of Symposia in Applied Mathematics, Vol 47, I*. Daubechies ed. Amer. Math. Soc., Providence, R. I., 1993, pp. 173-205.

7 Donoho, D. L. and Johnstone, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Stat. Assoc*. vol. 90, pp 1200-1224.

8 Page, E.S., Continuous Inspection Schemes, *Biometrika*, 41, (1954).

9 Chatfield, C. (2003) The Analysis of Time Series: An Introduction. Chapman & Hall texts in statistical science series.

10 MacGregor, J. F., and Nomikos, P. (1992), Monitoring Batch Processes, in Batch Processing Systems Engineering: Current Status and future Directions (NATO AS1 Series ?) eds. Reklaitis, Rippin, Hertacso, and Sunol, Heildeberg: Springer-Verlag.

11 Nomikos, P., and MacGregor, J. F. (1995). Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, February 1995, Vol 37, No. 1.

12 Smiths Aerospace report REP1682(2): *Proposal for Recommended Further HUMS Advanced Anomaly Detection Research Work*. August 2006.

13 Kullback. S and Leibler, R, A. (1951) On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

14 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

ANNEX D

Report on Second Operational Trial Period

Based on a report prepared for the CAA by GE Aviation Systems Limited

Table of Contents

Report

INTENTIONALLY LEFT BLANK

List of Figures

INTENTIONALLY LEFT BLANK

List of Tables

INTENTIONALLY LEFT BLANK

Glossary

5IA/5IT	5-Indicator Absolute / 5-Indicator Trend anomaly model
8IA/8IT	8-Indicator Absolute / 8-Indicator Trend anomaly model
AGB	Accessory Gearbox
CAA	Civil Aviation Authority (UK)
CI	Condition Indicator
DAPU	Data Acquisition and Processing Unit
FS	Fitness Score
HUMS	Health and Usage Monitoring System
IF	Influence Factor
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
LHA	Left Hand Accessory module
M6A/M6T	M6 Absolute / M6 Trend anomaly model
MGB	Main rotor Gearbox
OEM	Original Equipment Manufacturer
PA	Probability of Anomaly
RHA	Right Hand Accessory module
SOA/SOT	Shaft Order Absolute / Shaft Order Trend anomaly model
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring

INTENTIONALLY LEFT BLANK

Executive Summary

This report documents the results of the second in-service trial of an advanced HUMS anomaly detection system, developed for a CAA funded research programme to demonstrate the intelligent analysis of helicopter HUMS Vibration Health Monitoring (VHM) data. The goal of the programme is to improve the fault detection performance of HUMS.

The advanced anomaly detection capability, and system to implement it, were developed in Phase 1 of the CAA research programme, which also included an off-line demonstration using a database of historical AS332L HUMS data. In Phase 2 of the programme, the anomaly detection system was subjected to a six month in-service trial by Bristow Helicopters. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L fleet HUMS data being automatically transferred on a daily basis from Aberdeen. The results of Phases 1 and 2 are reported in references [2] and [3].

The CAA contract included an option for a six-month extension to the in-service trial of the HUMS anomaly detection system. Building on the success of the initial trial, four additional research tasks were carried out to research and develop further advanced HUMS data analysis capabilities prior to the second trial period. These included a data re-modelling task, and the introduction of new Probability of Anomaly (PA) and Influence Factor (IF) outputs, with a global anomaly alerting threshold applied based on the PA value. The results of these additional research tasks are presented in reference [4].

The second in-service trial period evaluated the enhanced advanced anomaly detection system incorporating the updated anomaly models and the new PA and IF outputs. To minimise Bristow's workload associated with the trial, this time GE Aviation also provided an analysis service, sending Bristow daily fleet health status reports. The second trial phase commenced at the beginning of January 2008, with the formal trial period completing on 30 June, however the trial was continued informally until 19 December 2008 while two further research tasks were completed.

The second trial period has provided further evidence to that contained in reference [3] demonstrating the effectiveness of the advanced HUMS anomaly detection system. The system has clearly highlighted anomalous HUMS VHM Condition Indicator (CI) trends associated with both aircraft problems and HUMS instrumentation faults that are more difficult to detect by traditional HUMS analysis. The two trial periods have proven the system's ability to improve the performance of HUMS. The web-based system implemented for the trial has been demonstrated to be a reliable and effective solution, and is a proven model for the wider implementation of advanced HUMS anomaly detection. Supplementing the system with an analysis service was a valuable help in minimising the operator's workload associated with the new capability.

It is recommended that a plan is now developed for implementation of the advanced HUMS anomaly detection capability on all helicopter fleets involved in UK Continental Shelf offshore support activities.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Health and Usage Monitoring Systems (HUMS), incorporating comprehensive drive train Vibration Health Monitoring (VHM), have contributed significantly to improving the safety of rotorcraft operations. However, experience has also shown that, while HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected using current threshold setting methods. Earlier research (see reference [1]), conducted as part of the CAA's helicopter main rotor gearbox seeded defect test programme, demonstrated the potential for improving fault detection performance by applying unsupervised machine learning techniques such as clustering to seeded fault test data. The CAA therefore commissioned a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841). GE Aviation was contracted to carry out this programme in partnership with Bristow Helicopters, analysing IHUMS data from Bristow's North Sea AS332L fleet.

The work under Contract No. 841 was structured as two phases. Phase I involved the research and development of an advanced anomaly detection capability to enhance the fault detection performance of HUMS. Anomaly models were built for all the AS332L rotor drive system components. These fused multiple input HUMS Condition Indicators (CIs) into a single output 'Fitness Score' (FS), indicating how well new data fit the model's view of normality. Phase 1 also included an 'off-line' demonstration of the advanced anomaly detection capability on a database of historical AS332L data, and the development of a system implementing this capability for evaluation in an in-service trial. The results of Phase 1 are reported in reference [2].

Phase 2 of the CAA research programme comprised a six-month in-service trial, conducted by Bristow Helicopters, of the advanced HUMS anomaly detection system developed in Phase I. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L IHUMS data being automatically transferred every night from Aberdeen, and Bristow Helicopters having a remote login to the server to check the anomaly detection system outputs. The results of Phase 2 are reported in reference [3].

The CAA contract included an option for a six-month extension to the in-service trial of the HUMS anomaly detection system. Building on the success of the initial trial, four additional research tasks were carried out to research and develop further advanced HUMS data analysis capabilities prior to the second trial period. A model tuning and re-modelling task benefited from other work on re-implementing and improving the modelling technique which provided an enhanced anomaly modelling facility. All the anomaly models were re-built using this improved facility. A probabilistic alerting policy was implemented, with the FS outputs from the anomaly models being converted into normalised Probability of Anomaly (PA) values with a scale of 0-1. This simplified threshold setting, with a global alerting threshold being set at a selected PA value. Finally, Influence Factors (IFs) were implemented to provide diagnostic information on the causes of an anomaly alert. The IFs are derived from an anomaly model, with one IF for each HUMS CI used to train the model, and provide an indication of how much each CI contributes to an anomaly alert. Again, they are normalised and so can be directly compared. The results of these four additional research tasks are presented in reference [4].

This report presents the results of the second in-service trial period, evaluating the enhanced advanced anomaly detection system incorporating the updated anomaly models and the new PA and IF outputs. To minimise Bristow's workload associated with the trial, this time GE Aviation also provided an analysis service, identifying any detected anomalies which Bristow should follow-up in daily fleet health status reports. Bristow then used the anomaly detection system to investigate the identified anomalies. The second trial phase commenced at the beginning of January 2008, with the formal trial period completing on 30 June, however the trial was continued informally until 19 December 2008 while two further research tasks were completed.

Section 2 of the report gives an overview of the IHUMS data analysed, the anomaly processing performed, and the web-based anomaly detection system used in the trial. Section 3 describes the trial operating procedures implemented by GE Aviation and Bristow, and the operational experience gained with the advanced anomaly detection system. A catalogue of significant trial findings, with 15 example cases, is presented in Section 4, and a statistical analysis of the anomaly detection system performance is contained in Section 5. Finally, some conclusions and recommendations are given in Section 6.

2 Advanced HUMS Anomaly Detection Processing and System

2.1 IHUMS Data Analysed

The anomaly detection system analysed IHUMS VHM data from the following assemblies in the AS332L drive train: Main rotor Gearbox (MGB), left and right Accessory Gearboxes (AGBs) and oil cooler fan, Intermediate Gearbox (IGB), and the Tail rotor Gearbox (TGB). Thirty-five drive train component analyses were defined, as listed in Table 2-1. The table also shows the aircraft sensor used to acquire the component data, and the equivalent IHUMS 'channel' (i.e. analysis number) allocation. The IHUMS performs thirty-two component analyses. However, three of the components have two gears on a single shaft (the MGB combiner gear and bevel pinion, and the left and right AGB hydraulic drive 47- and 81-tooth gears). For the purposes of anomaly modelling, each of these was split into separate components, with separate anomaly models built for each of the two gears on the shaft using the different gear mesh-specific CIs.

Table 2-1 AS332L drive train components analysed

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MGB
2	1	RH high speed input shaft	MGB
1	2	Left torque shaft - fwd end	MGB
2	3	Right torque shaft - fwd end	MGB
3	4	Left torque shaft - aft end	MGB
4	5	Right torque shaft - aft end	MGB
3	6	Combiner gear	MGB
3	6	Bevel pinion	MGB
4	7	Bevel wheel and oil pump drive	MGB
7	8	1st stage sun gear	MGB
7	9	1st stage planet gear	MGB
5	10	1st epicyclic annulus fwd (RH)	MGB
6	11	1st epicyclic annulus left	MGB
7	12	1st epicyclic annulus aft (RH)	MGB
7	13	2nd stage sun gear	MGB
7	14	2nd stage planet gear	MGB
5	15	2nd epicyclic annulus fwd (RH)	MGB
6	16	2nd epicyclic annulus left	MGB
7	17	2nd epicyclic annulus aft (RH)	MGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB

Table 2-1 AS332L drive train components analysed (Continued)

3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB
4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MGB

The CIs calculated by the IHUMS during each component analysis are listed in Table 2-2 (a more detailed description is presented in reference [2]). Not all of the indicators were used in the anomaly modelling. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data (the actual indicators used in the different anomaly models are listed in Section 2.2.2).

Table 2-2 IHUMS Condition Indicators

Condition Indicator	Description
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation (rms)
FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev (rms)
ESA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA_CVG	Signal average convergence

2.2 Data Processing for Anomaly Detection

A detailed description of the data processing performed is presented in references [2] and [4]. A brief summary of the three stages of the anomaly detection processing is given below.

2.2.1 Data Pre-Processing

Different pre-processing was applied to the input IHUMS CI data to enable two types of model to be built for each monitored component; an 'absolute' and a 'trend' model. The absolute models identify combined CI values that are anomalous in absolute terms, whereas the trend models identify anomalous combined CI trends, irrespective of the absolute values of the indicators.

The pre-processing for the absolute models was minimal, and involved the application of a two stage filter to each indicator: the first stage removed unreasonable values and the second stage applied a median filter to remove up to two successive spikes in the time series data.

For the trend models, the two stage filter for the absolute models was used, and then a 'moving median difference' algorithm was applied. Following each new acquisition, the median of the time history is re-calculated and subtracted from the newly acquired value to provide a normalised value. This technique reduced the impact of early post-installation trends, and also small step changes due to maintenance, because the normalised value would gradually recover back to the median base line level. Although there was some distortion of the time-history, this simple approach worked well. Its weakness was that, while it achieved the normalisation objective, it acted as a difference calculator and could not distinguish between step changes, short duration changes, and developing trends.

2.2.2 Anomaly Models

Using the pre-processed data, anomaly models were constructed for each drive train component. The models were constructed from a training database of historical data, and adapted so that they eliminated the effects of suspected anomalies existing in the training data. The adaption process is described in reference [4]. No data is actually removed in this process, and the resulting models can make predictions on all data, including that associated with suspected anomalies. Models were built to represent absolute data behaviour (between-component variability) and trend behaviour (within-component variability). These models were sophisticated statistical representations of the data generated from in-service experience, fusing sets of CIs (i.e. vibration features) to reduce a complex data picture into a single time history called a 'FS' trace. The FS measured the degree of abnormality in the input data and mirrored the shape of any significant data trends. It represented a 'goodness of fit' criterion, indicating how well data fitted the model of normality, i.e. the corresponding anomaly model. Therefore the FS exhibited a decreasing trend as the data became increasingly abnormal.

With the exception of two drive train components that did not include a gear on the monitored shaft, every component had anomaly models constructed from two sub-sets of the CIs listed in Table 2.2:

- '8-indicator' models, comprising the following eight indicators: {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.
- 'M6' models, comprising the following two indicators: {ESA_M6*, ESA_WEA}, (ESA_WEA is a derived indicator, calculated as the ratio ESA_SD/SIG_SD).

The indicators were split to reduce the potential for one indicator to be masked by another; the criterion for this split was derived from engineering knowledge of the indicator design and characteristic behaviour. The oil cooler fan and dual bearing module were monitored by only two models – absolute and trend models comprising the following '5-indicator' set: {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

For the MGB input shafts and TGB output, additional 'Shaft Order' (SO) models were provided. These models comprised the following two indicators: {FSA_SO1, FSA_SON, where N =2}.

Using the different forms of pre-processing, 'absolute' and 'trend' models were built for each set of indicators. Therefore the majority of drive train components were monitored by four anomaly models – '8-Indicator Absolute' and 'Trend' models, and 'M6 Absolute' and 'Trend' models. The exceptions to this were two components (the oil cooler fan and dual bearing module) monitored by two models, and three components (the MGB input shafts and TGB output) monitored by six models.

2.2.3 **Anomaly Model Outputs**

The FS output from the anomaly models contains useful trend information, however the absolute values vary between models. The anomaly detection output is enhanced if the measure of what is anomalous is normalized across models. Normalization allows model outputs to be compared, and such a measure could be fed into a secondary process such as automated reasoning to assess the nature of an anomaly. In addition, it would facilitate data mining of anomalous patterns in the search for new knowledge. A 'Probability of Anomaly' (PA) measure was therefore introduced, which is a normalized probability measure that ranges between 0 and 1 (see reference [4] for more information). For each model there is a PA distribution which is an extreme value distribution. An FS value is passed to the PA distribution and a PA value is returned. Most FS values will return a PA of 0 because most acquisitions will be normal. The PA values are used to define an alerting threshold, which can be common to all models and components (the FS equivalent of this PA threshold will be different for each model).

'Influence Factors' (IFs) are a type of model prediction that can provide diagnostic information about an anomaly model and its inputs (i.e. HUMS CIs). There are different types of IF, each type producing a different view on a model or input variable. An acquisition will generate a single IF predicted value for each CI used to train a model. IF time histories provide information regarding the influence of HUMS CIs on the fused FS but, unlike CIs, IFs are normalized and can be directly compared. In addition to further enhancing the usability of the HUMS anomaly detection system, like the PA values the IFs will also facilitate developments such as the application of automated diagnostic reasoning and data mining. Although multiple types of IF are generated for GE Aviation's internal use, with each type being designed to provide different information about a model or an input feature, only one type of IF is presented to the operator. This is the one that most closely matches the trends the operator would observe in the HUMS CI data.

2.2.4 **Anomaly Alerts**

An alerting threshold is defined using the PA values, which greatly simplifies the threshold setting process. A default PA threshold is normally applied that is common to all models and components (the FS equivalent of this PA threshold will be different for each model). The default threshold for an individual model can, however, be changed if it is considered that the alert rate is too high or too low. A global PA threshold of 0.90 was used for the trial.

An alert is triggered when M out of the last N points are above the threshold. The default values for M and N are 3 out of 4.

2.3 Anomaly Detection System Implementation

The system implementation for the anomaly detection trial is illustrated in Figure 2-1. This is a web-based system, hosted on a server at GE Aviation. HUMS data files are automatically copied every night from the HUMS ground station to a secure web portal at Bristow Helicopters in Aberdeen. The files are then automatically downloaded to GE Aviation in Southampton, imported into the anomaly detection system's data warehouse, and analysed. Bristow have a remote secure login to the system to view results at any time via a web browser.

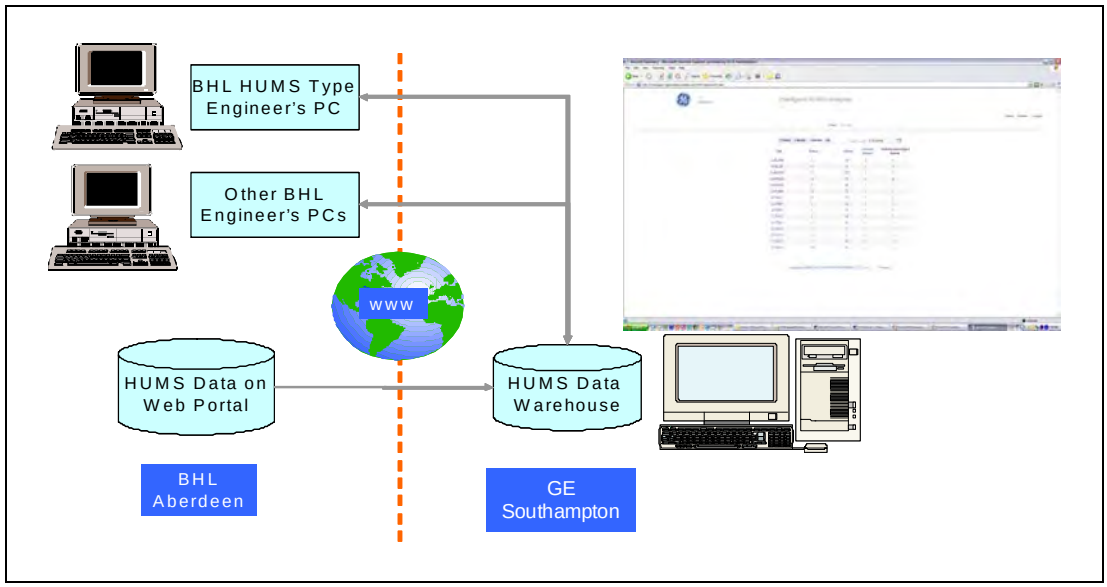

Figure 2-1 Trial system implementation

2.4 Anomaly Detection System Website

This section provides an overview of the system website that has been created for the trial. The website provides a flexible, but simple to use, interface which facilitates rapid access to any required information. The system separately tracks current alerts, and current unacknowledged alerts. Charts can be annotated with notes, and anomaly model outputs can be rapidly accessed from listings of alerts or notes, or via a conventional component hierarchy. A reports page allows the viewing of reports on anomalies that are being tracked by the system.

After logging into the system, the user is presented with a fleet view showing the number of aircraft currently operating, alerts (total, current, and unacknowledged), and user-entered notes (Figure 2-2). The user can limit the view of the data to one week, one month, three months, or can view the complete history. There is also a calendar that allows the data to be viewed as it was at some previous date. This option is necessary to access historical data relating to components that have been replaced because the system's 'latest date' view, which is the default, normally includes only those components that are currently fitted to the fleet.

For routine daily checks of new alerts across the fleet, the user would normally click on the alerts total to view new alerts generated from the anomaly processing of fleet data downloaded from the previous day's flying (Figure 2-3).

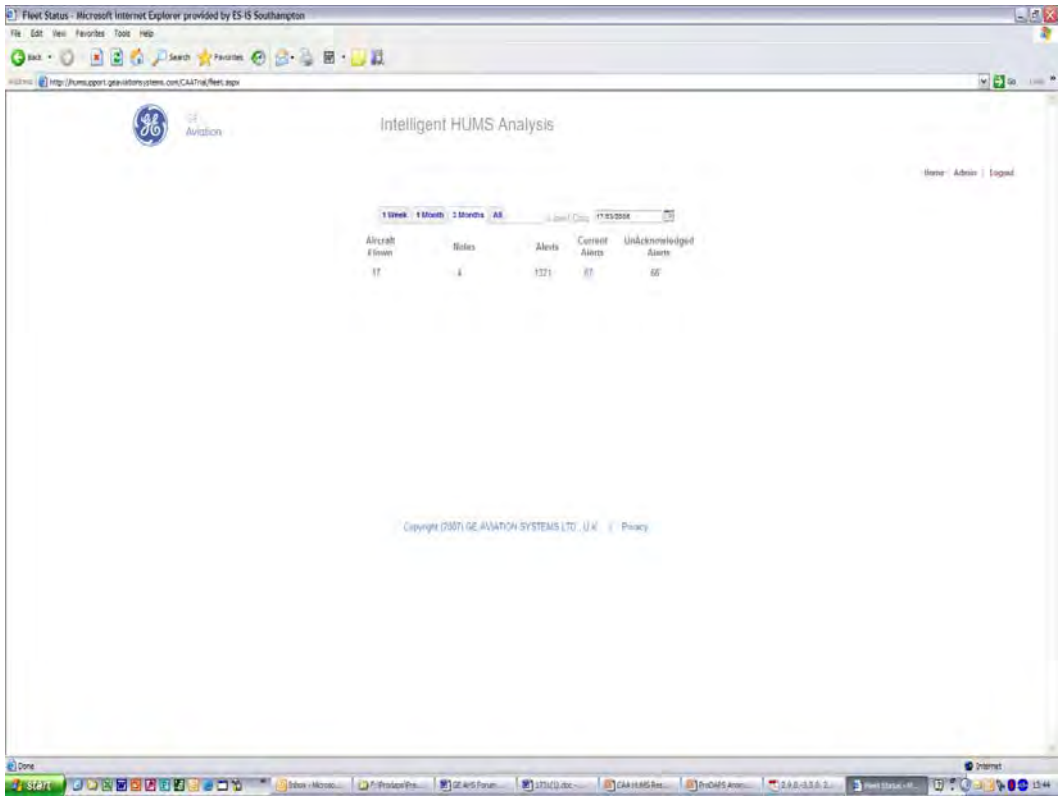

Figure 2-2 Fleet status page

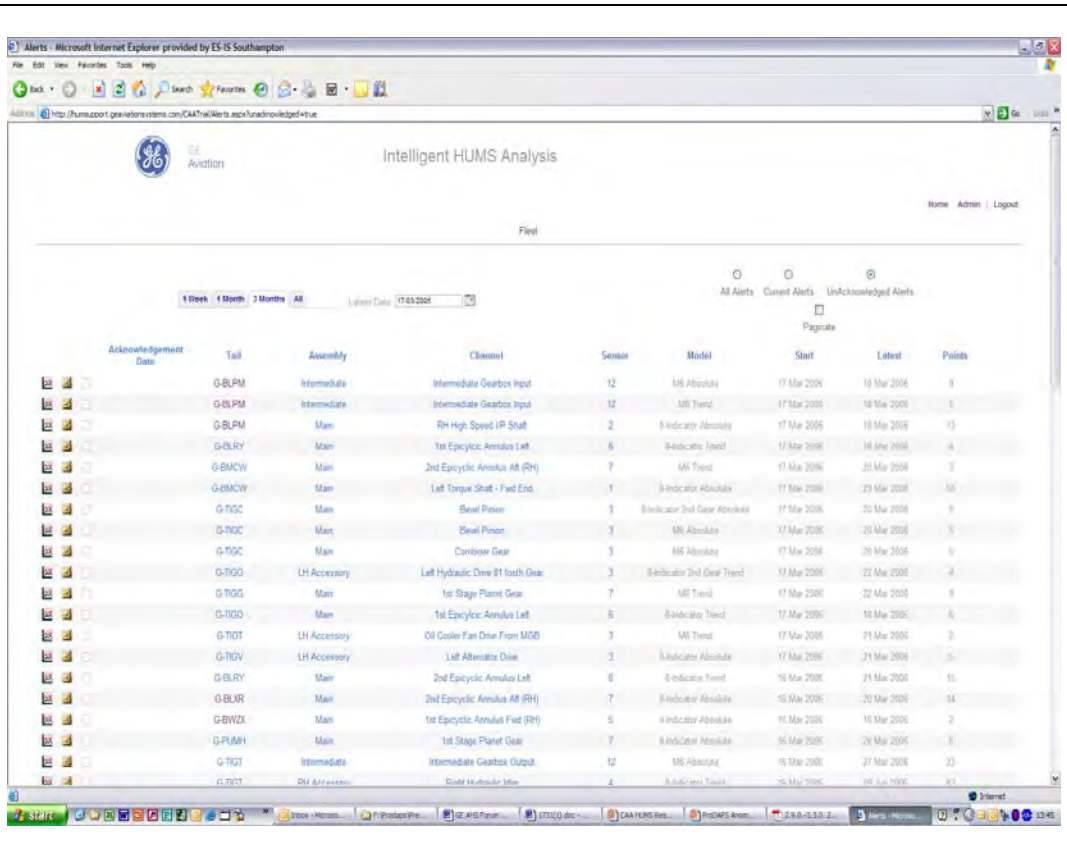

Figure 2-3 Alerts page, showing unacknowledged alerts

By clicking on the chart icon at the left hand side of an alert record, the user can view the anomaly model Probability of Anomaly (PA) and FS traces for the component in alert (Figure 2-4). These traces are annotated with alerts and any notes the operator has entered to document items such as the findings from any follow-up investigation or maintenance action (a note appears as a yellow triangle on the trace display). A maintenance action causing a step change in some of the HUMS CIs could trigger a trend model alert. Therefore a facility is provided to reset the trend pre-processing, which re-baselines the trend model data and clears the alert.

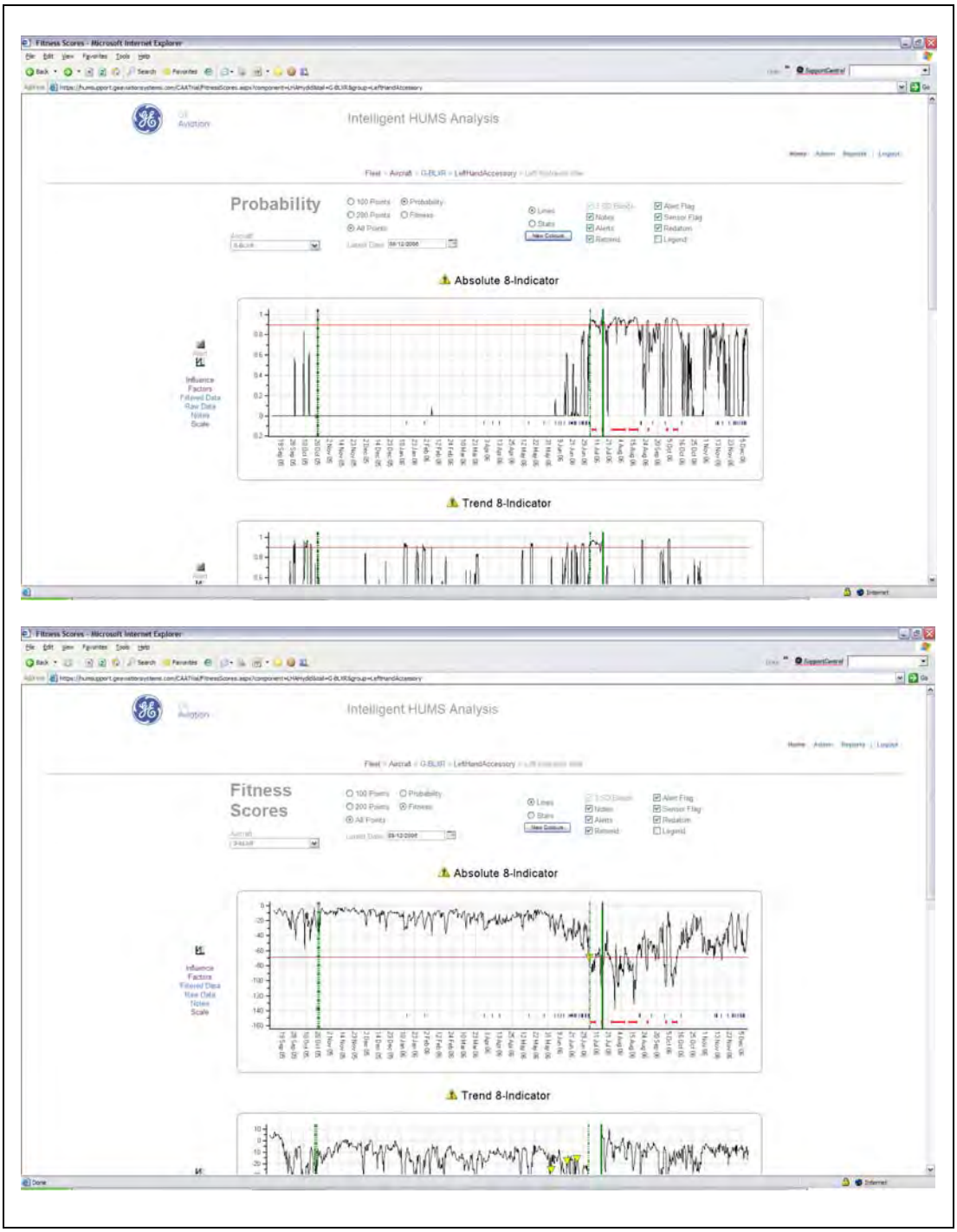

Figure 2-4 PA and FS traces

The user can obtain anomaly model diagnostic information by viewing Influence Factors (IFs) to identify which HUMS CIs are driving anomaly indications (Figure 2-5). An IF is calculated for each HUMS CI included in an anomaly model.

Figure 2-5 IF and HUMS CI displays

For further investigation, the user can drill down to either the filtered or raw HUMS CIs. For those indicators used to build models, the indicator plots show the upper and lower three standard deviation bands that have been derived from the anomaly model. These bands are different to those that would be computed directly from the fleet data because the anomaly models have been adapted to reduce the impact of outlying data.

It is also possible to view IF and HUMS CI data for a fleet of aircraft either as multiple trend plots or as statistical plots (Figure 2-6). Once an investigation and any associated follow-up action has been completed the user can acknowledge an alert and, if appropriate, enter a note.

Figure 2-6 Fleet data displays

The user can view alerts and notes by aircraft, assembly and component, and also drill down to the PA/FS, IF and HUMS CI traces via this component hierarchy. A display of user-entered notes is presented in Figure 2-7. Again, by clicking on the chart icon in a note record, the user can view the data traces to which the note applies.

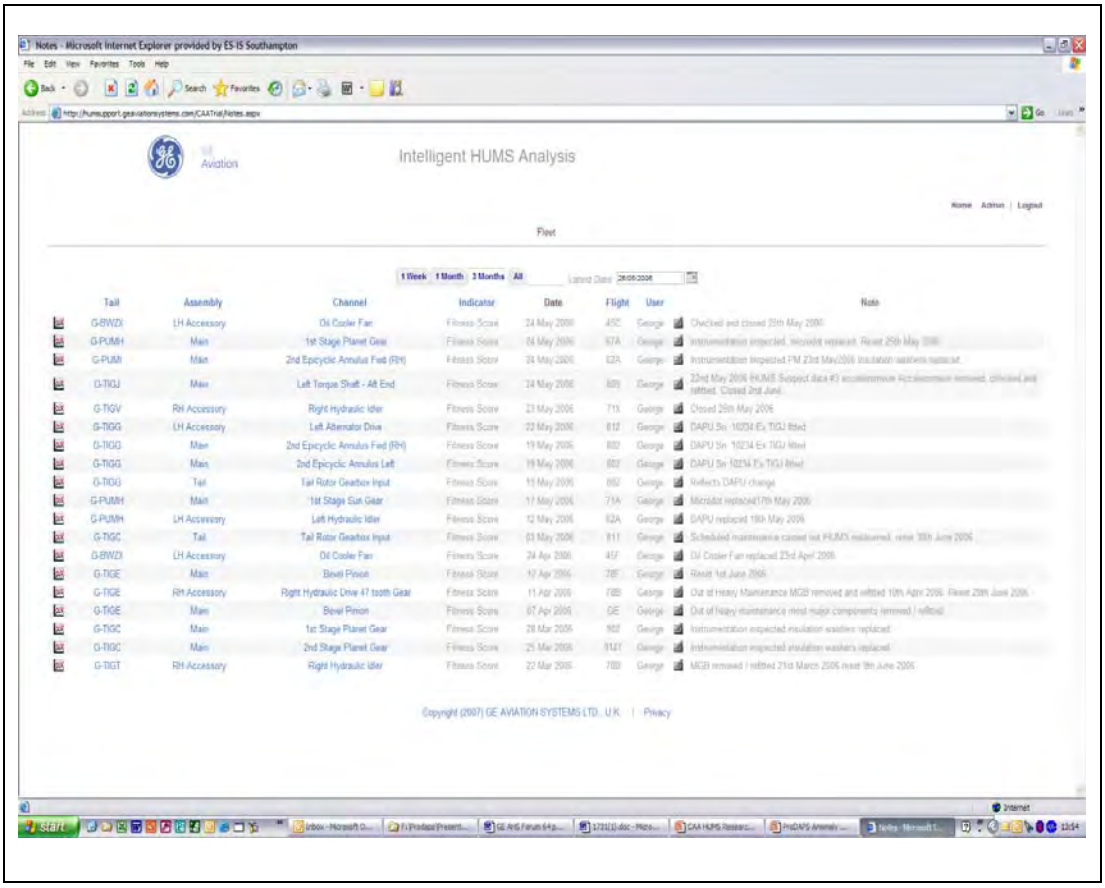

Figure 2-7 Notes display

The user can select 'reports' to view daily fleet health status reports created as part of the analysis service. A typical report is shown in Figure 2-8. Alerts are grouped into tracked items. New report entries appear in red, and entries carried forward from the previous report appear in black. Again, by clicking on an alert, the user is taken directly to the anomaly model PA/FS traces for the component in alert.

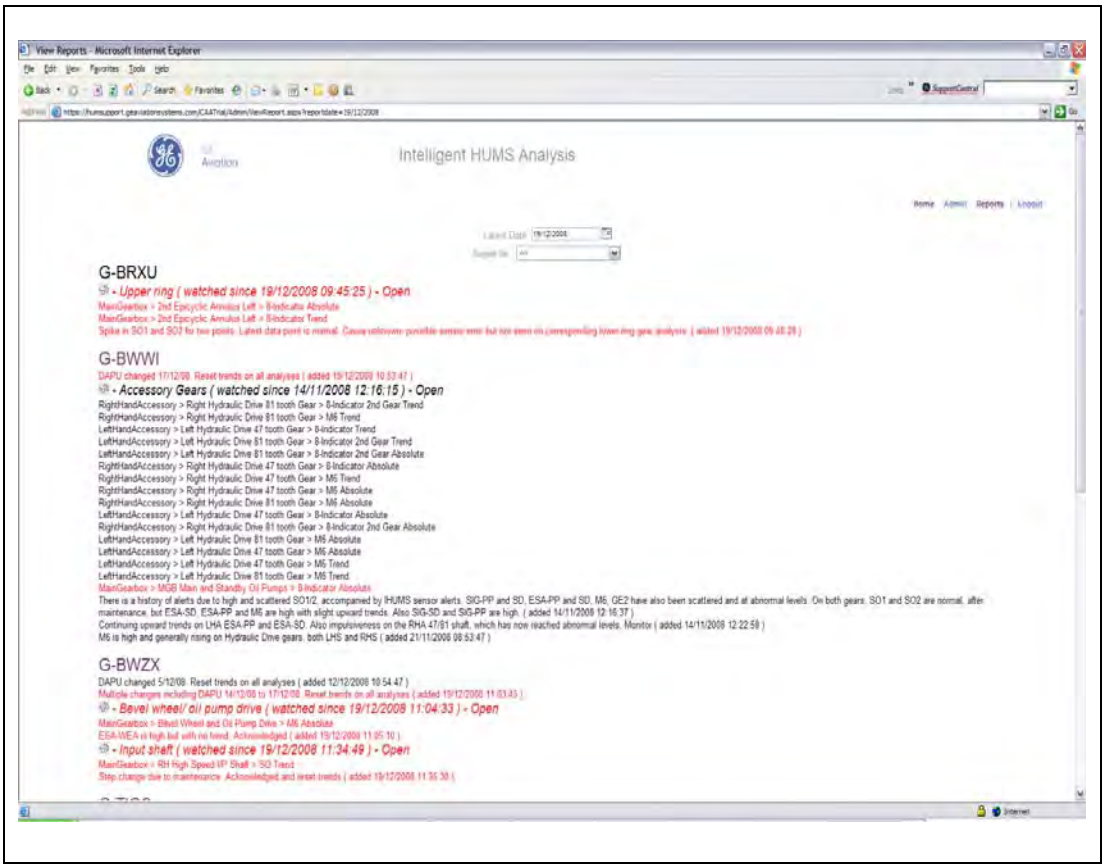

Figure 2-8 Health status report

Figure 2-9 provides a key to the documentary information shown on the anomaly detection system charts that are included in Section 4 of this report. The chart shown is a FS trace.

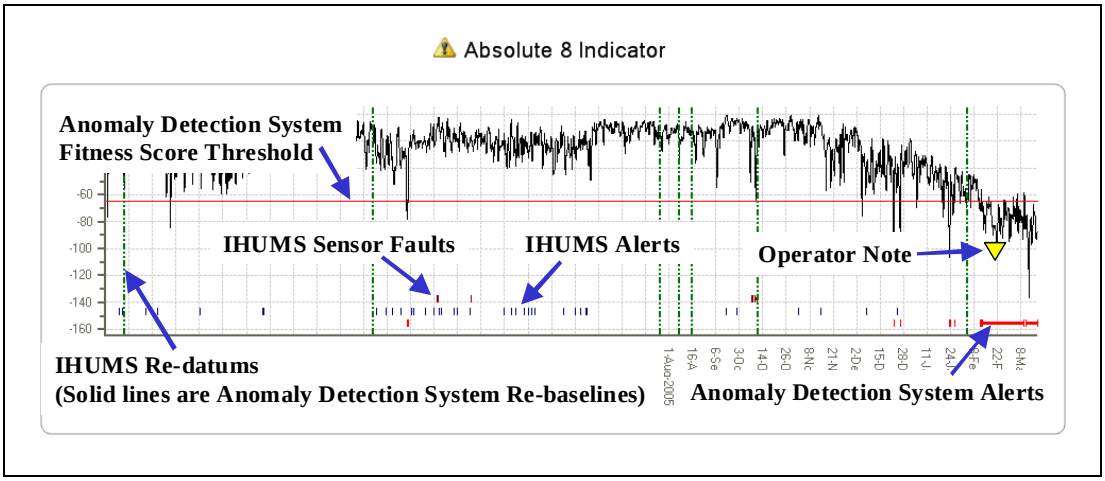

Figure 2-9 Key to documentary information shown on a chart display

3 Advanced HUMS Anomaly Detection Trial

The first six-month in-service trial of the advanced HUMS anomaly detection system on Bristow Helicopters' Super Puma fleet took place between May and November 2006. This trial formed part of the main CAA HUMS research programme, and the results were presented in reference [3]. Bristow's HUMS engineer operated the system, reviewing anomaly alerts triggered across the fleet on a daily basis.

As a result of the success of the trial, a number of additional research tasks were undertaken to further enhance system capabilities. These included rebuilding all the Super Puma anomaly models using a re-implemented and refined modelling facility, and also adding Probability of Anomaly (PA) and Influence Factor (IF) outputs to the system. These new capabilities are described in reference [4]. Bristow continued using the anomaly detection system during this period.

A second formal six-month in-service trial of the updated system took place from 7 January to 30 June 2008, however the trial was continued until 19 December 2008 while GE Aviation completed work on two final research tasks involving data mining and automated reasoning technology demonstrations. For this second trial period GE Aviation monitored the outputs from the anomaly detection system and provided an analysis and reporting service to Bristow. For the period up to the end of June GE Aviation provided daily fleet health status reports to Bristow during the normal working week. Reports were then produced on a weekly basis until the trial finished in December. Bristow still had full access to the system during this time, and used it to investigate anomalies identified by GE Aviation as requiring follow-up action.

The reason for trialling an analysis and reporting service in addition to the advanced HUMS anomaly detection system was the recognition that some helicopter operators have limited manpower and skills for HUMS support activities. Irrespective of how effective the anomaly detection system is, this still represents another system that must be regularly checked, and many operators are already having to support a number of different HUM systems. Some operators might therefore prefer to access the new capability through a service rather than an 'in-house' system.

3.1 Trial Operating Procedures

3.1.1 System Operated by Bristow (First Trial Period)

Bristow's HUMS engineer checked the anomaly detection system for new anomaly alerts and followed up the alerts by performing some or all of the following tasks:

1 Determined which IHUMS CIs had triggered the alert.

2 Viewed the data associated with the component in the IHUMS ground station, and looked for trends or abnormal indications in the IHUMS graphical displays.

3 Checked the IHUMS re-datum history for any actions that could correlate with the alert occurrence.

4 Searched the aircraft maintenance database for any maintenance actions or component replacements that could be associated with the alert.

5 Plotted IHUMS raw and enhanced vibration signal data to look for suspect instrumentation faults, and plotted IHUMS linear FFT data to check for abnormal peaks at certain frequencies or shaft orders.

6 When appropriate; acknowledged the alert, reset the trend pre-processing, or input a note on the relevant anomaly chart with reported findings.

If the data investigation suggested a problem on the aircraft (most commonly related to the IHUMS instrumentation), a maintenance task would be raised for investigation and/or rectification of the problem.

3.1.2 Analysis Service Provided by GE Aviation (Second Trial Period)

To maximize the value of the analysis service, Bristow provided GE Aviation with access to maintenance records for the Super Puma fleet in its web-based IFS (Industrial Financial Systems) maintenance system. This enabled GE Aviation to view maintenance on tracked parts and identify component replacements, faults and rectification actions, correlating these with anomaly alerts. The maintenance database was examined daily. Any gearbox changes were noted, and applied to the database as a component change. A check was also made for other relevant maintenance such as sensor changes, and removals/changes of components attached to gearboxes (e.g. engine, oil cooler fan, rotor head, drive shafts), and trends were reset where necessary.

A GE Aviation diagnostic engineer checked the anomaly detection system for new anomaly alerts and also reviewed the latest data for existing alerts. GE Aviation then:

1 Used the IF and HUMS CI displays to determine which CIs were driving any alerts.

2 Acknowledged those alerts that were considered not to be potentially fault related. Alerts that GE wished to bring to the attention of Bristow were left unacknowledged.

3 Searched the aircraft maintenance database for any maintenance actions that could be associated with the alert.

4 Reset any trends where step changes in data (e.g. associated with maintenance actions) had caused spurious trend model alerts.

5 Prepared a daily fleet health status report for Bristow, identifying components on particular aircraft with anomaly alerts that could be potentially fault related, and that should be brought to Bristow's attention. A potential fault could be related to the monitored component, or to the IHUMS instrumentation itself. The report separately identified on-going alerts for which no new information was available, and new alerts, or new information on existing alerts. This enabled Bristow to quickly identify new information arising from the previous day's IHUMS downloads.

When carrying out the follow-up investigation, Bristow still performed some of the tasks defined in Section 3.1.1. Both GE Aviation and Bristow used the website notes facility to record comments on anomaly alerts and the results of follow-up investigations.

The daily reports were initially MS Word documents. However, following the development of a website reporting capability on another programme, these were created on, and accessible from, the system website. A notification email was sent to Bristow when a new report had been added to the website.

For all trial periods a GE Aviation website administrator assigned website logins, with associated roles and responsibilities defining access permissions for each user.

3.2 Trial Operational Experience

Comments here apply to the second formal six-month in-service trial period starting in January 2008, together with the trial extension period to December 2008.

The only changes to the system User Interface from the first trial period were the addition of the new PA and IF displays providing new anomaly assessment and

diagnostic information. The only functional enhancement made to the system during the trial period was the addition in early November 2008 of the new web-based reporting functionality that had been developed on another programme.

The system website was considered to be very user friendly, with simple and intuitive displays and navigation facilities. The optimised structure of the system's data warehouse facilitated the rapid retrieval of large quantities of data. The fleet data displays enabled quick comparisons of data from an aircraft under investigation with the rest of the fleet. The daily fleet health status reporting facility enabled easy viewing of the daily reports and identification of new information contained in these, with hyperlinks enabling rapid access to the associated data.

The analysis service and associated daily fleet health status reports were effective in reducing Bristow's workload associated with the trial, enabling Bristow's effort to be concentrated on following up the anomalies that had been flagged by GE Aviation. Where possible, diagnostic information was included in the reports to suggest the cause of an anomaly, which was most frequently an IHUMS instrumentation problem. Again this helped to minimize the effort required by Bristow to investigate identified anomalies.

For a limited period of time during the trial Bristow conducted an evaluation of a new version of the IHUMS ground station software on part of the Super Puma fleet. There were no changes to the files containing VHM data, however while two versions of the IHUMS software were running concurrently it was necessary to copy two sets of data files to the web portal for downloading by GE Aviation. This required some changes to the data import process. The files were treated as coming from two separate aircraft databases, which were then merged in the data warehouse. GE Aviation also added a capability to automatically detect the movement of aircraft between databases, so that there would be no requirement for further changes as aircraft transitioned between the different versions of the IHUMS ground station software.

The reliability of the nightly automated data download from Aberdeen was extremely good, with no failures occurring. The reliability of access to the web server at GE Aviation was also good, with very few occasions when the site was down for more than a few minutes (an operationally insignificant period of time). The one exception to this was a hardware failure of the server at GE Aviation in Southampton that occurred in October 2008. The failure was the result of activities during the site's switchover to GE Aviation's IT infrastructure that had caused some significant disruption. Due to the high workload of IT personnel at this time, and the fact that the formal six-month trial period had already been completed, the failure resulted in the trial system being unavailable for a month. However, no data was lost and any issues arising when the system was down would have been picked up in the first report produced after it was reinstated.

4 Catalogue of Significant Findings

This section presents a selection of a further 15 example cases to provide additional evidence of the performance of the anomaly detection system to that given in reference [3]. Twelve of the cases are from the second trial period, and the remaining three are from the period of operations in 2007 between the two in-service trials when Bristow continued to use the system while GE Aviation carried out additional research tasks. Although these three cases occurred before the anomaly detection system was updated to include the re-built anomaly models and the PA and IF outputs, all historical data was reprocessed by the updated system prior to the start of the second trial. Therefore the system outputs for these cases presented in the report are the same as for 12 cases from the second trial period.

To provide consistency with reference [3] the example cases are described under the same category headings as used in that document:

1 Cases where anomaly detection has identified a fault not seen by the existing HUMS.

2 Cases where anomaly detection is corroborated by existing HUMS indications.

3 Cases where anomaly detection has failed to identify a fault that was seen by the existing HUMS.

4 Cases where anomaly detection has identified an existing HUMS false or premature alert.

5 Cases where anomaly detection has generated a false or premature alert.

6 Cases of anomaly indications with a currently unknown outcome.

Nine of the 15 cases presented here have been placed in Category 2. However, in most of these cases, the anomaly detection system gave an earlier and clearer indication of anomalous IHUMS CI trends, often flagging these up on more components than did the IHUMS.

In this analysis, faults are considered to include both aircraft component faults and HUMS instrumentation faults. In the majority of cases it is not possible to positively confirm the presence of a specific fault. For the purposes of categorisation, a correct indication is considered to have occurred if:

- The operator has determined that maintenance action is required (e.g. to replace an accessory module, or a HUMS accelerometer).
- The maintenance action performed has resulted in data returning to normal levels.

The cases are illustrated with a series of charts of PA and FS values, IFs, IHUMS CIs, and fleet views of CIs, that are taken directly from the anomaly detection system website described in Section 2.4. In the fleet views of CIs, data for the aircraft to which the example applies is always shown in black. Most charts of HUMS CIs show the median filtered data that is input to the anomaly models.

4.1 Cases where Anomaly Detection has Identified a Fault not Seen by the Existing HUMS

Example 1: G-BLXR LHA Oil Cooler Fan Drive

The '8-Indicator Absolute' model for the LHA gearbox oil cooler fan drive was in continuous alert from March 2007 (Figure 4-1). The '8-Indicator Trend' and 'M6 Absolute' models triggered alerts in October 2007, and went into almost continuous alert from late January 2008 (Figure 4-2 and Figure 4-3). Anomalies were also detected on other shafts monitored using sensor 3, and the RHA right hydraulic drive 81T gear (monitored from sensor 4). There were no IHUMS alerts, however data levels were anomalous for the indicators SIG_PP, SIG_SD, ESA_PP, and ESA_SD on these shafts. Most notably, there were very clear rising trends in ESA_SD and ESA_PP, coupled with an increase in ESA_M6, on the LHA oil cooler fan drive from January 2008 (Figure 4-4 to Figure 4-6).

Sensor maintenance was performed on 4 February 2008, however this had no effect on the rising trends. The LHA gearbox was rejected on 13 March 2008, and the data levels returned to normal. Although the IHUMS had not alerted, Bristow involved Meggitt Avionics in the investigation of this case, and Meggitt agreed with the rejection decision.

Bristow sent the IHUMS data to Eurocopter together with the LHA gearbox, and requested a strip report. Unfortunately, by the end of the trial no gearbox strip information had been received. This case again highlights the lack of an effective process for obtaining feedback from the OEM on component condition to correlate with HUMS alerts. (The first example was an MGB rejected for metal contamination, and on which rising trends were detected on the second stage epicyclic ring gear, documented in reference [3]).

The fact that the IHUMS did not alert to the clear rising trends in the data is probably due to the multiple re-datums that took place following sensor and other maintenance (these are shown as dashed green lines in the figures). A re-datum action results in the IHUMS recalculating its thresholds based on the most recent data. This case clearly highlights the greater reliability and robustness of the anomaly modelling approach in alerting to abnormal data trends.

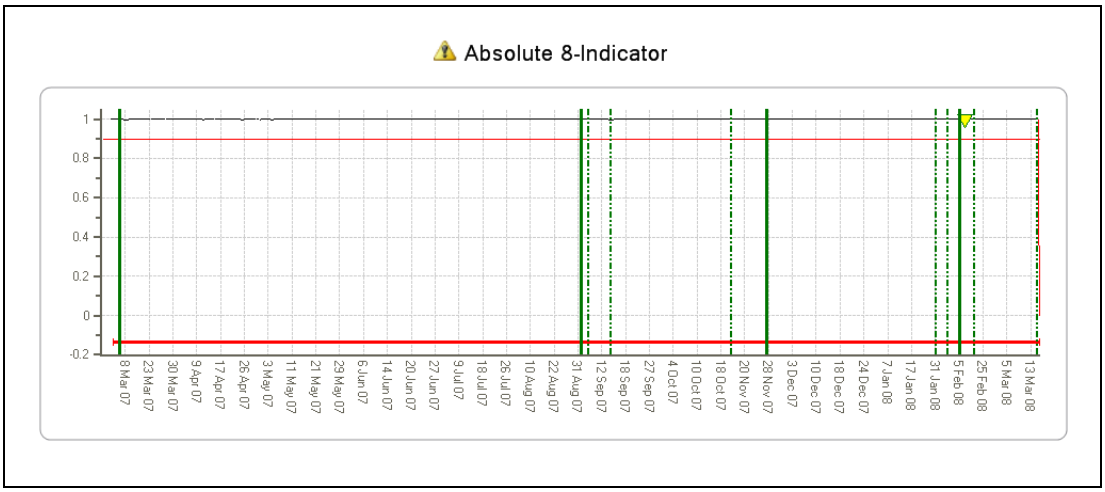

Figure 4-1 G-BLXR LHA oil cooler fan drive – '8-indicator' absolute model PA

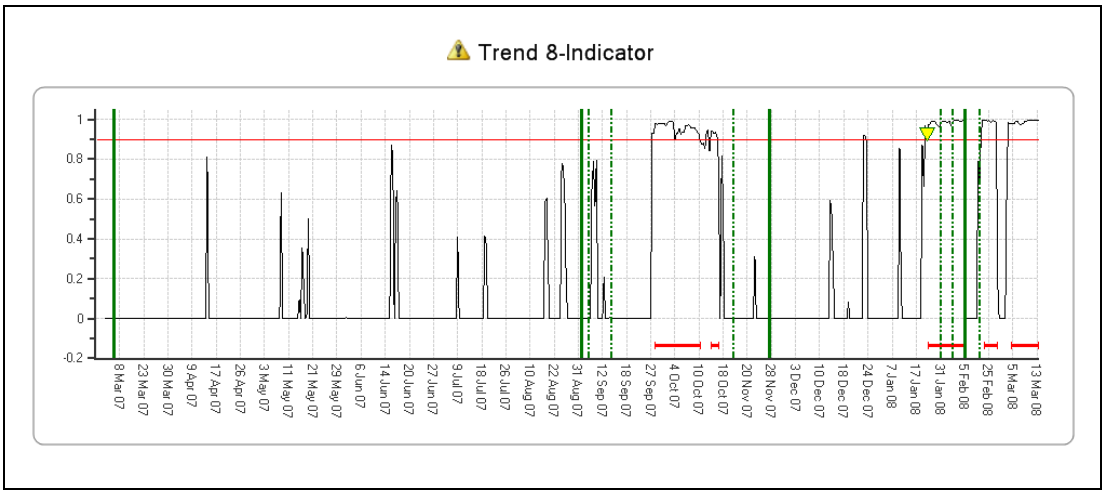

Figure 4-2 G-BLXR LHA oil cooler fan drive – '8-indicator' trend model PA

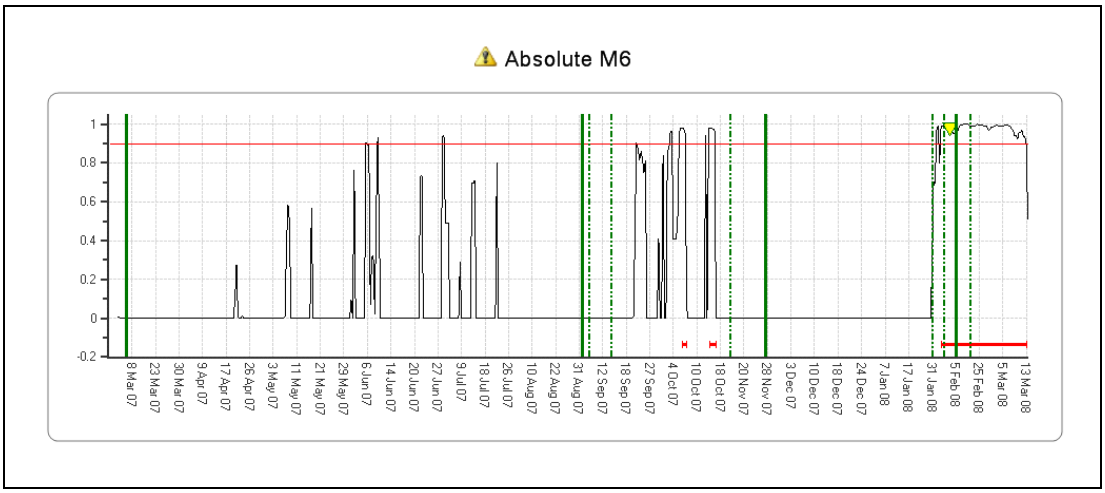

Figure 4-3 G-BLXR LHA oil cooler fan drive – M6A model PA

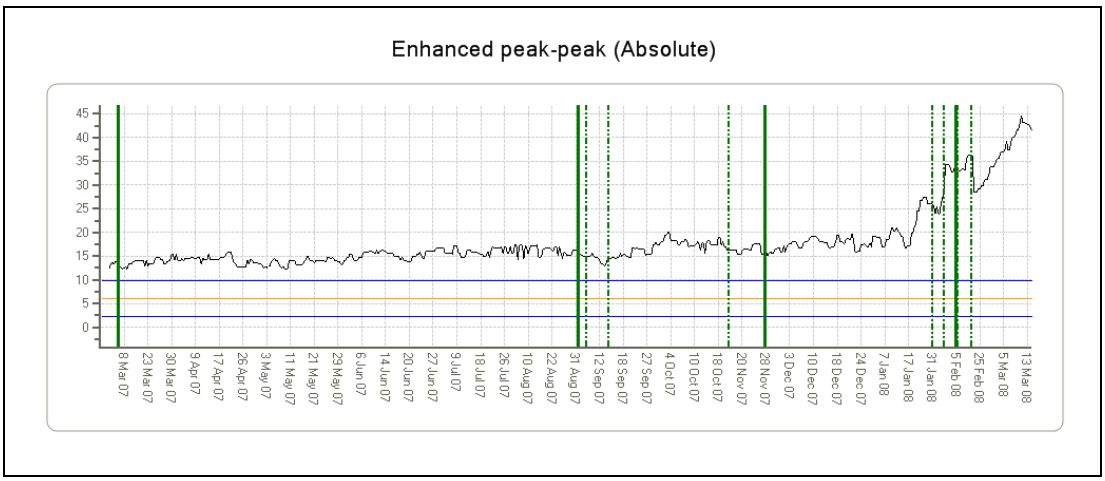

Figure 4-4 G-BLXR LHA oil cooler fan drive – ESA_PP

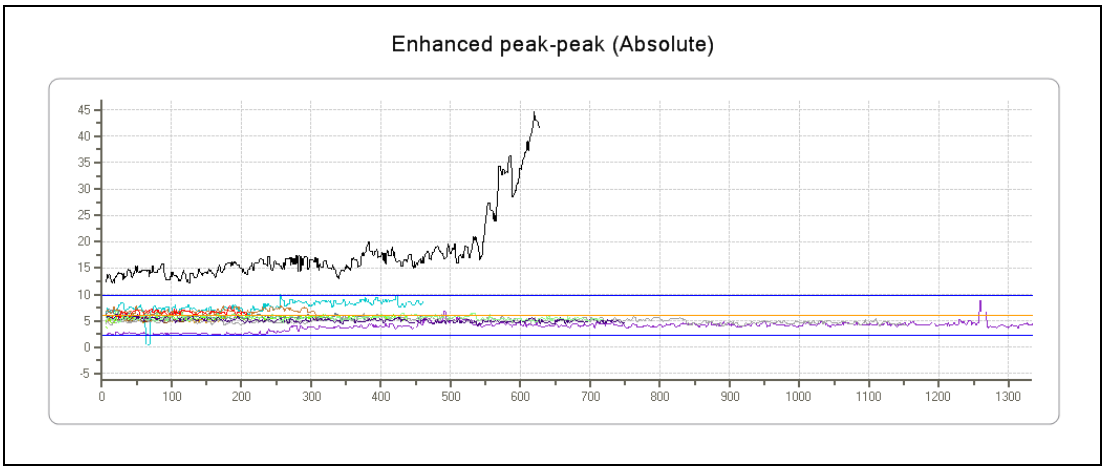

Figure 4-5 G-BLXR LHA oil cooler fan drive – ESA_PP – fleet view

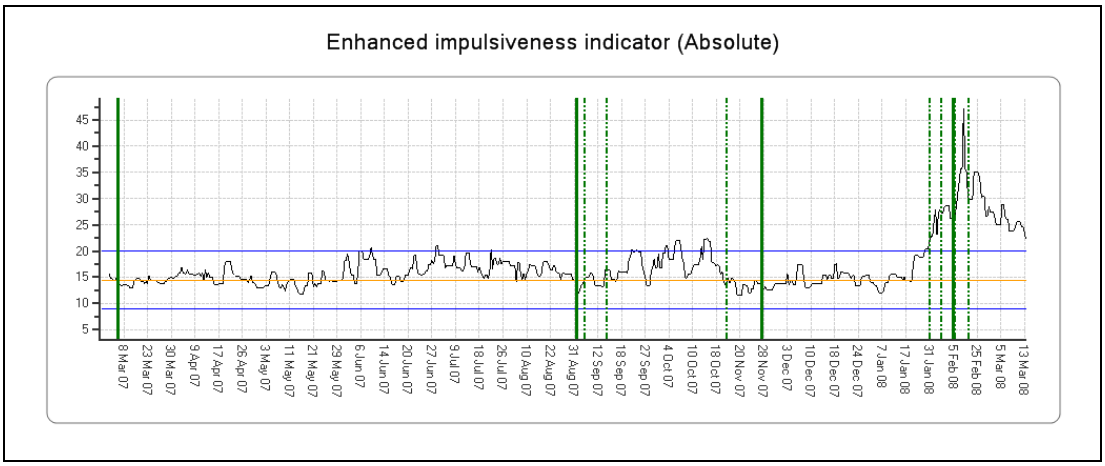

Figure 4-6 G-BLXR LHA oil cooler fan drive – ESA_M6

4.2 Cases where Anomaly Detection is Corroborated by Existing HUMS Indications

A total of nine example cases have been included in this category. Examples 7 to 9 are from the interim period of operations in 2007.

Example 1: G-BLPM LHA Left Hydraulic Idler

The left hydraulic idler in the LHA gearbox of G-BLPM consistently generated 'M6 Absolute' models alerts from June 2007 (Figure 4-7 shows the PA trace, and Figure 4-8 shows the related FS trace). The IHUMS also generated multiple alerts, as indicated by the blue markers above the red anomaly model alert bars in the figures. These alerts were triggered by a rising trend in ESA_M6 values, followed by a period of increased variability, with M6 values in the range 80-100 (Figure 4-9 and Figure 4-10).

The data was similar to previous fault cases, for example a LHA gearbox rejection on G-BLXR in December 2006 due to high ESA_M6 and ESA_PP values on the left hydraulic idler (reference [3]). Bristow placed the gearbox on close monitor, but noted that in other similar cases of gearbox removals Eurocopter had stated that these could have remained in service until their normal retirement life. The aircraft left the North Sea fleet in March 2008, and Bristow reported that most components would have been replaced at that time, therefore no further information on the gearbox anomaly would be available.

The anomaly modelling and IHUMS thresholds correctly alerted to clear trends in HUMS CIs, however this is another example of a case where the cause of the trends was not determined, and a gearbox strip report would have been very beneficial.

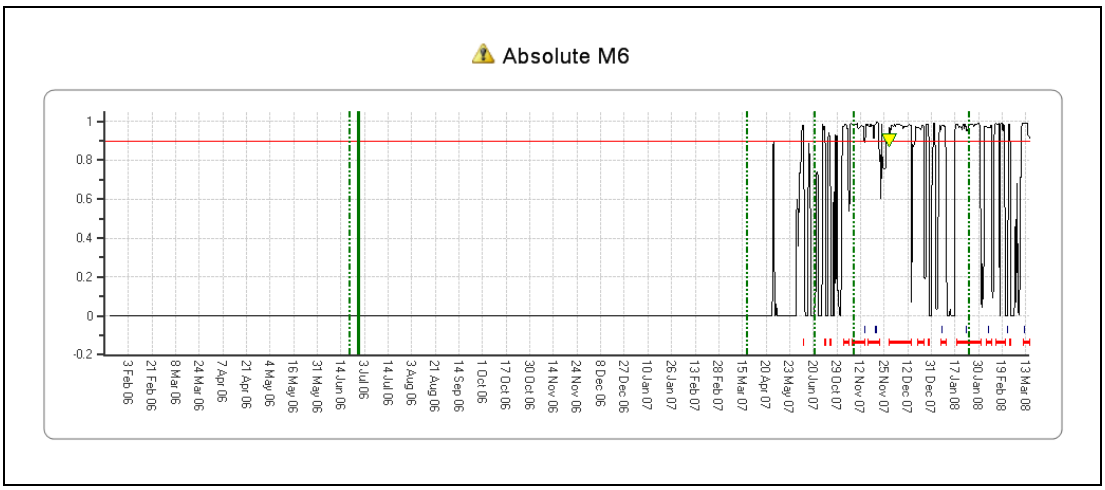

Figure 4-7 G-BLPM LHA left hydraulic idler – M6A model PA

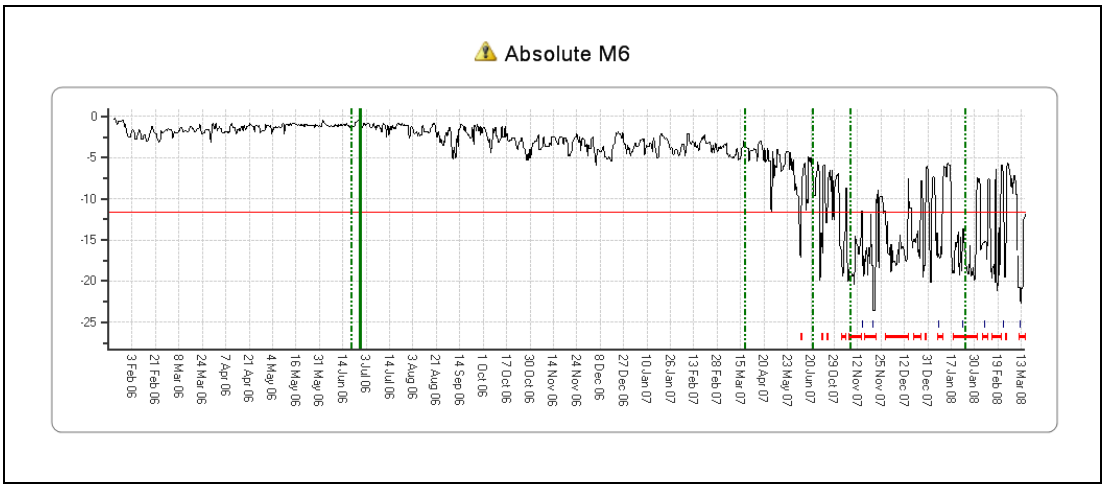

Figure 4-8 G-BLPM LHA left hydraulic idler – M6A model FS

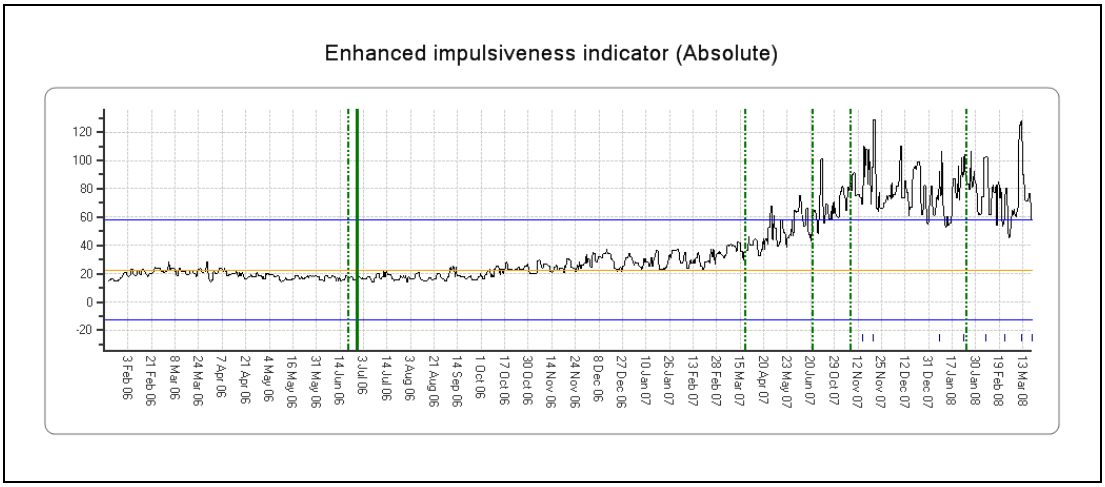

Figure 4-9 G-BLPM LHA left hydraulic idler – ESA_M6

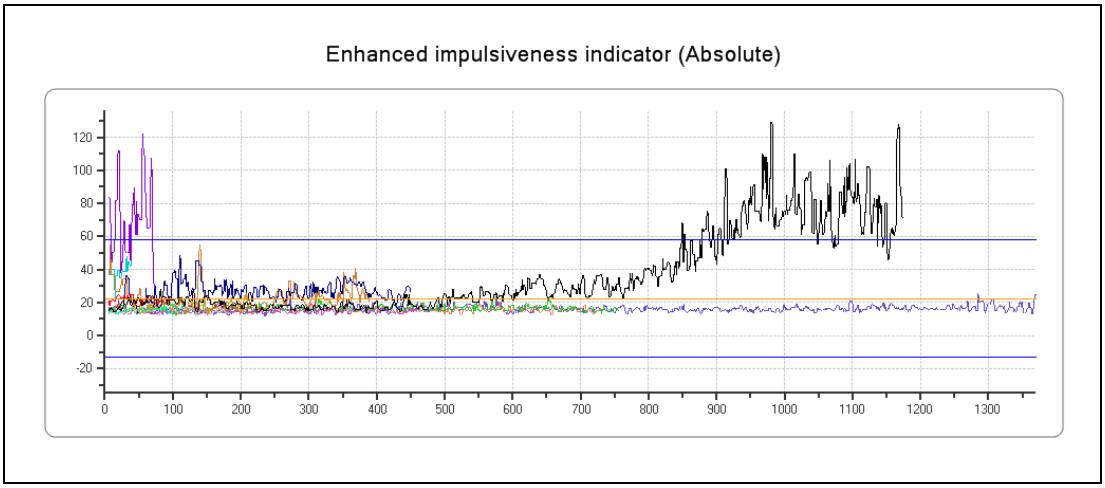

Figure 4-10 G-BLPM LHA left hydraulic idler – ESA_M6 – fleet view

Example 2: G-BWWI LHA Left Hydraulic Drive 47/81T Gears

The LHA gearbox left hydraulic drive 47/81T Gears consistently triggered '8-Indicator Absolute', '8-Indicator Trend' and 'M6 Absolute' model alerts from October 2007 (Figure 4-11). These alerts were due to initial clear rising trends in ESA_M6 and ESA_PP, which then stabilised at an increased level (Figure 4-12 to Figure 4-15). There were no IHUMS alerts at this time. The CI trends started increasing again in October 2008, finally triggering two IHUMS CI alerts in November 2008 and December 2008. There was also an IHUMS sensor alert in June 2008, which correlated with rapid rises in FSA_SO1 and FSA_SO2 (Figure 4-16). The CIs returned to normal when the sensor was replaced, however this had no effect on the levels of ESA_M6 and ESA_PP

The initial rising trends in ESA_M6 and ESA_PP were also mirrored on the RHA gearbox right hydraulic drive 47/81T Gears, but levels remained in the normal range. The second period of rising CI trends on LHA left hydraulic drive, starting in October 2008, was also mirrored on the RHA, this time causing the CIs to reach abnormal levels (Figure 4-17 to Figure 4-20). The '8-Indicator Absolute' anomaly model for the RHA gearbox right hydraulic drive 47/81T Gears generated continuous alerts from May 2008, and all the remaining models alerted from October 2008. However, no IHUMS CI alerts were triggered on the right hydraulic drive. The fact that the trends on the LHA were mirrored on the RHA (monitored from a different sensor) indicated that these were originating from the gearbox, and were not due to an instrumentation problem.

