Bayesian Extreme Value Statistics for Novelty Detection in Gas-Turbine Engines

David A. Clifton
Dept. of Engineering Science, University of Oxford, Oxford, UK, OX1 3PJ
Oxford BioSignals Ltd., 174F Milton Science Park, Abingdon, UK, OX14 4SE
01235 433578
davidc@robots.ox.ac.uk

Nicholas McGrogan
Oxford BioSignals Ltd.
174F Milton Science Park
Abingdon, UK, OX14 4SE
01235 433574
nicholas.mcgrogan@oxford-biosignals.com

Lionel Tarassenko
Dept. of Engineering Science
University of Oxford
Oxford, UK, OX1 3PJ
01865 273113
lionel@robots.ox.ac.uk

Dennis King, Steve King, Paul Anuzis
Rolls-Royce PLC
PO Box 31,
Derby, UK, DE24 8BJ
{dennis.king, steve.king, paul.anuzis}@rolls-royce.com

Abstract—We present a novel method for the identification of abnormal episodes in gas-turbine vibration data, in which we show 1) how a model of normal engine behaviour is constructed using signatures of "normal" engine vibration response; 2) how extreme value theory (EVT), a branch of statistics used to determine the expected value of extreme values drawn from a distribution, can be used to set novelty thresholds in the model, which, if exceeded, indicate an "abnormal" episode; 3) application to large data sets of modern gas-turbine flight data, which shows successful novelty detection results with low false-positive alarm rates.

The advantages of this approach over previous work are 1) a very low false-positive alarm rate, while maintaining sufficient sensitivity to detect known abnormal events; 2) the use of a Bayesian framework such that uncertainty in the distribution of "normal" data is modelled, giving a principled, probabilistic interpretation of results; 3) an implementation that is sufficiently "lightweight" in processing and memory resources that real-time, on-line novelty detection is possible in an "on-wing" engine health-monitoring system.[1][2]

TABLE OF CONTENTS

[1] 1-4244-1488-1/08/$25.00 ©2008 IEEE
[2] IEEEAC paper#1461, Version 3, Updated 2007:12:10

1. INTRODUCTION

Novelty Detection

Modern aero engines are designed to be extremely reliable and robust, typically operating for many thousands of hours before requiring major overhaul. High product reliability and the ability to minimise unplanned equipment downtime are therefore significant factors in service support required to guarantee availability throughout operational life. A major component of this capability is the provision of diagnostic and prognostic tools through advanced health monitoring. Condition monitoring, employing novelty methods, is therefore a key aspect of this process.

In novelty detection (or one-class classification), a model of normality is constructed from "normal" data with significant deviations from that model classified as "abnormal" [1]. In the field of condition monitoring this technique is particularly well-suited to the identification of event precursors in data sets where the number of abnormal examples is too small to adopt a conventional fault-detection (multi-class classification) approach. This is a typical scenario in high-integrity systems such as gas-turbine engines, in which faults are very rare in comparison with long periods of normal operation [2].

Existing Work

Current methods of novelty detection in aerospace gas-turbine engine data suffer from a number of disadvantages:

1) A *novelty threshold* is set on the model output such that data exceeding this threshold are deemed "abnormal" – this is the decision boundary. The setting of this novelty threshold often relies on a heuristic approach [1]. Such heuristics are typically derived from engineering experience, and have no principled, quantifiable measure of the

likelihood of false detections (which is required for making business decisions regarding the operation of condition monitoring). This reliance on heuristic expertise also makes transferral of knowledge between engineering projects difficult – typically, the setting of novelty thresholds is performed in a heuristic, time-consuming manner for each new engine design. This also precludes the use of the system in an automated on-line setting, as the model's novelty thresholds cannot be set without human intervention [3].

2) Many high-integrity systems are susceptible to the presence of noisy data, which can result in undesirable false-positive novelty detections. In the analysis of gas-turbine engines, for example, the cost associated with a false-positive alarm is particularly high, and must be reduced as much as possible while still being sufficiently sensitive to detect episodes of abnormality [2].

3) Existing approaches require an explicit training period, sometimes over 50 flights in duration for aircraft engines [4,5], during which the model cannot perform novelty detection. It is desirable that novelty detection is performed as early as possible, ideally from the entry of the system into service.

This paper presents the results of an investigation into a new method of novelty detection that overcomes these disadvantages.

Paper Overview

In Section 2, we describe the vibration data domain used in the investigation described by this paper, introducing challenges for the novelty detection process that arise due to the nature of aerospace gas-turbine vibration. Section 3 describes a proposed Bayesian extension to extreme value statistics that allows the setting of novelty thresholds to be performed in a principled, probabilistic manner, and which allows novelty detection to take place during model training. The application of the technique to several large data sets of vibration data recorded from a modern gas-turbine is described in Section 4, and results of novelty detection are presented. Finally, Section 5 discusses the advantages and limitations of the proposed technique, and introduces topics for further research.

2. Gas-Turbine Vibration Data

Modern aerospace gas-turbine engines divide the task of air compression from atmospheric pressure to that ultimately required within the combustion chamber into several stages. Many gas-turbine engines within the civil aerospace market involve three consecutive compression stages: the low pressure (LP), intermediate pressure (IP), and high pressure (HP) stages [6]. Air passes through each stage as it travels from the front of the engine to the rear, being further compressed by each, until it reaches the combustion chamber.

