An Approach to Reducing Input Parameter Volume for Fault Classifiers

Ann Smith Fengshou Gu Andrew D. Ball

Centre for Efficiency and Performance Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

Abstract: As condition monitoring of systems continues to grow in both complexity and application, an overabundance of data is amassed. Computational capabilities are unable to keep abreast of the subsequent processing requirements. Thus, a means of establishing computable prognostic models to accurately reflect process condition, whilst alleviating computational burdens, is essential. This is achievable by restricting the amount of information input that is redundant to modelling algorithms. In this paper, a variable clustering approach is investigated to reorganise the harmonics of common diagnostic features in rotating machinery into a smaller number of heterogeneous groups that reflect conditions of the machine with minimal information redundancy. Naïve Bayes classifiers established using a reduced number of highly sensitive input parameters realised superior classification powers over higher dimensional classifiers, demonstrating the effectiveness of the proposed approach. Furthermore, generic parameter capabilities were evidenced through confirmatory factor analysis. Parameters with superior deterministic power were identified alongside complimentary, uncorrelated, variables. Particularly, variables with little explanatory capacity could be eliminated and lead to further variable reductions. Their information sustainability is also evaluated with Naïve Bayes classifiers, showing that successive classification rates are sufficiently high when the first few harmonics are used. Further gains were illustrated on compression of chosen envelope harmonic features. A Naïve Bayes classification model incorporating just two compressed input variables realised an 83.3% success rate, both an increase in classification rate and an immense improvement volume-wise on the former ten parameter model.

Keywords: Fault diagnosis, classification, variable clustering, data compression, big data.

1 Problem specification

Condition monitoring (CM) is concerned with preventing, or at the very least predicting, impending component failure. Modern process monitoring is complemented with maintenance on demand, maintenance based on real time observations of operations with respect to expected normal behaviour^[1, 2]. With increasing pressure on excellence of process performance and product quality, CM becomes ever more vital. Quality management through continuous monitoring of process outputs aims to detect and identify deviations from normal operation at onset thus ensuring optimal performance and productivity. Reliable and timely indicators are therefore imperative to ensure processes continue running efficiently and without unplanned interruptions. Certainly, there is no shortage of dedicated data collection devices.

As machine complexity continues to increase with respect to both individual components and complete systems, so does the intricacy and cost of maintenance programmes. Reliable methods of assessing and continually monitoring process health are necessary. Incorporating transducers into complex systems to facilitate instantan-

eous information gathering provides a solution. Strategically positioned sensors capture essential process information, vibrations, temperatures, etc. Data can then be relayed to a suitable data acquisition system (DAS) for processing and subsequent real-time analysis. Multifunctional data collection interfaces are capable of recording waveform, digital and marker information and generating simultaneous output measurements for further computer processing and analysis. Sophisticated modelling of signal patterns requires advanced data processing capabilities and multivariate statistical modelling techniques^[3–5].

However, the volume of data collected is often vast and so requires untenable quantities of processing power for pertinent information gain. Resulting algorithms either grind out untimely information or fail to converge due to computational burdens. Plausible methods of restricting input data volumes are sought.

One solution is found in data reduction methods such as principal component analysis (PCA). PCA is frequently used in CM of industrial systems, prevalent in monitoring chemical processes^[6] and often incorporated with wavelet transform methods. Fault classification using PCA extensions^[7] and recently emerging Kernel PCA (KPCA)^[8–10] are increasingly employed. Kernel based PCA extends the method for use in non-linear or overlapping cluster applications. Wavelet transform methods are evolving as an alternative feature extraction method to

Research Article

Manuscript received January 5, 2018; accepted October 19, 2018; published online January 23, 2019

Recommended by Associate Editor Dong-Hua Zhou

© The Author(s) 2018

Fourier much favoured in [9]. Another extension to PCA, multi-scale PCA (MSPCA) is a natural progression in that it combines both waveform feature extraction and PCA variable reduction. Much previous research has emulated this technique but not as an embedded methodology. However, its embedded principle is also a weakness in that specific techniques cannot be tailored to application requirements. A bespoke coupling was used in [11] to diagnose reciprocating compressor (RC) valve faults, joining Teager-Kaiser and deep belief networks. The former is to estimate the envelope amplitudes, which are de-noised by wavelet transform methods and the latter is to establish the fault classifier. Likewise wave matching feature extraction methods along with support vector machines (SVM) were employed in [12] for classifying vibration signals in fault detection. MSPCA with prior data compression and signal simplification is shown to be effective in reducing input parameter volume^[13] with much improved classification rates for identifying RC component faults.

