Equipment Health Monitoring with Non-Parametric Statistics for Online Early Detection and Scoring of Degradation

Maizura Mokhtar[1], Joseph C. Edge[2] and Andrew R. Mills[1]

[1] *Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, United Kingdom*
{*m.mokhtar, a.r.mills*}*@sheffield.ac.uk*

[2] *Rolls-Royce plc, PO Box 31, Derby, England*
joseph.edge@rolls-royce.com

ABSTRACT

This paper develops a health monitoring scheme to detect and trend degradation in dynamic systems that are characterised by multiple parameter time-series data. The presented scheme provides early detection of degradation and ability to score its significance in order to inform maintenance planning and consequently reduce disruption. Non-parametric statistics are proposed to provide this early detection and scoring. The non-parametric statistics approximate the data distribution for a sliding time window, with the change in distribution is indicated using the two-sample Kolmogorov-Smirnov test. Trending the changes to the signal distribution is shown to provide diagnostic capabilities, with deviations indicating the precursors to failure. The paper applies the equipment health monitoring scheme to address the growing concerns for future gas turbine fuel metering valve availability. The fuel metering unit within a gas turbine is a complex electro-mechanical system, failures of which can be a major source of airline disruption. The application is performed on data acquired from a series of industrial tests performed on large civil aero-engine fuel metering units subjected to varying levels of contaminant. The data exhibits characteristics of degradation, which are identified and trended by the equipment health monitoring scheme presented in this paper.

1. INTRODUCTION

The assessment and trending of novelty within the measured parameters of a dynamic system may be used to diagnose and predict the performance and health of a system, and thus inform activities to reduce the impact of decreasing functional performance. The use of novelty as a measure of health has advantages in that the exact nature of fault characteristics are not required in advance, only a measure of departure from nominal conditions. To generate actionable information, signals are typically processed from raw measurements into a reduced dimension novelty summary value that may be more easily transferred to where it can be trended and interpreted by an asset manager. In line with the aspirations of the novelty detection and trending paradigm to determine any departure from nominal conditions, the novelty assessment scheme should be sensitive to all changes in the underlying system, not only deviations in particular characteristics of signals. A multi-variate equipment health monitoring (EHM) scheme is developed to address these novelty trending objectives.

Early warning of degradation is provided by a novelty scoring metric, which aims to detect the changes in the system dynamic response as the results of the degradation and to trend the degradation significance and severity. The changes in the dynamic response are visible when analysing the measured data distributions. For the work presented in this paper, novelty is defined as the change in the measured signal distribution when compared to a reference distribution, generated from a previous known condition or from its earlier behaviour. The principle of our novelty detection scheme is supported by Andrade et al. (2001) which states: “data derived from measurements taken from an undamaged system will have a distribution with an associated mean and variance; if the system is damaged, then, there may be a change to its mean, variance, or both”. Online indication and trending of the distribution change, with any order of statistical moments (Scheffer & Heyns, 2001), (Salgado & Alonso, 2006), enable the indication of the system health condition.

Because of this, the proposed EHM scheme does not require an explicit model of normality to be constructed as part of the design and development process. This is in contrast to the work published in (Sohn et al., 2001) and other similar works, (Andrade et al., 2001) (Hall & Mba, 2004), (Kar & Mohanty, 2006), (Subramaniam et al., 2006) and (Zhan & Mechefske, 2007). These papers compare the measured dynamic response against the model of normal(s) and/or fault conditions - therefore, causing a disadvantage because of the requirement for prior knowledge. Our work also enables the early detections of the onset of change in dynamic response, which is also indicative of the degradation.

The novelty scoring is achieved using online non-parametric statistics that approximates the data distribution for the time window consisting of the N number of samples at current time t and compare to the previous N number of samples separated by an interval of S samples. A non-parametric statistical approach is proposed so that this scheme will not be reliant on the prior training of normal and faults conditions. This is a major criteria for the development of the online and unsupervised EHM scheme, because, as indicated in (Modenesi & Braga, 2009), novelty detection is concerned with the identification of unexpected events or regime changes to the system that is not well understood - "The vagueness of the description is inherent to the novelty detection problem, in fact, it is the very centre of the problem: how to detect data whose only particular characteristic is that it has not appeared before?". Furthermore, when variations occur, the variations may cause the need to redesign and reconstruct any system models developed; the development process itself is time consuming, and may not reflect all normal or fault conditions (Zhan & Mechefske, 2007).

By using an online non-parametric statistics (Subramaniam et al., 2006), the approximation of the data distribution adapts over time. The characteristics of the distribution will differ when the conditions of the system have changed, thus changes that are resultant of degradations are identified. Novelty, defined by this work, is the identification of the changes to the distribution, which signifies when a change in the system's conditions have occurred, i.e. the measured dynamic response that is the outcome of degradations. Authors of (Marsland, 2003) and (Modenesi & Braga, 2009) also indicated that novel data or outliers have a large effect on the analysis of the system, which can result in the change to the measured data distribution.

One mechanism to monitor the distribution change is by trending the change in the distributions mean, standard deviation and other statistical moments (e.g. skewness or kurtosis). These summary statistics are not guaranteed to unambiguously measure all the different changes that may occur in the data. In addition, as the number of variables in the analyses increases, the co-relations between parameters should also be calculated, and thus the number of calculations increases non-linearly ($O(n^2)$). Modern complex systems have a combination of multiple sensed parameters that all may contribute to the efficacy of monitoring (Subramaniam et al., 2006). Therefore, an alternative generic measure of distribution change is advantageous and is proposed in this paper.

We apply this scheme to a component previously identified as a source of high disruption and service cost to aero-engine manufacturers (Eleffendi et al., 2012). The fuel metering unit (FMU) within a gas turbine is a complex electro-mechanical system. Failures to the FMU can be a major source of airline disruption. The system operates in a harsh environment where high temperatures and fuel impurities can lead to system degradation and functional failure. Fuel impurities, often categorised as contaminants, are one of the culprits that cause system degradation. Contaminants accumulate in fuel system filters, nozzles, the walls of control valves and other sliding components. These accumulations resulted in increased friction, which can, in addition to other failure mechanisms, result in valve seizure and in-flight shutdown. Early detection of this degradation can inform maintenance planning and avoid in-service events, which helps minimise disruptions.