Each of the compressor stages is driven by its own turbine assembly, resulting in three corresponding turbine units situated within the exhaust stream at the rear of the engine. Each compressor is linked to its corresponding turbine by a separate shaft, which are mounted concentrically. In three-compressor engines, these are named the *LP* shaft, the *IP* shaft, and the *HP* shaft. The operating point of the engine is often defined in terms of the rotational speed of these shafts [7].

Transducers are mounted on various points of the engine assembly for the measurement of engine vibration. Vibration data used for investigations described in this paper were acquired using the QUICK acquisition system [8] that computes spectral representation of engine vibration via high-resolution FFTs at rate 5 Hz, for each sensor output. Engine vibration is assumed to be pseudo-stationary over this measurement period such that the generated FFTs may be assumed to be close approximations of actual engine vibration power spectra.

Tracked Orders

A *tracked order* is the fundamental data type used in the investigation described by this paper, and is defined to be the amplitude of engine vibration measured within a narrow frequency band centred on the fundamental or a harmonic of the rotational frequency of a shaft [4]. During normal engine operation, most vibration energy is present within tracked orders centred on the fundamental frequency of each rotating shaft; we term these the fundamental tracked orders. Using the terms LP, IP, and HP to refer to engine shafts, we define fundamental tracked orders associated with those shafts to be *1LP*, *1IP*, and *1HP*, respectively.

Vibration Signatures

We define a *vibration signature* to be the vibration amplitude or phase of a tracked order measured over a range of speeds of the corresponding shaft. For example, a signature may be constructed from 1HP vibration data measured as a function of the speed ω of the HP shaft.

Tracked order vibration is measured across the speed range ω = [0 100] % maximum speed of the corresponding shaft. This speed range is subdivided into B equal bins, defining B sub-ranges of ω. Within each bin $b = 1 \ldots B$, vibration amplitude or phase values observed within the corresponding sub-range of shaft speeds are collected.

Figure 1 shows an example vibration signature, in which amplitude of vibration $|x|$ is shown against shaft speed ω, for B = 400 bins on the horizontal axis. Mean vibration amplitude μ can be seen to be generally increasing with shaft speed, while maximum variance of $|x|$ occurs at approximately ω = 50%.

In this paper, we consider only vibration amplitude; the use of phase is considered as future work in Section 5.

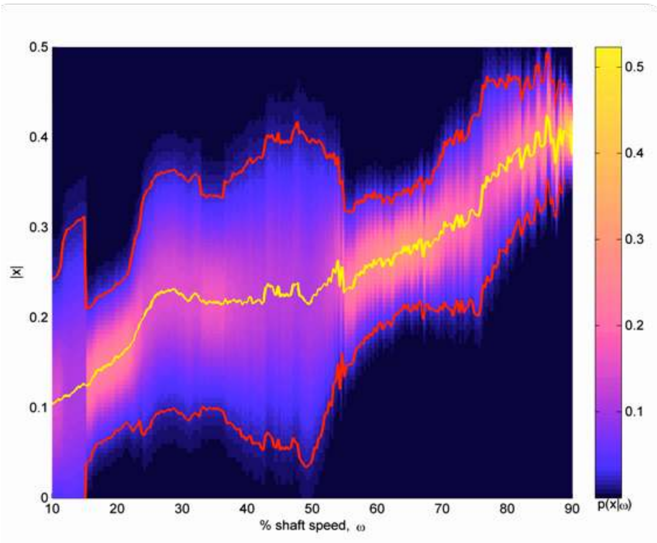


Figure 1 - A vibration signature showing vibration amplitude |x| against shaft speed ω. Mean vibration amplitude μ is shown as a yellow line plotted through the centre of each distribution. Example novelty thresholds are shown as red lines above and below μ.

Existing Use of Vibration Signatures

Previous approaches construct a speed-based vibration signature from vibration values observed during a single engine run, which could be either a single flight or a single ground-based test cycle [4,5,7,9]. The vibration values in each of the B bins are averaged, resulting in a vector of B average vibration amplitudes for each engine run. A model of normality is then constructed from these B-dimensional vectors, using clustering methods [10] or a Support Vector Machine [4]. This has been shown to allow the separation of vibration signatures from "normal" and "abnormal" flights.

However, these methods suffer from the three disadvantages noted in Section 1 (reliance on heuristic expertise, potential for over-sensitivity resulting in false alarms, and the requirement for a training period during which novelty detection does not take place). Furthermore, by only using the mean vibration value in each speed bin, useful information regarding the full distribution of vibration data is discarded.

For the purposes of the investigation described by this paper, we wish to consider the full distribution of vibration values within each of the B speed bins. For each bin b, we will find lower and upper novelty thresholds ($h_{\min}$ and $h_{\max}$, respectively) such that if a flight contains vibration amplitudes falling outside the range [$h_{\min}$ $h_{\max}$], it will be classified "abnormal". Example novelty thresholds are shown in Figure 1 as red lines.

3. THEORY

This section describes the proposed method of setting novelty thresholds in each speed bin b such that the previously-identified disadvantages are overcome.

First, conventional approaches to setting novelty thresholds are described, which are seen to result in oversensitivity to noisy in the "normal" data.

Secondly, the field of extreme value theory (EVT) is presented as a statistical method for setting novelty thresholds that can overcome this sensitivity to noise, but which has further disadvantages of its own when less than 50 engine runs of data are available.

Finally, a novel Bayesian extension to extreme value statistics is presented that overcomes this requirement for large amounts of data by modelling uncertainty in our estimate of the data distribution. This allows novelty thresholds to be set using as few as 2 engine runs of data, and which adapt to new data as more engine runs are completed.

