Inter-engine variation analysis for health monitoring of aerospace gas turbine engines

William R. Jacobs[1], Huw L. Edwards[2], Visakan Kadirkamanathan[3], and Andrew R. Mills[4]

[1,3,4] *Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, UK*
w.jacobs@sheffield.ac.uk
a.r.mills@sheffield.ac.uk

[3] *R^2 Data Labs, Rolls-Royce Plc, Derby, Uk*
Huw.Edwards@controlsdata.com

ABSTRACT

This study investigates the development of a new inter-engine variation analysis method for the purpose of equipment health monitoring, in which the similarity - in both system behaviour and external disturbances - across multiple (sister) engines is leveraged. The sister engine provides a baseline description of the engine under observation, such that the challenge becomes the differentiation between normal inter-engine variation and the anomalous behaviour, bypassing the need to describe highly complex engine dynamics. The inter-engine residuals are modelled directly with input data from both engines, using previous healthy data for training. The trained model is used to compensate known differences between real engines. Anomalous data is detected by comparison of the simulated output with the true residuals. The method is demonstrated on a real data set containing both nominal, healthy engine data, and engine data containing anomalies.

1. INTRODUCTION

Modern engineering systems are growing increasingly complex in order to achieve higher levels of performance. In response, equipment health monitoring (EHM) procedures must become more sophisticated to ensure reliable operation of the asset. EHM includes a variety of tasks, such as fault detection and isolation/localisation, monitoring of component degradation, and prediction of impending failures. EHM is key in assessing the health of systems/components of aero gas turbine engines (GTEs) in order to provide an early warning of impending failures.

Currently, EHM algorithms, operating both on-board and off-board, monitor engine data in order to detect both fault signatures and anomalous behaviour. The result of the EHM informs human operators of which parameters are of interest such that a decision/diagnosis can be made based on both engineering knowledge and the result of automatic EHM. The recent availability of high frequency time series data, collected throughout the flight, provides an opportunity to increase EHM coverage. However, it also dramatically increases the size of the data collected. Diagnosis of faults requires both detection and localisation of fault signifiers in the data, such that automatic algorithms as well as engineering judgement can be employed.

A standard approach to detecting known fault modes, i.e. those that have been previously observed, is to identify a fault signature in time series data and employ algorithms that can detect similar instances of the signature in future data. Algorithmic techniques may range from simple thresholding of parameters, to sophisticated machine learning techniques (Isermann, 2006; Yan & Yu, 2015). This approach has been shown to be affective across many applications (Miljković, 2011). Modern aero GTEs have a host of sensors collecting operational data for the purposes of EHM, however, examples of fault/impending fault symptoms are rarely observed because GTE components are not run to failure. The detection of fault symptoms therefore, often becomes the detection of novel behaviour in the data record. To detect novel data instances a good description of the nominal system behaviour must be achieved, however, the behaviour of a GTE is highly complex and subject to many unmeasured disturbances, the description of which offers a significant, possibly infeasible challenge. As such, standard approaches are not appropriate in this context, motivating the method developed in this work.

Aerospace GTEs are commonly used in pairs, named sister engines. Sister engines share the same hardware and control architecture, as well as sharing many external disturbances when in operation (such as changes in ambient temperatures and pressures). They therefore act as a physical models of their counterpart, potentially replacing the need for a mathematical model and producing an inter-engine residual as a base for analysis and fault detection. However, they also display heterogeneous behaviour in normal operation, this can be due to a difference in engine component age, heterogeneous external disturbances and behaviour due to faults that is to be detected. In order for analysis of the inter-engine residual to be successful, variations due to normal, explainable differences in engine behaviour must be eliminated in order that anomalous behaviour can be easily detected. This is the problem that is addressed in this work.

A method for inter-engine variation analysis is presented, based on the description of nominal inter-engine variations with a data driven model, the mixture of experts (MoE) model. The MoE model is able to automatically switch between different regression experts based on the distribution of the current input data sample, (Jacobs & Hinton, 1991). It can well describe complex systems that operate over multiple different operating conditions, and is hence well-suited to modelling the behaviour of GTEs.

