An Alternative Approach to Anomaly Detection in Health and Usage Monitoring Systems – Mixture Modeling

Submitted to:
American Helicopter Society
Airworthiness, CBM, and HUMS Specialists’ Meeting
February 2013

Submitted by:
Dean Westerfield – QinetiQ North America, Inc., Lifecycle Solutions
Christopher Doktor – Frontier Technology, Inc.

QinetiQ
North America
An Alternative Approach to Anomaly Detection in Health and Usage Monitoring Systems – Mixture Modeling

Abstract: A pilot project was conducted by QinetiQ North America, Inc. in cooperation with Frontier Technology, Inc. to evaluate a new, early fault identification technology for use in Health Monitoring and Usage Systems (HUMS). The pilot evaluation developed a statistical Mixture Model using Frontier Technology’s NormNet® Prognostic & Health Management (PHM) software tool and employed integrated vibration, parametric, and maintenance data extracted from the Condition Based Maintenance Data Warehouse (CBM-DW), developed by QinetiQ North America, Inc. under contract to U.S. Army Aviation and Missile Life Cycle Management Command (AMCOM) G-3 office. The focus of the pilot was the AH-64D Apache helicopter’s Tail Rotor Gearbox Assembly.

The ability of the Mixture Model based technology to predict hardware faults was compared to that obtained employing Condition Indicator (CI) Limits under identical assumptions and conditions. The comparison showed the probability that an actual anomaly would be detected by the Mixture Model was 2.5 times better than that produced using CI Thresholds, as fielded at the time of this paper. The False Negative (Missed Fault) and False Positive (False Alarm) rates of the Mixture Model were half those obtained using CI Thresholds.

Executive Summary: In March of 2012 QinetiQ North America, Inc. in cooperation with Frontier Technology, Inc. began a pilot evaluation of a new pattern recognition technology for detection of hardware anomalies for use in Health Monitoring and Usage Systems (HUMS). The evaluation focused upon AH-64D Tail Gearbox and Rotor System. The required data, used with permission of the U.S. Army Apache Project Management Office, was extracted from the Army Aviation Condition Based Maintenance Data Warehouse (CBM-DW). The CBM-DW was developed by QinetiQ North America under contract to the U.S. Army Aviation and Missile Life Cycle Management Command (AMCOM) G-3 office. This data was modeled and analyzed using Frontier Technology’s NormNet® Prognostic & Health Management (PHM) software tool.

The pilot project addressed the central problem of condition based maintenance – accurate, early fault identification (FID) in critical subsystems based upon digital source collection (DSC) data. In HUMS, vibration data is typically reduced to a set of condition indicators (CIs) where each CI is computed as a function of a single sensor’s data. The simplest and most common FID technique employs a static threshold such that when a CI value exceeds its threshold an anomalous hardware condition is indicated. In practice, it is difficult to find a threshold value that produces an acceptably high anomaly detection rate while simultaneously achieving a reasonably low false alarm rate. There is no guarantee for many CIs that an acceptable threshold exists.

An alternative approach to CI Thresholds for anomaly detection employs statistical baseline modeling of the HUMS data. In intuitive terms, this statistical technique uses a population of data describing anomaly free, system behavior to generate a mathematical model representing normal behavior. Actual system behavior is then compared to normal behavior, as predicted by the model under similar conditions, to measure how close to normal the actual system
behavior is. Abnormality is determined by the statistical improbability of the measured values against the predicted system behavior over time. The NormNet® Prognostic Health Management (PHM) tool uses a Bayesian statistical algorithm to produce a Probability of Abnormality (POA) value to describe the state of the system as data is processed over time. While the sensitivity of this algorithm can be tuned, it assesses system health largely by the inherent statistics of the baseline model.

The statistical baseline modeling approach used for this project is referred to Mixture Modeling (Hastie, et. al., pg 188). The mixture modeling approach is a technique for modeling complex (including multi-modal, non-linear and non-stationary) statistical distributions as a linear combination of simpler, flexible, multi-variate distributions. A mixture model does not employ one dimension at a time like CI Thresholds. Rather, as a multi-variate statistical distribution, it models the relationships (i.e., covariances) across the entire measurement set combining both vibration and parametric data into a single model. The resulting model essentially defines a normal operating envelope of the monitored system, along with the expected system variances. The advantage of this approach is that the normal or expected value of any measure is predicted in the context of all other measurements, i.e. in the context of what the entire system is doing at the time of the measurement.

The AH-64D Tail Gearbox pilot evaluation employed 24 synchronous measurements including four vibration and 20 parametric measures that were readily available in statistically significant quantities from the CBM-DW. The baseline model, representing normal behavior of an AH-64D Tail Rotor Gearbox, was developed using 129 case histories that were identified as nominal based upon their maintenance records, and review of each case’s content. This data set was referred to as the “Training Data.”

An additional 173 Tail Gearbox case histories represented the “Test Data” were employed to evaluate the Mixture Model’s performance. The Test Data contained cases representing both nominal and known anomalous conditions (based upon maintenance records) – the point being to test the model’s ability to tell them apart. Each test, Tail Rotor Gearbox case history was analyzed with the use of the NormNet® based model for the existence and timing of anomalous behavior indicating the presence of an incipient hardware fault. The result of this analysis was compared to maintenance records to score each test case as True Positive, False Negative (Missed Alarm), False Positive (False Alarm), or True Negative. Only sensor detectable faults were counted (i.e. scratches, leaking seals, and corrosion were excluded). A case was scored as True Positive only if the model detected an anomaly and it was confirmed by the existence and timing of a relevant maintenance action. Additionally, if sensor data was available subsequent to the maintenance, it must have shown a decline in PoA after the maintenance action was complete.