The MGB (together with both accessories) was removed on 22 December 2008 because of metal contamination. Bristow requested a strip report from Eurocopter to identify the source of the debris, however no report had been received by the end of the trial. All CI levels returned to normal after the MGB change. Although the source of the metal contamination is unknown, it is possible that this could have originated from the LHA gearbox.

Because the IHUMS did generate some alerts, this example has been placed in the category of anomaly model alerts corroborated by the IHUMS. However, the IHUMS alerts on the left hydraulic drive occurred much later than the anomaly model alerts, and no IHUMS alerts were triggered on the right hydraulic drive. Therefore the anomaly models gave an earlier and clearer indication of the abnormal CI trends on these components.

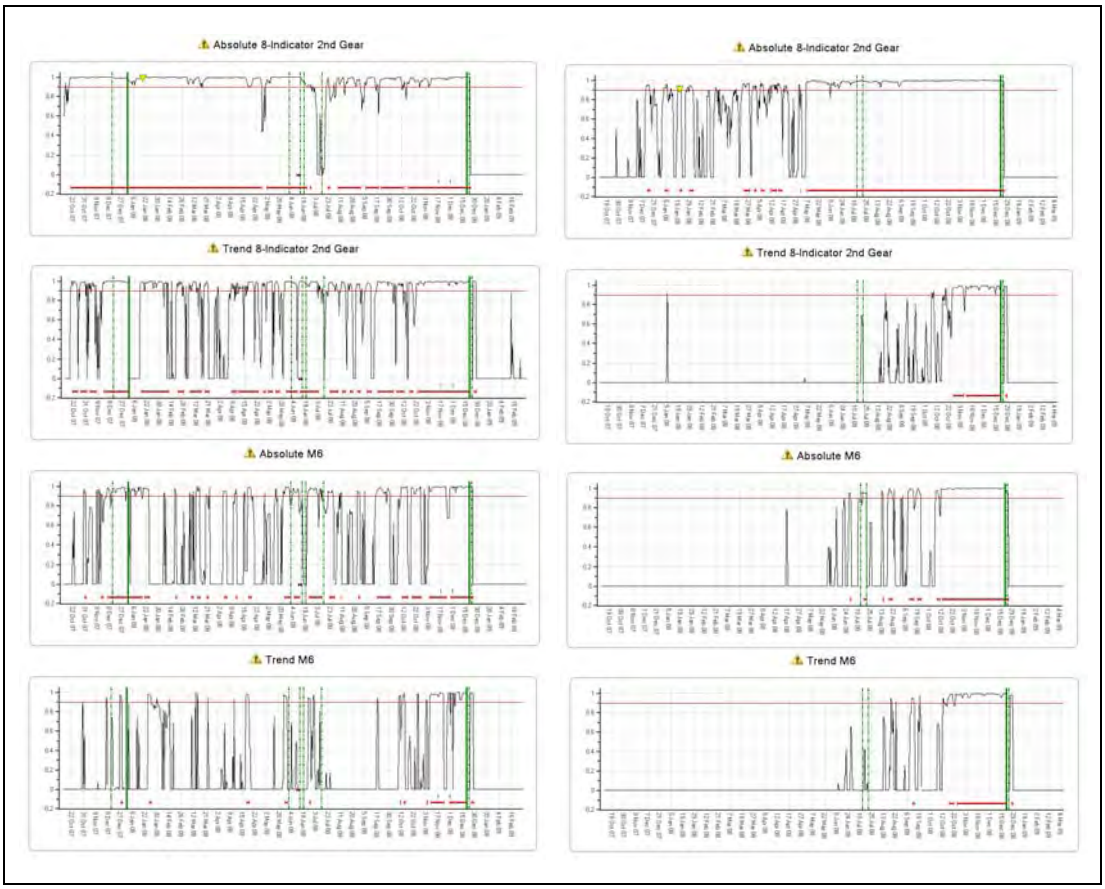

Figure 4-11 G-BWWI LHA and RHA hydraulic drive 81-tooth gears; LHS top to bottom – LHA 8IA, 8IT, M6A, M6T model PA; RHS top to bottom – RHA 8IA, 8IT, M6A, M6T model PA

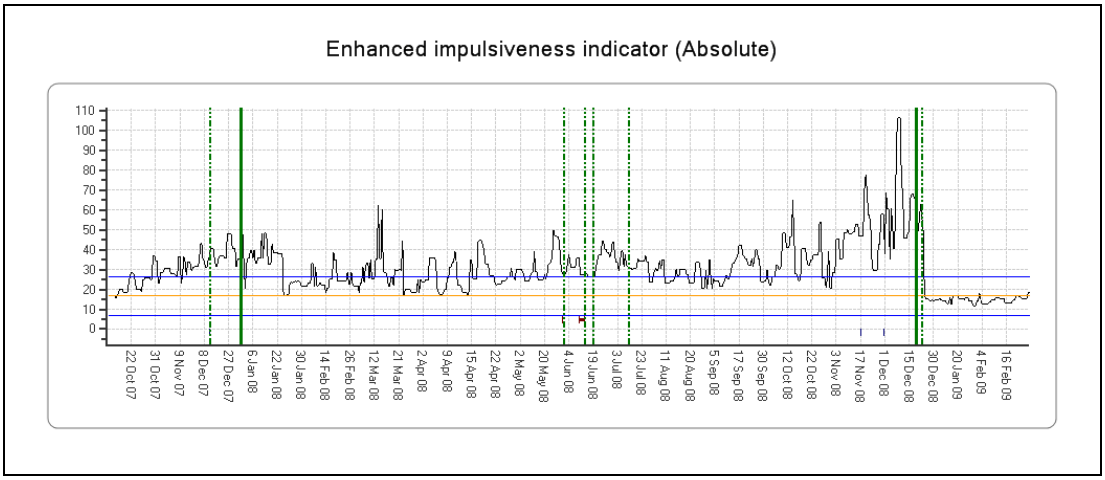

Figure 4-12 G-BWWI LHA left hydraulic drive 81-tooth gear – ESA_M6

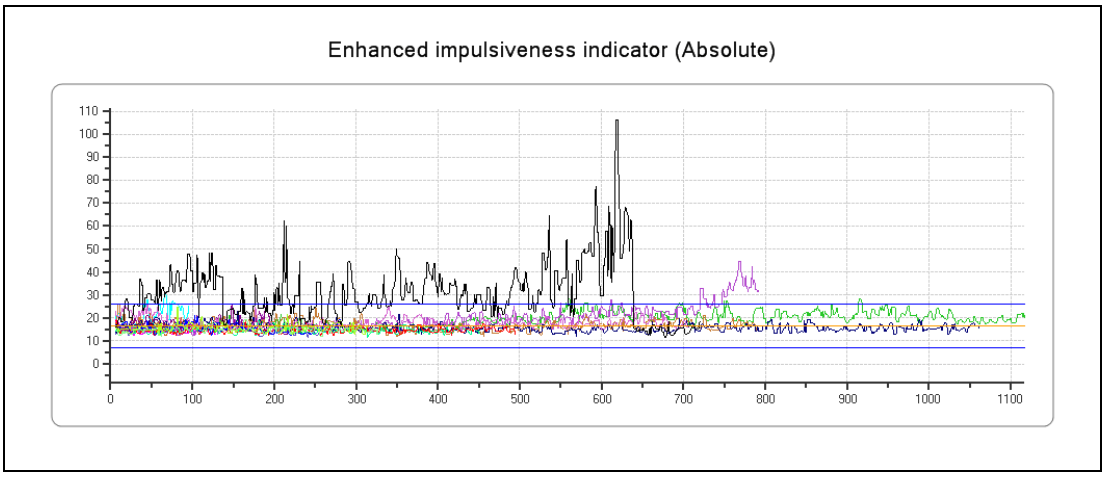

Figure 4-13 G-BWWI LHA left hydraulic drive 81-tooth gear – ESA_M6 – fleet view

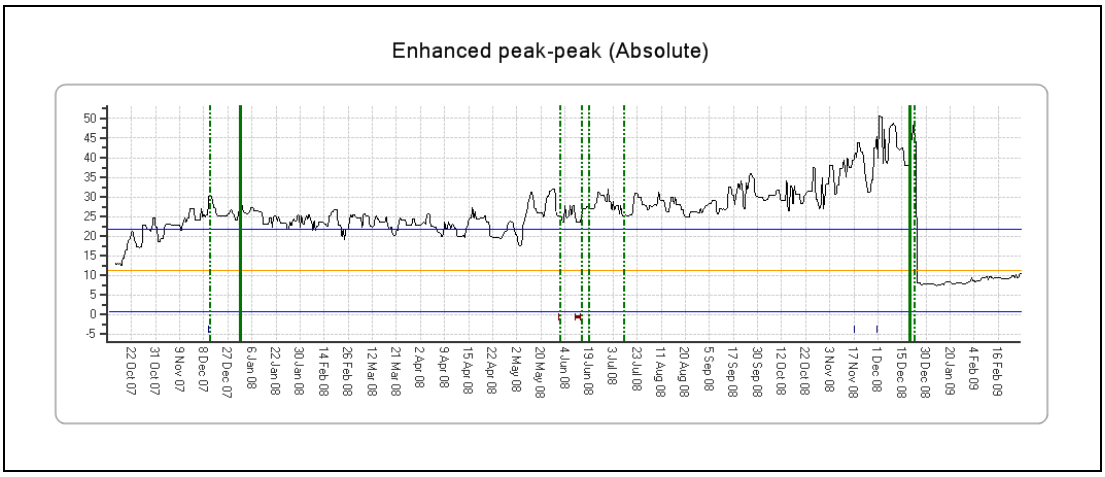

Figure 4-14 G-BWWI LHA left hydraulic drive 81-tooth gear – ESA_PP

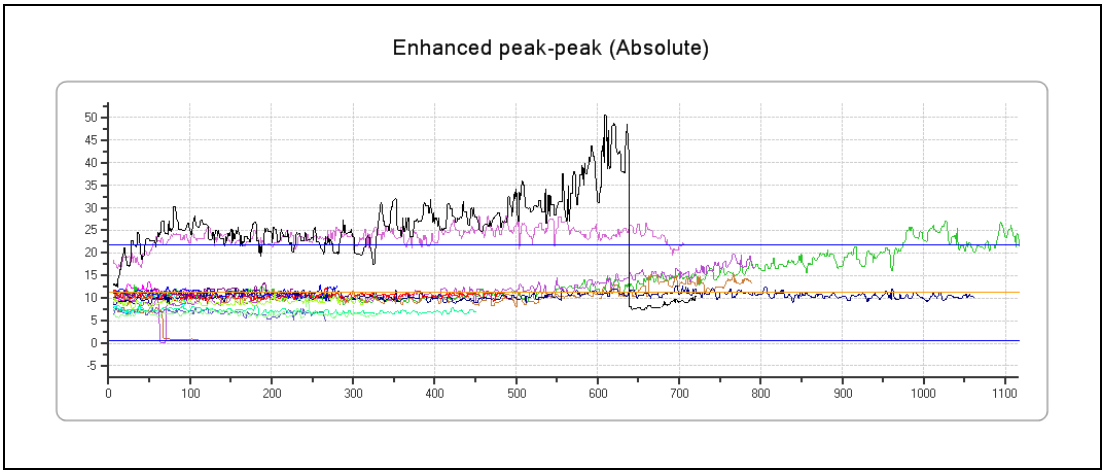

Figure 4-15 G-BWWI LHA left hydraulic drive 81-tooth gear – ESA_PP – fleet view

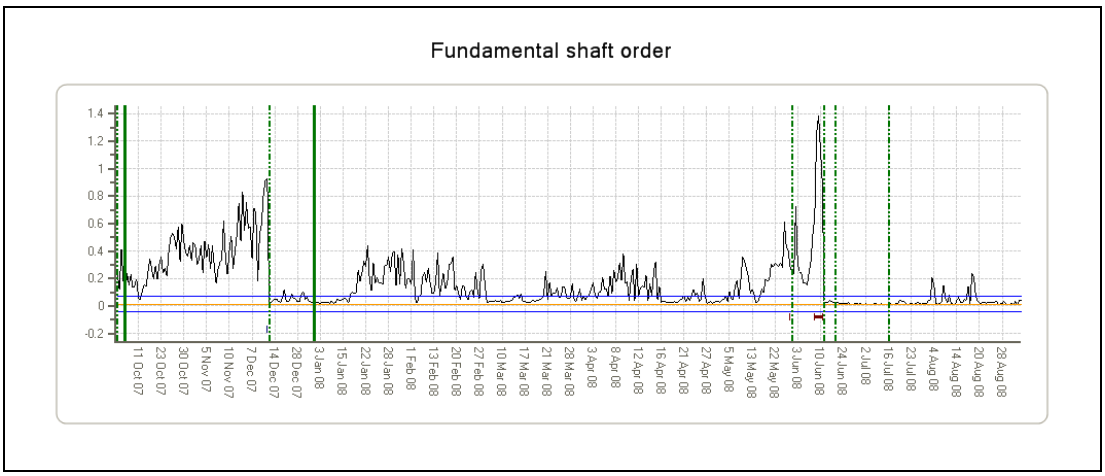

Figure 4-16 G-BWWI LHA left hydraulic drive 81-tooth gear – FSA_SO1

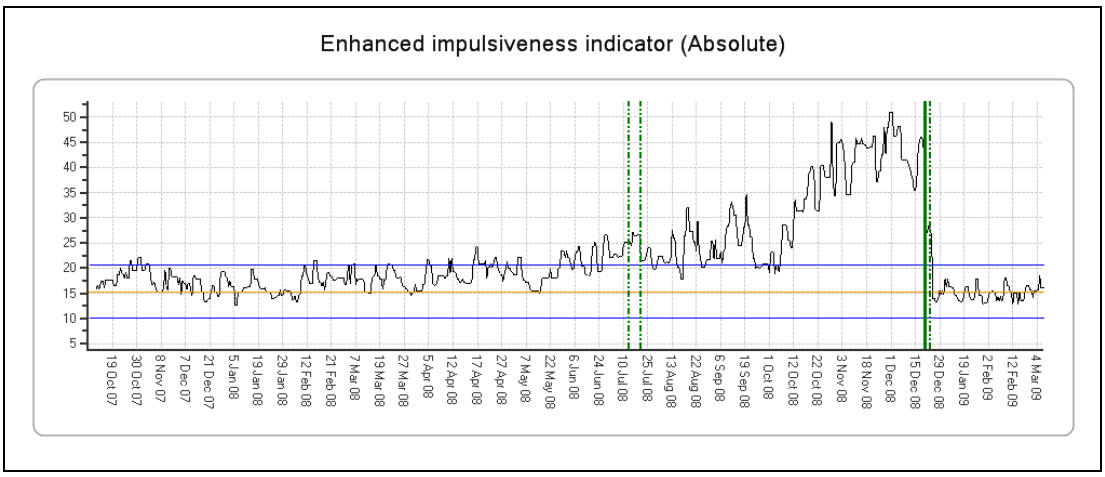

Figure 4-17 G-BWWI RHA right hydraulic drive 81-tooth gear – ESA_M6

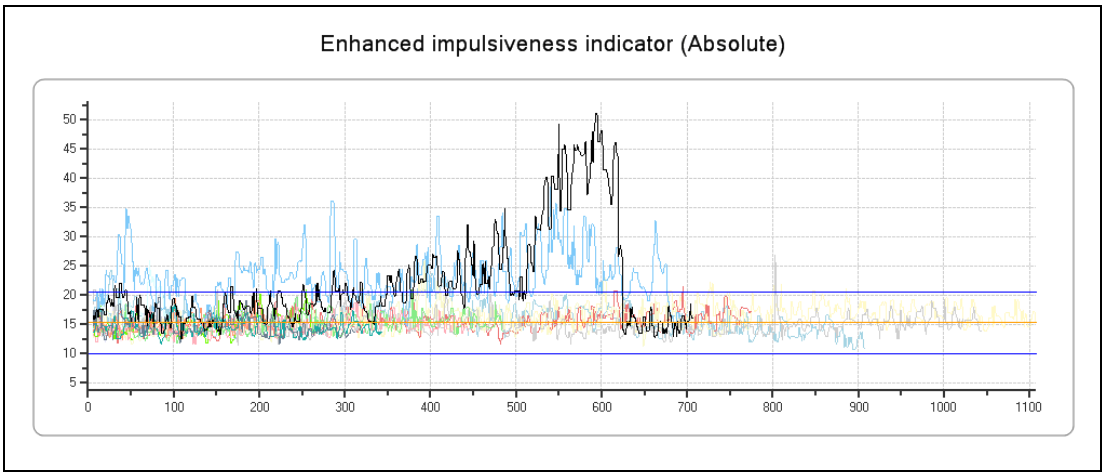

Figure 4-18 G-BWWI RHA right hydraulic drive 81-tooth gear – ESA_M6 – fleet view

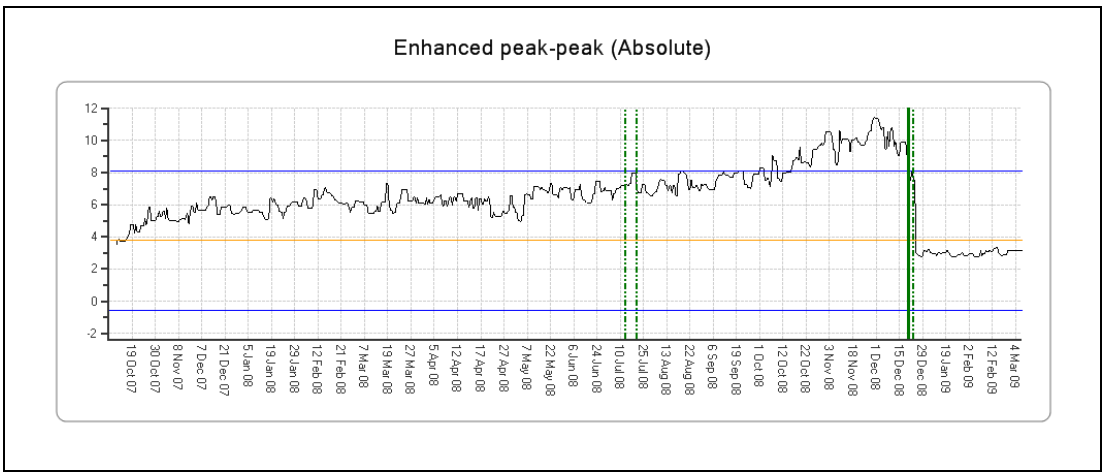

Figure 4-19 G-BWWI RHA right hydraulic drive 81-tooth gear – ESA_PP

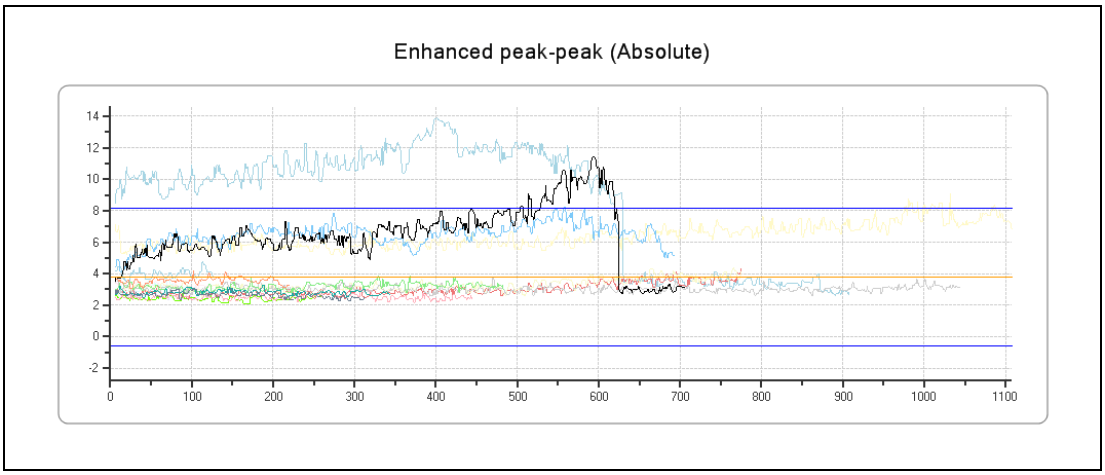

Figure 4-20 G-BWWI RHA right hydraulic drive 81-tooth gear – ESA_PP – fleet view

Example 3: G-BMCW Sensor 7 – Upper Planets

On 25 January 2008 the '8-Indicator Absolute' model went into alert, and this was followed shortly after by the '8-Indicator Trend' model (Figure 4-21 and Figure 4-22). These alerts were caused by a sudden and very large increase in the levels and variability of FSA_SO1 and FSA_SO2, with some other sensor 7 analyses also being affected (FSA_SO1 values are shown in Figure 4-23 and Figure 4-24). These characteristics have previously been associated with sensor 7 faults.

There were no IHUMS CI alerts, however an IHUMS sensor fault was recorded on 3 February 2008. The data returned to normal on 10 February 2008. Although there was no information in the maintenance database, it is believed that sensor maintenance had been performed at that time (maintenance actions that do not involve replacement of the sensor are not recorded).

This is a typical example of a number of IHUMS instrumentation faults detected during the trial. Although the IHUMS detected a sensor fault, the anomaly models again give an earlier and clearer indication of the problem.

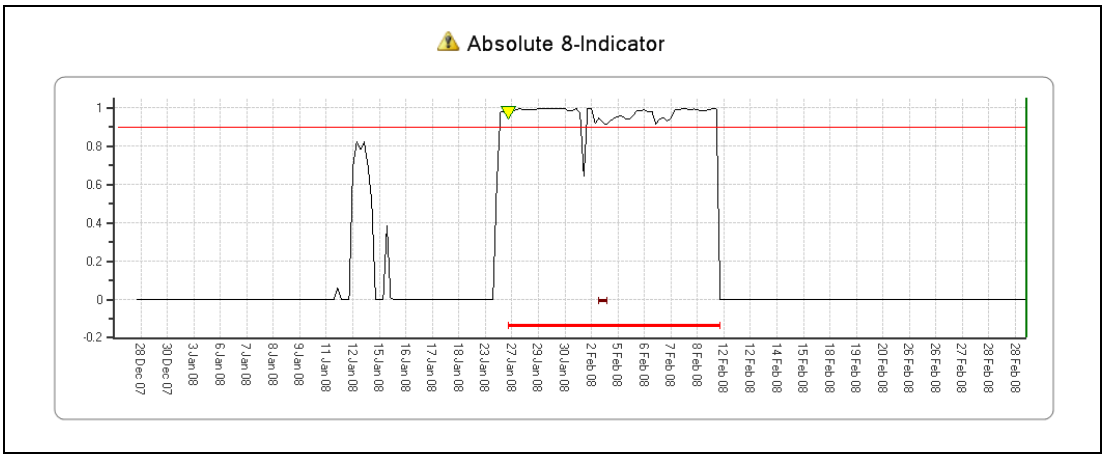

Figure 4-21 G-BMCW 2nd stage planet gears – '8-Indicator' absolute model PA

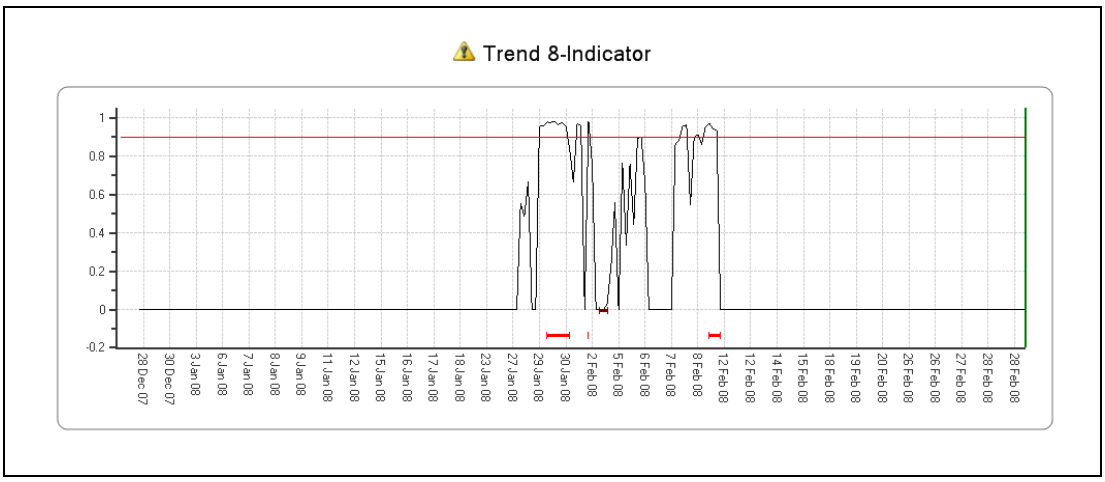

Figure 4-22 G-BMCW 2nd stage planet gears – '8-Indicator' trend model PA

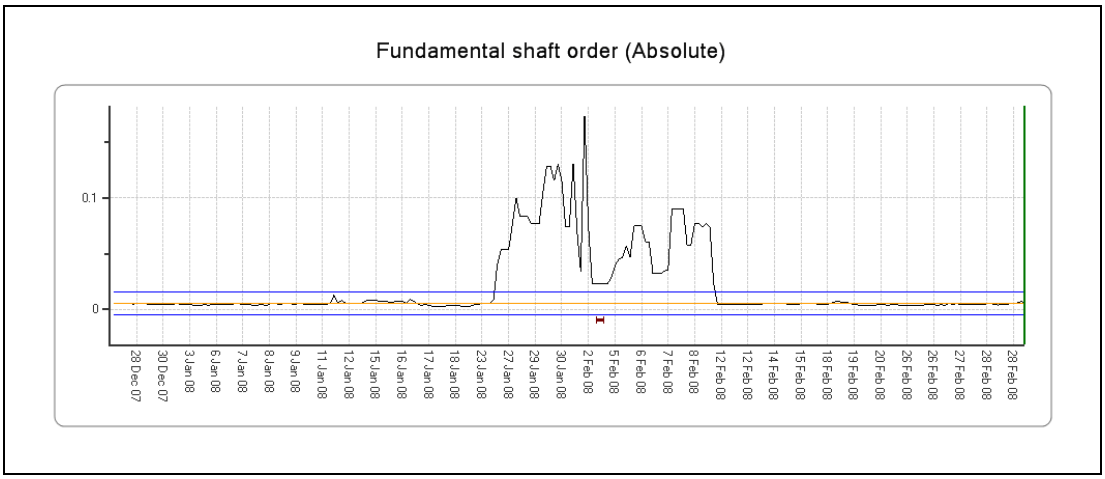

Figure 4-23 G-BMCW 2nd stage planet gears – FSA_SO1

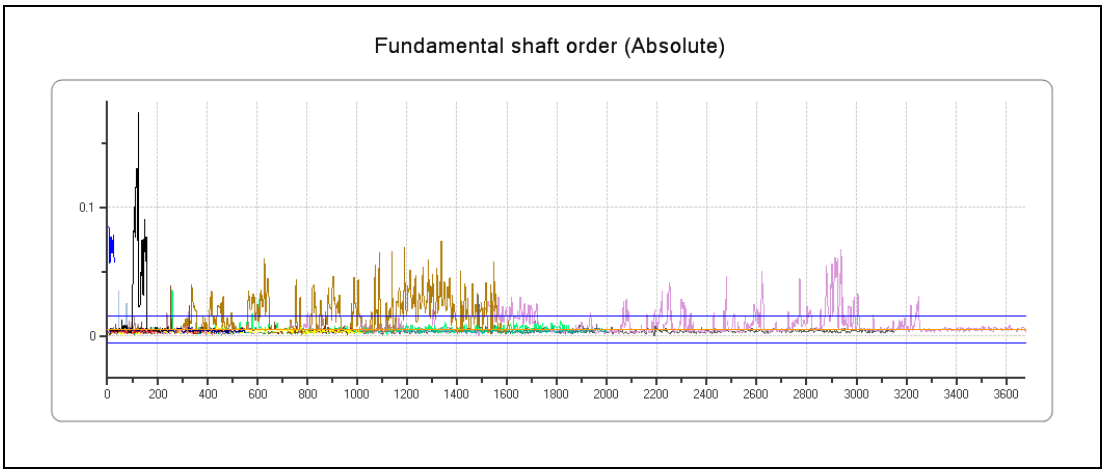

Figure 4-24 G-BMCW 2nd stage planet gears – FSA_SO1 – fleet view

Example 4: G-BWMG Tail Rotor Gearbox Output

Between April 2008 and July 2008 there were a large number of 'Shaft Order Absolute' model alerts, and also many IHUMS alerts and sensor fault indications (Figure 4-25). These were caused by multiple periods of high and variable FSA_SO1 values, with a number of related maintenance actions (Figure 4-26). The aircraft transitioned from Aberdeen to Nigeria in that period, receiving a new registration (5N-BGP).

The early 'Shaft Order' anomaly model alerts were due to rising FSA_SO1 levels, which also caused IHUMS alerts, though at a later date. These were not cleared by balance operations on the tail rotor, but did clear after maintenance on the spindles and flapping hinge bearings.

Further anomaly model alerts were generated a few weeks later due to the occurrence of extremely high and variable FSA_SO1 values, this time accompanied by IHUMS sensor fault indications. Unlike the earlier FSA_SO1 trend, the extreme behaviour of the later data suggested an instrumentation problem. Multiple maintenance actions on the sensor and the Tail Rotor in June 2008 and July 2008 did not correct the abnormal FSA_SO1 behaviour. However, FSA_SO1 levels returned to normal after a sensor replacement on 4 August 2008. Bristow confirmed that the alerts were due to an instrumentation problem. The initial efforts to correct this by changing the sensor had been unsuccessful, and it was eventually discovered that the sensor cable had been shorting out.

This example shows that the 'Shaft Order' models implemented as a result of experience gained during the first six-month trial (reference [3]) are successfully detecting faults, with good correlation between anomaly model alerts and IHUMS alerts. The example represents a relatively complex maintenance scenario, with both aircraft faults and IHUMS instrumentation faults present. The aircraft had transitioned to Nigeria in late April 2008 before the appearance of the instrumentation problem. If the engineers in Nigeria had been aware of the abnormal characteristics of the FSA_SO1 trend (i.e. highly variable, with many exceptionally high values), they might have more rapidly diagnosed and rectified this problem.

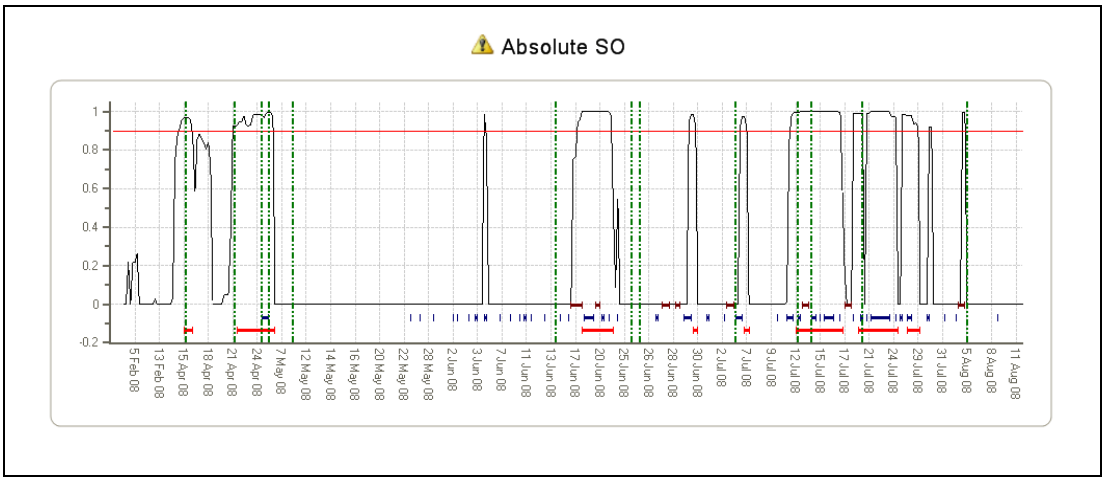

Figure 4-25 G-BWMG tail rotor gearbox output – SOA model PA

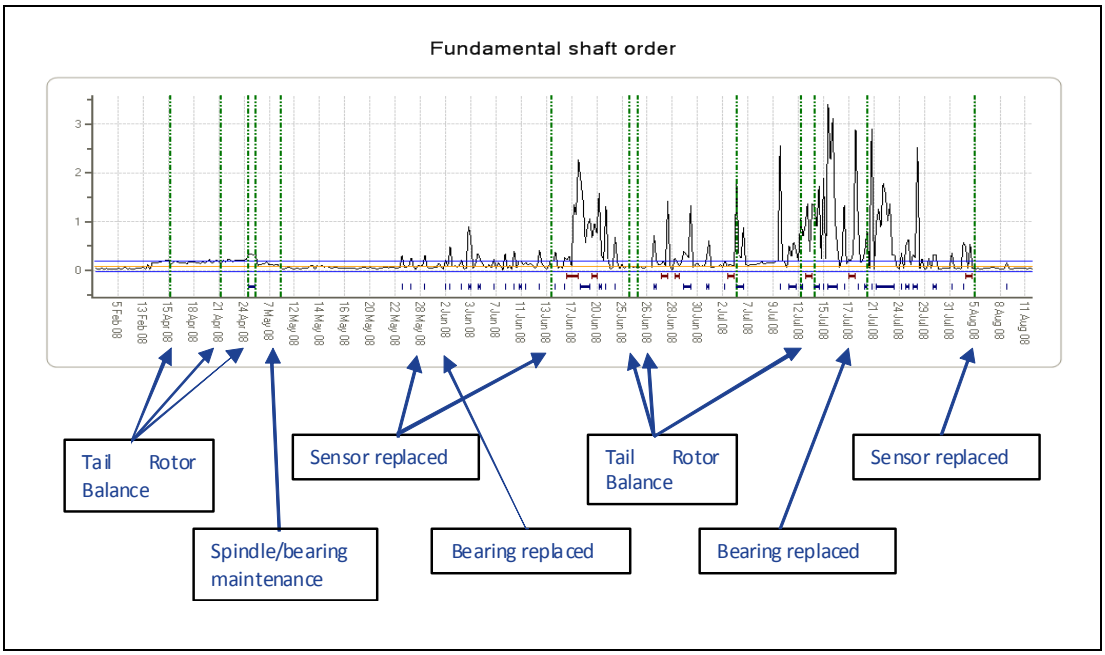

Figure 4-26 G-BWMG tail rotor gearbox output – FSA_SO1

Example 5: G-TIGG Tail Rotor Gearbox Output

Similar to the previous example, 'Shaft Order Absolute' anomaly model alerts were initially generated from 1 May 2008, and there was a good correlation of these with IHUMS alerts (Figure 4-27). The alerts were caused by increases in FSA_SO1 (Figure 4-28). Bristow reported that this aircraft had a known tail rotor problem. There had been multiple rebalancing operations, with associated IHUMS re-datums. The alerts finally cleared after maintenance around 17 May 2008, when worn tail rotor pitch link bearings were found and replaced.

However the 'Shaft Order Absolute' model produced later alerts, which were not accompanied by a rise in FSA_SO1 (Figure 4-29). In this case the alerts were being triggered by low FSA_SO2 values. The FSA_SO2 levels were the lowest in the fleet, and there were also some IHUMS sensor alerts (Figure 4-30 and Figure 4-31). Bristow reported that the IHUMS showed occasional signal dropouts on all accelerometer channels, which may indicate a possible DAPU issue.

The aircraft was in heavy maintenance between 16 October 2008 and 17 November 2008, during which time various components were removed and replaced. Following the maintenance period there was a step increase in FSA_SO2 values into the normal range, and the anomaly model alerts ceased. It is not known whether the later anomaly model alerts were due to an IHUMS instrumentation problem, or if the model was possibly being over sensitive to low values of FSA_SO2.

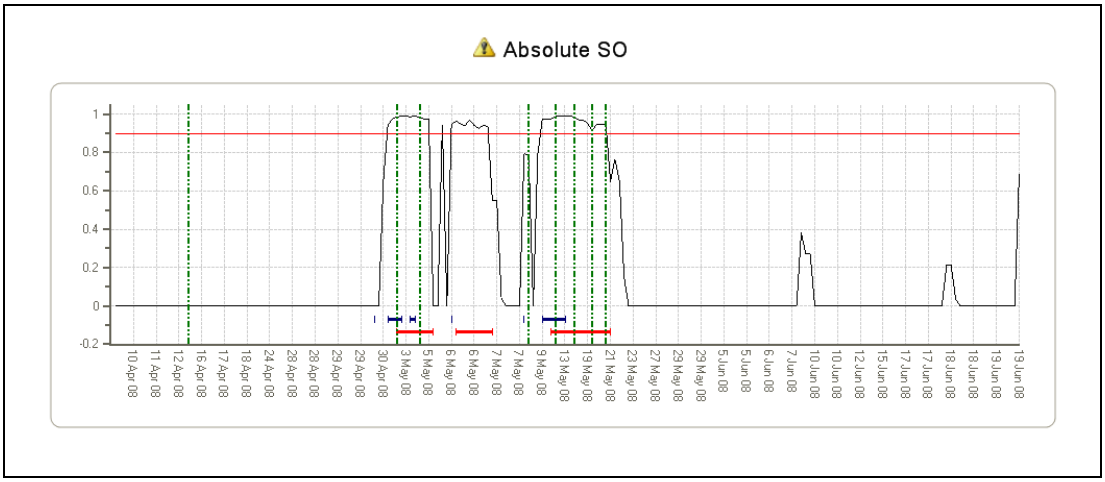

Figure 4-27 G-TIGG tail rotor gearbox output – SOA model PA

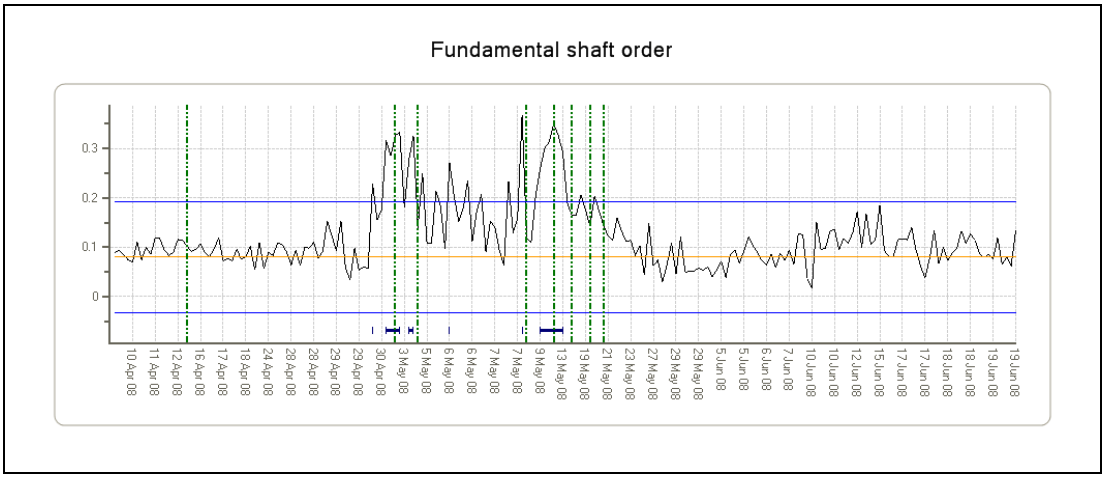

Figure 4-28 G-TIGG tail rotor gearbox output – FSA_SO1

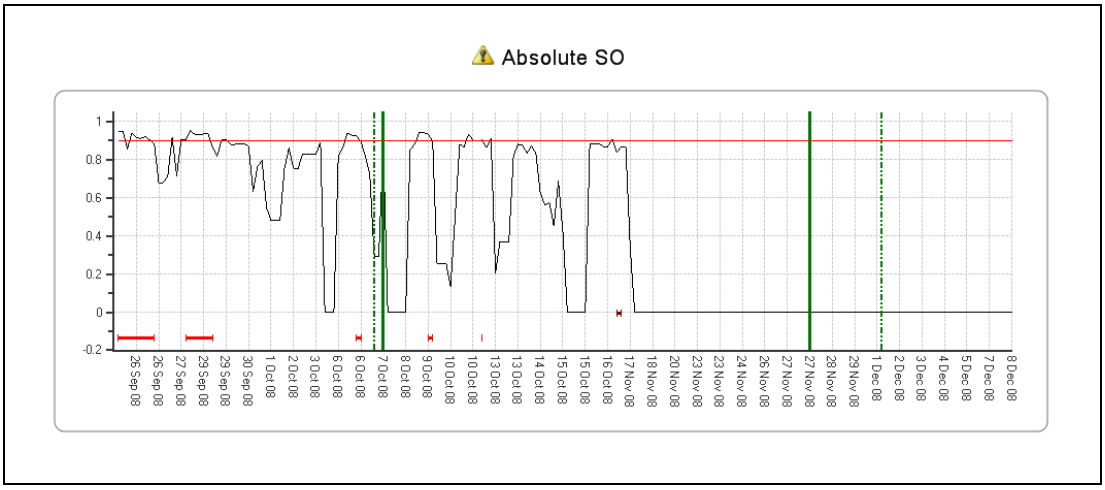

Figure 4-29 G-TIGG tail rotor gearbox output – SOA model PA

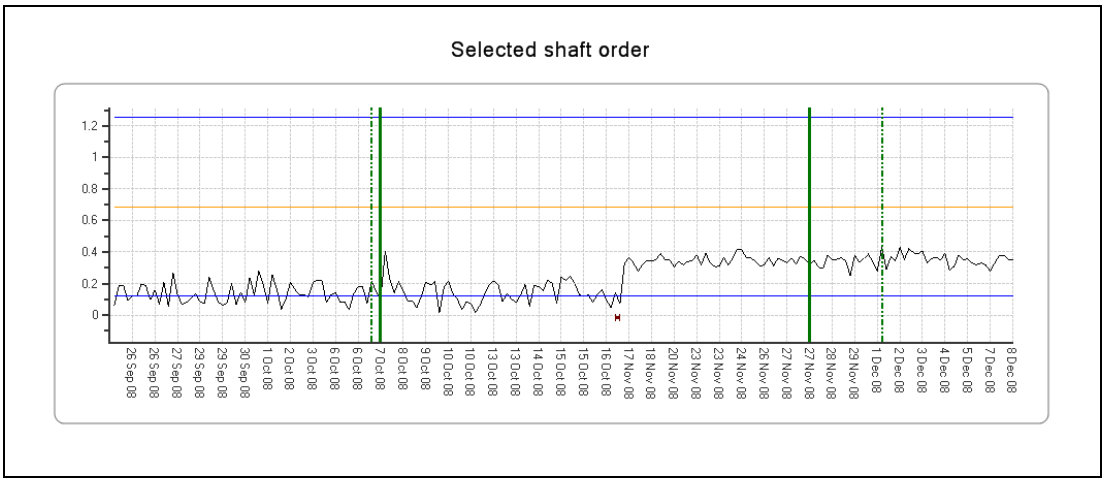

Figure 4-30 G-TIGG tail rotor gearbox output – FSA_SO2

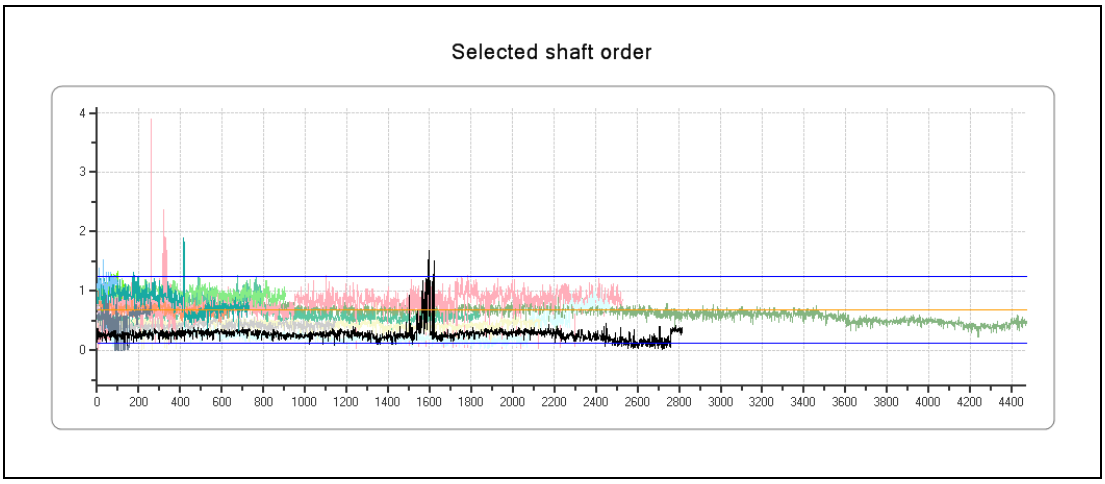

Figure 4-31 G-TIGG tail rotor gearbox output – FSA_SO2 – fleet view

Example 6: G-TIGC MGB RH Input Shaft

Frequent intermittent 'Shaft Order Absolute' model alerts were triggered from March 2008 (Figure 4-32), caused by increased levels and variability of FSA_SO1 (Figure 4-34 and Figure 4-35). A short time later '8-Indicator Absolute' and 'M6 Absolute' model alerts started to be triggered (Figure 4-33). These alerts were caused by increasing trends in ESA_M6 (shown in Figure 4-36 and Figure 4-37) and ESA_PP. There were also many IHUMS alerts. A left and right sensor swap on 11 May 2008 did not alter the variable behaviour of FSA_SO1.

During a period of heavy maintenance the engine and bendix shaft were changed. This resulted in step decreases in ESA_M6 and FSA_SO1 when the aircraft came out of maintenance on 21 December 2008, although ESA_SD remained at an elevated level. The 'Shaft Order Absolute' and 'M6 Absolute' model alerts ceased at this time. When the engine was removed the inboard bearing mount was found to be worn, causing some movement of the bearing. This may explain the high and variable CI values causing the anomaly model alerts, which returned to normal after maintenance. However, similar anomalous behaviour of ESA_M6 has been observed on the MGB input shafts on other aircraft, with no cause for this being found (see Section 4.6).

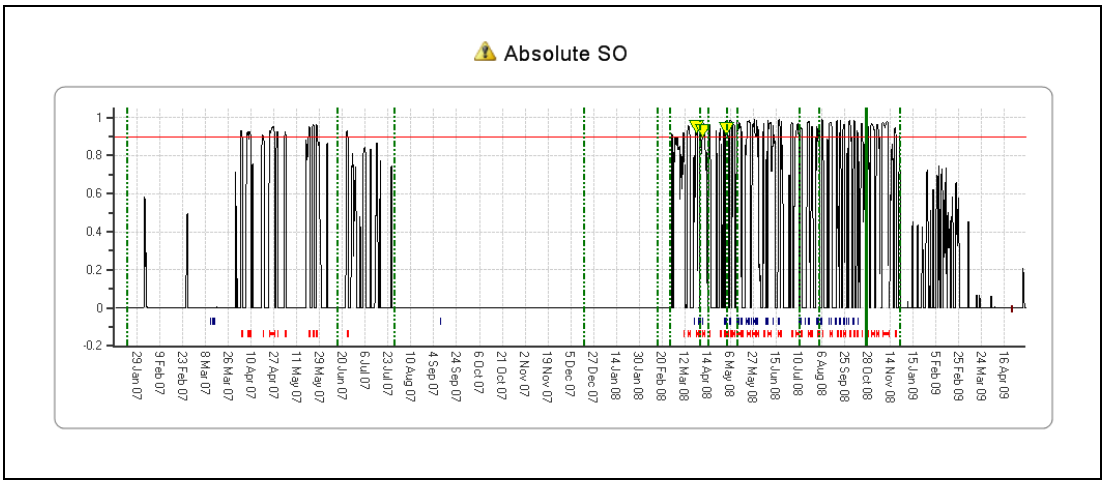

Figure 4-32 G-TIGC MGB RH input shaft – SOA model PA

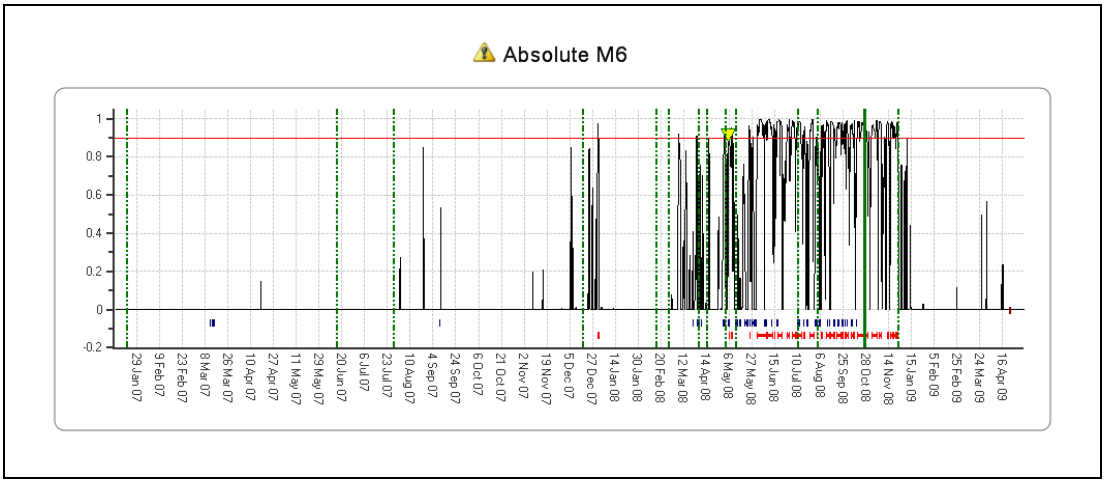

Figure 4-33 G-TIGC MGB RH input shaft – M6A model PA

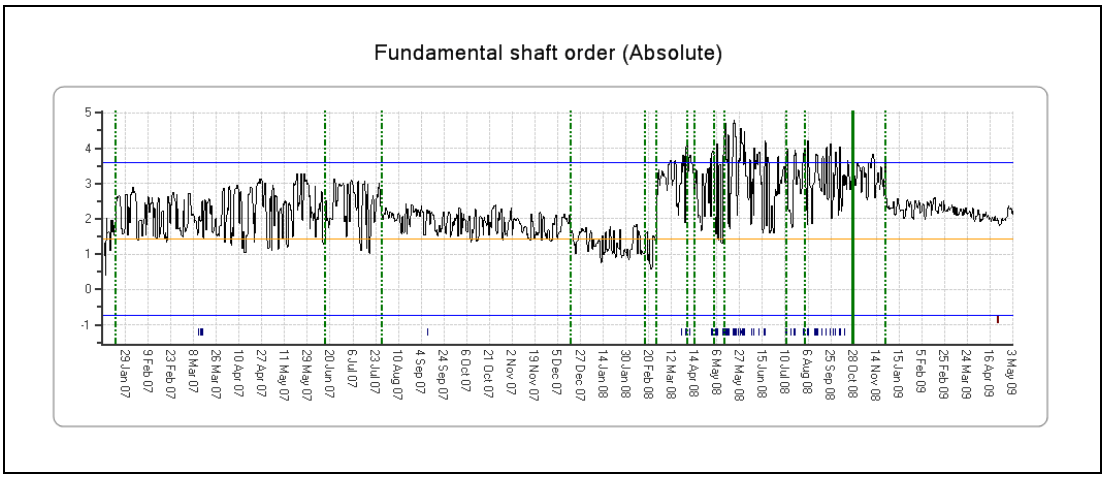

Figure 4-34 G-TIGC MGB RH input shaft – FSA_SO1

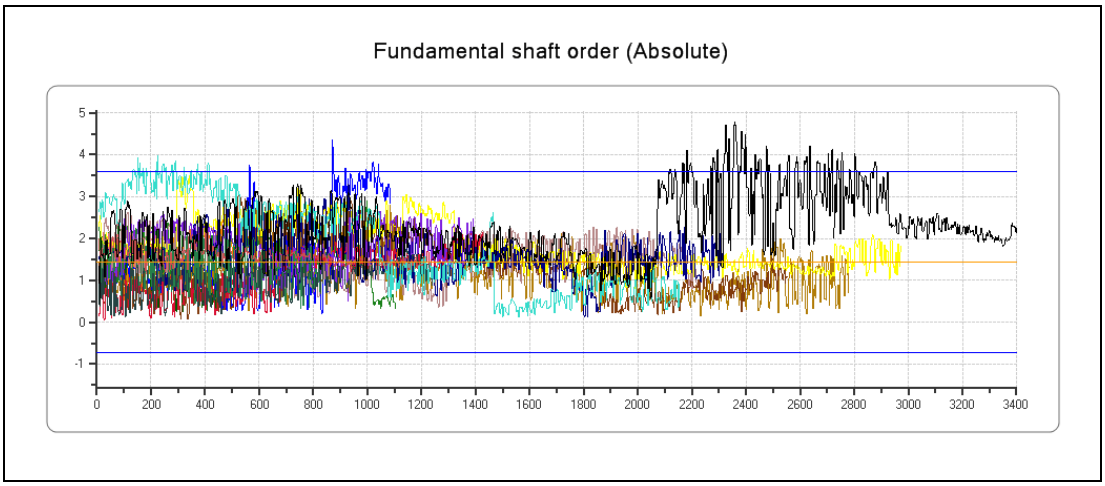

Figure 4-35 G-TIGC MGB RH input shaft – FSA_SO1 – fleet view

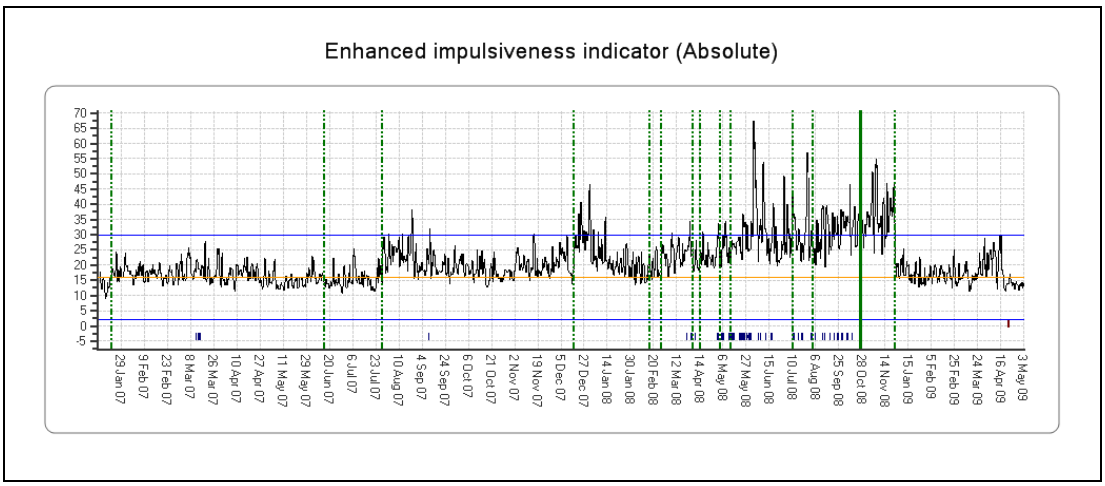

Figure 4-36 G-TIGC MGB RH input shaft – ESA_M6

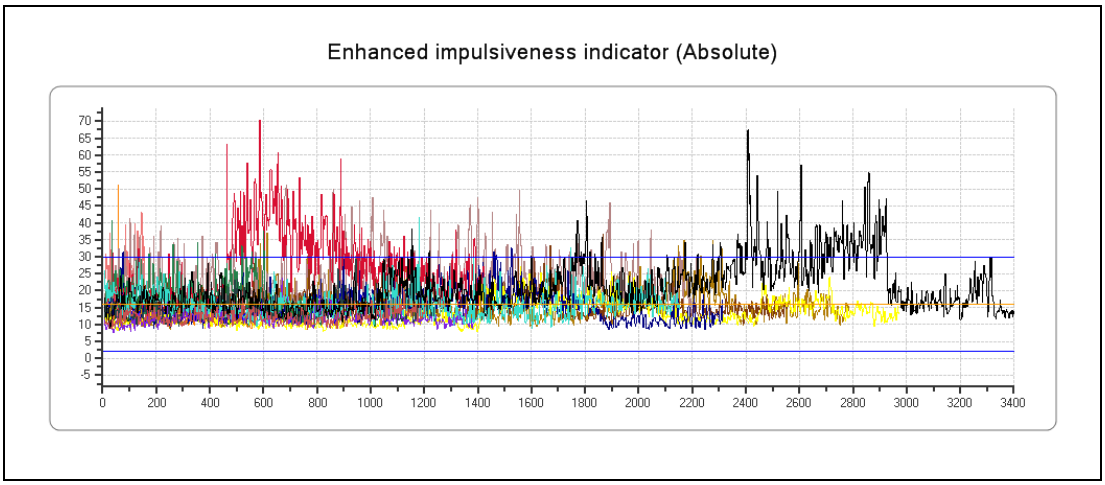

Figure 4-37 G-TIGC MGB RH input shaft – ESA_M6 – fleet view

Example 7: IHUMS DAPU

After the IHUMS DAPU in G-TIGE was changed on 2 February 2007 there were absolute anomaly model alerts on the TGB Input and all three 2nd stage epicyclic gears (Figure 4-38 and Figure 4-41). These alerts were caused by step increases in multiple CIs, including ESA_SD (Figure 4-39 and Figure 4-42) and FSA_SO2 (Figure 4-40 and Figure 4-43). There were also a few IHUMS alerts.

The multiple anomaly model alerts were thought to be connected to the DAPU change, and a faulty DAPU was suspected. This was therefore swapped with the DAPU in G-BWZX on 15 March 2007. All the anomaly model alerts moved with the DAPU, which again caused step increases in multiple CIs. Figure 4-44 to Figure 4-49 show the equivalent data for G-BWZX to that in Figure 4-38 to Figure 4-43. The DAPU swap confirmed the fault, therefore this was removed as unserviceable.

Although the anomaly model alerts were corroborated by a few IHUMS alerts, it can be seen from the figures that the anomaly models gave a clearer and more consistent indication of the anomalous data generated by the faulty DAPU.