Alternative means of extracting features are delivered by evolutionary genetic algorithms (GA) which are used for clustering or partitioning groups or for dividing items into distinct groups. Particle swarm optimisation (PSO) is another population-based search method inspired by observation of the collaborative behaviour of biological populations such as birds or bees. Specifically, these populations are seen to demonstrate a collective intelligence^[14–17]. Many current studies incorporate GA feature selection^[14–18] a major advantage of the method being that prior knowledge of the process is not required to establish the best fit. The main criticisms of the method are tendency to over fit and limitations of input parameters. Furthermore, whilst there are several popular algorithms for aiding input parameter choices^[14, 15, 18–21], they themselves have limited capabilities with respect to the number of input parameters permitted to enable convergence. Typically input of around a dozen variables at the most is computable. In addition, original physical characteristics of the input variables are not maintained during analysis. Fault prevalence is thus not directly attributable to specific input parameters.

Classifiers are multivariate models which allocate cases to predetermined classes by incorporating measurements on a number of explanatory variables. For algorithmic convergence, it is generally necessary to restrict the number of input parameters to as few as 10 to 15. To optimise the explanatory power of a model input parameters require rigorous scrutiny. Selection of a few highly informative variables offers increased modelling capabilities hence greater classification power. The key to success for being a reliable measure is to assess variable information and so facilitate the subsequent selection of an input parameter set with maximum explanatory power.

Once characteristics of operational behaviour are iden-

tified, future data readings can be measured and scrutinised for deviations from the norm, hence fault blueprints are established and inform diagnostic practice.

In this paper, techniques will be developed which allow the volume of data to be reduced prior to model construction. Rather than consider means to further expand algorithmic and computational capabilities, the focus is on prior analysis of the input variables. Input parameters are thus assessed in terms of their uniqueness and their ability to explain system variability. Available information is thus rated by explanatory relevance before it is incorporated in analytical modelling processes.

A novel solution is offered with mutual benefit as input parameter groups of reduced number yet increased coverage are identified, simultaneously maintaining classification accuracy and avoiding bias.

Input parameters considered are the envelope harmonic features, a considerable data reduction in themselves requiring transfer of approximately 10% only of the full frequency spectrum^[18]. Further reductions in input parameter volume are considered by prior selection of a small number of heterogeneous parameters identified through variable cluster analysis (CA). Ultimately, selected parameter sets are incorporated in statistical models and classification successes compared. Section 2 explains the theoretical background applied.

2 Theoretical background

2.1 Envelope harmonic spectra

Envelope analysis is a highly effective method of extracting fault signatures and is widely used in condition monitoring of planetary gears, bearings and mechanical systems^[18, 21–24]. Data is initially processed by passing through a band-pass filter to remove extraneous noise, mainly due to set up imbalances, which mask crucial signal trends. The complex envelope can be calculated by application of a suitable transform from which the envelope spectrum can be captured. Spectra are generally normalised to remove bias and a window is applied to stem spectral leakage.

Applying the Fourier transform (FT) to a measured signal, $x(t)$, the repetitive pattern hidden in the data emerges with the Fourier coefficients which allows key features of signals to be easily recognized. The spectrum is the FT coefficients at corresponding frequencies. To identify frequency spectra, the fast Fourier transform (FFT) is the most efficient method of calculation. The FFT, $X(f)$, a function of the frequency, f , a continuous function of time, t , is given by

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi j f t} dt. \quad (1)$$

For digital signals, the discrete Fourier transform (DFT) gives a numerical approximation and is widely

used in engineering.

$$X(t) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-j\left(\frac{2\pi kt}{N}\right)} \quad (2)$$

where $t = 0, 1, 2, \dots, N-1$, N is the number of samples taken, $X(t)$ the value of the signal at time t and k the current frequency, 0 to $(N-1)$ Hz.

Finding the frequency spectrum provides valuable information about the underlying frequency characteristics of signal outputs and so informs definition of system characteristics. Hilbert transforms or wavelet transforms can be applied with similar effect^[22, 25, 26].

Prior research^[18, 22, 25, 26] has shown features extracted from envelope spectra in the frequency domain have superior deterministic properties over their time domain equivalents in CM. Envelope spectra show only the amplitude profile of original signals hence providing clearer insight in to underlying behaviour. Signal variations due to noise are filtered out leaving machine health fluctuations alone.

The envelope harmonic spectra form the group of potential input variables for the model. A means of prioritising variable quality and so selecting a reduced number of input parameters is therefore required.

2.2 Cluster analysis

Cluster analysis (CA) creates data or variable groupings such that objects in the same cluster are similar and objects in different clusters are distinct. This will therefore facilitate the necessary reduction in the number of variables. However, different measures of similarity can present differing results so a rigorous methodology must be applied to the similarity analysis. Assessment of variable likeness can be judged either on an individual to individual basis or by comparison of individuals to a group statistic. Similarity measures are application dependent and include Euclidean distance and Mahalanobis distance^[27–30]. Euclidean distance is the Pythagorean metric, the straight line distance from one point to another in Euclidean space.