The MoE model is used to perform a mapping between inter-engine residuals considered as system inputs and outputs. The difference between measured and predicted output inter-engine residuals is demonstrated to be a good target for anomaly detection. The new analysis method is demonstrated on a real data case study and is shown to be successful in producing a variable with reduced influence from explainable behaviour as well as increased sensitivity to anomalous behaviour.

The paper proceeds as follows. In Section 2 the concept of inter-engine variation analysis is introduced. The MoE mdoel is then introduced in Section 3. A real world case study is introduced in Section 4 and the results are presented in Section 4.3. Finally, concluding remarks are given in Section 5.

2. INTER-ENGINE VARIATION ANALYSIS

Sister engines, numbering 2 or greater, on a single aircraft are generally controlled to the same control reference, and are therefore attempting to produce the same amount of thrust. Inter-engine variation is defined as the difference in behaviour between the two engines. The variation can be quantified by computing the residuals between engine parameters recorded from various on-board sensors. Sources of this variation are attributed to the following;

1. Heterogeneous (expected) degradation in the health of engine components (e.g. loss of efficiency due to component age)
2. Heterogeneous measured external disturbances (e.g. temperatures at engine inlet)
3. Heterogeneous unmeasured external disturbances (e.g. airflow at engine inlet)
4. Abnormal behaviour (to be detected)

Degradation in the health of engine components is expected to occur slowly, over a long time scale, and is approximately invariant between consecutive flights. It is therefore assumed that the difference in degradation can be captured by some mapping from the parameters of sister engines to the inter-engine residuals. Similarly, the effects of differing measured inlet conditions can be mapped to the inter-engine residuals. The mapping may be non-linear and dynamic, such that it is dependent on the current operating condition of each engine. Given a sufficiently flexible model structure, and representative training data, variations due to these causes are hence describable.

To achieve such a mapping a statistical model can be identified such that

$$y_n^r = f(\boldsymbol{x}_n^r) + r_n^{A,B} + e_n \tag{1}$$

where $y_n^r = y_n^A - y_n^B$, $\boldsymbol{x}_n^r = \boldsymbol{x}_n^A - \boldsymbol{x}_n^B$ are residuals of an engine parameter of interest and the measured inlet conditions respectively, $r_n^{A,B}$ is process noise that contains unmeasured disturbances and abnormal behaviour, and e_n is i.i.d. Gaussian white noise. A and B indicate data originating from engine A and B respectively. n is the current time step where $n = 1, 2, \ldots, N$

The quantity $R = [r_1^{A,B}, r_2^{A,B}, \ldots, r_N^{A,B}] + \boldsymbol{e}$ is defined as the *normalised inter-engine residual* for the current flight under investigation, where

$$r_n^{A,B} + e_n = y_n^r - f(\boldsymbol{x}_n^r). \tag{2}$$

Under the nominal case, with no different unmeasured disturbances acting on the sister engines such that $r_n^{A,B} = 0$, $R = \boldsymbol{e}$ is zero mean i.i.d. Gaussian white noise. Any data that is distinguishable from Gaussian white noise is therefore resultant from either an unmeasured disturbances or abnormal behaviour and can be detected.

The normalised inter-engine residual, $\mathbf{r}^{A,B}$, is therefore more amenable to anomaly detection than either the raw data or the inter-engine residual, $\mathbf{y}^r$. Multiple methods are available to perform anomaly detection on $\mathbf{r}^{A,B}$. The simplest approach is to threshold the absolute value, $|\mathbf{r}^{A,B}|$, any value that is observed above the threshold is a potential fault. A further method is to apply a sliding window across $\mathbf{r}^{A,B}$ and $\mathbf{y}^r$ and compare the variance of the windowed data, resulting in an anomaly score. Many more advanced methods are available (Chandola, Banerjee, & Kumar, 2009). This work is intended to highlight the concept of inter-engine variation analysis and as such the anomaly detection step is considered beyond the scope of this paper.