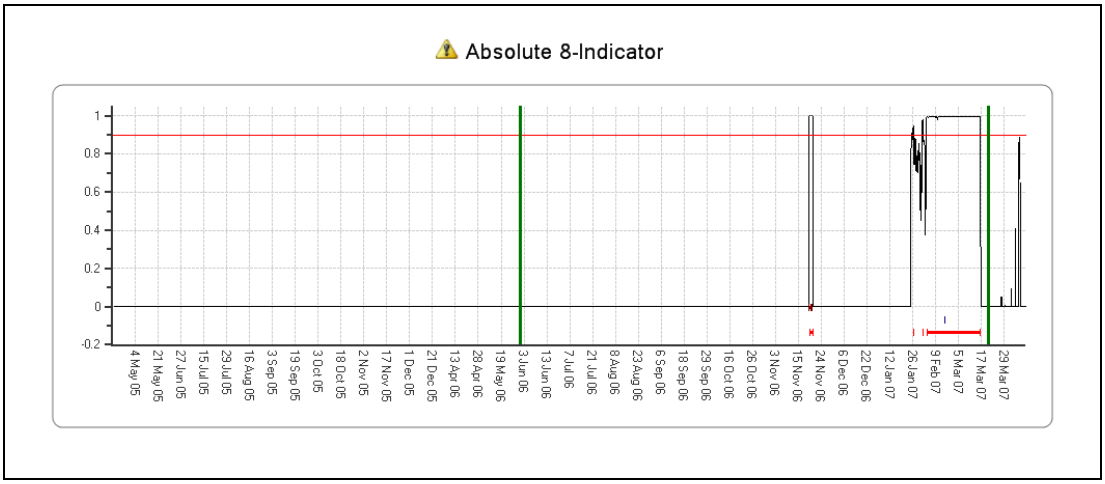

Figure 4-38 G-TIGE TGB input – 8IA model PA

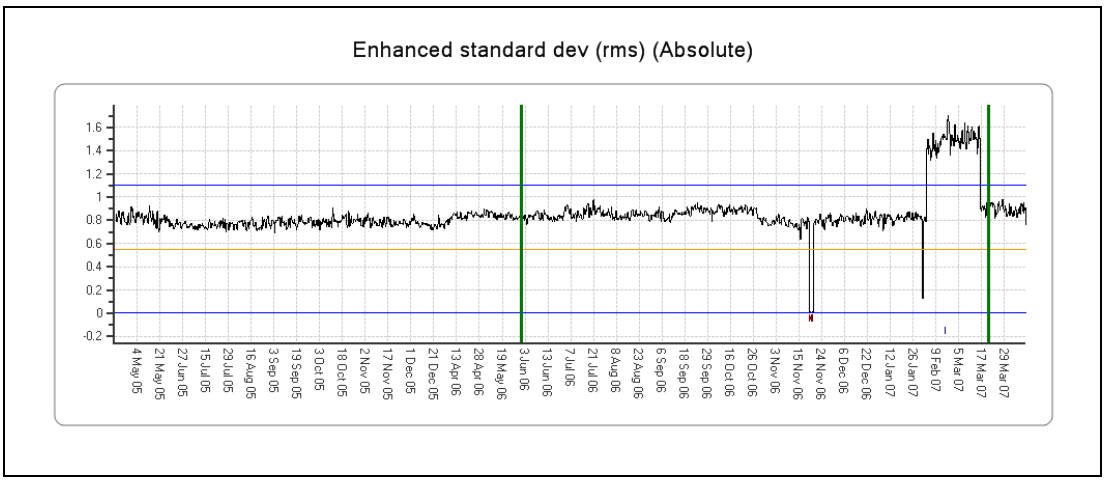

Figure 4-39 G-TIGE TGB input – ESA_SD

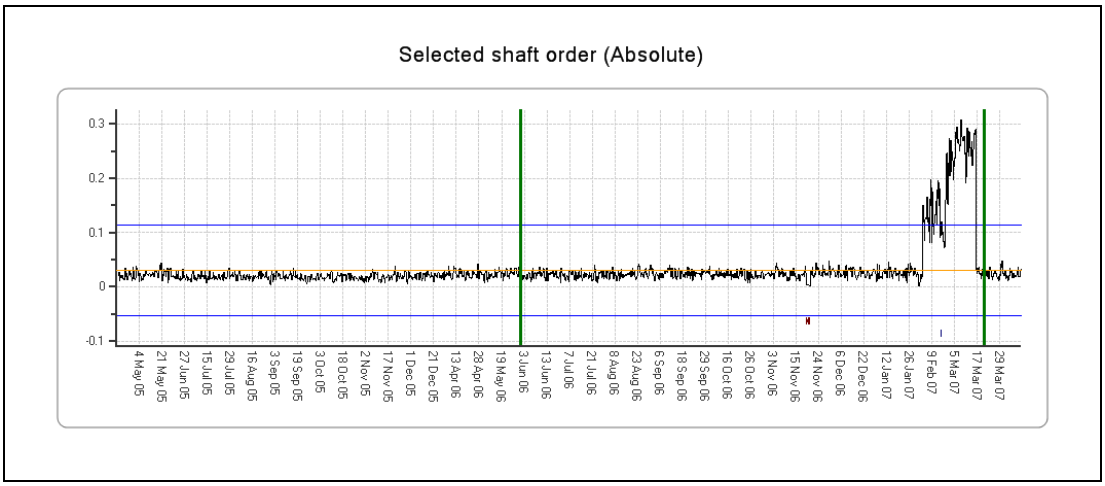

Figure 4-40 G-TIGE TGB input – FSA_SO2

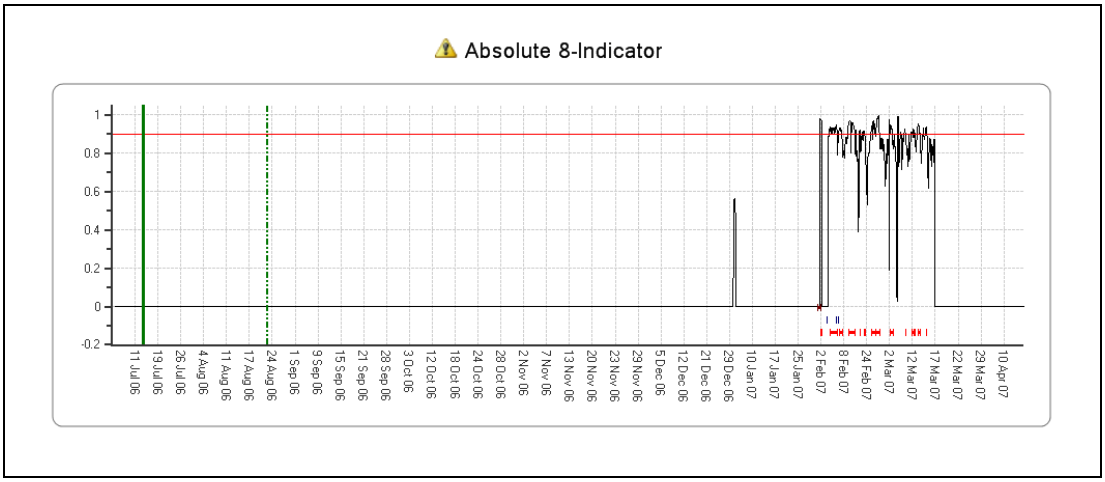

Figure 4-41 G-TIGE MGB 2nd stage planet gear – 8IA model PA

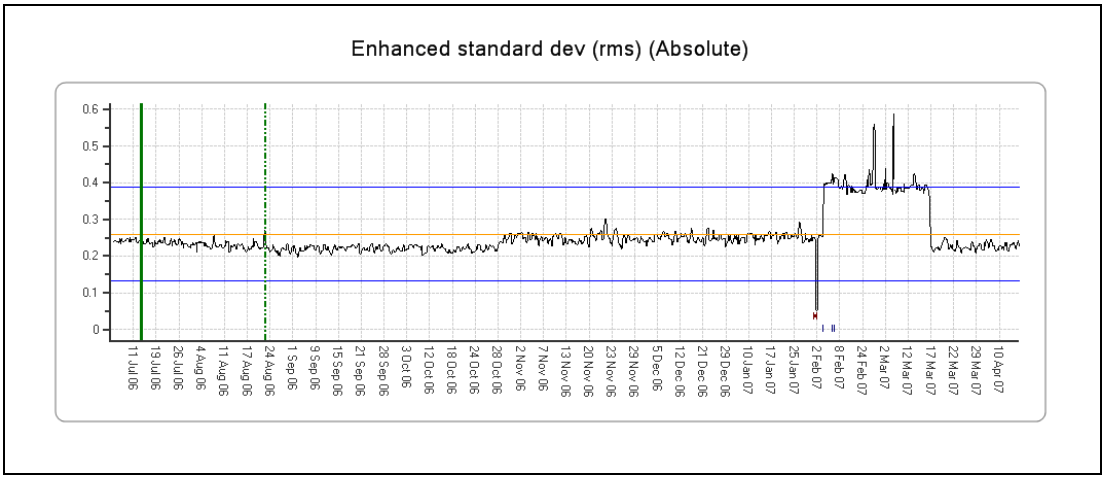

Figure 4-42 G-TIGE MGB 2nd stage planet gear – ESA_SD

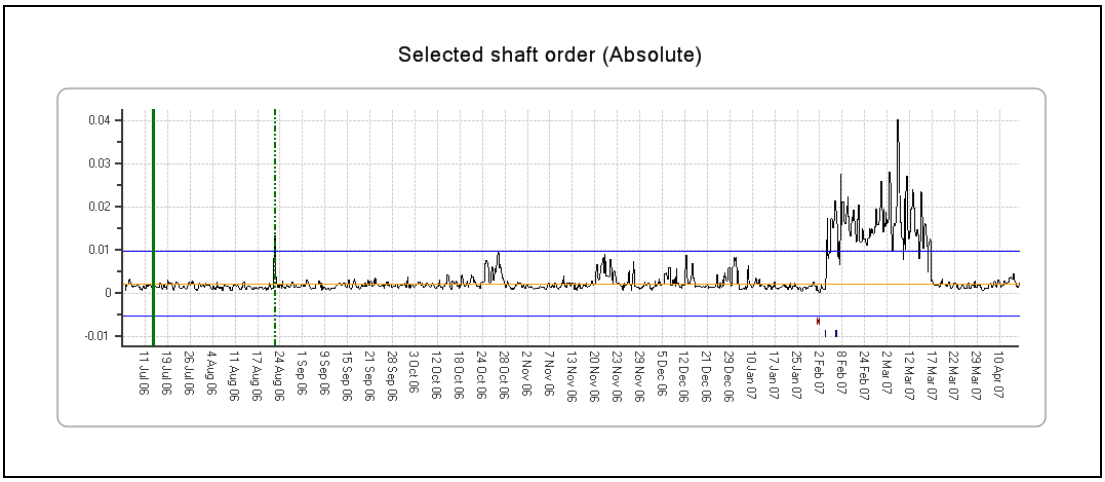

Figure 4-43 G-TIGE MGB 2nd stage planet gear – FSA_SO2

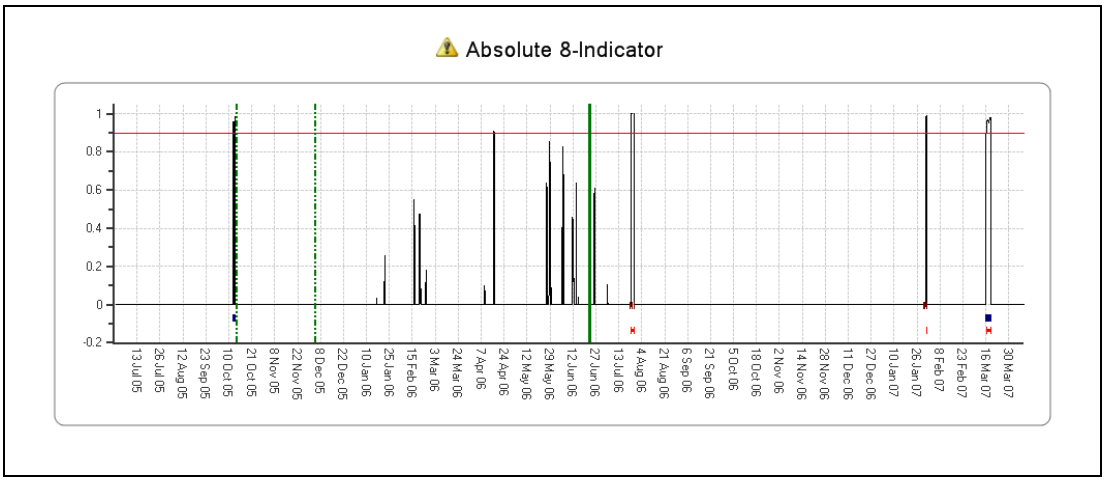

Figure 4-44 G-BWZX TGB input – 8IA model PA

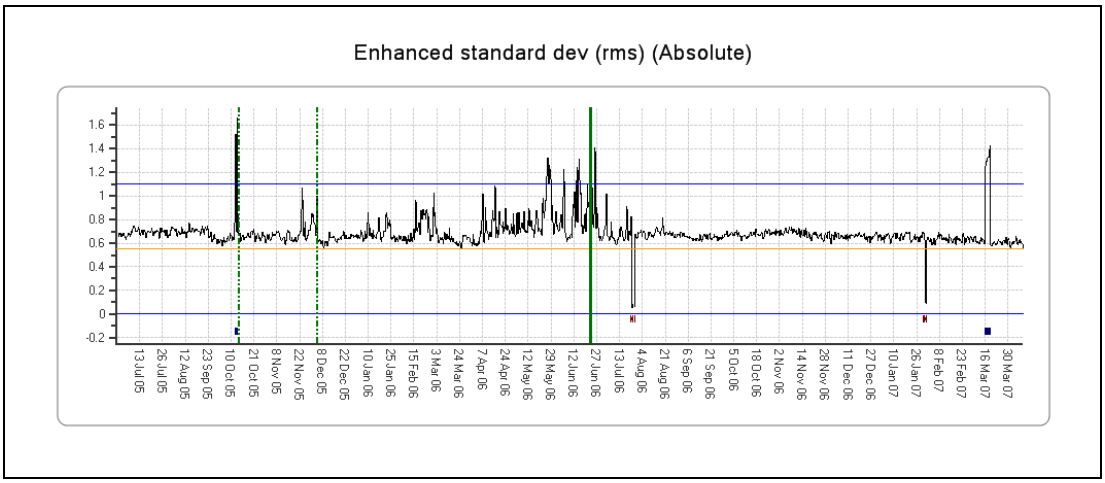

Figure 4-45 G-BWZX TGB input – ESA_SD

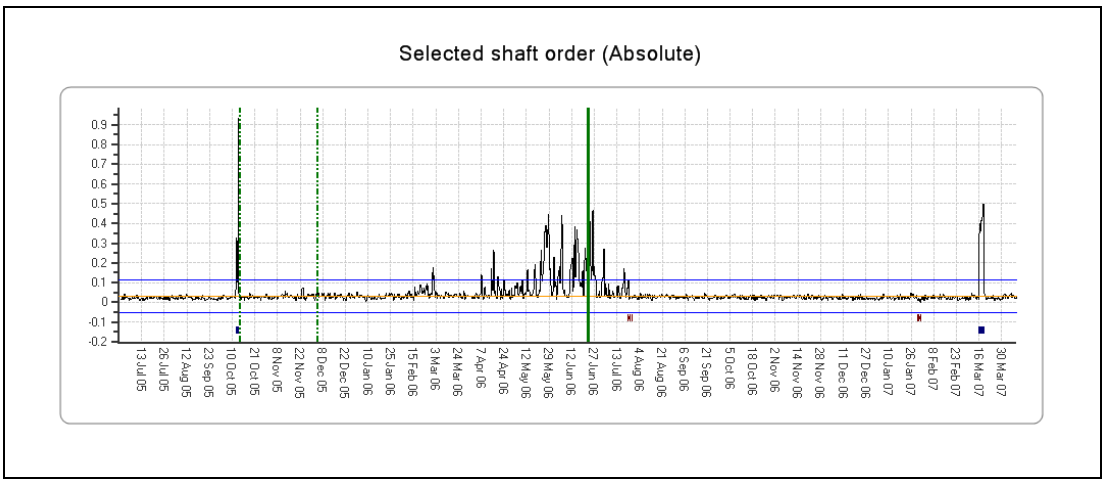

Figure 4-46 G-BWZX TGB input – FSA_SO2

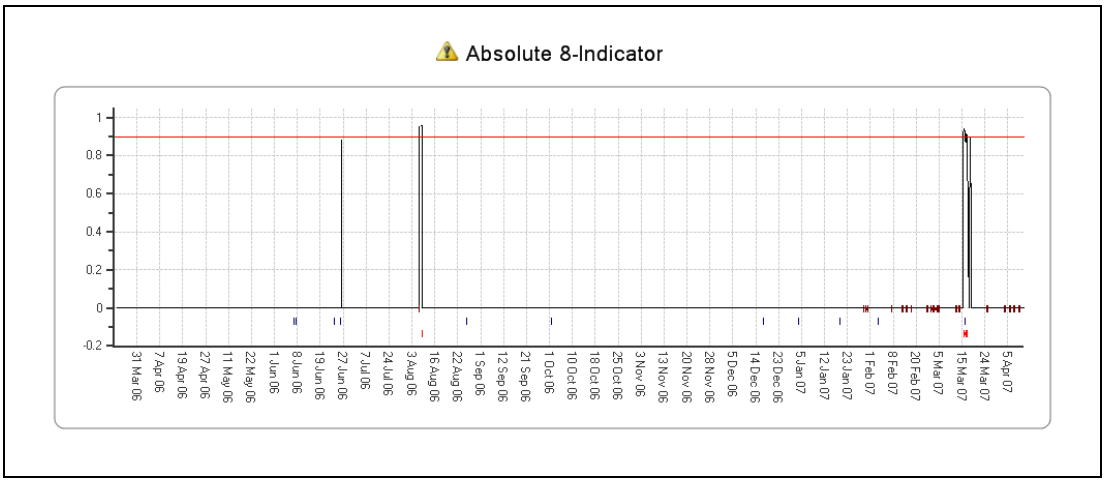

Figure 4-47 G-BWZX MGB 2nd stage planet gear – 8IA model PA

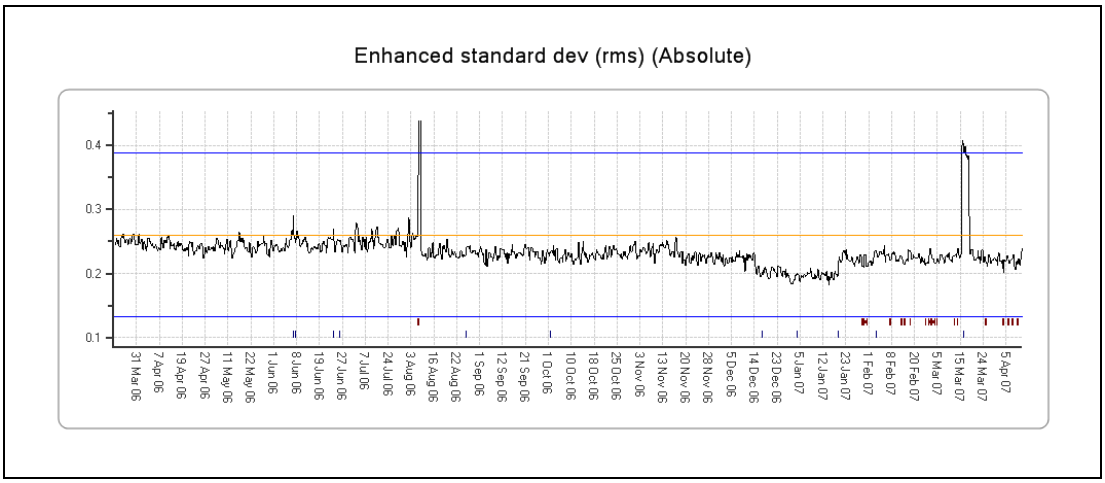

Figure 4-48 G-BWZX MGB 2nd stage planet gear – ESA_SD

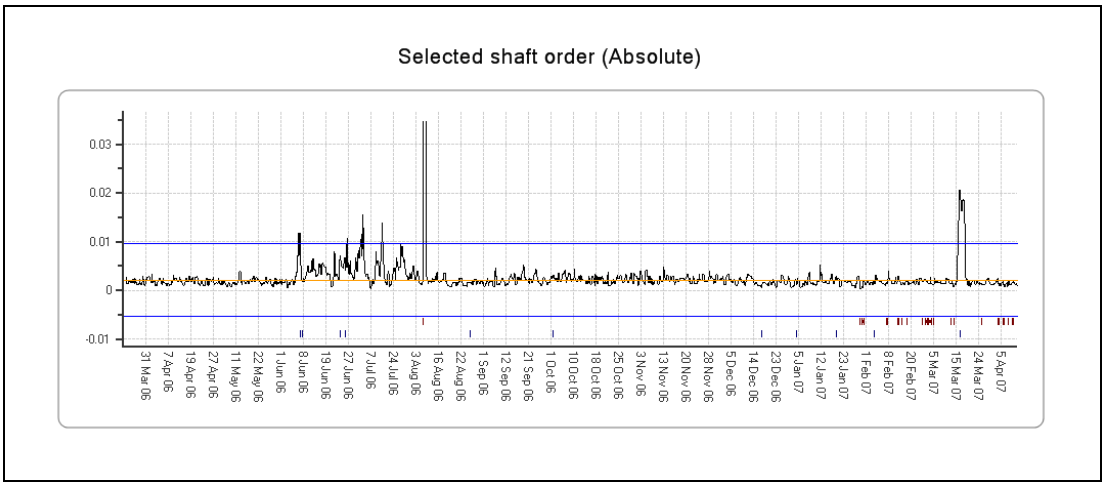

Figure 4-49 G-BWZX MGB 2nd stage planet gear – FSA_SO2

Example 8: G-BMCW LHA Left Hydraulic Idler

The LHA gearbox was replaced on 27 April 2007 following an IHUMS alert on the left hydraulic idler. The alert was caused by two consecutive high values in multiple CIs, including ESA-PP and ESA_M6 (Figure 4-50 and Figure 4-51). These were mirrored on the right hydraulic idler, but at a lower level (Figure 4-52 and Figure 4-53). The '8-Indicator Trend' model had previously alerted on the left hydraulic idler (Figure 4-54), however this was due to short periods of abnormally high FSA_SO1 values (Figure 4-55), most likely caused by an instrumentation problem.

According to the data, the two consecutive high CI values on the left hydraulic idler were followed by two normal values before the LHA gearbox was replaced. The primary reason that the anomaly models did not alert is that the two high values were suppressed by the median filtering pre-processing designed to remove signal outliers. As the data appears to have returned to normal prior to the LHA gearbox being removed, it is possible that the two high CI values were indeed outliers due to some unknown cause, and that the gearbox removal was unnecessary.

There were no further anomaly alerts until a scheduled MGB replacement for overhaul on 2 June 2007. All the MGB accelerometers were replaced at this time, probably for a scheduled calibration. 'M6 Absolute' and '8-Indicator Absolute' anomaly model alerts were generated on the left hydraulic idler from the start of the life of the new LHA gearbox (Figure 4-56 and Figure 4-57), and there were also several IHUMS alerts. These were caused by high values of multiple CIs, including ESA_M6 and ESA_PP (Figure 4-58 and Figure 4-59). The gearbox was placed on close monitor, and was then replaced on 5 August 2007, after which the data returned to normal. There were further '8-Indicator Absolute' anomaly model and IHUMS alerts between September 2007 and December 2007, most of these being driven by high values of FSA_SO1 (Figure 4-60), which are believed to be due to instrumentation problems. Some sensor maintenance actions were performed during this time.

The anomaly modelling suppressed the initial IHUMS alerts resulting in rejection of the LHA gearbox on 27 April 2007. This gearbox might have continued operating satisfactorily until the MGB replacement on 2 June 2007. However the anomaly models corroborated the IHUMS alerts occurring on the new LHA gearbox, resulting in a second gearbox replacement on 5 August 2007. Both the anomaly models and the IHUMS also then flagged the subsequent sensor problems.

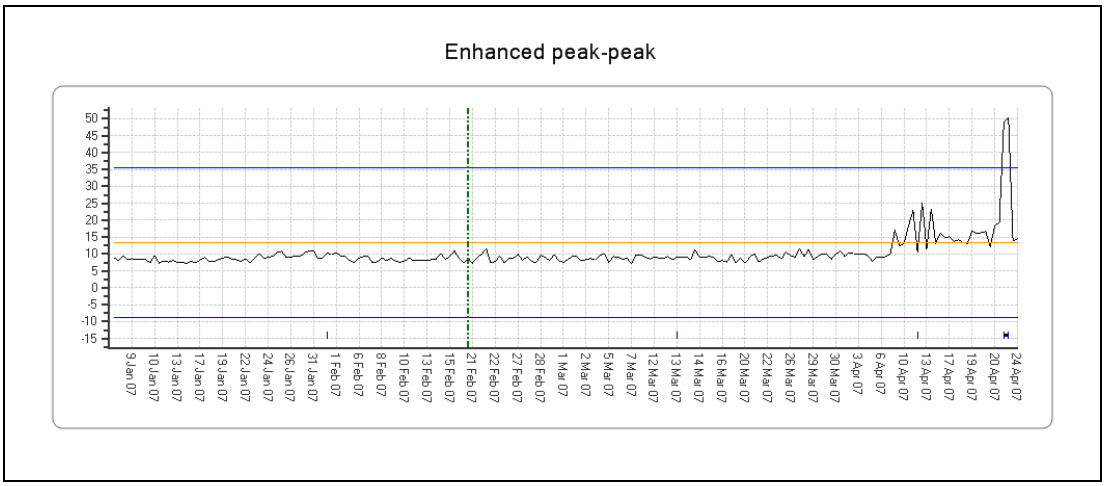

Figure 4-50 G-BMCW LHA left hydraulic idler – ESA_PP

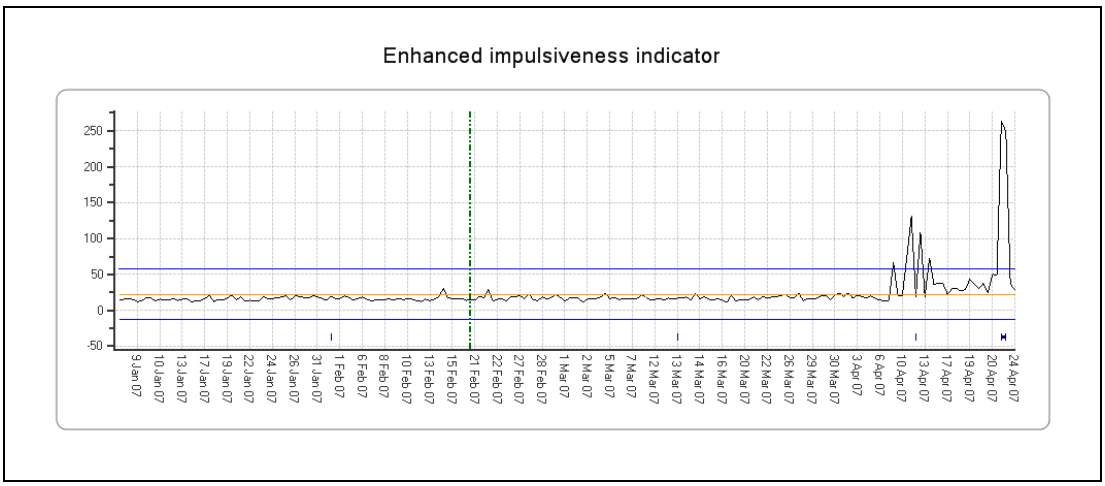

Figure 4-51 G-BMCW LHA left hydraulic idler – ESA_M6

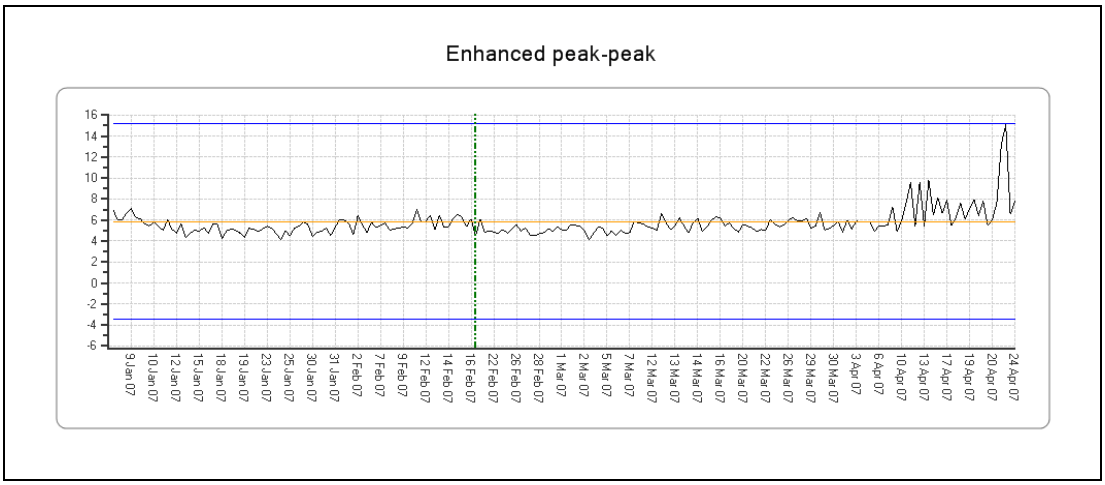

Figure 4-52 G-BMCW RHA right hydraulic idler – ESA_PP

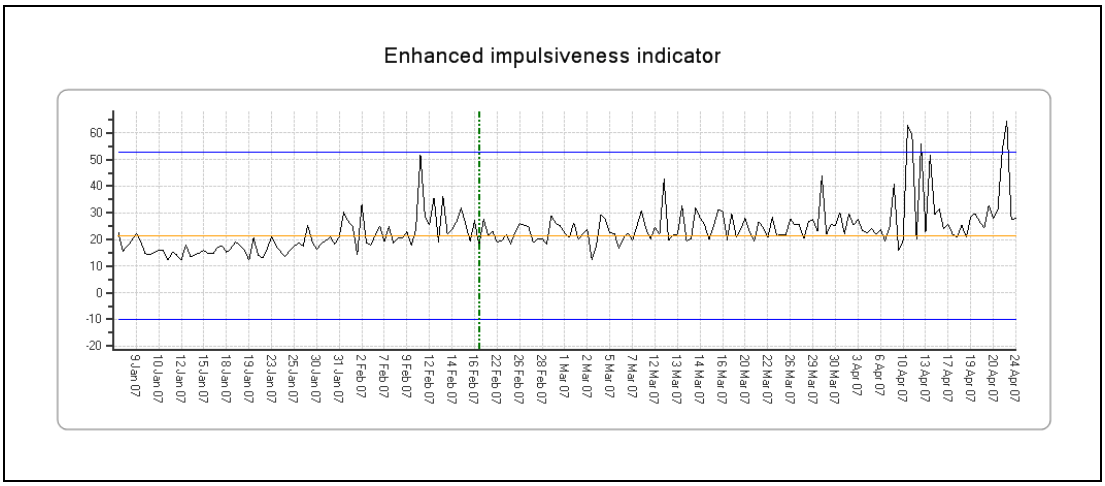

Figure 4-53 G-BMCW RHA right hydraulic idler – ESA_M6

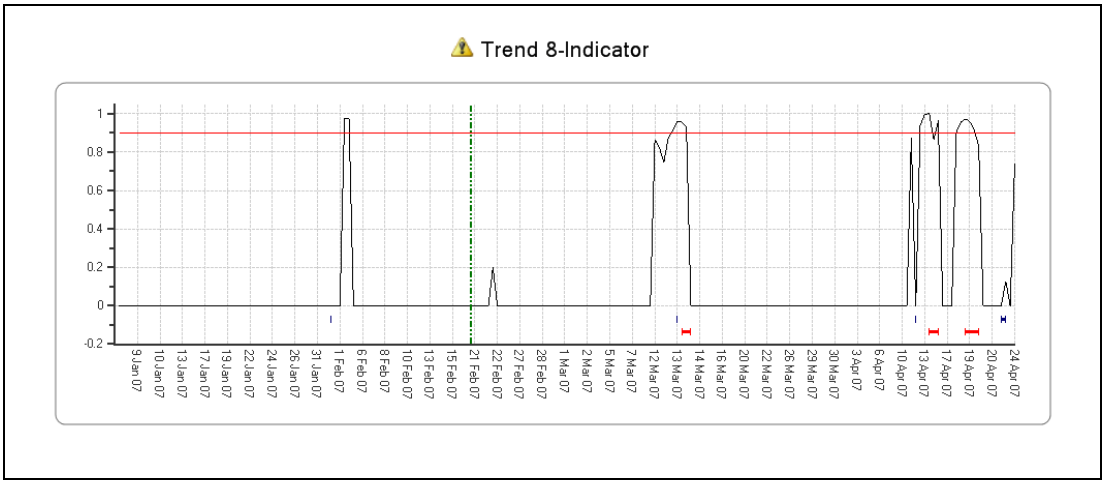

Figure 4-54 G-BMCW LHA left hydraulic idler – 8IT model PA

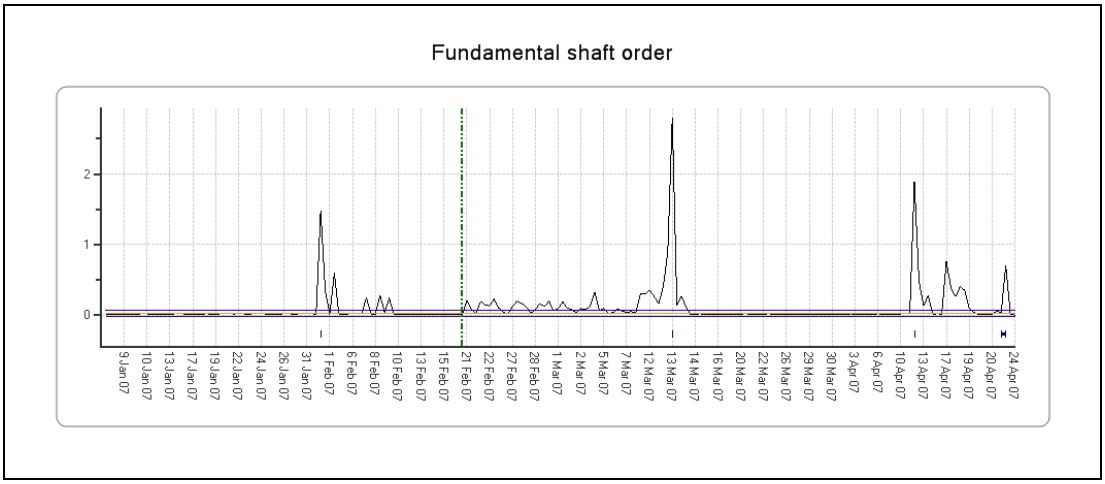

Figure 4-55 G-BMCW LHA left hydraulic idler – FSA_SO1

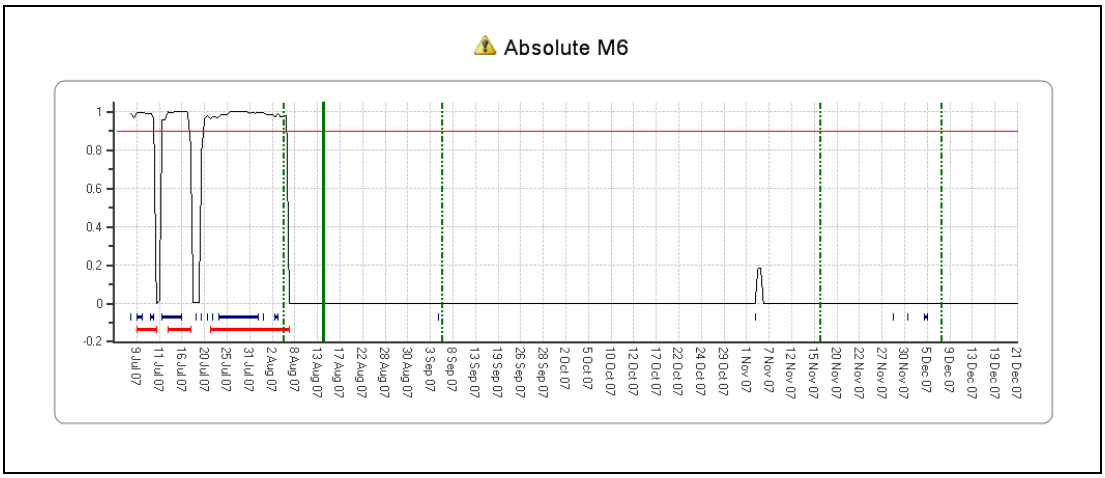

Figure 4-56 G-BMCW LHA left hydraulic idler – M6A model PA

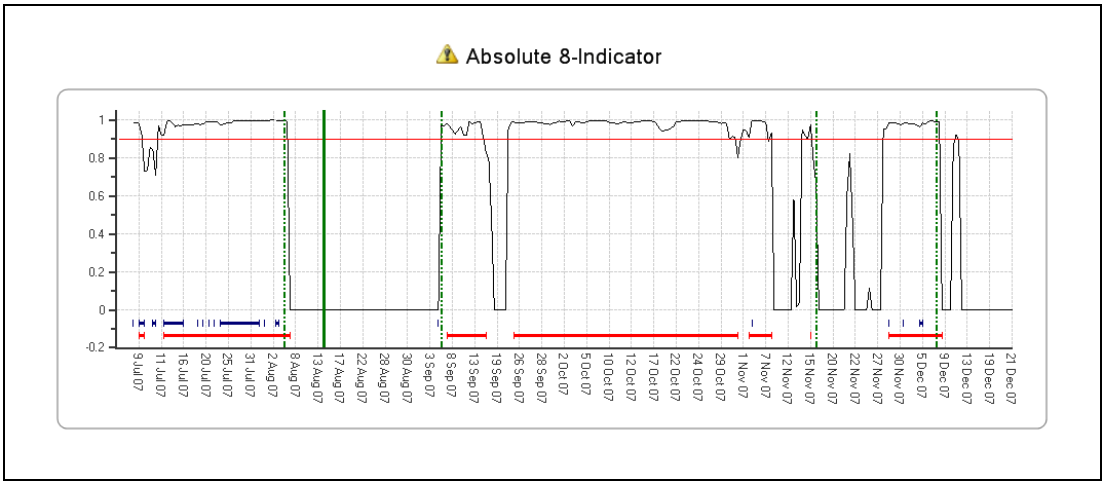

Figure 4-57 G-BMCW LHA left hydraulic idler – 8IA model PA

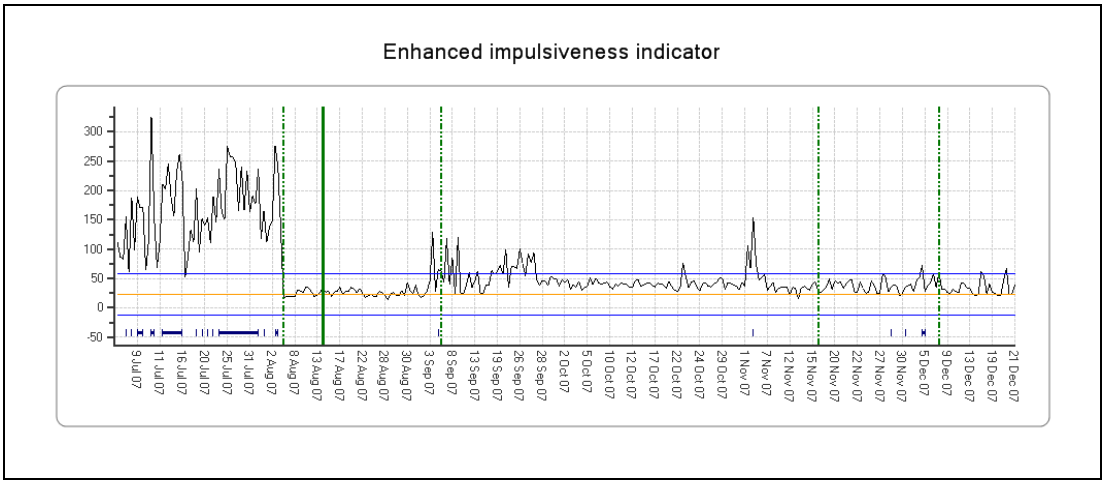

Figure 4-58 G-BMCW LHA left hydraulic idler – ESA_M6

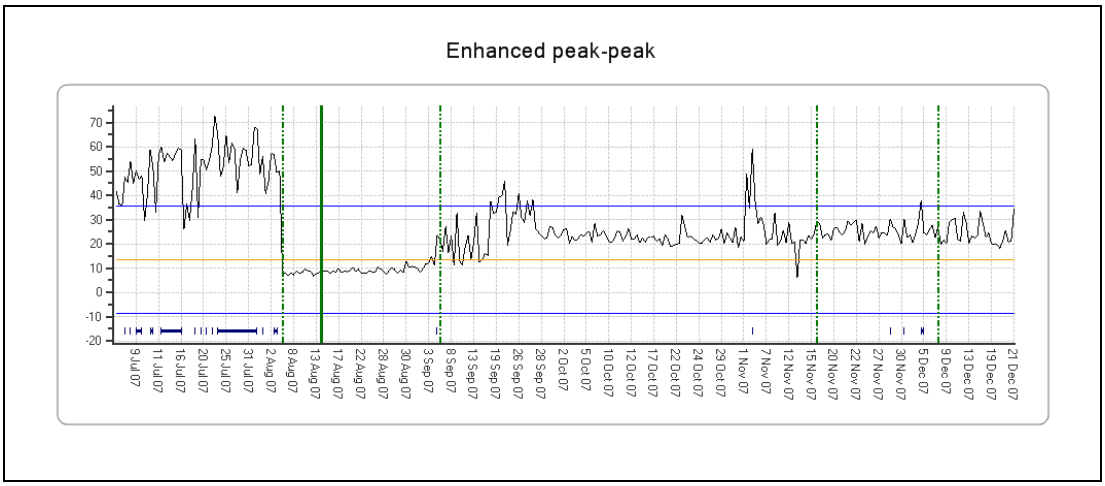

Figure 4-59 G-BMCW LHA left hydraulic idler – ESA_PP

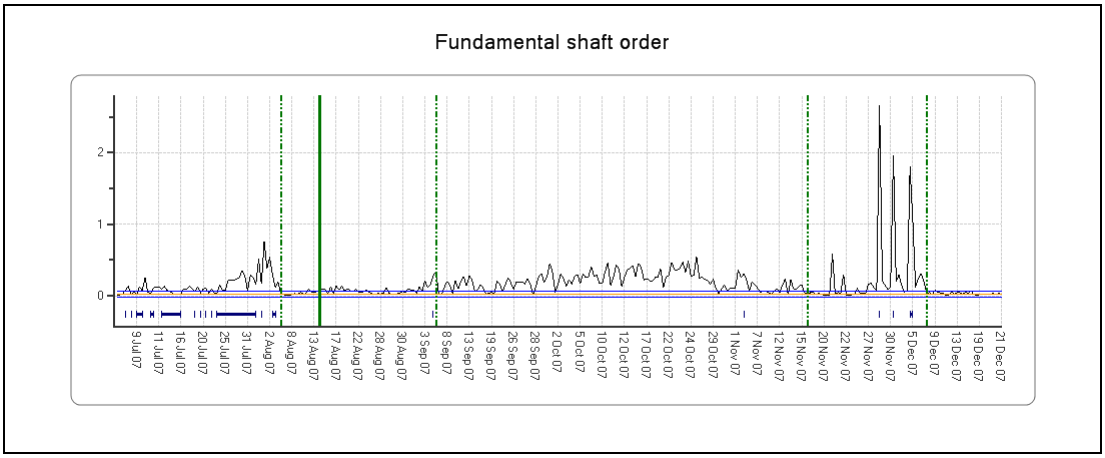

Figure 4-60 G-BMCW LHA left hydraulic idler – FSA_SO1

Example 9: G-BWWI Oil Cooler Fan

The oil cooler fan duct on G-BWWI was found to be cracked, and was replaced on 17 March 2007. The maintenance action resulted in increased FSA_S01 levels (Figure 4-62, which shows the raw CI values), triggering both IHUMS and '5-Indicator Absolute' anomaly model alerts (Figure 4-61). The oil cooler fan was replaced on 23 March 2007, and further adjustments were carried out 27 March 2007. FSA_SO1 returned to normal at that time, clearing the alerts.

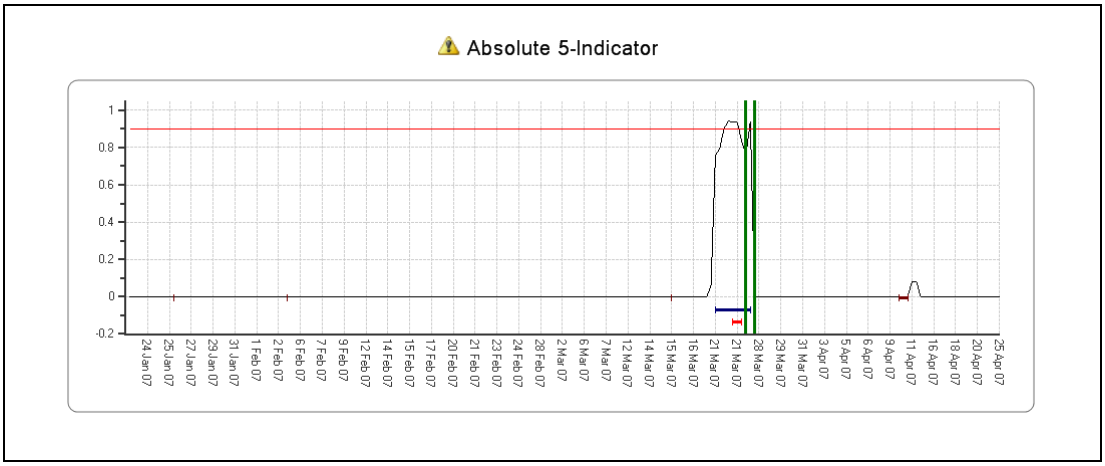

Figure 4-61 G-BWWI oil cooler fan – 5IA model PA

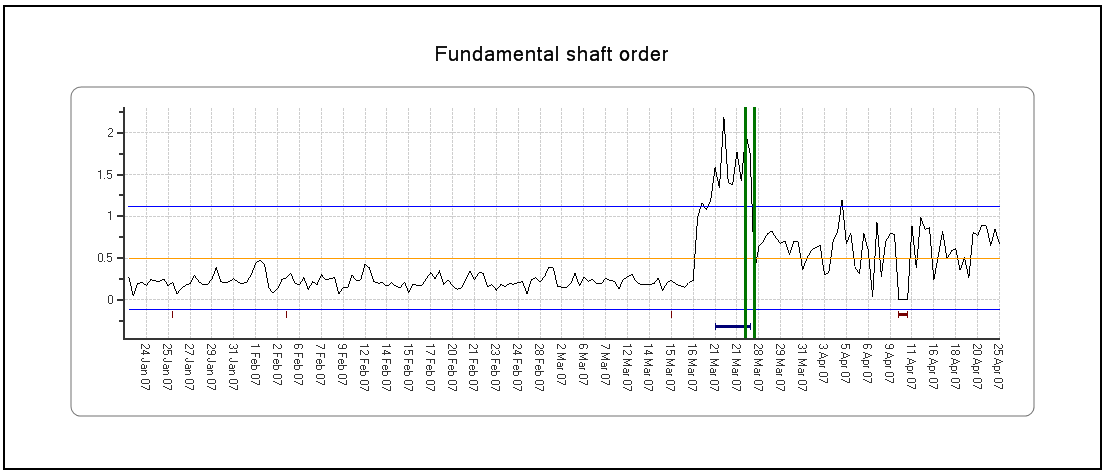

Figure 4-62 G-BWWI oil cooler fan – FSA_SO1

4.3 Cases where Anomaly Detection has Failed to Identify a Fault that was Seen by the Existing HUMS

In the second trial period no cases were identified of the anomaly modelling failing to identify a fault detected by the IHUMS.

4.4 Cases where Anomaly Detection has Identified an Existing HUMS False or Premature Alert

The only possible case of the anomaly modelling identifying a false or premature IHUMS alert was the first LHA gearbox rejection on 27 April 07 in Example 8 in Section 4.2. The anomaly modelling process did not corroborate the IHUMS alerts, and the gearbox might have continued operating satisfactorily until the MGB replacement on 2 June 2007.

4.5 Cases where Anomaly Detection has Generated a False or Premature Alert

Example 1: G-BLPM MGB RH Input Shaft

'8-Indicator Absolute' model alerts were consistently triggered on this shaft from December 2006, however there were no IHUMS alerts (Figure 4-63). Examination of the HUMS CIs showed no levels that were abnormal compared to the rest of the fleet. Similarly, examination of the standard IFs presented to the operator did not identify any CIs that were driving the anomaly indication.

GE Aviation then reviewed the other two variants of IF that had been implemented primarily to provide additional information on anomaly model behaviour. The IF providing information on the correlation between CIs did show increased levels for SIG_SD and FSA_MS, indicating that an abnormal relationship between these two CIs was the primary driver of the anomaly model alerts (Figure 4-64). Further investigation showed an abnormal correlation between the two indicators, with SIG_SD being higher than normal for a given FSA_MS (Figure 4-65). The reason for this was unknown, but no action was considered to be required. It was noted that the abnormal relationship could not have been detected using the standard method of setting thresholds on individual HUMS CIs.

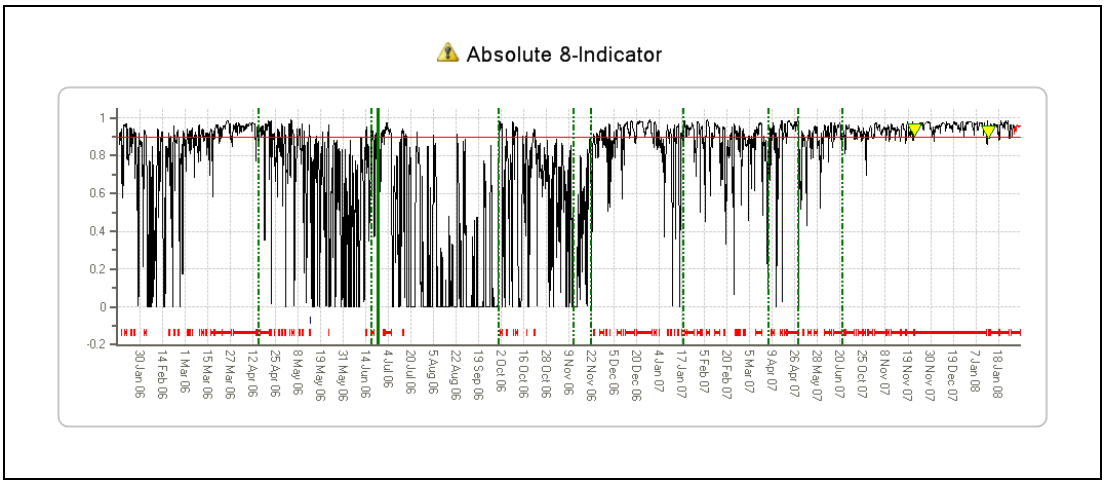

Figure 4-63 G-BLPM MGB RH input shaft – 8IA model PA

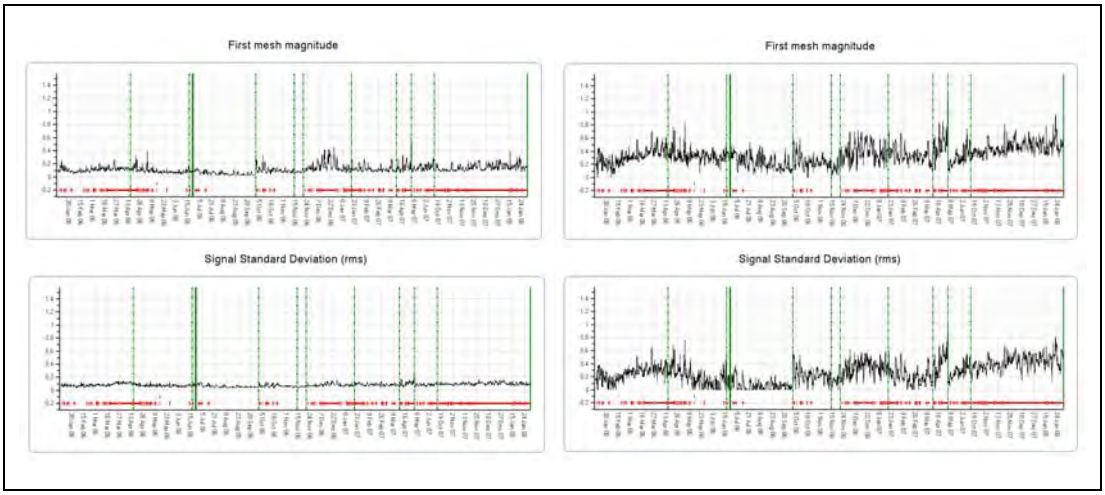

Figure 4-64 G-BLPM MGB RH input shaft; LHS top and bottom – standard IFs for FSA_MS and SIG_SD; RHS top and bottom – IFs Indicating abnormal correlation between CIs for FSA_MS and SIG_SD

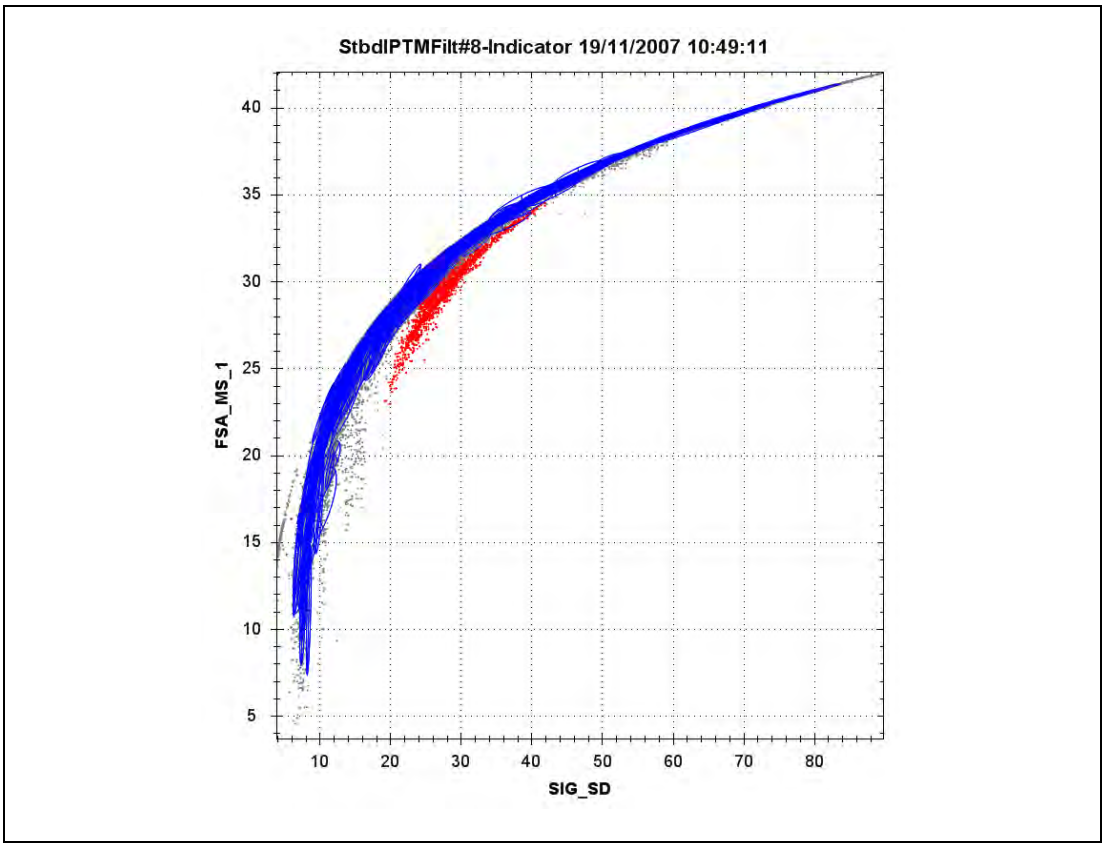

Figure 4-65 Plot of FSA_MS vs SIG_SD for MGB RH input shaft. Anomaly model cluster locations shown in blue, fleet data points (excluding G-BLPM) shown in grey (nearly all their points lie below the blue clusters), G-BLPM data points shown in red.

4.6 Cases with an Unknown Cause

The four example cases included in this section illustrate the difficulties involved in the analysis of HUMS VHM data. In all cases the advanced anomaly detection correctly detected an abnormal data trend, but it was not possible to determine the cause of the trend, and it was therefore difficult to determine whether any maintenance action should be performed.

Example 1: G-TIGV MGB LH Input Shaft

The 'M6 Absolute' model triggered a consistent alert from 25 March 2008 (Figure 4-66). This was due to a step increase in ESA_M6 to abnormal levels compared to the rest of the fleet (Figure 4-67 and Figure 4-68). The start of the alert period coincided with an engine removal and replacement for a 750 hour check. The left and right sensors were swapped, but this made no difference to the data. Bristow examined the acquired vibration signals and observed that these had an unusual characteristic, however some other aircraft were producing similar signals.

The 'M6 Absolute' model alert cleared on 25 November 2008, due to a step decrease in ESA_M6 following removal and replacement of the engine for an inspection. The cause of the increased signal impulsiveness as measured by ESA_M6 in the period between the two engine removals and replacements could not be identified.

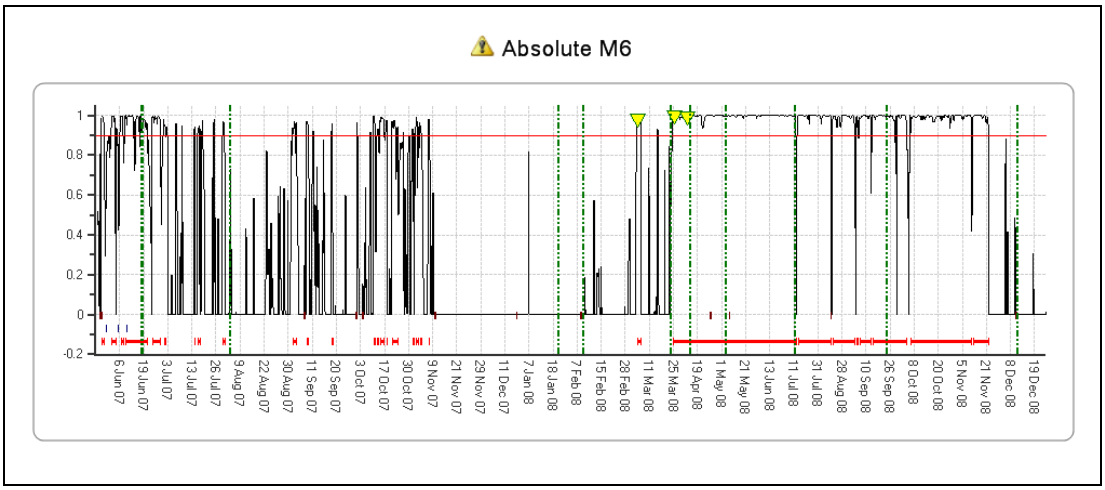

Figure 4-66 G-TIGV MGB LH input shaft – M6A model PA

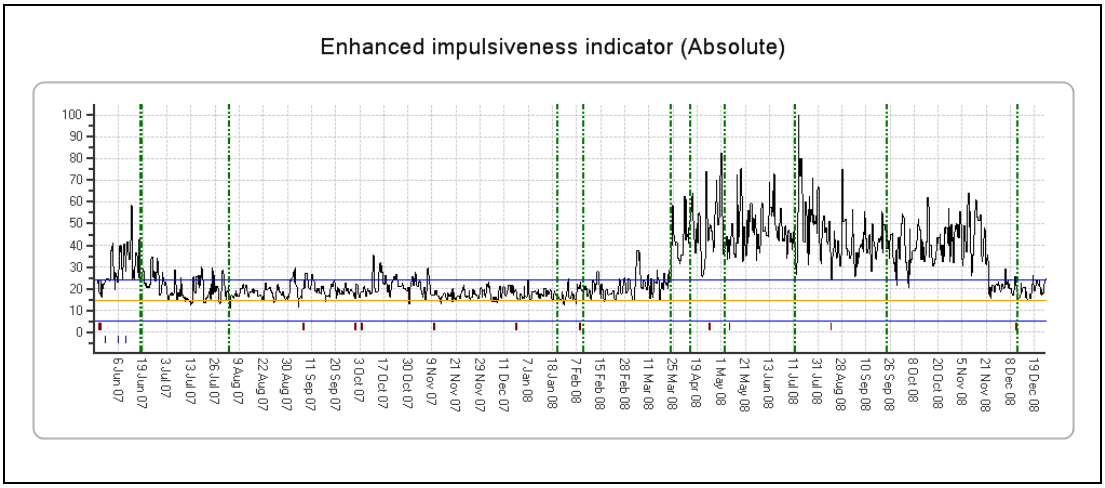

Figure 4-67 G-TIGV MGB LH input shaft – ESA_M6

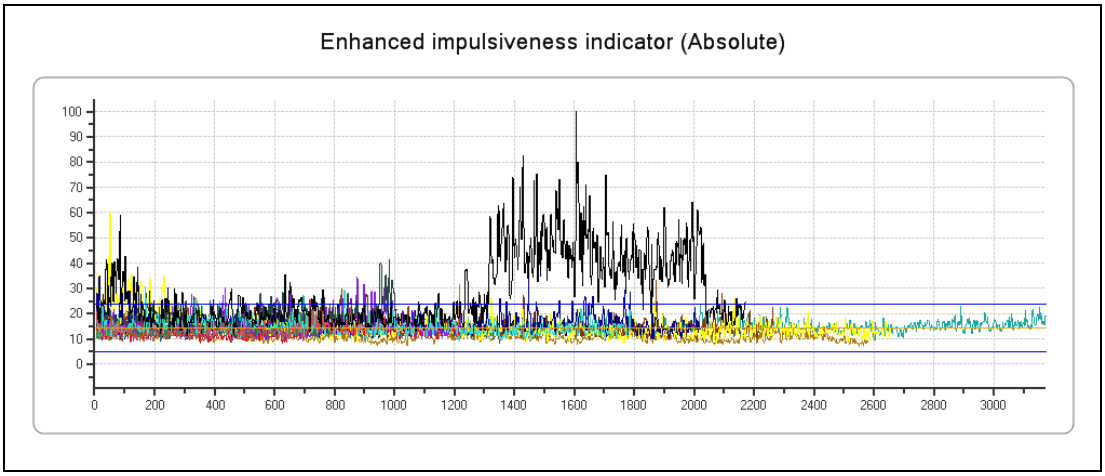

Figure 4-68 G-TIGV MGB LH input shaft – ESA_M6 – fleet view

Example 2: G-TIGJ MGB RH Input Shaft

Around the time of an engine replacement on 17 May 2008 there was an upward trend in ESA_M6, triggering 'M6 Absolute' model alerts (Figure 4-69 and Figure 4-70). ESA_M6 values were high compared to the rest of the fleet (Figure 4-71), but the trend stabilised, and then started reducing, eventually returning to normal levels. As for the previous example, the cause of this behaviour could not be identified.

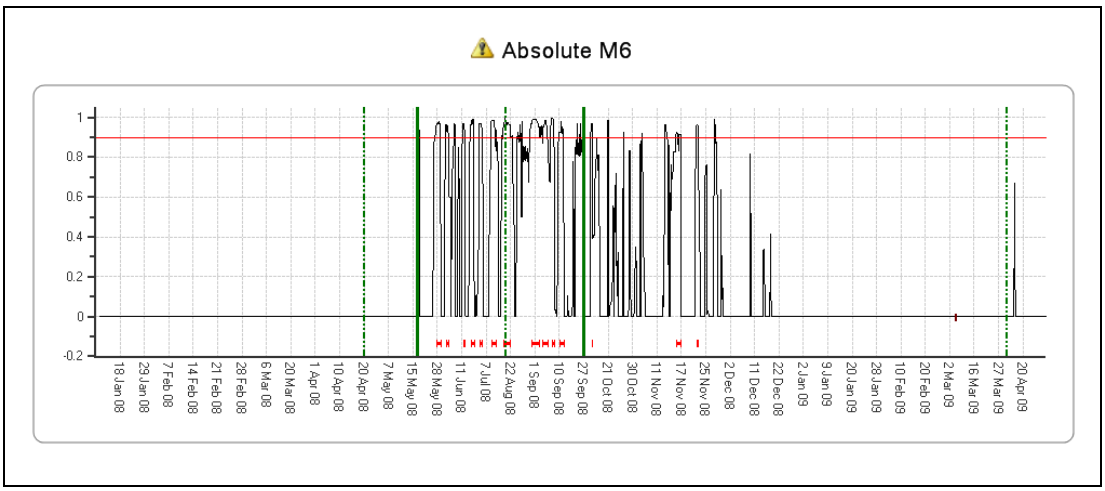

Figure 4-69 G-TIGJ MGB RH input shaft – M6A model PA

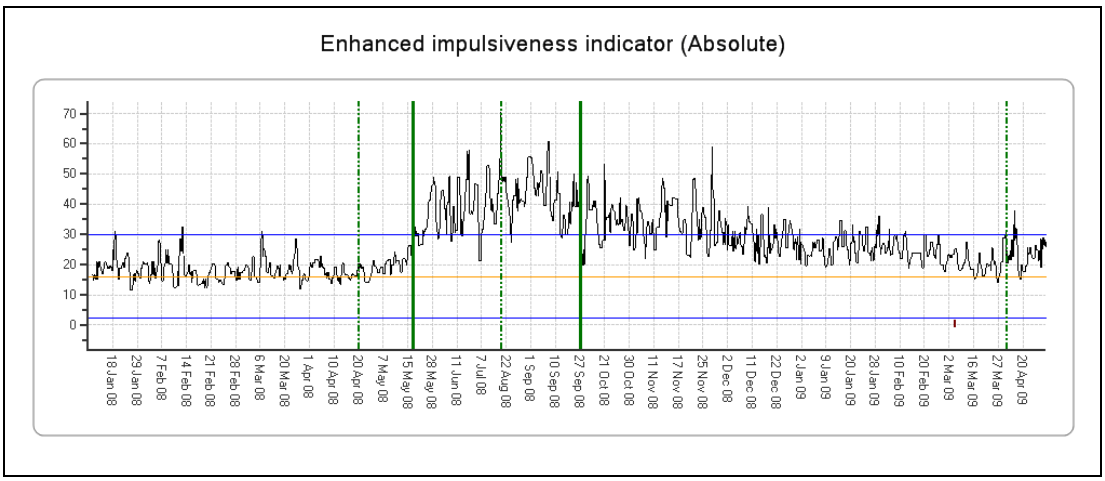

Figure 4-70 G-TIGJ MGB RH input shaft – ESA_M6

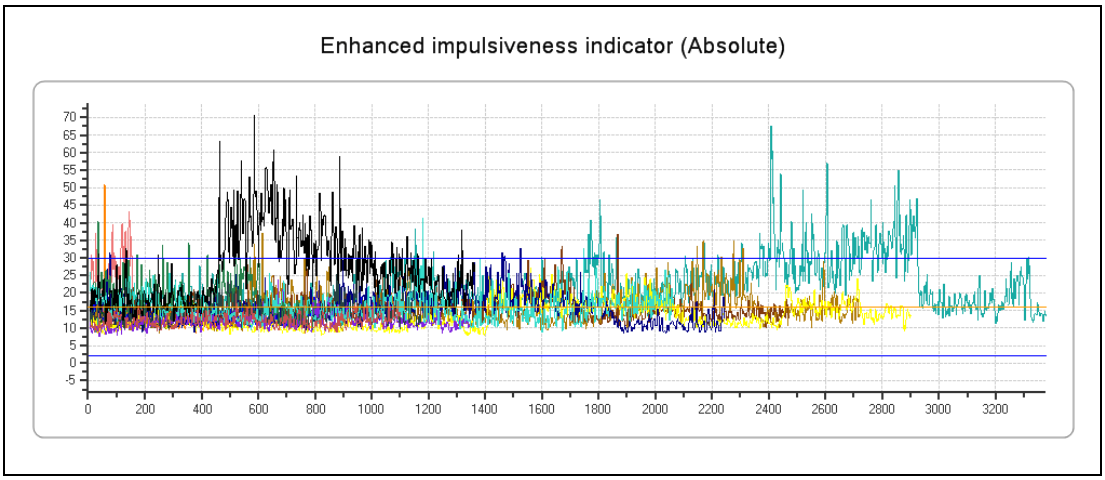

Figure 4-71 G-TIGJ MGB RH input shaft – ESA_M6 – fleet view

Example 3: G-TIGF MGB Bevel Pinion

A new Main rotor Gearbox was installed in early May 2008, and the '8-Indicator Absolute' and 'M6 Absolute' anomaly models for the bevel pinion and combiner gear were in alert from the start of the life of this gearbox, indicating abnormal data levels (Figure 4-72 and Figure 4-73). The following indicators were abnormally high compared to the fleet: ESA_PP (Figure 4-74 and Figure 4-75), ESA_SD, ESA_M6 (Figure 4-76 and Figure 4-77), FSA_SO1 (Figure 4-78 and Figure 4-79), and FSA_SO2. Although at high levels, the indicators did not show any clear rising trends. The IHUMS was also generating multiple CI and sensor alerts.

Bristow checked the sensor, and also repositioned the accelerometer mount, however this had no effect on the data. The acquired vibration signal averages were observed to contain intermittent spikes, and it was believed that the data may be indicating an instrumentation problem. After further investigation it was reported that there had been oil ingress into the microdot connector, and the connector was replaced on 28 November 2008. This caused FSA_SO1 and FSA_SO2 to return to normal levels (Figure 4-78), and cleared the IHUMS alerts. However, ESA_PP, ESA_SD and ESA_M6 remained at abnormally high levels, although ESA_M6 values gradually declined over the following months. The reason for this could not be determined, but it was speculated that it could possibly be due to a small gear tooth profile or pitching error.

