# Aircraft Engine Health Monitoring using Density Modelling and Extreme Value Statistics

Srini Sundaram, Iain G.D. Strachan
Oxford BioSignals Ltd.
174, Milton Park, Abingdon, Oxfordshire, OX14 4SE, UK
srini.sundaram@oxford-biosignals.com

David A. Clifton, Lionel Tarassenko
Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford
Old Road Campus, Roosevelt Drive, Oxford, OX3 7DQ, UK

Steve King
Rolls-Royce plc, EHM Global Capability Group, PO Box 31,
Derby, DE24 8BJ.

## Abstract

Current practice in the operation and maintenance of an aircraft fleet requires analysis of data obtained from in-service engines in order to identify engine deterioration and provide preventative maintenance. Typically large quantities of engine vibration and performance data are available from various engine-mounted sensors. The analysis of such data requires techniques for modelling these multivariate data allowing fleet specialists to establish profiles of engine behaviour under different operating conditions. Additionally, such techniques can be used to identify precursors of engine events to avoid loss of engine service.

This paper describes density modelling techniques for the estimation of the multivariate unconditional data density of performance and vibration parameters acquired from aerospace gas-turbine engines. We set a probabilistic threshold using Extreme Value Theory (EVT). This framework is used to generate reliable, timely alerts concerning abnormal engine operation. Finally, case studies are presented that analyse performance and vibration data obtained from a representative set of civil aircraft engines. Our results show that such techniques can provide reliable identification of abnormal engine events.

## 1 Introduction

In order to reduce operational and maintenance costs, Engine Health Monitoring (EHM) experts apply various tools that can learn the behaviour of engines and alert before components reach a critical stage of failure. This paper presents a novelty detection based approach that constructs a model of normality using examples of “normal” engine behaviour and then detects deviations from the model. Novelty detection is ideally suited for condition monitoring of engines as most of the data available are typically from “normal” engine behaviour and failure examples are very rare. This approach allows users to construct models of normality, calibrate them to the available engine data, and use the resultant models to identify abnormal events [1,2]. A systematic investigation of visualisation and novelty detection techniques is presented in this paper using a representative set of engine data. The investigation is divided into four stages: (i) pre-processing of data for feature extraction, (ii) data understanding using methods of high-dimensional visualisation , (iii) construction of models of normality (multivariate Gaussian Mixture Models and Parzen window estimators), and (iv) principled setting of decision thresholds for novelty detection using multivariate extreme value theory (M-EVT).

## 2 Feature Extraction and Visualisation

In order to prepare a dataset for analysis, pre-processing must be performed in order to remove artefacts from the data due to errors in sensor measurements. Subsequently, data are then normalised using component-wise normalisation (i.e., a zero-mean, unit-variance transformation) to ensure that all parameters vary over similar ranges. Next, the multidimensional data are inspected using the NeuroScale visualisation technique[3,4] in order to assist in model construction The NeuroScale neural network[1] allows the visualisation of high-dimensional vectors by mapping them to lower numbers of dimensions (typically two, for visual inspection). Feature vectors are extracted from the data that capture the difference between "normal" and "abnormal" engine operation.

For every high-dimensional feature vector, the NeuroScale network provides a corresponding pair of $(i, j)$ co-ordinates. This is a projection from $D > 2$ dimensions to $D' = 2$. The training algorithm of the NeuroScale network attempts to preserve the inter-point Euclidean distances of high-dimensional vectors after projection into 2-dimensional space by minimizing the Stress metric, $E$ , as defined in equation (1).

$$E = \sum_{i}^{N} \sum_{j>1}^{N} (d_{ij}^{*} - d_{ij})^{2} \qquad (1)$$

where is $d_{ij}^{*}$ the Euclidean distance between vectors $i$ and $j$ in data space, and $d_{ij}$ is the Euclidean distance between corresponding vectors in the visualisation space. The objective function, $E$ , is minimised by adjusting the locations of the visualisation vectors. The NeuroScale algorithm adjusts the output weights of an RBF network in order to reduce the value of $E$. Thus, $n$-dimensional feature vectors which are similar (i.e., close together in the original high-dimensional data space) should be kept close together after projection into 2-dimensional space. Conversely, $n$-dimensional vectors that are significantly different from one another (i.e., far apart in high-dimensional space) should remain well-separated after projection into 2-dimensional space. The goal is to allow clusters of feature vectors corresponding to "normal" behaviour to be evident, with feature vectors corresponding to "abnormal" behaviour to be far removed from them (and thus detectable by some later analysis technique).