3. BAYESIAN MIXTURE OF EXPERTS

The MoE model is capable of describing complex dynamic systems that operate over multiple states. The model probabilistically divides up the input space using a gating network, a regression expert operates over each of regions selected by the gating network. The parameters of MoE model are estimated within an elegant Bayesian inference framework, named variational Bayesian inference, by iterating through closed form update equations, (Ueda & Ghahramani, 2002; Baldacchino & Rowson, 2016).

3.1. The Mixture of Experts Model

A multi-input single-output system at time instant n can be described by a MoE model, formed by the sum of the product of the k'th gating network $g_k(\cdot)$ and the k'th expert function $f_k(\cdot)$, such that

$$y_n = \sum_{k=1}^{K} g_k(\boldsymbol{x}_n, \pi_k, \theta_k^g) f_k(\boldsymbol{x}_n, \boldsymbol{w}_k) \;, \tag{3}$$

where $\boldsymbol{x}_n = [x_n^1, \dots, x_n^{d^x}]$ is the d_x dimensional system input, and y_n is the system output. The k'th expert function, $f_k(\cdot)$ is restricted to to be linear-in-the-parameters, such that it can be decomposed into the sum of M weighted basis functions,

$$f_k(\boldsymbol{x}_n, \boldsymbol{w}_k) = \sum_{m=1}^{M} w_{k,m}\phi_n(\boldsymbol{x}_n) \tag{4}$$

$$= \Phi_n \boldsymbol{w}_k \tag{5}$$

where

$$\Phi_n = [\phi_1(\boldsymbol{x}_n), \phi_2(\boldsymbol{x}_n), \dots, \phi_M(\boldsymbol{x}_n)], \quad \Phi_n \in \mathcal{R}^{1\times M}$$
$$\boldsymbol{w}_k = [w_{k,1}, w_{k,2}, \dots, w_{k,M}]^T, \quad \boldsymbol{w}_k \in \mathcal{R}^{M\times 1}$$

and Φ_n is the n'th row of the matrix $\boldsymbol{\Phi}$,

$$\boldsymbol{\Phi} = [\Phi_1^T, \Phi_2^T, \dots, \Phi_M^T]^T, \quad \boldsymbol{\Phi} \in \mathcal{R}^{N\times M}.$$

where N is the total number of data points.

The k'th gating network $g_k(\cdot)$ takes the form of a normalised Gaussian function,

$$g_k(\boldsymbol{x}_n, \theta_k^g) = \frac{\pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Lambda_k^{-1})}{\sum_i^K \pi_i \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_i, \Lambda_i^{-1})} \tag{6}$$

where $\theta_k^g = [\boldsymbol{\mu}_k, \Lambda_k]$, and $\boldsymbol{\mu}_k$ and Λ_k are the mean and covariance matrix of the k'th gating network respectively. π_k is the k'th mixing coefficient, satisfying the conditions $\pi_k \geq 0$ and $\sum_{k=1}^K \pi_k = 1$. The gating network gives the probability that the output is described by the k'th expert function.

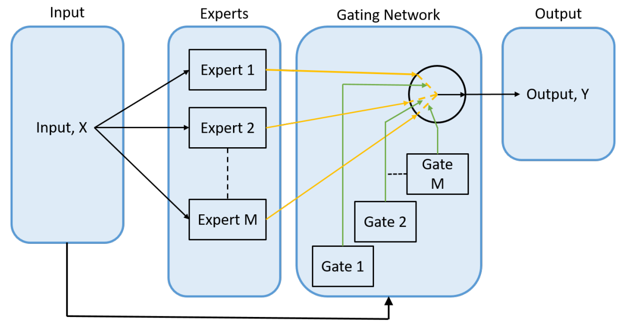


Figure 1. Schematic of the Mixture of Experts model.