The paper presents the multivariate EHM scheme that performs early diagnosis and trending of the FMU degradation as a result of friction increase. The EHM scheme uses non-parametric statistics. The non-parametric status is discussed in Section 2. Section 3 describes the FMU used to test and analyse the capabilities of the EHM scheme and Section 4 discusses the results produced. Section 5 concludes the paper.

2. NON-PARAMETRIC STATISTICS FOR NOVELTY DETECTION AND SCORING

The novelty detection scheme proposed is performed by comparing the differences between the two distributions: the current distribution and the previous distribution measured. If the system is in nominal conditions and at non-transient operations, the change should be minimal. If the system performance degrades, a change in the distribution between current and previous is indicated. Changes in the distribution are indicated using a multivariate two-sample Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test signifies the probability whether the two underlying probability distributions differs. The test compares two empirical cumulative distribution functions (ECDFs) and for the work presented in this paper, the two ECDFs are the current and previous distributions. This enables trending of any system change.

2.1. Multivariate Two-sample Kolmogorov-Smirnov Test

Since different data sets, or different distribution functions, have differing cumulative density functions, one can establish the likelihood that two sets of data are originating from the same distribution function by measuring the differences between their ECDFs. The ECDF for the N samples of variable v is defined by Eq. (1), and provides a measure of the relative number of samples for v, $v = \{u_1, u_2, ..., u_N\}$, less than or equal to x. $\mathbf{1}\{u_i \leq x\}$ is the indicator of such an event.

$$\mathrm{ECDF}_v(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{u_i \leq x\} \tag{1}$$

The two-sample Kolmogorov-Smirnov test compares the two ECDFs by calculating the statistical distance D between the two distributions. The statistical distance D is given by Eq. (2), where $F(x)$ and $R(x)$ are the samples from the ECDFs of $F(\mathrm{x}_1)$ and $R(\mathrm{x}_2)$ respectively (Andrade et al., 2001).

$$D = \max_{-\infty<x<\infty} |F(x) - R(x)| \tag{2}$$

The statistical distance D is converted into a similarity probability using the Kolmogorov-Smirnov p value, defined by Eqs. (3)–(4) (Greenwell & Finch, 2004) (Kar & Mohanty, 2006). The p value provides the metric for novelty scoring.

$$p = Q_{KS}(z) = 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2z^2) \tag{3}$$

$$z = D\sqrt{\frac{N_1N_2}{N_1+N_2}} \tag{4}$$

N_1 is the number of points in $F(\mathrm{x}_1)$ and N_2 is the number of points in $R(\mathrm{x}_2)$. Equation (3) is for when N_1 and N_2 tends to infinity (Kar & Mohanty, 2006).

p-value is a monotonic function with limiting values of:

$$p = Q_{KS}(z) = \begin{cases} 1 & \text{if } z \to 0 \\ 0 & \text{if } z \to \infty \end{cases} \tag{5}$$

If the two distributions are statistically similar (similar ECDFs), Q_{KS} tends towards 1. If the distributions are different, i.e. varied, Q_{KS} will go towards 0. A variation between the two distributions indicates that a novelty has occurred.

In the work presented in this paper, the $F(\mathrm{x}_1)$ and $R(\mathrm{x}_2)$ are the product of the single variate ECDF_v in the multivariate data, calculated using Eq. (6).

$$F(x) = \prod_{v=1}^{V} \mathrm{ECDF}_v(x) \tag{6}$$

where V is the number of variables considered.

Novelty is indicated when $p < 0.90$. $p < 0.90$ is chosen because, based on the critical value approximation which indicates:

$$H = \begin{cases} 0 & \text{if } D < D_{\text{critical}} \\ 1 & \text{if otherwise} \end{cases} \tag{7}$$

$H = 0$ when the two distributions are the same and $H = 1$ if otherwise, D_{critical} is equated using Eq. (8) (Kar & Mohanty, 2006).

$$D_{\text{critical}} = \alpha\sqrt{\frac{N_1+N_2}{N_1N_2}} \tag{8}$$

Assuming that the distribution of the D-values produced is normal and the sizes of $F(\mathrm{x}_1)$ and $R(\mathrm{x}_2)$ are N_1 and N_2 respectively, novelty is indicated when D (Eq. (2)) is above the 2.698σ or the upper quartile of the D distribution. $\alpha = 0.57$ produces the D_{critical} value, for which any values of D beyond or equal to D_{critical} will produce $p < 0.9$.

2.2. Offline and Online Novelty Trending

In order to trend degradation, the capability provided by the previous section must be augmented with the ability to look at parameter distribution change over time. The distributions under comparison should therefore be sampled as two windows of data separated by an appropriate time interval. Two modes of operation are outlined in this paper to provide this measure of change as a function of time:

1. *Offline:* This strategy compares the distributions from the first flight to all other complete flights. In effect, the first flight is used to build a model of normal, and the *offline* test observes the divergence of the system over its lifetime as an analogue to deterioration. Therefore, the analysis performed compares how the subsequent cycles differ from the first cycle: N_1 = number of samples in the x-th cycle and N_2 = number of samples in the 1-st cycle. This methodology will only detect deterioration at a period of complete flights.
2. *Online:* A sliding window approach is employed to enable un-delayed detection and scoring of novelties, therefore allowing indication of novelty occurring during a flight. The sliding window approach is further discussed in the next section.

2.3. Online Trending the Changes to the Distributions: The Sliding Window Approach

The online strategy addresses the trending of novelty by accumulating parameter distribution changes occurring in a time period much less than the typical prognostic horizon of the system degradation. It has been observed that the frequency of distribution changes is indicative of the deterioration for the failure modes explored in this paper. Measures of this trend are termed 'health metrics' and are calculated in two ways: as an average probability of change and as a count of changes per cycle.