---

[1] A Radial Basis Function (RBF) neural network.

# 3 Density Modelling using Gaussian Mixture Models and Parzen Window Estimators

The next stage of the investigation is to construct a model of normality using "normal" training data. The model of normality is provided using two candidate techniques, Parzen window estimation and Gaussian Mixture Modelling (GMM). Both approaches estimate the unconditional probability density of the training data, $p(x)$. If we consider $N$ training points from the input data $x$, the data density for the Parzen window estimator is defined to be

$$p(x) = \frac{1}{N\sigma}\sum_{i=1}^{N} K\left(\frac{x-\mu_i}{\sigma}\right) \quad (2)$$

where σ is the width or bandwidth parameter and $K(.)$ is the Gaussian kernel given by

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \quad (3)$$

The Parzen window model places an identical Gaussian kernel $K$ on each of the $N$ training data and uses a common width for each kernel(5,6). The width is calculated as the average distance to the 10 nearest neighbours(1).

The GMM computes $p(x)$ using a linear combination of basis functions(7). The data density is defined to be

$$p(x) = \sum_{j=1}^{M} P(j)\,p\left(x \mid j\right) \quad (4)$$

where $p(x \mid j)$ is the probability of $x$ with respect to kernel $j$, and $P(j)$ is the prior probability that $x$ was generated by kernel $j$.

To ensure that the result is a probability density function, $p(x)$ must satisfy the following criteria:

i) the function should be non-negative throughout, and

ii) the function should integrate to 1.

Choosing a normal density for $p(x \mid j)$ results in a proper probability density function, as desired.

# 4 Multivariate Extreme Value Statistics

In most the real-world problems, $p(x)$ is multimodal and multivariate. The "classical" approach to EVT is based on finding the extreme values from a distribution, where "extreme" is defined in terms of the magnitude of the data $x$. It also states that each Gaussian density $p(x)$ in the mixture of Gaussians modelling the multivariate space along radius $r$ varies according to univariate Gaussian and thus their EVD can be modelled using univariate Gumbel distributions(8).

(8) showed that this assumption does not hold in multivariate space. In order to use EVT for novelty detection, the aim is to identify events that are extreme in probability, with respect to some normal model, rather than detect events that are extreme in magnitude.

Though outside the scope of this paper, multivariate EVT can be used to determine where the boundary of "normal" behaviour should lie, with respect to a model of normality, under "normal" engine conditions. Then, if data are observed beyond this boundary, they are classified "abnormal", and an alert is provided to the user. Further reading is provided in (8).

# 5 Experiments and Results

We consider a set of 9 aerospace gas-turbine engines. From each engine, a set of performance parameters (such as pressures and temperatures) and vibration characteristics were acquired. These parameters are conventionally used as descriptive parameters of engine condition by domain experts. Histograms of the distribution of performance and vibration parameters obtained for engine 1 are shown in Figures 1 and 2. It may be seen from the figures that the tails of each distribution differ for each parameter. In multivariate space, this will require a more accurate method of modelling the extreme values distribution than a simple expert threshold such as 3 standard deviations with an assumption of Gaussian distribution.