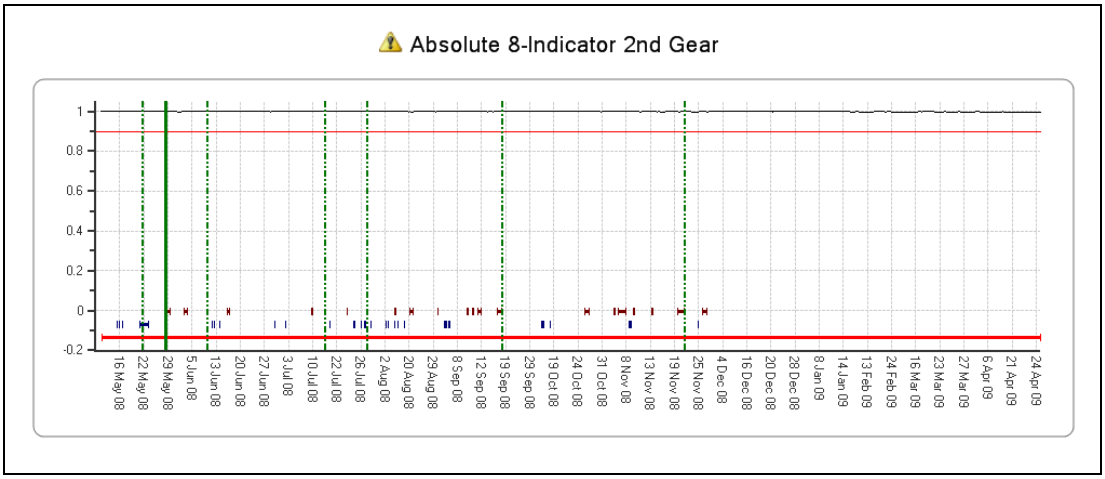

Figure 4-72 G-TIGF MGB bevel pinion – 8IA model PA

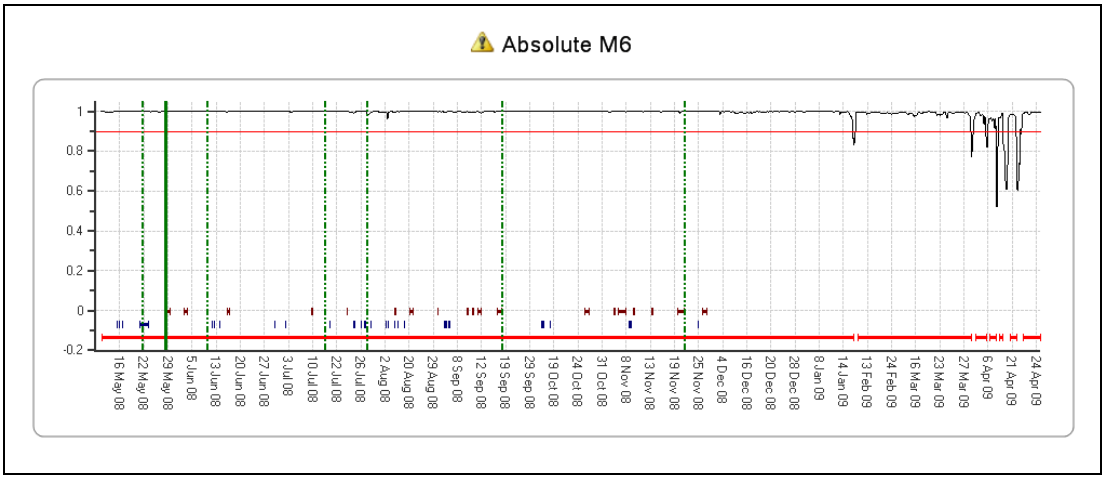

Figure 4-73 G-TIGF MGB bevel pinion – M6A model PA

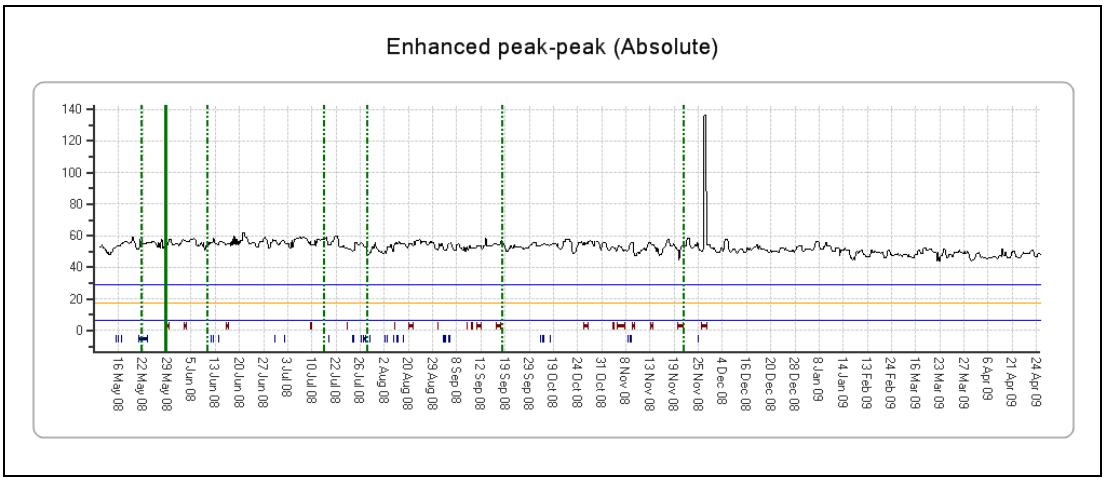

Figure 4-74 G-TIGF MGB bevel pinion – ESA_PP

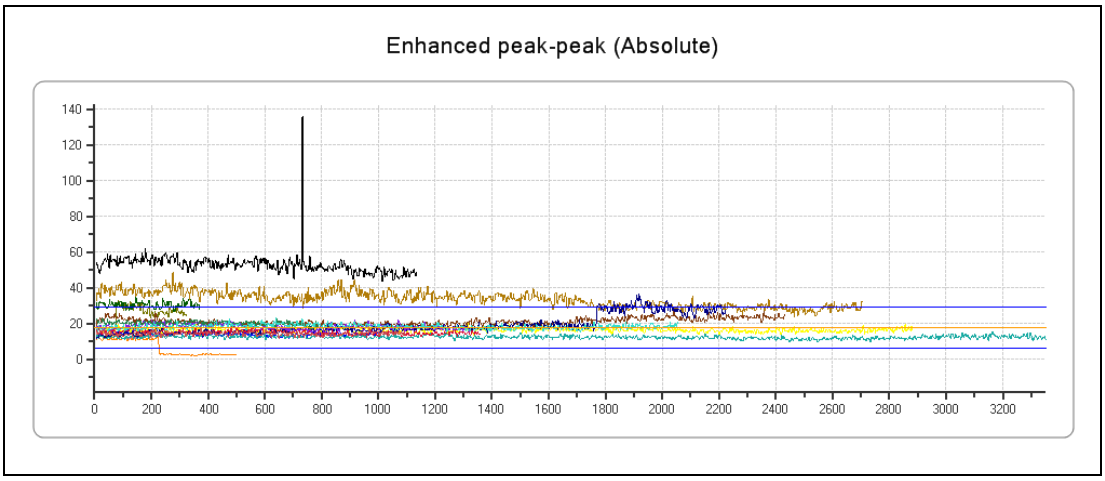

Figure 4-75 G-TIGF MGB bevel pinion – ESA_PP – fleet view

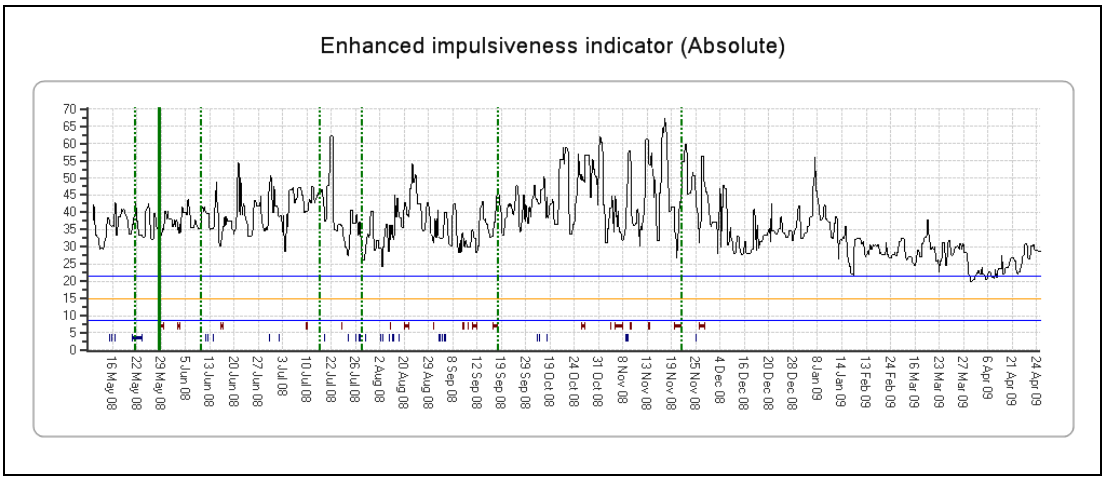

Figure 4-76 G-TIGF MGB bevel pinion – ESA_M6

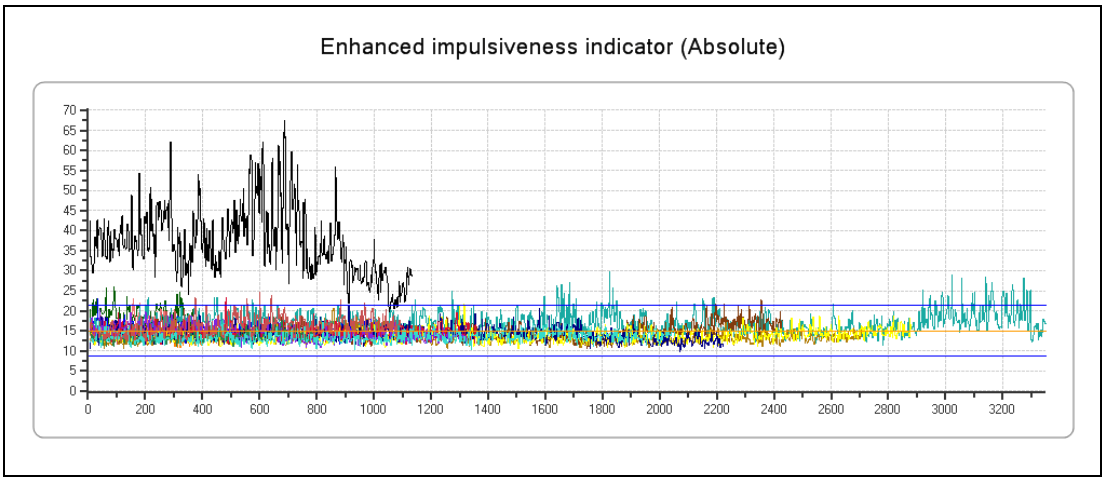

Figure 4-77 G-TIGF MGB bevel pinion – ESA_M6 – fleet view

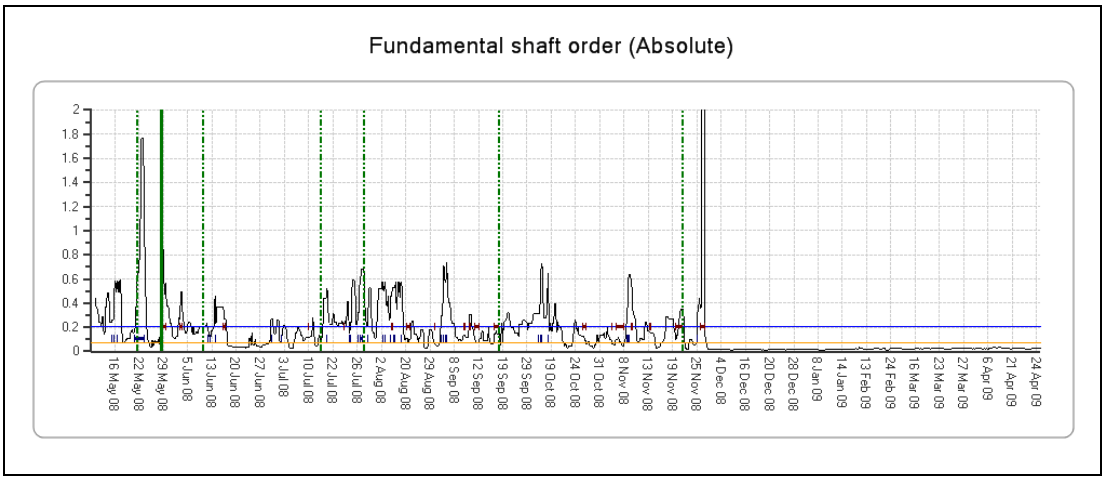

Figure 4-78 G-TIGF MGB bevel pinion – FSA_SO1

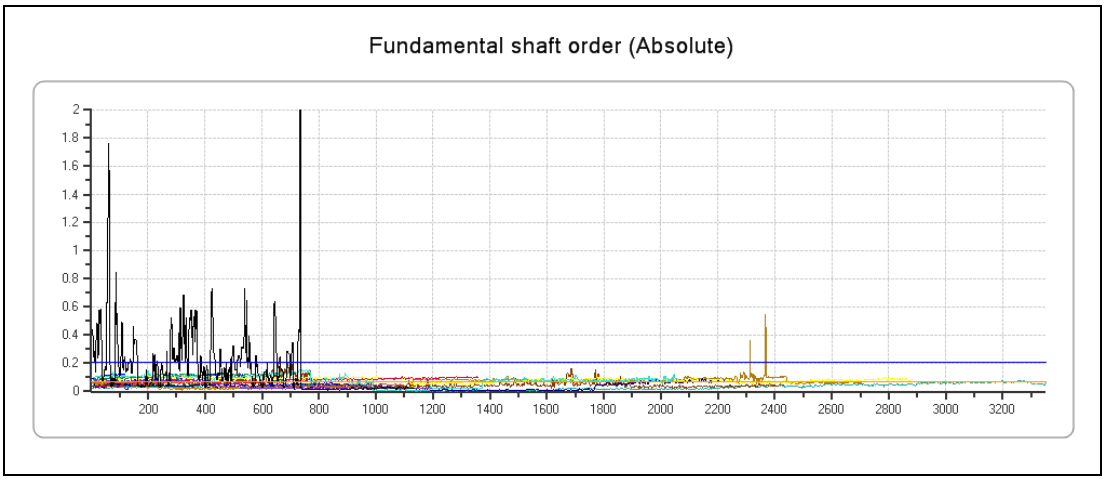

Figure 4-79 G-TIGF MGB bevel pinion – FSA_SO1 – fleet view

Example 4: G-BWZX MGB Bevel Wheel and Oil Pump Drive

Repeated 'M6 Absolute' anomaly model alerts were generated from the start of monitoring of this gear in March 2008 (Figure 4-80). The '8-Indicator Absolute' anomaly model was also generating elevated PA values (Figure 4-81). The cause of the 'M6 Absolute' model alerts was abnormally high values of the derived CI WEA, which is the ratio of ESA_SD to SIG_SD (Figure 4-82). These high values were not caused by a high ESA_SD, but an abnormally low SIG_SD (Figure 4-83). However the most notable feature of the data was the very low amplitude of the bevel wheel mesh tone, as measured by FSA_MS (Figure 4-84 and Figure 4-85). The abnormal data did not trigger any IHUMS alerts.

A similar case reported in reference [3] was attributed to the wiring for sensors 3 and 4 having been swapped, so that the gear was being monitored from the wrong sensor. The wiring was checked, but this time was found to be correct. In addition, the data from the right torque shaft – aft end, monitored from the same sensor, was checked and found to be ok. This indicated that there was no problem with the sensor.

Over a period of approximately one month from early January 2009 the amplitude of the bevel wheel mesh tone progressively increased, until it was within the bounds of normality for the fleet (Figure 4-84). The anomaly model alerts cleared at that time. No explanation could be found for the initial abnormally low gear mesh tone, or why it then progressively increased.

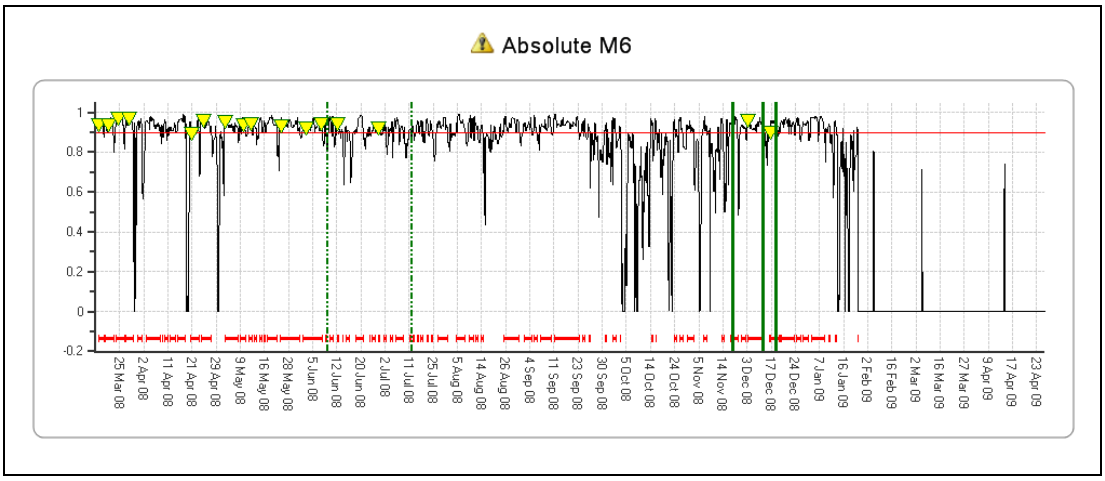

Figure 4-80 G-BWZX MGB bevel wheel and oil pump drive – M6A model PA

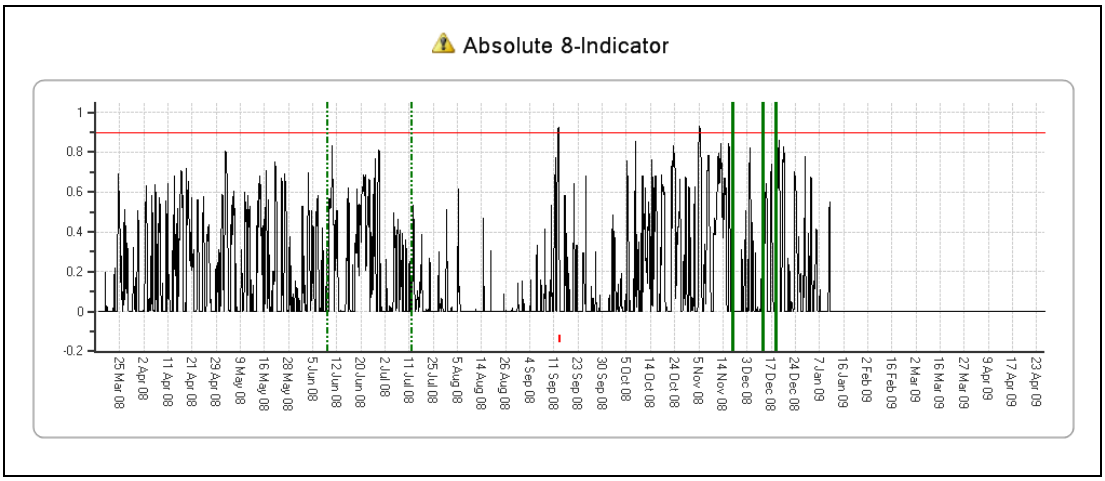

Figure 4-81 G-BWZX MGB bevel wheel and oil pump drive – 8IA model PA

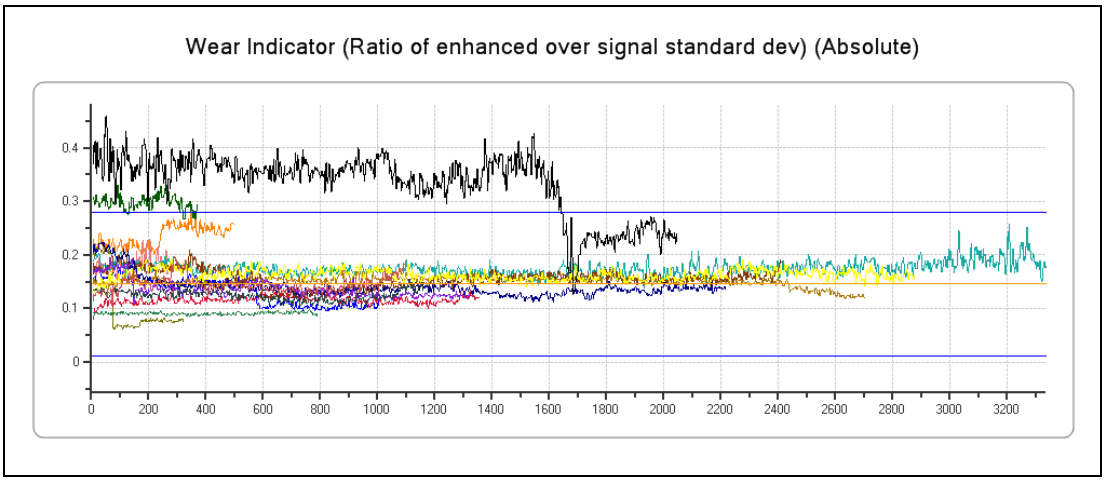

Figure 4-82 G-BWZX MGB bevel wheel and oil pump drive – WEA – fleet view

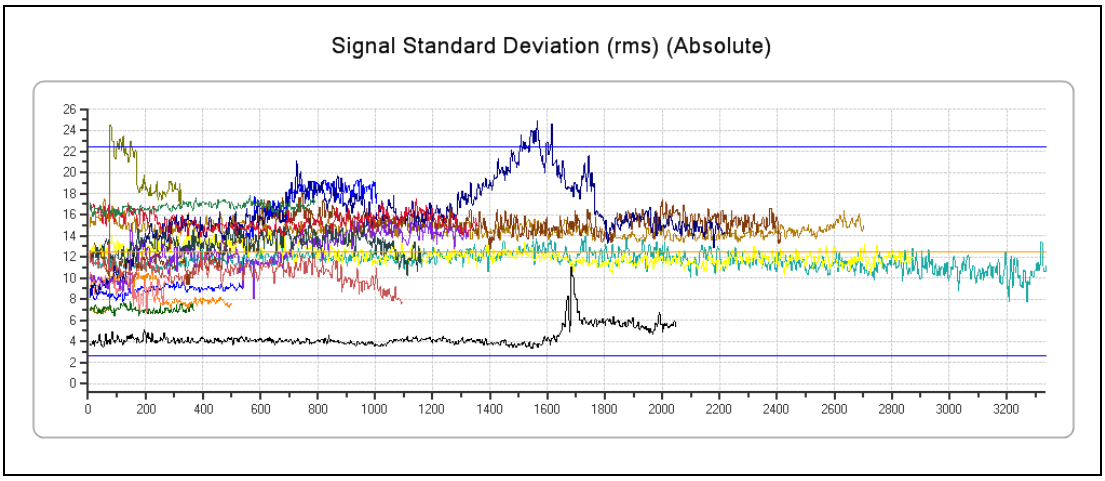

Figure 4-83 G-BWZX MGB bevel wheel and oil pump drive – SIG_SD – fleet view

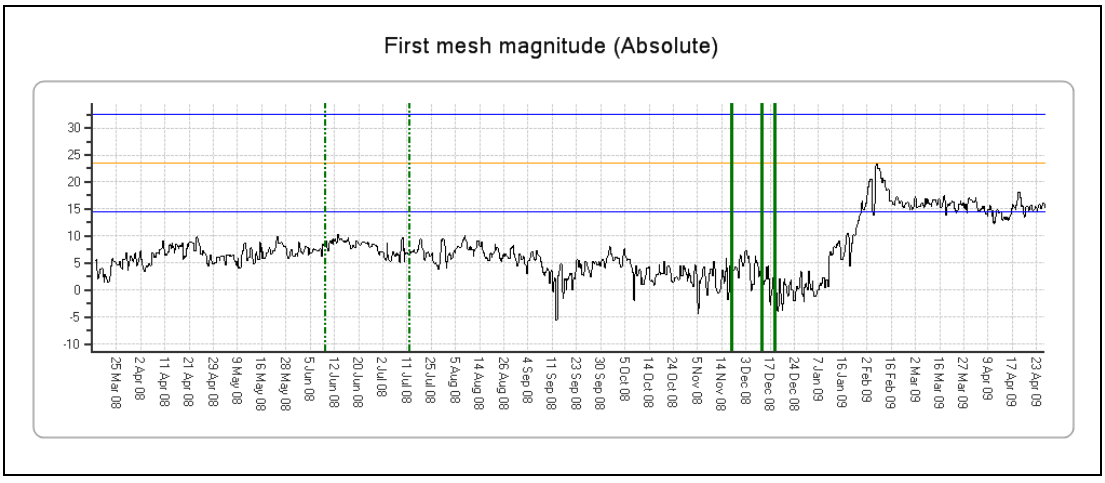

Figure 4-84 G-BWZX MGB bevel wheel and oil pump drive – FSA_MS

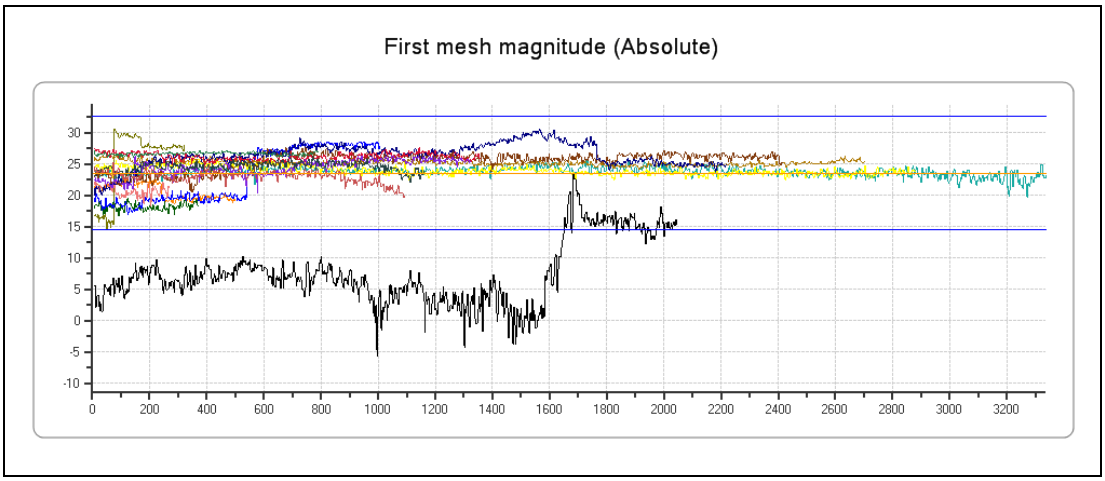

Figure 4-85 G-BWZX MGB bevel wheel and oil pump drive – FSA_MS – fleet view

5 Statistical Analysis of System Performance

A statistical analysis has been performed of the alerts generated during the trial to assess the performance of the advanced anomaly detection system. The statistics presented apply to the second trial period from 7 January to 19 December 2008. However, for comparison purposes, some of the statistics presented in reference [3] for the first six-month trial period from 1 June to 30 November 2006 are repeated in Section 5.1.

As part of the assessment of anomaly detection system performance, the number of alerts generated by this system has been compared to those generated by the IHUMS. The IHUMS alert figures combine both sensor and CI (or 'defect') alerts, as the anomaly detection system alerts do not differentiate between these. The IHUMS alert information has been obtained from the CI data files extracted by Bristow from the IHUMS ground station. However, it should be noted that a new 'm out of n' alert filtering process was implemented in the IHUMS ground station between the two trial periods. Unfortunately the outputs from this process are not included in the CI data files, and so are not available for analysis. Therefore the IHUMS alert figures for the second trial period represent the 'raw' alert data only, not what may be presented to the IHUMS operator. (It is understood that operators have a choice as to whether or not to use the new alert filtering process).

In 2008, while the second trial was in progress, GE Aviation undertook another project for the UK Ministry of Defence (MoD), configuring and trialling an advanced HUMS anomaly detection system for the UK fleet of Chinook helicopters. The trial involved the evaluation of a HUMS data analysis service provided to the MoD's Materials Integrity Group (MIG), and had a duration of three months. Based on the CAA trial experience, it was considered that an alert threshold set at a PA value of 0.90 may be too sensitive, and that there should be a higher probability that data is anomalous before an alert is triggered. Following an investigation into the effects of varying the PA threshold on the alert rate and fault detection performance, a PA threshold of 0.98 was selected for the Chinook trial. The number of findings during the relatively short trial confirmed that the system is still an effective detector of anomalous HUMS CI data using this PA threshold. Therefore, again for comparison purposes, statistics have been calculated for the number of alerts that would have been generated in the second CAA trial period if the PA threshold had been set at 0.98 instead of 0.90.

5.1 Total Anomaly Detection System and IHUMS Alerts

Table 5-1 presents a summary of the anomaly detection system and IHUMS alerts generated during the two trial periods. Anomaly detection system alert statistics are shown for PA thresholds of both 0.90 and 0.98.

The IHUMS performs a total of 32 drive train component analyses. There are three drive train components that have two gears on the monitored shaft. For each of these, separate anomaly models have been created for the two gears, using different gear mesh-related CIs in each of the models. As a result, the anomaly detection system performs 35 drive train component analyses. The comparison between the anomaly detection system and the IHUMS is therefore made on the basis of both the total number of alerts generated, and the alerts per analysis.

Although the first trial period had a duration of six months, and the second period a duration of almost one year, there was a decrease in the flying rate of Bristow's AS332L fleet between the two periods as new aircraft such as the EC225 were introduced into North Sea operations, and a few of the AS332Ls were transitioned overseas. The total flying hours for the AS332L fleet included in the anomaly

detection trial were 9,951 hours for the first trial period and 11,096 hours for the second period. Therefore the most meaningful comparison of the alert rate between the two trial periods is on the basis of alerts per flying hour, rather than alerts per day.

For the first trial period, the advanced anomaly detection system alert thresholds were set on the FS values output from the anomaly models. The thresholds were calculated automatically, based on a probability distribution fitted to the FS values derived from the training database, and varied from one model to the next. For the second trial period the threshold was based on the PA output, with a global threshold at a PA of 0.90 being applied to all models. This PA threshold value was chosen to give a similar sensitivity to the FS thresholds used in the first trial period.

The 'IHUMS' data presented in Table 5-1 are the 'raw' alert statistics referred to above. The 'Anomaly FS / PA 0.90' data shows the alerts generated by the anomaly detection system based on FS thresholds in the first trial, and the 0.90 PA threshold in the second trial. The 'Anomaly PA 0.98' data is for the second trial period only, and shows the number of alerts that would be generated if a 0.98 PA threshold had been applied.

Table 5-1 Alert summary

Alert Statistics	First Trial Period	Second Trial Period
Total Alerts		
IHUMS	1381	1749
Anomaly FS / PA 0.90	1484	1509
Anomaly PA 0.98		587
Alerts per Analysis		
IHUMS	43.2	54.7
Anomaly FS / PA 0.90	42.4	43.1
Anomaly PA 0.98		16.8
Alerts per Flying Hour		
IHUMS	0.139	0.158
Anomaly FS / PA 0.90	0.149	0.136
Anomaly PA 0.98		0.053

As discussed in reference [3], for the first trial period the alert rate of the anomaly detection system was comparable to that of the IHUMS, with the anomaly detection system generating slightly more alerts in total, and slightly fewer alerts per analysis. In both cases, the total number of alerts increased in the second trial period, however this was of longer duration, and involved a greater number of flying hours.

When the data is factored by the total flying hours, the IHUMS still generated more 'raw' alerts in the second trial period than the first. The reason for this is unknown, and could be due to the change in operations, with some aircraft transitioning overseas. However, where the new 'm out of n' alert filtering is applied, the IHUMS operator would see fewer alerts than the numbers shown.

The model rebuilding performed between the two trial periods resulted in approximately a 10% reduction in the alert rate of the anomaly detection system from 0.149 to 0.136 per flying hour. The figure for the second trial period was approximately 15% lower than the 'raw' IHUMS alert rate. Given that the IHUMS alert rate actually increased between the trial periods, if there had been no change in aircraft operations the reduction in the anomaly detection system alert rate could have been greater than 10%. Table 5-1 shows that increasing the threshold to a PA value of 0.98 would have a significant effect on the alert rate, reducing this by over 60% to 0.053 per flying hour.

With a PA threshold of 0.98, 14 of the 15 example cases described in Section 4 would still trigger anomaly alerts. The only one that would not now trigger an alert is Example 9 in Section 4.2, where there was a short period of increased FSA_S01 values on the oil cooler fan following oil cooler maintenance. This resulted in three PA values in the range 0.935-0.940, which exceeded the 0.90 PA threshold and triggered a single anomaly alert.

5.2 Anomaly Detection System and IHUMS Alerts by Sensor

Figure 5-1 shows the number of IHUMS and anomaly detection system alerts that have been triggered on data from each IHUMS sensor, and also the number of anomaly detection system alerts that would have been triggered with a PA threshold of 0.98. As the number of component analyses performed using data from each sensor varies, the alert figures per sensor have been normalised by the number of analyses using that sensor to give a more meaningful comparison between sensors.

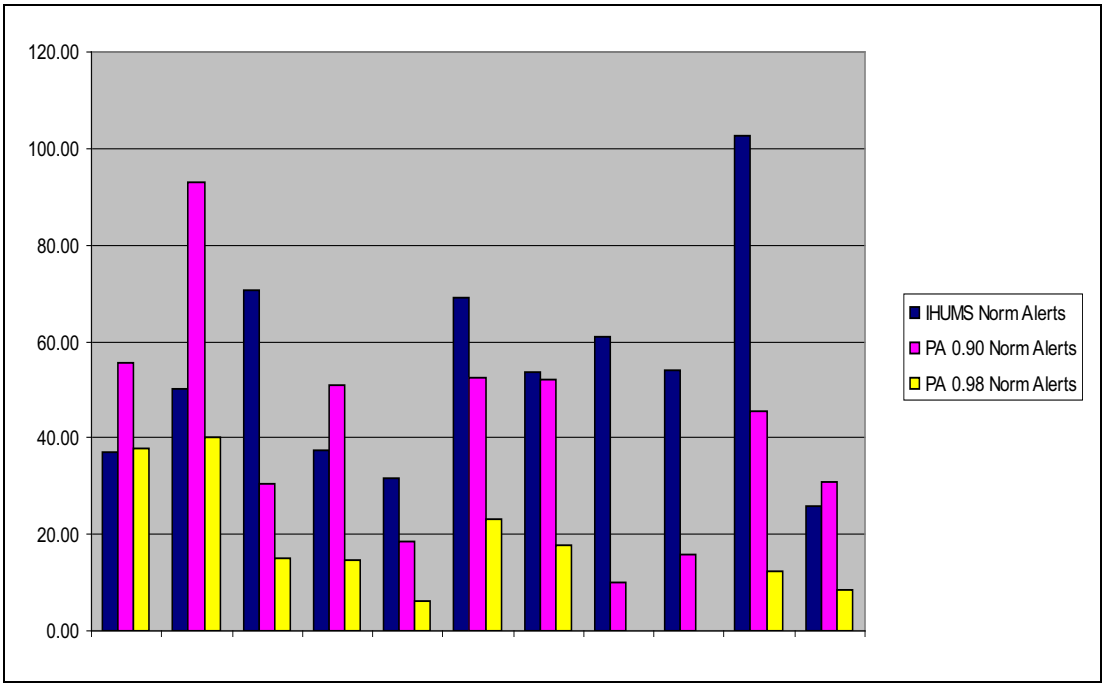

Figure 5-1 Anomaly detection system and IHUMS alerts by sensor (normalised)

Figure 5-2 compares the number of individual data points in alert for each sensor, with the data again being normalised by the number of analyses performed. The fact that there are more data points in alert for the anomaly detection system than the IHUMS is due to the longer duration of the anomaly model alerts. There are several reasons for this difference. The IHUMS is the "executive system", and therefore IHUMS alerts must be responded to. Also the IHUMS uses datum thresholds, which are calculated based on a limited number of data samples. As a result, the IHUMS may not alert on data that is consistently anomalous compared to the fleet norm. The

system is also normally re-datumed after an alert has been investigated, which would re-set the threshold and clear the alert. However, the anomaly detection system will continue to generate alerts if the data remains abnormal.

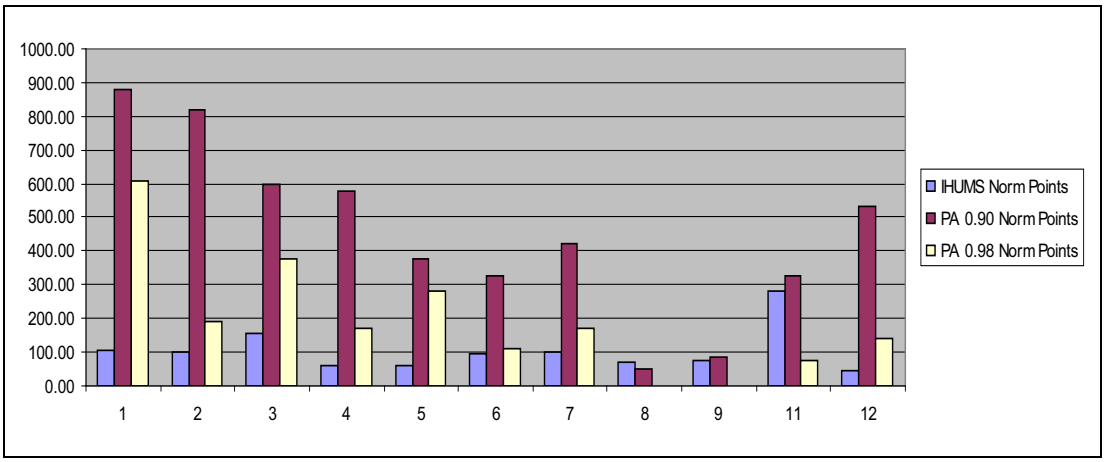

Figure 5-2 Anomaly detection system and IHUMS data points in alert by sensor (normalised)

The greatest number of anomaly detection system alerts have been triggered on sensors 1 and 2, monitoring the MGB LH and RH input shafts and left and right torque shafts. These sensors have also generated the greatest number of data points in alert. The statistics are consistent with the evidence presented in Section 4, where four of the 15 example cases were on the MGB input shafts, with alerts being present for a considerable period of time as the causes of the anomalous IHUMS CI data were unknown. The data in Figure 5-2 in particular can be skewed by a small number of aircraft generating long periods of anomalous data. For example, in the first trial period the greatest number of data points in alert occurred on sensor 7, and this was attributed to wiring harness problems on a couple of aircraft (reference [3]). The greatest numbers of IHUMS alerts were triggered on sensor 11, monitoring the Tail rotor Gearbox (Figure 5-1). This again is consistent with the evidence presented in Section 4, where there were two example cases of protracted Tail Rotor and TGB sensor issues, with many associated IHUMS alerts.

To confirm the source of the alerts on sensors 1 and 2, Figure 5-3 and Figure 5-4 present pie charts showing, for each sensor, the proportions of anomaly detection system data points in alert from each aircraft included in the trial. The data points in alert from sensor 1 are dominated by two aircraft; G-TIGC (45% of total), and G-TIGV (41% of total) (Figure 5-3). Raised and variable ESA_M6 levels on the MGB left torque shaft – fwd end generated many 'M6 Absolute' model alerts on G-TIGC. Similarly, a period of high ESA_M6 levels between two engine removals and replacements generated persistent 'M6 Absolute' model alerts on G-TIGV (see Example 1 in Section 4.6). The two aircraft generating the highest number of data points in alert from sensor 2 are G-TIGC (32% of total) and G-BLPM (22% of total) (Figure 5-4). Both of these aircraft are included in the example cases presented in Section 4 (see Example 6 in Section 4.2, and Example 1 in Section 4.5).

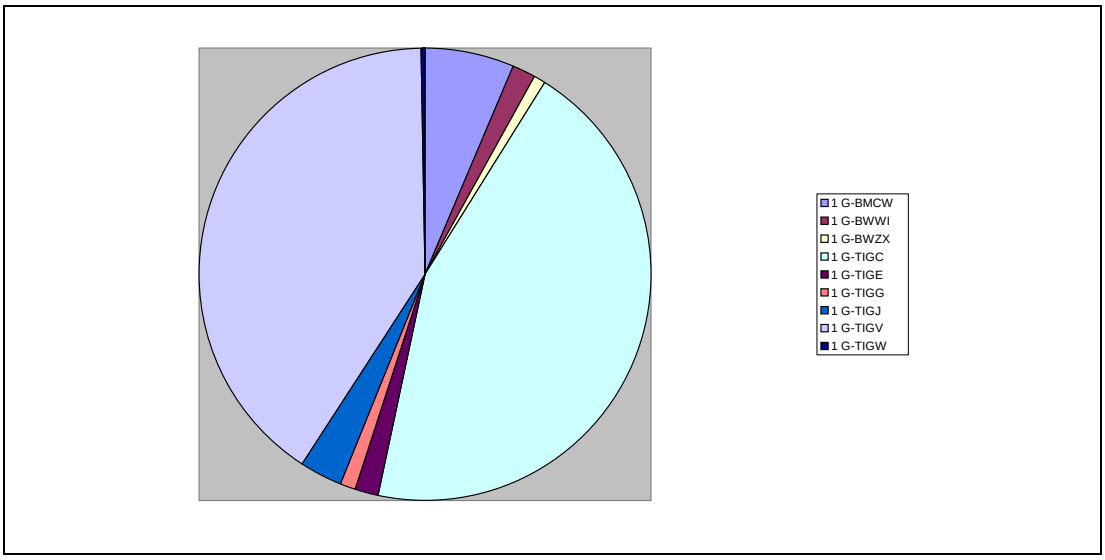

Figure 5-3 Anomaly detection system data points in alert by aircraft for sensor 1

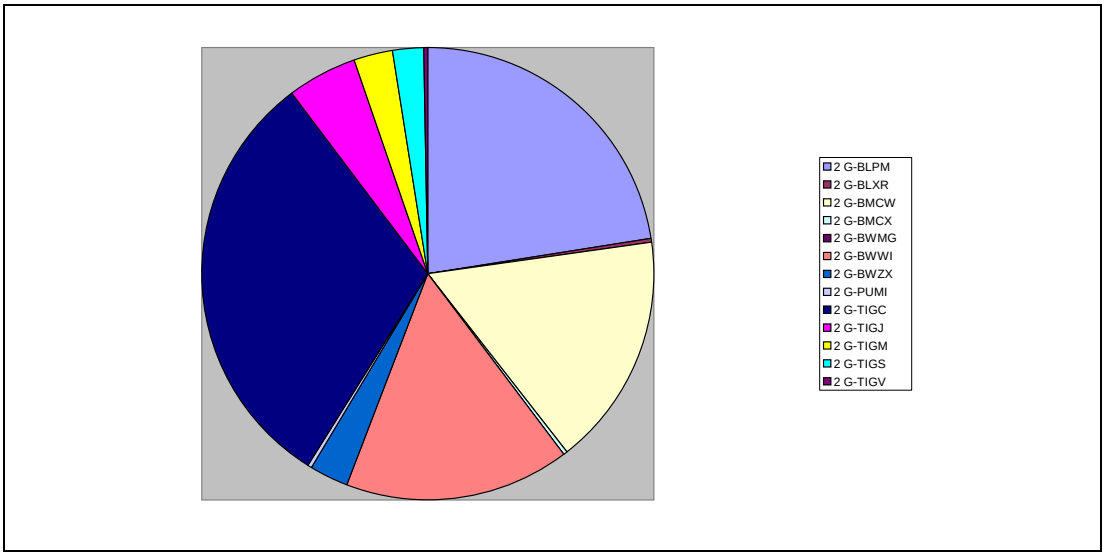

Figure 5-4 Anomaly detection system data points in alert by aircraft for sensor 2

5.3 Anomaly Detection System and IHUMS Alerts by Analysis

Figure 5-5 shows the number of anomaly detection system and IHUMS alerts that have been triggered from the analysis of each drive train component. It also includes the number of anomaly detection system alerts that would have been triggered with a PA threshold of 0.98. Figure 5-6 shows the number of individual data points in alert by component analysis.

The anomaly detection system (PA = 0.9) has triggered most alerts on the MGB RH input shaft (sensor 2), confirming the evidence presented in the previous section. In the first trial period the most anomaly detection system alerts occurred on the MGB left torque shaft – aft end (reference [3]). This again suggests that the relative analysis alert rates are influenced by particular aircraft issues that are present during the period of monitoring. Most IHUMS alerts were triggered on the TGB output shaft (sensor 11), also confirming the previous evidence.

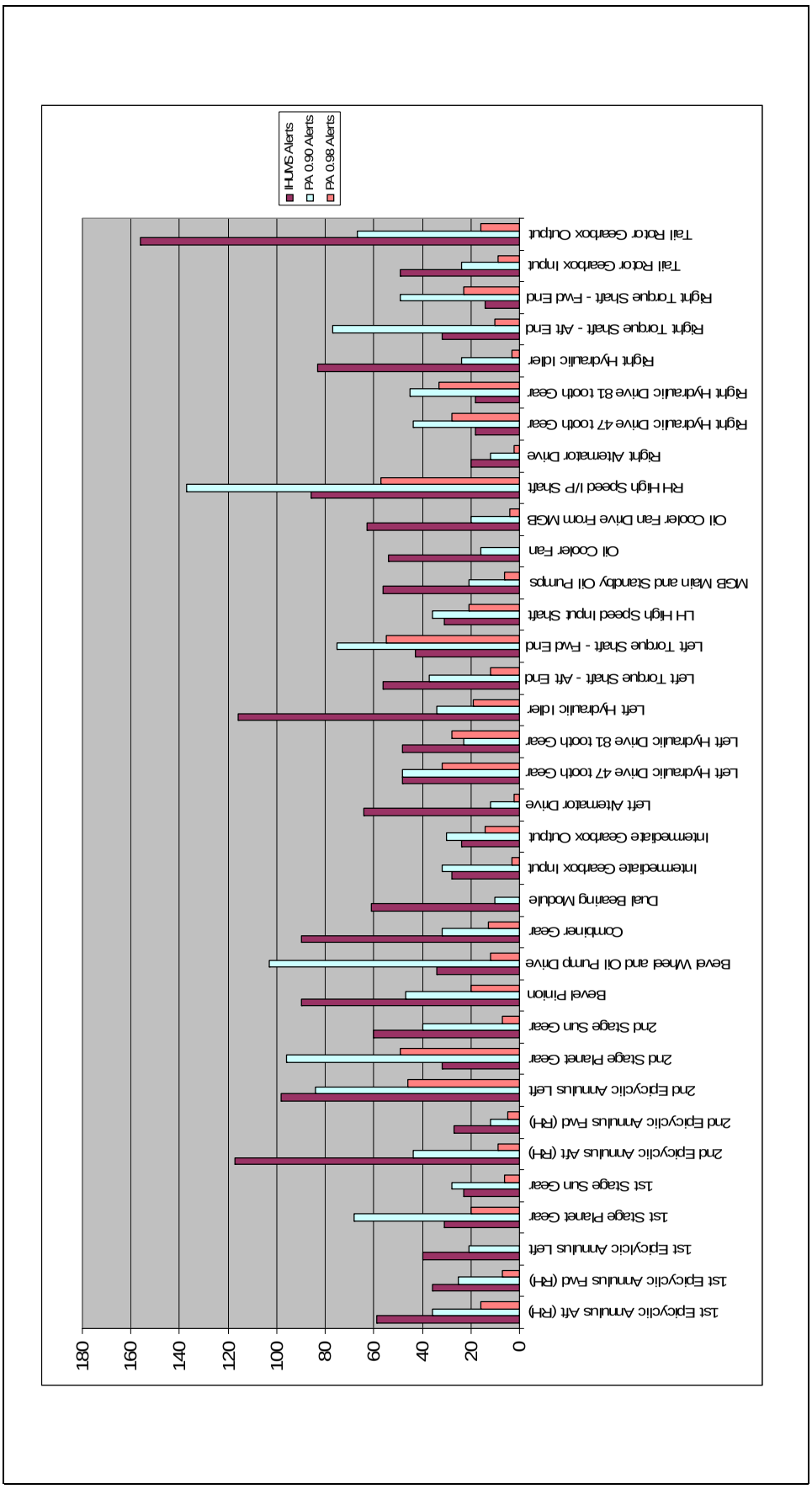

Figure 5-5 Anomaly detection system and IHUMS alerts by component analysis

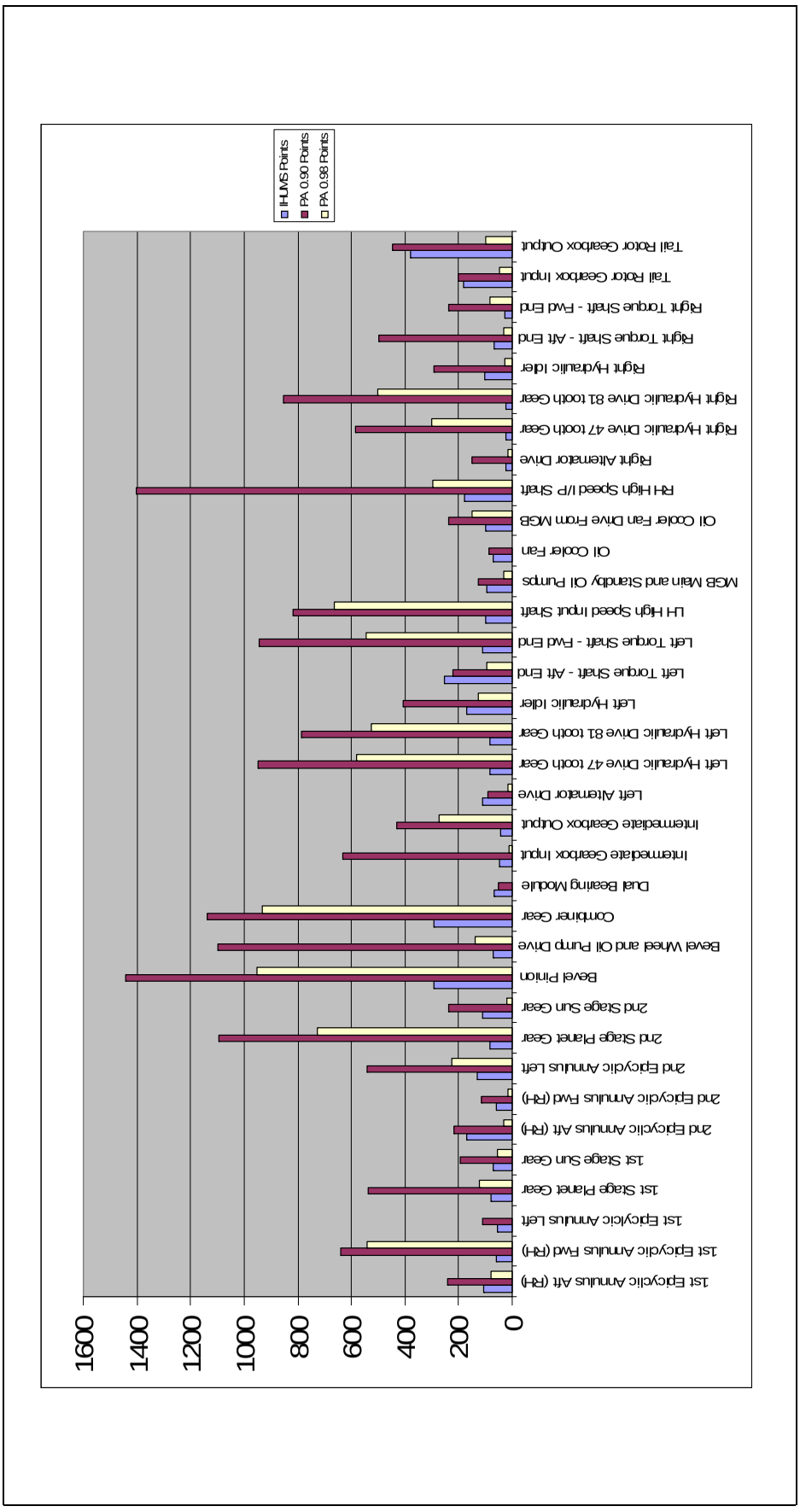

Figure 5-6 Anomaly detection system and IHUMS data points in alert by component analysis

For the anomaly detection system (PA = 0.9), the greatest numbers of data points in alert occurred on the MGB bevel pinion and MGB RH input shaft. G-TIGF was responsible for most of the MGB bevel pinion alert data points, due to the issue described in Example 3 of Section 4.6. Again there is a change in the pattern of the alert data from the first trial period, where there were high numbers of data points in alert on the six 1st and 2nd stage epicyclic gear analyses using sensor 7.

5.4 Anomaly Model Performance

In total, there were eight different types of anomaly model. Four types of anomaly model (the '8-Indicator Absolute' and 'Trend' models, and the 'M6 Absolute' and 'Trend' models) were applied to 33 of the 35 drive train components. The remaining two components (the oil cooler fan and dual bearing module) were monitored by '5-Indicator Absolute' and 'Trend' models. Three of the 33 components (the MGB input shafts and TGB output) also had 'Shaft Order Absolute' and 'Trend' models applied. An anomaly detection system alert was triggered whenever one or more of the models went into an alert state. Counting the individual model outputs separately, these triggered a total of 2,195 alerts in the second trial period, which relates to the 1,509 system alerts. The difference of 686 is a result of multiple models going into alert at the same time.

Figure 5-7 shows the number of alerts triggered by each of the eight different types of anomaly model. The figure also indicates the numbers of alerts that would have been triggered with a PA threshold of 0.98. The 'M6 Absolute' models generated the greatest number of alerts, followed by the '8-Indicator Absolute' models. This finding is somewhat different that from the first six-month trial period, where the '8-Indicator Absolute' and 'Trend' models generated the most alerts (reference [3]). Although the anomaly models were rebuilt between the two trial periods using an enhanced modelling technique, the difference is believed to be primarily due to the changing issues seen on the aircraft in the two periods of operation.

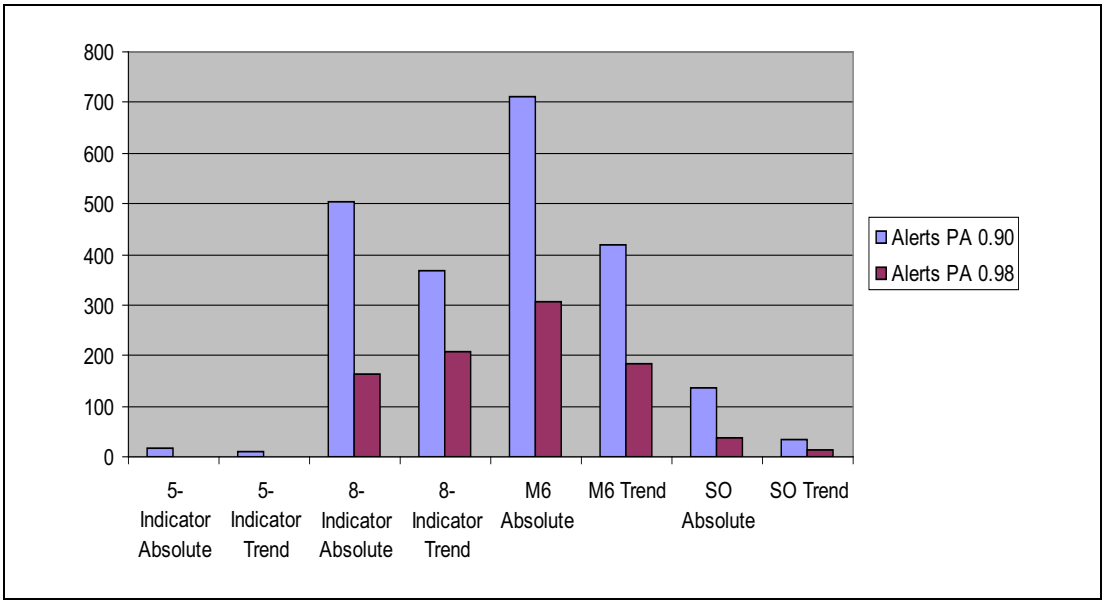

Figure 5-7 Anomaly alerts by model type

Figure 5-8 shows the number of alerts triggered by each of the different types of anomaly model normalised by the number of component analyses in which the model type is utilised. The 'Shaft Order Absolute' models are now responsible for the greatest number of alerts.

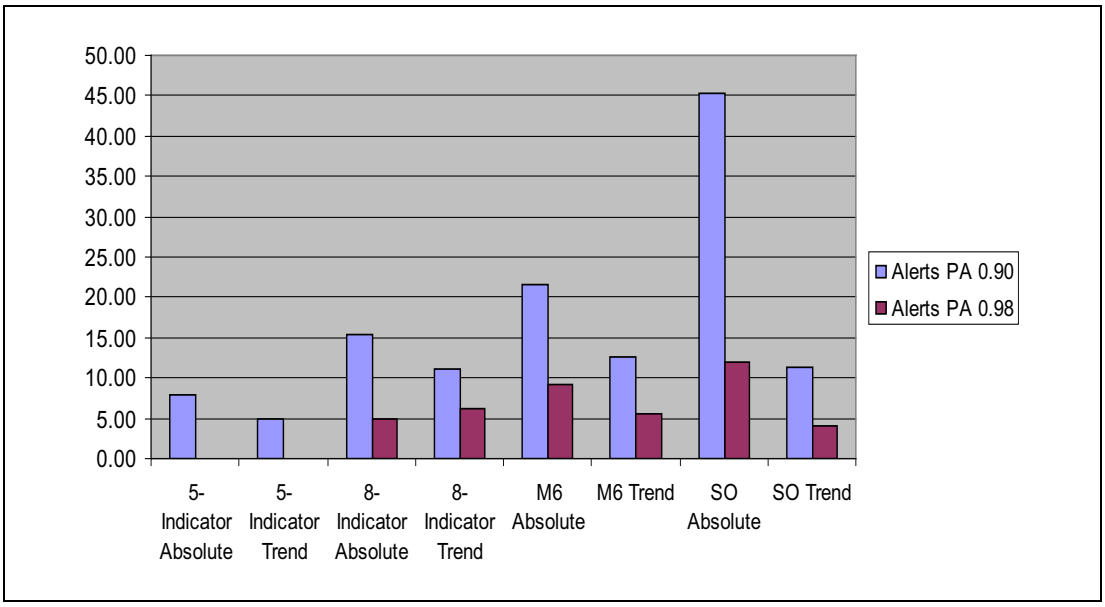

Figure 5-8 Anomaly alerts by model type (normalised)

The number of data points in alert by model type are shown in Figure 5-9 and Figure 5-10, normalised by the number of analyses in the case of the latter. The numbers of data points that would have been generated with a PA threshold of 0.98 are also indicated. The picture is similar to that for the number of alerts, but with the absolute models being more prominent. This is probably due to the fact that the system identified a significant amount of anomalous IHUMS data due to instrumentation problems.

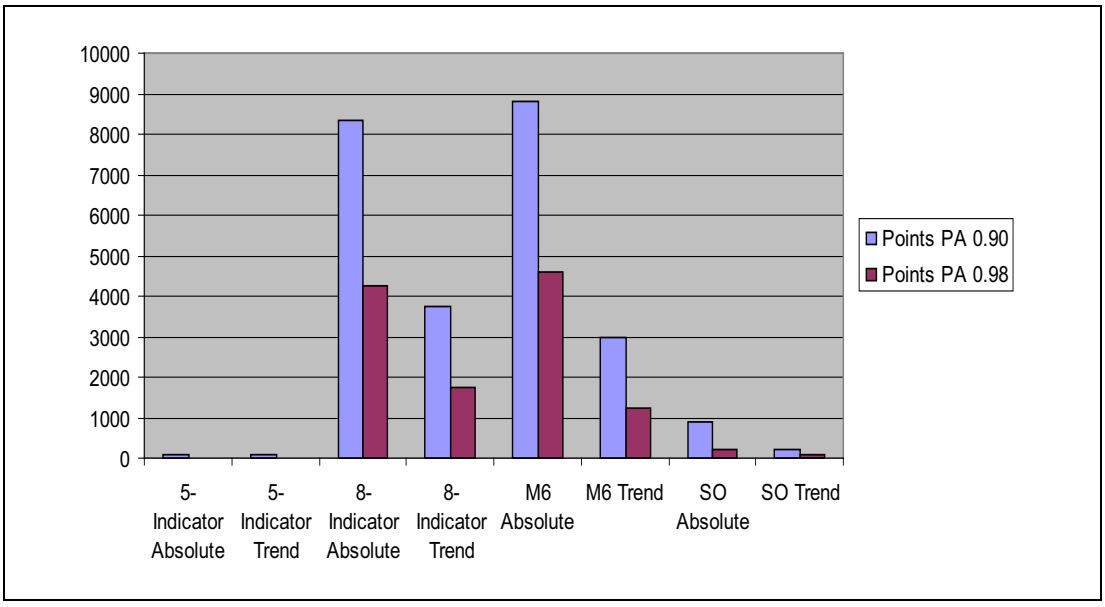

Figure 5-9 Anomaly data points in alert by model type

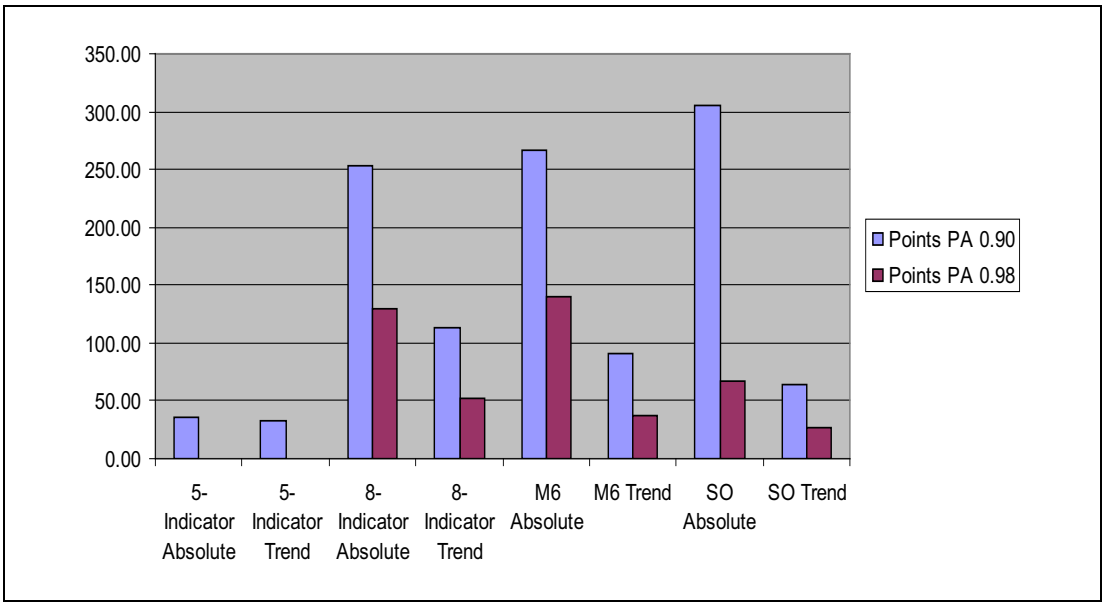

Figure 5-10 Anomaly data points in alert by model type (normalised)

To provide further information on the performance of the individual models, Figure 5-11 to Figure 5-19 present breakdowns of alerts and data points in alert by component analysis for the following six model types; '8-Indicator Absolute' and 'Trend', 'M6 Absolute' and 'Trend', and 'SO Absolute' and 'Trend'.