Conventional Methods for Setting Novelty Thresholds

Given the distribution of vibration amplitudes within a speed bin b, conventional methods of setting novelty thresholds could be used in which a statistical model is fit to the data, giving an estimate of the underlying data distribution $p(x)$. Approaches to this vary from parametric approaches, in which the data are assumed to be generated from a known distribution (such as the Gaussian distribution), to semi-parametric approaches, such as using a Gaussian Mixture Model [11]. The novelty threshold is then set in the tail of this distribution $p(x)$, often using the cumulative distribution such that $P(x) \leq 1\text{-}H$, where H is the probability of observing an abnormal event (e.g., $H = 10^{-6}$).

An example is shown in Figure 2, in which a Gaussian distribution has been fitted to "normal" vibration data in a speed bin b, and an upper novelty threshold $h_{\max}$ set in its tail, here using the cumulative probability distribution $P(x) \leq 1\text{-}10^{-6}$.

However, because the novelty threshold is set far into the tail of the distribution, it occurs where $p(x)$ takes very small values, and where the gradient of $p(x)$ and $P(x)$ is almost zero. Thus, small changes in the "normal" data can result in significant changes in the location of the novelty threshold, making it oversensitive to noise.

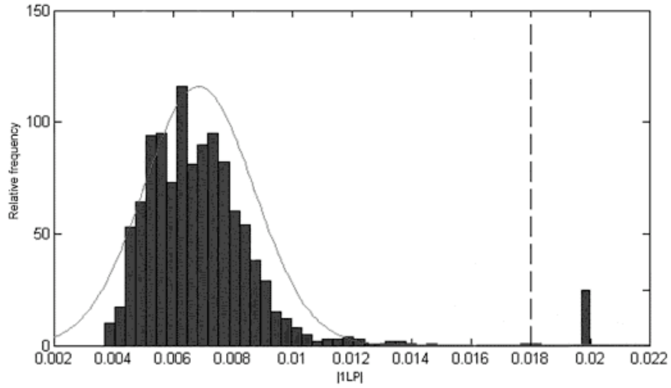


Figure 2 – Vibration amplitudes within an example speed bin, used to fit a Gaussian distribution (curved line). A novelty threshold has been set in its tail (dotted line).

Extreme Value Statistics for Setting Novelty Thresholds

Extreme value statistics is a branch of statistics that effectively models the tails of distributions, describing where extreme values drawn from the distribution of "normal" data are expected to lie.

One of the fundamental theories in "classical" extreme value statistics, the Fisher-Tippett theorem [12], states that if we draw $\mathbf{X} = \{x_1 \ldots x_m\}$ samples from a Gaussian distribution $p(x) = N(\mu, \sigma^2)$, the probability distribution $p_E(x)$ describing where we expect the most extreme of those m samples to lie tends towards the Gumbel distribution:

$$p_E(y_m) = \exp(-\exp(-y_m)) \tag{1}$$

which we term the Extreme Value Distribution (*EVD*), where y_m is termed the *reduced variate*,

$$y_m = \frac{x' - c_m}{d_m} \tag{2}$$

for normalised data $x' = (x - \mu) / \sigma$, and where c_m and d_m are, respectively, the location and scale parameters [13]

$$\begin{aligned} c_m &= \sqrt{2\ln m} - \frac{\ln\ln m + \ln 4\pi}{2\sqrt{2\ln m}} \\ d_m &= \left(\sqrt{2\ln m}\right)^{-1} \end{aligned} \tag{3}$$

In the same way that conventional methods can use the cumulative distribution $P(x)$ to set a novelty threshold at $P(x) \leq H$ (for some threshold probability H), so the cumulative distribution $P_E(x)$ associated with (1) can be used to set a novelty threshold [14]:

$$P_E(x) = \exp\{-\exp(-y_m)\} \tag{4}$$

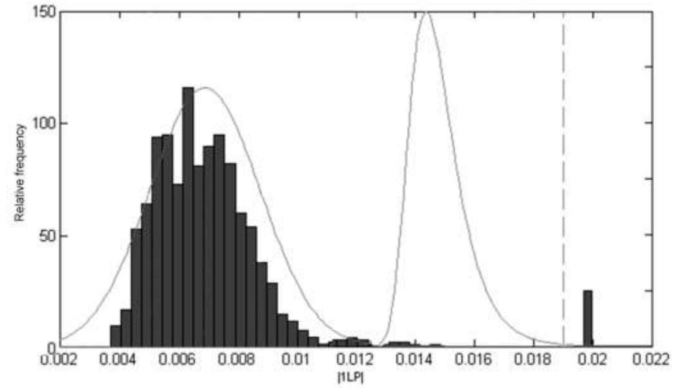


Figure 3 – The distribution $p_E(x)$, shown as the right-most curve, describes where the most extreme of m = 100 samples drawn from the data distribution should lie. A novelty threshold is set in the tail of $p_E(x)$ which is tolerant to noise in the "normal" data.

An example is shown in Figure 3, in which $p_E(x)$ describes our belief in where the most extreme of m = 100 points drawn from the distribution of the "normal" data will lie. A novelty threshold has been set at $P_E(x) \leq 1\text{-}10^{-6}$.

This novelty threshold set using extreme value statistics is tolerant to noise, because it depends only on the number of samples drawn m (which is independent of the "normal" data), and the parameters μ and σ, which are insensitive to small changes in "normal" data. Thus, extreme value statistics can overcome the disadvantage encountered by the use of conventional methods in setting novelty thresholds.