The performance of the Tail Rotor Gearbox Mixture Model was compared to that obtained using CI Thresholds for all CIs defined for the same sensors used to develop the Mixture Model. The results of the comparison showed the probability that an actual anomaly would be detected by the Mixture Model was about 2.5 times that obtained using CI Thresholds. The False Negative (Missed Fault) and False Positive (False Alarm) rates produced by the Mixture Model were about half that produced using CI Thresholds.
Contents

1 Introduction ........................................................................................................................................... 4
2 Methodological Overview ...................................................................................................................... 5
3 Mixture Models .................................................................................................................................... 5
   3.1 Application of Mixture Modeling within NormNet® PHM tool ....................................................... 6
4 Pilot Project Methodology ................................................................................................................... 7
   4.1 Source Data .................................................................................................................................... 8
   4.2 Component Selection ....................................................................................................................... 9
   4.3 Channel Selection .......................................................................................................................... 10
   4.4 Case History Generation ............................................................................................................... 12
   4.5 Model Generation .......................................................................................................................... 13
   4.6 Case History Evaluation ............................................................................................................... 15
   4.7 Example Case Evaluation – Pattern Recognition ........................................................................ 16
5 Issues Affecting the Experimental Outcome - Lessons Learned .......................................................... 18
   5.1 Missing and Inaccurate Maintenance Records Drive up False Positive and False Negative .... 18
   5.2 Tail Rotor Gearbox Failures Rarely Detectable ............................................................................. 19
   5.3 Pressure Altitude Generates False Anomaly .................................................................................. 19
   5.4 Failures of Interest were Primarily Tail Rotor Faults .................................................................... 19
   5.5 Sensor Channel Failures ............................................................................................................... 20
   5.6 Experiment Actually Measured the Tail Rotor System ................................................................ 20
6 Distribution of Actual Failures in Test Dataset .................................................................................... 20
7 Experimental Results ............................................................................................................................ 21
8 Analysis ................................................................................................................................................ 23
9 Conclusion ............................................................................................................................................ 24

Appendix A – Reconciliation of Performance Matrix Tables ................................................................... 26
References ................................................................................................................................................. 27
Acronyms ................................................................................................................................................. 27
1 Introduction

In March of 2012 QinetiQ North America, Inc. in cooperation with Frontier Technology, Inc. began a pilot project to evaluate a new pattern recognition technology for detection of hardware anomalies for use in Health Monitoring and Usage Systems (HUMS). The evaluation focused upon AH-64D Tail Rotor Gearbox Assembly and subsequently added the Rotor System Assembly. Data required for the analysis and used with permission of the U.S. Army Apache Project Management Office, was extracted from the Army Aviation Condition Based Maintenance Data Warehouse (CBM-DW). The CBM-DW was developed by QinetiQ North America under contract to the U.S. Army Aviation and Missile Life Cycle Management Command (AMCOM) G-3 office. Using this data, and employing Frontier Technology’s NormNet® software, a statistical Mixture Model was developed. The pilot evaluation was completed in about four person-months.

The pilot project addressed the central problem of condition based maintenance – accurate component prognostics based upon digital source collector (DSC) data. In HUMS, vibration data is typically reduced to a set of condition indicators (CIs) where each CI is computed as a function of a single sensor’s data. The simplest and most common prognostic technique employs a static threshold such that when a CI value exceeds its threshold an anomalous hardware condition is indicated. In practice, it is difficult to find a threshold value that produces an acceptably high anomaly detection rate while simultaneously achieving a reasonably low false alarm rate. There is no guarantee that for many CIs, a threshold that meets these criteria exists.

An alternative approach to CI Thresholds for anomaly detection employs a Mixture Model. In intuitive terms, this statistical technique employs a population of data describing anomaly free system behavior, to generate a model representing normal behavior. Actual system behavior is then compared to normal behavior, as predicted by the model under similar conditions, to measure how close to normal the actual system behavior is. Abnormality is determined by the statistical improbability of the measured values against the predicted system behavior over time.

Unlike CI Thresholds, a mixture model employs more than one dimension at a time. It is a multi-dimensional model describing an envelope of normal behavior in n-space. Vectors of simultaneous measurements are collected during system operation that describe the performance of the target system in the context of the environment in which it is operating. This context includes all relevant vibration and parametric measures (e.g. speed, torque, temperature, etc.) that are synchronous and relevant to the performance of the target subsystem. The distance, in n-space, of these vectors from normal behavior, as predicted by the model, is computed and the joint probability of that distance is determined. In short, the greater the distance from “normal” a vector of measurements is, the greater the probability that there exists an anomaly. A single Probability of Abnormality (PoA) threshold is selected such that when that threshold is exceeded a component’s condition is flagged as anomalous.
An obvious advantage of a Mixture Model approach to anomaly detection is that no individual CI thresholds are employed. The methodology is entirely data driven requiring only that a significant population of synchronous historical data be available and that the data can be classified as describing normal versus abnormal behavior. This approach reduces the cost and effort inherent in generating individual thresholds for each CI (although it does not eliminate the CIs themselves.).

The Mixture Model approach has the additional advantage that the environment in which the component is operating is part of the equation; hence there is reduced propensity to mistake measurements collected under extreme conditions or unusual regimes as anomalies. This reduces the false positive rate without sacrificing true positive accuracy.

Of course, the claims of improved accuracy and reduced false positive rates over those achieved with CI Threshold limits must be demonstrated. To that end, the pilot project documented by this paper was conducted. The goals of the pilot were to demonstrate that the Mixture Modeling technique can

1) Detect hardware anomalies
2) Detect hardware anomalies more reliably than CI Thresholds
3) Detect hardware anomalies earlier than CI Thresholds

2 Methodological Overview

There are two sections to the discussion of methodology. The first section addresses the technical elements of Pattern Recognition, in particular statistical Mixture Models and their application via the NormNet® software to Prognostic Health Management. The second section describes the methodology employed by the pilot project to compare the performance of Pattern Recognition to CI Thresholds as applied to a single, selected subsystem.

3 Mixture Models

As previously mentioned, the methodology used in this study operates on the general principle of monitoring system performance variables in the context of a system's operating environment. In this case, variables such as torque, control position, vibration and temperature are monitored relative to the environment of airspeed, altitude, air temperature, pitch rate, yaw rate, etc. That is to say, that if operating at some speed, altitude, etc. one would expect a healthy system’s torque, vibration amplitudes (and frequencies), etc. to fall within some definite (but perhaps ‘fuzzy’) range. Likewise, as the operating environment changes, one would expect the system parameters to operate in some different, but still specific range. The goal is to allow a computer system to observe these measurements, determine the expected patterns of behavior for a healthy system, and then reliably determine when a system is no longer following those patterns. The expectation is that a system that violates these patterns is developing some underlying fault. This is an application of statistical pattern recognition.

There are many different types of algorithms, serving a variety of purposes, within the field of statistical pattern recognition. The NormNet® PHM tool that was used in this study uses several
techniques in combination to develop and apply models of system behavior. But, the core functionality is based around the concept of a multi-variate, statistical, mixture model.