The construction of the health metrics involves first applying the multivariate Kolmogorov-Smirnov test to two consecu-

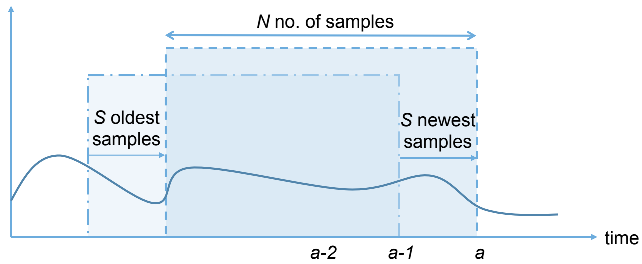


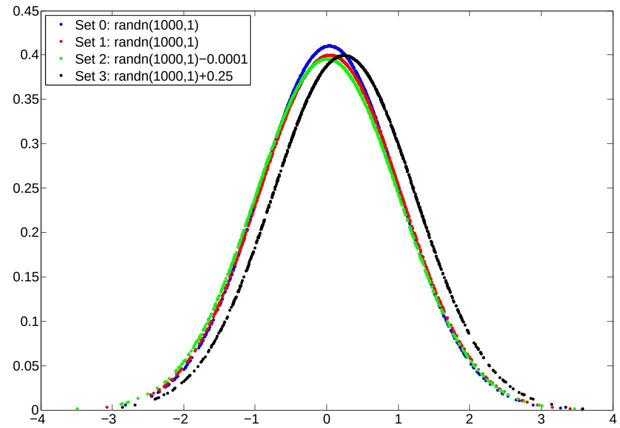
Figure 1. Sliding window. The p values are calculated at a by comparing the distributions of N samples at $a-1$ and at a when the oldest S samples values are replaced with S newest values.

tive sliding windows of data containing N number of samples. The two sliding windows, separated by S number of samples, are used to construct individual multivariate ECDFs (Eq. (6)). The change in distribution is then constructed. The first p probability is calculated when the first and second distributions are obtained, with $N+S$ number of samples, and are subsequently calculated at every interval a when S new samples are obtained. This is as illustrated in Fig. 1.

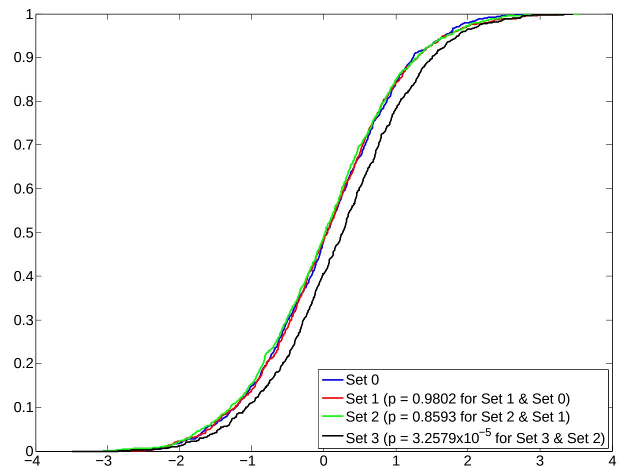
Figure 2 illustrates the concept on synthetic data. The four distributions in the figure are each generated from a window of $N = 1000$ samples selected at different times from Gaussian distributed data with time-increasing mean offset, (Fig. 2a). Formation of an ECDF from the data and calculating the maximum distance D (Eq. (2)) allows changes in distributions to be indicated by the p probability calculated using Eq. (3). A comparison between Set 0 and 1, shows a high similarity ($p = 0.98$), becoming progressively lower as the distance between distributions increases. The lowest p value (approximately 0) occurs when a significant change in the distribution is indicated between Set 2 and 3. Indicating and trending the changes in distribution are useful to identify the deteriorating conditions of the system.

Equations (3)–(4) show the relationship between the D value and its associated p probability of distribution change. The p value decreases exponentially with the increase in the D-value, therefore only when a significant change in the distribution is detected will there be a decrease in the probability of similar distributions. The confidence of novelty is the complement of the probability given by the p value (i.e. $1-p$). When the system is at nominal conditions and at non-transient operations, the change in distribution should be minimal, with the probability of change given by Eq. (3), $p \geq 0.90$.

The p value, therefore, can be used to visualise the measure of health for the system at any given time. The trend in p may be observed by calculating the running average of the p values at every a during the period of interest (for example a flight), Eq. (9).



(a) The Gaussian PDF of the generated data.



(b) The respective ECDFs and the associated p values when a set is compared against another set.

Figure 2. Kolmogorov-Smirnov p values (Eq. (3)) indicating the change in the distributions.

$$p_{rAve}(a) = \frac{\sum_{i=1}^{a} p(i)}{a} \tag{9}$$

When no novelty is occurring in the system, $p_{rAve} \approx 1$. The value of p_{rAve} decreases with the increase in the rate of novelty detection. Trending the change to the p_{rAve} value shows the severity of the system degradation.