Disadvantages with Classical Extreme Value Statistics

In practice, the distribution of vibration amplitudes within a speed bin b is often not Gaussian. In both flight data and ground-based data, the distribution can be multimodal [15]. However, following [16], rather than using the entire data set, if we instead only consider the N = 3 maximum and N = 3 minimum vibration amplitudes in a speed bin for each engine run, the resultant maxima and minima distributions are approximately Gaussian, and tend towards the Gumbel distribution (1), as shown in Figure 4.

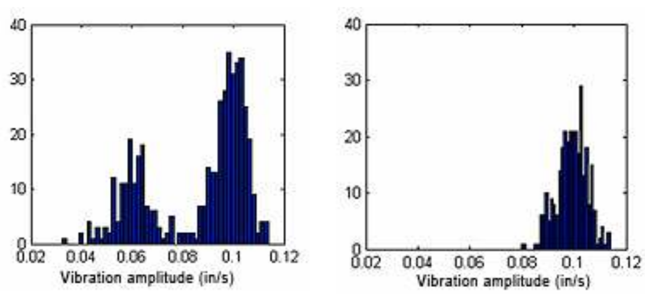


Figure 4 – The data distribution for all vibration amplitudes within an example speed bin, across all engine runs, are bimodal (left). The distribution formed from taking the N = 3 maxima from each engine run is approximately Gaussian (right).

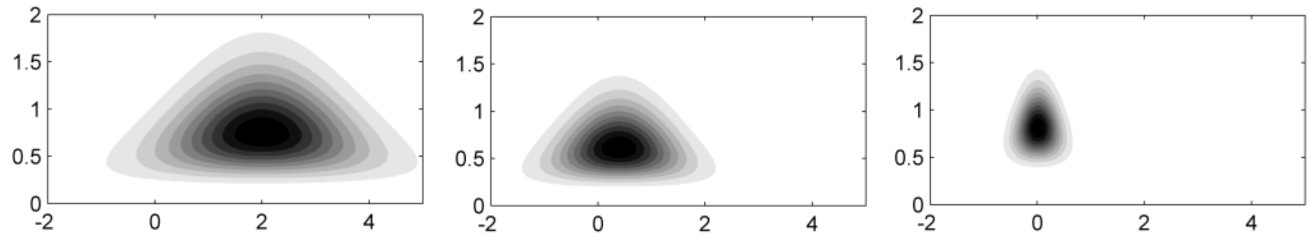


Figure 5 – Distributions $p(\mu|\mathbf{X})$, shown on the horizontal axis, and $p(\lambda|\mathbf{X})$, shown on the vertical axis, give the joint probability distribution $p(\mu,\lambda|\mathbf{X})$. From left to right, the joint distribution is plotted for 0, 10, and 25 observed data x. As more data x are observed, the joint distribution becomes more peaked around the true values of μ and λ.

Thus, we find speed bin b's upper novelty threshold $h_{\max}$ using the distribution of maxima (formed by taking the $N = 3$ maxima from bin b in each engine run), and likewise find the lower novelty threshold $h_{\min}$ using the distribution of minima (formed by taking the $N = 3$ minima from bin b in each engine run).

A second disadvantage of classical extreme value statistics is that the resultant novelty threshold is correct only if we have a reliable estimate of parameters μ and σ. This is because the novelty threshold is dependent on $P_E(x)$, which is ultimately dependent on μ and σ (via y_m and x'). In previous work [17,18], the maximum likelihood (*ML*) measures μ_{ML} and σ_{ML} are used to estimate μ and σ; i.e., the mean and standard deviation of the "normal" data observed so far. In an on-line context, these ML measures are good estimates of the true μ and σ when large numbers of data are observed, but can be very poor estimates when only smaller numbers of data are available. This is particularly true for application to distributions of maxima (or minima): if each engine run results in $N = 3$ new observed values, the ML measures may only be accurate estimates of the true μ and σ after a large number of engine runs, and thus the novelty threshold may be unreliable for small numbers of runs.

Bayesian Extreme Value Statistics

In order to overcome this problem associated with unreliable estimates of the parameters μ and σ, we set the problem within a Bayesian framework, explicitly modelling the uncertainty in our estimates of μ and σ.

We form probability distributions $p(\mu)$ and $p(\lambda)$ that model our belief in the value of the parameters of the data distribution, μ and λ (where we used precision $\lambda = 1/\sigma^2$ for later notational convenience). That is, instead of a single pair of ML estimates μ_{ML} and σ_{ML}, which may be inaccurate, we now consider a range of possible values for μ and λ.

These distributions $p(\mu)$ and $p(\lambda)$ can be updated online as we observe more data $\mathbf{X}$, using Bayesian updating [19]. That is, after observing data $\mathbf{X}$, our current belief in the values of μ and λ is given by the distributions $p(\mu|\mathbf{X})$ and $p(\lambda|\mathbf{X})$. An example is shown in Figure 5, in which an initial *prior* estimate for $p(\mu)$ and $p(\lambda)$ is shown in the left subplot, noting that this joint distribution is $p(\mu,\lambda|x)$. As more data $\mathbf{X}$ are observed, $p(\mu,\lambda|\mathbf{X})$ becomes more peaked around the true values of μ and λ. Thus, for large numbers of observed data, $p(\mu,\lambda|\mathbf{X})$ tends towards the ML estimate – but for smaller numbers of data, when the ML estimate may be poor, a wider range of possible μ and λ values is considered.