Euclidean distance, $d(p, q)$ between the two points p and q is given in (3).

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}. \quad (3)$$

Mahalanobis distance, $D_M(\underline{x})$, measures the proximity of a point p to a distribution or cluster mean. A generalisation in multivariate space of the normal distribution measuring the number of standard deviations a point p is from the cluster mean. Thus, as further points join the cluster, the mean is recalculated. A point p at the cluster mean has a distance of zero. Distances increase for points placed along each principal axis. As Mahalanobis distance is a function of the data correlations, it is scale

invariant whereas Euclidean distance is not. However, scaling axes to unit variance would equate the two measures.

The Mahalanobis distance, $D_M(\underline{x})$, of an observation $\underline{x} = (x_1, x_2, \dots, x_n)^T$ from a set of observations with mean $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ and covariance matrix S is defined in (4).

$$D_M(x) = \sqrt{(\underline{x} - \underline{\mu})^T S^{-1} (\underline{x} - \underline{\mu})}. \quad (4)$$

Whilst there are many different CA algorithms: single linkage, complete linkage, average linkage, to name a few; there are two main types. Agglomeration techniques whereby all objects originate as individuals and are systematically joined until all belong to a single common group and the reverse process, division. Agglomeration commences by joining the two most like objects, the “nearest neighbours” whereas the division algorithm would first select the least like, the “farthest neighbours” for separation. Distances in both cases calculated according to the linkage method employed.

Pairwise Euclidean difference, d_{ij} , between the i -th and j -th observations is given in (5), a square matrix is generated of order m with each entry (i, j) being the Euclidean distance between the observations i and j .

$$d_{ij}^2 = (x_i - x_j)(x_i - x_j)'. \quad (5)$$

Average linkage, $d_{(r, s)}$, is calculated from the average distance between all pairs of objects in any two clusters.

$$d_{(r, s)} = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist(x_{ri}, x_{sj}) \quad (6)$$

where n_r is the number of objects in cluster r , x_{ri} the i -th object in cluster r and x_{sj} the j -th object in cluster s .

Both agglomeration and division are hierarchic methods so directly facilitate representation in pictorial dendrogram format. Inspection of the dendrogram offers a direct visual impression of group proximity and object nearness. The resulting tree is not a single definitive set of clusters but rather a multilevel hierarchy from which a sensible degree of separation or number of clusters can be identified.

Variable clustering offers a means of organising and sifting variables into heterogeneous groups. variables within a cluster group display similar characteristics whereas variables from different cluster groups are not alike. It is thus possible to remove duplicated variables from consideration. A reduced number of distinct variables are evident and fewer representative variables replicate the cluster group behaviour. Consequently, a reduced number of heterogeneous input parameters are established by selection across all groups. Group representation is verified by inspection of separation capabilities on the envelope spectrum. Thus input parameters which pos-

sess optimal explanatory power are extracted with their original characteristics preserved. In addition, direct links are evident between operating condition and specific input harmonics. As input volume is also considerably reduced, algorithmic convergence is faster with computational savings.

2.3 Naïve Bayes (NB)

To explore the efficiency of variable selections, classifiers were constructed using varying numbers of input parameters to discriminate between differing numbers of known classes. Due to its relative simplicity and ability to incorporate higher numbers of classes, Naïve Bayes classification was considered with varying numbers and sets of input parameters. In addition, this technique is generally reliable should normality assumptions be violated.

Naïve Bayes is a relatively straightforward technique for constructing classifiers. Although based on Bayes conditional probability, it is not strictly speaking a Bayesian statistical method. Features within a class are assumed to be independent although good results are often achieved when this assumption is violated. Data is partitioned into training samples and prediction samples and a model established based on the known training set classes. Posterior probabilities for each sample dictate group classifications^[31, 32].

Classification is based on estimating the conditional probability $p(C_k|x_1, \dots, x_n)$ for n independent variables or features $\underline{x} = x_1, \dots, x_n$.

$$p(C_k|\underline{x}) = \frac{p(C_k)p(\underline{x}|C_k)}{p(\underline{x})}. \quad (7)$$

Since the evidence, $z = p(\underline{x})$, is not dependent on class and is effectively constant under Naïve conditional independence assumptions, the probability model becomes

$$p(C_k|x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i|C_k) \quad (8)$$

where the evidence, $z = p(\underline{x})$, is a constant scaling factor dependent only on $\underline{x} = x_1, \dots, x_n$.

NB lends itself to increased numbers of groups and input parameters. Although a classification tree becomes overly complex, data is readily presented in matrix or graphical formats. Constructing a confusion matrix (class number by class number) of detailed case allocations records classification patterns. A 3D bar chart of the confusion matrix data offers a clear visual display^[31-34].