The idea behind mixture modeling is simply that complex statistical distributions may be reasonably approximated by superposition of a group, or “mixture” of simple base distributions [ref, Bishop]. Any statistical distribution or combination of distributions can be used for the base distribution, but the most common is the Gaussian distribution. Each of these base distributions captures the variances and covariances between channels over a portion of the full operating range. Hence, the data collected from helicopter sensors can be thought of as being generated by a complex combination of statistical and deterministic processes, and therefore represents sample data from a complex underlying distribution. This distribution can be represented by a series of simpler distributions, or “centers,” that capture the relationships, or correlations, between channels over specific sub-regions of the operating envelope.

Mixture modeling has application across a broad spectrum of fields, including pattern recognition, signal processing, control system design, data mining and statistical modeling. This is because it is easily combined with almost any other signal processing/simulation/analysis/modeling technique (e.g., neural networks, kernel methods, physics-based models, wavelet transformations, Kalman filters, etc.) to create powerful modeling, prediction and simulation algorithms.

3.1 Application of Mixture Modeling within NormNet® PHM tool

The NormNet® PHM algorithms are built on the premise that complex engineering systems (mechanical, fluid, electrical, electro-mechanical, etc. such as engines, pumps, controllers, robots, etc.) are complex in both process and time. By process complexity we mean that the physical principles (and therefore physics equations) by which they operate can be highly non-linear. By temporal complexity we mean that the physics and/or external inputs that dominate a system’s behavior changes over time. In engineering parlance, the temporal complexity of a system is usually discussed in terms of transient versus steady-state operation when referring to the physics of a system. When referring to the external system inputs, engineers typically refer to the ‘duty cycle’ to which a system is subjected. In statistical terms, a system that has transient behavior and/or a non-steady duty cycle is referred to as a non-stationary. Therefore, from a statistical standpoint, the problem being solved is the modeling of a non-linear, non-stationary, statistical process. It is because of the complex nature of this statistical process that a mixture model is employed. The mixture model method enables the mathematical characterization of complex behavior as a set of linear sub-regions referred to as “centers.”

The NormNet® PHM modeling and analysis process requires a set of sample data, called the “training data,” to build the baseline model of system behavior. This data must be collected from a population of system instances that are known to be healthy, and must cover the entire expected operating envelope. The size of the healthy system population is determined by the intended application, and the desired sensitivity of the models. The population size can be a single example of the system, but this would not account for natural variation across the
population of healthy systems. Therefore, such a model would apply only to the specific system from which it was built. However, this model would be the most sensitive to changes in that system. On the other hand, a model can be built from a statistically significant sample of system instances, which can then be applied across the entire population of systems. However, this model will be less sensitive than the sample-specific model.

Once supplied with sufficient training data, the NormNet® tool builds a model that conforms to this data. This process combines an algorithm called an Expectation-Maximization (EM) with other specialty algorithms to automatically optimize the number and location of centers within the data-space. The baseline model is a compact representation of the relationships between the various parameters over the healthy system's operational envelope.

Upon completion, the model is used to assess data collected during operation of the target system. Each data sample (composed of n-channels) is compared to the baseline model to determine the most likely center. The data is then compared to the most likely data expected from that center under those conditions to determine the statistical plausibility of the incoming data. This information is fed into separate algorithm that determines the Probability of Abnormality (POA) of the system generating the data.

The POA algorithm uses a Bayesian hypothesis test to determine if a rolling window of data came from a healthy system. While the specifics of this algorithm are beyond the scope of this discussion, intuitively it is designed to look for repeated blocks of statistically implausible data across a consistent set of channels. This is based on the theory that when a system truly has a fault, that fault will manifest itself in some consistent, deterministic fashion. This assumption and resulting algorithm therefore provide for rejection of randomly generated false alarms. The POA algorithm outputs a value between zero and one, representing the probability that the underlying system is currently operating in an unhealthy condition. A POA of zero represents a high confidence that the system is healthy. A POA equal to one represents a high confidence that the system is unhealthy.

Because NormNet® performs analysis and prediction on a single sample, and assesses POA using a rolling window of data, it can be used for real-time analysis of streaming system data.

4 Pilot Project Methodology

An experiment was designed to test the hypothesis that the Pattern Recognition approach to early fault detection could demonstrate greater accuracy and false result rejection than CI Thresholds. Actual DSC and maintenance data collected from deployed aircraft was employed to generate case histories of a target subsystem. The effectiveness of both Pattern Recognition and CI Thresholds were then evaluated using those case histories to score the performance of the two approaches.

The steps employed for the pilot project were as follows:
1) Select a relatively simple target subsystem for which both sensor data and maintenance data were available.
2) Identify data channels for the target subsystem and down-select a subset of channels that meet the technical and statistical criteria of the experiment.
3) Generate case histories for the selected subsystem that integrate the selected channels and maintenance data.
4) Generate a Mixture Model from a subset of the case histories (the Training Data).
5) Test the Mixture Model using the remaining case histories (the Test Data).
6) Score the performance of the Mixture Model and accumulate the scores into a Performance Matrix.
7) Score the performance of the CI Threshold approach using the same cases and identical assumptions.
8) Compare the performance of the Mixture Model approach to that of CI Thresholds to address the goals of the analysis.

4.1 Source Data

The source data required to test a fault detection technique can be generated specifically for that experiment, or existing data can be exploited. The familiar bench test approach runs a component or subsystem under controlled circumstances, often with a seeded fault, to detect and characterize a fault signature. Measurements taken during the course of a bench test are used as a basis to test a FID technique. The advantage of the bench test approach is control – the fault is known; the scope of the test is controlled; the conditions are measured; the necessary data is collected. The disadvantages are that the component is tested in isolation rather than as part of a functioning system; and it is often prohibitively expensive to perform a statistically significant number of tests.