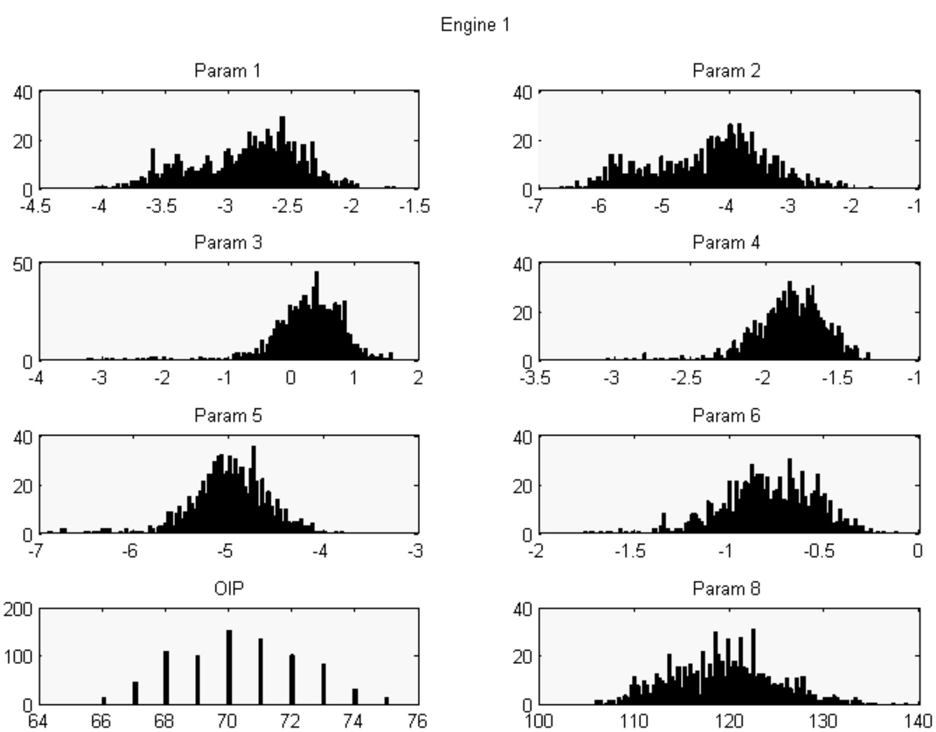


**Figure 1 – Histograms Representation of Performance Parameters of engine 1**

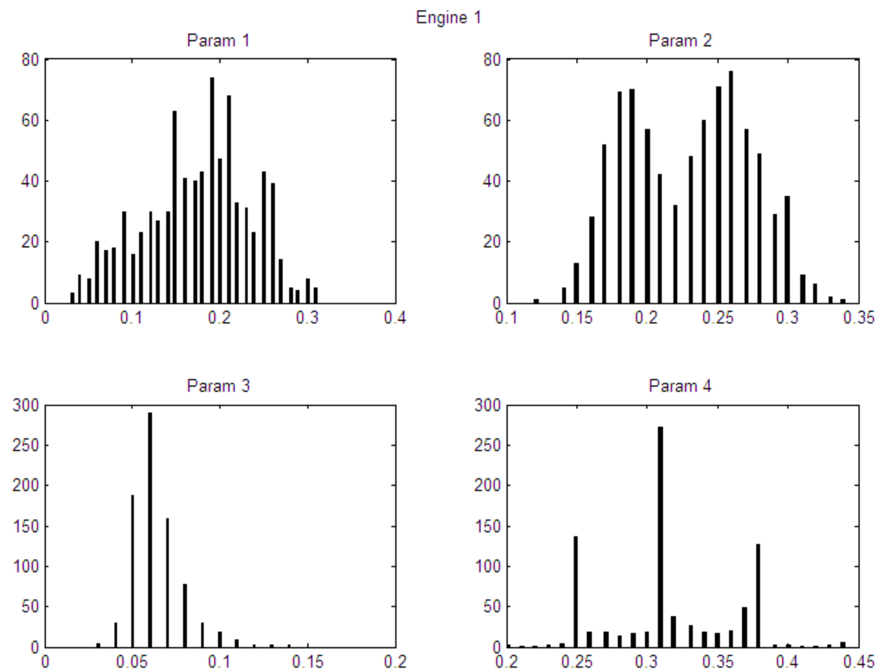


**Figure 2 – Histogram Representation of Vibration Parameters of engine 1**

## High-Dimensional Visualisation

Each feature vector is defined to be a multivariate vector comprised of the input parameters. We construct separate feature vectors for performance and vibration data. (In this study, we have chosen 8-dimensional feature vectors in the case of performance data and 4-dimensional feature vectors in the case of vibration data). The overall behaviour of data in their native high-dimensional space may be visualised in two dimensions using a NeuroScale network, as described previously. The following criteria were adopted for training the NeuroScale visualisation network:

- If a dataset contains > 400 flights of data, 200 flights of data were used for training the model, and the number of hidden nodes in the NeuroScale RBF network was set to be 50.
- If the engine dataset contains < 400 flights of data, 50% of the flights are taken for training, and the number of hidden units was set to be 20% of the number of training data points.