Our goal is to find a final EVD $p_F(x)$, after observing data $\mathbf{X}$, that takes into account the current range of possible μ and λ values. This can be achieved by integrating (1) over all values of μ and λ:

$$\begin{aligned} p_F(y_m) &= \int_{-\infty}^{\infty}\int_{0}^{\infty} p_E(y_m)\, p(\mu,\lambda\,|\,\mathbf{X})\; d\lambda\; d\mu \\ &= \int_{-\infty}^{\infty}\int_{0}^{\infty} \exp\left(-\exp\left(-y_m\right)\right) p(\mu,\lambda\,|\,\mathbf{X})\; d\lambda\; d\mu \end{aligned} \tag{5}$$

which has no solution in closed form. However, we can write the integration as a weighted sum individual EVDs $p_E(x)$

$$p_F(y_m) = \sum_{i=1}^{S} w_i\, p_E(y_m) \tag{6}$$

where each of the S weights is determined by sampling the joint distribution $p(\mu,\lambda|\mathbf{X})$. This method is computationally lightweight and accurately provides the final EVD $p_F(x)$.

In order to set the novelty threshold as before, we require the cumulative distribution $P_F(x)$, which from (6) will be a linear sum of individual cumulative distributions $P_E(x)$, given by (4):

$$\begin{aligned} P_F(y_m) &= \int_{-\infty}^{\infty} p_f(y_m)\, dy_m \\ &= \int_{-\infty}^{\infty}\left[\sum_{i=1}^{S} w_i\, p_E(y_m)\right] \\ &= \sum_{i=1}^{S}\left[w_i \int_{-\infty}^{\infty} p_E(y_m)\right] \\ &= \sum_{i=1}^{S} w_i P_E(y_m) \end{aligned} \tag{7}$$

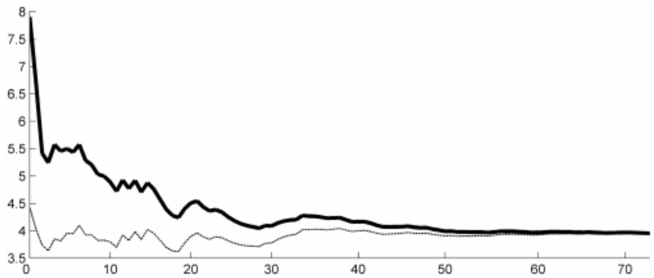


Figure 6 – Novelty thresholds h_{max} (vertical axis) for an example data set of increasing number of engine runs (horizontal axis), comparing the threshold for the proposed Bayesian method (thick line) with the threshold set using only the ML parameter estimates (thin line). Units have been anonymised.

Finally, we set the novelty threshold using (7) such that $P_F(x) \leq 1 - H$.

A comparison of setting novelty thresholds using the proposed Bayesian-EVT method and the existing ML-based method is shown in Figure 6, using an example data set. It can be seen that the proposed method sets novelty thresholds high when only small numbers of engine runs are available. At this point, there is maximum uncertainty in our estimates of the true parameters of the data distribution μ and λ, and so the thresholds are set at their most conservative. As more engine runs are completed, and we observe more data, the uncertainty in our estimate decreases, and the novelty threshold similarly becomes less conservative, reflecting our increased confidence in our estimates.

In contrast, the ML-based method does not take into account this uncertainty for smaller numbers of engine runs, and so sets novelty thresholds much lower – this increases the probability of the ML-based method generating false-positive alarms, because the estimates may be poor for smaller numbers of engine runs.

Note that, for large numbers of engine runs, the proposed method and the ML-based method converge, as desired.

We note that previous approaches to using Bayesian methods with EVT have been restricted to direct estimation of the location and scale parameters c_m and d_m [20], without accounting for the uncertainty in the underlying data distribution μ and λ, which is necessary for the purposes of on-line novelty detection, and does not result in the behaviour of the novelty threshold shown in Figure 6.

4. Application to Flight Data

This section describes the application of the proposed Bayesian extension to EVT for novelty detection in several large data sets derived from flights of modern civil gas-turbine engines. Results from this retrospective analysis are presented comparing the performance of the proposed method to the existing ML-based method, and to a heuristic method currently used by domain experts for engine vibration analysis.

Data Description

The data sets considered within the investigation described by this paper fall into two groups:

A) Abnormal data, which contain known engine events, and precursors to those engine events.

B) Normal data, which contain no known engine events.

Thus, data from the first group are used to test the *sensitivity* of the various novelty detection methods (ensuring that they can correctly identify engine events, and precursors to those engine events), while data from the second group are used to test the *specificity* (ensuring that the false-positive alarm rate is sufficiently low).

Table 1 describes the data sets used in this investigation (divided into groups A and B, described above), each of which was recorded during flights of different engines. The average length of each flight is greater than 1 hour, resulting in a total data set size of greater than 449 hours.

The data were obtained as described in Section 2. Thus, each flight resulted in three speed-based signatures of vibration amplitude, corresponding to the amplitude of the three fundamental tracked orders, 1LP, 1IP, and 1HP. Each signature was constructed by dividing the speed range into $B = 20$ sub-ranges.

Table 1 – Description of data sets used.

Data Set	Length (in flights)	Comments
A1	72	Engine event occurs in flights 71-72 Precursors occur in flights 65-70
A2	27	Engine event occurs in flight 23 No precursors
B1	87	No events
B2	96	No events
B3	126	No events
B4	41	No events
	449	(Total number of flights)

Experimental Methodology

Given that each data set contains data from a separate engine, an engine-specific approach was simulated by considering each data set separately.

To simulate the on-line novelty detection process, we step through each data set flight-by-flight - each flight resulting in a 20-bin signature for each of the three fundamental tracked orders.

Stepping through flights from [2 n], at flight n, we will compare data from the current flight against a model of normality constructed from the previous [1 n-1] flights.