3.2. Variational Bayesian Inference

Variational Bayesian inference is used to estimate the parameters of the MoE model. The likelihood function for Equation (3) is given by

$$p(y_n|\boldsymbol{x}_n, \boldsymbol{\pi}, \theta_k^g, \theta_k^e) = \sum_{k=1}^{K} p(k|\boldsymbol{x}_n, \pi_k, \theta_k^g) p(y_n|\boldsymbol{x}_n, \theta_k^e). \tag{7}$$

where $p(k|\boldsymbol{x}_n, \pi_k, \theta_k^g) = g_k(\boldsymbol{x}_n, \theta_k^g)$ is the posterior conditional probability that $\boldsymbol{x}_n$ is assigned to the k'th expert, and $p(y_n|\boldsymbol{x}_n, \theta_k^e)$ is the probability distribution describing the k'th expert, parametrised by θ_k^e where $w_k \in \theta_k^e$. Given N training data points, the joint likelihood is given by

$$P(Y, \boldsymbol{\Phi}|\boldsymbol{\pi}, \boldsymbol{\theta}^g, \boldsymbol{\theta}^e) = \prod_{n=1}^{N} \sum_{k=1}^{K} p(k|\boldsymbol{x}_n, \pi_k, \theta_k^g) p(y_n|\boldsymbol{x}_n, \theta_k^e) \tag{8}$$

where $Y = [y_1, \dots, y_N]^T$. The joint likelihood given by Equation (8) implies a soft switching between experts such that the gating network determines the contribution of each expert to the output. In order to enforce hard switching, such that one expert is dominant at each sample time, the latent variable $Z \in \mathcal{R}^{N\times M}$ is introduced. The elements $z_{n,k}$ of Z are chosen as $z_{n,k} = 1$ if the k'th expert is dominant at sample time n, and 0 otherwise. The incorporation of the latent variable Z simplifies Equation (8) by allowing the sum to be replaced by a product, such that (8) can be re-written

$$P(Y, \boldsymbol{\Phi}, Z|\boldsymbol{\pi}, \boldsymbol{\theta}^g, \boldsymbol{\theta}^e) = \prod_{n=1}^{N} \prod_{k=1}^{K} \Big(p(k|\boldsymbol{x}_n, \pi_k, \theta_k^g) \times p(y_n|\boldsymbol{x}_n, \theta_k^e) \Big)^{z_{n,k}}. \tag{9}$$

Conjugate exponential distributions are assigned to the probability distributions in Equation (9), with appropriate prior distributions. The choice of distributions and priors are based on those chosen in (Baldacchino & Rowson, 2016) and (Baldacchino, 2018), interested readers are referred to the former for a complete mathematical description. Of interest here is the choice of a Gaussian mixture model (GMM) for $p(k|\boldsymbol{x}_n, \pi_k, \theta_k^g)$, such that the gating network is able to probabilistically partition the potentially high dimensional input space.

The joint distribution of all the random variables in the model is given by

$$p(Y, \boldsymbol{\Phi}, Z, \boldsymbol{\theta}^g, \boldsymbol{\theta}^e, \boldsymbol{a}|\boldsymbol{\pi}) \tag{10}$$

where $\boldsymbol{a}$ is a hyper-parameter introduced by the choice of prior distribution.

In order to infer the posterior distribution of the model parameters $p(\boldsymbol{\theta}^g, \boldsymbol{\theta}^e, \boldsymbol{a}|Y)$, an approximate Bayesian framework must be leveraged, because the marginal likelihood $P(Y)$ is intractable. The choice of the latent variables and conjugate distributions in Equation (9) allows for the inference to be performed via variational Bayesian expectation maximisation (VBEM). Variational inference places the assumption that the posterior distribution over the latent variables and parameters can be approximated by factorised distributions. The Kullback-Leibler (KL) divergence between the variational posterior distribution and the true posterior is used to define a lower-bound on the marginal likelihood and Closed form update equations can be derived by performing a functional differentiation on the lower-bound. The update equations are a Bayesian equivalent to the well known expectation maximisation (EM) algorithm,

VBE-step: The latent variables z are updated by evaluating the distributions over the random variables

VBM-step: The distributions over the random variables using the values of the latent variables, z.