We note that, though we have chosen large training sets for illustration of this technique, comprising many flights, practical systems would be able to construct models using much smaller numbers of flights. Typically, a minimum of 5 flights is required in order to perform novelty detection[8].

The output of the NeuroScale network is a 2-dimensional map. Expert labels were provided for the 9 datasets used by the investigation describes by this paper, and we adopt the following colour scheme corresponding to these labels:

- data used for training are shown in green;
- data occurring after those used for training are shown in black or grey*;*
- data occurring after conventional EHM systems identified an engine event are shown in red.

Figures 3 and 4 show the visualisation of data from the 9 engines, showing performance and vibration data, respectively.

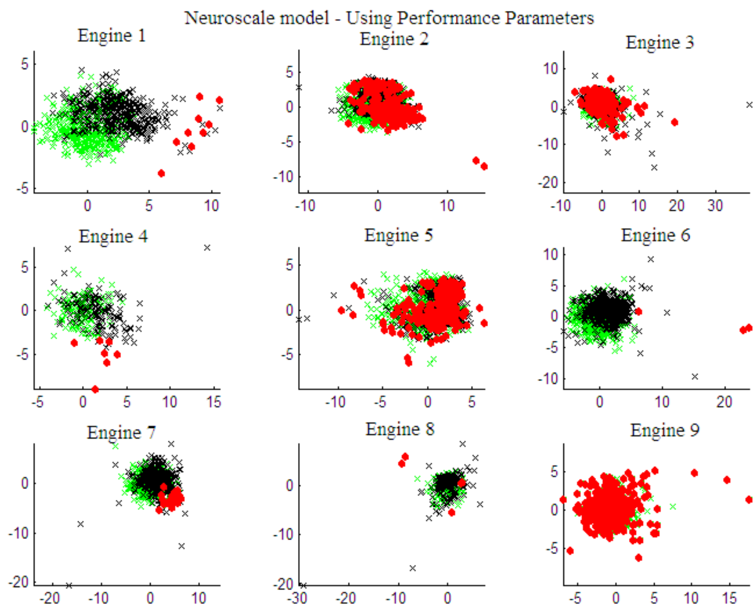


**Figure 3 — Neuroscale Visualisation of Performance Parameters for nine engines (Green indicates data points taken for training, Black indicates post training data points until Event and Red indicates data points after Event)**