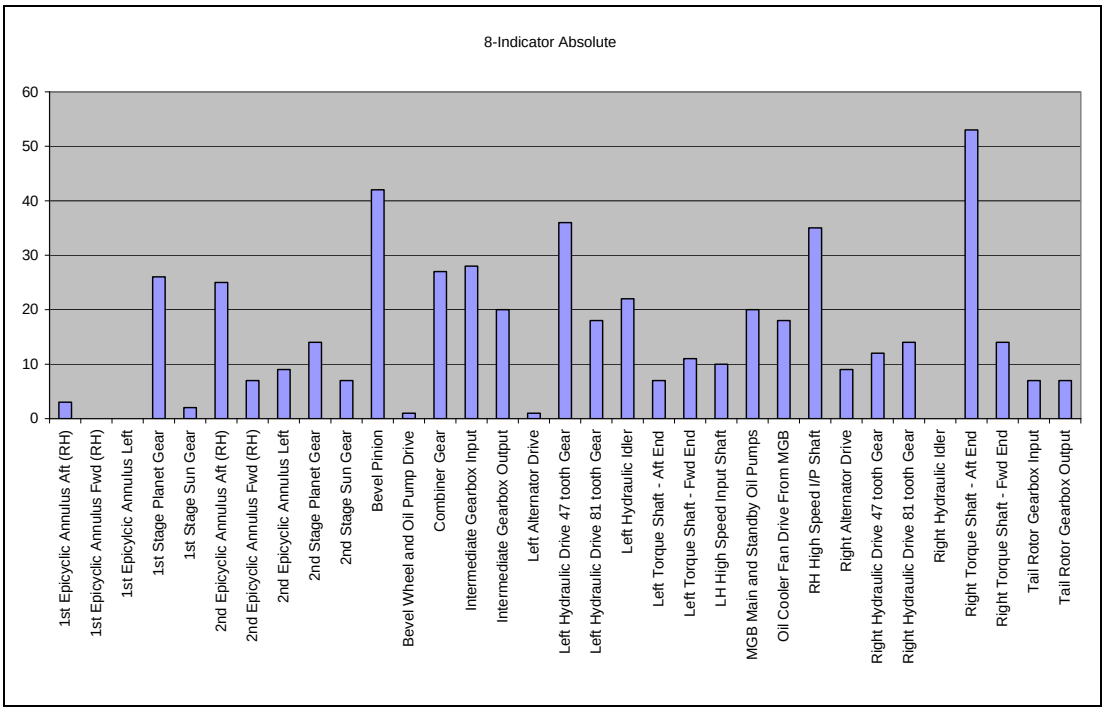

Figure 5-11 '8-Indicator absolute' model alerts by component analysis

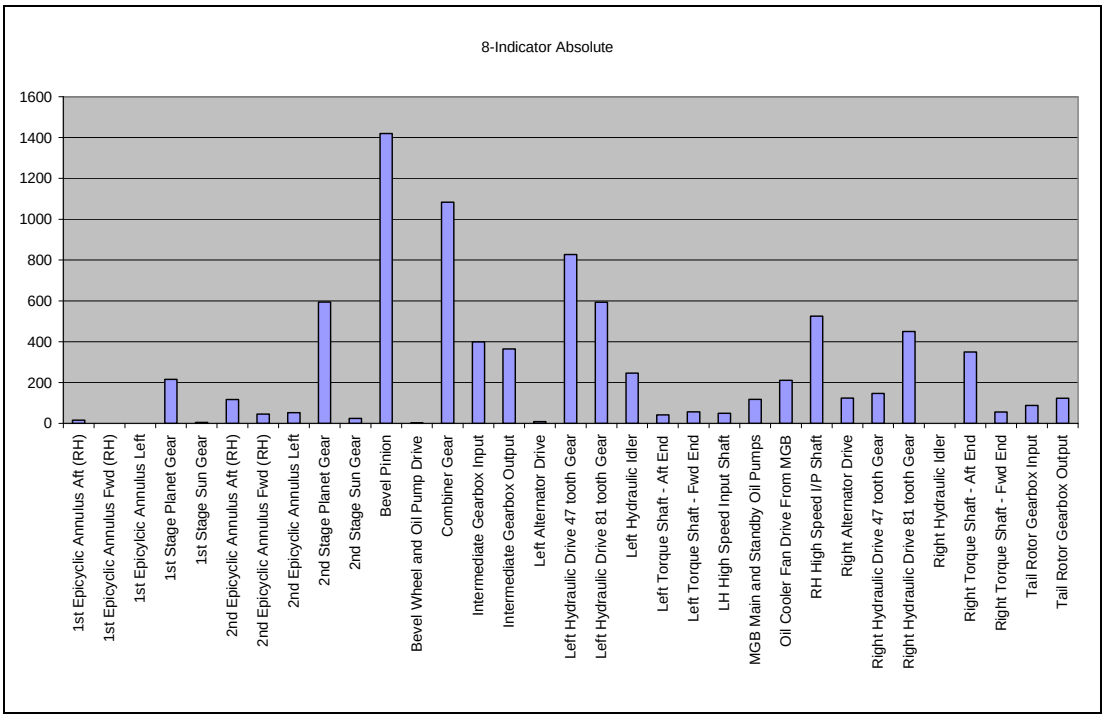

Figure 5-12 '8-Indicator absolute' model data points in alert by component analysis

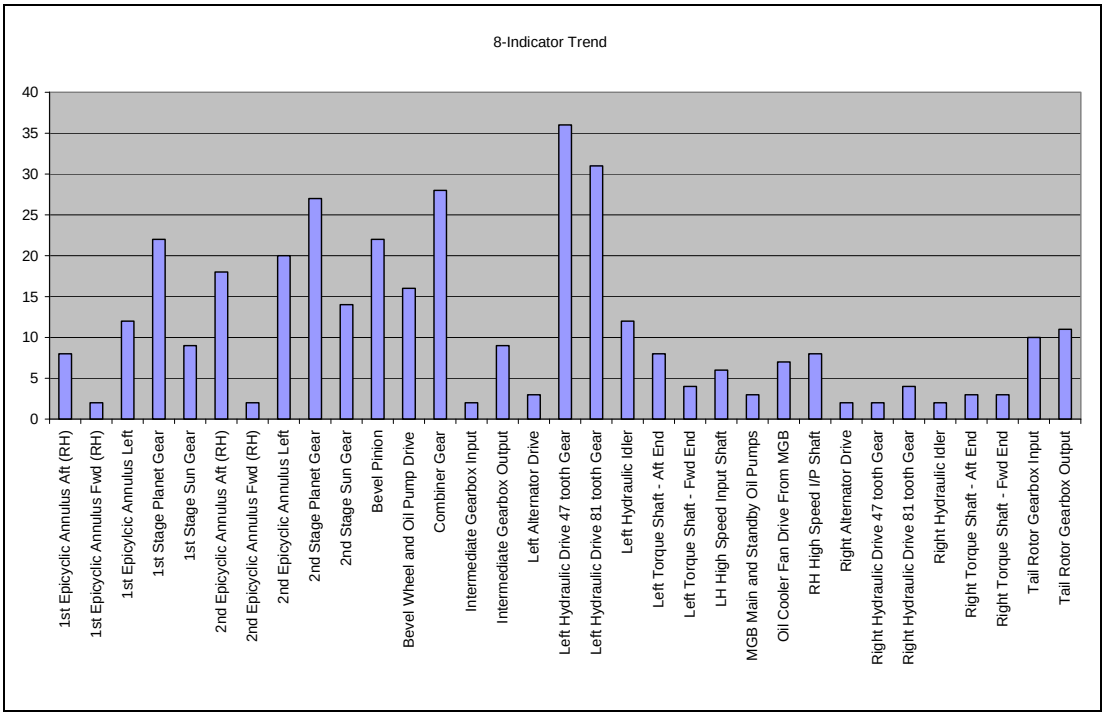

Figure 5-13 '8-Indicator trend' model alerts by component analysis

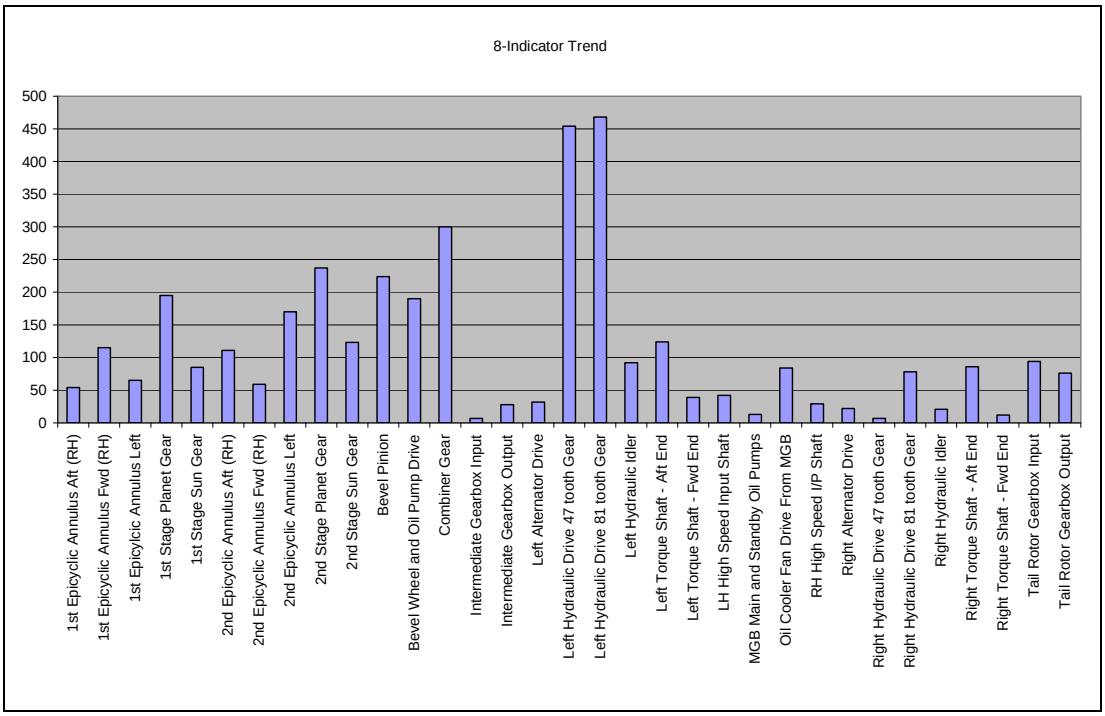

Figure 5-14 '8-Indicator trend' model data points in alert by component analysis

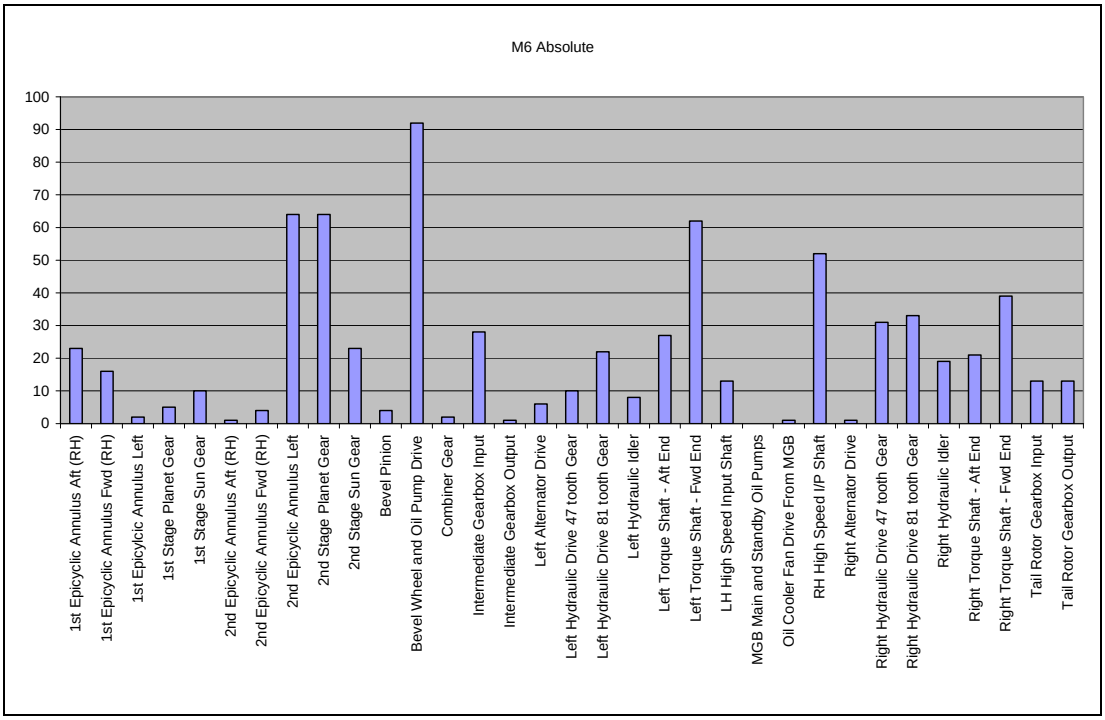

Figure 5-15 'M6 absolute' model alerts by component analysis

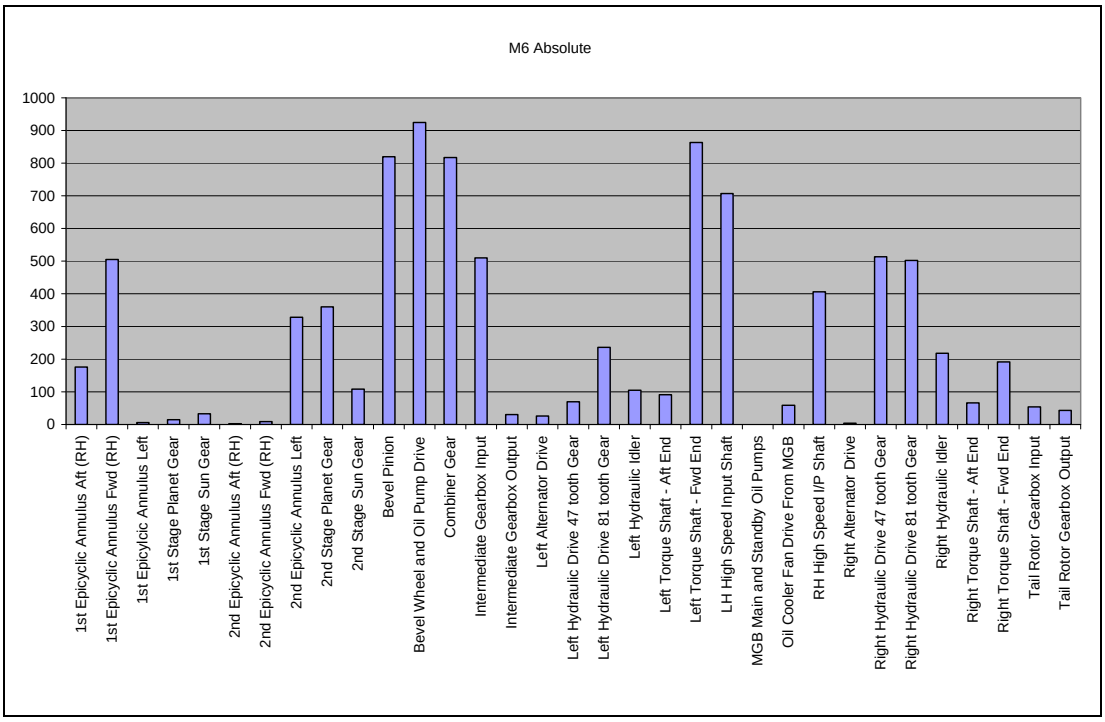

Figure 5-16 'M6 absolute' model data points in alert by component analysis

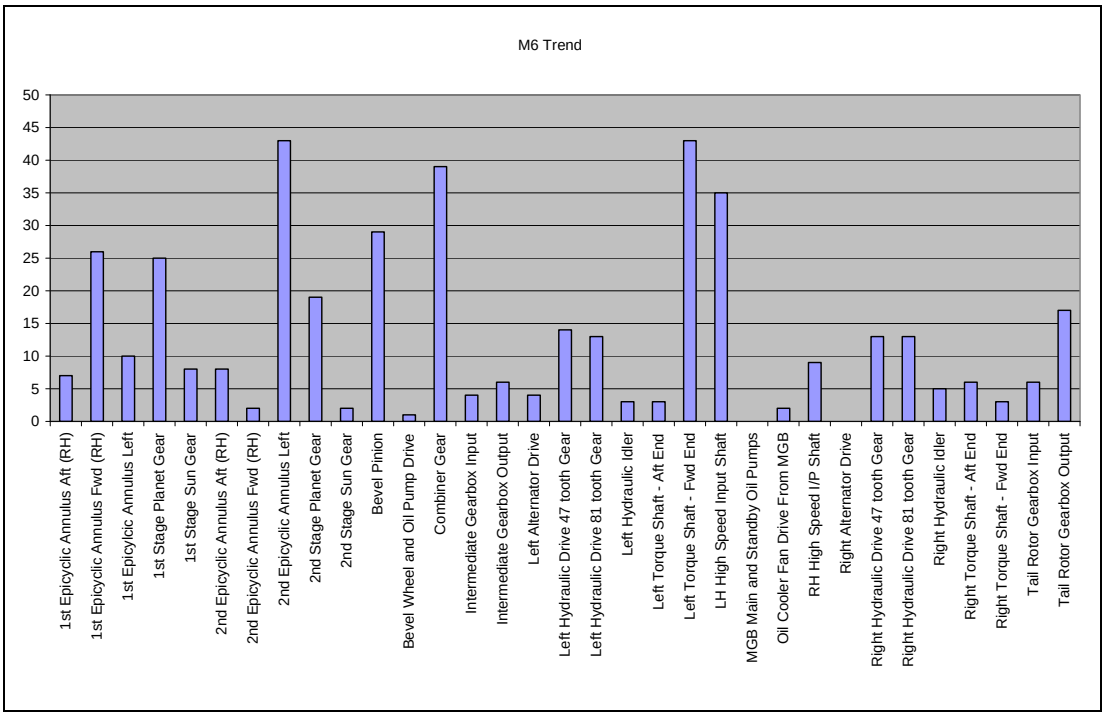

Figure 5-17 'M6 trend' model alerts by component analysis

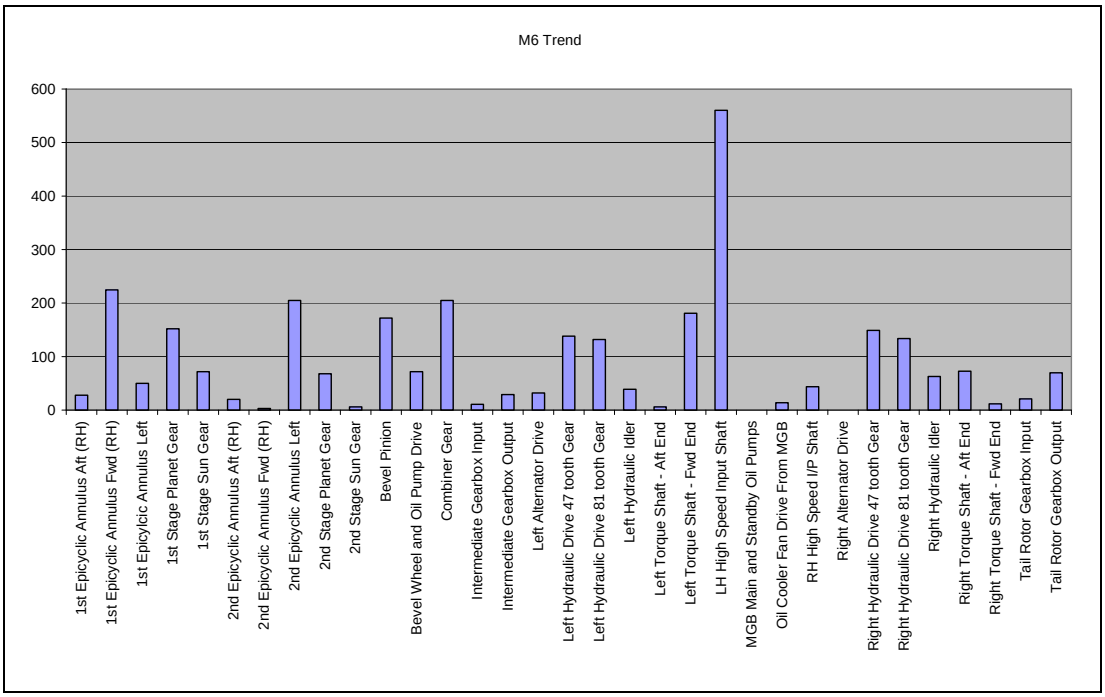

Figure 5-18 'M6 trend' model data points in alert by component analysis

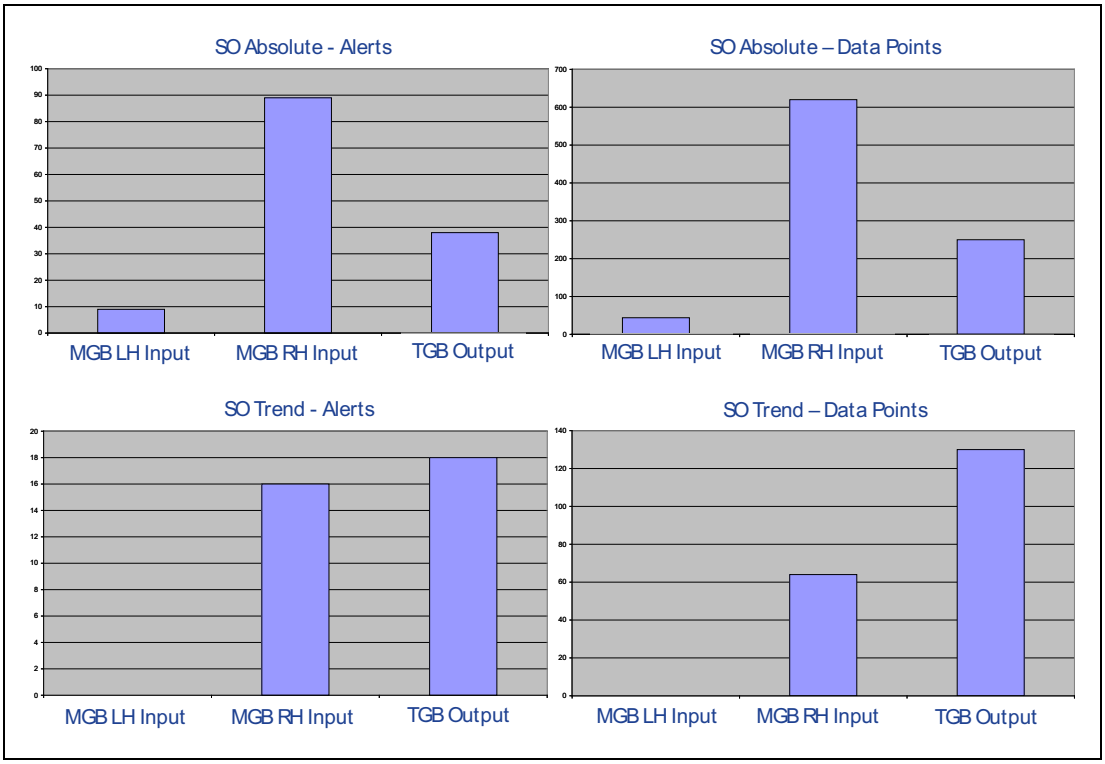

Figure 5-19 'SO' model alerts and data points in alert by component analysis

High numbers of alerts and data points in alert for a particular analysis are typically driven by specific aircraft issues, a number of which have already been covered in Section 4 and the earlier parts of Section 5. One general observation is that there are differences between the 'Absolute' and 'Trend' models in the relative numbers of alerts and data points in alert that are produced on the different component analyses. This indicates that the two types of model are providing complementary anomaly detection capabilities, and that both types are contributing to the overall effectiveness of the system.

6 Conclusions and Recommendations

6.1 Conclusions

The second trial period has provided further evidence to that contained in reference [3] demonstrating the effectiveness of the advanced HUMS anomaly detection system. The system has clearly highlighted anomalous HUMS Vibration Health Monitoring (VHM) Condition Indicator (CI) trends that are more difficult to detect by traditional HUMS analysis. The two trial periods have proven the system's ability to improve the performance of HUMS.

The web-based system implemented for the trial has been demonstrated to be a reliable and effective solution, and is a proven model for the wider implementation of the advanced HUMS anomaly detection capability. Supplementing the system with an analysis service supplying daily fleet health status reports was a valuable help in minimising the operator's workload associated with the new capability.

The new Probability of Anomaly (PA) and Influence Factor (IF) outputs were very useful enhancements. These provided clear information to assist the operator in identifying abnormal HUMS CI trends and diagnosing their possible causes.

The advanced HUMS anomaly detection system continued to out-perform the traditional HUMS analysis in successfully highlighting both aircraft problems and HUMS instrumentation faults. The alert rate was lower than the 'raw' alert generation rate of the traditional HUMS analysis, and had reduced slightly from the first trial period as a result of the data remodelling and new threshold setting process. However it is now considered that the PA threshold of 0.90 is too sensitive, and a PA threshold of 0.98 has been successfully trialled on another programme. Increasing the PA threshold to 0.98 would result in a substantial reduction in the anomaly detection system alert rate, whilst still detecting 14 of the 15 example cases reviewed in the report.

Although anomalous HUMS CI trends can now be reliably and effectively highlighted, the example cases presented in this report provide further evidence of the difficulties involved in interpreting this data and determining appropriate actions. A number of cases have been presented for which the causes of the abnormal vibration characteristics could not be determined, and this is typical of HUMS VHM experience in general.

At the end of the trial period, none of Bristow's requests for strip reports on components rejected for anomalous HUMS CI trends or metal contamination had resulted in any feedback being received on component condition from the Repair and Overhaul process. The lack of feedback information on the condition of drive train components removed from aircraft, such as from gearbox strip reports, often prevents any meaningful interpretation of CI trends in terms of component condition. This type of feedback is critical to the on-going development of HUMS data analysis capabilities, and the absence of such information is seriously hindering this development.

6.2 Recommendations

A plan should be developed for implementation of the advanced HUMS anomaly detection capability on all helicopter fleets involved in UK Continental Shelf offshore support activities. While it is recognised that there may be a few aircraft type and HUMS combinations for which the numbers are too small to make implementation feasible, the plan should cover as many aircraft fleets as possible.

Some guidance documentation should be developed to support this fleetwide implementation. The documentation should include a definition of advanced HUMS anomaly detection, and a specification of the requirements that must be met to achieve this. It should also address certification and security issues such as ensuring effective control of anomaly model development and updating, protecting sensitive data, and the validation of performance through a Controlled Service Introduction.

For the further development of HUMS anomaly detection and VHM capabilities in general, there should be an effective mechanism to enable operators to request feedback on the condition of rejected components. To facilitate this, helicopter OEMS and overhaul agencies should have a reliable process for recording component condition information when gearboxes are stripped for overhaul.

Building on the proven advanced HUMS anomaly detection capabilities, research and development effort should now be concentrated on data mining and automated diagnostic reasoning. These analytical tools can minimise the operator workload associated with HUMS, and provide further assistance in the difficult task of interpreting HUMS VHM data to determine appropriate responses to abnormal behaviour.

7 References

1 Harrison, N., Baines, N. C. (1999). *Intelligent Management of HUMS Data: The use of Artificial Intelligence Techniques to Detect Main Rotor Gearbox Faults*. Study II, CAA Paper 99006.

2 Smiths Aerospace report REP1697(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Interim Report on Phase 1 of the Research Project*. May 2007.

3 Smiths Aerospace report REP1712(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Phase 2 of the Research Project: Six-Month Operational Trial*. July 2007.

4 GE Aviation report REP1731(3): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Additional Research Work (Tasks 1 to 4)*. July 2009.

ANNEX E

Report on Additional Research Work (Tasks 5 and 6)

Based on a report prepared for the CAA by GE Aviation Systems Limited, UK

Table of Contents

Report

INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK

Glossary

8IA/8IT	8-Indicator Absolute / 8-Indicator Trend anomaly model
AGB	Accessory Gearbox
CAA	Civil Aviation Authority (UK)
CBR	Case Based Reasoning
CI	Condition Indicator
DAPU	Data Acquisition and Processing Unit
FS	Fitness Score
HUMS	Health and Usage Monitoring System
IF	Influence Factor
IGB	Intermediate Gearbox
IHUMS	Integrated HUMS (Meggitt Avionics Ltd)
LHA	Left Hand Accessory module
M6A/M6T	M6 Absolute / M6 Trend anomaly model
MGB	Main rotor Gearbox
PA	Probability of Anomaly
ProDAPS	Probabilistic Diagnostic and Prognostic Solutions
Radio button	A control where the selection of one option results in other options being deselected
RHA	Right Hand Accessory module
SOA/SOT	Shaft Order Absolute / Shaft Order Trend anomaly model
TGB	Tail rotor Gearbox
VHM	Vibration Health Monitoring

INTENTIONALLY LEFT BLANK

Executive Summary

This report documents the results of two final research tasks, involving data mining and automated reasoning technology demonstrations, conducted as part of a CAA research programme to demonstrate the intelligent analysis of helicopter HUMS Vibration Health Monitoring (VHM) data. The demonstrations were performed on the outputs of an advanced anomaly detection system, developed earlier in the programme. The goal of the CAA research programme is to improve the fault detection performance of HUMS.

The advanced HUMS anomaly detection capability, and web-based system to implement it, were developed in Phase 1 of the CAA research programme. In Phase 2 of the programme, the anomaly detection system was subjected to a six-month in-service trial by Bristow Helicopters. The results of Phases 1 and 2 are reported in references [2] and [3]. Building on the success of the initial trial, four additional research tasks were carried out to research and develop further advanced HUMS data analysis capabilities. These included a data re-modelling task, and the introduction of new Probability of Anomaly (PA) and Influence Factor (IF) outputs. The results of these additional research tasks are presented in reference [4]. A second in-service trial was then conducted, evaluating the enhanced advanced anomaly detection system incorporating the updated anomaly models and the new PA and IF outputs. The results of the second trial are presented in reference [5].

Having demonstrated the HUMS performance improvements that can be achieved through the application of advanced anomaly detection, the greatest scope for further improvement in the combined performance of the system and operator is through the development of a second level of advanced data analysis, based on data mining and automated reasoning technologies. This report presents the results of two final additional research tasks, demonstrating new system concepts and capabilities for this second level of advanced data analysis, and the data mining and reasoning technologies that could provide them.

The data mining demonstration extracted new information from the anomaly model output data for post-processing of anomaly alerts and use in automated diagnostic reasoning. It focused on automated change detection, trend analysis, anomaly significance assessment, and data analysis to support case-based reasoning. The reasoning demonstration focussed on data fusion and fault diagnosis. The data fusion utilised physical relationships between HUMS sensors and component analyses, generated diagnostic summary information, and supported the anomaly significance assessment. The fault diagnosis element demonstrated an ability to automatically identify explainable patterns due to maintenance actions and instrumentation faults, and explored the application of Case Based Reasoning. Mock-ups of possible new anomaly detection system alert displays were created to illustrate how the resulting new information could be presented to the operator.

The technology demonstrations have clearly shown how secondary analysis of the current anomaly detection system outputs can provide further benefits to operators, reducing workload, automatically highlighting significant anomaly alerts, and providing new information to support decision making.

INTENTIONALLY LEFT BLANK

Report

1 Introduction

Health and Usage Monitoring Systems (HUMS), incorporating comprehensive drive train Vibration Health Monitoring (VHM), have contributed significantly to improving the safety of rotorcraft operations. However, experience has also shown that, while HUMS has a good success rate in detecting defects, not all defect related trends or changes in HUMS data are adequately detected using current threshold setting methods. Earlier research (see reference [1]), conducted as part of the CAA's helicopter main rotor gearbox seeded defect test programme, demonstrated the potential for improving fault detection performance by applying unsupervised machine learning techniques such as clustering to seeded fault test data. The CAA therefore commissioned a further programme of work titled "Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service" (CAA Contract No. 841). GE Aviation was contracted to carry out this programme in partnership with Bristow Helicopters, analysing IHUMS data from Bristow's North Sea AS332L fleet.

The work under Contract No. 841 was structured as two phases. Phase I involved the research and development of an advanced anomaly detection capability to enhance the fault detection performance of HUMS. Anomaly models were built for all the AS332L rotor drive system components. These fused multiple input HUMS Condition Indicators (CIs) into a single output 'Fitness Score' (FS), indicating how well new data fit the model's view of normality. Phase 1 also included an 'off-line' demonstration of the advanced anomaly detection capability on a database of historical AS332L data, and the development of a system implementing this capability for evaluation in an in-service trial. The results of Phase 1 are reported in reference [2].

Phase 2 of the CAA research programme comprised a six-month in-service trial, conducted by Bristow Helicopters, of the advanced HUMS anomaly detection system developed in Phase I. The system was implemented as a web server located at GE Aviation in Southampton, with AS332L IHUMS data being automatically transferred every night from Aberdeen, and Bristow Helicopters having a remote login to the server to check the anomaly detection system outputs. The results of Phase 2 are reported in reference [3].

Building on the success of the initial trial period, four additional research tasks were carried out to research and develop further advanced HUMS data analysis capabilities. First, an investigation was carried out into the use of additional pre- and post-processing techniques to characterise data trends. Although the investigation did not result in the implementation of any new functionality in the anomaly detection system, it did provide a good foundation for the future development of new capabilities for data characterisation. Second, a model tuning and re-modelling task benefited from other work re-implementing and improving the modelling technique which provided an enhanced anomaly modelling facility. All the anomaly models were re-built using this improved facility. Third, a probabilistic alerting policy was implemented, with the FS outputs from the anomaly models being converted into normalised Probability of Anomaly (PA) values with a scale of 0-1. On this scale, a value of 0 represents normal data while a value of 1 indicates a clear anomaly. This simplified threshold setting, with a global alerting threshold being set at a selected PA value. Finally, Influence Factors (IFs) were implemented to provide diagnostic information on the causes of an anomaly alert. The IFs are derived from an anomaly

model, with one IF for each HUMS CI used to train the model, and provide an indication of how much each CI contributes to an anomaly alert. Again, they are normalised and so can be directly compared with each other.

The results of these four additional research tasks (Tasks 1 to 4) are presented in reference [4]. The new PA and IF data was evaluated in a second trial period (reference [5]). This commenced at the beginning of January 2008, with the formal six-month trial period completing on 30 June, however the trial was continued informally until 19 December 2008 while the research work described in this report was completed.

Tasks 5 and 6 covered the following two topics:

1 Data mining of anomalous trend features: The IFs generated from the anomaly models can point to the HUMS CIs that are driving anomaly alerts. The objective of data mining is to identify trend patterns in large quantities of data that would be impractical or impossible to detect by manual inspection. It would be very informative to mine the FS and IF data features to identify trends, automatically assess the significance of anomalies, test established HUMS diagnostic knowledge and further develop this knowledge. The discovered knowledge can then be utilised in the automated reasoning process described in the second work package.

2 Automated reasoning: The next major step improvement in the effectiveness and usability of HUMS is expected to be achieved through the application of automated reasoning to the output from the anomaly detection process. When an anomaly is flagged a manual assessment is currently made to answer a number of questions such as:

   - How significant is the trend?
   - Which CIs are driving this?
   - Does the sensor's data look reliable?

   Significant operator benefits can be obtained by feeding anomaly detection outputs into probabilistic reasoning networks to automatically answer some of these questions. The operator would then be provided with a clear and concise summary of key information that can be deduced from the anomaly model outputs. Using the IFs, reasoning can be performed on the indicators driving a trend to provide more detail on the significance of anomalies, and also derive some fault diagnostic information. Case Based Reasoning can also be performed to search for any similar previous cases.

Because of the relative immaturity of the application of automated data mining and reasoning to health monitoring, and the magnitude of the task to research, develop, test and implement new automated data mining and reasoning functions, the outputs of Tasks 5 and 6 were concept demonstrations only. The dual aims of the work were to:

- Research various approaches to the application of data mining and reasoning tools to Vibration Health Monitoring (VHM) data.
- Show how, by applying an additional layer of automated processing to the current anomaly detection system outputs, valuable new information could be provided to enable an operator to take the most appropriate action in response to anomaly alerts.

Section 2 of this report provides some background information by giving a brief overview of the HUMS anomaly detection system on which the work has been

conducted. Section 3 presents concepts for new types of information that could be provided to operators as a result of the new technology described in the following sections. Section 4 presents the results of the data mining demonstration, while Section 5 describes the automated reasoning demonstration. Finally, some conclusions and recommendations are presented in Section 6.

1.1 Database of Anomaly Model Outputs Used for the Demonstrations

During the second six-month in-service trial period GE Aviation introduced a further refinement to the anomaly modelling technique, involving a minor change to the cluster removal method in the model adaptation process. Testing of the updated approach using synthetic data showed that this could give a small improvement in performance.

As the in-service trial was already in progress, the trial system was not updated with new models created using the refined technique. However a copy of the database of IHUMS CIs was taken at the end of the formal trial period on 30 June. New models were created using this data, and the PA, FS and IF predictions were recalculated for all of the historical data. This parallel off-line database was used for the data mining and reasoning demonstrations described in this report.

A comparison of the anomaly model outputs showed that there were no significant differences between the results obtained using the trial system and those using the modified system held in the parallel off-line database. Therefore the plots of anomaly model FS and IHUMS CI data included in this report are taken from the website of the in-service trial system, and show exactly what the operator would currently see in that system.

2 Background Information on the HUMS Anomaly Detection System

2.1 IHUMS Data Analysed

The anomaly detection system analysed IHUMS VHM data from the following assemblies in the AS332L drive train: Main rotor Gearbox (MGB), left and right Accessory Gearboxes (AGBs) and oil cooler fan, Intermediate Gearbox (IGB), and the Tail rotor Gearbox (TGB). Thirty-five drive train component analyses were defined, as listed in Table 2-1. The table also shows the aircraft sensor used to acquire the component data, and the equivalent IHUMS 'channel' (i.e. analysis number) allocation. The IHUMS performs thirty-two component analyses. However, three of the components have two gears on a single shaft (the MGB combiner gear and bevel pinion, and the left and right AGB hydraulic drive 47- and 81-tooth gears). For the purposes of anomaly modelling, each of these was split into separate components, with separate anomaly models built for each of the two gears on the shaft using the different gear mesh-specific CIs.

Table 2-1 AS332L drive train components analysed

Sensor	Channel	Shaft/Gear	Assembly
1	0	LH high speed input shaft	MGB
2	1	RH high speed input shaft	MGB
1	2	Left torque shaft - fwd end	MGB
2	3	Right torque shaft - fwd end	MGB
3	4	Left torque shaft - aft end	MGB
4	5	Right torque shaft - aft end	MGB
3	6	Combiner gear	MGB
3	6	Bevel pinion	MGB
4	7	Bevel wheel and oil pump drive	MGB
7	8	1st stage sun gear	MGB
7	9	1st Stage planet gear	MGB
5	10	1st epicyclic annulus fwd (RH)	MGB
6	11	1st epicyclic annulus left	MGB
7	12	1st epicyclic annulus aft (RH)	MGB
7	13	2nd stage sun gear	MGB
7	14	2nd stage planet gear	MGB
5	15	2nd epicyclic annulus fwd (RH)	MGB
6	16	2nd epicyclic annulus left	MGB
7	17	2nd epicyclic annulus aft (RH)	MGB
12	18	Intermediate gearbox input	IGB
12	19	Intermediate gearbox output	IGB
11	20	Tail rotor gearbox input	TGB
11	21	Tail rotor gearbox output	TGB
3	0	Left alternator drive	AGB

Table 2-1 AS332L drive train components analysed (Continued)

3	1	Left hydraulic idler	AGB
3	2	Left hydraulic drive 47-tooth gear	AGB
3	2	Left hydraulic drive 81-tooth gear	AGB
4	3	Right alternator drive	AGB
4	4	Right hydraulic idler	AGB
4	5	Right hydraulic drive 47-tooth gear	AGB
4	5	Right hydraulic drive 81-tooth gear	AGB
3	6	Oil cooler fan drive from MGB	AGB
3	7	MGB main and standby oil pumps	MGB
9	8	Oil cooler fan	AGB
8	9	Dual-bearing module	MGB

The CIs calculated by the IHUMS during each component analysis are listed in Table 2-2 (a more detailed description is presented in reference [2]). Not all of the indicators were used in the anomaly modelling. A subset of indicators was chosen on the basis that they would include all of the key diagnostic information contained in the data (the actual indicators used in the different anomaly models are listed in Section 2.2.2).

Table 2-2 IHUMS Condition Indicators

Condition Indicator	**Description**
SIG_MN	Signal mean (DC offset)
SIG_PK	Signal peak
SIG_PP	Signal peak-peak
SIG_SD	Signal standard deviation (rms)
FSA_SO1	Fundamental shaft order
FSA_SON	Selected shaft order
FSA_SE1	Shaft eccentricity/imbalance
FSA_MS_1	First mesh magnitude
FSA_MS_2	Second mesh magnitude
FSA_GE11	First gear narrowband mod.
FSA_GE12	Second gear narrowband mod.
FSA_GE21	First gear wideband mod.
FSA_GE22	Second gear wideband mod.
ESA_PP	Enhanced peak-peak
ESA_SD	Enhanced standard dev (rms)
ESA_M6*	Enhanced impulsiveness indicator
SIG_AFH	Airframe (flying) time
SIG_HIS	Synchronization histogram (CG)
SA_CVG	Signal average convergence

2.2 Data Processing for Anomaly Detection

A detailed description of the data processing performed is presented in references [2] and [4]. A brief summary of the three stages of the anomaly detection processing is given below.

2.2.1 Data Pre-Processing

Different pre-processing was applied to the input IHUMS CI data to enable two types of model to be built for each monitored component; an 'absolute' and a 'trend' model. The absolute models identify combined CI values that are anomalous in absolute terms, whereas the trend models identify anomalous combined CI trends, irrespective of the absolute values of the indicators.

The pre-processing for the absolute models was minimal, and involved the application of a two stage filter to each indicator: the first stage removed unreasonable values and the second stage applied a median filter to remove up to two successive spikes in the time series data.

For the trend models, the two stage filter for the absolute models was used, and then a 'moving median difference' algorithm was applied. Following each new acquisition, the median of the time history is re-calculated and subtracted from the newly acquired value to provide a normalised value. This technique reduced the impact of early post-installation trends, and also small step changes due to maintenance, because the normalised value would gradually recover back to the median base line level. Although there was some distortion of the time-history, this simple approach worked well. Its weakness was that, while it achieved the normalisation objective, it acted as a difference calculator and could not distinguish between step changes, short duration changes, and developing trends.

2.2.2 Anomaly Models

Using the pre-processed data, anomaly models were constructed for each drive train component. The models were constructed from a training database of historical data, and adapted so that they eliminated the effects of suspected anomalies existing in the training data. The adaptation process is described in reference [4]. No data is actually removed in this process, and the resulting models can make predictions on all data, including that associated with suspected anomalies. Models were built to represent absolute data behaviour (between-component variability) and trend behaviour (within-component variability). These models were sophisticated statistical representations of the data generated from in-service experience, fusing sets of CIs (i.e. vibration features) to reduce a complex data picture into a single time history called an FS trace. The FS measured the degree of abnormality in the input data and mirrored the shape of any significant data trends. It represented a 'goodness of fit' criterion, indicating how well data fitted the model of normality, i.e. the corresponding anomaly model. Therefore the FS exhibited a decreasing trend as the data became increasingly abnormal.

With the exception of two drive train components that did not include a gear on the monitored shaft, every component had anomaly models constructed from two sub-sets of the CIs listed in Table 2.2:

- '8-indicator' models, comprising the following eight indicators: {ESA_PP, ESA_SD, FSA_GE21/22, FSA_MS_1/2, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.
- 'M6' models, comprising the following two indicators: {ESA_M6*, ESA_WEA}, (ESA_WEA is a derived indicator, calculated as the ratio ESA_SD/SIG_SD).

The indicators were split to reduce the potential for one indicator to be masked by another; the criterion for this split was derived from engineering knowledge of the

indicator design and characteristic behaviour. The oil cooler fan and dual-bearing module were monitored by only two models – absolute and trend models comprising the following '5-indicator' set: {ESA_M6*, FSA_SO1, FSA_SON, SIG_PP, SIG_SD}.

For the MGB inputs and TGB output, additional 'Shaft Order' (SO) models were provided. These models comprised the following two indicators: {FSA_SO1, FSA_SON, where N =2}.

Using the different forms of pre-processing, 'absolute' and 'trend' models were built for each set of indicators. Therefore the majority of drive train components were monitored by four anomaly models – '8-Indicator Absolute' and 'Trend' models, and 'M6 Absolute' and 'Trend' models. The exceptions to this were two components (the oil cooler fan and dual-bearing module) monitored by two models, and three components (the MGB input shafts and TGB output) monitored by six models.

2.2.3 **Anomaly Model Outputs**

The FS output from the anomaly models contains useful trend information, however the absolute values vary between models. The anomaly detection output is enhanced if the measure of what is anomalous is normalized across models. Normalization allows model outputs to be compared, and such a measure could be fed into a secondary process such as automated reasoning to assess the nature of an anomaly. In addition, it would facilitate data mining of anomalous patterns in the search for new knowledge. A 'Probability of Anomaly' (PA) measure was therefore introduced, which is a normalized probability measure that ranges between 0 and 1 (see reference [4] for more information). For each model there is a PA distribution which is an extreme value distribution. An FS value is passed to the PA distribution and a PA value is returned. Most FS values will return a PA of 0 because most acquisitions will be normal. The PA values are used to define an alerting threshold, which can be common to all models and components (the FS equivalent of this PA threshold will be different for each model).

'Influence Factors' (IFs) are a type of model prediction that can provide diagnostic information about an anomaly model and its inputs (i.e. HUMS CIs). There are different types of IF, each type producing a different view on a model or input variable. An acquisition will generate a single IF predicted value for each CI used to train a model. IF time histories provide information regarding the influence of HUMS CIs on the fused FS but, unlike CIs, IFs are normalized and can be directly compared. In addition to further enhancing the usability of the HUMS anomaly detection system, like the PA values the IFs will also facilitate developments such as the application of automated diagnostic reasoning and data mining. Although multiple types of IF are generated for GE Aviation's internal use, with each type being designed to provide different information about a model or an input feature, only one type of IF is presented to the operator. This is the one that most closely matches the trends the operator would observe in the HUMS CI data.

2.2.4 **Anomaly Alerts**

An alerting threshold is defined using the PA values, which greatly simplifies the threshold setting process. A default PA threshold is normally applied that is common to all models and components (the FS equivalent of this PA threshold will be different for each model). The default threshold for an individual model can, however, be changed if it is considered that the alert rate is too high or too low. A global PA threshold of 0.90 was used for the trial.

An alert is triggered when M out of the last N points are above the threshold. The default values for M and N are 3 out of 4.

2.3 Anomaly Detection System

The anomaly detection system was implemented as a web-based system, hosted on a server at GE Aviation in Southampton.

The web server has a simple-to-use interface, enabling the user to quickly access anomaly alerts, PA and FS traces, IF and HUMS CI trends, and notes documenting follow-up actions. After clicking on a chart icon in an alert record, the user is presented with the anomaly model PA traces for the component in alert or, using a radio button (see glossary), the display can be switched to FS traces (Figure 2-1). This represents the primary system output display associated with an anomaly alert.

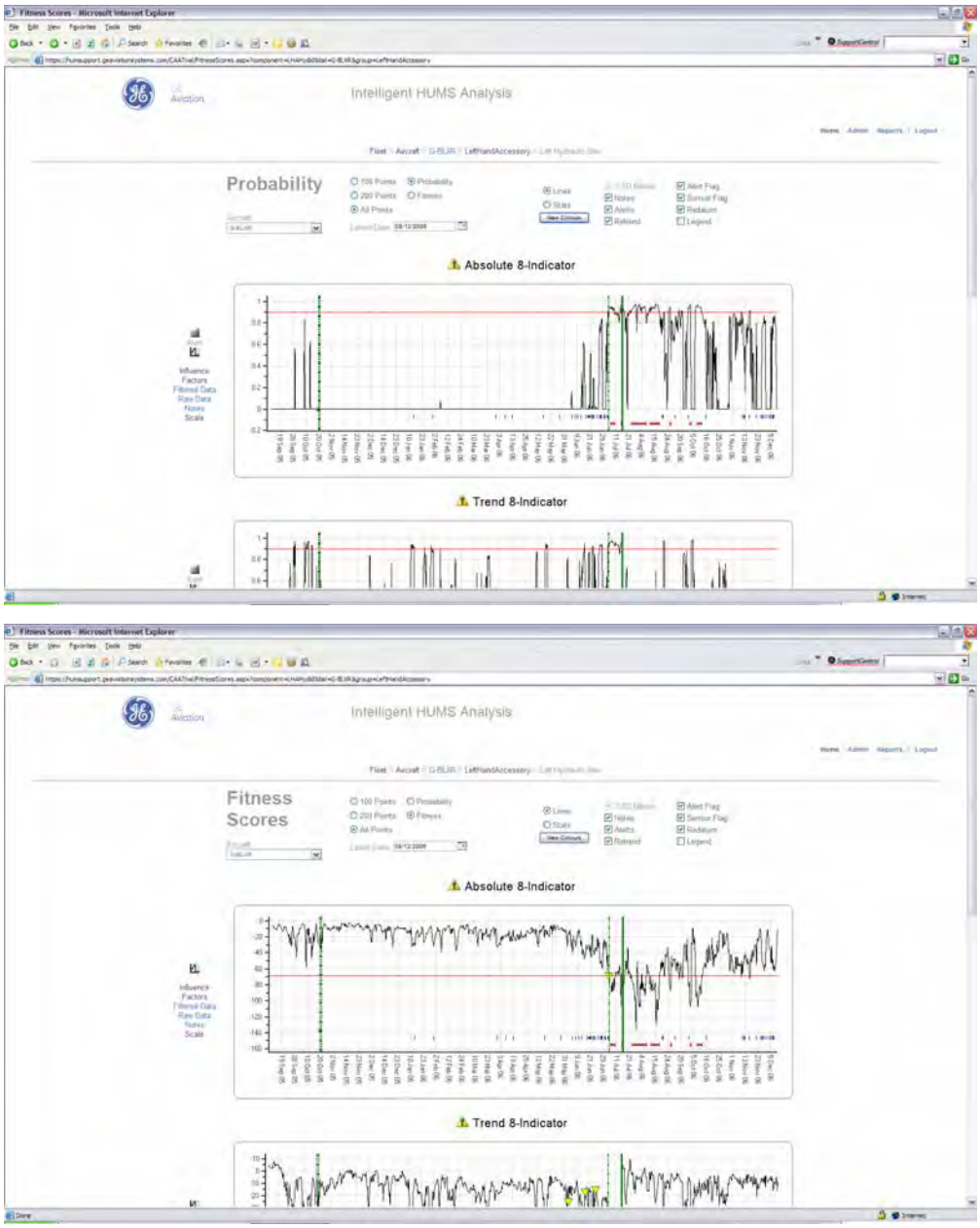

Figure 2-1 PA and FS traces

For investigation, the user can access charts of IF and HUMS CI trends (Figure 2-2). For those CIs used to build models, the indicator charts showed the upper and lower three standard deviation bands that had been derived from the anomaly model.

Figure 2-2 IF and IHUMS CI displays

It is also possible to view IF and HUMS CI data for a fleet of aircraft either as multiple trend plots or as statistical plots (Figure 2-3). Once an investigation and any associated follow-up action had been completed the user can acknowledge an alert and, if appropriate, enter a note.

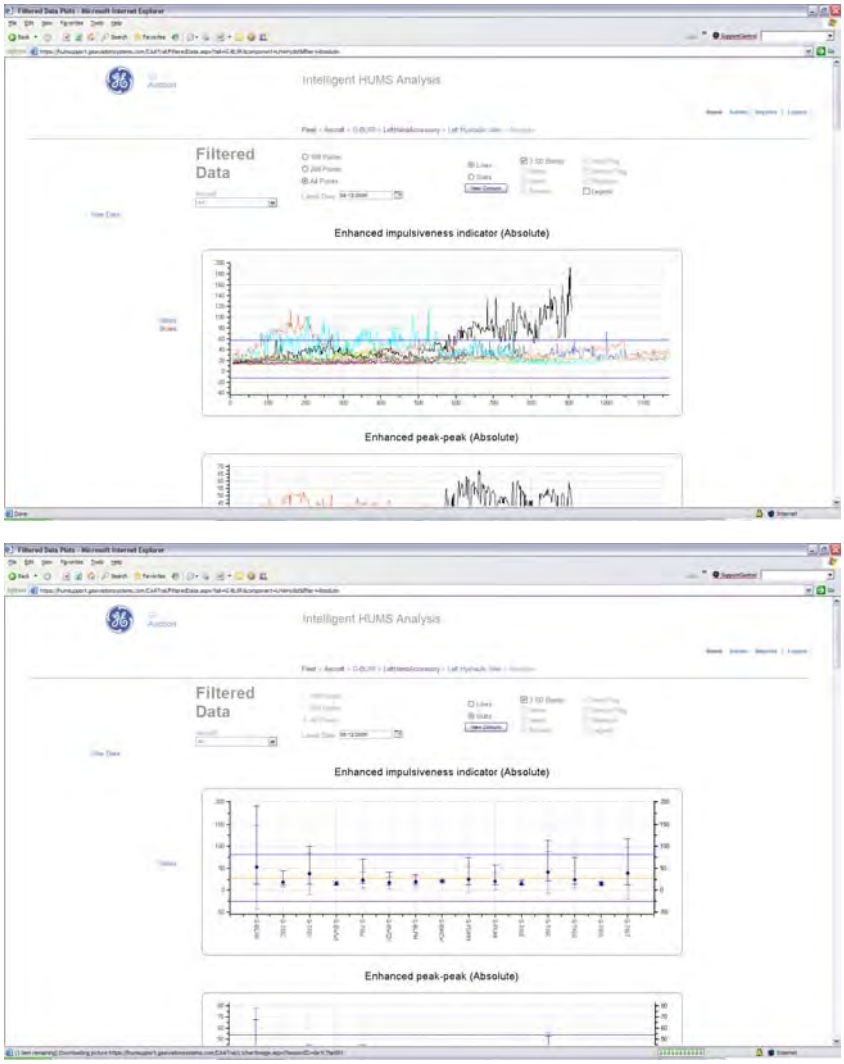

Figure 2-3 Fleet data displays

While the system outputs shown in the above figures contain information on an anomaly, the user must extract this manually by reviewing a series of charts, and then using existing knowledge to interpret these. A key objective of research tasks 5 and 6 was to demonstrate processes that could automatically extract this information and present it to the user in a clearly understandable form.

3 Concepts for New Information Displays

Two key objectives of research tasks 5 and 6 were to:

1 Enhance the effectiveness of the advanced HUMS anomaly detection system by providing new information to aid the operator's data interpretation and maintenance decision making processes.

2 Reduce the operator's workload involved in using the system and responding to anomaly alerts.

Mock-ups of possible new anomaly detection system alert displays were created to illustrate the new capabilities and output information that the data mining and reasoning technologies are aimed at providing, and how these might be presented to the operator. It was not an objective of this work to design new system displays. The mocked up displays were purely intended to help communicate the type of new information that could be generated, and to prompt discussion on how this could be presented. For example, some operators may only want a simple output, however a knowledgeable operator may want a comprehensive display bringing together multiple pieces of information. Using the analogy of a cockpit instrument display, this would look bewildering to a passenger, but a pilot can quickly scan the displays and understand the status of the aircraft. The key point is that the anomaly models can provide a lot of information on alerts, and it is clearly beneficial to extract this and make it available to the user.

Sections 4 and 5 describe examples of the new types of information that were introduced. The output of research tasks 5 and 6 was a technology demonstration, involving the demonstration of new system concepts and capabilities, and the technologies that can provide them. Further work is required to develop these technologies to a readiness level suitable for implementing in an operational system.

Figure 3-1 shows the existing aircraft level alert summary display (similar displays summarise the number of alerts by fleet, assembly and component).

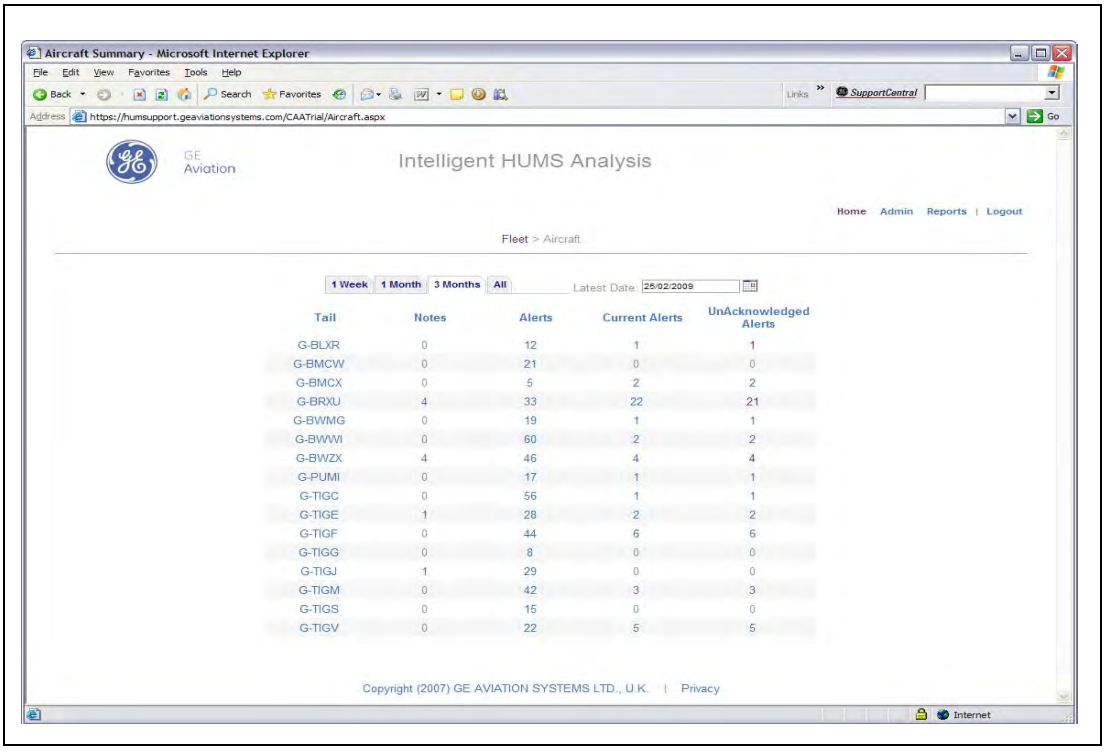

Figure 3-1 Existing alert summary display

A mock-up display illustrating how this could be updated is presented in Figure 3-2. Although new information is provided, the primary new capability being illustrated is 'change detection'. This has the twin objectives of reducing workload by suppressing repeated occurrences of previously acknowledged alerts, while improving monitoring performance by highlighting changes in data for existing acknowledged and unacknowledged alerts.

The change detection function would automatically acknowledge repeat occurrences of alerts that had previously been acknowledged, and that had been re-triggered by only a small change in the data, removing the need for the operator to review these again. At the same time, it would constantly monitor continuing alerts that had been acknowledged by the operator when they first occurred, and flag up any for which the data had changed significantly since the time of acknowledgement. A second workload reduction feature is provided by an automated reasoning capability which suppresses alerts triggered by small step changes (i.e. the data remains within normal bounds) that are probably caused by benign maintenance actions.

Aircraft	Current alerts	Acknowledged alerts	Repeat ack alerts	Change	Unacknowledged alerts	Change	Prob benign maint alert	New alerts
G-BRXU	2	2	1	1	0	0	0	0
G-BWWI	5	0	0	0	4	0	0	1
G-BWZX	5	2	1	0	2	1	1	1

Figure 3-2 Mock-up of possible new alert summary display

The current alerts in Figure 3-2 are the sum of the previously acknowledged and unacknowledged alerts plus the new alerts. If a new alert occurs and the change detection function determines that it is a repeat occurrence of a previously acknowledged alert, re-triggered by only a small data change, the function automatically acknowledges this, and moves the alert from the 'New alerts' column to the 'Repeat ack alerts' column. If it detects a significant change in a continuing alert (whether previously acknowledged or not), it flags this up in the appropriate 'Change' column, indicating to the operator that the alert should be reinvestigated. If a new alert occurs and the automated reasoning function determines that it is due to benign maintenance, the function automatically acknowledges this, and moves the alert from the 'New alerts' column to the 'Prob benign maint alert' column.

Figure 3-3 shows the existing alerts display. This is a listing of all the alerts in the time window selected (either current or unacknowledged, depending on which radio button is selected).

Figure 3-4 presents a mock-up of an updated display of alert records, with an additional 'automated assessment' column. This includes high level outputs from the combined data mining and reasoning functions. The 'Ab Maint' (Abnormal Maintenance) and 'Inst' (Instrumentation) columns show the outputs from a reasoning function responding to the presence of recognised patterns in the data to indicate the probability that the alert was triggered by a non-benign maintenance action or a HUMS instrumentation problem. The 'Significance' column shows the fused output from combined data mining and reasoning functions, which is an automated overall alert significance assessment. The alert significance is seen as a key piece of information to ensure that an operator pays close attention to an alert that the system considers has high significance. This significance assessment is based on multiple criteria; the rarity of the identified patterns (a significant aircraft component fault is a rare occurrence), the magnitude of the data change triggering an alert (as indicated by the combined magnitude of the normalized IF values), and data

trend information. The final piece of information is an output from the change detection function indicating whether the data has changed since the alert was first triggered.

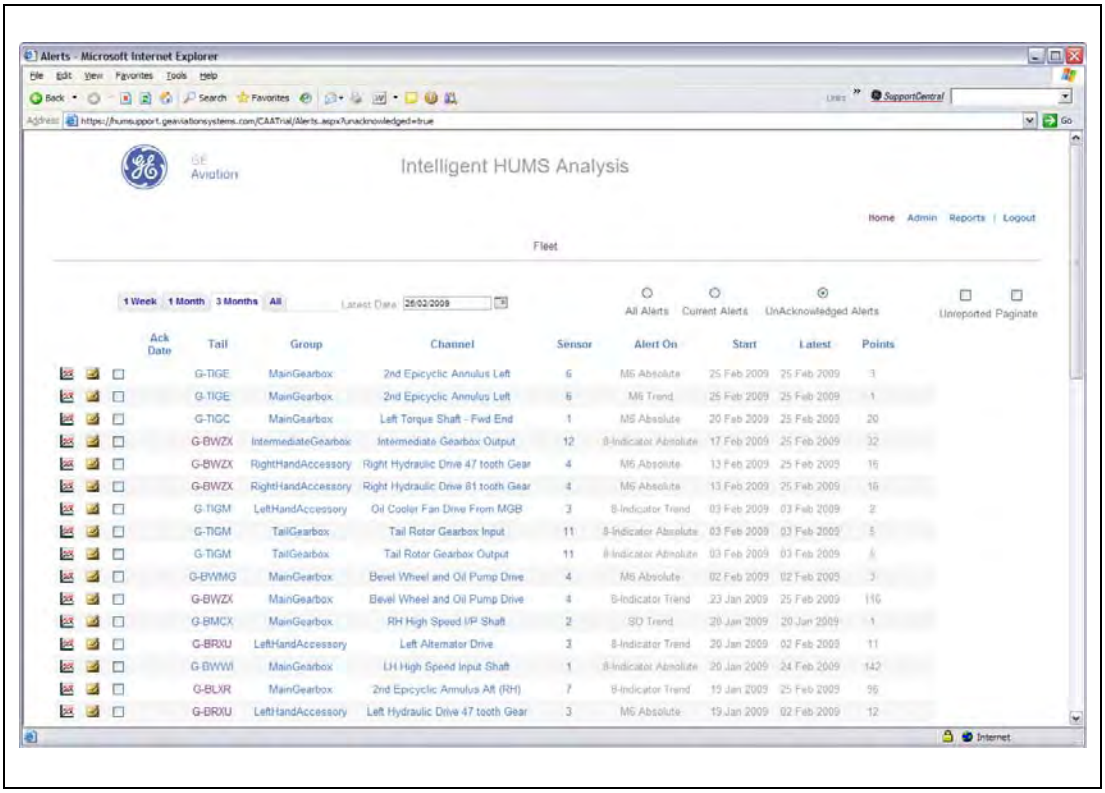

Figure 3-3 Existing alerts display

Ack Date	Aircraft	Group	Channel	Sensor	Models	Start	Latest	Automated assessment Ab Maint	Inst	Significance	Change	Rev
	G-BRXU	Main Gearbox	RH High Speed IP	2	8IA 8IT	10 Dec 08	14 Dec 08					
	G-BWWI	LH Accessory	Hyd Drive 47T	3	M6A M6T	10 Dec 08	14 Dec 08					
	G-BWWI	LH Accessory	Hyd Drive 81T	3	M6A M6T	10 Dec 08	14 Dec 08					

Figure 3-4 Mock-up of possible new alert information

A final mocked-up display presented in Figure 3-5 shows a possible new alert information page. The intent is to summarise all the information that the system could provide on an alert within one display. This would remove the need for an operator to manually create a picture of an alert by viewing many chart displays, and would also provide more alert information than the operator could obtain from reviewing the existing system data.

The new information that could potentially be provided by the data mining and reasoning functions demonstrated consists of:

1 Summary information – an alert summary showing which anomaly models are in alert, the current PA levels for all models, a data trend severity assessment for each model type, and the overall significance assessment (a repeat of that shown in Figure 3-4).

2 Trend classification – determining the extent to which the data trend fits classifications of ‘Clear trend’, ‘Noisy’ and ‘Step’.

3 Significance assessment based on the mining of a database of historical alerts – showing both the rarity of the identified patterns and the magnitude of the data change triggering an alert. (The overall significance assessment also takes into account trend information).

4 Recognised patterns – indicating the probability that the alert was triggered by a non-benign maintenance action or a HUMS instrumentation problem (a repeat of the information shown in Figure 3-4), and identifying the extent to which any documented fault cases match the data characteristics of the current alert.

5 Diagnostic summary information – summarising the information provided by the sets of IFs calculated for each anomaly model in the form of the significance of six different features of the vibration data as measured by the HUMS CIs. All but one of these features ('Impulsiveness') are based on pairings of HUMS CIs ('Signal energy' = SIG_SD+PP, 'Base energy' = ESA_SD+PP, 'Shaft order energy' = FSA_SO1+2, 'Gear mesh' =FSA_MS+GE, and 'Wear' = ESA_SD/SIG_SD).

Figure 3-5 illustrates how the new alert information display would appear for an alert caused by an example HUMS instrumentation problem. This causes some extremely abnormal data values in the HUMS CIs related to vibration signal 'Base energy', but the data is not trending and, as instrumentation problems are relatively common, the data patterns are not unusual. It therefore is given a low overall significance. The system recognises that these patterns can be caused by instrumentation faults, with a wiring harness fault giving the closest match. (The example case is contained in reference [3]).

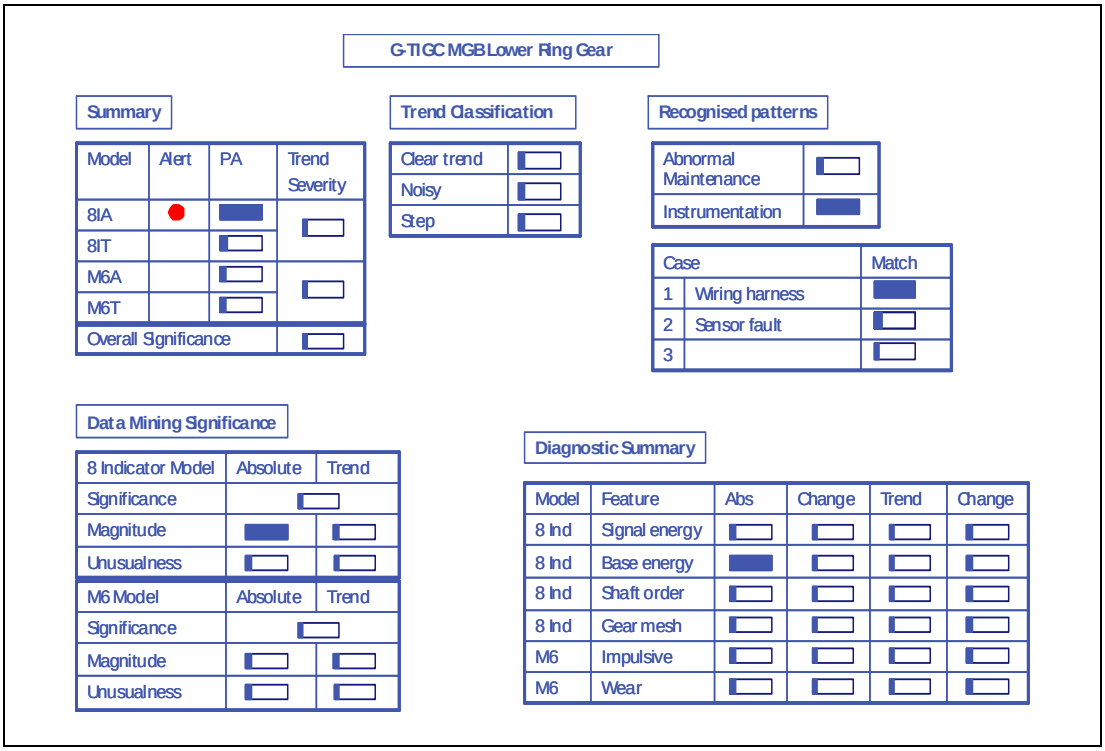

Figure 3-5 Mock-up of possible new alert information display

In contrast, Figure 3-6 shows how the display would appear for the cracked MGB bevel pinion fault case (see references [2] and [4]). Both the absolute and trend models are in alert, a clear trend is detected, multiple features of the vibration signal are changing, and the data patterns are very unusual. Therefore the system gives this case a high overall significance assessment. The data patterns do not match any documented fault cases, indicating that these have not been seen before. All this

information should prompt an operator to pay very close attention to the alerts, and take appropriate maintenance actions.

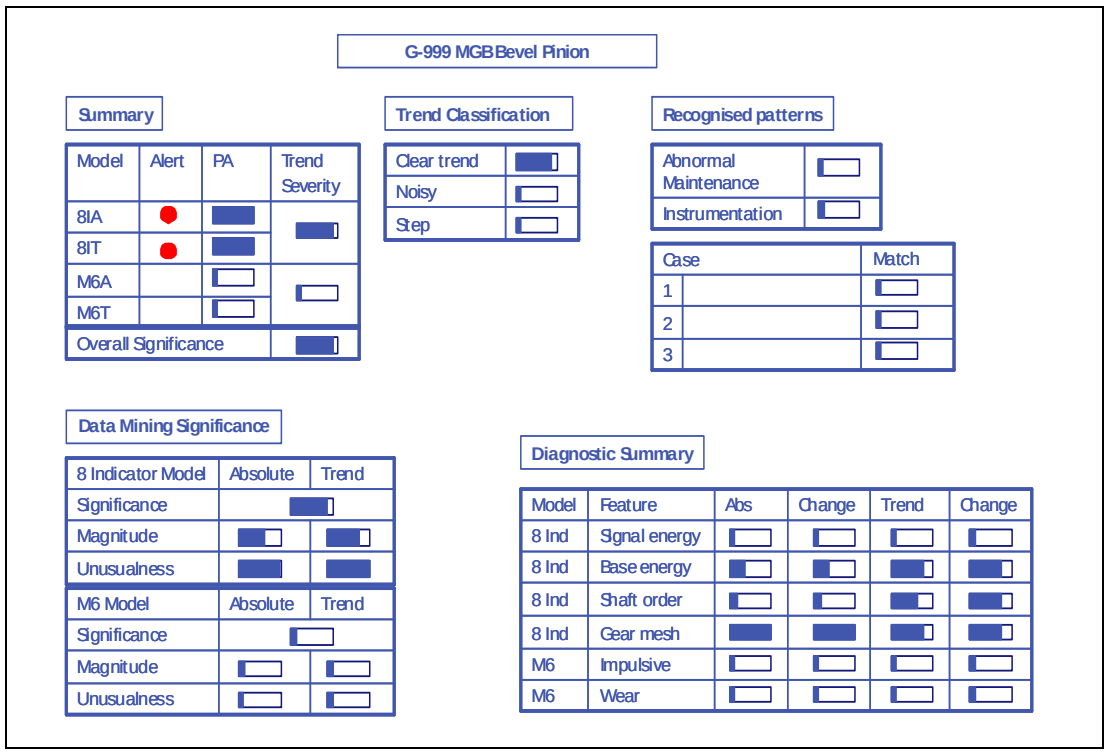

Figure 3-6 Alert information display for cracked MGB bevel pinion

4 Data Mining Demonstration

The primary purpose of the data mining task was to explore the data generated in the HUMS anomaly detection trial to test the modelling capability, and also to extract new knowledge from the data for use in automated diagnostic reasoning. The output could also be used for post-processing of anomaly alerts to reduce operator workload. The data mining effort focused on automated change detection, trend analysis, anomaly significance assessment, and data analysis to support Case Based Reasoning. The last two items utilised GE Aviation's ProDAPS (Probabilistic Diagnostic and Prognostic Solutions) data modelling tool, which is briefly introduced in the following section.