Iterating between the E and M steps is guaranteed to increase the variational lower bound and hence the algorithm is guaranteed to converge to a local maxima.

4. CASE STUDY: ANOMALY DETECTION IN MEASURED ENGINE SIGNALS

In this section, the inter-engine variation analysis framework developed above is applied to a real world data set. The case study demonstrates how the method can reduce the bias in the residuals as well as increase the sensitivity to anomalous engine behaviour.

4.1. Example anomaly

An example of an anomaly that has been observed in real engine data is shown in Figure 2. The signature of this anomaly is characterised by a sharp drop in measured pressure in engine A (orange line, Panel E) followed by a short time-scale increase in high power shaft speed (N2) (orange line, panel B). Following this, multiple engine parameters then operate around a changed set point, here shown for N2 and P30 (Top and Middle panels respectively). This anomaly was successfully detected by automated, on-board EHM, operating on snapshot data (short sections of time series data that are recorded periodically during a flight).

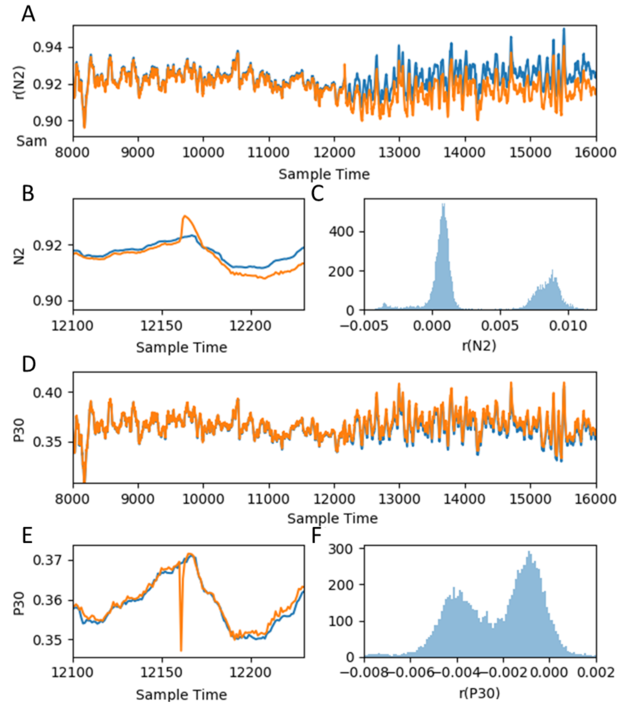


Figure 2. Example anomalous behaviour detected in real engine data. A) N2 time series for both sister engines, B) zoom in on anomaly location in N2, C) Histogram of $r(N2)$, D) P30 time series for both sister engines, E) Zoom in on anomaly location in P30, F) Histogram of $r(P30)$.

A data set containing a range of measured engine parameters across a series of consecutive flights has been collected. The data set contains 18 flights of data, the anomaly was detected at flight 15.

4.2. Implementation

In order to perform automatic anomaly detection on the described data set the inter-engine variation analysis procedure described above is performed. Inter-Engine residuals are computed for the measured engine data for each flight. In order to train the MoE model the model inputs and outputs must first be defined. The model output is chosen as r(P30), where $r(\cdot)$ indicated the inter-engine residual. The model inputs are chosen as r(T20), r(P20) and r(N1). T20 and P20 are the temperature and pressure at the inlet to the combusted respectively and N1 is the low power shaft speed. The inputs are enriched by the inclusion of second order polynomials and a DC term.