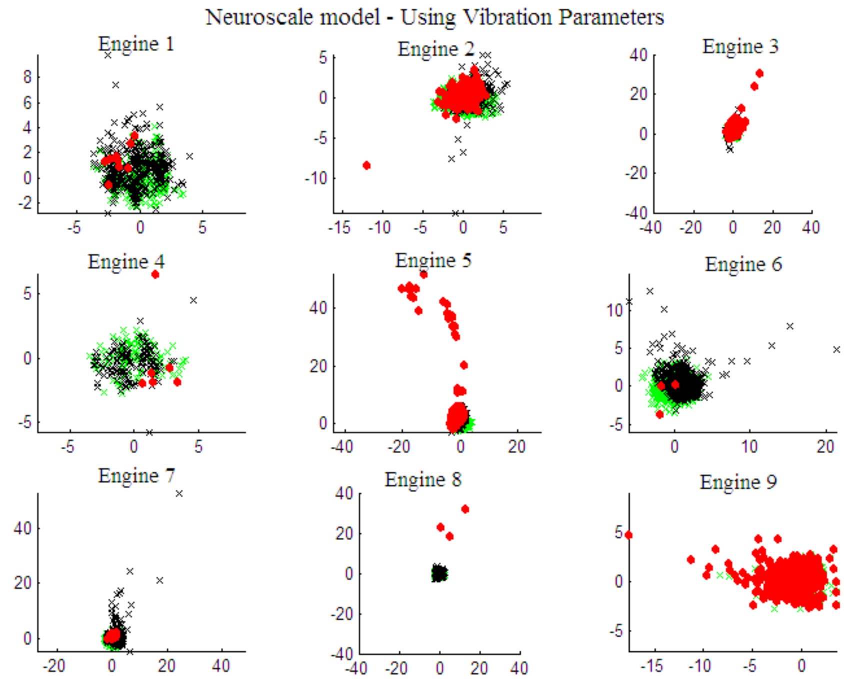


**Figure 4 — Neuroscale Visualisation of Vibration Parameters for nine engines (Green indicates data points taken for training, Black indicates post training data points until Event and Red indicates data points after Event)**

For datasets in which one or more parameters deviate from the model of normality, it is expected that data corresponding to the event (shown in red) would be separated from the "normal" training data (shown in green). Any precursors to this event in the data shown in black could be visible as excursions from the "normal" data.

In Figure 3, visualisation of data from engines 1, 4, 6 and 7 shows that the data shown in red are significantly separated from the "normal" cluster. This may indicate that one or many performance parameters for these engines are indicative of abnormal behaviour, compared to the "normal" data from that engine. In the case of the visualisation of data from engine 1, many of the data coloured in black, corresponding to the period after training, show significant separation from the "normal" data, and this may correspond to a precursor of the eventual engine event. Similarly in Figure 4, the visualisation of data from engines 3, 6, and 8 shows similar separation between engine event and "normal" data.

## Density Models

The visualisation gives an approximation of the separation of feature vectors in their native high-dimensional data space. In order to study engine behaviour in more detail, the two candidate density estimation models, based on Parzen window estimation and GMMs, were constructed using the "normal" feature vectors for each engine. Note that a separate model was trained for each engine.

We define a novelty score $z(x)$ = -ln p($x$), such that improbable events, which have low unconditional probabilities $p(x)$, take high novelty scores, $z(x)$.

Figures 5 and 6 show novelty scores obtained using the Parzen window estimator for performance and vibration parameters, respectively

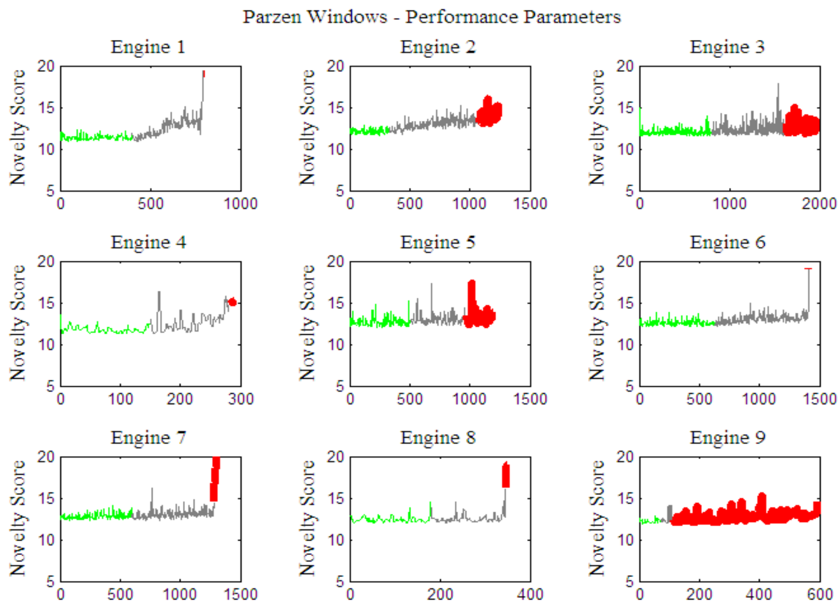


**Figure 5 — Parzen Windows Model Novelty Scores obtained using dataset containing Performance Parameters for nine engines (Green indicates data points taken for training, Grey indicates post training data points until Event and Red indicates data points after Event)**