4.1 ProDAPS Data Modelling Tool

The ProDAPS data modelling tool is a key part of GE Aviation's advanced data analysis toolset, and provides the core functionality for the building of the anomaly models implemented in the CAA trial system. The tool utilises elements of a state-of-the-art data mining framework which provides a number of core components for data source connection, data pre-processing, data model creation and storage, model query and predictions. The core data mining algorithm utilised is a probabilistic mixture model cluster algorithm that has many parameters available to control model complexity. The algorithm is based upon the best techniques developed by academic and industrial research communities, and incorporates proprietary extensions to satisfy GE Aviation's specific application requirements.

Following data checking, cleansing and pre-processing, a mining process includes three stages. The first is model creation, for which a number of items must be specified, including the data source, the input and predictable data items, and the algorithm configuration parameters. The second is model training, where the tool obtains the training data, generates the mining model, and saves it. The final stage is model prediction, where the tool obtains the data to be mined, and executes the specified type of prediction, for example cluster membership, prediction of missing values, or case likelihood. The models built using the mixture model cluster algorithm record various information, and the flexibility of the tool allows this information to be utilized to adapt the models for anomaly detection. The tool facilitates the batch processing of models, rapid prototyping of new modelling approaches, and implementation of the model adaptation and anomaly detection processing.

4.2 Change Detection

The purpose of change detection is to reduce workload associated with the anomaly detection system by suppressing repeat alerts triggered by small changes in data that are close to the threshold, while flagging up significant changes in data that are constantly exceeding a threshold (particularly where an alert may have previously been acknowledged), indicating that a fresh review is required. An example of the frequent repeat triggering of alerts over a considerable period of time that the change detection is intended to suppress is presented in Figure 4-1. This shows almost two years of data from the MGB left torque shaft – fwd end on G-TIGC. The alerts were caused by increased levels of ESA_M6, which had no trend for a long period (Figure 4-2).

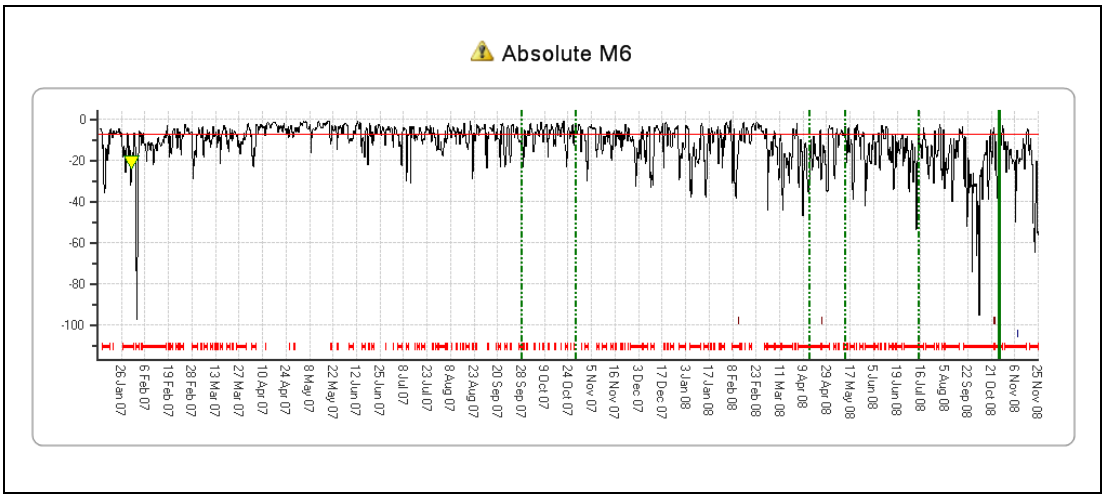

Figure 4-1 G-TIGC MGB left torque shaft – forward end – M6A model FS values and alerts (shown as red bars)

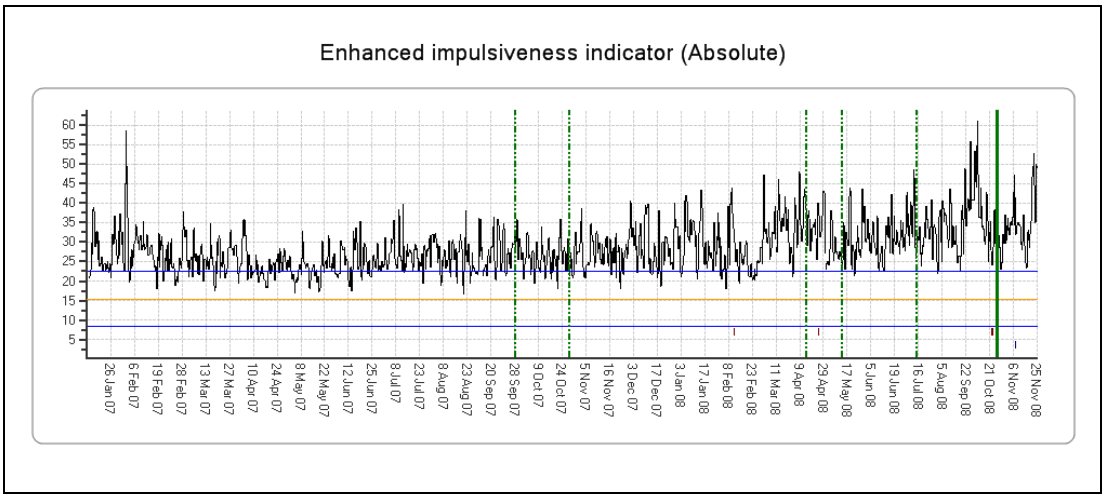

Figure 4-2 G-TIGC MGB left torque shaft – forward end – ESA_M6

A change detection algorithm was developed for analysing the '8-Indicator Absolute' (8IA) anomaly model IF values obtained from data points in alert. The eight IFs were combined into a normalised 4-dimensional vector with components based on natural pairings of the IFs calculated for the HUMS CIs ('Signal energy' = SIG_SD+PP, 'Base energy' =ESA_SD+PP, 'Shaft order energy' = SO1+2, and 'Gear mesh' = FSA_MS+GE). These indicator pairings normally exhibit some degree of correlation. A 4-D vector was used in preference to an 8-D vector based on the individual IFs as it was considered to be more robust, and there were advantages in limiting the dimensionality of the feature space as, the greater this dimensionality, the more data that is required to provide adequate modelling.

The vector angle between each new alert data point and the last acknowledged alert point was calculated. The data for the new alert point was then determined to show no change if both the change in vector angle and the change in magnitude of the FS remained below pre-set thresholds. For this exercise, the thresholds were based on the results of a historical data review, assessing the levels at which the data should be considered to exhibit a noticeable change.

Repeat alert suppression was demonstrated using example results from 18 months of data from the 2nd stage sun gear on G-TIGC. The results were based on the assumption that the system could automatically suppress repeat occurrences of a previously acknowledged alert provided the change detection algorithm indicated that

there was no significant change in the data. A total of 117 alerts were triggered on this component. The high number of repeat alerts was due to frequent small changes in data that was very close to the threshold at which the component model classed this as anomalous (Figure 4-3).

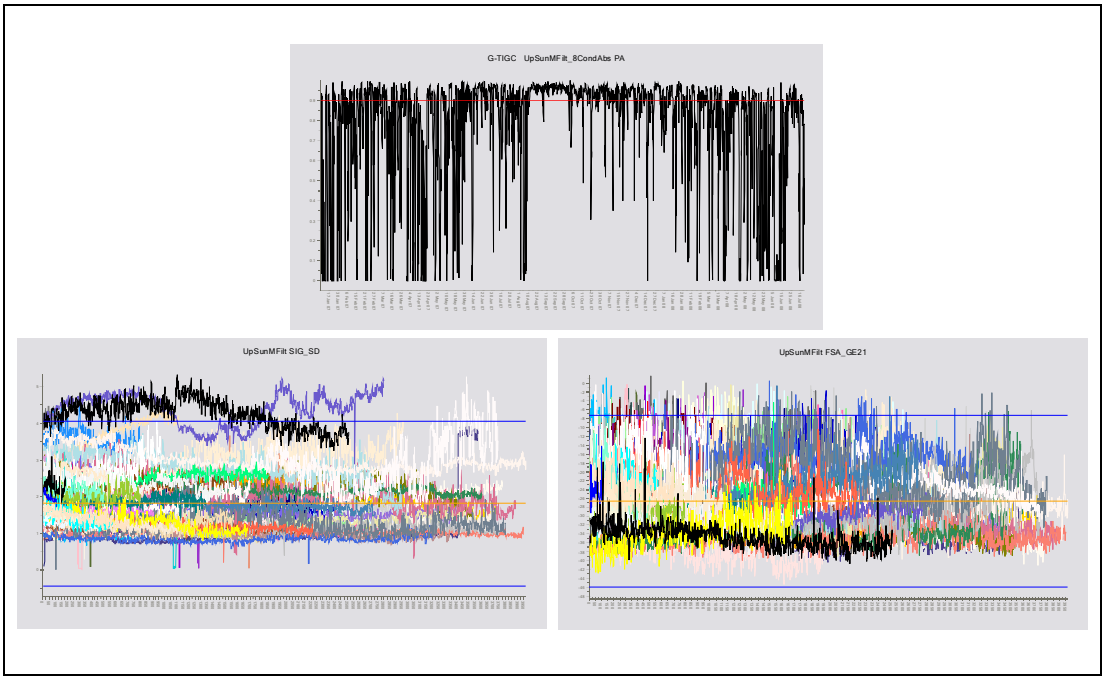

Figure 4-3 G-TIGC 2nd stage sun gear – PA (top), SIG_SD (bottom left) and FSA_GE (bottom right) (G-TIGC values shown in black)

The historical database showed that the alert had been initially acknowledged by the system user, and then re-acknowledged two further times, during this period (the change detection algorithm reset itself at each of these manual acknowledgements). Of the remaining total of 114 unacknowledged alerts, 111 were automatically acknowledged by the algorithm as repeats of a previously acknowledged alert. The algorithm also detected one short duration change in the data during the period an alert was acknowledged (due to a short duration data spike). The algorithm classified three short alerts (out of the total of 114) as new alerts because it detected a change in the data since the previously acknowledged alert.

The detection of a change in alert characteristics was demonstrated using data from the right torque shaft – fwd end on G-PUMI. There were a total of eight alerts. One alert was manually acknowledged, and five of the remaining seven alerts were automatically acknowledged by the algorithm as repeat alerts, with the remaining two being classed as new alerts. A change was also detected during the alert period following an acknowledgement. A large step change in the data occurred and, although there was some variability in PA values, the data remained in alert before and after the step change (Figure 4-4). Although PA values largely remained above the threshold, the HUMS CI ESA_PP switched from abnormally high values to abnormally low values. Currently, due to the pre-existing acknowledged alert, the step change would go undetected. However the change detection algorithm flagged this up, and identified all subsequent data points as changed. No further changes were identified once the alert associated with the new data after the step change was acknowledged.

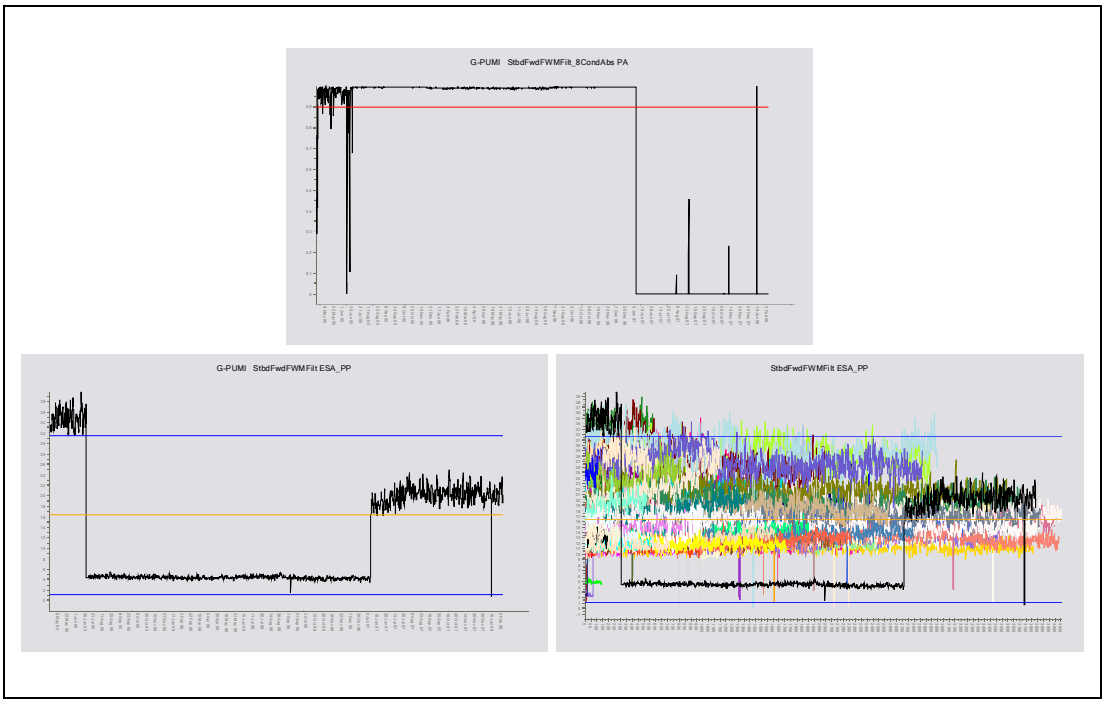

Figure 4-4 G-PUMI right torque shaft forward end – PA (top), ESA_PP (bottom left) and ESA_PP fleet view (bottom right)

4.3 Trend Analysis

The second topic addressed in the data mining task was trend analysis, with investigations being carried out into methods for automated trend detection, classification and severity assessment. Trend information is critical to determining the nature and significance of an anomaly alert.

4.3.1 Short Term Trend Detection and Classification

An investigation was carried out into the extraction of short term trend features, followed by an assessment of their significance. The basic approach used was to histogram the data and then detect patterns of transition between different bins of the histogram. A steady trend will gradually traverse bins in a consistent manner from one end of the histogram to the other, with the rate of traverse depending on the trend severity. However a noisy trend will jump up and down between bins.

The histogram was applied to a range of FS values, with the histogram window starting in the normal data range before the alerting threshold is reached, and extending to a multiple of the threshold. Ideally, the window range should include the proportion of the FS distribution capable of showing trend information (i.e. what the anomaly model is sensitive to). Logic-based rules were then applied to the histogram to characterise the behaviour of the data as it transitioned between bins. These rules looked for features such as the number of bins jumped during successive acquisitions, the time spent within a specific bin, the direction of travel (i.e. random, increase or decrease), etc. Rules were also generated for assessing the IFs to provide further information. These rules used the IFs to determine what was driving the trend (trends in CIs that are known to be reliable measurements could give more emphasis to an alert).

Example results from applying the histogram technique to a set of MGB bevel pinion data are presented in Table 4-1 (sorted by 'Total Bin Transitions') and Table 4-2 (sorted by 'Total Duration'). These tables show the top ten trends identified for different 'gearbox fits' (i.e. combinations of a particular gearbox installed in a particular aircraft)

where the data was in alert (i.e. generating absolute model PA values of greater than 0.90). The 'Abs Bin Trans' and 'Trend Bin Trans' values are counts of the total number of histogram bins transitioned by the FS from the 'Absolute' and 'Trend' models respectively during an identified continuous trend. A continuous trend was defined as a progressive transition of one or more bins within a moving window of data points. (If fewer than two data points within a 5-point window transition bins, or if any point transitions bins in the opposite direction, the trend is considered to have ended or reversed in direction). The 'Total Bin Transitions' is the sum of the total bin transitions from the 'Absolute' and 'Trend' models. The 'Total Duration' is the sum of the number of data points from the 'Absolute' and 'Trend' models during which the continuous trend was determined to exist.

The histogram trend analysis has identified the clear developing FS trend generated by the cracked MGB bevel pinion fault case (gearbox fit 999) as significant, and ranked this joint 7th in terms of magnitude (Table 4-1) and 2nd in terms of duration (Table 4-2). The FS trend is shown in Figure 4-5 (note that the dates in the chart are artificial to separate this data set from the Bristow 332L fleet data).

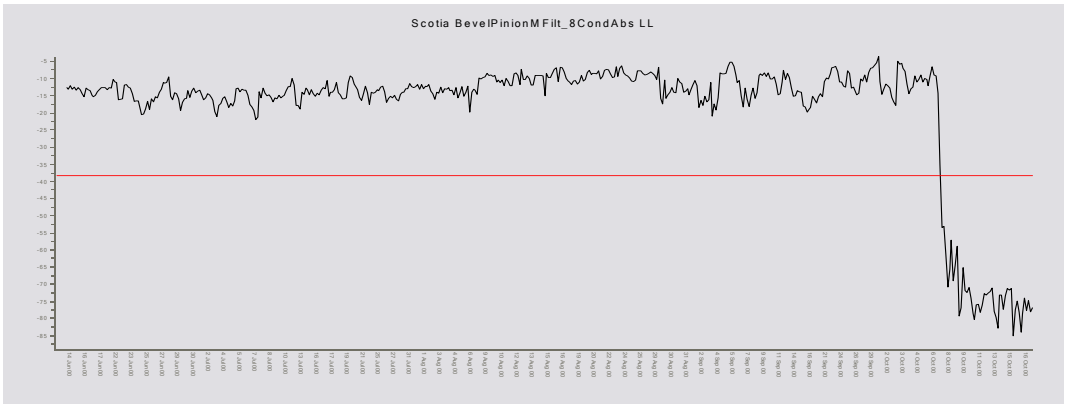

Figure 4-5 Cracked MGB bevel pinion (gearbox fit 999) – 8IA model FS values

The IF flags show that the top six entries in Table 4-1 are all related to signal energy CIs (the 'Energy_flag', based on a combination of ESA_SD+PP and SIG_SD+PP, is set). An inspection of the CI data showed that, in all these cases, there had been a large short duration drop in signal energy levels, possibly due to an instrumentation problem.

The example results show that the histogram technique could successfully be used to detect and classify short term trends. However, while the technique was effective over short period trends and there is scope to refine its implementation, variability in the data (i.e. signal noise), made it more difficult to implement over longer periods.

Table 4-1 Bevel pinion – histogram trend analysis, results sorted by total bin transition (see key on following page)

Gearbox fit	Abs_FS	Trend_FS	Abs_PA	Trend_PA	SO_flag	Energy_flag	GE_MS_flag	Abs Bin Trans	Trend Bin Trans	Total Bin Trans	Total Duration
1059	-67.16	-141.41	0.99	1.00	0	1	0	11	27	38	3
132	-85.90	-36.36	0.99	0.98	0	1	0	13	5	18	5
1028	-84.44	-29.94	0.99	0.96	0	1	0	15	3	18	3
1029	-104.42	-40.51	1.00	0.99	0	1	0	12	5	17	5
1053	-140.46	-33.98	1.00	0.97	0	1	0	13	4	17	5
1053	-134.78	-27.87	1.00	0.95	0	1	0	12	2	14	5
999	-53.48	-41.40	0.98	0.99	0	0	1	7	6	13	6
1056	-71.27	-40.16	0.99	0.99	1	0	0	8	5	13	7
1132	-60.58	-40.28	0.98	0.99	1	0	0	8	5	13	5
152	-62.60	-31.79	0.99	0.97	1	0	0	8	4	12	5

Table 4-2 Bevel pinion – histogram trend analysis, results sorted by total duration (see key on following page)

Gearbox fit	Abs_FS	Trend_FS	Abs_PA	Trend_PA	SO_flag	Energy_flag	GE_MS_flag	Abs Bin Trans	Trend Bin Trans	Total Bin Trans	Total Duration
1056	-71.27	-40.16	0.99	0.99	1	0	0	8	5	13	7
999	-53.48	-41.40	0.98	0.99	0	0	1	7	6	13	6
132	-85.90	-36.36	0.99	0.98	0	1	0	13	5	18	5
1029	-104.42	-40.51	1.00	0.99	0	1	0	12	5	17	5
1053	-140.46	-33.98	1.00	0.97	0	1	0	13	4	17	5
1053	-134.78	-27.87	1.00	0.95	0	1	0	12	2	14	5
1132	-60.58	-40.28	0.98	0.99	1	0	0	8	5	13	5
152	-62.60	-31.79	0.99	0.97	1	0	0	8	4	12	5
1059	-67.16	-141.41	0.99	1.00	0	1	0	11	27	38	3
1028	-84.44	-29.94	0.99	0.96	0	1	0	15	3	18	3

Key to Tables 4.1 and 4.2

Gearbox fit:	An instance of a gearbox installed in an aircraft
Abs_FS:	'Absolute' model FS at the end of a continuous trend
Trend_FS:	'Trend' model FS at the end of a continuous trend
Abs_PA:	'Absolute' model PA value at the end of a continuous trend
Trend_PA:	'Trend' model PA value at the end of a continuous trend
SO_flag:	Set if the alert is due to FSA_SO1 or FSA_SO2
Energy_flag:	Set if the alert is due to ESA_PP+SD or SIG_PP+SD
GE_MS_flag:	Set if the alert is due to FSA_GE or FSA_MS
Abs Bin Trans:	Total number of histogram bins transitioned by the 'Absolute' model FS during the detected trend
Trend Bin Trans:	Total number of histogram bins transitioned by the 'Trend' model FS during the detected trend
Total Bin Trans:	Sum of the 'Absolute' and 'Trend' model bin transitions
Total Duration:	Sum of the number of 'Absolute' and 'Trend' model data points during the detected trend

4.3.2 Longer Term Trend Detection and Classification

Further trend analysis work was undertaken, resulting in an algorithm more suited to the analysis of longer term trends (i.e. those continuing over a larger number of data points). This algorithm first detects the presence of a trend in anomaly model FS or HUMS CI values, and determines whether this is positive or negative. It then assesses the severity and duration of the trend. The algorithm will typically identify multiple trends (occurring at different periods of time) in the complete time history of the FS or CI values for a monitored component. The algorithm design priorities were the quick detection of developing trends (i.e. with a minimum of lag), and good resilience to noisy, non-trending signals to prevent the generation of false alerts.

The algorithm first smoothes the signal using a process that acts to reduce the difference between the smoothed signal (*S*) and the current actual signal (*x*) by an amount that varies as a function of the magnitude of the error between the two. Small, noisy fluctuations have little effect on the smoothed signal, but trends rapidly cause an increased error that drives the smoothed signal to react and follow the trend shown by the input. The filter algorithm for this first stage of smoothing is shown below (n = the current data sample and n-1 = the previous sample):

$$S_n = S_{n-1} + f(y)$$

where

$$y = (x_n - S_{n-1})$$

The smoothed signal is 'reset' following the end of a detected trend, with the gradient being set back to zero.

The gradient of the smoothed signal is calculated as the difference between successive data points (i.e. the first derivative of the signal). The gradient is normalised, and then smoothed by a second filter algorithm. The function *f*(*y*) used in the smoothing is typically a linear function for the raw signal, but for smoothing the gradient it has an exponential form where '*k*' and '*c*' are configurable constants:

$$f(y) = y \times 2^{k|y|^c}$$

Finally, a five point moving median filter is applied to further smooth the output.

The normalisation of the gradient is carried out so that a constant trend detection threshold value can be applied regardless of the range of the input signal, and also to reduce the false detection of trends by adjusting the scaling based on the noise of the signal. The factor used in the scaling process continuously varies to cope with signals that start off stable but become noisy with no clear trends. The normalised output is then used to detect trends when it exceeds the threshold value.

Two severity measures were generated for each trend, and used in the ranking; the initial severity (best fit gradient over trend_start – five data points to trend_start + 20 data points), and the final severity (over trend_start - five data points to trend_end + five data points). The initial severity gives an approximation to a 'real-time' use of the algorithm but tends to give high severity for trends that are later revealed to be just short period spikes in the data, and may indicate a direction which is opposite to the actual trend in noisy signals. The final severity measure reduces the significance ranking of spikes that return to initial levels, and prioritises those trends that continue to develop over a period of time. (For real time applications, the severity is calculated as trend_start - five data points to the latest data point if the trend is continuing).

The trend detection algorithm was demonstrated on data from the MGB 2nd stage epicyclic annulus aft (RH) on G-BWZX. The anomaly detection system had detected multiple anomalous CI trends on the MGB 2nd stage epicyclic annulus, and the gearbox was removed for metal contamination in March 2006 (the example is included in reference [3]). The FS trend is shown in Figure 4-6. The cause of the anomaly alert was a rising trend in SIG_SD (Figure 4-7).

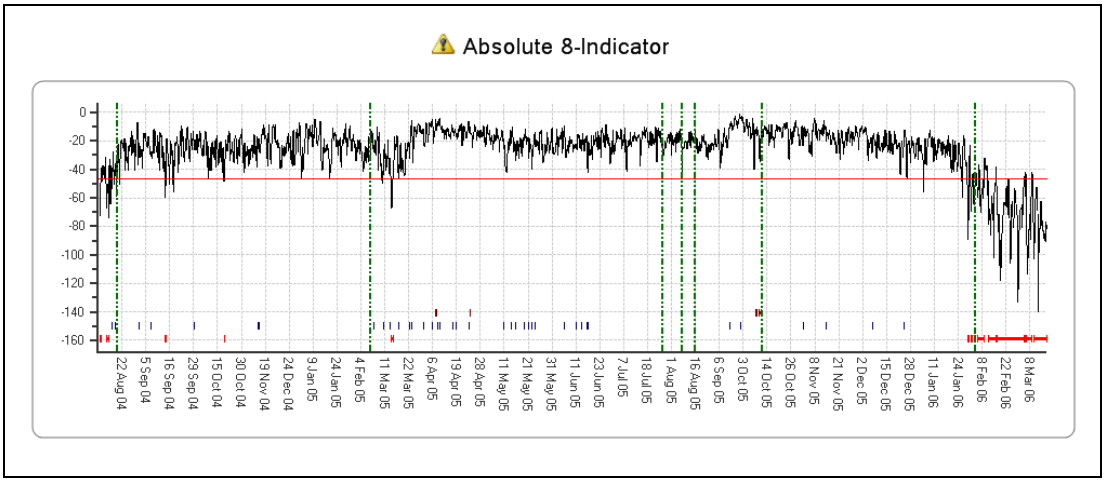

Figure 4-6 G-BWZX MGB 2nd epicyclic annulus aft (RH) – 8IA model FS

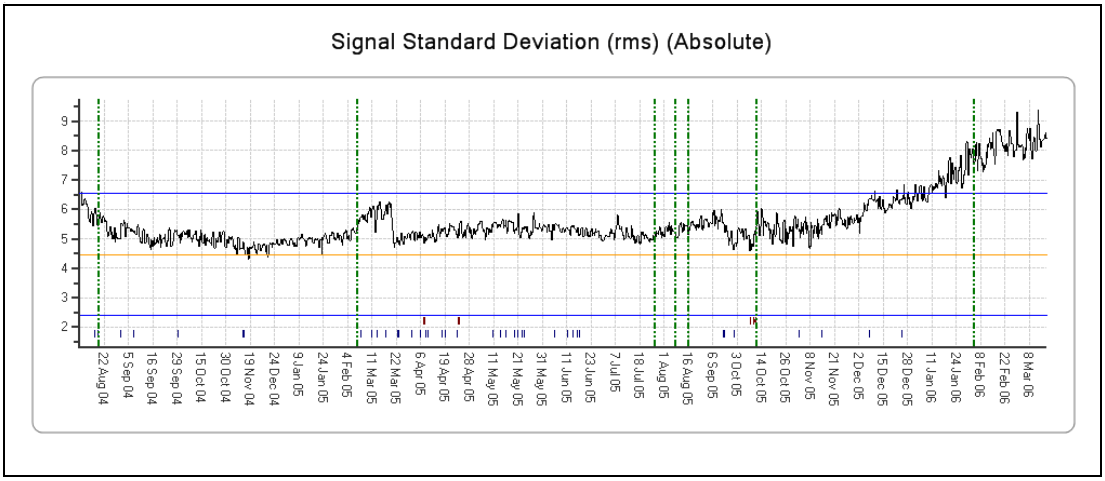

Figure 4-7 G-BWZX MGB 2nd epicyclic annulus aft (RH) – SIG_SD

The trend detection algorithm was applied to the SIG_SD HUMS CI trend, producing the outputs shown in Figure 4-8 and Figure 4-9. The pink line in Figure 4-8 shows the output from the smoothing filter and the yellow line shows detected trends, with a step down from the mean data level indicating a negative trend, and a step up indicating a positive trend. It can be seen from the figure that, over the life of the gearbox, one negative trend and two positive trends have been detected. (The magnitude of the steps in the yellow line does not have any significance)..

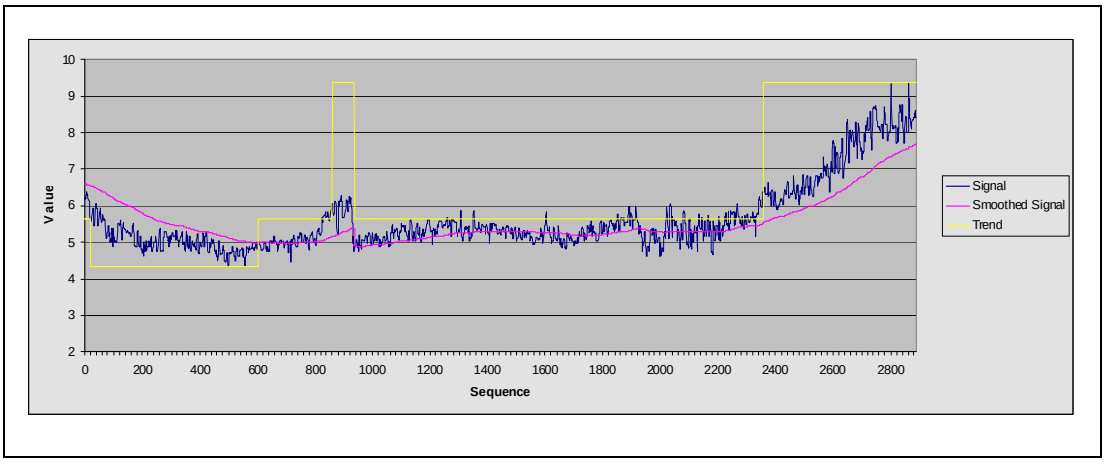

Figure 4-8 G-BWZX MGB 2nd epicyclic annulus aft (RH) – SIG_SD trend analysis

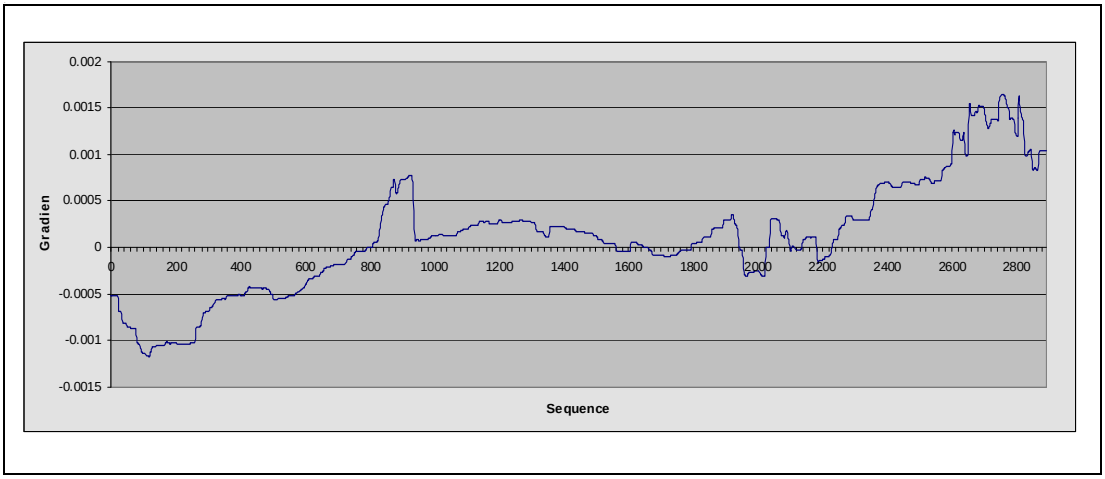

Figure 4-9 G-BWZX MGB 2nd epicyclic annulus aft (RH) – SIG_SD trend gradient

The SIG_SD trends detected are shown in Table 4-3 together with their severities and durations. The sign attached to the trend number indicates whether a positive or negative trend has been detected. The sign attached to the severity calculation also indicates whether the result is positive or negative. It is noted that although trend 2 is positive, the final severity value (calculated over the period trend_start - five data points to trend_end + five data points) is negative as the data has stepped back down again.

Table 4-3 G-BWZX MGB 2nd epicyclic annulus aft (RH) – SIG_SD trends, severities and durations

Trend	Initial Severity	Final Severity	Start	End	Duration
-1	-0.00366	-0.00261	34	611	578
2	0.00303	-0.00002	870	944	75
3	0.00401	0.00431	2369	2899	531

The trend detection algorithm was applied to FS data from the MGB 2nd stage epicyclic annulus aft (RH) on G-BWZX (Figure 4-6), and to all the other gearbox fits in the database. A ranking of the top ten gearbox fits based on trend severity is shown in Table 4-4. Cases from the first six-month trial period included in reference [3] are identified by the green shading. G-BWZX (gearbox fit 816) is ranked 10th and G-TIGC (gearbox fit 1053) with an epicyclic stage wiring harness problem is ranked 5th. The FS trace for this component is shown in Figure 4-10. The trend is less visually apparent due to the scale of the trace, with the extremely low FS values being recorded (the models used in the first trial period gave a flat-line FS output at -700).

Table 4-4 MGB 2nd epicyclic annulus aft (RH) – 8IA model FS trends, severities and durations

Gearbox fit	Trend No.	Severity	Duration
186	-4	-1.278	151
1041	-4	-0.586	600
1132	-3	-0.302	105
243	-2	-0.296	84
1053	-9	-0.273	228
1056	-7	-0.085	322
172	-12	-0.084	42
183	-1	-0.079	350
798	-3	-0.077	149
816	-2	-0.048	240

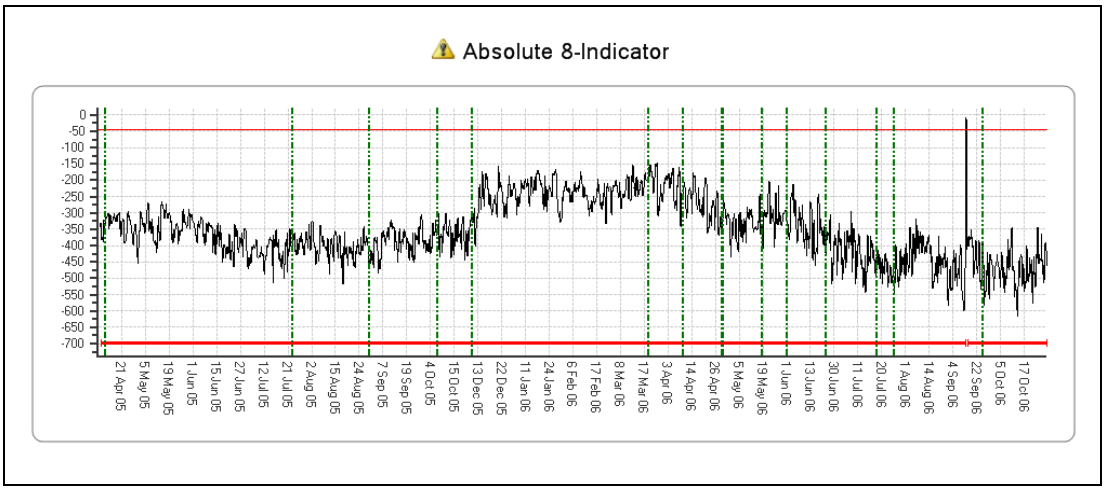

Figure 4-10 G-TIGC MGB 2nd epicyclic annulus aft (RH) – 8IA model FS

Of particular interest is gearbox fit 186, ranked as having the most significant FS trend in Table 4-4. This is from G-TIGE, and the identified trend occurred before the start of any of the in-service trials in September 2003 (Figure 4-11). There was a rapid and very large decreasing trend in the FS values, caused by a rapid increase in SIG_SD (Figure 4-12). The reason for the trend is unknown, however it is clearly significant. The example highlights the benefits of data mining tools such as automated trend detection and severity assessment in identifying significant trends in large volumes of data.

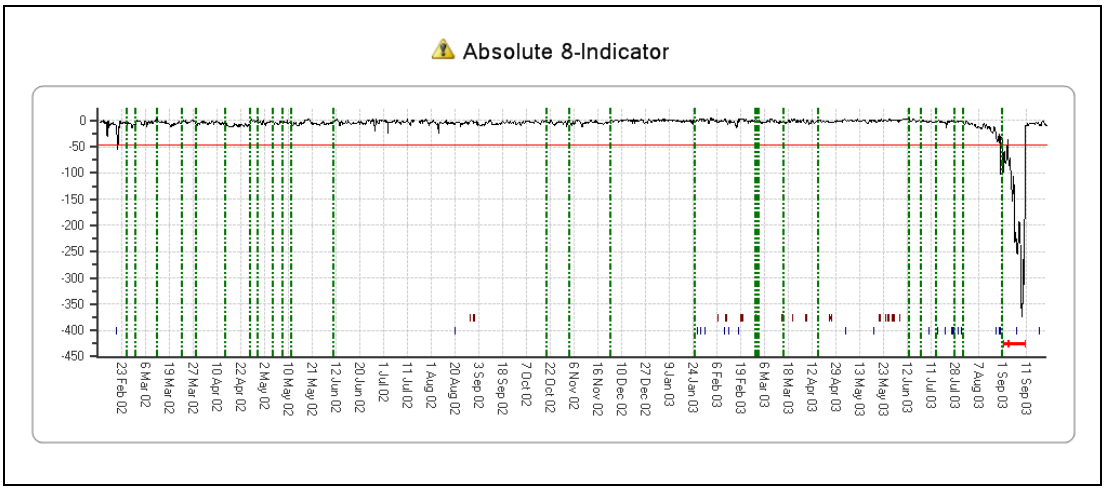

Figure 4-11 G-TIGE MGB 2nd epicyclic annulus aft (RH) – 8IA model FS

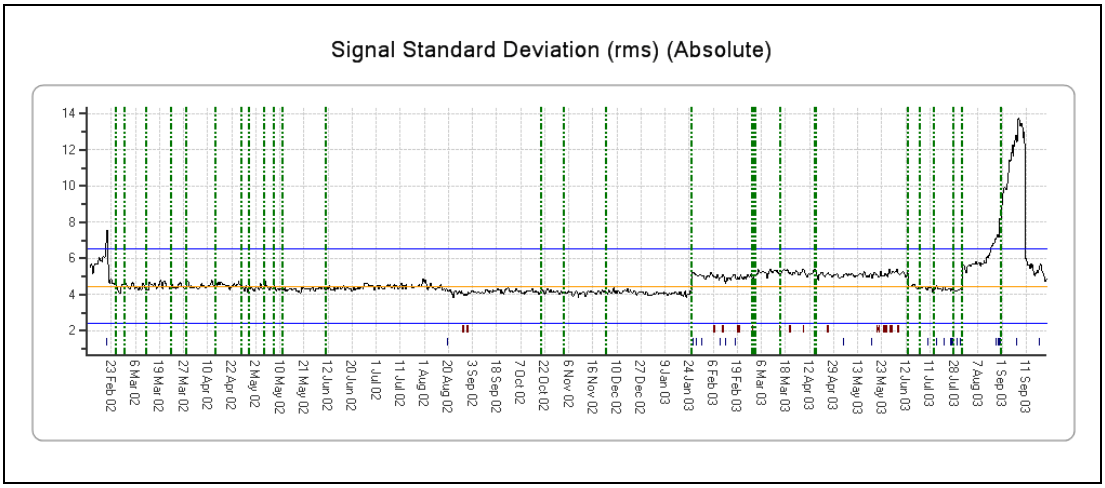

Figure 4-12 G-TIGE MGB 2nd epicyclic annulus aft (RH) – SIG_SD

A ranking of the top five gearbox fits based on FS trend severities detected on the MGB bevel pinion is shown in Table 4-5. The trend detection algorithm has identified the cracked bevel pinion fault case (gearbox fit 999) as having the most significant FS trend in the database.

Table 4-5 MGB bevel pinion – 8IA model FS trends, severities and durations

Gearbox fit	Trend No.	Severity	Duration
999	-2	-0.71714	47
1053	-8	-0.17011	227
133	-1	-0.05948	45
810	-1	-0.05337	126
182	-5	-0.03773	237

The trend detection algorithm has successfully identified significant trends in FS and HUMS CI data. However some step changes in the data can also trigger a high trend severity, therefore the algorithm should be used in conjunction with a step change detector.

4.3.3 Signal Step Change and Noise Detectors

GE Aviation had previously developed a step change detector on another project. This searches for combined steps in both the current signal value (based on a delta between successive points) and between forward and backward windows of this. The step detection threshold automatically adjusts itself based on the range and standard deviation of the input signal.

The step change detection algorithm was applied to anomaly model FS values for several different components. Table 4-6 shows steps detected in the FS data for the MGB 2nd stage planet gear. The step down in FS on G-TIGC (gearbox fit 183) is shown in Figure 4-13. This occurred in February 2004 before the start of the trial, and its cause is unknown. Step changes were also detected on G-TIGG and G-TIGJ (gearbox fits 924 and 1056) associated with the installation and removal of a faulty IHUMS DAPU (Figure 4-14). This is one of the documented cases from the first six-month trial period (reference [3]). It can be seen from the comments in Table 4-6 that two steps were observed in the data that were not detected by the algorithm due to a high level of signal noise.

Table 4-6 MGB 2nd stage planet gear – 8IA model FS steps

Aircraft	Gearbox fit	Sequence	FS	Detected Step
G-TIGC	183	80	-83.61	Step down
G-TIGJ	924	830	-13.57	Step up - DAPU removal (no step down detected due to noise)
G-TIGG	1056	2202	-254.62	Step down - DAPU installation
G-TIGG	1056	2325	-9.75	Step up - DAPU removal
G-BMCW	1161	160	3.24	Step up (previous step down not detected due to noise)

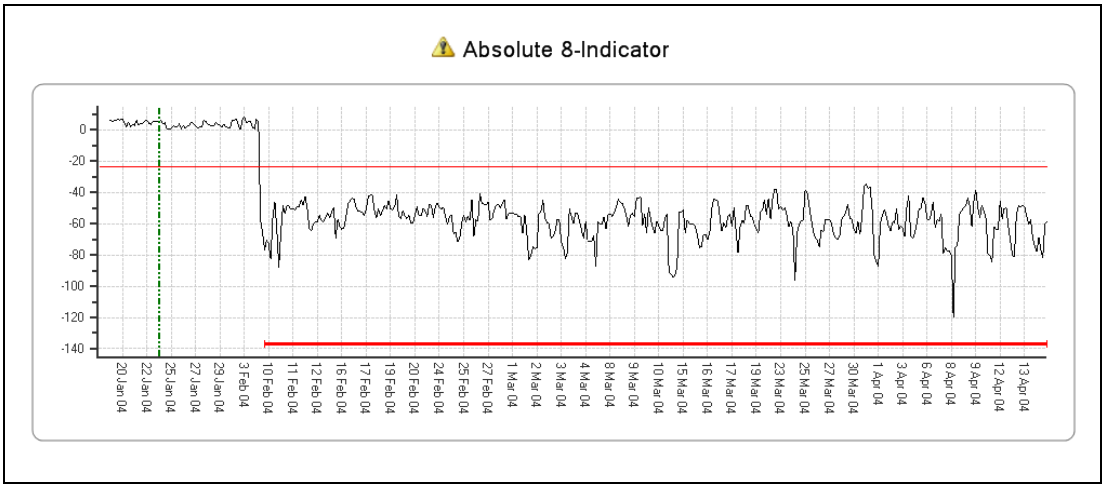

Figure 4-13 G-TIGC MGB 2nd stage planet gear – 81A model FS

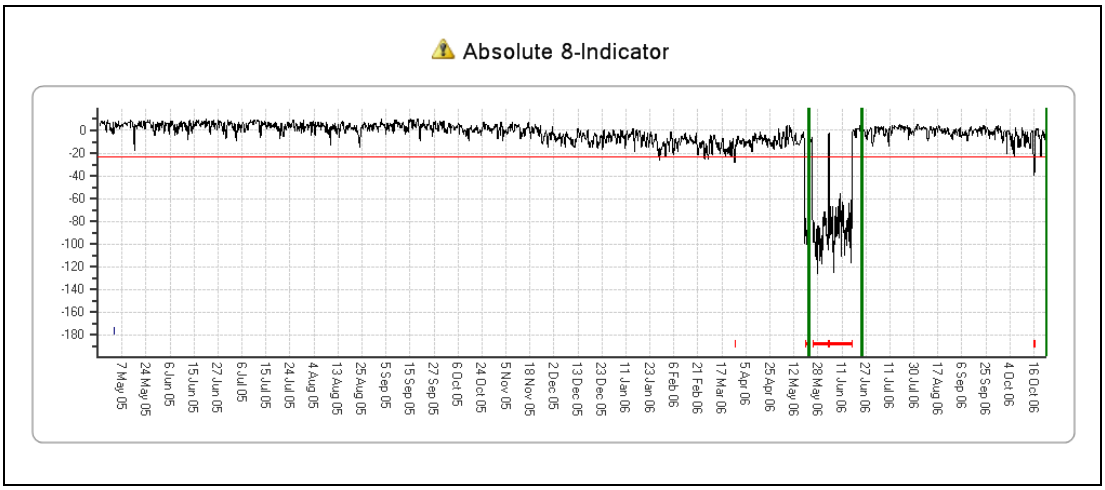

Figure 4-14 G-TIGG MGB 2nd stage planet gear – 81A model FS

The results showed that the step change detector is able to identify data steps, and is a useful addition to the data mining toolset. However very noisy signals can mask steps or trigger false step detections, therefore the algorithm should be used in conjunction with a signal noise detector.

To provide a comprehensive trend analysis capability, it is considered that two final trend processing elements would be required, however these have not yet been implemented. The first is a signal noise detector, which would probably be based on the standard deviation of the signal and a calculation of the error between the actual signal and the smoothed trend signal. The second is a reasoning network to fuse the results from the various independent trend analysis tools to provide a single overall signal trend classification.

4.4 Anomaly Significance Assessment

A potentially very useful data mining capability would be an automatic assessment of the significance of a detected anomaly on the basis of a combination of a severity assessment and a measure of uniqueness or rarity. The measure of rarity may be of most significance as instrumentation faults can generate highly abnormal data, but are common in comparison to the occurrence of a real aircraft fault. This type of analysis does require a sizeable database of historical data to enable criteria such as rarity to be properly evaluated.

The data modelling performed to demonstrate an example automated anomaly alert significance assessment used the standard IFs (i.e. those presented to the operator) as inputs. The IFs provide information on the HUMS CIs driving an anomaly alert however, unlike the CIs, they are normalised and so enable data from different components to be combined in the modelling process. One item of information currently lacking from the IF data is direction, indicating whether a CI driving an anomaly indication is abnormally high or low. The results of this data mining exercise would be enhanced if that information can be provided in the future.

The objective of the modelling was to assess the significance of detected anomalies, therefore the input IF data was limited to that associated with PA values of greater than 0.90, i.e. that which would have been in alert during the in-service trial. To limit the amount of any very extreme data being input to the modelling process, that associated with IHUMS-detected sensor faults was filtered out using the recorded IHUMS sensor fault flag.

Initial modelling work showed that, due to the large quantity of associated data, the output could be biased by a 'gearbox fit' that was permanently in alert. To prevent this, the data was limited to the last 50 alert points for each gearbox fit.

Different types of 'mixture model' (i.e. cluster model) and anomaly model were built and evaluated. Because they had the greatest numbers of CIs as inputs, the work concentrated on the '8-Indicator Absolute' and 'Trend' anomaly models. Separate data mining models were built using the IFs from these 'Absolute' and 'Trend' anomaly models.

The modelling approach found to be most suitable for the data mining demonstration was a type of anomaly model, built using data from multiple Main and Accessory Gearbox components, that was configured to generate 15 clusters per component. Therefore the models built using data from 17 components contained a total of 255 clusters. The clusters within these models were associated with components, with a component being represented by a subset of the model's clusters. The models were unfiltered, so no outlying data was removed. Bringing multiple components together in one model had the effect of normalising the cluster distances across the 17 components (distance is explained in the next paragraph).

Three cluster statistics were used in the automated significance assessment; distance, entropy and support. Cluster distance is a measure of how far a cluster is from the centre of the modelling space, with extreme IF values being represented by clusters with a high distance. Therefore distance gives a measure of the severity of an anomaly. Cluster entropy is a function of the distribution of gearbox fits contained in the cluster. A cluster containing data from many gearboxes will have a high entropy, and indicate a pattern of IF values that is relatively common, for example one associated with a HUMS instrumentation problem. However a cluster with low entropy would contain data predominantly from a single gearbox. Therefore entropy provides a measure of anomaly uniqueness. The most significant clusters would be expected to be those with a relatively large distance measure and also a low entropy measure.

The final cluster statistic used in the analysis was support, indicating the number of data points included in a cluster. A minimum support value was used to filter out clusters associated with only a small number of data points, which may be associated with data spikes or other signal corruption.

Two different approaches were taken to fusing the outputs from the data mining models created for the automated significance assessment. The first, discussed here, was a rule-based approach to ranking alerts. The second, discussed in Section 5.4, used a reasoning network.

In the rule-based approach, insignificant clusters were first filtered out based on distance and entropy, and the remaining clusters were given both an entropy and a distance rank across each component type. The rankings were then combined, with more weighting given to entropy, to produce an overall model score. The primary gearbox fit in each cluster was identified, and the scores from corresponding clusters in the 'Absolute' and 'Trend' models were multiplied together, with the final ranking being performed on the resulting score.

This rule-based approach produced the overall significance rankings for the anomaly alerts associated with particular gearbox fits shown in Table 4-7 to Table 4-9. The tables show the top 25 rankings for the individual 'Absolute' and 'Trend' models, and also the combined ranking. Items shaded in green had been previously identified in the two six-month trial periods (the cracked MGB bevel pinion case is also highlighted). The first two tables show the cluster name, together with the three

cluster statistics of support, distance and entropy, and the component name and the primary gearbox fit associated with that cluster. The component distance and entropy ranks are the component-specific distance and entropy ranks for the cluster. An overall weighted ranking for the cluster is then calculated, which gives greater emphasis to the entropy ranking than the distance ranking. It also factors gearbox fits that appear in more than one cluster. The final combined ranking shown in Table 4-9 is simply a multiplication of the individual 'Absolute' and 'Trend' model rankings.

The anomaly significance assessment process has included seven of the previously identified cases in the top 15 of the ranking. It should be noted that the assessment was only based on the '8-Indicator' anomaly model results, and therefore does not include previous cases detected by the 'Shaft Order' and 'M6' models. Only one previous case detected by the '8-Indicator' model did not appear high in the rankings. The cracked MGB bevel pinion fault case (gearbox fit 999) is ranked highest of the previously identified cases in Table 4-9, and 6th overall. Obviously relative rankings are affected by the details of the rules used to calculate them however this is still a significant finding. The case is ranked joint 14th in the results from the '8-Indicator Absolute' model, and joint 9th in the results from the 'Trend' model.

The top two cases in the combined ranking highlighted by the data mining exercise were alerts contained in the processed historical data, but that had occurred prior to the start of the first trial period in May 2006. Top of the ranking was an alert on the 2nd epicyclic annulus aft (RH) of G-TIGE (gearbox fit 186, September 2003), that was also highlighted by the trend analysis (see Section 4.3.2). Ranked second was the alert on the LHA oil cooler fan drive of G-TIGS (gearbox fit 859, October 2005). Figure 4-15 and Figure 4-16 show the very clear 'Absolute' and 'Trend' model FS trends associated with this alert. These were caused by clear trends in many HUMS CIs, including SIG_SD (Figure 4-17). The related alert on the LHA left hydraulic drive 81-tooth gear appears 4th in the combined ranking.

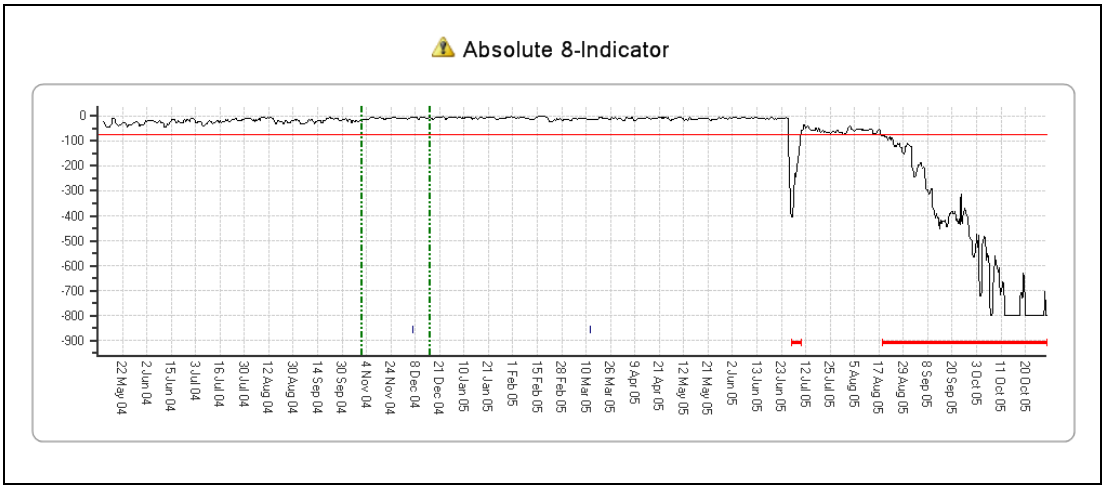

Figure 4-15 G-TIGS LHA oil cooler fan drive – 81A model FS

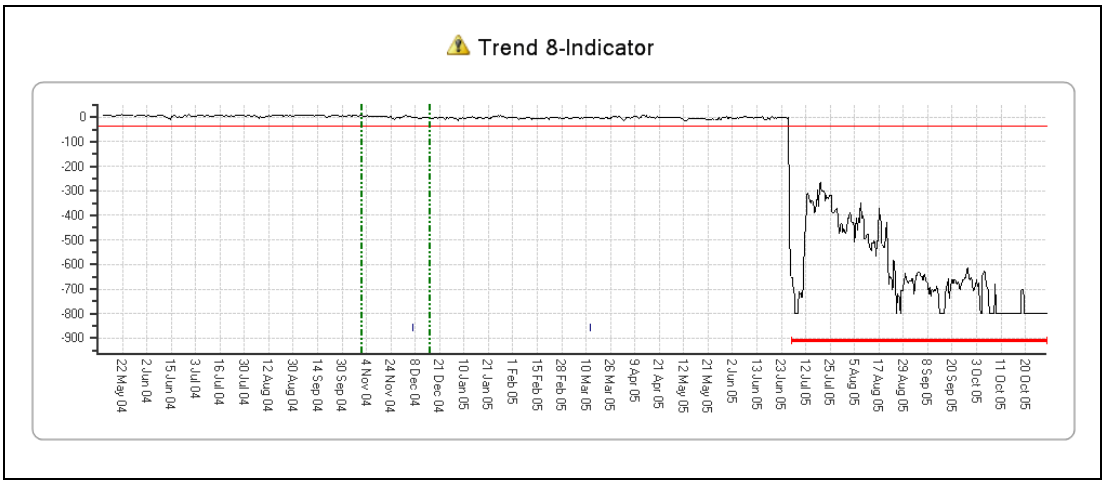

Figure 4-16 G-TIGS LHA oil cooler fan drive – 81T model FS

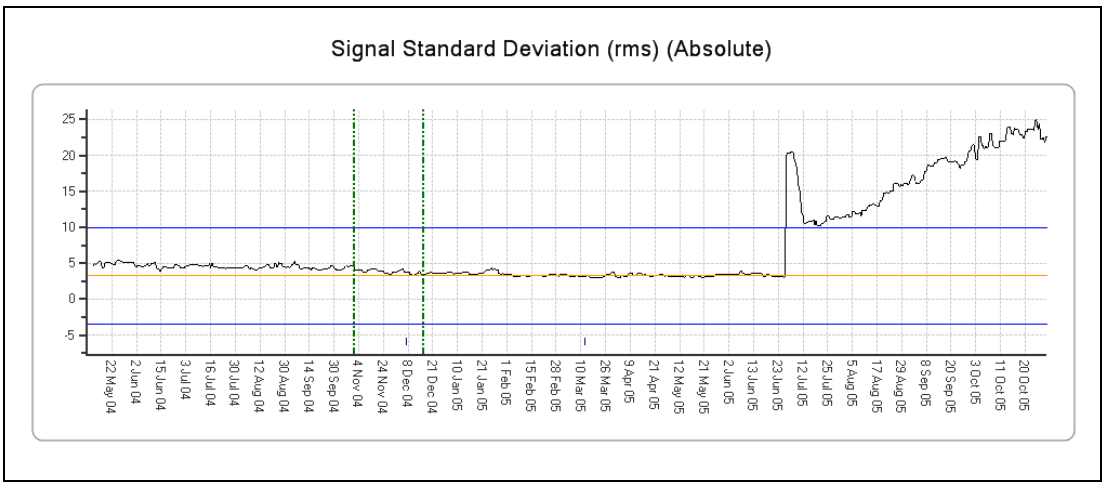

Figure 4-17 G-TIGS LHA oil cooler fan drive – SIG_SD

The results of the anomaly significance assessment showed that this can flag up the most significant anomalies, ensuring that these receive very close attention from the operator. The limitations of the rule-based approach to fusing the data mining model outputs were, however, demonstrated by the fact that, due to the extreme absolute values of the associated data, some cases of instrumentation faults or other suspect data also appeared high in the ranking. An example of this was the 2nd stage sun gear of G-TIGC (gearbox fit 1053), which was ranked 9th overall. This alert was due to an epicyclic stage wiring harness problem (see Section 4.3.2 and reference [3]). Another example is the 2nd epicyclic annulus fwd (RH) of G-TIGF (gearbox fit 1055), where there was a short period of low FS values (Figure 4-18), caused by a step decrease in SIG_SD (Figure 4-19) and increase in ESA_M6 (Figure 4-20). This was ranked 5th overall, but the most probable cause of the step change in the data is an instrumentation problem.

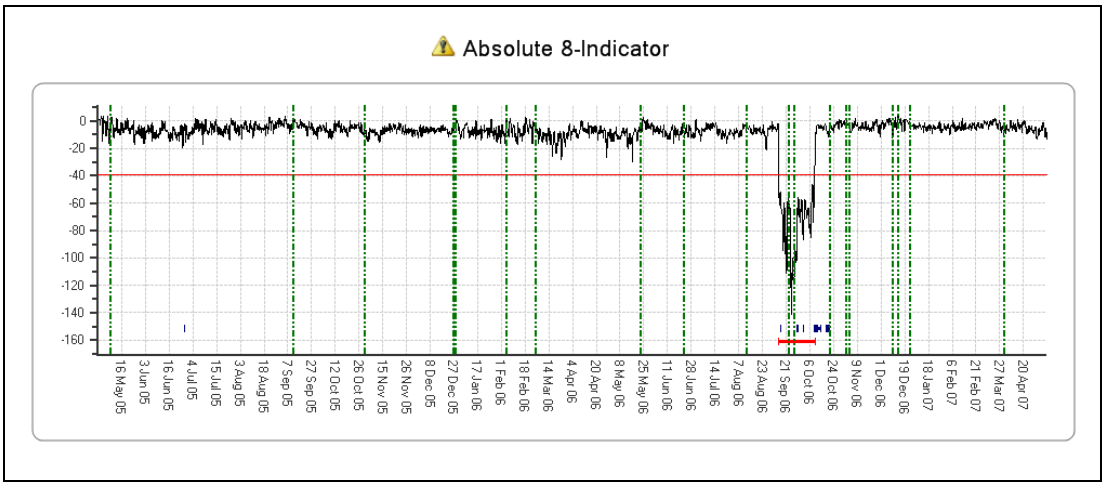

Figure 4-18 Gearbox fit 1055 – G-TIGF 2nd epicyclic annulus forward (RH) – 8IA model FS

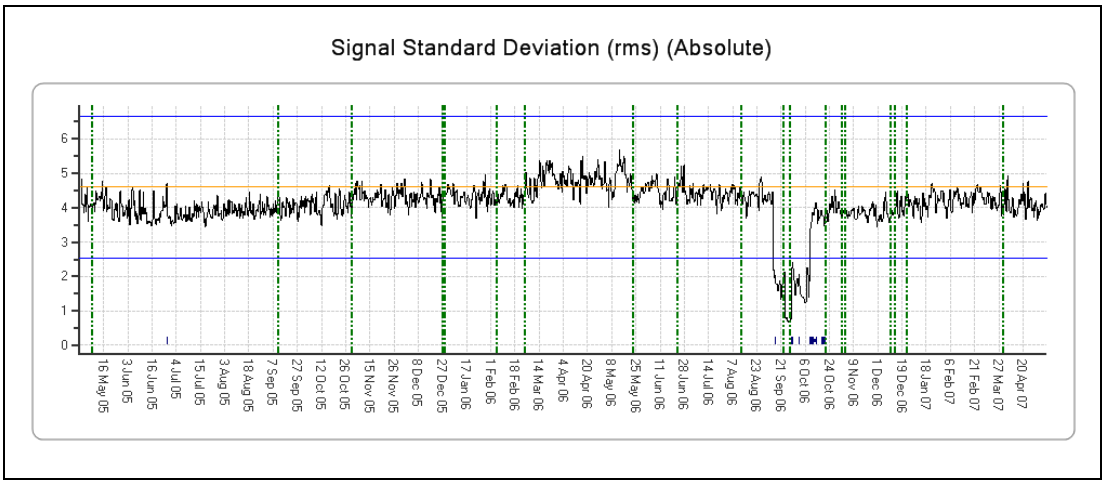

Figure 4-19 Gearbox fit 1055 – G-TIGF 2nd epicyclic annulus forward (RH) – SIG_SD

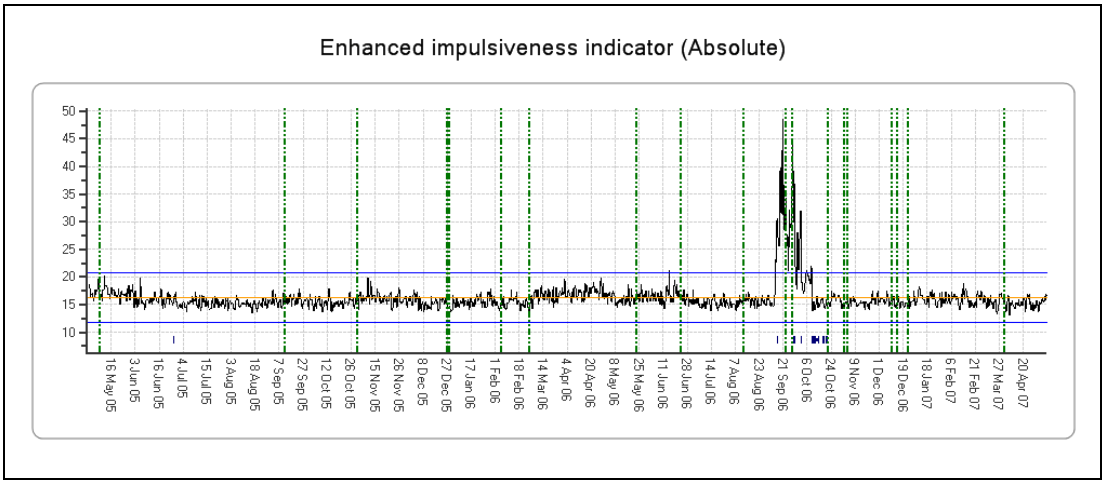

Figure 4-20 Gearbox fit 1055 – G-TIGF 2nd epicyclic annulus forward (RH) – ESA_M6

Table 4-7 Alert ranking for '8-indicator absolute' model

Cluster	Support	Distance	Entropy (Gearbox fit)	Component	Primary Component FitID	Comp Distance Rank	Comp Entropy Rank	Final Weighted Comp Rank
Cluster 158	51.00	59.88	0.23	UpSun	1053	1	1	1
Cluster 193	37.03	80.41	0.08	PortFwdFW	1049	1	1	2
Cluster 187	42.05	129.37	1.19	LHAHydid	1133	1	1	3
Cluster 171	50.00	87.01	0.01	UpAnn6	789	1	1	3
Cluster 227	17.00	43.83	2.37	TailOutput	1097	1	1	3
Cluster 222	20.00	32.33	0.04	PortAftFW	809	1	1	3
Cluster 150	53.00	24.58	1.61	StbdIPT	261	1	1	3
Cluster 242	8.00	23.70	1.71	LHAAltDr	1083	1	1	3
Cluster 240	8.99	23.58	1.26	StbdFwdFW	250	1	1	3
Cluster 176	50.00	74.84	0.02	UpAnn7	1041	2	1	10
Cluster 174	50.00	22.11	0.03	BevelPinion	1172	2	1	11
Cluster 181	49.00	22.91	0.06	LHAHyddr81Tooth	1115	3	1	12
Cluster 215	23.00	55.31	0.91	LHAHyddr81Tooth	1153	1	2	13
Cluster 169	50.00	22.72	0.02	UpAnn5	2999	3	1	13
Cluster 172	50.00	126.12	0.05	UpAnn5	190	1	2	14
Cluster 204	31.00	89.87	0.39	BevelGear	800	1	2	14
Cluster 160	51.00	27.22	0.58	BevelPinion	999	1	2	14
Cluster 239	9.01	45.93	0.07	BevelGear	817	3	1	14
Cluster 149	53.02	24.53	0.26	LHAHyddr47Tooth	850	3	1	14
Cluster 67	101.19	30.43	1.08	LHAHyddr47Tooth	1115	2	2	20
Cluster 229	15.97	71.86	1.01	PortFwdFW	1049	2	2	21
Cluster 152	52.22	20.07	0.77	UpSun	1056	2	2	21
Cluster 168	50.00	42.26	0.03	UpPlan	808	4	1	21
Cluster 173	50.00	24.42	0.01	LHAOilCoolDr	859	4	1	21
Cluster 175	50.00	61.19	0.03	LHAOilCoolDr	1076	2	2	22

Table 4-8 Alert ranking for '8-indicator trend' model

Cluster	Support	Distance	Entropy (Gearbox fit)	Component	Primary Component FitID	Total Rank	Comp Distance Rank	Comp Entropy Rank	Final Weighted Comp Rank
Cluster 164	36.00	66.89	0.02	UpAnn7	186	1	1	1	1
Cluster 131	50.00	150.83	0.07	LHAOilCoolDr	859	3	1	1	2
Cluster 133	50.00	68.82	0.05	LHAHyddr81Tooth	859	2	1	1	2
Cluster 130	50.00	21.96	0.78	StbdFwdFW	261	22	2	1	4
Cluster 183	30.99	20.02	0.25	UpAnn5	1055	21	2	1	4
Cluster 132	50.00	49.54	0.06	LHAHyddr47Tooth	859	4	2	1	5
Cluster 150	45.00	22.53	0.08	UpPlan	183	12	2	1	6
Cluster 148	45.85	17.86	0.65	StbdIPT	1051	26	2	1	7
Cluster 196	27.00	69.73	1.02	LHAHyddr47Tooth	545	16	1	2	9
Cluster 107	62.98	18.52	1.26	TailOutput	319	48	3	1	9
Cluster 141	48.00	16.51	0.02	BevelPinion	999	18	3	1	9
Cluster 127	52.44	17.39	0.24	PortFwdFW	156	23	3	1	10
Cluster 134	50.00	23.01	0.05	UpSun	147	9	3	1	11
Cluster 120	53.94	24.23	1.17	UpPlan	1132	29	1	2	12
Cluster 137	49.00	24.36	0.07	UpAnn6	183	7	3	1	12
Cluster 147	46.00	66.86	0.08	BevelGear	1051	5	1	2	13
Cluster 106	63.35	12.31	1.38	LHAAltDr	504	82	5	1	17
Cluster 194	27.12	17.80	0.89	StbdIPT	261	36	3	2	18
Cluster 152	44.00	21.28	0.06	UpAnn7	183	11	3	2	19
Cluster 202	22.85	28.11	1.22	UpSun	924	27	1	3	20
Cluster 203	22.83	15.92	1.34	LHAHydid	863	55	2	3	21
Cluster 142	48.00	15.64	0.74	LHAHydid	1083	37	4	2	21
Cluster 140	48.40	14.60	0.78	UpAnn5	168	43	4	2	21
Cluster 174	33.66	12.42	1.39	LHAAltDr	1109	79	4	2	21
Cluster 124	53.00	25.69	0.49	UpAnn6	816	13	2	3	22

Table 4-9 Combined alert ranking

Component and Gearbox Fit	Abs Rank	Trend Rank	Total
UpAnn7 186	40	1	40
LHAOilCoolDr 859	21	2	42
StbdIPT 261	3	18	54
LHAHyddr81Tooth 859	30	2	60
UpAnn5 1055	31	4	124
BevelPinion 999	14	9	126
UpPlan 183	45	6	270
BevelGear 800	14	22	308
UpSun 1053	1	500	500
UpPlan 1132	45	12	540
LHAHyddr81Tooth 1153	13	62	806
UpAnn6 816	40	22	880
UpAnn7 183	47	19	893
PortFwdFW 1049	2	500	1000
LHAHyddr47Tooth 1153	28	36	1008
UpPlan 1053	27	42	1134
LHAHydid 1133	3	500	1500
UpAnn6 789	3	500	1500
TailOutput 1097	3	500	1500
PortAftFW 809	3	500	1500
LHAAltDr 1083	3	500	1500
StbdFwdFW 250	3	500	1500
BevelGear 1050	37	80	2960
LHAOilCoolDr 520	43	69	2967
UpAnn7 1041	10	500	5000

4.5 Data Analysis to Support Case Based Reasoning

An investigation was performed into data analysis to generate information for Case Based Reasoning (CBR). Following an anomaly alert, the objective of CBR is to indicate whether there have been previous cases for which data behaviour was similar to the current case. Knowing the cause of any such previous alerts can indicate a possible cause for the current alert. While component faults are rare, HUMS instrumentation problems are more common and it should be possible to identify related cases of instrumentation problems from the similar data characteristics.