Thus, at flight n, for each of the B speed bins, we form a data set from the N = 3 maxima from the [1 n-1] flight signatures observed so far, treating each tracked order separately. I.e., at flight n, we have a distribution of 3(n-1) maxima for each of the B speed bins, for each of the three tracked orders. This is repeated separately for minima. As described in Section 3, we can now set lower and upper novelty thresholds ($h_{\min}$ and $h_{\max}$) in each speed bin. We set thresholds using three methods for comparison:

Method 1: conventional ML-based EVT, as described in Section3.

Method 2: the proposed Bayesian-EVT method

Method 3: a heuristic method, currently used by domain experts.

This latter method sets novelty thresholds $h_{\min}$ and $h_{\max}$ for observed data **X** as follows:

$$\begin{aligned} h_{\max} &= \mu_{\mathbf{X}} + K\left[\max\left(\mathbf{X}\right) - \mu_{\mathbf{X}}\right] \\ h_{\min} &= \mu_{\mathbf{X}} - K\left[\mu_{\mathbf{X}} - \min\left(\mathbf{X}\right)\right] \end{aligned} \qquad (8)$$

where $\mu_{\mathbf{X}}$ is the mean of **X**, and K is some constant, chosen heuristically on a model-by-model basis to best separate "normal" data from "abnormal" data, typically taking values in the range K = [1.2 1.5].

Classification

Thus, for each of the three methods, at flight n, we have set novelty thresholds $h_{\min}$ and $h_{\max}$ in each of the B = 20 speed bins, for each of the three fundamental tracked orders. A method classifies a flight tracked order signature "abnormal" if it contains vibration data that exceed novelty thresholds in *two or more* speed bins. This latter restriction removes the potential for single-point noise spikes to cause a tracked order signature to be classified "abnormal".

We define true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) classifications according to the agreement between the classification of a tracked order signature output by the novelty detection method ("normal" or "abnormal") and presence or absence of an engine event (or precursor event), as described in Table 1.

The *sensitivity* is the proportion of tracked order signatures containing actual engine events that were correctly classified as "abnormal":

$$Sensitivity = \frac{TP}{TP + FN} \qquad (9)$$

The *specificity* is the proportion of tracked order signatures containing no actual engine events that were correctly classified as "normal":

$$Specificity = \frac{TN}{TN + FP} \qquad (10)$$

These two metrics are often used for comparing the performance of classification schemes [1], with the ideal novelty detection method having sensitivity and specificity of 1.0.

Results

Results of setting novelty thresholds using the three methods are shown in Table 2, given as sensitivity and specificity. Note that, because data sets {B1…B4} contain no abnormal events, the sensitivity will always be zero, because TP = 0. These "normal" data sets are used only to evaluate the specificity of each method.

Discussion

When applied to the "abnormal" data sets A1 and A2, all three methods detected both of the engine events and the same set of precursor events. Several of the precursor events from data set A1 were not detected by any of the methods, due to the low magnitude of vibration amplitude observed during those flights. However, all three methods provided the same amount of advance warning of the main engine event (which took place in flight 71), detecting the first precursor event during flight 65. Thus, the sensitivity can be seen in Table 1 to be identical for all three methods.

The specificity of the three methods is significantly different, with the heuristic method attaining specificity 0.73, the ML-based EVT method attaining 0.83, and the proposed Bayesian-EVT method attaining 0.99.

This significant improvement in specificity is the key advantage of the proposed method over the original ML-EVT method and the heuristic method, which were suspected *a priori* to suffer from large numbers of false-

Table 2 - Results

Data Set	Method 1: Bayesian EVT		Method 2: ML-Based EVT		Method 3: Heuristic	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
A1	0.75	1.00	0.75	0.97	0.75	0.93
A2	1.00	1.00	1.00	0.90	1.00	1.00
B1	-	1.00	-	0.99	-	0.99
B2	-	1.00	-	0.85	-	0.45
B3	-	0.99	-	0.70	-	0.68
B4	-	1.00	-	0.70	-	0.65
Overall	0.78	0.99	0.78	0.83	0.78	0.73

positive alarms. The ML-EVT method, as shown in Figure 6, keeps novelty thresholds low even when uncertainty in the model is at its maximum (when there are small numbers of flights). This results in many "normal" flights being incorrectly classified "abnormal", particularly within the first 20-30 flights of each data set. The heuristic method, which uses max($\mathbf{X}$) to set the novelty threshold $h_{\max}$ and min($\mathbf{X}$) to set the novelty threshold $h_{\min}$, is entirely dependent on the single largest and smallest values in the data set, respectively. This results in particularly poor performance, because the novelty thresholds are maximally sensitive to changes in those single points – no density estimation is takes place in the heuristic scheme, and information concerning the distribution of points around the data mean is disregarded.

5. Conclusions

This paper has presented a method of novelty detection that correctly identifies engine events and provides advance warning of those events by detecting precursor events. In comparison to existing methods, it has been shown to provide similar levels of sensitivity to event detection, but does not suffer from the particularly high false-positive alarm rates that are problematic in the use of conventional systems.

Performance of the method in comparison to existing techniques has been verified using large quantities of in-flight vibration data from a number of modern aerospace gas-turbine engines.

The proposed method is suitable both for off-line engine monitoring, in which vibration signatures are constructed at the end of each flight (such as may be performed in a ground-based tracking station), and on-line monitoring, taking place within an engine-mounted health monitoring system. The method is lightweight in memory and processing requirements, such that it may be implemented "on-wing".