2.4. Nominal and Non-Transient Operations

The nominal and non-transient phase of operations are only investigated at presence. This is to enable the proof of concept of the novelty scoring ability for the proposed method. Furthermore, current aircrafts use an Aircraft Condition Monitoring System (ACMS) to acquire the data for the EHM, and the acquisition of the data is perform at the three defined

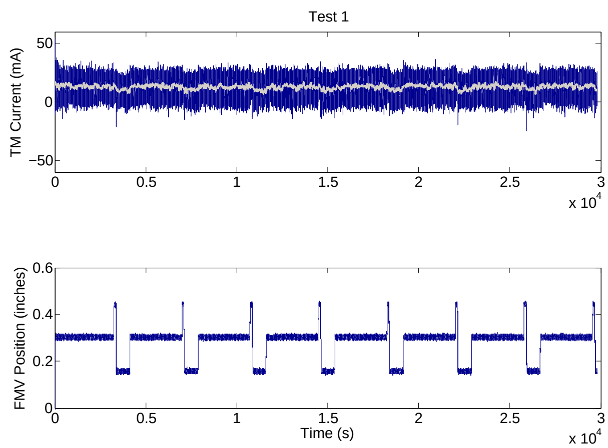


Figure 3. Test 1 (Baseline): Minimal contaminant detected. Small changes in the mean of the TMC are indicated for this test. The mean is indicated by the grey line. The FMV position is averaged at 0.3 inches

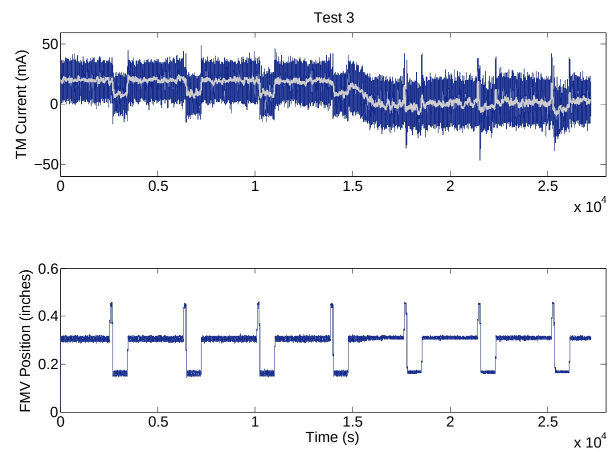


Figure 5. Test 3: Contaminant resulted in stiction. Changes to the TMC mean and FMV positions are shown with the increase in the contaminant level.

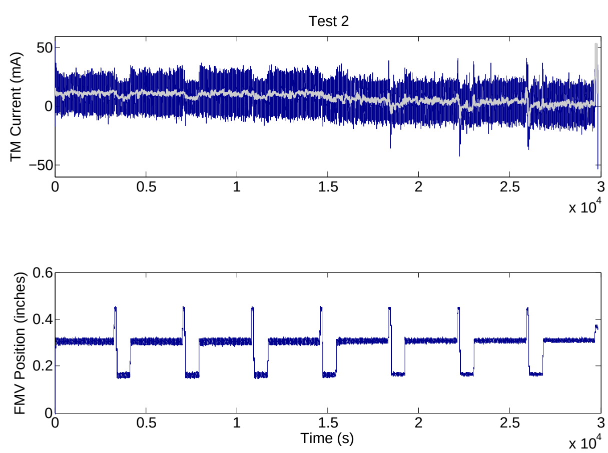


Figure 4. Test 2: Contaminant causing stiction. There are changes to the TMC mean and the FMV positions as the result of the degradation.

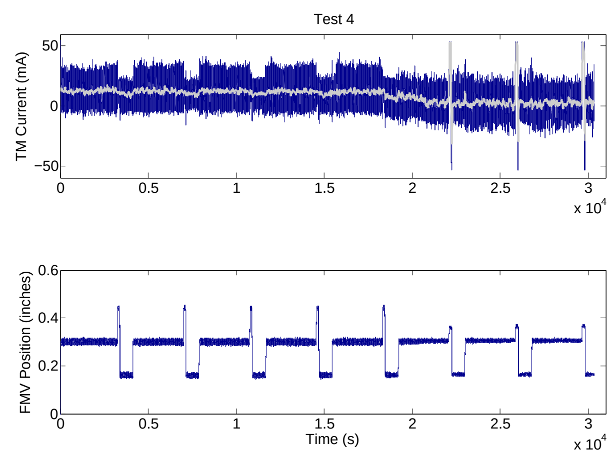


Figure 6. Test 4: Contaminant causing stiction.

phases: take-off, cruise and landing. The EHM then summaries the health of the engine at these phases separately (Waters, 2009). Therefore, we envisage the novelty scoring to be calculated separately at each of these phases. Future work will include understanding the data distribution trends when operating at the transient phase, and to derive the novelty scoring metrics for the nominal and transient operations.

3. EXPERIMENT: NOVELTY DETECTION OF FUEL METERING UNIT

The presented equipment health monitoring scheme is used to detect and trend the degradation of a gas turbine fuel metering unit (FMU). The primary function of a FMU is to regulate fuel flow in response to the Electronic Engine Control (EEC) demand required to deliver commanded engine thrust. This is achieved through position control of two-stage servo fuel metering valve (FMV), which alters the pressure drop across the valve and flow rate through it.

The functional failures associated with the FMU are the loss of FMV bandwidth with poor demand tracking, leading to the inability to control valve position and fuel flow. These, as indicated in Section 1, may be due to debris ingestion resulting in valve friction/stiction or filter clogging.

Data has been collected from fuel system rig tests, which were subject to the introduction of fuel contaminant, and run over up to 8 cycles of cruise, idle and take-off phases. These

Table 1. Mass (as percentage of maximum test amount) of contaminant introduced per test cycle.

Cycle #	Test 1: Baseline	Test 2	Test 3	Test 4
1	13.29	2.85	12.03	0.00
2	18.99	0.95	4.43	4.11
3	22.15	5.06	3.80	2.85
4	20.25	8.54	7.59	7.59
5	24.05	9.49	11.71	4.75
6	36.39	37.34	23.10	19.30
7	35.44	43.04	28.80	34.18
8	69.62	100.00	31.96	98.73

tests exhibited functional failures from loss of metering valve control at high contaminant levels and serve as a basis for evaluating the outlined novelty trending schemes.

The EHM scheme presented is used to indicate how the system degrades as the result of the contaminant introduction. At present, the analysis of the scheme is to indicate the degradation only when the engine is supposedly at the cruise phase. This is shown in Figs. 3–6 when the FMV position is averaged at 0.3 inches. In all three tests, the mean at cruise of the TMC reduces as the control system compensates for the effects of the increase in the contaminant level.