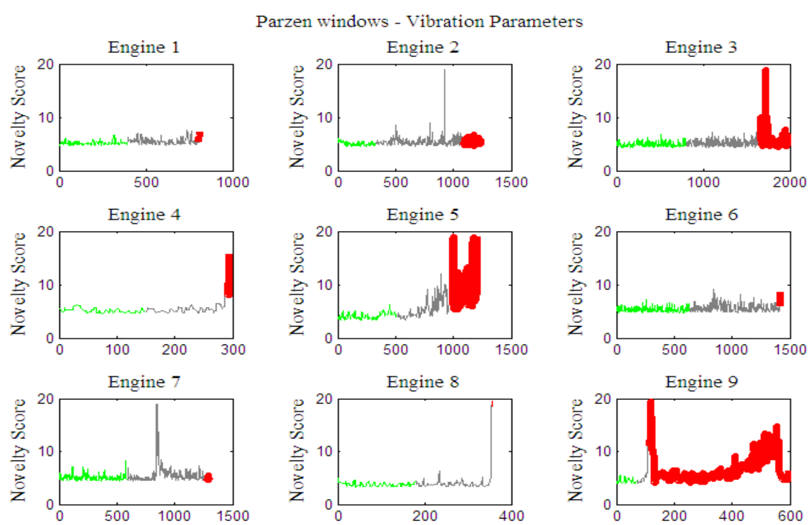


**Figure 6 — Parzen Windows Scores obtained using dataset containing Vibration Parameters for nine engines (Green indicates data points taken for training, Grey indicates post training data points until Event and Red indicates data points after Event)**

In Figure 6, novelty scores from the Parzen window estimator clearly show an increase in the case of data from engines 3, 4, 6, 8 and 9. This indicates that the corresponding vibration parameters in each case contain evidence of "abnormal" engine behaviour. In some cases, the novel events are observed in only one of the vibration or performance models. For example, engine 1 shows increases in novelty score for the performance model and not the vibration model. Engine 3 shows a significant rise in novelty score output by the performance model considerably earlier than the rise in novelty score output by the vibration model. We note that in the case of engine 3, the engine event was detected by conventional methods using vibration data. Novelty scores determined using the GMMs are shown in Figures 7 and 8 for performance and vibration parameters, respectively, where similar trends were observed. (We note that both Parzen window estimation and GMMs provide similar estimates of $p(x)$ and thus similar novelty scores for each engine.)

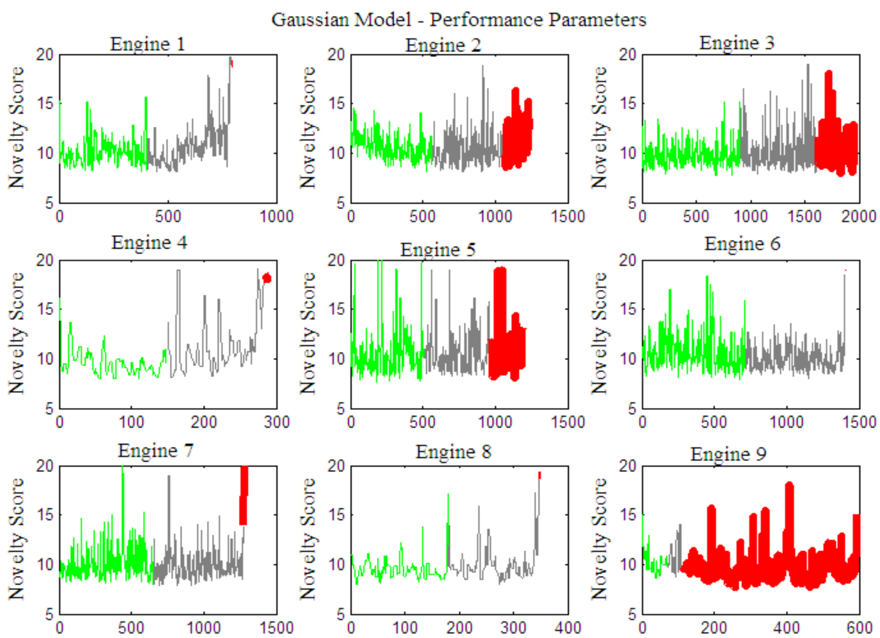


**Figure 7 — Gaussian Model Novelty Scores obtained using dataset containing Performance Parameters for nine engines (Green indicates data points taken for training, Grey indicates post training data points until Event and Red indicates data points after Event)**