2.4 Principal component analysis

As an alternative to cluster analysis and variable selection, models can be constructed without pre-selection of variables. Generally referred to as data reduction tech-

niques, all 32 variables are incorporated initially and a new set of fused parameters is generated. The newly fabricated variable set is usually ordered automatically by explanatory power. One such technique is principal component analysis.

Principal component analysis (PCA) is a statistical procedure that generally uses an orthogonal transformation to convert a set of highly correlated variables into a set of linearly uncorrelated variables called principal components (PCs)^[27-30].

The method is designed to reduce the number of correlated independent variables, X , to a much smaller number of uncorrelated PCs, Z , which are weighted combinations of the original variables. Each case can then be described by a reduced number of PCs, which will ideally account for most of the variance. Higher correlations between the original variables give greater benefit from this method.

For an n by p matrix X consisting of n observations for each of p variables, a set of p -dimensional weights or loadings vectors

$$w(p) = (w_1, \dots, w_p)_p \quad (9)$$

map each row vector, x_i , of X , a new vector of principal component scores

$$Z_{(i)} = (Z_1, \dots, Z_p)_{(i)} \quad (10)$$

is given by

$$Z_{k(i)} = X_{(i)} \cdot W_{(k)}. \quad (11)$$

The full principal component decomposition of X is given by $Z=XW$, where W is the p by p matrix whose columns are the eigenvectors of $X^T X$.

No data assumptions are required hence its attraction for use with non-interval data or data of unknown distribution^[27-30].

Initially a set of uncorrelated PCs is produced from the original correlated variables. The first PC accounts for the largest proportion of the variance in the sample, the second PC accounts for the second highest proportion of variance, etc. Note the PCs are uncorrelated. Initially as many PCs as original variables are generated together accounting for the total variance in the sample. However, the vast majority of the total variance can be assigned to the first few PCs alone with only a negligible amount ascribed to the remainder. Hence, these latter PCs can be dropped from further analysis, reducing the “dimensionality” of the data set. PCA is mostly used as a tool in exploratory data analysis prior to construction of predictive models. In practice, PCA is executed either by eigenvalue decomposition of a data covariance or correlation matrix or by singular value decomposition of a data matrix. The latter usually means centering normalised Z -scores of the data matrix for each attribute. PCA results

are usually discussed in terms of their factor scores and loadings. Factor, or component scores, are the transformed variable coefficients corresponding to particular data points. The factor loadings are the weights by which each standardised original variable is multiplied to achieve the component score.

$$\text{var}(Z_i) = \lambda_i \quad (12)$$

$$\sum_{i=1}^p \text{var}(Z_i) = \text{tra}(C) \quad (13)$$

where $\text{tra}(C)$ is the sum of diagonal elements of matrix C , the covariance matrix, with the corresponding eigenvector (Z_i) , for each eigenvalue λ_i , is given in (14).

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p. \quad (14)$$

Operation of PCA can be thought of as revealing the internal structure of the data to best explain the variance in the system.

2.5 Confirmatory factor analysis

To gain further quantitative measures of variable attributes a confirmatory factor analysis can be conducted.

This is similar in purpose to PCA, however, distinct in that factor analysis (FA) is founded on a true mathematical model being based on the row ratios of the correlation matrix of a set of original variables. Discounting elements in the leading diagonal, the self-correlations, correlation matrices have the property that elements in any two rows are almost exactly proportional. Spearman first proposed the model over a hundred years ago (1904) on analysing standardised preparatory school exam scores and finding the common ratio for each of the subjects, e.g., classics and French, French and music, etc. to be approximately equal to 1.2, hence proposing the model used today^[30].

$$X_i = a_i F + e_i \quad (15)$$

where X_i is the i -th standardised score, mean zero and standard deviation one, a_i is the factor loading which is a constant, F is the factor value and e_i is the portion of X_i specific to the i -th test only.

Thus, there is a constant ratio between the rows of the variable correlation matrix and this is a plausible model for the data. It also follows that the variance of X_i is given by

$$\text{var}(X_i) = \text{var}(a_i F + e_i) = a_i^2 + \text{var}(e_i). \quad (16)$$

Further, since the variables are standardised,

$$\text{var}(X_i) = 1 = a_i^2 + \text{var}(e_i). \quad (17)$$

Thus, the square of the factor loading equates to the proportion of the variance of X_i that is accounted for by the factor. The sum of the squared factor loadings, $\sum a_i^2$ is the communality of X_i and describes the part of variance related to the common factors. The remaining variance, which is not accounted for by the common factors, is given by $\text{var}(e_i)$, the specificity of X_i . Although there are no specific or widely accepted guidelines, it is a generally accepted rule that loadings between ± 0.3 to ± 1.0 represent salient loadings with the interpretation that the original variable is meaningfully related to that particular factor. Should the factor loadings be difficult to generalise being neither close to zero or ± 1.0 , a rotation of the solution could be considered. It should be noted that factor rotation is a mathematical aid to interpretation rather than a refitting of the model, hence will not affect the overall goodness of fit of the model, simply the arbitrary axes along which the factors are measured^[30].