The mass of contaminant introduced in each cycle is listed in Table 1. It should be noted this is not a measure of degradation, and only indicates the mass of particles introduced to the system at each cycle presented as a percentage of maximum cycle dosage over all tests.

Two signals are initially chosen for use to monitor the degradation level in response to the introduction of the contaminant. They are:

1. The torque motor current (TMC), and
2. The fuel metering valve (FMV) position.

The TMC values and the FMV position values are sampled at 40Hz, and are normalised (Eq. (10)) so that their values are between -1 to +1 prior to the analysis.

$$x_n = (b - a) \times \frac{x_o - x_{min}}{x_{max} - x_{min}} + a \qquad (10)$$

x_n is the normalized value and x_o is the value to be normalized. a and b are the minimum and maximum value of the range to be normalized to, which in this case is $a = -1$ and $b = +1$. x_{max} and x_{min} are the maximum and minimum values of the range of x_o.

Figure 3 indicates the values of these variables when the majority of the contaminants introduced are captured by the low pressure (LP) filter. The LP filter traps the contaminant upstream of the metering valve, therefore preventing stiction and degradation. Physical analysis of this test also indicates that only a small amount of contaminant is detected in the FMU as the results of the filtering, too small to cause degradation. Because of this, minimal changes in the system dynamic response are shown, despite the contaminant introduction. This test acts as the baseline test (Test 1: Baseline) to evaluate the capabilities of the presented EHM scheme.

In tests 2–4 (Fig. 4–6), the system degrades over time and with the increase in the contaminant level introduced per flight cycle.

For the sliding window approach, three different window sizes are considered: 60 seconds of data, $N = 2400$ number of samples; 120 seconds of data, $N = 4800$ samples; and 300 seconds with $N = 12000$ samples. The distribution of the sensors values are updated and compared when the oldest S sample values were replaced with the newest S values, at every a. Six different sets of N and S are analysed:

1. For 60 and 120 seconds of data ($N = 2400$ and $N = 4800$ samples): $S = 40$ samples (1 second of data).
2. For 60 and 120 seconds of data ($N = 2400$ and $N = 4800$ samples): $S = 80$ samples (2 seconds of data).
3. For 300 seconds of data ($N = 12000$ samples): $S = 200$ samples and $S = 400$ samples (5 seconds and 10 seconds, respectively).

4. RESULTS

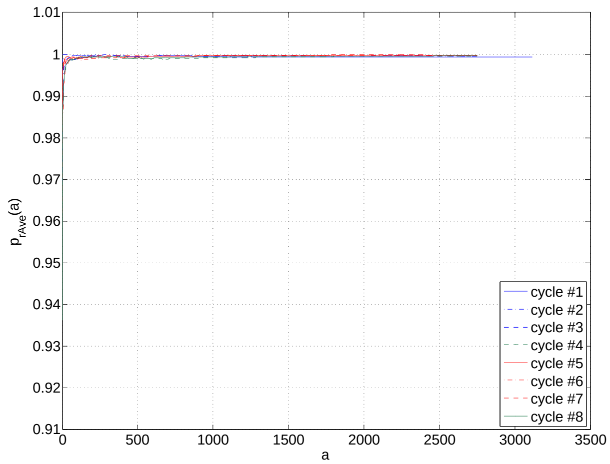
4.1. Offline novelty trending

Figure 7 represents the D values produced by the analysis using the offline strategy comparing each subsequent cycle to the first cycle. The figure shows, for all tests other than the baseline (i.e. not Test 1), a large change to the cycles' distributions are shown when they are compared to their initial cycle. The significance in change is indicated by the large increase in the D value for each cycle, caused by the change to the FMU system dynamics. The D-values remain approximately the same for Test 1: Baseline.

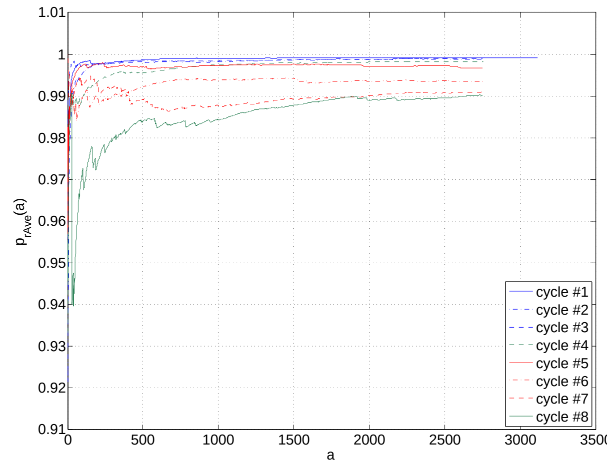
The D values are directly used in the offline analysis. The significant changes between each cycle result in large D values being produced by the comparative analysis, these all result in the the probability of no change tending to a very small value ($p \to 0$). Therefore, for the offline cycle-to-cycle mode of comparison, the novelty detection is made based on the D values instead of its $p-$values.

4.2. Online novelty trending

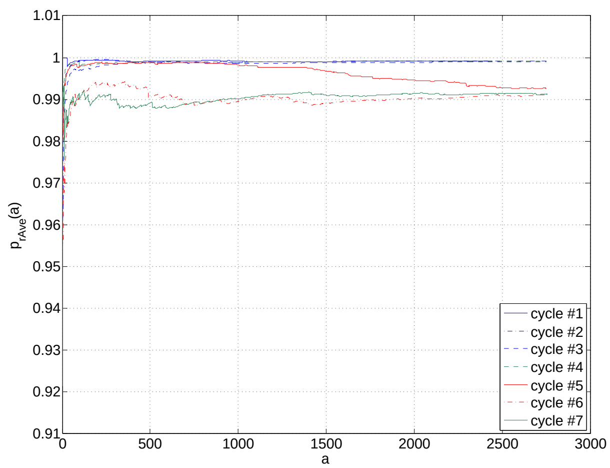
This section evaluates the performance of the two on-line trending approaches as described in Section 2.3. Tables 2–4 show the number of changes to the distributions cycle (i.e. the count of evaluations when $p < 0.90$). Low detection rates are shown for Test 1: Baseline, as most of the contaminants are filtered prior to the metering valve. Frequent changes in distribution is shown by higher count values when the con-