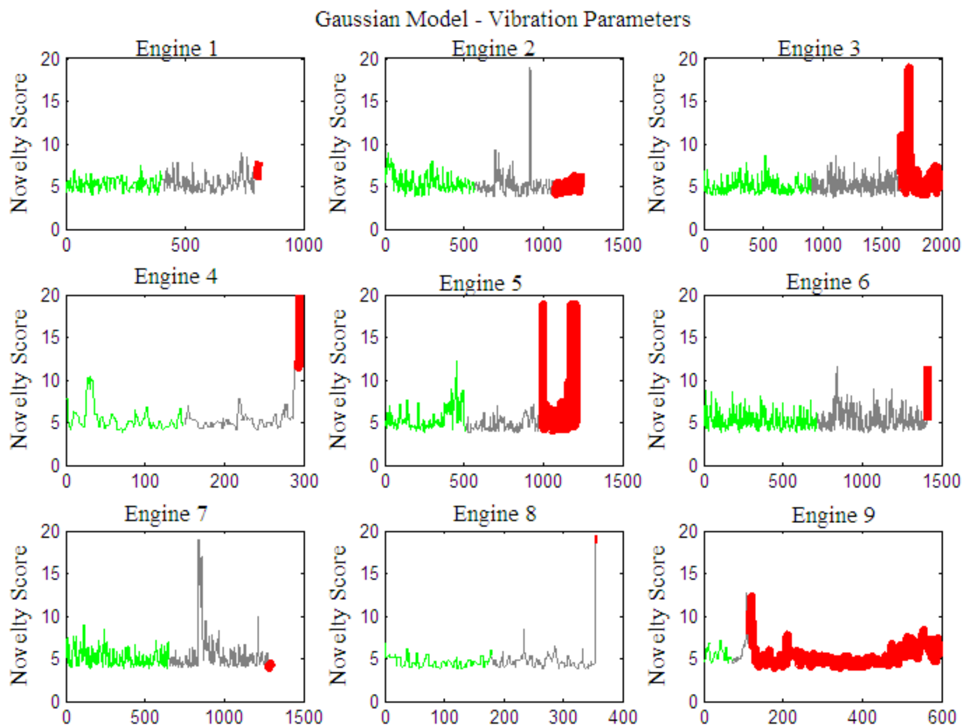


**Figure 8 — Gaussian Model Novelty Scores obtained using dataset containing Vibration Parameters for nine engines (Green indicates data points taken for training, Grey indicates post training data points until Event and Red indicates data points after Event)**

The results from the density models shown above demonstrate that precursor events can be identified in performance and/or vibration data. The next stage is to set a principled threshold using M-EVT, described in section 4.

## Models with M-EVT Threshold

The model output (novelty scores) with M-EVT applied to performance data from engines 1 and 2 are shown in Figures 9 and 10, respectively. It can be seen that in Figure 9, the threshold exceedance is observed some 200 points (approximately 50 flights) before the actual event was observed (around data point 800).

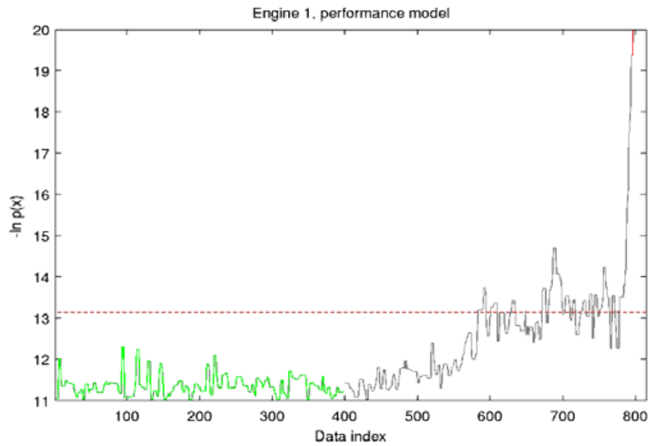


**Figure 9 – Parzen Windows model Novelty Scores for Engine 1 Performance Parameters with M-EVT Threshold (Red dotted line)**