Whilst FA has its limitations^[27], it is of particular benefit in gaining insight into the nature of underlying variables in multivariate data. The worth of FA is as a descriptive tool to uncover or describe underlying data structures albeit with consideration of methodological limitations. Thus, whilst FA is largely, an exploratory technique, substantive and practical considerations should strongly guide the analytical process.

2.6 Data compression

Further volume reductions are possible by trimming signals using data compression techniques. Typically, such compressions can be done by a wavelet method whereby a signal is decomposed at a specified number of levels. Coarse approximations are made using details from each level and simplified signals then reconstructed. Simplified signals are smoother with local fluctuations removed. Although a crude de-noising process, if salient information is retained, the process offers further reductions in input parameter volume.

Multiscale PCA at 5 levels was conducted using Matlab sym4 with PCs retained according to Kaisers rule, $\lambda_i > \bar{\lambda}$, where λ_i is the i -th eigenvalue. Further signal simplification was achieved by retaining only the final 4 PCs of the 7 components generated. The impact of the compression was not uniform distribution-wise. Harmonic feature 4 became more skewed on compression whereas the reverse was seen for feature 6.

To effectively demonstrate the aforementioned principles, techniques were applied to experimental data taken from a reciprocating compressor rig.

3 Data acquisition

Due to their prevalence and importance in industrial processes, there is naturally much interest in the detection and diagnosis of reciprocating compressor (RC) faults^[35–37]. Of the two major faults groups, due to fail-

ure of mechanical moving parts and those due to loss of elasticity in sealing components resulting in leaking gas, the latter is most prevalent and forms the focus of this illustration^[36].

A broom wade TS9 RC rig was utilised for the experimental investigations, as shown in Fig.1. The TS9 is a two-stage RC with compression cylinders arranged in V-shape formation. The rig has a maximum working pressure of 1.379MPa (13.8bar) and a crank speed of 440rpm. Further specification details are stipulated in Table 1. Operational characteristics of mechanically driven machinery, such as the RC, are effectively illuminated by analysis of vibration signals^[38, 39].

Fig. 1 Broom wade TS9 reciprocating compressor rig

Table 1 Two stage broom wade RC specification

Type	TS9
Maximum working pressure	1.379MPa
Number of cylinders	2 (90 degrees opposed in V shape)
First stage piston diameter	93.6mm
Second stage piston diameter	55.6mm
Piston stroke	76mm
Crank speed	440rpm
Motor power	2.2kW
Supply voltage	380/420V
Motor speed	1420rpm
Current	4.1/4.8A

Vibration signals were collected via transducers from the second stage cylinder head. Type YD-5-2 accelerometers with a frequency range between 0 and 15kHz, sensitivity 45mVms⁻², temperature tolerance up to 150°C and acceleration maximums of 2000ms⁻². Measurements of interest were thus comfortably within the range of sensor capabilities whilst sensors were hardy enough to withstand any extreme shock and temperature loadings they might encounter.

Second stage cylinder head pressure signals were collected simultaneously via GEMs type 2200 dynamic strain gauge pressure transducers inserted at the cylinder

head. These sensors also gave adequate range coverage having an output of 100mV when used with a 10Vd.c. power supply, a range up to 4MPa (600psi) and upper frequency limit of 4kHz. Pressure sensors were similarly well specified for purpose. Since no amplification is required, the sensors can be connected directly to the CED and PC. National instruments laboratory Windows software Tm/CV1/Version 5.5, written in the C language, enabled data storage and conversion into Matlab bin files for export and analysis.

The RC rig was operated under healthy conditions and with four independently seeded faults (suction valve leakage (SVL), discharge valve leakage (DVL), intercooler leakage (ICL) and a loose drive belt (LB)). Each experimental run was repeated 24 times. Thus, a total of 120 observations were recorded at each of six pressure loads^[40–42].

Vibration capture devices are ideal for gleaning highly informative non-intrusive data measurements and making continuous process monitoring feasible. Although more problematic to secure, the pressure measurements provide a useful comparative measure in the ensuing discussion.

Signal amplitudes of the FFT are widely recognized as a means of efficient pattern recognition. The basic concept of FFT analysis is to reduce a complex signal in the time domain to its component parts in the frequency domain. Salient features of the signal thus become apparent as confusion due to noise is removed^[36–42]. A combined approach using time-frequency analysis of vibration signals in conjunction with image-based pattern recognition techniques also realised high classification success rates^[43], the vibration signals with prior feature reduction using PCA and a Bayesian classification approach were successful in assessment of discharge valve deterioration with 90% accuracy.