An alternative to the bench test is the use of a “virtual environment.” For the purpose of this discussion, the term “virtual environment” refers to a statistically significant population of actual, flight sensor data integrated with maintenance data. This is distinct from a “simulated environment” in that a virtual environment is not a mathematical construct, such as a Monte Carlo Model, but is actual flight and maintenance data collected from the field under real rather than modeled conditions. The advantages of a virtual environment are that a statistically significant population of cases can be employed; there is no additional expense required to collect the data (although one must still manipulate it into an appropriate form); and the full range of actual conditions and environments (including unexpected ones) are more likely represented in that data than in a bench test. The disadvantage is that the conditions of the experiment are not controlled – the set of data one might have liked to analyze is not necessarily available; descriptions of failure modes and maintenance actions are sometimes incomplete, inaccurate, or missing. Basically, you get what you get, and the methodology must be adapted to the available data.

For this analysis a virtual environment was employed. The CBM Data Warehouse, developed by QinetiQ North America, Inc. for Army Aviation, provided the data. The CBM-DW integrates DSC sensor data with maintenance data from both the DA-2410 system (installs and removes) and ULLS-A (maintenance log book). The CBM-DW employs a true dimensional structure, (as
opposed to a large file system or relational database that is sometimes referred to a data warehouse), hence it had the functionality required to extract and generate subsystem case histories with the requisite characteristics from the large volume of available data.

The source data was employed to generate case histories, each case representing one statistical trial. A case history is a chronology of a serial numbered instance of the target subsystem, beginning with installation (or first available data), and containing all the maintenance records and DSC data for that subsystem. The case ends when any part of the subsystem or the subsystem itself is removed from a vehicle, or no more data is available. Each case history was tested for the existence of an anomaly, and the associated maintenance data was examined to determine the nature and confirm the existence of that anomaly.

4.2 Component Selection

Since the experiment was a pilot analysis, the selection of a single, relatively simple target subsystem was desirable. From a number of candidate subsystems, the AH-64D Tail Rotor Gearbox Assembly was selected. This component has data visibility sufficient to support the analysis. Installations and removals are tracked by serial number (form DA-2410). The component has an assigned Work Unit Code (WUC 06G01) and maintenance actions are recorded. Additionally, the Tail Rotor Gearbox Assembly is instrumented with horizontal and vertical accelerometers and there are a number of CIs that measure its vibration.

The Tail Rotor Gearbox Assembly was selected with input from the U.S. Army Apache Project Management Office, because it met the criteria of the analysis plus there were other data and analyses available that could be compared to this analysis (by the Project Office). Additionally, there was only one configuration (corresponding to basic part number 7-311340001) for the time period of the analysis, obviating complications if such configurations behaved differently. The Tail Rotor Gearbox Assembly is pictured in Figure 4.2-1.
4.3 **Channel Selection**

A “channel” refers to any metric describing the performance, energy state, or environment of the target subsystem. A channel is not limited to vibration data. It can describe systems other than the target, whose performance correlates with that of the target. Or, it can characterize the environment in which the target is operating. An instance of a set of synchronous (or nearly synchronous) channel data forms a vector which is the input to the Pattern Recognition model.

Important constraints on the data required for the analysis were that the measurements contained in a vector must be synchronous, and all channels in a vector instance must be non-null (although the non-null constraint could be relaxed in future analyses). Additionally, there must be at least 50 vector instances in a case history. The AH-64D employs the MSPU digital source collection (DSC) system which does not collect data synchronously nor continuously throughout a use\(^1\). Rather, the MSPU has different modes that determine the frequency of collection and which channels are recorded. Data might be collected periodically throughout a use (Monitor Mode), or collection might be triggered by an event (Survey Mode). Moreover, during collection each sensor is polled in sequence, or only a particular subset of sensors might be polled. Since the Pattern Recognition approach requires synchronous (or nearly synchronous) data, the selection of channels employed in the model was constrained by these characteristics.

Examination of the Tail Rotor Gearbox sensor data revealed four, vibration related CI metrics that were more consistently available than other CIs derived from the sensors of interest. These CIs were not strictly synchronous. However, measurements were often recorded together in a 3 second window. Since these CIs represented the best approximation to synchronous vibration data that was available, they were treated as if they were synchronous for the purposes of this experiment. The results of the analysis subsequently supported this approximation.

Non-vibration metrics are referred to a parametric data. The available AH-64D parametric data includes 22 attributes that were often recorded during the same time intervals as the four selected CIs. Moreover, the frequency of data collection for the parametric data was sufficiently high that synchronization of the CI data with the parametric data was not a significant issue.

Vectors of the 26 measurements (four CIs plus 22 parametric metrics) were constructed from the CBM-DW in the following manner. Synchronous parametric data was associated with each CI measurement instance forming a vector of a single CI and 22 parametric measures. There are four CIs of interest, so the vectors associated with each of the four CIs occurring within a three second window were identified. These four vectors were then combined into single, vector by mapping the CI measures into the final vector and mapping the arithmetic mean of each parametric measure into the final vector.

---

\(^1\) I use the term “use” rather than “flight” since a particular period of operation may not involve the aircraft leaving the ground.
A list of the 26 metrics is given in Table 4.3-1. In the final model, the Main_Rotor_Speed, and Vertical_Speed (which turned out to be always null), were dropped. The CI attribute 64D-TR-22, which reports phase, was retained in the model but (not unexpectedly) did not provide predictive value. The earliest sensor data was dated 9/10/2009 and the latest 3/27/2011, which brackets the period of relevant data employed in this analysis.

Data employed in the analysis was collected from the AH-64D MSPU digital source collection (DSC) system as fielded during the 2009 through 2011 time period. The MSPU configuration, which specifies CI definitions and thresholds, is identified as Version 75.
Table 4.3-1 – Data Channels Included in Analysis