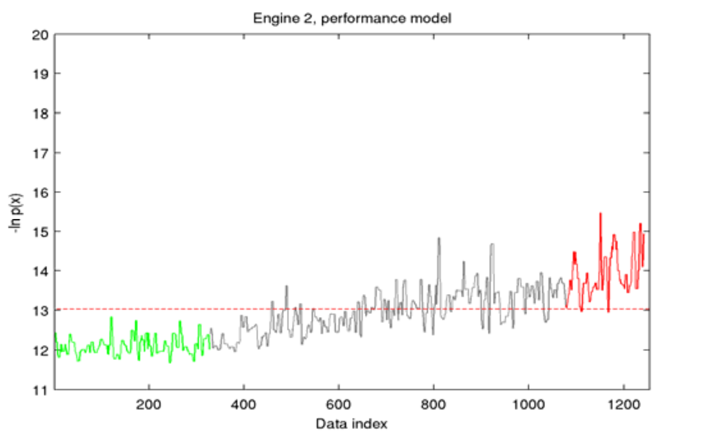


**Figure 10 – Parzen Windows Novelty Scores for Engine 2 Perforamce Parameters with M-EVT Threshold (Red dotted line)**

Similarly in Figure 10, the M-EVT threshold exceedance occurs around data point 620, and continued to show exceedances thereafter. We note that conventional methods observed the event much later, at around data point 1100. Similar results are obtained for data from other engines, which are not shown here.

# 6 Conclusion

This paper has presented a novelty detection methodology for aircraft engine health monitoring using visualisation, density modelling and extreme value statistics. First, a visualisation model was used to understand the multivariate data. Two candidate approaches to modelling normality, namely Parzen window estimation and GMMs, were constructed using "normal" data from each engine. These models were then shown to be used for detecting abnormal events in both vibration and performance parameters. Finally, a principled decision threshold was set using multivariate EVT. Results show that this methodology can be used to provide early warning of engine events, typically far in advance of conventional EHM systems.

In the study described by this paper, we have made an *a priori* assumption that the first flights of the engine were "normal", such that models of normality could be constructed for each engine. We note that, in practice, data from a much smaller number of flights is required in order to construct such models, and the large numbers of flights used for training models in this study are for the purposes of illustration only.

Future work includes setting this multivariate methodology into a Bayesian framework, such that uncertainty in our data may be quantified in our model. This can further help to drive down the false-positive novelty detection rate, which is already low in the case of robust multivariate techniques(8).

## Acknowledgements

The authors acknowledge the support of Oxford BioSignals Ltd. and the Hidden Patterns research programme, funded by Rolls Royce PLC. The authors also would like thank OSyS for providing the dataset during the extended validation activity period of the project.

## References

[1] D.A. Clifton, L.A. Clifton, P.R. Bannister, and L. Tarassenko, 'Automated Novelty Detection in Industrial Systems', in Y. Liu et al. (eds), Advances in Computation Intelligence in Industrial Systems, Springer-Verlag, Berlin, 2008.

[2] Tarassenko, L., Clifton, D.A., Bannister, P.R., King, S., and King, D.: Novelty Detection. Worden, K., et al. (eds): Encyclopaedia of Structural Health Monitoring. Wiley, 2009. In press.

[3] Sundaram, S., Strachan, I.G.D., Clifton, D.A., King, S., and Palmer, J.: A Data Mining Approach to Reveal Patterns in Aircraft Engine and Operational Data. Proc. of 5th IET Int. Conf. on Condition Monitoring, Edinburgh, U.K., 2008, pp. 1387-1397.

[4] D. Lowe and M.E. Tipping, 'Feed-Forward Neural Networks and Topographic Mappings for Exploratory Data Analysis', Neural Computing and Applications, Vol. 4, pp 83-95, 1996.

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

[6] E. Parzen. On the estimation of a probability density function and mode. Annals of Mathematical Statistics, 33:1064–1076, 1962.

[7] I. Nabney, 'Netlab: Algorithms for Pattern Recognition', Springer-Verlag, Berlin, 2002.

[8] D.A. Clifton, "Novelty Detection with Extreme Value Statistics in Jet Engine Vibration Data". DPhil thesis, University of Oxford, 2009.