An 87835 point filter revealed signal trends in the measured signal from which a Hilbert transform was applied to calculate the signal envelope. Hanning windows were subsequently employed to reduce spectral leakages in computing the envelope spectrum. For each captured signal, the first 32 harmonic features were extracted from the envelope spectrum, thus a classification matrix X (120 observations by 32 variables) was constructed per signal.

Signal amplitude profiles for each signal, illustrated in Fig.2, are stored for each of the machine states considered. Thus, a three dimensional data array (120 cases $\times$ 32 harmonics $\times$ 7 signals) is stored, with the 32 harmonics being the potential input variables.

4 Experimental validation

4.1 Input parameter peduction

As the main focus of the CA was to identify variable similarities, an agglomerative cluster method was em-

Fig. 2 Harmonic spectrum of second stage vibration signal per class (Color versions of the figures in this paper are available online)

ployed. Thus group formations, visibly observed through dendrograms, were readily identifiable. The proximity measure utilised was the Euclidean distance since the data represents harmonic amplitudes.

Dendrograms provide a visual illustration of variable similarities, as illustrated in Figs.3 and 4, which are based on vibration and pressure signals respectively. Note the cluster linkage differences are variable specific hence the lack of parity between scales for the second stage vibration and pressure variables. Also, although measure-

Fig. 3 Clustering of second stage vibration signal envelope harmonics

Table 2 Second stage vibration cluster groups	
	Group members
Group 1	4 selected as representative of large group of like features.
Group 2	(13, 14), (11, 16)
Group 3	(3, 5), 6, 7, 8, 10, 15
Independent features	2, 9 (1 omitted having been shown to have little discriminating power).

Fig. 4 Clustering of second stage pressure signal envelope harmonics

ments were simultaneously captured from the RC rig, harmonics form entirely unique groupings.

For the second stage vibration signal, harmonic features 23, 24, 27 and 28 are most alike whilst the pairing (3, 5) are least like all other pairs, see Fig.3. Strong similarities are easily identified from early linkage. Homogeneous pairings are identifiable from the dendrogram, for example (11, 16) and (13, 14) in addition to (3, 5). It would seem reasonable to suggest three main harmonic groupings and perhaps 5 “semi-independent” harmonic features, shown in Table 2. The latter, are the variables, 6, 7, 8, 10 and 15, which are late additions to the third cluster group ($T \approx 10$).

Cluster formation for the second stage pressure signals follows a very different pattern being especially uniform as shown in Fig.4. Early, close proximity, pairings are common with (18, 19), (1, 10), (5, 11), (9, 24) and (14, 15), the latter two pairs also quickly forming a homogeneous set along with harmonic 12. Pressure envelope harmonic features could be considered as 5 early formed groups ($T < 0.2$) or 2 groups ($T < 0.35$). However, since harmonic 25 is not linked until $T = 0.9$, it should be considered an individual that is considerably different to all the rest, thus remaining unconnected to the other groups.

Subsequent scatter plot analysis in this section focuses on the second stage vibration signals since they are collected non-intrusively.

Visual observation of class separation in Figs.5 and 6, emphasises the importance of input variable choice in statistical modelling. Features 4 and 7 give reasonable grouping in the two class case, whereas features 4 and 6 demonstrate superior classification characteristics displaying a clear tract between the differing class types.

Confirmation of homogeneity within group harmonic features is demonstrated via the scatter plots in Figs.7 and 8. Harmonic feature 6 has been shown to demonstrate superior deterministic properties over feature 7. Both these homogeneous pairs are observed in Fig.7 to almost isolate the ICL class when used in conjunction with feature 4, harmonic 6 producing the greater degree

Fig. 5 Scatter plot of healthy and ICL groups using harmonic features 4 and 7

Fig. 6 Scatter plot of healthy and ICL groups using harmonic features 4 and 6

Fig. 7 Scatter plots illustrating similarity for homogeneous features 6 and 7 in comparison to heterogeneous feature 12

Fig. 8 Homogeneity of within group characteristics demonstrated using harmonics 3 and 5

of separation. Feature 12, on the other hand, is incapable of that distinction.

Fig. 8 illustrates the near identical properties of harmonic features 3 and 5, a most homogeneous pair of envelope harmonics taken from the second stage vibration data. Evidence of input variable duplicity, hence of the power of CA to inform input variable choice in reducing input volume, is clearly visible.

4.2 Application to classifier construction

A Naïve Bayes model for the two class case (healthy and ICL) required just 5 input parameters for 100% success rate in classification. However, as the number of classes considered increases so must the model complexity and generally so will the number of input parameters necessary for a high degree of classification accuracy.