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>64D-TG-25</td>
<td>Tail gearbox lateral input shaft 1R vibration throughout the flight.</td>
</tr>
<tr>
<td>64D-TR-12</td>
<td>Tail rotor 1R vibration measured from tail gearbox lateral measured throughout the flight.</td>
</tr>
<tr>
<td>64D-TR-20</td>
<td>Tail rotor 1R vibration measured from tail gearbox vertical measured throughout the flight.</td>
</tr>
<tr>
<td>64D-TR-22</td>
<td>Tail rotor 1R phase measured from tail gearbox vertical measured throughout the flight.</td>
</tr>
<tr>
<td>AX</td>
<td>Longitudinal acceleration</td>
</tr>
<tr>
<td>AY</td>
<td>Lateral acceleration</td>
</tr>
<tr>
<td>AZ</td>
<td>Vertical acceleration</td>
</tr>
<tr>
<td>CAL_AIR_SPEED</td>
<td>Calibrated Airspeed</td>
</tr>
<tr>
<td>MAIN_ROTOR_SPEED</td>
<td>Main Rotor Speed (not employed in final model)</td>
</tr>
<tr>
<td>PA</td>
<td>Pressure Altitude</td>
</tr>
<tr>
<td>PITCH</td>
<td>Pitch</td>
</tr>
<tr>
<td>PITCH_ACCEL</td>
<td>Pitch acceleration</td>
</tr>
<tr>
<td>PITCH_RATE</td>
<td>Pitch Rate</td>
</tr>
<tr>
<td>ROLL</td>
<td>Roll</td>
</tr>
<tr>
<td>ROLL_ACCEL</td>
<td>Roll acceleration</td>
</tr>
<tr>
<td>ROLL_RATE</td>
<td>Roll Rate</td>
</tr>
<tr>
<td>TEMP_E1</td>
<td>Temperature Engine #1</td>
</tr>
<tr>
<td>TEMP_E2</td>
<td>Temperature Engine #2</td>
</tr>
<tr>
<td>TQ_E1</td>
<td>Torque Engine #1</td>
</tr>
<tr>
<td>TQ_E2</td>
<td>Torque Engine #2</td>
</tr>
<tr>
<td>VERTICAL_SPEED</td>
<td>Vertical Speed (not employed in final model)</td>
</tr>
<tr>
<td>VX</td>
<td>Longitudinal Air Speed</td>
</tr>
<tr>
<td>VY</td>
<td>Lateral airspeed</td>
</tr>
<tr>
<td>VZ</td>
<td>Vertical Airspeed</td>
</tr>
<tr>
<td>YAW_ACCEL</td>
<td>Yaw acceleration</td>
</tr>
<tr>
<td>YAW_RATE</td>
<td>Yaw Rate</td>
</tr>
</tbody>
</table>

4.4 Case History Generation

A case history is a chronology of a subsystem instance, in this case the Tail Rotor Gearbox Assembly. It begins with installation of a Tail Rotor Gearbox into a helicopter and contains all the qualified channel data from that point until the case is terminated by a maintenance action involving the removal of some part of the subassembly or the subassembly itself. Alternately, a case can terminate when no more data is available. A case history also contains records of all
maintenance and inspection actions not involving part removal occurring between the start and end dates of the case.

Each case history represents a single element of a statistical population. A case history serves as a trial that is evaluated for the presence (or absence) of an anomaly using both Pattern Recognition and CI Thresholds, and then confirmed by the existence (or absence) and timing of a relevant maintenance action. (Case evaluation is discussed more thoroughly in Section 4.6.)

Case histories were extracted from the CBM-DW subject to the constraints that 1) nearly synchronous CI and parametric data must exist; 2) there must be at least 50 vectors of channel data; and 3) maintenance data for the relevant time period must exist. Software was written to extract the required data from the CBM-DW, construct the necessary vectors, and integrate the maintenance data. Table 4.4-1 summarizes the quantities of data involved at each step of the case history generation process.

<table>
<thead>
<tr>
<th>Table 4.4-1 – Case History Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of separate instances of the four condition indicators</td>
</tr>
<tr>
<td>Number of separate instances of the four CIs for which synchronous parametric data was available</td>
</tr>
<tr>
<td>Number of complete groups of four CIs with parametric data within 3 second windows</td>
</tr>
<tr>
<td>Number of Tail Gearbox Case Histories constructed from the complete groups</td>
</tr>
<tr>
<td>Number of Case Histories with sufficient Sensor and Maintenance Data</td>
</tr>
<tr>
<td>Training Cases</td>
</tr>
<tr>
<td>Test Cases</td>
</tr>
</tbody>
</table>

As noted in Table 4.4-1, the process produced a statistically significant 302 cases. These cases were divided into two, independent groups. The first group, called the Training Data, contained no (known) faults. Hence, it represented normal behavior and was used to construct the Mixture Model. The second group is called the Test Data. It contained case histories depicting both anomalous and nominal behavior, the point being to test the model’s ability to tell these apart.

### 4.5 Model Generation

A Mixture Model was generated using the 129 case histories in the Training Data. This was accomplished using the NormNet® software installed on a quad-processor, Windows XP host running at 2.67 GHz. Model generation required about three days of run time. No significant software issues were encountered.

Figure 4.5-1, below, is a visualization of three of the 24 dimensions employed to generate the Mixture Model. The figure shows the modeled relationships among the Pressure Altitude,
Calibrated Airspeed, and Tail Gearbox Lateral Input Shaft 1R Vibration. The two views depict the rotation of the model in a counter-clockwise direction.

Figure 4.5-1 – Three Dimensions of the 24 Dimensional Tail Rotor Gearbox Mixture Model

The Mixture Model appears in Figure 4.5-1 as a set of intersecting planes, which is an approximation in three dimensions, of the nonlinear relationships among the three aforementioned channels. It represents the normal or expected behavior of the subsystem in these three dimensions. Just as a non-linear function in two dimensions can be approximated as set of line segments, in three dimensions the function is approximated as set of planes. Each plane is referred to as a center (or component), and represents the maximum likelihood estimates for a particular cluster of data. A cluster is a subset of the population that is strongly correlated and often has some functional explanation for this high correlation.
The graphic of Figure 4.5-1 is useful for performing a visual check of the modeled relationships. The check is accomplished by superimposing the source data over the expected values (depicted by the planes). The planes should capture the “central tendency” of the source data.

### 4.6 Case History Evaluation

Each case history in the Test Data was evaluated using Pattern Recognition, and also CI Thresholds, to determine if an anomaly was detected and identify the timing of that anomaly. The maintenance history of each case also was reviewed to ascertain the existence of a detectable fault and the timing of that fault. The timing and nature of detected faults were compared to the timing and nature of maintenance actions to score each case. Each scored case was classified into one of four categories, and the scores were accumulated into a Performance Matrix.

A performance matrix is a simple two-by-two matrix in which the True Positive, False Positive, False Negative, and True Negative case counts are accumulated, as shown in Figure 4.6-1. From these counts the average and overall success rates were calculated, as well as the Missed Fault Rate (Bayesian probability of detecting an actual fault) and the False Alarm Rate (Bayesian probability of a fault detection given that no fault exists).