The objective of this element of the data mining task was to investigate techniques for identifying similar cases in the database. Two approaches to case based analysis were investigated, the first was based on unfiltered anomaly models and the second used a rule-based method.

4.5.1 Anomaly Modelling Approach

Two types of model were investigated. Both utilised the IFs output by the '8-Indicator Absolute' models. Separate unfiltered anomaly models were built for six different components using the IFs generated from the last 50 alert points for each gearbox fit.

The number of clusters was initially varied, but limiting this to one for each gearbox fit and period of alert was found to be necessary for good results. Once the models were trained, the IFs from all the data in alert were analysed and the models were used to predict the best matching cluster associated with a different gearbox fit. Typically, a few cases dominated the predictions. This was normally due to data spikes that generated large clusters. However, in a number of cases, some clear matching patterns in the IF data were identified

Some example results are presented for the 2nd epicyclic annulus aft (RH) (sensor 7) model. 17 gearbox fits were predicted as similar to fit 1048 (G-BLXR), which had a sensor fault resulting in low frequency noise. All cases exhibited high and noisy FSA_SO1/FSA_SO2 values, suggesting that these may have had a similar sensor problem. Five other gearbox fits (183, 798, 808, 1041, and 1053 which is G-TIGC) were also predicted as having some similar anomalous data characteristics. These are the five red and green clusters in Figure 4-21, which shows the locations of the model clusters on the axes of the IFs for SIG-SD and SIG_PP. Although there are variations in the locations and orientations of the five clusters, on both axes they are all well separated from the blue clusters representing the remainder of the data. All cases had high SIG_SD and ESA_SD (Figure 4-22) values that were believed to be associated with a wiring harness problem.

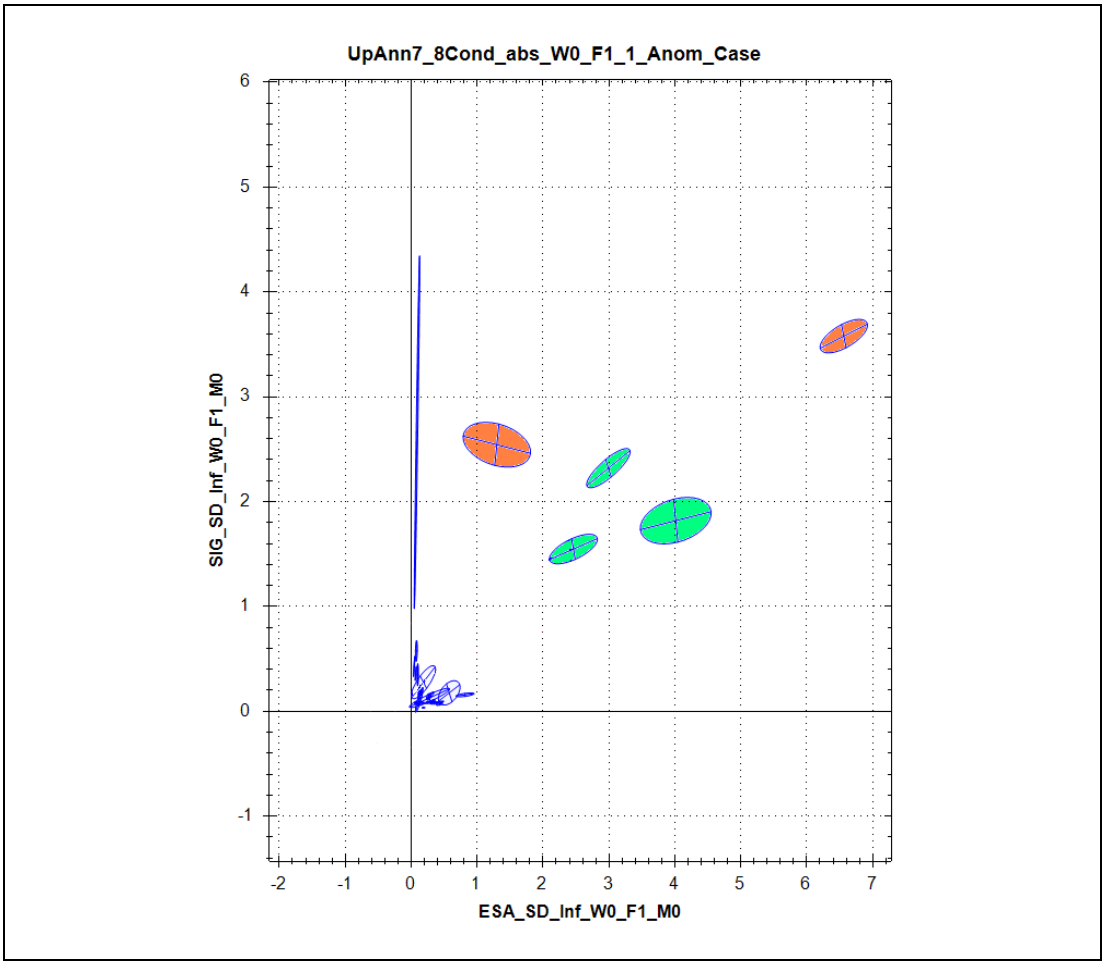

Figure 4-21 2nd epicyclic annulus aft (RH) model – SIG_SD IF vs ESA_SD IF

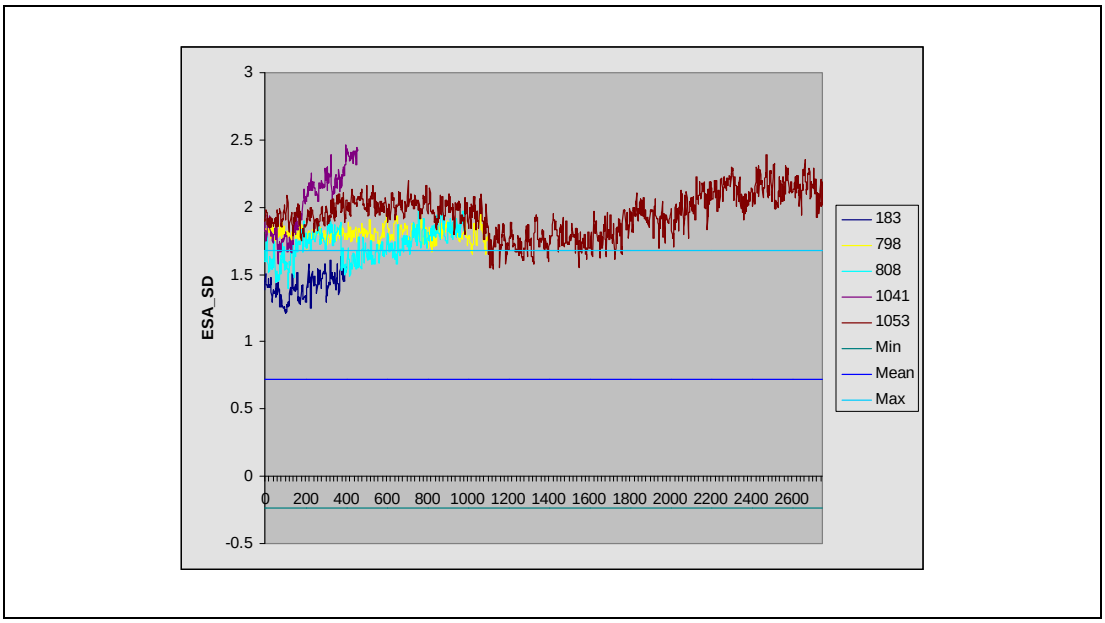

Figure 4-22 2nd epicyclic annulus aft (RH) – ESA_SD

An alternative analysis used the 255 cluster anomaly model created previously for the anomaly significance assessment. Clusters with high entropy were identified (i.e. clusters containing data from several different gearbox fits), and then the similar gearbox fits within these clusters were noted. This produced comparable results to those presented above.

The case based analysis successfully identified similar cases of anomaly alerts that could be used to assist the diagnosis of different types of HUMS instrumentation problem. However, the work showed some limitations with the approach. The models use IFs, and these cannot indicate the directionality of a change in CI values. Therefore a case with a very high CI value could be classed as similar to one with a very low value. The analysis needs to be extended to include IF directionality information. For the unfiltered component anomaly models there is a need to control the cases that predictions are performed against, as clusters having a high variance as a result of alerts caused by noisy data can act as a catch-all, and therefore bias results. For the unfiltered 255 cluster multiple component anomaly model, using a fixed number of clusters per component type does not guarantee similar cases being grouped together, and sometimes non-similar cases may be grouped depending on the data and modelling resources.

4.5.2 Rule-Based Approach

For comparison, a second approach to case-based analysis was investigated, using a rule-based method.

An example that is to be included in the set of known fault cases is identified and, for the component data in alert, the mean and standard deviation of each CI included in the anomaly model are calculated. When a new case triggering alerts is detected, again the mean and standard deviation of each CI are calculated for the duration of the alert. Deltas in the means and standard deviations from the current example and the known fault cases are calculated and normalised. These are then summed and finally factored by the number of deltas exceeding a defined threshold. The output is a score representing the closeness of fit of the new example to the different known fault cases, and scores can be ranked to identify the closest fit.

To demonstrate the method, means and standard deviations were calculated for the CI data from the 2nd epicyclic annulus aft (RH) (sensor 7) for all the different gearbox fits triggering anomaly model alerts. The rule-based approach was then used to rank the gearbox fits on the basis of similarity to 1053 (G-TIGC). The results are presented in Table 4-10. The top four matching gearbox fits are 183, 798, 808, and 1041, which is the same group predicted by the anomaly model in Section 4.5.1.

The primary disadvantage of this rule-based approach is that it uses the HUMS CIs, and so can only be applied to cases from the same component. However it could be extended to utilise the anomaly model IFs if some directional information was also provided to identify whether an IF was identifying an abnormally high or low HUMS CI value.

Table 4-10 Ranking of similar cases for 2nd epicyclic annulus aft (RH)

Component Fit	Delta SIG_SD	Delta ESA_SD	Delta SIG_PP	Delta ESA_PP	Delta FSA_SO1	Delta FSA_SON	Delta FSA_GE2	Delta FSA_MS	Total
1053	0	0	0	0	0	0	0	0	0
798	0.077	0.106	0.115	0.079	0.005	0.001	0.153	0.141	3.425
183	0.227	0.348	0.145	0.348	0.000	0.011	0.066	0.133	3.460
808	0.073	0.157	0.102	0.182	0.032	0.018	0.124	0.052	3.878
1041	0.199	0.041	0.230	0.042	0.009	0.027	0.067	0.122	7.180
190	0.680	0.358	0.669	0.277	0.018	0.023	0.004	0.076	16.986
2999	0.459	0.353	0.469	0.211	0.010	0.055	0.251	0.201	18.883
243	0.751	0.359	0.761	0.185	0.007	0.113	0.040	0.137	20.640
177	0.525	0.425	0.425	0.371	0.006	0.065	0.029	0.093	22.285
146	0.664	0.381	0.596	0.304	0.057	0.037	0.005	0.059	23.264
197	0.172	0.845	0.379	0.856	0.022	0.103	0.019	0.006	23.815

5 Automated Reasoning Demonstration

The purpose of this task was to demonstrate an automated reasoning capability, performing secondary processing of anomaly detection alerts to help the operator determine the most appropriate action. The reasoning technology is primarily based on 'Bayesian' probabilistic networks, which provide a widely recognised capability for reasoning in conditions of uncertainty (i.e. with limited or possibly conflicting information). Once the anomaly detector has flagged abnormal data, reasoning networks can then fuse multiple features of the anomalous data, together with other information, to automatically indicate the anomaly's significance and possible cause.

A range of potential reasoning inputs are available. These include outputs from the anomaly detection processing such as anomaly model alerts, PA values, FS trend information, and the IFs. They can also include outputs from the data mining tasks, such as measures indicating the significance of alert data, identifying significant trends, and identifying similar cases. Other inputs are aircraft configuration data and existing diagnostic knowledge. The output of the reasoning process should be a summary of key information that can assist an operator determine the appropriate response to an anomaly alert. This can include an anomaly significance assessment, evidence of an instrumentation problem or a maintenance-induced alert, diagnostic information from the IFs, and the identification of any similar previous cases.

GE Aviation's ProDAPS reasoning tool (described in Section 5.1) was used to configure demonstration reasoning networks, with real data from the CAA in-service trial being used to test these. The task investigated a number of issues, such as; what reasoning capability is required (e.g. what are the desired outputs); how best to address the problem; what inputs to use; how best to design and construct reasoning networks to ensure robustness and supportability; and how would a reasoning capability perform in practice. The reasoning task must carefully consider what type of information should be presented to an operator, as it can be dangerous to provide leading information that could result in an incorrect conclusion.

A set of probability tables must be configured for each reasoning network, and the larger the network (in terms of the number of node associations), the larger and more complex these tables become. Therefore the reasoning approach adopted was based on a concept of 'multiple independent experts'. These are independent targeted networks looking for either evidence of a particular potential cause of patterns identified in the HUMS data (e.g. a maintenance action or a HUMS instrumentation problem), or providing a particular piece of information (e.g. the significance of an anomaly alert). This targeted approach enables network size to be limited and benefits robustness and supportability.

Ideally statistical data from in-service experience would be used to calculate probabilities for populating the probability tables of networks performing diagnostic functions. However, this is often not available, and it is necessary to use approaches to probability assessments based on engineering knowledge, with testing to check that networks behave as expected. Initially configured probability tables can then be updated from experience, which is a key advantage of Bayesian networks over rule-based approaches. Another benefit of the probabilistic reasoning models is that these are compatible with the anomaly models, therefore anomaly model predictions can be directly used by the reasoning networks. This means that sophisticated reasoning models can be constructed from much simpler and well understood sub-models, relating to different component anomaly models.

The reasoning demonstration has focussed on data fusion and fault diagnosis. The data fusion utilised physical relationships between HUMS sensors and component

analyses, generated diagnostic summary information, and provided enhanced anomaly significance assessment capabilities. The fault diagnosis element demonstrated an ability to automatically identify explainable patterns due to maintenance actions and instrumentation faults, and explored the application of Case Based Reasoning.

5.1 ProDAPS Reasoning Tool

The ProDAPS reasoning tool is a state-of-the-art application programming interface to create, manage and update probabilistic (Bayesian) reasoning networks. Having knowledge and control of the software source code enables GE Aviation to control the detailed implementation of reasoning networks, and adapt the implementation to meet specialist requirements. This control is crucial to address issues associated with scaling up to real world health management data sets, and to develop the network auto-generation techniques required for creating large reasoning networks from existing design data. It is also necessary to be able to generate efficient code for embedded reasoning within an on-aircraft system (for example to reduce memory requirements).

The tool can be used to create probabilistic networks for diagnostic and prognostic reasoning. These networks typically model statistical relationships between nodes representing system components, failure modes, and items of health management information. The probabilistic networks offer advantages that are essential for in-field applications. They provide a theoretically sound framework for modelling uncertainty, and can handle missing or conflicting data. They can also be updated with new knowledge to allow reasoning capabilities to develop as experience is gained.

Probabilistic networks are a special case of a more general class of probabilistic models that includes mixture models, Markov models, factor models, etc. This means that there is a general unifying framework for all the ProDAPS tools, which is important for information integration between the different tools. For example, a capability now exists to perform reasoning directly on anomaly model outputs. The across-model fusion of anomaly information is naturally handled in the probabilistic network inference engine. The anomaly model has a direct formal mapping into a probabilistic (Bayesian) network. It is therefore possible to use the ProDAPS data modelling tool to learn the anomaly model, and then use the ProDAPS reasoning tool to output the same FS (as produced by the anomaly detection component) for a given acquisition. This ability means that the outputs from multiple models can be fused, and other information can be fused with anomaly FSs. This higher level fusion for reasoning further enhances the diagnostic and prognostic information that can provided. For the reasoning demonstration described in this report, anomaly models were not directly implemented in reasoning networks, however these networks did process anomaly model outputs such as FS and IF values.

Examples of ProDAPS reasoning tool displays are shown in Figure 5-1. Networks can be displayed graphically as a set of interconnected nodes. These nodes can represent items such as components, faults, or health information. Networks can also be displayed in tabular form, for example one table may show the input health data, with another listing the faults than can be diagnosed, and others showing assemblies and components on which a fault has occurred.

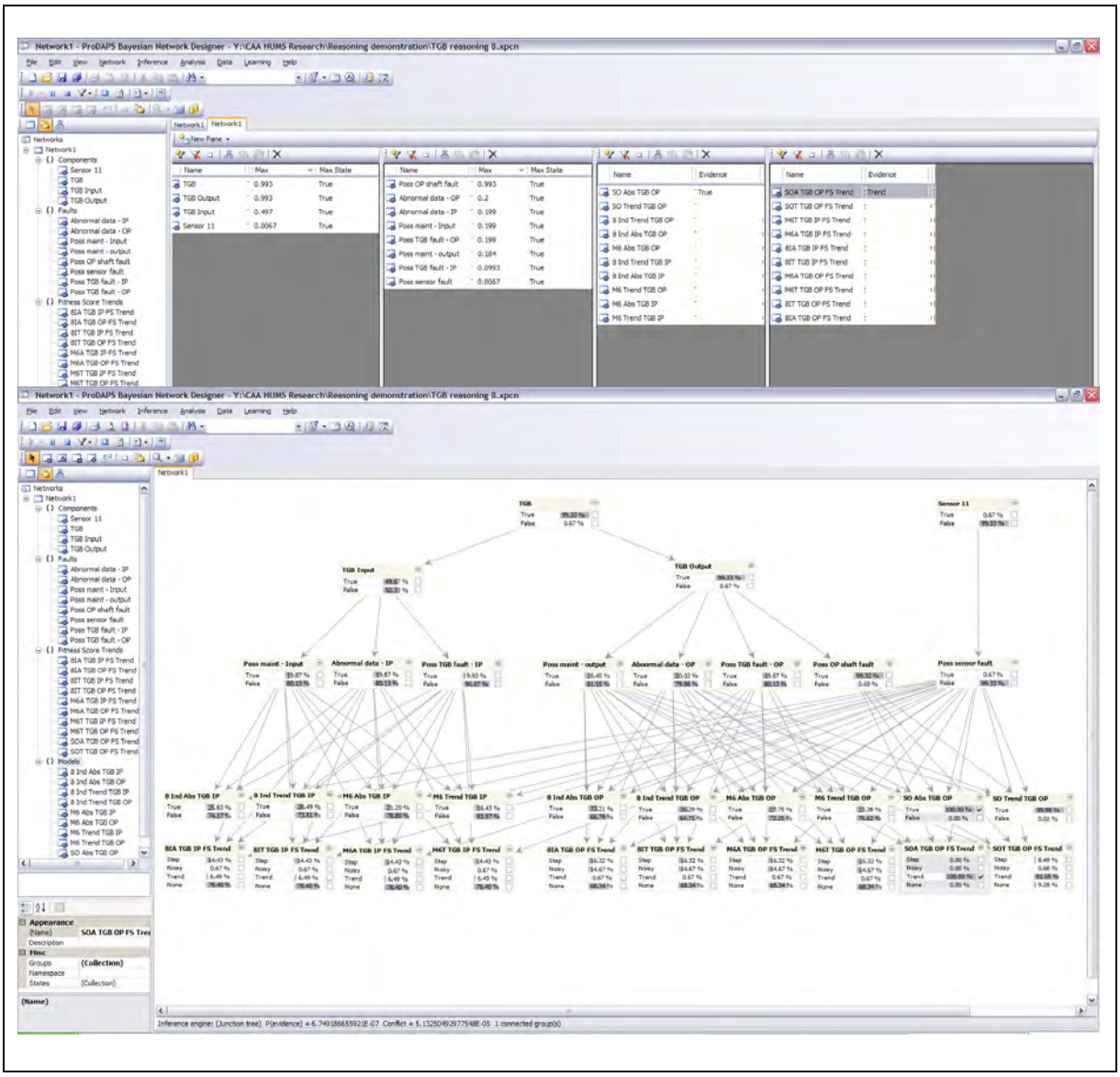

Figure 5-1 Example ProDAPS reasoning tool displays

5.2 Data Fusion by Physical Modelling in Reasoning Networks

Reasoning networks provide an effective way of fusing items of HUMS data from different sensors, or from different components monitored by the same sensor, to increase the confidence in an output. This is a relatively simple form of data fusion, where a reasoning network models the physical relationship between HUMS sensors and monitored components.

To demonstrate the concept an example epicyclic stage reasoning network was created, fusing data from multiple sensors and gears to give increased confidence of both a gear fault and a sensor fault as multiple items of evidence are fused. For the IHUMS equipped AS332L, three accelerometers are used to monitor the 1st and 2nd stage annulus gears. The input nodes used in this reasoning network were 'discrete' nodes, for example being set to mutually exclusive states such as 'true' or 'false'.

Part of the epicyclic stage network is shown in Figure 5-2. The part of the network that is visible is modelling five component analyses – upper and lower annulus monitored from sensor 5, upper and lower annulus monitored from sensor 6, and upper annulus monitored from sensor 7. The outputs from the visible part of the

network are the probability of occurrence of faults on the upper annulus, lower annulus, both annulus gears, sensor 5 and sensor 6.

Figure 5-2 to Figure 5-4 show the detection of a fault on the upper annulus gear (circled in red). In Figure 5-2 the first piece of evidence has been entered – an alert has been generated on the data from sensor 5 (the analysis is labelled 'UpAnn5'), and other reasoning networks have determined it is not related to a sensor fault or maintenance action (see Section 5.5), therefore the node state is set to '100% other'. With this single piece of evidence, the network indicates a 55% probability of an upper annulus gear fault. Based on the sensor 5 alert, the network also has an expectation that there is a 49% probability that gear fault-related alerts will also be triggered on sensors 6 and 7 (the analyses are labelled 'UpAnn6' and 'UpAnn7', with the fault-related alert probability being indicated by the 'other' node states). The evidence from these two sensors can then be entered. When it is confirmed that sensor 6 is also in alert (Figure 5-3), the probability of an upper annulus gear fault increases to 96%, and the probability that sensor 7 is also in alert increases to 82%. If sensor 7 actually is in alert, the network diagnoses an upper annulus gear fault with a 100% probability (i.e. confidence, Figure 5-4). However, if sensor 7 is not in alert, the probability drops to 75%.

Figure 5-5 and Figure 5-6 show an equivalent scenario for diagnosing a sensor fault (sensor 5, again circled in red), based on the results from two independent component analyses. When an alert is triggered on the upper annulus gear analysis from sensor 5, and an independent reasoning network has diagnosed a sensor problem (see Section 5.5), the network gives a probability of an actual sensor fault of 67% (Figure 5-5). If a similar alert is also triggered on the lower annulus gear analysis from sensor 5, this probability increases to 97% (Figure 5-6).

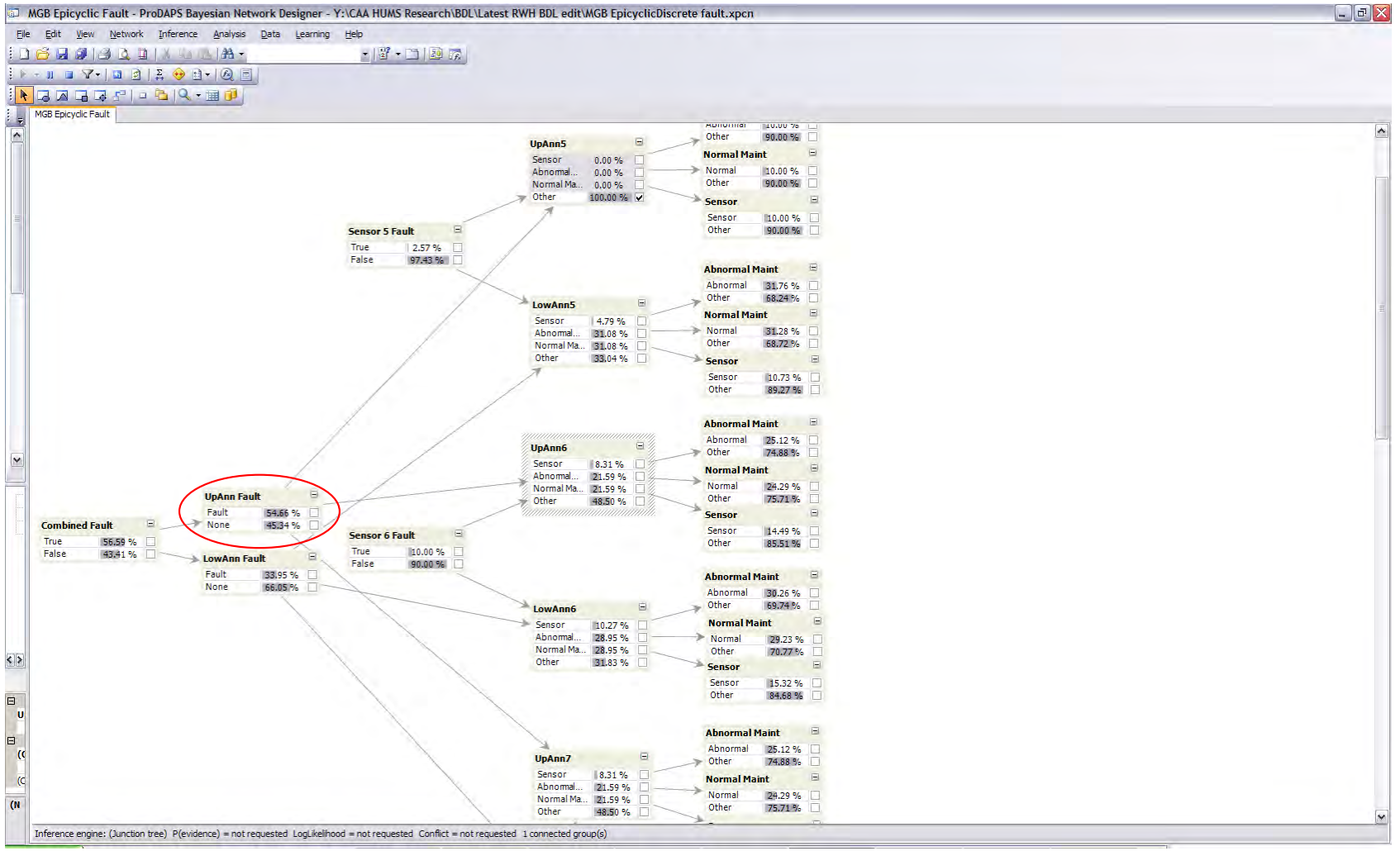

Figure 5-2 Epicyclic stage reasoning network – one item of gear fault evidence (not all nodes shown)

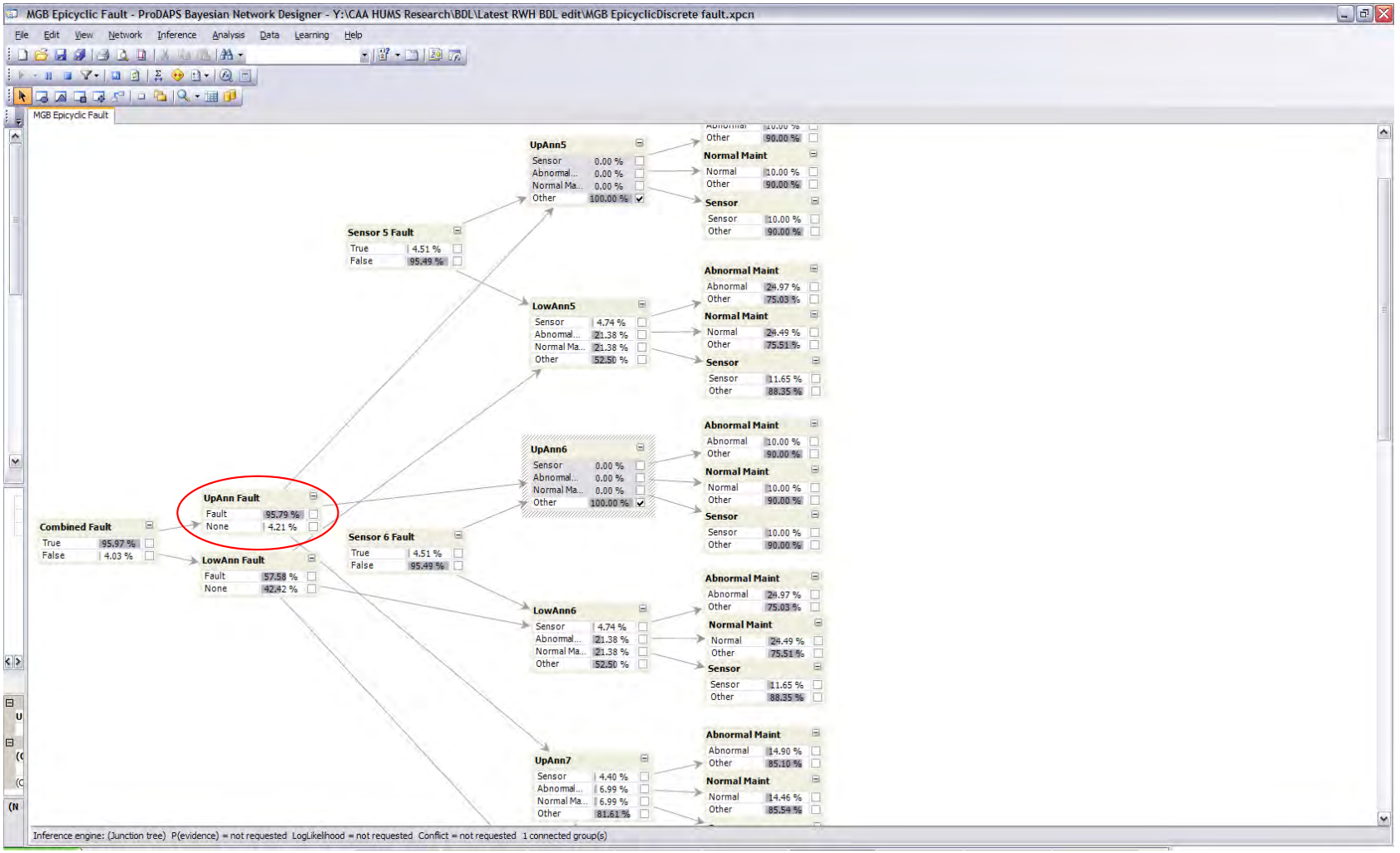

Figure 5-3 Epicyclic stage reasoning network – two items of gear fault evidence

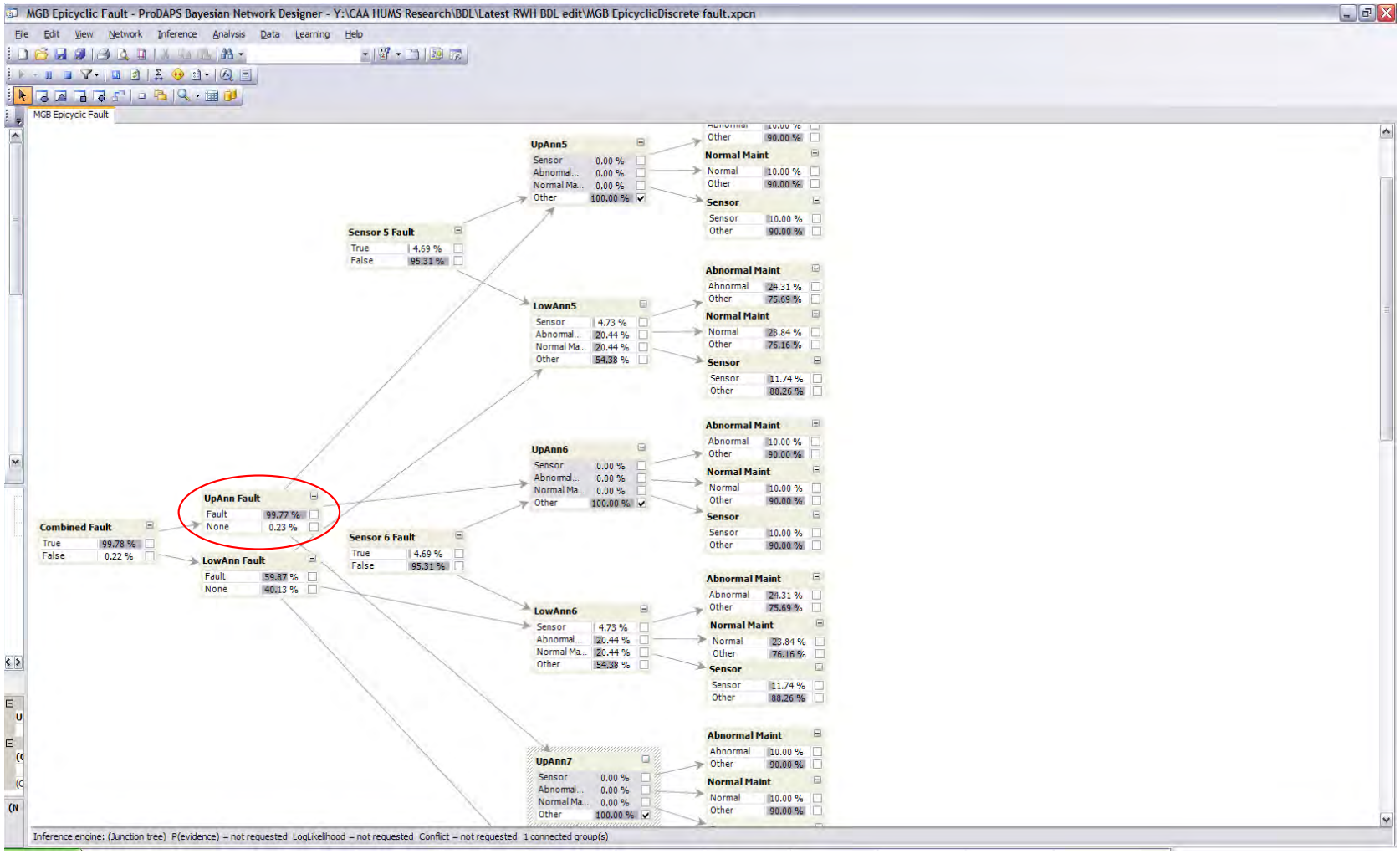

Figure 5-4 Epicyclic stage reasoning network – three items of gear fault evidence

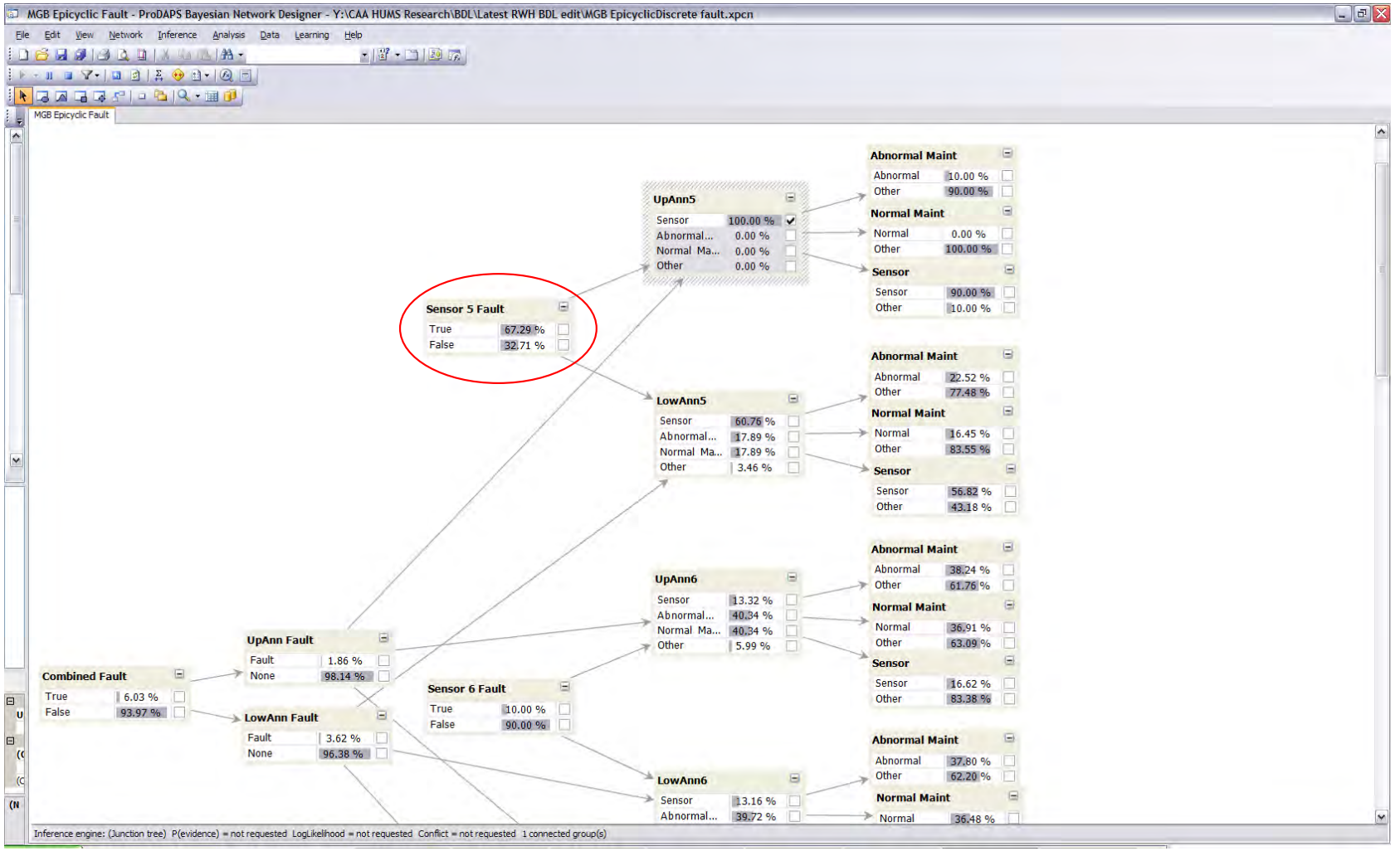

Figure 5-5 Epicyclic stage reasoning network – one item of sensor fault evidence

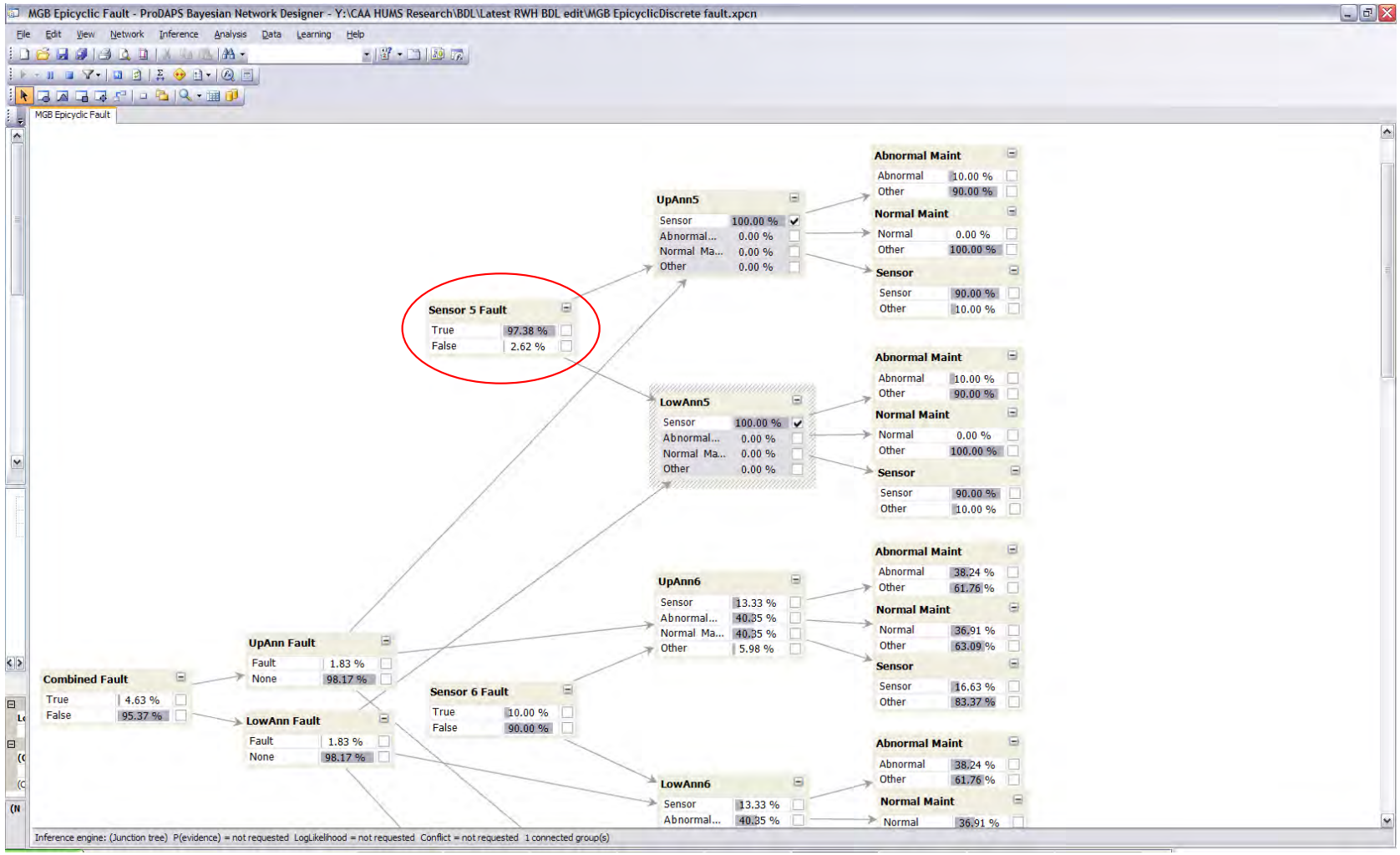

Figure 5-6 Epicyclic stage reasoning network – two items of sensor fault evidence

An investigation was performed into the practicality of a more generic use of this type of physical modelling of the aircraft configuration for fusing HUMS data in reasoning networks. This clearly has value, however the modelling would need to be aircraft and HUM system specific, and Figure 5-7 shows that, just for the IHUMS equipped AS332L, five different component/sensor combinations would need to be modelled. This, combined with the fact that nearly all components are monitored using a single sensor, resulted in a conclusion that there is limited scope for the use of physical modelling in a generic reasoning process.

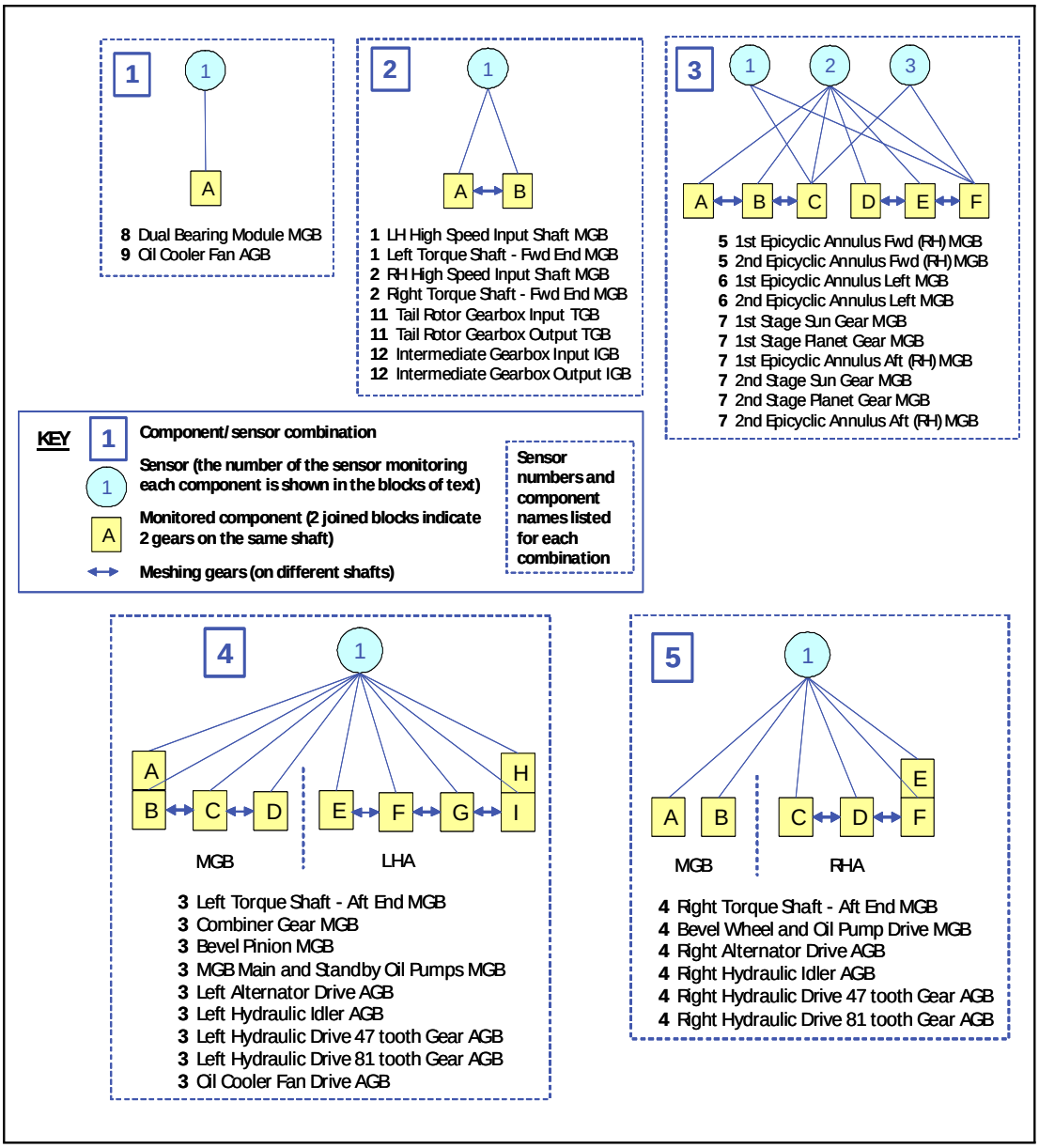

Figure 5-7 AS332L IHUMS component/sensor combinations (the number of the sensor monitoring each component is shown in the blocks of text)

The in-service trial experience described in references [3] and [5] did highlight areas where this type of automated data fusion could be applied. The first was in diagnosing the many senor faults causing anomaly alerts, where the fault affected the results from multiple component analyses performed from that sensor. There were also two examples of gear fault detections where this type of fusion would be beneficial. The first was in the diagnosis of a possible upper annulus gear fault that caused trends in the data from all three sensors monitoring this gear. The second was for certain LHA

and RHA components, where there was an equivalent shaft in the other accessory gearbox rotating at exactly the same speed. As the two gearboxes were monitored from different sensors, confirming evidence of a fault was sought by checking for the mirroring of data trends on the equivalent component in the other gearbox.

5.3 Diagnostic Summary Information

The IFs output from the anomaly models provide diagnostic information on the HUMS CIs that are driving an anomaly alert. With two variants ('Absolute' and 'Trend') of three different types of anomaly model ('8-Indicator', two-indicator 'M6', and two-indicator 'Shaft Order') being used in the CAA trial, a total of 24 different IFs were generated. Therefore, to access the diagnostic information provided by the IFs, an operator must view up to 24 different charts and then manually assimilate this information. To enhance system usability, there is a desire to minimise the need for operators to view and interpret multiple charts when investigating an anomaly alert.

A diagnostic network was created, using the anomaly model IFs as inputs, to generate diagnostic summary information for an alert in one operator report table, as described in Section 3. This simplifies the operator's data interpretation task, provides more meaningful descriptive information, and could be used in the future to drive specific diagnostic outputs.

The network uses 'continuous' nodes so that the input evidence is actual IF values. A continuous node represents a range of values defined by a Normal distribution with parameters for mean and variance. A continuous node can have discrete or continuous parent nodes. A mean and variance is defined for each configuration of the discrete parents. A regression coefficient is specified for each continuous parent. Suppose node Y has parent node X and both nodes are continuous. This is an example of the familiar linear regression function where Y is predicted from X. Furthermore, the model can characterise the variation of the dependent variable (Y) around the regression function. Although a continuous node is a Normal distribution, it is possible to approximate any type of distribution using a mixture of continuous nodes. Continuous nodes enable the construction of a broad range of statistical models that include Principal Component Analysis, Mixture models, Markov models, etc. Reference [6] provides an introduction to these model types.

The network converts the IF / HUMS CI names into more meaningful descriptors and, where two IFs provide related diagnostic information, combines these into a single diagnostic descriptor. Therefore, for the '8-Indicator' model, the eight IFs SIG_SD, SIG_PP, ESA_SD, ESA_PP, FSA_SO1, FSA_SO2, FSA_MS and FSA_GE are converted into the four descriptors 'Signal Energy', 'Base Energy, 'Shaft Order' and 'Gear Mesh'.

All the nodes in the diagnostic network are shown in Figure 5-8. The top two rows are the network outputs for the 'Shaft Order' (left), '8-Indicator' (centre), and 'M6' (right) models. The top row contains the outputs from the absolute versions of these models, and the second row the outputs from the trend versions. The next block of eight rows is the input and modelling nodes for the '8-Indicator' 'Absolute' and 'Trend' models. The bottom two blocks of two rows are the input and modelling nodes for the 'M6' and 'Shaft Order' 'Absolute' and 'Trend' models respectively.

The key input nodes for the '8-Indicator' anomaly model are shown in Figure 5-9 (not all nodes are shown). The other nodes associated with these are automatically generated modelling nodes (representing data distributions associated with the continuous nodes) that would be hidden in any user display. The nodes on the left of Figure 5-9 are associated with the 'Absolute' model (i.e. 8IA), and those on the right with the 'Trend' model (i.e. 8IT). The outer columns of nodes show the actual IF

values that have been entered. The inner columns show these mapped onto the continuous nodes, with IF values now converted into a probability that the CI associated with the IF is contributing to an anomaly indication.

The output nodes for the '8-Indicator' 'Absolute' and 'Trend' anomaly models are shown in Figure 5-10. No evidence has been entered at this stage. The following two figures show the states of the output nodes for two previously documented example fault cases.

In Figure 5-11, anomalous IF data has been entered for the 2nd stage sun gear of G-TIGC (gearbox fit 1053). The associated fault was determined to be an epicyclic stage wiring harness problem (reference [3]). The key IF characteristics summarised by the diagnostic network are: '8-Indicator Absolute' model – 'Signal Energy' 100% anomalous, 'Base Energy' 100% anomalous; '8-Indicator Trend' model – no anomalies. In contrast, Figure 5-12 shows the results of entering the anomalous IF data for the cracked MGB bevel pinion (gearbox fit 999). The diagnostic network has summarised key features as: '8-Indicator Absolute' model – 'Gear Mesh' 99% anomalous; '8-Indicator Trend' model – 'Base Energy' 98% anomalous, 'Shaft Order' 20% anomalous, and 'Gear Mesh' 92% anomalous.

These examples show how the diagnostic network can summarise data contained in up to 24 different IF charts into summary information giving a clear picture of the HUMS CIs that are driving anomaly alerts that could be presented in a single table.

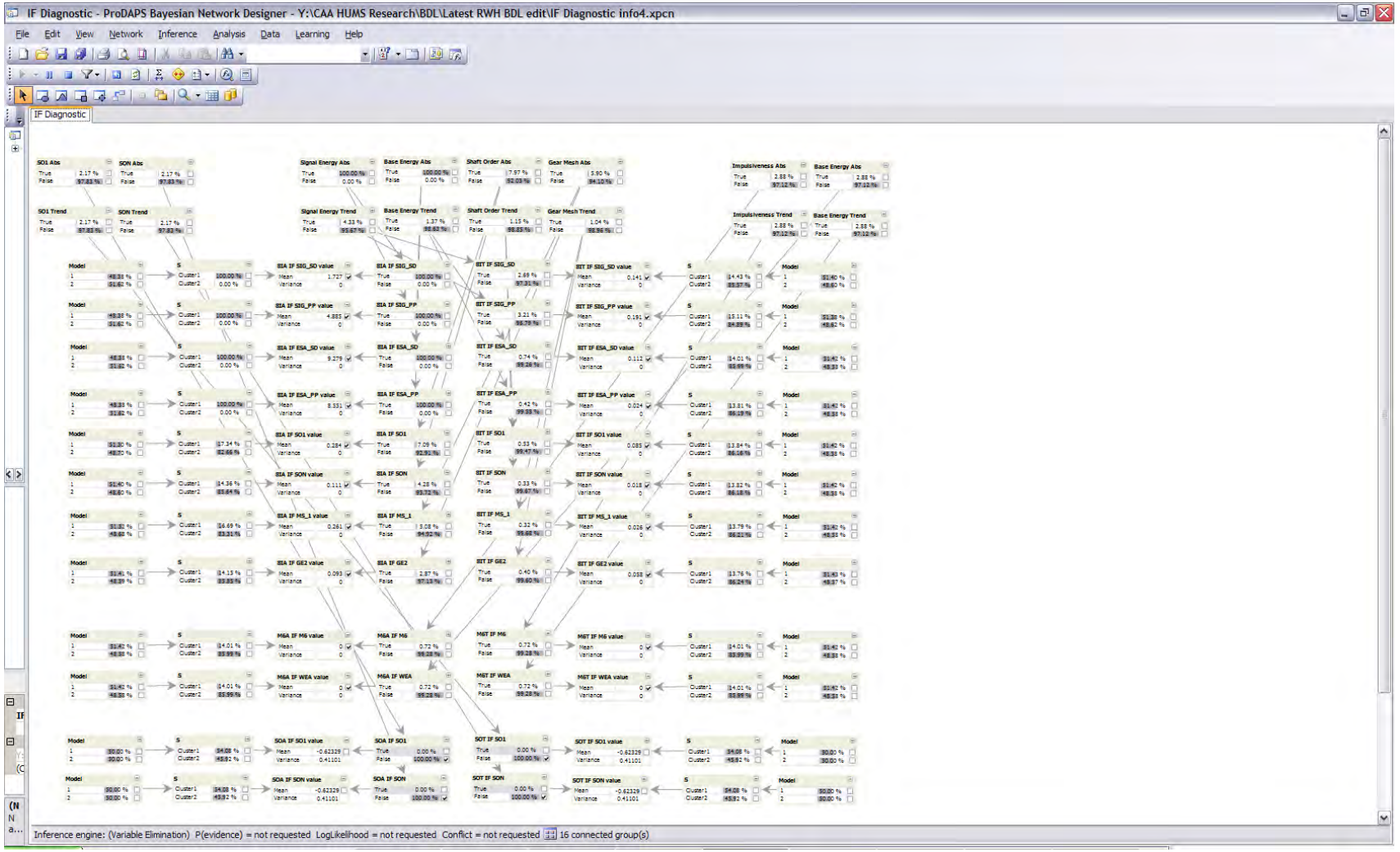

Figure 5-8 Diagnostic summary network – all nodes

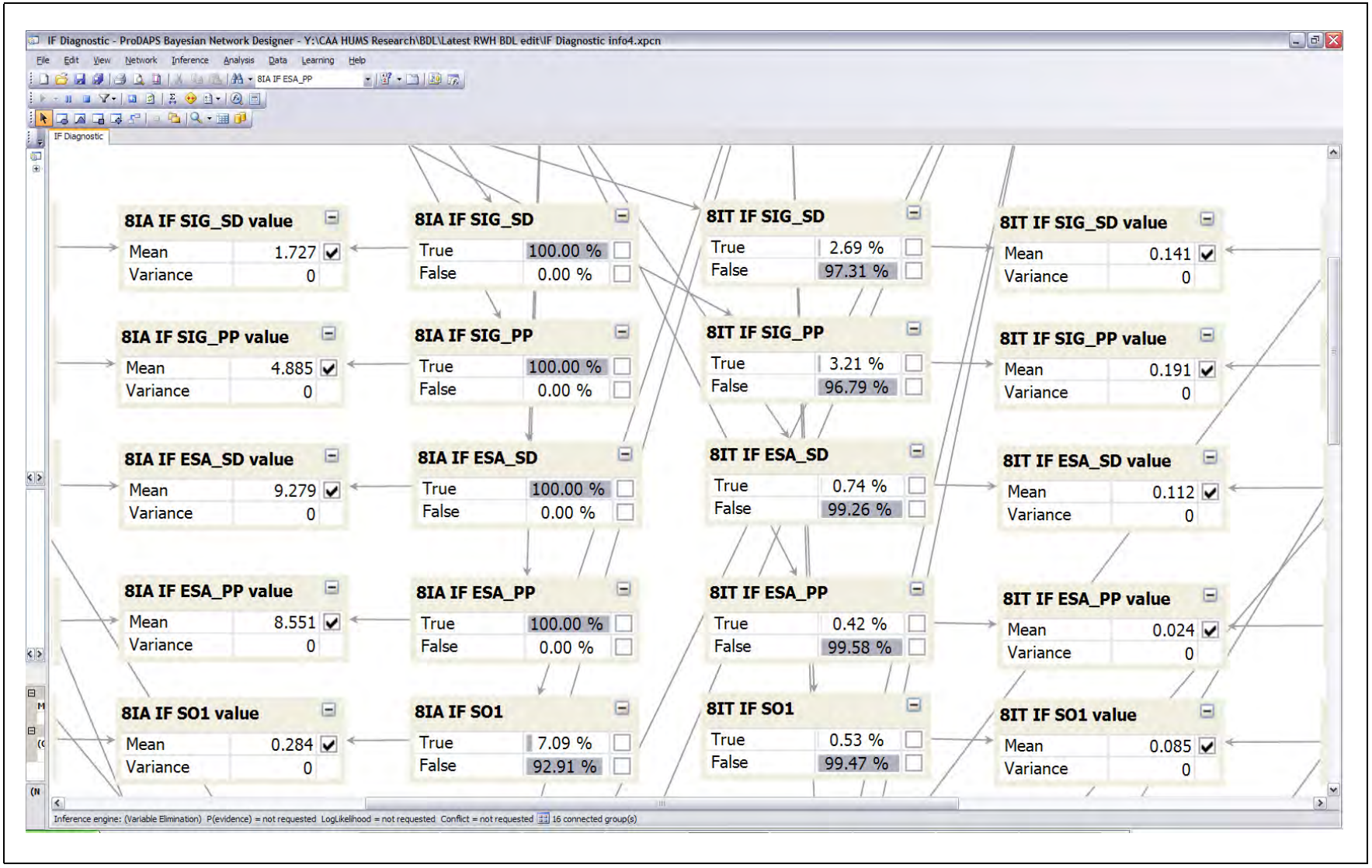

Figure 5-9 Diagnostic summary network – input nodes for 8IA and 8IT models (not all nodes shown)

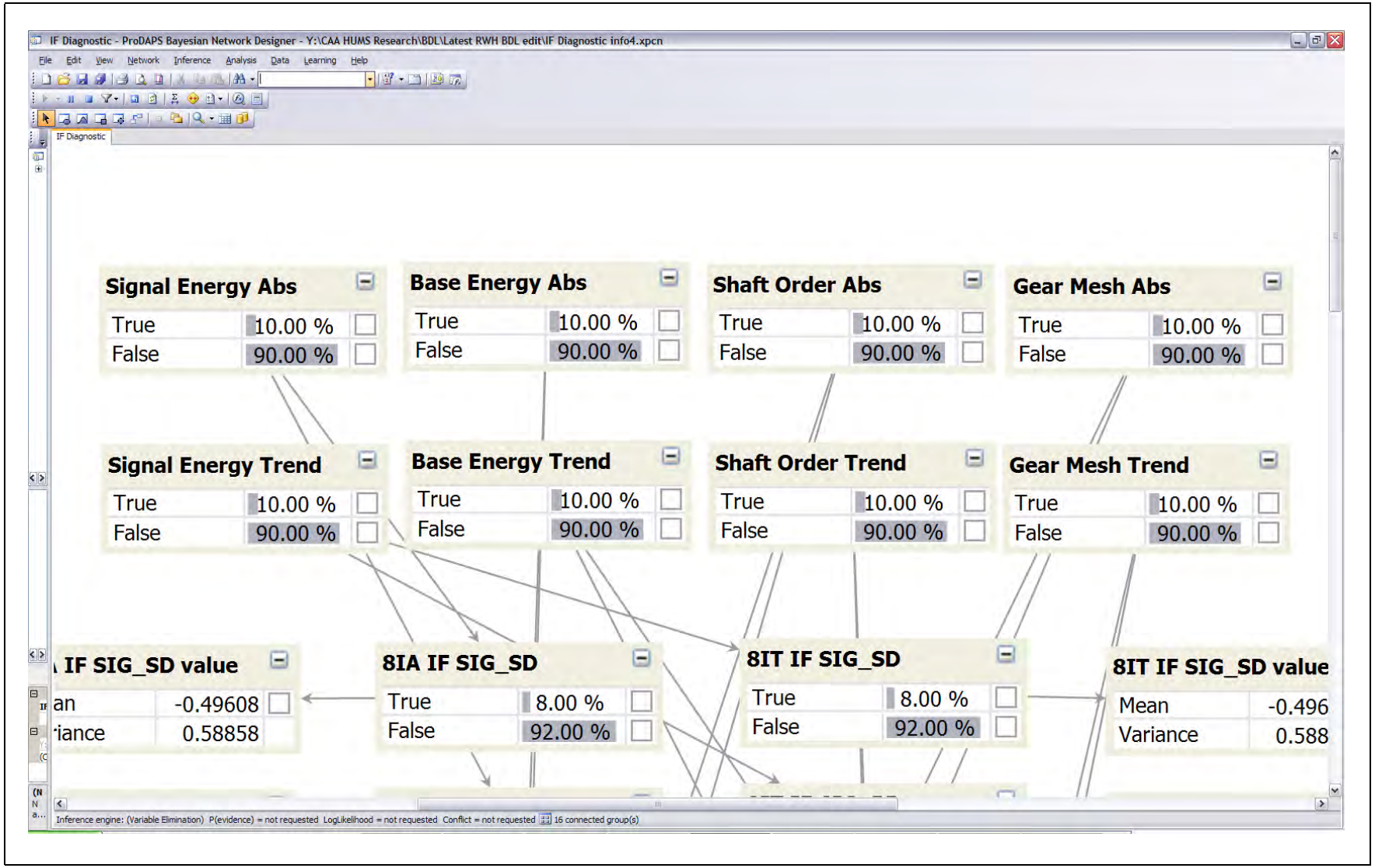

Figure 5-10 Diagnostic summary network – output nodes for 8IA and 8IT models – no evidence entered

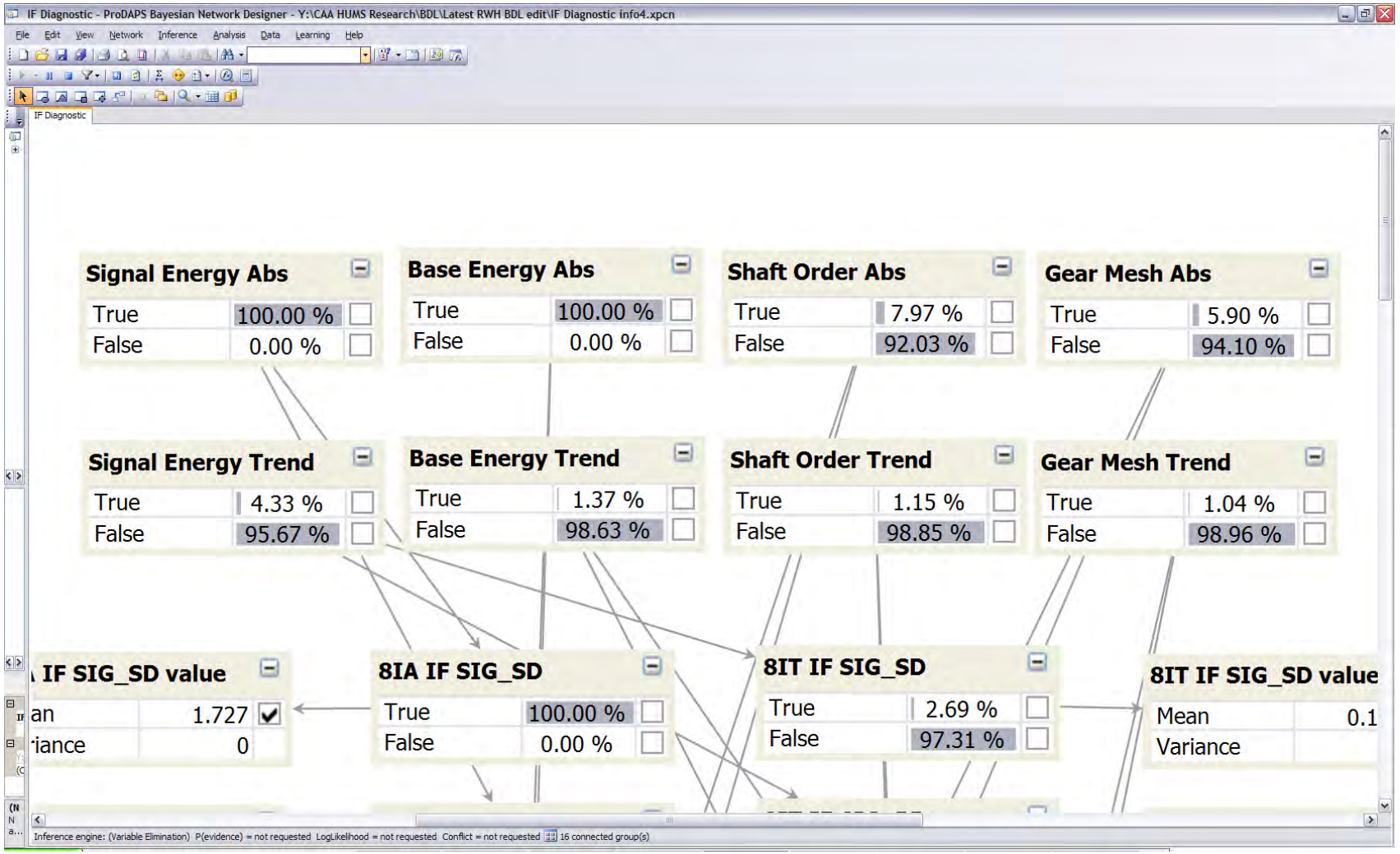

Figure 5-11 Diagnostic summary network – output nodes for 8IA and 8IT models – evidence entered for gearbox fit 1053

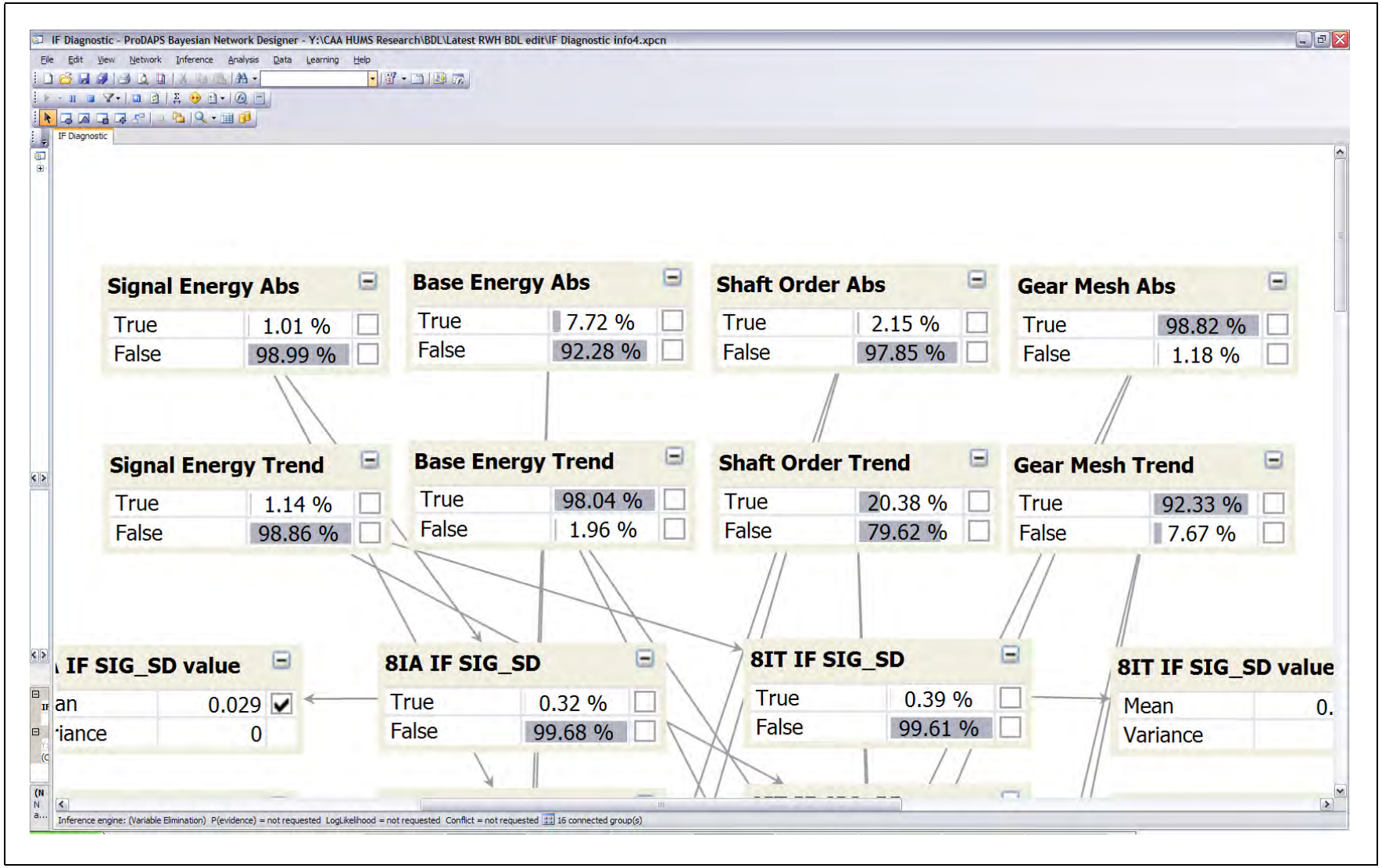

Figure 5-12 Diagnostic summary network – output nodes for 8IA and 8IT models – evidence entered for gearbox fit 999

5.4 Data Fusion for Anomaly Significance Assessment

The next step in the diagnostic process is determining the significance of an anomaly. The data mining task to determine the significance of anomalies in terms of anomaly cluster 'distance' (magnitude) and 'entropy' (uniqueness) was described in Section 4.4. Separate measures were created for IF data from the '8-Indicator Absolute' and 'Trend' models, and the results were then combined using algorithmic rules to give an overall significance assessment.