The MoE model is initialised by assigning each data point to an expert at random. The model is trained using the first data set (the first flight in the series). The model is then used to

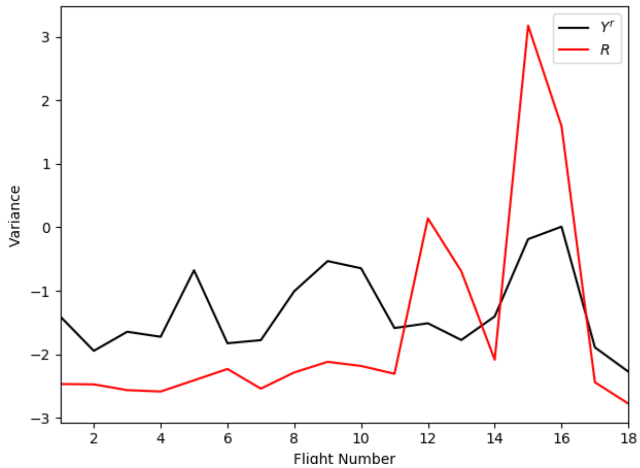


Figure 3. Variance of inter-engine residuals and normalised inter-engine residuals across a series of consecutive flights.

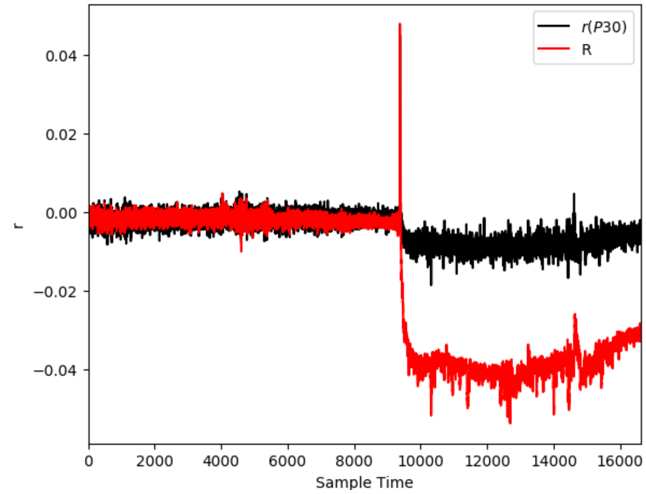


Figure 4. Time series inter-engine residuals and estimated normalised inter-engine residuals for flight 15.

predict the chosen output inter-engine residual for each of the remaining data sets using the respective model inputs in order to compute the normalised inter-engine residual.

In order to show that the bias in the residuals has been reduced histograms of the predicted output parameter are produced. The method considered to be successful if 1) $var(R) < var(y)$ for healthy data sets, and 2) if $var(R) > var(y)$ for data sets containing anomalies.

4.3. Results and discussion

It is observed that the variance of the normalised inter-engine residual is less than the variance of the output inter-engine residual for the majority of the flights (Excluding flights 12, 13, 15 and 16), see Figure 3. For these cases the method is successful in reducing the explainable variation in the inter-engine residuals, resulting from the difference in engine condition or from measured external disturbances. Histograms of the inter-engine residuals and normalised inter-engine residuals for the first four flights show that the procedure successfully normalises the data and reduces the DC offset that is present in the measured residuals, see Figure 5, top panels.

The four flights which coincide with the instances where the normalised inter-engine residual is greater than the variance of the output inter-engine residual have observably different histograms to the other flights. In each case, a significant portion of the data is distributed away from the origin, indicating large residuals, see Figure 5, bottom panels. A manual investigation into the data sets for flights 12, 13 and 16 has discovered instances of the signature originally observed in flight 15. These anomalies were not originally detected using current EHM methods. This may be because they were either of a smaller magnitude, or because they fell outside of the snapshot data.

The detection of new anomalies in the data set indicates that the normalised inter-engine residual has greater sensitivity to the presence of anomalies. The histograms of the inter-engine residuals and normalised inter-engine residuals for flights 12, 13, 15 and 16 show that the nominal data remains distributed around zero while the anomaly has caused new distributions located at an offset. see Figure 5, bottom panels. In Flight 15, for which current methods were successful in identifying the anomaly, significant peaks are seen in both the inter-engine residuals and normalised inter-engine residual, although the effect is more severe in the latter case.