<table>
<thead>
<tr>
<th>Case Counts</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>True Positive</td>
<td>False Negative (Missed Alarm)</td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>False Positive (False Alarm)</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

**Figure 4.6-1 – Performance Matrix**

The Performance Matrix classifications rules are summarized as follows.

1) An anomaly is considered detected by the Pattern Recognition approach if the Probability of Anomaly (POA) exceeds 75%. For CI Thresholds, an anomaly is detected if any CI exceedance threshold (“red” threshold), associated with the two Tail Rotor Gearbox sensors, is exceeded.

2) An Actual Anomaly exists if there is a maintenance action involving the adjustment or replacement of a relevant subsystem, component, or part resulting from a detectable fault (detectable fault defined below). An inspection is not an actual anomaly.

3) A True Positive exists if an anomaly is detected, and it is confirmed by the existence and timing of an actual anomaly. If channel data is available after the maintenance action, then the POA must show an immediate decline after that action.

4) A False Negative (Missed Alarm) occurs when an actual anomaly involving the Tail Gearbox or Tail Rotor exists, but it is not detected.
5) A False Positive (False Alarm) exists when an anomaly is detected but there is no Actual Anomaly.

6) A True Negative occurs when no anomaly is detected and there is no Actual Anomaly.

The 75% PoA threshold level employed by Pattern Recognition for this experiment was selected using sensitivity analysis to optimize the tradeoff between True Positive and False Negative in the context the actual case data.

A detectable fault, as mentioned in the definition of Actual Anomaly, is one that generates a signal that is detectable by a sensor. In general, this means faults that cause vibration in the Tail Gearbox Assembly and/or the Tail Rotor Assembly. (The inclusion of the Tail Rotor Assembly is discussed in Section 5 – Issues Affecting the Experimental Outcome.) Faults that are not detectable were excluded from the scoring. The rationale for this exclusion was that faults that produce no sensor signal cannot reasonably be counted as a failure of the algorithm to detect those faults. The purpose of the pilot was to test the efficacy of the algorithm, not the sensor hardware. Hence, Tail Rotor Gearbox and Tail Rotor faults that do not cause vibration such as leaking fluid, corrosion, external scratches and gouges on the gearbox, and small defects such as dents in the Tail Rotor Blade, were not counted as actual anomalies.

Faults that reasonably can be expected to produce a sensor signal are included in the performance counts. In particular, any fault that would cause vibration such as play in bearings or links; worn bushings; tail rotor balance; loose bolts, nuts, and screws; and any fault described in the maintenance record as producing vibration.

Pattern Recognition employs many channels of data, some of which describe the environment of the target rather than the target itself. Pattern Recognition detects anomalies sensed by any of these channels. For example, a sensor fault (e.g., flat-line) is a true fault, but of course, the sensor itself is not the target of the model. These types of faults were counted as True Positives. However, for purposes of comparison with the Threshold Limit performance, faults caused by anything other than the Tail Rotor Gearbox or Tail Rotor assemblies were excluded. Finally, failure to detect an anomaly outside the Tail Rotor Gearbox or Tail Rotor assemblies was not counted as a False Negative.

### 4.7 Example Case Evaluation – Pattern Recognition

The set of 173 case histories in the Test Data were individually evaluated and scored. In general, when an anomaly was detected it could be traced to the deviation of a particular channel from the expected value predicted for that channel by the Mixture Model. Two charts were generated for each anomaly. The first, an example of which is given in Figure 4.7-1 below, shows the expected value of the channel (the green line in the example), the actual value (shown in red), and the Probability of Anomaly (shown in blue). This chart has no time scale – the channel data is reported as if flight data were continuous across uses. The example illustrates channel data for the tail rotor 1R vibration measured from the tail gearbox vertical sensor throughout a flight. The actual data is clearly deviating from normal through the first portion of the chart. In response to the deviation, the Pattern Recognition Model generates an
increasing POA that eventually exceeds 75% and flags the anomaly as a fault. An actual fault was confirmed from the maintenance records, and the channel data abruptly declines into the normal range with timing that is consistent with that of the maintenance. This particular case involved a Tail Rotor delamination which was detected by the model 7 uses and 25 days before the fix documented in the maintenance log. The case was scored as True Positive.

![Figure 4.7-1 – Example Case Analysis (No Time Scale)](image)

The next chart, Figure 4.7-2 below, is the same data as in Figure 4.7-1, but is time scaled. Data collected during each use is compressed into a vertical bar representing the relatively short time period of a use compared to the overall time scale. This style chart was employed to compare the timing of maintenance actions to the timing of changes in channel data, in order to confirm that a maintenance action did, in fact, address the anomaly seen by the model.
5 Issues Affecting the Experimental Outcome - Lessons Learned

There were some interesting discoveries and lessons encountered during course of the pilot analysis. Constraints arising from the selection of the Tail Rotor Gearbox as the target of the analysis became apparent. The use of an LRU defined target, as opposed to a physics based target definition, and the use of actual maintenance data imposed limitation on the results that could be achieved by the analysis. There were also some important lessons regarding how to model small, but normal subpopulations.

5.1 Missing and Inaccurate Maintenance Records Drive up False Positive and False Negative

An important assumption employed to score the cases was that maintenance records were correct and complete. While it was sometimes obvious from comparison of maintenance records to corresponding sensor data that some maintenance records were missing or incorrectly diagnosed a fault, for the purpose of this analysis maintenance records were nevertheless assumed to be the “source of truth.” This assumption was required because there is no better alternative when actual field data is employed. The effect of the assumption was that both False Positive and False Negative rates were higher than they would otherwise be if perfect information were available. The False Positive rate was driven up when an actual fault occurred but the accompanying maintenance action was not available. The False Negative rate
was exaggerated when a fault was misdiagnosed as caused by a detectable Tail Gearbox or Rotor fault when it was, in fact, not.

Since the “Correct Maintenance Record” assumption was applied to the same set of cases used to evaluate both the Pattern Recognition and CI Threshold approaches, both were affected equally and the relative performance of the two systems could be determined. However, the actual values of the False Positive and False Negative rates were inflated (and the True rates correspondingly deflated) by unknown amounts.