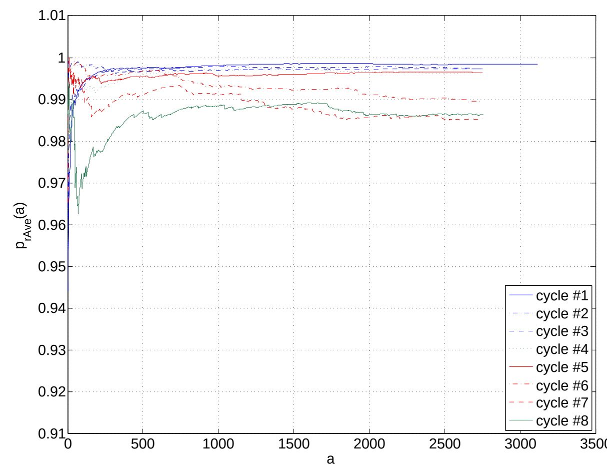
(a) Test 1



(b) Test 2



(c) Test 3



(d) Test 4

Figure 8. The EHM health metric provided by the $p_{rAve}(a)$ values with $N = 2400$ samples (60 seconds of data) and $S = 40$ samples (1 second of data).

Table 2. Number of occurrences when $p < 0.90$, when $N = 2400$ number of samples (60 seconds of data).

Cycle #	N = 2400 samples, S = 40 samples				N = 2400 samples, S = 80 samples			
	Test 1: Baseline	**Test 2**	**Test 3**	**Test 4**	**Test 1: Baseline**	**Test 2**	**Test 3**	**Test 4**
1	**0**	**0**	**0**	**0**	144	168	110	294
2	**0**	**0**	**0**	**0**	100	177	117	332
3	**0**	**0**	**0**	**1**	59	190	148	356
4	**0**	**0**	**0**	**2**	100	216	110	352
5	**0**	**17**	**59**	**1**	57	223	354	379
6	**0**	**53**	**70**	**81**	79	314	384	395
7	**0**	**72**	**68**	**128**	79	352	454	452
8	**0**	**84**	**N/A**	**114**	81	290	N/A	376

taminants were not filtered from the unit (Tests 2–4).

The two tables show that the optimal N:S ratio for the EHM scheme is 60:1 (indicated in bold). Any increase to the ratio will result in a higher number of false detection, i.e. higher number of false detection for the baseline test (Test 1) when alternative ratios are used. Results also show that the window with 60 seconds of samples ($N = 2400$ samples) and 1 second interval ($S = 40$ samples) is sufficient for detection of

Table 3. Number of occurrences when $p < 0.90$, when $N = 4800$ samples (120 seconds of data).

Cycle #	N = 4800 samples, S = 40 samples				N = 4800 samples, S = 80 samples			
	Test 1: Baseline	**Test 2**	**Test 3**	**Test 4**	**Test 1: Baseline**	**Test 2**	**Test 3**	**Test 4**
1	0	0	0	0	**20**	**20**	**13**	**57**
2	0	0	0	0	**19**	**24**	**15**	**73**
3	0	0	1	0	**8**	**30**	**19**	**72**
4	0	0	1	0	**9**	**34**	**15**	**87**
5	0	0	35	1	**3**	**50**	**120**	**84**
6	0	0	40	51	**10**	**87**	**122**	**128**
7	0	2	38	58	**8**	**95**	**180**	**122**
8	0	6	N/A	55	**10**	**94**	**N/A**	**116**

Table 4. Number of occurrences when $p < 0.90$, when $N = 12000$ samples (300 seconds of data).

Cycle #	N = 12000 samples, S = 200 samples				N = 12000 samples, S = 400 samples			
	Test 1: Baseline	**Test 2**	**Test 3**	**Test 4**	**Test 1: Baseline**	**Test 2**	**Test 3**	**Test 4**
1	**25**	**48**	**36**	**82**	61	76	58	106
2	**15**	**46**	**15**	**92**	37	75	39	105
3	**15**	**48**	**32**	**89**	37	73	60	99
4	**11**	**46**	**12**	**81**	37	77	33	89
5	**11**	**43**	**114**	**99**	33	64	128	114
6	**16**	**62**	**105**	**92**	32	86	124	107
7	**7**	**78**	**160**	**123**	28	92	155	116
8	**10**	**74**	**N/A**	**107**	35	98	N/A	111

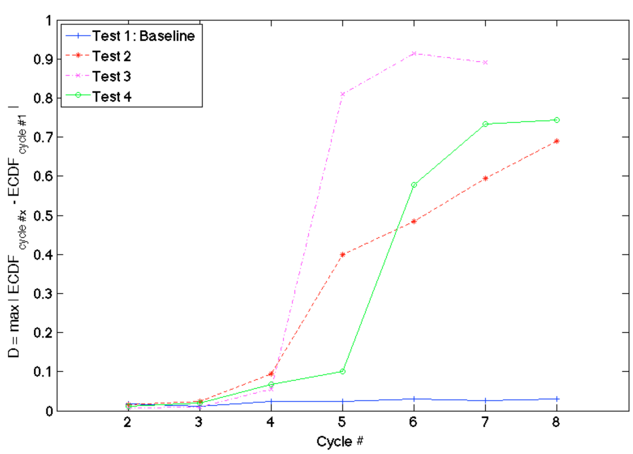


Figure 7. The D values when comparing a cycle to its first cycle.

the degradation. This because of the no (zero count) detections for Test 1: Baseline, as well as the ability to detect the changes to the TMC and FMV positions for Test 2–4.