Directly constructing a Naïve Bayes classifier utilising all 32 envelope harmonic features from the second stage vibration signal spectrum achieved 75% classification success rate across all 5 classes, as shown in Fig. 9. However, this was much improved by restricting input parameters for data reduction to those identified via variable clustering. That fewer input parameters are more able to accurately describe system variation is a surprising yet key finding of this research. Subsequent classification using a 10 parameter model, as shown in Fig. 10, achieved 82% success rate. Cross classification matrices are given in Table 3, note perfect 100% classification is indicated by $24I$ with " I " being the 5×5 identity matrix.

4.3 Variable reduction

Principal component analysis (PCA) is a variable reduction technique. The focus of the analysis is to seek underlying principal components (PCs) to define the vari-

Fig. 9 Naïve Bayes classifier using all 32 input parameters

Fig. 10 Naïve Bayes classifier using 10 input parameters

Table 3 Cross classification matrices

	10 parameter model					32 parameter model				
Healthy	17	4	0	3	0	16	4	0	4	0
DVL	0	24	0	0	0	1	23	0	0	0
ICL	0	3	19	0	2	0	2	20	1	1
LB	5	1	0	17	1	6	6	0	12	0
SVL	1	2	0	0	21	1	4	0	0	19

ation in the system. The PCs could then be used as input variable in the construction of classifiers to identify machine health. Ideally a smaller number of PCs will account for the vast majority of the total variance. Hence, a reduced number of highly representative new variables are established each of which incorporates elements of all the original variables.

Since only the first three PCs had eigenvalues greater than one, hence contribute substantially towards the total variation in the system, a two and three PC model were investigated in this analysis. Any PC with an eigenvalue greater than one is considered to contribute “more than

its share” towards explaining the variance in the system. With $\lambda_1 = 12.3416$, the majority of the variance, almost 60%, was incorporated in PC1 with an additional 14% from the second PC and approximately 8% from the third, results are summarised in Table 4.

Thus, the first two PCs accounted for approximately 73% of the total variation in measurements. Clearly, when all the cases are plotted against these first two PCs, the SVL group, as shown in Fig. 11, is seen to be entirely separate from all other classes having the lowest scores on both the 1st and 2nd principal components.

Identifying the SVL fault is thus particularly straightforward, the first two PCs form a sufficiently sophisticated model for successful classification. Even using just two PCs, all other cases are reasonably well grouped by class. With the addition of a third PC, which increases the cumulative sum of the model to 81%, classification rates improve further still. Class grouping is clearly evident as displayed in Fig. 12.

The fourth PC has a variance very close to one and could reasonably be incorporated to further improve model accuracy. However, the remaining PCs all have variances less than one and so offer increasingly negligible contributions in deterministic terms thus further classification improvements are not realistic using PCA. The cumulative sums for the first 14 principal components are reported in Table 5. Whilst 81% of the total variation in the system is accounted for by the first three PCs alone, the first ten PCs are required to achieve 95%.

Table 4 Eigenvalues for the first three PCs

Variance (PC1)	$\lambda_1 = 12.3416$
Variance (PC2)	$\lambda_2 = 2.9228$
Variance (PC3)	$\lambda_3 = 1.7662$
Total variance (PC1+PC2+PC3)	$\lambda_1 + \lambda_2 + \lambda_3 = 17.0321$
Total variance in system	$\sum \lambda_i = 21.201$
Cumulative sum of variance	$17.0321/21.201 = 0.8103$

Fig. 11 Class clusters using the first two PCs

Fig. 12 Fault clusters using the first three principal components

Table 5 Cumulative sums for the first 14 principal components

(1)	(2)	(3)	(4)	(5)	(6)	(7)
0.5872	0.7262	0.8103	0.8556	0.8824	0.9037	0.9198
(8)	(9)	(10)	(11)	(12)	(13)	(14)
0.9328	0.9449	0.9545	0.9616	0.9673	0.9713	0.9750

Table 6 NB cross classification matrix using compressed harmonics 4 and 6

H	16	3	0	5	0
DVL	0	24	0	0	0
ICL	0	1	22	0	1
LB	3	2	0	17	2
SVL	1	1	0	1	21

4.4 Confirmatory factor analysis

Since there appear to be underlying generic health conditions governed by collective groups of harmonic features, a confirmatory factor analysis was conducted.

Inspection of the factor loadings, Table 7, on the first two factors shows high factor 1 loadings for harmonic features 6 and 7 thus these two harmonics are highly correlated as might be expected. Also, both have negligible factor 2 loadings. On the other hand, harmonic feature 4 has a high factor 2 loading with much lower factor 1 loading so is less correlated with features 6 and 7 but more highly correlated with features 3 and 5. These find-

ings confirm that inclusion of both harmonics 6 and 7 is unnecessary for modelling purposes as they “explain” similar variability, whereas, harmonic 4 makes a distinct contribution.