5.2 Tail Rotor Gearbox Failures Rarely Detectable

The Tail Rotor Gearbox, unfortunately for this analysis, did not have many failures for vibration producing causes. To validate this observation, a review of 166 Tail Rotor Gearbox teardown analyses was conducted. This data including all AH-64 Tail Rotor Gearbox analyses with removal dates occurring in the period 09/20/2001 through 12/14/2011. Of the 166 cases only 14 (8%) mentioned vibration as a cause of removal, and only 5 (3%) were confirmed as vibration related. This rate of detectable failure is comparable to that found in the Test Data in which only 3 of 47 failures (6%) involved a vibration related failure. Of the 47 true anomalies counted in the Pattern Recognition Performance Matrix, 44 (94%) were attributable to sources other than the Tail Rotor Gearbox itself – mostly various sensor channel flat-line failures, and Tail Rotor System events.

5.3 Pressure Altitude Generates False Anomaly

The final Pattern Recognition model developed for this analysis employed 24 data channels, 20 of which describe something other than the Tail Rotor Gearbox itself. Hence, some anomalies detected by the Pattern Recognition model were for failures of subsystems other than the Tail Rotor Gearbox.

One of the channels employed in the model was Pressure Altitude. A small number aircraft in the sample population were stationed at altitudes that were significantly higher or lower than the normal range. The number of cases that included this phenomenon was not sufficient to have been recognized as normal during the model development process, resulting in seven cases being erroneously classified as anomalous (False Positive). This issue would be easily corrected though the addition of two centers to the model - one for normally high and another for normally low altitude. The model itself was not updated to include these additional centers due to time constraints. However, the Performance Matrix given in Table 7-1 assumes this correction.

5.4 Failures of Interest were Primarily Tail Rotor Faults

The target of this analysis was the Tail Rotor Gearbox Assembly. Hence, only data from the sensors installed in the gearbox were collected and analyzed. Vibrations from system elements picked up by the Tail Rotor Gearbox sensors include the Tail Rotor Assembly as well as the Tail Rotor Gearbox Input Shaft. Moreover, most of the actual anomalies of interest “seen” by the model involved the Tail Rotor rather than the Tail Rotor Gearbox itself. A physics based, as opposed to LRU based, definition of the target system would have produced a better result. In
particular, broadening the definition of the target system to include the Tail Rotor would have incorporated a third sensor located on the Tail Rotor Swashplate. The inclusion of the Tail Rotor sensor, designed to detect Tail Rotor anomalies, which were in fact what the model found, would presumably have produced a better model.

5.5 Sensor Channel Failures

A significant number of actual anomalies were traced to various flat-lined sensor data – most often the engine temperature and torque channels. Flat-line failures accounted for 28 (60%) of the 47 true anomalies detected by Pattern Recognition. These failures were counted as True Positive in the Pattern Recognition performance matrix provided in Table 7-1. However, flat-line channel failures were excluded from the comparison of Pattern Recognition to the CI Threshold approaches.

5.6 Experiment Actually Measured the Tail Rotor System

The Test data contained 173 cases, of which 47 contained at least one true anomaly. Of these 47, 44 or 94% are related to sources other than the Tail Rotor Gearbox. In fact, only one of nine False Negative outcomes and none of the False Positive outcomes actually related to the Tail Rotor Gearbox.

The intent of the analysis was to evaluate the performance of Pattern Recognition versus CI Thresholds using the Tail Rotor Gearbox as the target subsystem. However, given the low number of Tail Rotor Gearbox failures and actual sources of failures detected by both approaches, what the analysis actually accomplished was a measurement of the ability of Pattern Recognition and CI Thresholds to detect Tail Rotor anomalies using only Tail Gearbox sensors.

6 Distribution of Actual Failures in Test Dataset

The distribution by cause of anomalies documented in the Test Data is shown in Figure 6.0-1. Various channel flat-line failures accounted for 28 (60%) of the 47 true anomalies. The next largest group, with 15 occurrences, was associated with the Tail Rotor. The actual Tail Rotor Gearbox accounted for only 3 of the anomalies. All 47 True Anomalies were included in the Pattern Recognition Performance Matrix provided in Table 7-1. However, True Anomalies that could not reasonably be expected to be detected using CI Thresholds (i.e., sensor flat-lines) were excluded from the comparison with Pattern Recognition. The number of True Anomalies that could have been detected by CI Thresholds is 18. A reconciliation of the total counts reported in the performance matrices for Pattern Recognition and CI Threshold is provided in Appendix A.
 Experimental Results

There are two parts to the experimental results. The first part is the Performance Matrix resulting from the Pattern Recognition approach. Secondly, there is a comparison of Pattern Recognition versus CI Threshold approaches evaluated using identical case histories and assumptions.

As mentioned in Section 5.1, both the Pattern Recognition and CI Threshold results reflect higher False Alarm (False Positive) and Missed Fault (False Negative) rates (and correspondingly lower true rates) than would be achieved with more complete maintenance information. Additionally, the exclusion of the Tail Rotor sensor from the analysis, presumably, also drives the false rates up. These limitations of the experiment constrain the interpretation of the data. The False Positive and False Negative rates should be viewed as upper bounds rather than absolute measures of performance. However, since both Pattern Recognition and CI Thresholds were evaluated under identical conditions and assumptions, it is reasonable to employ the results for comparison of the two approaches.

Note also that the Pattern Recognition performance has been corrected for Pressure Altitude (as discussed in Section 5.3).

Table 7-1, below, reports the absolute results for Pattern Recognition and includes sensor channel flat-line failures. A meaningful comparison of Pattern Recognition with CI Thresholds requires the exclusion of sensor channel failures, so Table 7-2 provides the same data less the sensor channel failures. Table 7-3 depicts the performance of CI Thresholds employing the same cases histories and assumptions as that of Pattern Recognition.
Table 7-1 – Performance Matrix, Pattern Recognition  
(Missed Fault and False Alarm are Upper Bounds)

<table>
<thead>
<tr>
<th>Performance Matrix - Key</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>True Positive</td>
<td>False Negative (Missed Alarm)</td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>False Positive (False Alarm)</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case Counts and Accuracy with Correction for Altitude</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>38</td>
<td>9</td>
<td>47</td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>8</td>
<td>118</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>127</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bayesian Probability of Detected State given Actual State</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>Missed Fault Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>0.809</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>0.063</td>
<td>0.937</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bayesian Probability of Actual State given Detected State</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>False Alarm Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>0.826</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>0.174</td>
<td>0.929</td>
<td></td>
</tr>
</tbody>
</table>