In flights 12, 13 and 16 the anomalies have little effect on the distribution of inter-engine residuals and so the anomalies may not be detectable. The variance of the inter-engine residuals is also comparable to the nominal flights, such that it is not possible to differentiate the anomalies from other sources of variation, see Figure 3. In comparison, there is a clear difference in the magnitude of the variance of the normalised inter-engine residuals between nominal flights and those containing anomalies. Observing the residuals as a time series this effect is clear, see Figure 4. Before the anomaly occurs the normalised inter-engine residual (red) has a smaller variation than the inter-engine residual (Black), the anomalous engine behaviour (starting at around sample time 9000) causes a much larger difference in the normalised inter-engine residual than in the inter-engine residual.

5. CONCLUSION

The purpose of this study is to demonstrate a method for reducing the influence of both normal differences in engine health and measured external disturbances in the inter-engine residuals, while being sensitive to anomalous behaviour. To this end, inter-engine variation analysis is introduced, making use of inter-engine residuals as input-output data to train a MoE model. The MoE model can make predictions on the

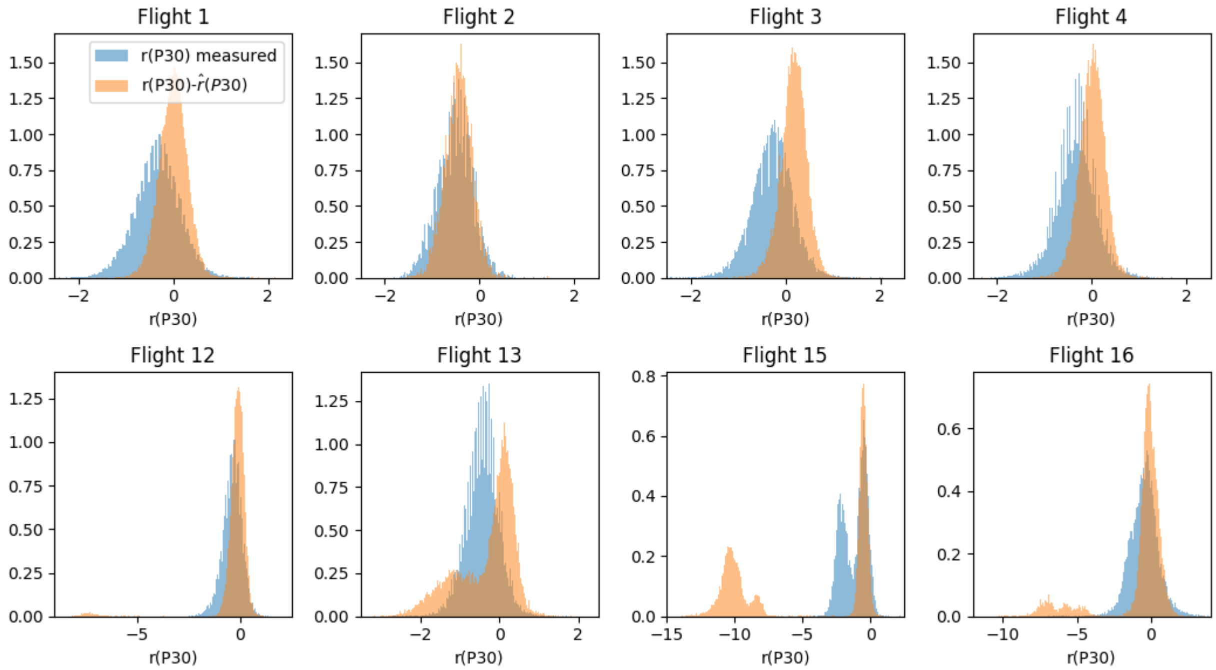


Figure 5. Healthy data figure.

output for new input data. A normalised inter-engine residual is then computed which is used as a new variable for the purpose of anomaly detection.