The alternative health metric, $p_{rAve}(a)$, for each test for $N =$ 2400 samples and $S = 40$ samples is shown in Fig. 8. For all non-baseline test (Test 2–4), the values of the $p_{rAve}(a)$, which is an indicative of the health of the system, reduces overtime. This shows that the health of the system has degraded with time with the increase in contaminant per flight cycle. $p_{rAve}(a)$ are constant and are ≈ 1 for Test 1: Baseline, indicating no system degradation because the LP filter has trapped the contaminant upstream of the metering valve, therefore preventing degradation. The decrease in health also indicates the increase in the novelty detection rate.

4.3. Univariate vs Multivariate

As indicated in Section 1, a gas turbine is complex electro-mechanical system. Determining the most effective parameter for analysis is not always apparent. If one is to perform univariate analysis, the incorrect selection of sensing parameter will lead to a different outcome. For example, if one chooses the FMV position to indicate novelty for N = 2400 samples (60 seconds worth of data) and S = 40 samples (1 second worth of data), as shown in Table 5, no trending of degradation is achievable. Similar observation is shown when the analysis is performed using the TMC's distributions for N = 2400 and S = 40. The short time interval between time windows is not sufficient to identify the changes in these co-related variables when they are treated in isolation.

The analysis presented earlier in this paper is for bivariate analysis ($V = 2$ in Eq. (6)). Figure 9 and Table 6 show the results when increasing the number of variables, V, analysed from the measured rig test data, at the optimal N:S ratio of 60:1 (N = 2400 samples of data and S = 40 samples). The last recorded p_{rAve} values for $V = \{3, 4, 5\}$, i.e. the cycle average p value, decrease with the increase in the level of contaminants for all non-baseline tests (Fig. 9). The detection event count also increases with the increase in contaminants (Ta-

Table 5. Number of occurrences when $p < 0.90$, when $N = 2400$ samples and $S = 40$ samples.

Cycle #	Variable # 1 TMC				Variable # 2 FMV position			
	Test 1: Baseline	Test 2	Test 3	Test 4	Test 1: Baseline	Test 2	Test 3	Test 4
1	0	0	0	0	33	11	15	8
2	0	0	0	0	23	11	41	7
3	0	0	0	0	18	5	33	15
4	0	0	0	0	23	19	28	26
5	0	0	4	0	14	88	49	21
6	0	0	4	4	22	97	14	153
7	0	2	0	4	7	52	39	14
8	0	0	N/A	4	22	7	N/A	12

Table 6. Number of occurrences when $p < 0.90$ for the univariate test with $N = 2400$ samples and $S = 40$ samples.

Cycle #	$V = 3$ (+ MV Downstream Pressure)				$V = 4$ (+ LP Pump Outlet)				$V = 5$ (+ SPR Pressure)			
	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4
1	3	0	0	0	8	3	0	0	6	2	0	0
2	1	1	0	1	0	1	0	0	0	1	0	0
3	0	2	0	0	0	0	0	1	0	0	0	1
4	0	5	0	3	0	4	2	2	0	4	1	1
5	0	73	72	2	1	17	16	0	2	75	25	0
6	0	159	107	118	0	48	19	27	0	119	32	59
7	1	307	99	163	0	120	18	57	1	290	39	118
8	3	366	N/A	158	1	226	N/A	56	2	396	N/A	118

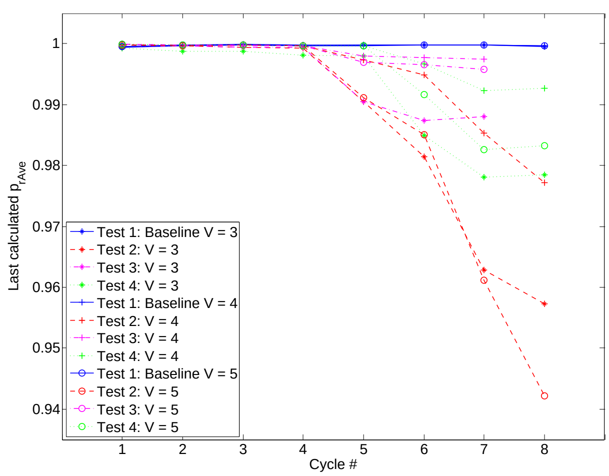


Figure 9. The last $p_{rAve}(a)$ calculated at the end of each cycle for the multivariate analysis related to Table 6.

ble 6). Additional variables included are, for $V = 3$, the normalised metering valve (MV) downstream pressure sensor values, and, for $V = 4$, the fourth variable is the normalised low pressure (LP) supply pressure data. The third analysis is performed with $V = 5$, which adds the servo-pressure.

An analysis of the sensitivity to degradation from these results can be made with respect to the physical interpretation of Figure 9. The average p value for tests where $V = 3$ or $V = 5$ are consistently lower (a change, thus degradation, more likely) than for the test with 4 variables. From this, we conclude that adding the LP supply pressure parameter (in $V = 4$) makes the change to multi-variate distributions less significant, not adding to the ability to determine degradation. The LP pump, thought to be robust to containment itself, is upstream from the valves which are impacted by the containment and therefore is not affected by system degradation. On the other-hand, servo-pressure is controlled by an additional valve to the FMV and therefore introduces sensitivity to another element of the system. The downstream flow pressure (added in $V = 5$) is dependent on the supply pressure (affected by a spill valve) and the valve position, again this valve introduces another candidate source of degradation from the spill valve. It is plausible that these observations could be used to aid fault isolation in future work.

These results corroborate the hypothesis that a combination of multiple sensing parameters is powerful for novelty detection analysis and health scoring of a system, as more dynamics are captured as part of the analysis. The non-parametric two-sample Komolgorov-Smirnoff test provides the mechanism to perform multivariate analysis with minimal pre- or post-processing of the provided data (aside from the normalisation of the data so that their values are between -1 to 1).