An engine-specific approach to novelty detection has been shown to be possible, in which the characteristics of data from individual engines is learned – this often provides better performance than generic "fleet-wide" models of normality [21,22], in correctly identifying behaviour that is abnormal *for that engine*, while minimising the number of false-positive alarms generated.

The proposed method is also "adaptive", with constant on-going updating of its model of normality as engine runs are completed. This allows novelty detection to occur throughout the service life of the engine, without the need for a special training period.

Limitations and Future Work

This paper has described the application of the proposed method to tracked order amplitudes collected within speed-based vibration signatures. However, the proposed Bayesian extension to EVT is applicable to general estimation tasks. Note, however, that this system has been devised specifically for the application to systems in which the cost of false-positive alarms is particularly high (as in the monitoring of aerospace gas-turbine engines). The proposed method sets novelty thresholds to be high when uncertainty in the model is high (for smaller amounts of observed data), and may thus be insensitive during this initial period – the novelty thresholds are initially set conservatively. This may be unsuitable for different application domains in which the cost of false-positive alarms is low in comparison to the cost of missing an abnormal episode – for example, in the detection of cancers in medical images, the specificity is much less critical than the sensitivity. However, within the domain of high-integrity system condition monitoring, the proposed method has been shown to be advantageous in comparison to existing methods.

Future research will include the application of the technique to vibration phase, and to non-vibration data sets, such as "performance parameters" (engine temperatures, pressures, etc.), which have previously been used with off-line density-estimation approaches [23].

ACKNOWLEDGEMENTS

The authors acknowledge the support of the HECToR research project, a consortium sponsored by the UK Department of Trade and Industry.

REFERENCES

[1] L. Tarassenko, A Guide to Neural Computing Applications, Arnold, UK, 1998

[2] I.K. Jennions, "Cross-platform challenges in engine health management", Proc. of Int. Conf. of Integrated Condition Monitoring, Anaheim, CA, USA, 2006

[3] D.A. Clifton, P.R. Bannister, and L. Tarassenko, "Novelty Detection in Large Vehicle Turbochargers", in: H.G. Okuno, M. Ali (eds.): New Trends in Applied Artificial Intelligence, Lecture Notes in Computer Science 4750, Springer-Verlag, Berlin, 2007

[4] P. Hayton, B. Scholkopf, L. Tarassenko, and P. Anuzis, "Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra", Proc. Neural Information Processing Systems, 2000

[5] D.A. Clifton, P.R. Bannister, and L. Tarassenko, "Learning Shape for Jet Engine Novelty Detection", in: J. Wang et al. (eds.): Advances in neural networks III., Lecture Notes in Computer Science, Springer-Verlag, Berlin, 3973:828–835, 2006

[6] Rolls-Royce PLC, The Jet Engine, UK, 1996

[7] D.A. Clifton, P.R. Bannister, and L. Tarassenko, "Visualisation of Jet Engine Vibration Characteristics for Novelty Detection", Proc. NCAF, London, UK, 2007

[8] The QUICK Monitoring System, Oxford BioSignals Ltd, 2007

[9] D.A. Clifton, P.R. Bannister, and L. Tarassenko, "Application of an Intuitive Novelty Metric for Jet Engine Condition Monitoring", In: M. Ali, R. Dapoigny (eds.): Advances in Applied Artificial Intelligence, Lecture Notes in Artificial Intelligence, Springer-Verlag, Berlin 4031: 1149-1158, 2006

[10] A. Nairac, N. Townsend, R. Carr, S. King, P. Cowley, and L. Tarassenko, "A System for the Analysis of Jet Engine Vibration Data", Integrated Computer-Aided Engineering 6(1):53–65, 1999

[11] C.M. Bishop, "Pattern Recognition and Machine Learning", Springer-Verlag, Berlin, 2006

[12] R.A. Fisher and L.H.C. Tippett: "Limiting Forms of the Frequency Distributions of the Largest or Smallest Members of a Sample", Proc. Camb. Philos. Soc. 24, 1928

[13] P. Embrechts, C. Kluppelberg, and T. Mikosch, Modelling Extremal Events, Springer-Verlag, Berlin, 1997

[14] L. Tarassenko, P.R. Bannister, and D. A. Clifton, "Novelty Detection", in K. Worden (ed.), Encylopaedia of Structural Health Monitoring, Wiley, 2008

[15] D.A. Clifton, "Density Estimation for Novelty Detection in Jet Engine Vibration Data", Technical Report, Oxford BioSignals Ltd., UK, 2006

[16] R.L. Smith, "Extreme Value Analysis of Environmental Time Series", Statistical Science 4(4), pp.367-377, 1989

[17] S.J. Roberts, "Novelty Detection Using Extreme Value Statistics", IEE Proceedings 146(3), 1999

[18] S.J. Roberts, "Extreme Value Statistics for Novelty Detection in BiomedicaDataprocessing", IEE Proceedings on Science, Measurement and Technology 147(6), 2000

[19] A. R. Webb, Statistical Pattern Recognition. Wiley, 2002

[20] S.G. Coles and E. A. Powell, "Bayesian Methods in Extreme Value Modelling: A Review and New Developments", International Statistical Review / Revue Internationale de Statistique, Vol. 64, No. 1., pp. 119-136, 1996

[21] D.A Clifton, P.R. Bannister, and L. Tarassenko, "A Framework for Novelty Detection in Jet Engine Vibration Data", in: L. Garibaldi, S. Surace, K. Holford (eds): Key Engineering Materials 347:305-312, 2007