Methods for the detection of anomalies are not discussed, rather it is demonstrated that the normalised inter-engine residual is more amenable to such methods. It hence replaces the traditional method of novelty detection that requires the generation of a normality model. Such a normality model may be hard to achieve for a system that displays highly complex behaviour such as a GTE. The case study demonstrates the application of the developed method on a real data example. The method is successful in both reducing inter-engine variation due to explainable behaviour, as well as increasing sensitivity to anomalous behaviour. The detection of previously undetected fault signatures in the data set indicates an increase in detection performance in comparison to current methods. The method has been implemented by our industrial partners and has been demonstrated to perform well in the detection of anomalous behaviour with a significant reduction in false positives.

REFERENCES

Baldacchino, T. (2018). Simultaneous force regression and movement classification of fingers via surface emg within a unified bayesian framework. *Frontiers in Bioengineering and Biotechnology*, *6*, 13.

Baldacchino, T., & Rowson, J. (2016). Variational bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems. *Mechanical Systems and Signal Processing*, *66*, 178–200.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 15.

Isermann, R. (2006). *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media.

Jacobs, R. A., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural computation*, *3*(1), 79–87.

Miljković, D. (2011). Fault detection methods: A literature survey. In *Mipro, 2011 proceedings of the 34th international convention* (pp. 750–755).

Ueda, N., & Ghahramani, Z. (2002). Bayesian model search for mixture models based on optimizing variational bounds. *Neural Networks*, *15*(10), 1223–1241.

Yan, W., & Yu, L. (2015). On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach. In *Proceedings of the annual conference of the prognostics and health management society.*

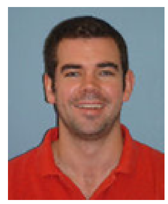
BIOGRAPHIES

William R. Jacobs received the M.Phys. degree in physics with mathematics from the University of Sheffield, Sheffield, UK, in 2010 and the PhD degree in nonlinear system identification from the University of Sheffield in 2016. He is currently working as a Research Associate in the Rolls Royce University Technology Centre, Department of Automatic Control and Systems Engineering at the University of Sheffield. His research interests include nonlinear system identification, fault detection and health management.

Huw L. Edwards received his Masters in Electrical and electronic engineering from the university of Bristol, UK in 2006. He has worked at Rolls-Royce for 10 years in a wide range of fields, initially focused in electrical engineering blue sky concepts including electrically powered aircraft and electric engine accessories. His recent focus has been in measurement systems and predictive health monitoring analytics for gas turbine engines. He is a Member and chartered engineer with the IET.

Visakan Kadirkamanathan received the B.A. and Ph.D. degrees in Electrical and Information Engineering from the University of Cambridge, UK. He held Research Associate positions at the University of Surrey and the University of Cambridge before joining the Department of Automatic Control and Systems Engineering, The University of Sheffield, as a Lecturer in 1993. He is currently the Head of the Department and a Professor of Signal and Information Processing, affiliated to the Centre for Signal Processing and Complex Systems. His research interests include nonlinear signal processing, system identification, intelligent control and fault diagnosis with applications in systems biology, aerospace systems and wireless communication. He has coauthored a book on intelligent control and has published more than 140 papers in refereed journals and proceedings of international conferences.

Professor Kadirkamanathan is the Co-Editor of the International Journal of Systems Science and has served as an Associate Editor for the IEEE Transactions on Neural Networks and the IEEE Transactions on Systems, Man, and Cybernetics, Part B. He was also the Conference Chair of the 4th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2009) held in 2009 at Sheffield, UK.



Andrew R Mills, CEng, PhD, is a Senior Researcher and Research Programme Manager at the University of Sheffield. He has worked in the defence industry on aerospace and automotive applications before his current post within the University Technology Centre supported by Rolls-Royce. Research interests are in a broad range of aspects related to control and health management systems including: signal processing algorithms, instrumentation, advanced processing technologies and model based design.