5. CONCLUSION

This paper presents the results of a multivariate equipment health monitoring (EHM) scheme that utilises non-parametric statistics. The scheme was developed to provide early detection of the gas turbine fuel metering valve faults, and to enable the scoring of the significance of the degradation. Degradation assessment can inform maintenance planning and consequently reduce disruption.

The scoring is achieved using non-parametric statistics that approximates the data distribution for the time window consisting of the current and the previous samples. The data distribution estimate adapts over time, and a generic measure of difference, a multivariate two-sample Kolmogorov-Smirnov test, is shown to provide diagnosis capabilities. The equipment health monitoring scheme is able to trend the degradation of the fuel metering valves, degradation resulted from the varying levels of contaminant introduced to the engine. Results indicate that the level and rate of detection increases with the increase in the contaminant level, which resulted in the degradations.

As indicated in Section 2.4, the analysis is restricted to cruise phase or at the non-transient phase of flight operations. This is to enable us to present the proof-of-concept capabilities of the novelty scoring metric using the non-parametric multivariate two-sample Komolgorov-Smirnoff test. Future work will include, but not limited to, the analysis of the capabilities of the algorithm to cope with transient phases. We envisioned that a different scoring metric is required to indicate for novelty when the system is in the transient phase of operations.

Two methods for novelty trending are presented in this paper: online sliding window approach and the offline cycle-by-cycle approach (Section 2.2). Schemes to fuse the outputs of these approaches together, along with schemes to trend the outputs over time, may provide advantages in detecting different failure modes and will be investigated.

As presented, the use of multivariate two-sample Kolmogorov-Smirnov test for the EHM scheme simplifies and enhances novelty detection, eliminating the need to choose variables or summary statistics for health analysis prior to system deployment.

ACKNOWLEDGMENT

The authors would like to thank Rolls-Royce plc for their support and direction of this research.

REFERENCES

Andrade, F. A., Esat, I., & Badi, M. N. M. (2001). A new approach to time-domain vibration condition monitoring: Gear tooth fatigue crack detection and identification by the kolmogorov-smirnov test. *Journal of Sound and Vibration*, *240*(5), 909-919. doi: 10.1006/jsvi.2000.329

Eleffendi, M. A., Purshouse, R., & Mills, A. R. (2012). Gas Turbine Fuel Valve Diagnostics. In *Proceedings of 2012 ieee aerospace conference.*

Greenwell, R. N., & Finch, S. J. (2004). Randomized rejection procedure for the two-sample kolmogorov-smirnov statistic. *Computational Statistics & Data Analysis*, *46*(2), 257-267.

Hall, L. D., & Mba, D. (2004). Acoustic emissions diagnosis of rotor-stator rubs using the ks statistic. *Mechanical Systems and Signal Processing*, *18*, 849-868. doi: 10.1016/S0888-3270(03)00050-5

Kar, C., & Mohanty, A. R. (2006). Multistage gearbox condition monitoring using motor current signature analysis and kolmogorov-smirnov test. *Journal of Sound and Vibration*, *290*(1-2), 337-368. doi: 10.1016/j.jsv.2005.04.020

Marsland, S. (2003). Novelty detection in learning systems. *Neural Computing Survey*, *3*, 157-195.

Modenesi, A. P., & Braga, A. P. (2009). Analysis of time series novelty detection strategies for synthetic and real data. *Neural Processing Letters*, *30*(1), 1-17. doi: 10.1007/s11063-009-9106-4

Salgado, D. R., & Alonso, F. J. (2006, February). Tool wear detection in turning operations using singular spectrum analysis. *Journal of Materials Processing Technology*, *171*, 451-458.

Scheffer, C., & Heyns, P. S. (2001). Wear monitoring in turning operations using vibration and strain measurements. *Mechanical Systems and Signal Processing*, *15*(6), 1185-1202.

Sohn, H., Farrar, C. R., Hunter, N. F., & Worden, K. (2001). Structural health monitoring using statistical pattern recognition techniques. *Journal of Dynamic Systems, Measurement, and Control*, *123*(4), 706-711. doi: 10.1115/1.1410933

Subramaniam, S., Palpanas, T., Papadopoulos, D., Kalogeraki, V., & Gunopulos, D. (2006). Online outlier detection in sensor data using non-parametric models. In *Proceedings of the 32nd international conference on very large data bases (vldb '06)* (p. 187-198).

Waters, N. (2009, June). Engine health management. *The Ingenia Magazine*(39), 37-42.

Zhan, Y., & Mechefske, C. K. (2007). Robust detection of gearbox deterioration using compromised autoregressive modeling and kolmogorov-smirnov test statistic-part i: Compromised autoregressive modeling with the aid of hypothesis tests and simulation analysis. *Mechanical Systems and Signal Processing*, *21*, 1953-1982.

BIOGRAPHIES

Maizura Mokhtar holds a BEng (Hons) in Electronics and Computing from Nottingham Trent University, an MSc(Eng) in Control Systems from the University of Sheffield and a PhD in Electronics from the University of York. She has previously worked as a Research Associate in the Department of Electronics at the University of York, and a Postdoctoral Research Assistant in the School of Computing, Engineering and Physical Sciences, University of Central Lancashire (UCLan). Her research interests focus on the development of intelligent autonomous systems, notably the use of computational intelligence algorithms (artificial neural networks, evolutionary algorithms, artificial immune systems) for achieving intelligent and adaptive behaviours in engineering systems.

Joseph C Edge, CEng is a Fuel & Heat Management Specialist at Rolls Royce. He graduated from Warwick University with a BEng System Engineering in 1995. He worked in the automotive industry for 8 years as an Aerodynamicists and Cooling Airflow Engineer before his current role at Rolls Royce.

Andrew R Mills, CEng is a Research Fellow at the University of Sheffield, from where he graduated with a MEng in Control System Engineering. He has worked in the defence industry on aerospace and automotive applications before his current role as Programme Manager and leading Health Management Technologies research within the University Technology Centre supported by Rolls-Royce. Research interests are in a broad range of aspects related to gas turbine system health management and controls.