Table 7-2 – Performance Matrix, Pattern Recognition Excluding Sensor Channel Failures  
(Missed Fault and False Alarm are Upper Bounds)

<table>
<thead>
<tr>
<th>Case Counts Excluding Sensor Channel Failures</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>13</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>8</td>
<td>118</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>127</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bayesian Probability of Detected State given Actual State</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>Missed Fault Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>0.591</td>
<td>0.409</td>
<td></td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>0.063</td>
<td>0.937</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bayesian Probability of Actual State given Detected State</th>
<th>Detected Anomalies</th>
<th>Detected Nominal</th>
<th>False Alarm Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Anomalies</td>
<td>0.619</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>Actual Nominal</td>
<td>0.381</td>
<td>0.929</td>
<td></td>
</tr>
</tbody>
</table>
8 Analysis

A comparison of the detection rates for Pattern Recognition versus CI Thresholds, achieved for the target subsystems, is presented in Figure 8-1. The comparison, of course, excludes sensor channel failures. The figure shows an Actual Anomaly Accuracy for Pattern Recognition that is about 2.5 times higher than that achieved with CI Thresholds. The CI Thresholds employed for the Tail Rotor Gearbox and Tail Rotor are intentionally configured to relatively high values in order to reduce the number of false alarms. However, the accuracy of Pattern Recognition outperformed that of CI Thresholds by a substantial margin. The False Rate comparison given in Figure 8-2 shows Pattern Recognition with a rate half that experienced with the CI Thresholds. One could infer that a reduction in the CI Threshold values to produce better accuracy would be offset by an increase in the false rates. Thus a lowering of CI Thresholds to improve accuracy would increase the margin by which Pattern Recognition outperformed CI Thresholds in rejection of false indications.

A one-tailed Difference of Proportions Test, using non-pooled variance, showed that the difference in False Alarm Rejection rates between Pattern Recognition and CI Thresholds was significant at the 5% level, but not at 1%. The CI Threshold approach produced only 4 True Positive results, necessitating the direct calculation of significance for the True Positive rate using the Binomial Distribution. Assuming a Null Hypothesis that the True Positive rate for CI Thresholds is equal to that of Pattern Recognition, there was less than a 1% chance of 4 or fewer True Positive instances. Hence, the difference in the True Positive rate between Pattern Recognition and CI Thresholds was significant at a level greater than 1%.
9 Conclusion

This pilot analysis compared the effectiveness of Pattern Recognition to CI Thresholds for early detection of vibration inducing faults in the AH-64D Tail Rotor Gearbox and Tail Rotor. Pattern Recognition uses a fundamentally different approach to early fault detection than CI Thresholds. CI Thresholds, as currently deployed, employ a physics based approach in which static upper limits are set for each CI metric and each CI is computed as a function of a single sensor’s data. This contrasts with the multi-dimensional, Pattern Recognition approach in which abnormality is determined by comparing an actual set of measures to a predicted set of measures representing normal behavior as determined by the model.
The pilot project demonstrated substantially better performance for the Pattern Recognition approach than produced by CI Thresholds in the Tail Rotor Gear Box and Tail Rotor when compared using identical case histories and assumptions. Pattern Recognition produced both better accuracy and false result rejection than CI Thresholds in the target subsystems, as fielded at the time of this analysis.

There were not a sufficient number of True Positive cases generated by the CI Threshold approach to draw any inferences regarding whether or not one approach detected anomalies earlier than the other.

While the experiment was successful in measuring the relative performance of the two approaches, the absolute rates measured by the experiment were of less significance. False positive and negative rates were affected by inaccuracies in the actual maintenance log data. True positive rates were constrained by the exclusion of the Tail Rotor Sensor, when in fact most of the anomalies detected in the experiment traced to the Tail Rotor Assembly rather than the Tail Rotor Gearbox. These limitations reflect faults in the experimental design and accuracy of the source data rather than the Pattern Recognition or the CI Threshold approaches to anomaly detection.
Appendix A – Reconciliation of Performance Matrix Tables

Table 7-1 depicts the performance of Pattern Recognition. There are 47 actual anomalies counted in this matrix. Table 7-3 depicts the performance of CI Thresholds and shows 18 actual anomalies. A reconciliation of the actual anomaly counts between Tables 7-1 and 7-3 is given below.

Table A-1 – Reconciliation of True Positive Counts in Tables 7-1 and 7-3

<table>
<thead>
<tr>
<th></th>
<th>Total Actual Anomalies counted in CI Threshold Performance Matrix (Table 7-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+18</td>
<td>Sensor Flat-line anomalies excluded from CI Threshold Performance Matrix</td>
</tr>
<tr>
<td>+25</td>
<td>Sensor Flat Line anomalies retained in CI matrix due to other CI exceedances</td>
</tr>
<tr>
<td>+3</td>
<td>Pattern Recognition is False Negative, but CI Threshold is False Positive</td>
</tr>
<tr>
<td>+1</td>
<td>Pattern Recognition is False Negative, but CI Threshold is False Positive</td>
</tr>
<tr>
<td>47</td>
<td>Total Actual Anomalies counted in Pattern Recognition Performance Matrix (Table 7-1)</td>
</tr>
</tbody>
</table>

Tables 7-2 and 7-3 depict the performance matrices for Pattern Recognition and CI Thresholds that were employed to compare the two approaches. Both show a total of 148 cases, which reflects the 173 cases in the Test Dataset (and Table 7-1) less 25 cases classified as True Anomalies solely because of sensor flat-line. These 148 cases are divided between Actual Anomalies and Actual Nominal cases. However, the division between these two categories is not the same for both Tables 7-2 and 7-3. Some cases moved between the “actual” classes because of multiple phenomena in each case, the nature of the phenomena, and the success of the methods in detecting those phenomena. For example, in Case 9 Pattern Recognition detected an anomaly but the PoA failed to reach 75% before the issue was fixed, hence the case was scored as False Negative. In the same case for CI Thresholds, multiple exceedances were recorded, but were not consistent with the timing of a maintenance action, thus the case was scored as False Positive.
References


Acronyms

CBM-DW: Condition Based Maintenance Data Warehouse
CI: Condition Indicator
DSC: Digital Source Collector
FID: Fault Identification
HUMS: Health and Usage Monitoring System
MSPU: Modern Signal Processing Unit
PHM: Prognostic and Health Management
POA: Probability of Anomaly