A network was created to demonstrate that reasoning networks provide a more effective way of fusing the different items of information to produce an overall significance assessment. These can also fuse additional information, such as trend data. The network created fused all the results of the data modelling (i.e. the distance and entropy values from the absolute and trend models), and then fused these with the results of the trend analysis (trend severity assessment and step change detection). Again, continuous nodes were used in the network to enable the entering of actual cluster distance and entropy values.

The network is shown in Figure 5-13. As indicated above, six items of information are entered: 'Absolute' model cluster distance and entropy, 'Trend' model cluster distance and entropy, trend severity, and whether a step change has been detected. Example cases are shown to demonstrate the performance of the reasoning network. The actual significance values generated are dependent on the details of the configuration of the probability tables in the reasoning network. Therefore what is of primary interest is the relative significance of the different example cases.

The data for the cracked MGB bevel pinion fault case (gearbox fit 999) is shown in Figure 5-13. This generated a high anomaly significance value of 70.3%. The top two alerts in the rule-based ranking in Section 4.4 that had occurred prior to the start of the first trial period also generated very high significance values. The 2nd epicyclic annulus aft (RH) of G-TIGE (gearbox fit 186) had a significance value of 88.0% (Figure 5-14), and the LHA oil cooler fan drive of G-TIGS (gearbox fit 859) had a value of 85.5%.

In contrast, the highly anomalous CI values from the 2nd stage sun gear of G-TIGC (gearbox fit 1053) which were believed to have been caused by a wiring harness problem only generated a significance value of 23.61%. The data was not trending significantly, and similar data characteristics had been seen on other aircraft. These similar characteristics, related to instrumentation faults, reduced the anomaly significance value, however they could enable case based reasoning to provide a diagnosis of the fault type (see the examples in Section 4.5). The data for the 2nd stage epicyclic annulus fwd (RH) of G-TIGF (gearbox fit 1055) also only had a significance value of 0.4% (Figure 5-15). Here the significance was suppressed by the fact that a step change had been detected. These two cases were ranked 9th and 5th in the rule-based significance ranking shown in Table 4-9.

The reasoning network has automatically highlighted the significance of a genuine component fault such as the cracked bevel pinion, while giving more common HUMS instrumentation faults a low significance. The key factors driving the result for the cracked bevel pinion were alerts from both the absolute and trend anomaly models, the clear FS trends, and the highly unusual data characteristics. The reasoning network provided a more effective method of fusing this data for an anomaly significance assessment than a rule-based approach. Anomaly diagnosis is considered in the following sections.

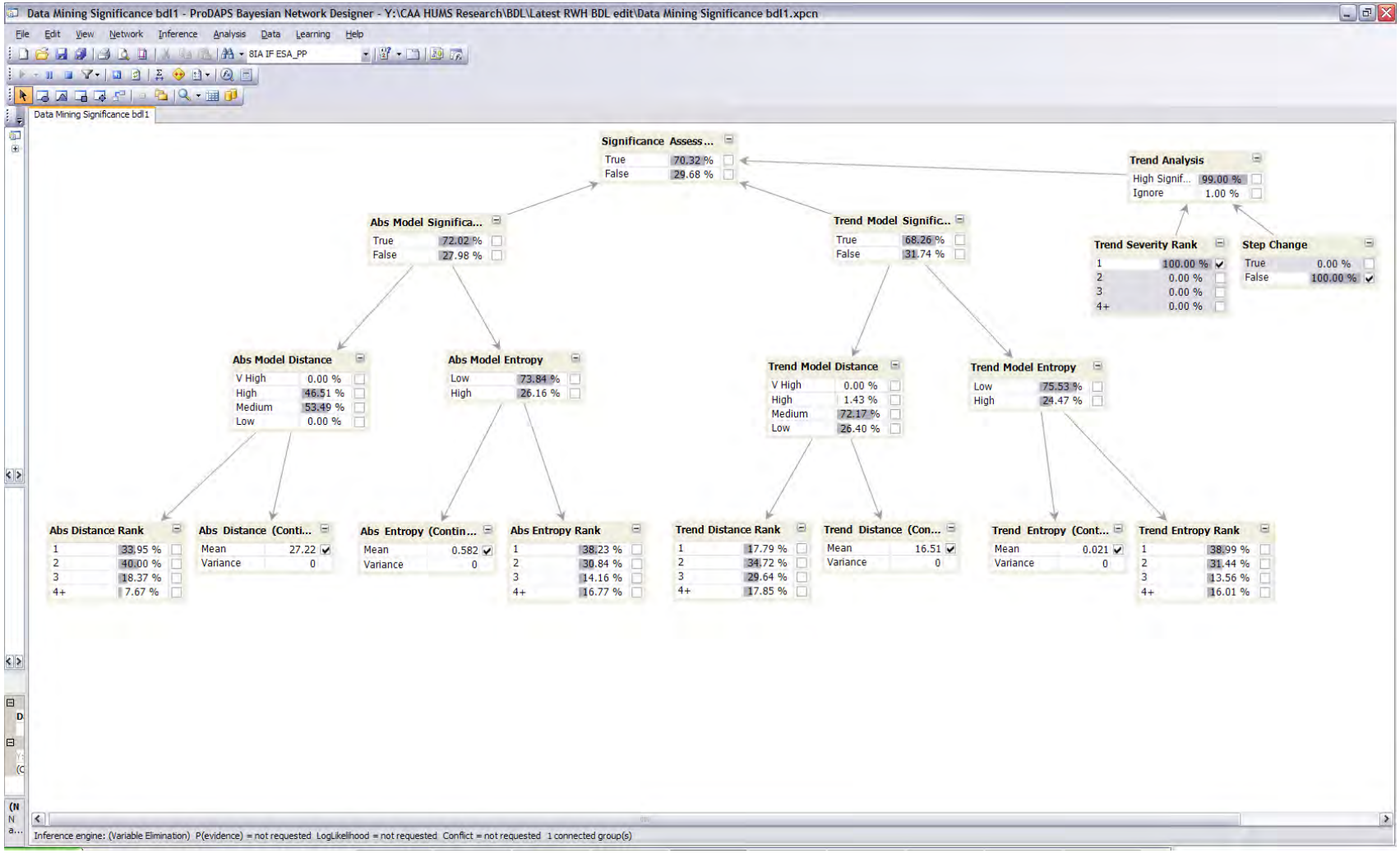

Figure 5-13 Anomaly significance network – evidence entered for gearbox fit 999

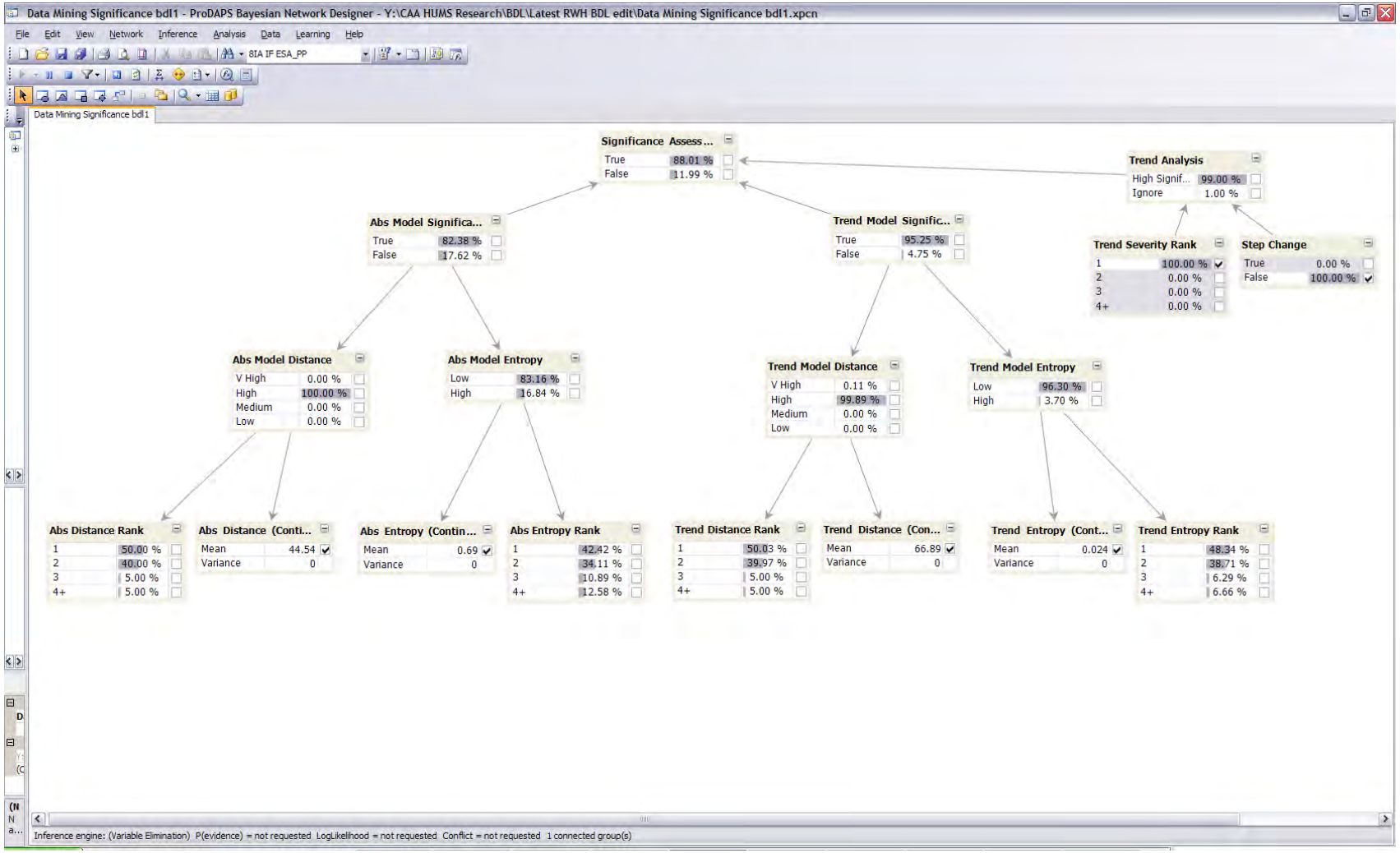

Figure 5-14 Anomaly significance network – evidence entered for gearbox fit 186

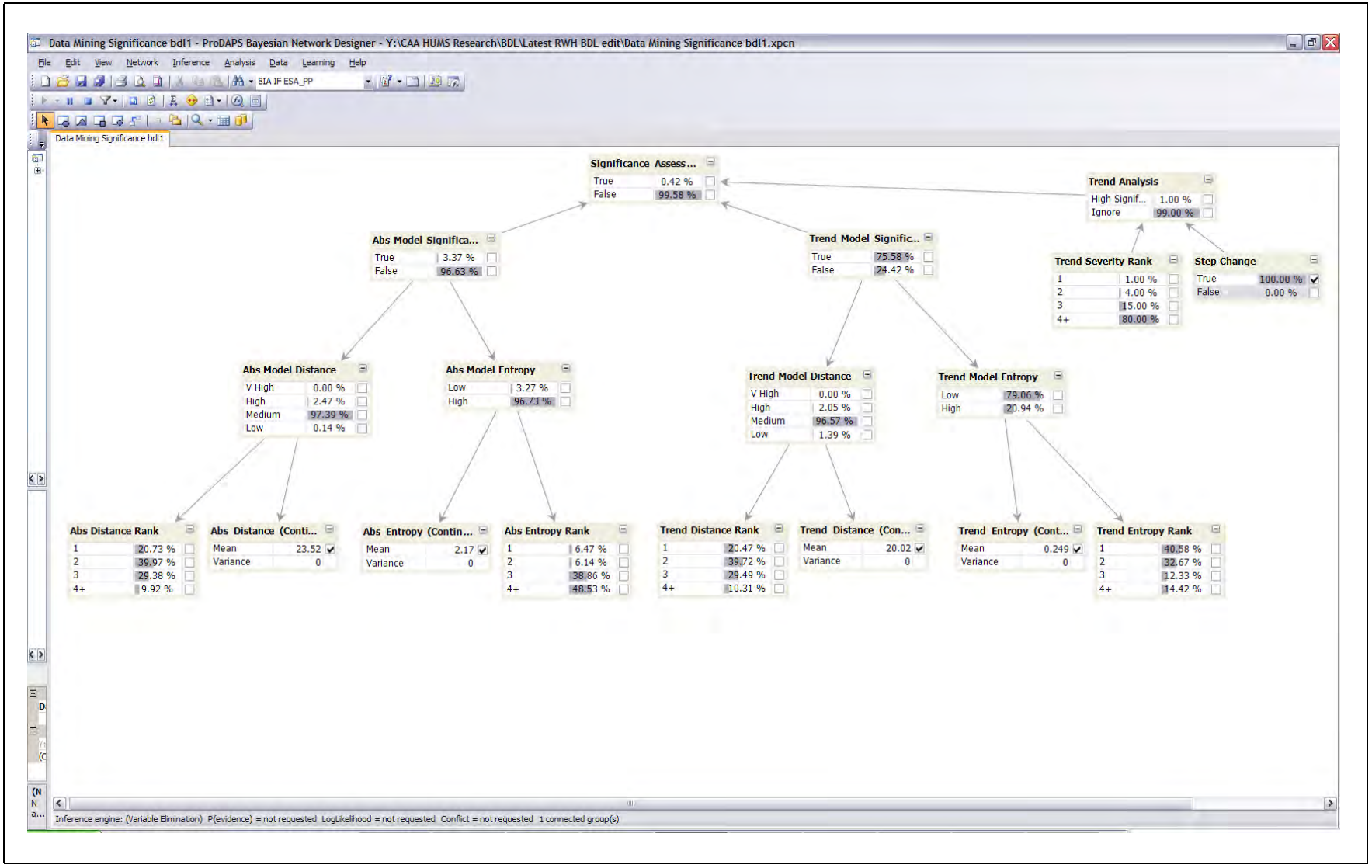

Figure 5-15 Anomaly significance network – evidence entered for gearbox fit 1055

5.5 Identifying Explainable Patterns

The previous reasoning networks are examples of data fusion, fusing sensor and component analysis information in a network utilising 'discrete' nodes, and fusing anomaly model IF outputs and trend information in networks utilising 'continuous' nodes, so that IF values could be entered directly. These networks increased the confidence in an alert, and provided diagnostic summary information and an anomaly significance assessment. The next step in the reasoning process should be automated fault diagnosis. However, aircraft faults can be rare events and even if two similar faults do occur, there can be differences in their specific characteristics. Most anomaly alerts are caused by maintenance actions or HUMS instrumentation problems. Therefore the first step in fault diagnosis should be to identify explainable patterns in the HUMS data due to maintenance or an instrumentation fault. Two example reasoning networks have been developed to demonstrate a reasoning capability to meet this requirement.

The first network detects evidence of a maintenance action (classifying this as normal or abnormal maintenance), and the second detects evidence of a HUMS instrumentation problem (classifying four different problem related signal characteristics). The networks again utilise 'continuous' nodes that enable PA and IF values to be directly entered as evidence. The networks also assume that there is an available output from a trend classifier providing FS trend information.

5.5.1 Maintenance Actions

The 'maintenance reasoning network' is shown in Figure 5-16. Anomaly alerts triggered by step changes in the HUMS CI data could be caused by normal maintenance, incorrect (or abnormal) maintenance, or another cause. If an alert is not caused by a step change in the data, it is unlikely to be due to a maintenance action. If the data after a step change is still within normal bounds, then the trend pre-processing should be re-set, but no further action is required. However, if the data after the step change is now abnormal, then incorrect maintenance may have been performed and further rectification maintenance could be required. The outputs from the reasoning network are therefore probabilities that normal maintenance has triggered the alert, that this is due to incorrect (or abnormal) maintenance, or that there is another cause. The first input to the network is a FS step change detector – if the alert is not due to a step change, the network will set the output to 'other' (i.e. no maintenance action detected). Other inputs to the network are '8-Indicator Absolute' and 'Trend', and 'M6 Absolute' model PA values, and the 'M6 Trend' model IF for ESA_M6. (The 'M6' information is included to indicate if there is abnormal impulsiveness in the signal after a step change, as no normal maintenance action should cause this).

Figure 5-16 shows the response of the network to data from the 2nd stage epicyclic annulus fwd (RH) of G-TIGG (gearbox fit 1055, see Figure 4-18 to Figure 4-20 in Section 4.4). A step change has been detected, resulting in elevated '8-Indicator' and 'M6' model PA values. The network has therefore classified this as an abnormal maintenance event, requiring further investigation. The step may not have been caused by a maintenance action, however this is an appropriate response indicating a possible cause.

Figure 5-17 and Figure 5-18 show network responses to data from the 1st stage planet gear of G-TIGG (gearbox fit 1056, see Section 4.3.3 and also reference [3]). Figure 5-17 shows the output following the installation of a faulty IHUMS DAPU on the aircraft - abnormal maintenance is diagnosed. Figure 5-18 shows the output following the replacement of this with a functioning DAPU - normal maintenance is now diagnosed, requiring no further action.

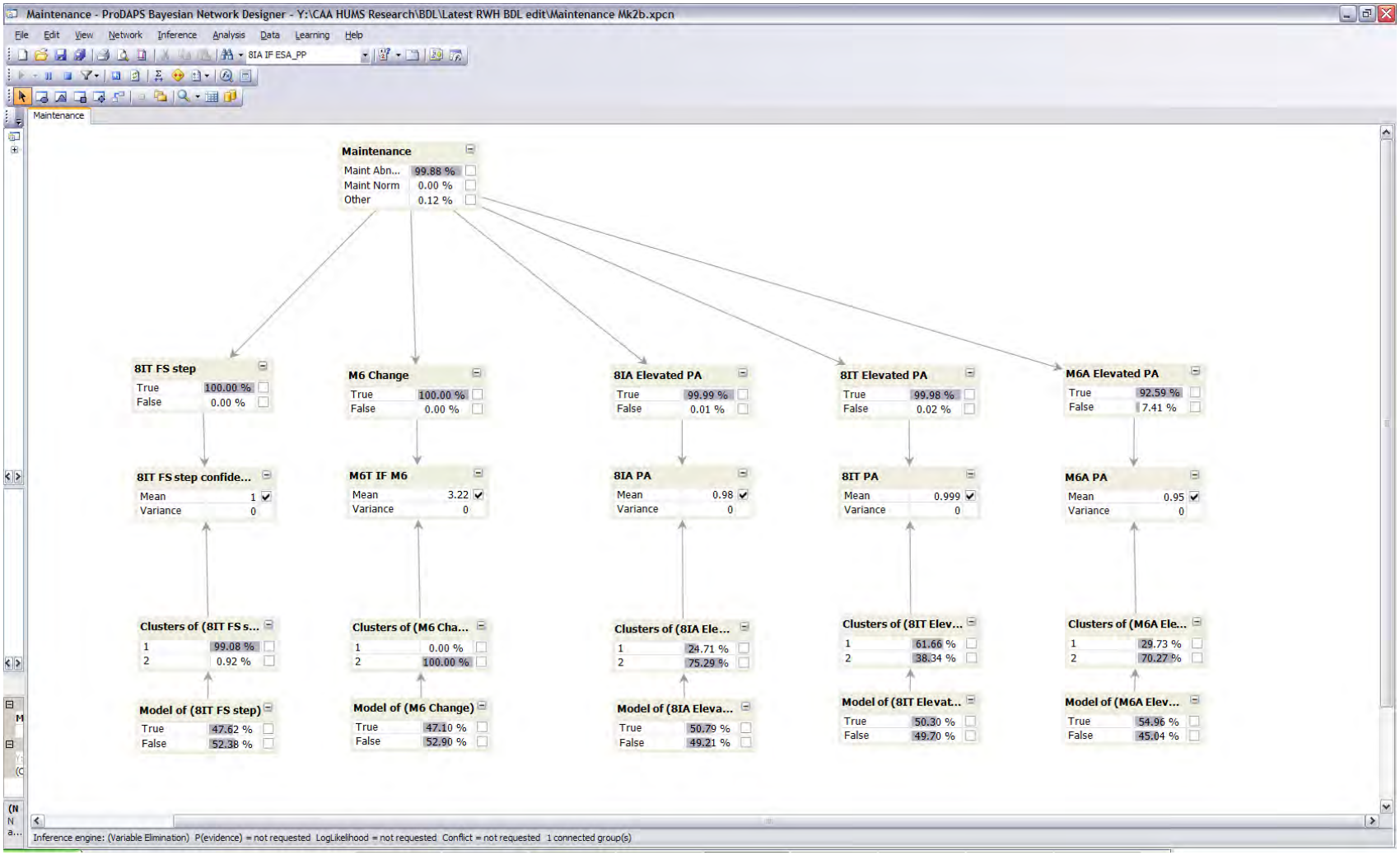

Figure 5-16 Maintenance network – evidence entered for gearbox fit 1055

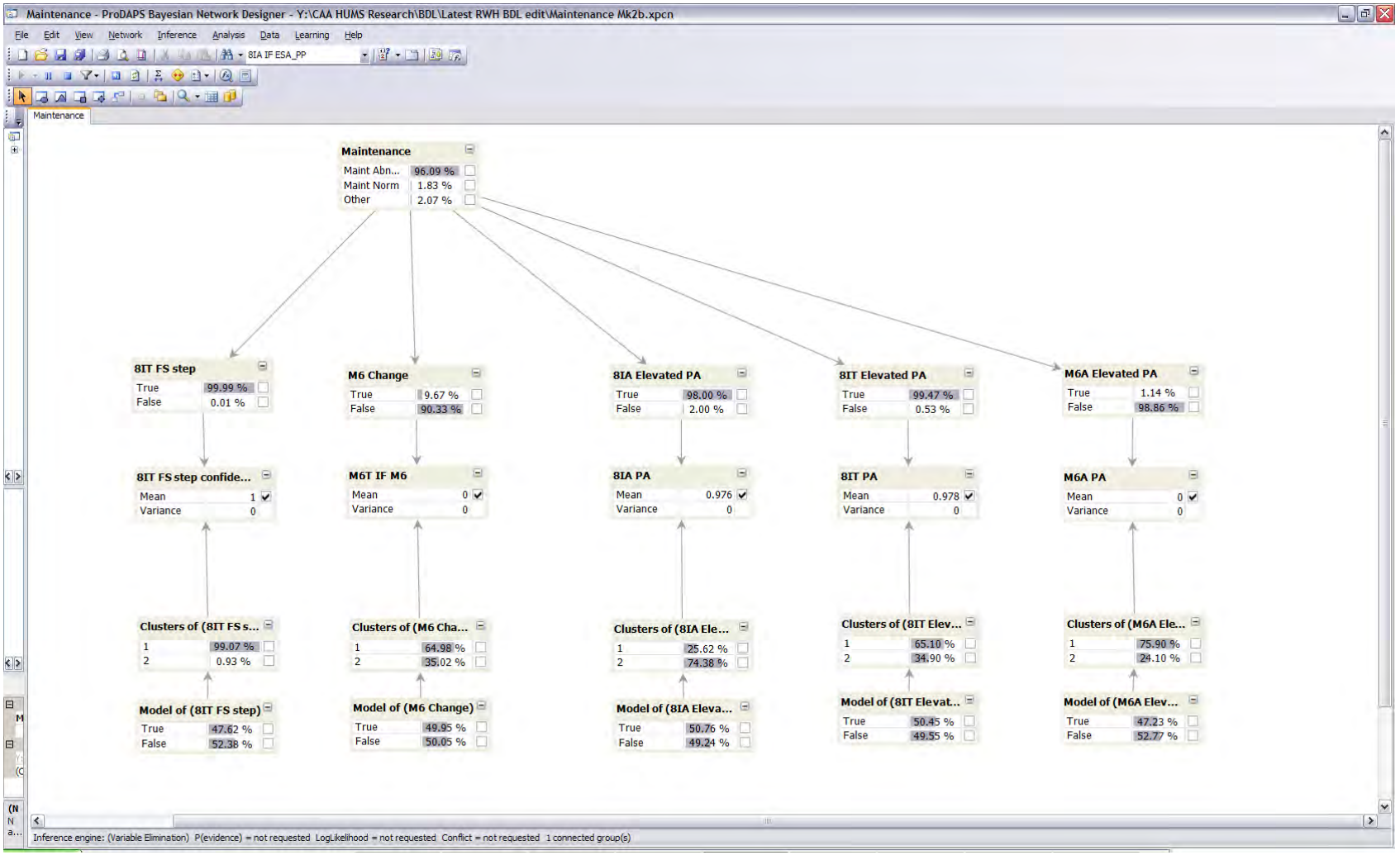

Figure 5-17 Maintenance network – evidence entered for gearbox fit 1056(a)

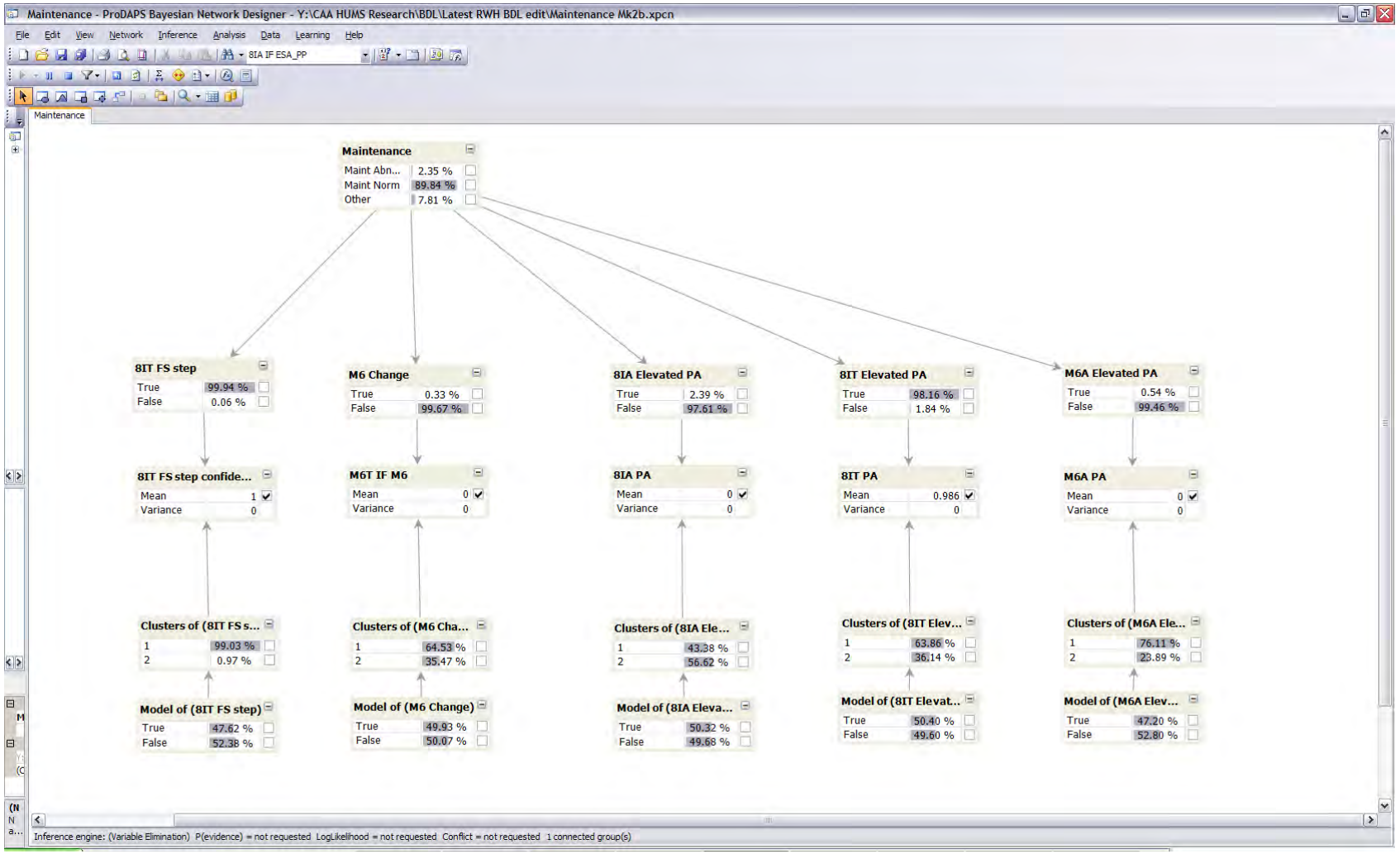

Figure 5-18 Maintenance network – evidence entered for gearbox fit 1056(b)

5.5.2 HUMS Instrumentation Faults

The 'instrumentation fault network' is shown in Figure 5-19, and the key elements of the network can be seen in Figure 5-20, where the network output is circled in red. The experience documented in references [3] and [5] shows that different types of instrumentation fault can generate different signal characteristics, such as increases in low frequency noise, high base energy (i.e. general signal noise), high overall signal energy, and loss of visibility of gear mesh frequency vibration. In addition to detecting evidence of a HUMS instrumentation problem, the network therefore identifies the presence of these four patterns (if no instrumentation fault is detected, the network output is 'other'). Two inputs to the network are '8-Indicator Absolute' model FS level and trend. Instrumentation faults can generate extreme data, generating a very low FS, and the data may either be consistently abnormal (i.e. not trending), or noisy. Input IF information is grouped according to the signal characteristics associated with the different types of instrumentation fault (e.g. 'Gear Mesh Tone', 'Shaft Order' etc.) Directionality information is also included to show whether a high IF is caused by a high or low HUMS CI value. Again, the network uses 'continuous' nodes so that IF values can be input directly.

Figure 5-20 shows the response of the network to data from the 2nd stage epicyclic annulus aft (RH) of G-TIGS (gearbox fit 1041), with a wiring harness problem. This generated extremely low FS values (Figure 5-21), caused by abnormally high ESA_SD/PP and SIG_SD/PP values (Figure 5-22) with little data trending. The network has diagnosed an instrumentation problem with the characteristics of high signal noise and overall energy. Figure 5-23 shows the network response to data from the left hydraulic drive 47-tooth gear of G-TIGG (gearbox fit 1080), with an accelerometer problem. This has caused a noisy FS trend and FSA_SO1 values (Figure 5-24 and Figure 5-25). The network has diagnosed an instrumentation problem with the characteristic of low frequency noise.

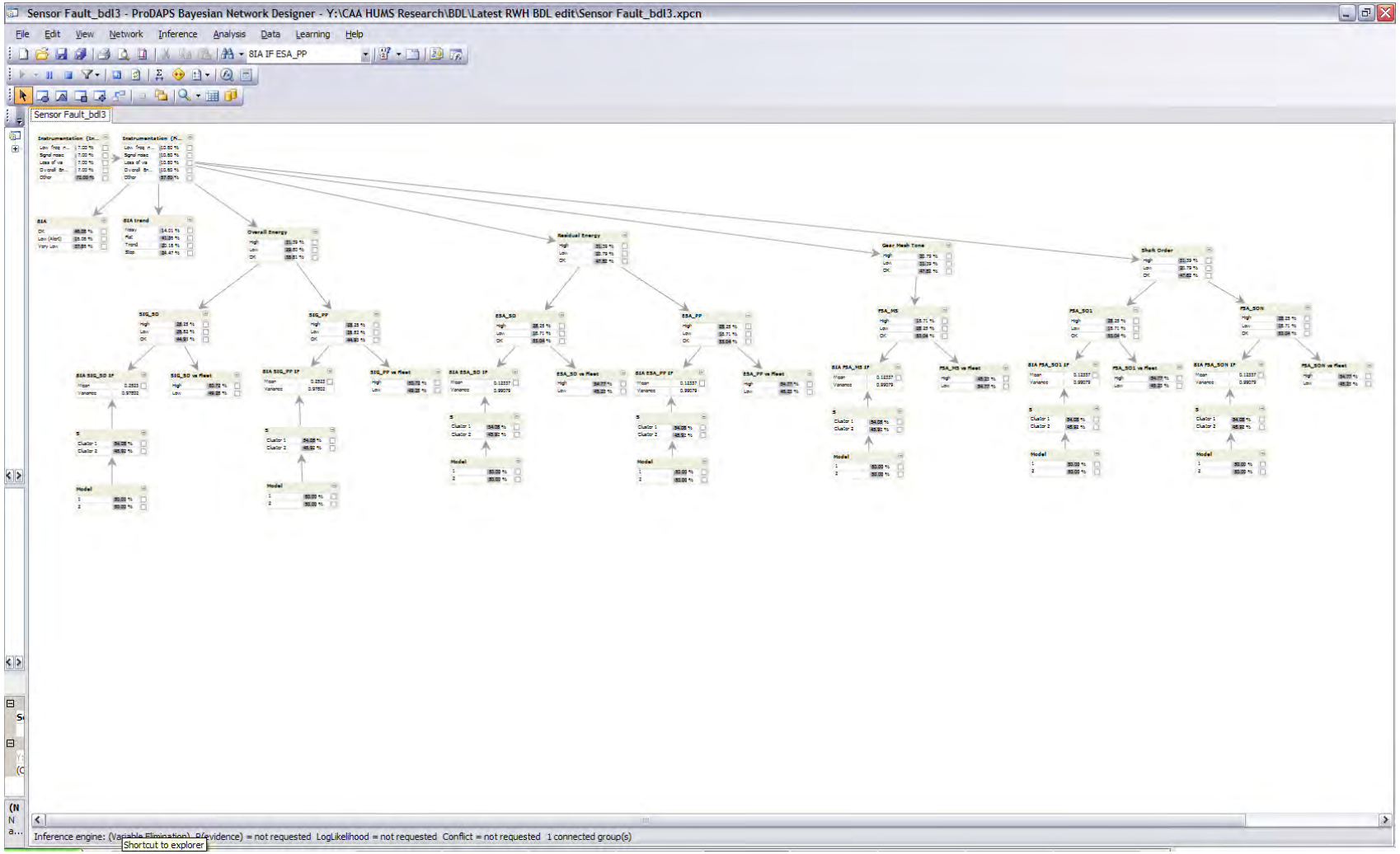

Figure 5-19 Instrumentation fault network

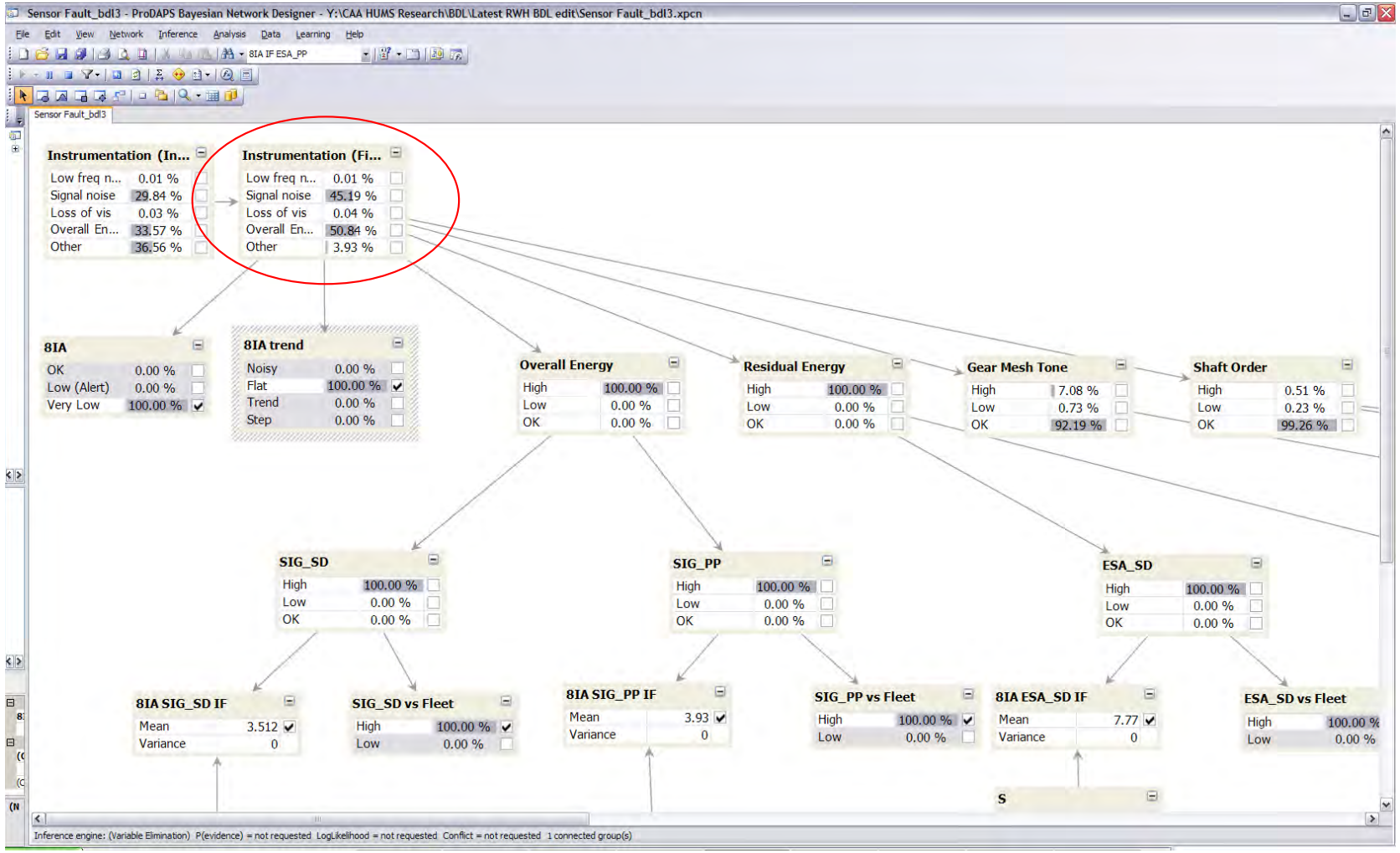

Figure 5-20 Instrumentation fault network – evidence entered for gearbox fit 1041

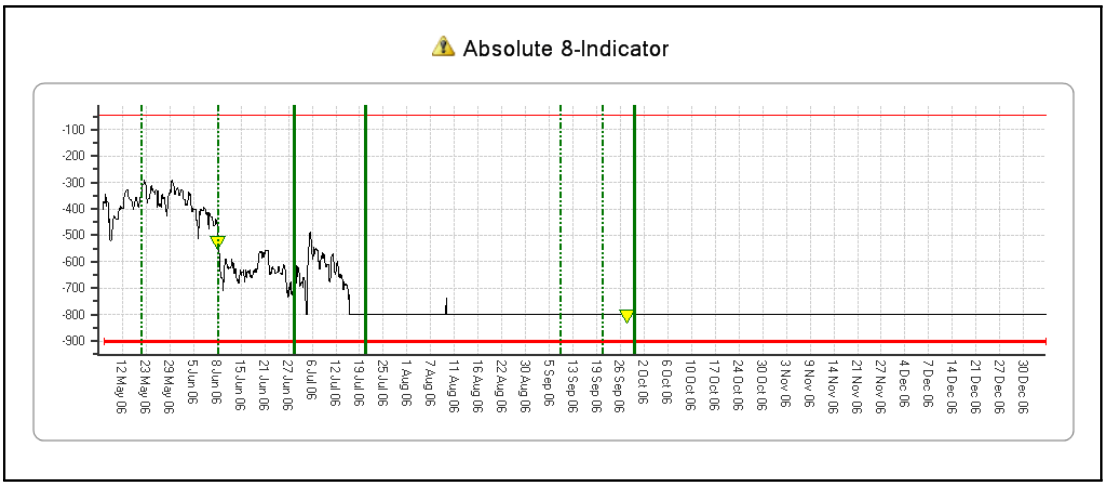

Figure 5-21 Gearbox fit 1041 – G-TIGS 2nd epicyclic annulus aft (RH) – 8IA model FS

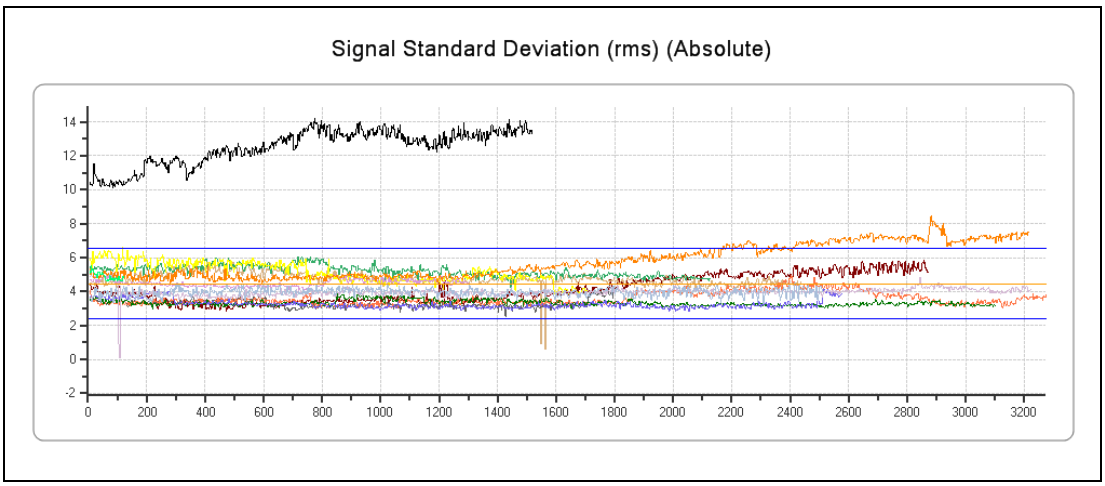

Figure 5-22 Gearbox fit 1041 – G-TIGS 2nd epicyclic annulus aft (RH) – SIG_SD – fleet view

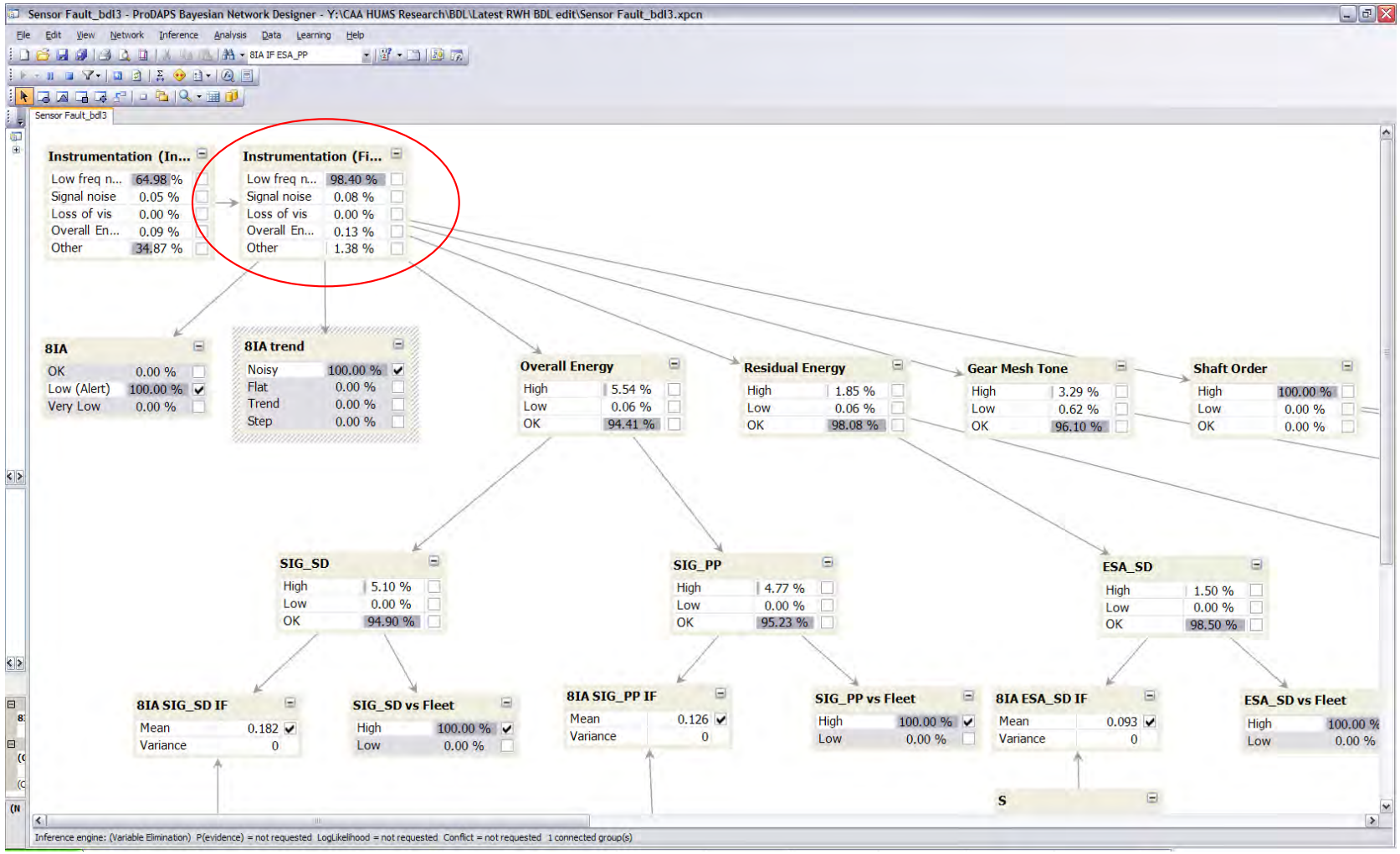

Figure 5-23 Instrumentation fault network – evidence entered for gearbox fit 1080

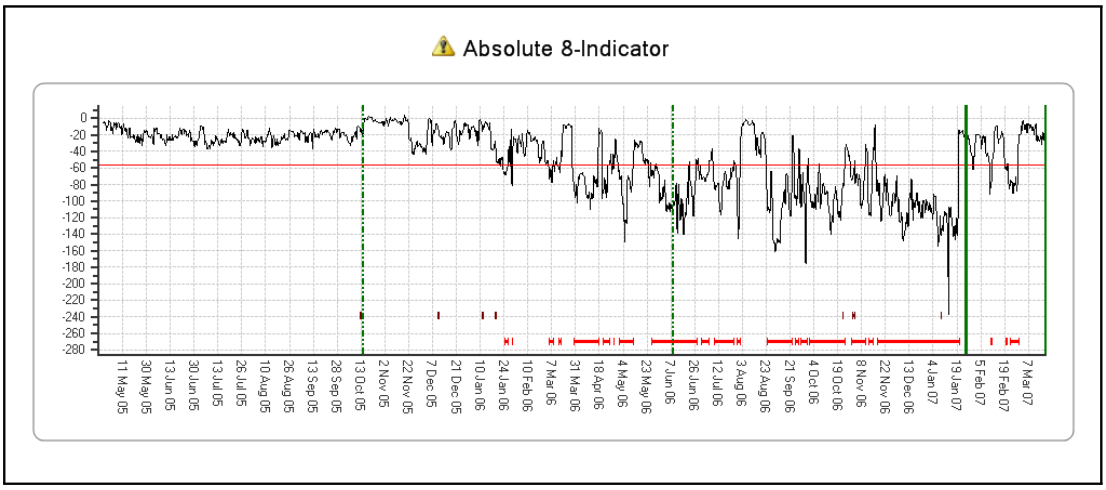

Figure 5-24 Gearbox fit 1080 – G-TIGG – LHA left hydraulic drive 47-tooth gear – 8IA model FS

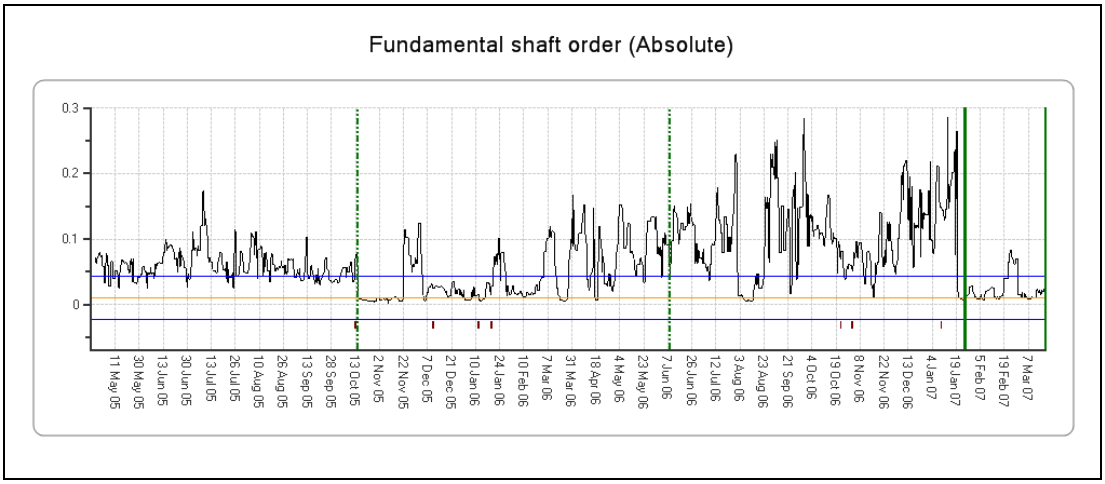

Figure 5-25 Gearbox fit 1080 – G-TIGG – LHA left hydraulic drive 47-tooth gear – FSA_SO1

5.6 Case Based Reasoning

Although no actual demonstration was performed, the final diagnostic element considered in the work was Case Based Reasoning (CBR). Fault diagnostic systems can utilise physics-based models to generate relationships between faults and monitored symptoms. However, the complexities of applying VHM to in-service aircraft means that it is impractical to model all the factors that could affect a component's vibration signature, and how this would be modified by the many different faults that could occur. Utilising a set of diagnostic rules based on engineering knowledge and previous experience is unreliable, as a new fault could have certain previously unseen features that contradict earlier experience. While only providing a partial diagnostic solution, with the current state-of-the-art of VHM technology, CBR provides a potentially valuable additional element of an overall automated reasoning capability.

CBR is based on the creation of a library of documented fault cases, containing information on both the fault that had been detected, and how this affected the HUMS CI trends. The information to support this can be generated by data mining tools as described in Section 4.5. When a new anomaly is detected, it can be compared to this library of documented fault cases, and any matching cases identified. Knowing the cause of the matching case provides a possible explanation

for the current anomaly. This is an extension of the reasoning to identify explainable patterns related to maintenance actions and instrumentation faults described in Section 5.5. These instrumentation faults could actually be included as cases in a CBR system. If the characteristics of a new anomaly do not match the explainable patterns, or any of the CBR cases, the automated anomaly significance assessment capability described in Section 5.4 can still provide useful information to determine the most appropriate response.

CBR is the concept of using past experience to solve a new problem. It does not denote a single algorithm or even class of algorithm to solve a problem but rather a framework for structuring a reasoning system that emphasises the role of previous cases for solving a new problem. Each case describes a problem, its solution if available and outcome. A CBR tool will execute four main processes: retrieve, reuse, revise and retain. Retrieve searches for a stored case that is relevant to the current situation; reuse applies the knowledge of the retrieved case to the new case; revise will test the solution and adapt it if necessary; retain updates the library of cases.

CBR as a concept is applicable to many situations such as medical diagnosis, legal proceedings and decision planning etc. Each task domain presents specific challenges and for this reason it is difficult to source an off the shelf approach that can be directly used to model a specific application. While CBR is undoubtedly a highly desirable asset, a CBR system will need to draw on many topics from Artificial Intelligence including knowledge representation, search, learning and planning. Many applications may also need to draw on uncertainty modelling and natural language models when data is stored as text.

CBR for vehicle health monitoring may involve a complex array of technologies. Troubleshooting a problem can involve multiple sensor types, modelling features over time, maintenance history, vehicle usage, etc. Consider vibration monitoring in isolation. Knowledge can be structured in multiple ways and may demand many forms of representation. The knowledge base will include: the physical make up of mechanical components, their relationship to other components both in a physical and functional sense; the physics of vibration features; the representation of vibration patterns across multiple features and over time; sensor location and component visibility. Knowledge about a previous case has to be stored with a sufficiently rich description to test its match with a new case and to permit discrimination between other stored candidate cases. Stored cases will be indexed by features and these features have to assist efficient retrieval and support measures of nearest match between cases. Indexing of cases could be done in a way that permits simultaneous matching but this might lead to added complexity because indexes would need constant adjustment as the historical case library is updated with new cases. Also some types of information are difficult to represent for case matching. For example, time histories of CIs are difficult to compare between different cases.

So CBR presents some difficult challenges but the importance of capturing case histories is seen as essential in many areas of business. Many business scenarios do not necessarily present classical CBR tasks but they do emphasise the need to capture experience and for an ability to relate disparate types of data. A large amount of knowledge gets lost or becomes difficult to source because it is never stored electronically, or when it is it remains hidden from information extraction tools. The message here is that the philosophy of CBR needs to be embraced and not delayed until a mature solution is in place. The internet has had a significant impact on technology development because it facilitates the dissemination of research results. The electronic infrastructure of the World Wide Web allows rapid communication and sophisticated search engines trawl millions of sites indexing trillions of items of data.

The work reported in this document has made a good start in showing how some difficult to describe information features can be represented. A lot of information about anomalies is compressed in the FS; this score gives a measure of abnormality and captures, in a single feature, trend information over multiple vibration variables. IFs normalise model outputs, permit comparisons between vibration features measured on completely different scales, and represent complex data in a way that is suited to automated information processing. All of these measures facilitate describing the complex data that accompanies cases found in a HUM system.

FSs and IFs are being stored electronically. The in-service trial also illustrated the capture of free form text in the format of reports. These reports captured observations, actions and outcomes. The next stage is to construct a formal procedure to ensure that data about any case that has caused an alert is captured electronically. The questions to start with are: What data / information should be captured? How do we ensure it is always captured? A web based service should aim to capture experience from a broad field of vehicle operations. The experience should be periodically analysed and lessons learnt made available as information through the service.

6 Conclusions and Recommendations

6.1 Conclusions

The advanced anomaly detection system developed in support of the CAA research programme has improved the performance and usability of HUMS. The system has continued to out-perform the traditional HUMS analysis in successfully highlighting both aircraft problems and HUMS instrumentation faults. The new PA and IF outputs provide clear information to assist the operator in identifying abnormal HUMS Condition Indicator (CI) trends and diagnosing their possible causes. However, the system operator must still view multiple PA, IF and HUMS CI charts when investigating an anomaly alert. In addition to the workload issues, this manual process requires HUMS knowledge and experience to ensure that the operator arrives at the correct conclusions, and determines the most appropriate actions.

The anomaly detection system represents the first level of a new advanced HUMS data analysis capability. The normalised PA and IF outputs can be further processed in a second level of advanced data analysis, utilising data mining and automated reasoning tools. This second level of analysis provides the greatest scope for further improving the performance and usability of HUMS, giving valuable additional information to assist the operator in determining the most appropriate response to an anomaly alert.

This report presents the results of additional research tasks 5 and 6, involving data mining and reasoning technology demonstrations exploring the approach to, and potential benefits of, a second level of advanced data analysis. The work has demonstrated new system concepts and capabilities, and the technologies that could provide them. Further work is required to develop these technologies to a readiness level suitable for implementing in an operational system.

The data mining demonstration extracted new information from the anomaly model output data for post-processing of anomaly alerts and use in automated diagnostic reasoning. It focused on automated change detection, trend analysis, anomaly significance assessment, and data analysis to support Case Based Reasoning. The reasoning demonstration focussed on data fusion and fault diagnosis. The data fusion utilised physical relationships between HUMS sensors and component analyses, generated diagnostic summary information, and supported the anomaly significance assessment. The fault diagnosis element demonstrated an ability to automatically identify explainable patterns due to maintenance actions and instrumentation faults, and explored the application of Case Based Reasoning. Mock-ups of possible new anomaly detection system alert displays were created to illustrate how the resulting new information could be presented to the operator.

The data mining and reasoning demonstrations have clearly shown how secondary analysis of the current anomaly detection system outputs can provide further benefits to operators. This can reduce operator workload, automatically highlight significant patterns that the operator should pay careful attention to, and provide new information to support decision making. While the techniques demonstrated are not the only ones that could be applied, the work has illustrated the type of approach that can be taken to ensure that robust and reliable capabilities are produced. Taking the wrong approach could result in misleading information being presented to the operator, which is clearly unacceptable.

Several of the techniques demonstrated are designed to capture new knowledge during system operations, and are based on the on-going evaluation of data. It would be difficult to apply these when initially qualifying equipment, however their purpose

is solely to provide a secondary evaluation of the data. Understanding will develop as the database of experience grows, and there should be an on-going secondary processing development task to maximise the benefits of this understanding. Some of the capabilities demonstrated may be more for in-house use by an analysis service provider, rather than for use by an operator. This would certainly be the case where a service is being provided to multiple small operators, with information and experience being pooled in one location.

6.2 Recommendations

Having demonstrated the HUMS performance improvements that can be achieved through the application of advanced anomaly detection, the greatest scope for further improvement in the combined performance of the system and operator is through the development of a second level of advanced data analysis, based on data mining and automated reasoning technologies. Therefore it is recommended that, building on the results of the technology demonstrations described in this report, work continues on the development of this second level of advanced data analysis.

The key focus of the second level of advanced data analysis is assisting the operator in determining the most appropriate action, given the information that is currently available. Effective operator interaction with the system will be critical. Therefore it is recommended that, before further development work is undertaken, operator feedback is obtained on the contents of this report.

The work has highlighted the value of the PA, Fitness Score (FS) and IF outputs from the anomaly models. However it has identified a need for the IF outputs to be enhanced with directionality information, indicating whether a HUMS CI that is driving an anomaly indication is increasing or decreasing. It is recommended that this information is added to the IF outputs.

Data mining is an effective tool for extracting new information from the database of HUMS experience. However, further development of fault diagnostic capabilities also requires information on component condition to correlate with HUMS data trends. Therefore an effective mechanism should be put in place to obtain feedback on component condition from the Repair and Overhaul process.

7 References

1 Harrison, N., Baines, N. C. (1999). *Intelligent Management of HUMS Data: The use of Artificial Intelligence Techniques to Detect Main Rotor Gearbox Faults*. Study II, CAA Paper 99006.

2 Smiths Aerospace report REP1697(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Interim Report on Phase 1 of the Research Project*. May 2007.

3 Smiths Aerospace report REP1712(2): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Phase 2 of the Research Project: Six-Month Operational Trial*. July 2007.

4 GE Aviation report REP1731(3): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Additional Research Work (Tasks 1 to 4)*. June 2009.

5 GE Aviation report REP1771(1): *Intelligent Management of Helicopter Vibration Health Monitoring Data: Application of Advanced Analysis Techniques In-Service - Report on Second Operational Trial Period*. May 2009.

6 Christopher M Bishop (2006): *Pattern Recognition and Machine Learning*. Springer.

INTENTIONALLY LEFT BLANK