[22] D.A. Clifton, B. Haskins, P.R. Bannister, and L. Tarassenko, "Specific and Generic Models for Gas-Turbine Novelty Detection", Proc. of 4th IET Int. Conf. on Condition Monitoring, Harrogate, UK, 2007

[23] S.W. Utete, D.A. Clifton, and L. Tarassenko, "Trending Performance Parameters for Aircraft Condition Monitoring", Proc. of 4th IET Int. Conf. on Condition Monitoring, Harrogate, UK, 2007

BIOGRAPHY

David A. Clifton *is a DPhil candidate in Engineering Science at Oxford University, and Senior Engineer at university spin-out company Oxford BioSignals Ltd. His research interests include statistical pattern recognition for health monitoring of biomedical and industrial systems, for which he has been the awarded many of the UK's leading student engineering awards, including those from the IEE, the Institute of Physics, the Royal Academy of Engineering, the European Commission, Hewlett-Packard Europe, Shell UK, and the UK Engineering Council. His current research, part-funded by the UK Department of Trade and Industry, involves devising novelty detection systems for modern civil and military aircraft engines, including the Eurofighter EJ200, the Airbus A380 Trent 900, and the Boeing Dreamliner Trent 1000.*

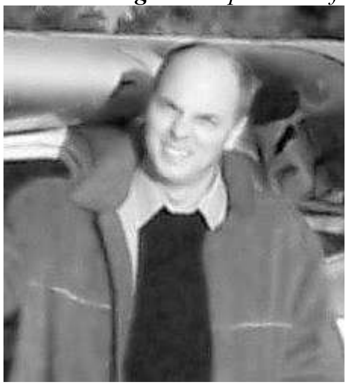
Nick McGrogan *is Technical Manager for the Industrial division of Oxford BioSignals Ltd. He holds an M.Sc. degree in Engineering and Computing Science and a D.Phil. degree in Engineering Science, both from Oxford University. He is also a member of the Institution of Engineering and Technology (IET). His D.Phil. research was in the area of advanced signal processing for the detection of epileptic seizures in long-term recordings of brain activity. While completing his D.Phil. degree he joined Oxford BioSignals when the company was founded to progress the development of medical signal processing products and progressed to leading the industrial division in the development of techniques for health monitoring of aircraft engines. In his current role he is responsible for Oxford BioSignal's technology strategy and manages the company's research projects.*

Lionel Tarassenko *Lionel Tarassenko was born in Paris, France, in 1957. He received the B.A. degree in engineering science in 1978, and the Ph.D. degree in medical engineering in 1985, both from Oxford University, Oxford, U.K.. After graduating, he worked for Racal Research Ltd. on the development of digital signal processing techniques, principally for speech coding. He then held a number of positions in academia and industry, before taking up a University Lecturership at Oxford in 1988. Since then, he has devoted most of his research effort to the development of neural network techniques and their application to signal processing, diagnostic systems, and parallel architectures. He has held the Chair in Electrical Engineering at Oxford University since October 1997. He was elected to a Fellowship of the Institution of Electrical Engineers (IEE) in 1996, when he was also awarded the IEE Mather Premium for his work on neural networks, and to a Fellowship of the Royal Academy of Engineering (RAE) in 2000.*

Dennis M. King *is the Rolls-Royce PLC. Company specialist in Engine Health Monitoring specialising in the vibration aspects of EHM. He was awarded an MSc in Engineering Mechanics in 1976 by the Cranfield Institute of Technology, UK. and Chartered Engineer in 1988 (RAeS). After graduating he joined Rolls-Royce in Derby, working in Stress analysis, and design before moving to the vibration laboratories and taking on the role of consulting on the design, build and balance of aero engines. Over the last twenty years he has been actively involved in developing EHM technology including Hardware, software, fault recognition and diagnostics. He has the experience of applying this technology globally within the R-R business. He has been awarded the R-R Component Engineering Quality award (1997), R-R Chairman's team award for technical innovation (2001).*

Paul Anuzis *has spent five years of his working career at Rolls-Royce in Germany, helping to establish a new and highly sucessful Rolls-Royce business unit in Germany, based initially in Munich now established in Dahlewitz on the outskirts of Berlin. Paul returned in late 1997 to take up his current position of Chief Reliability Engineer for Rolls-Royce Civil Aerospace – Airlines. He holds a degree in Physics and second degree in Electronics, is a Chartered Physicists, Chartered Scientist, Chartered Engineer and a Fellow of the Institute of Physics, and has twice won the Rolls-Royce Chairman's award for Technical Innovation. He has extensive applied research experience from several years of runing DTI technology programs. The most sucessful programs spanning eight years working with the Universities of Oxford, Nottingham and Cranfield on engine health monitoring technologies – which are about to enter service with the latest Trent 900/A380 application.*

Steve P.King *is a Specialist for advanced Engine Health*



Monitoring Methods at Rolls-Royce plc. He holds a degree in Mathematics and Computer Science and a PhD in the use of Expert Systems for Vibration Analysis. He is a Chartered Engineer and a member of the Institute for Engineering Technology (IET). Having worked in Rolls-Royce since 1979, he has built up many years experience working in advanced signal processing methods, Data Mining, Neural Networks and equipment health monitoring. Steve's current post is within the Company's Global EHM Capability Group. Prior to this he spent 12 years working in the Company's Corporate Strategic Research Centre with responsibility for the development and application of Computational Intelligence techniques across all business areas of Rolls-Royce. He was awarded the Rolls-Royce Chairman's award for Technical Innovation in 2001, and is named on three patents. He is also a member of the International Scientific Committee of the International Conference for Condition Monitoring.