The specific variance of harmonic 6 is 0.0632 which is close to zero so implies the variable is almost entirely determined by its common factors, in fact 93% of the variance of harmonic 6 is accounted for by factor one hence its superior deterministic power. 70% of the variance of harmonic 7 is explained by factor one. In comparison, harmonic 9 has a specific variance of 0.9710 almost 100% which implies there is practically no common factor component in the variable. Indeed harmonic 9 possesses just 2% common variance in factor one and considerably less in factor two. Consequently, harmonic 9 is deemed unlikely to add useful detection power if included.

4.5 Signal compression

The effect of compression for key harmonics is illustrated in Fig. 13.

Outstandingly a NB classification model for all classes, using compressed 4th and 6th harmonics alone, realised an 83.3% classification success which far exceeds prior rates. Previously comparable classifiers required 10 plus input parameters. The classification matrix, Table 6, shows the specific numerical details of group allocations. The healthy group, row 1, has the greater number of cases allocated to other groups with 3 to the DVL and 5 to the LB group. All DVL cases were correctly allocated. False positive allocations of healthy to fault states are inconvenient although less critical than false negative classifications.

False negative errors, or type I errors, give a measure of the significance level, α . Sensitivity of a test is defined as $(1-\alpha)$, i.e., the proportion of correctly identified cases. Specificity or power of a test is the proportion of healthy cases correctly identified as healthy. Sensitivity and specificity are complementary measures intrinsic to the test not dependent on fault prevalence. A balance between the two measures is sought to maximise information gain. Affording equal weight to the true positive and false positive rates optimises test information and leads to a convenient measure of worth, the information gained. Table 8 displays the classification frequencies from which the information gain is calculated to be 0.484, by defining an informed decision. Note equivalence to zero implies

Table 7 Factor loadings and specific variance for key harmonics

Factor loadings	Envelope harmonic									
	2	3	4	5	6	7	9	12	13	14
Factor 1	0.6679	0.1804	0.3615	0.1735	0.9659	0.8343	0.1459	0.2905	-0.1865	-0.1239
Factor 2	0.6448	0.7963	0.7810	0.8083	-0.0622	0.3980	-0.0877	0.7282	0.8267	0.7816
Specific variance	0.1381	0.3333	0.2593	0.3166	0.0632	0.1455	0.9710	0.3854	0.2817	0.3738

Fig. 13 Comparison of original and compressed signal for key harmonics

Table 8 Classification frequencies, sensitivity and specificity calculations techniques and methodology

		Predicted condition	
		Healthy	Faulty
True condition	Healthy	$TP = 16$	$FP = 8$ Type 11 Error (β)
	Faulty	$FN = 12$ Type 1 Error (α)	$TN = 84$

Sensitivity (power), $1 - \beta = \frac{16}{28} = 0.571$; Specificity, $1 - \alpha = \frac{84}{92} = 0.913$.

chance-level performance and less than zero indicates per-verse use of information.

Information gain = specificity + sensitivity - 1 =
 $0.913 + 0.571 - 1 = 0.484 > 0$.

5 Conclusions

Clearly envelope harmonic features have individual properties and differing attributes in terms of fault identification potential.

Envelope harmonic feature groupings are signal specific in terms of both group membership and uniformity of group formation. Clustering threshold distances are variable dependent hence require standardisation to enable

direct comparisons between signals.

Variable clustering facilitates identification of homogeneous groupings of envelope harmonics. Within group variables have been demonstrated to possess similar powers in discerning specific fault characteristics, whereas between group variables were shown to be heterogeneous. Thus variable clustering offers a means of selecting a complete and complimentary set of model input parameters avoiding duplicity hence reducing input parameter volume.

Computational burdens are relieved and faster algorithmic convergence feasible. In addition, refining the potential algorithmic input beforehand proffers demonstrably improved classification accuracy.

Application to classifier construction methodologies, illustrated with Naïve Bayes, further consolidates the findings. Models with greater accuracy are achievable from selective variable input sets.

Confirmatory FA supplied further quantitative evidence in support of individual harmonic feature supremacy. Harmonic 6 was demonstrated to be quite literally in a league of its own.

Preliminary investigations into data compression offers greatly improved success rates and further investigation is recommended.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- [1] R. Barron. *Engineering Condition Monitoring: Practice, Methods and Applications*, Essex, UK: Addison Wesley Longman, 1996.
- [2] A. Bhattacharya, P. K. Dan. Recent trend in condition monitoring for equipment fault diagnosis. *International Journal of System Assurance Engineering and Management*, vol. 5, no. 3, pp. 230–244, 2014. DOI: 10.1007/s13198-013-0151-z.
- [3] E. P. Carden, P. Fanning. Vibration based condition monitoring: A review. *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, 2004. DOI: 10.1177/1475921704047500.
- [4] A. Davies. *Handbook of Condition Monitoring: Techniques and Methodology*, London, UK: Chapman and Hall, 1998.
- [5] B. Flury, H. Riedwyl. *Multivariate Statistics: A Practical Approach*, London, UK: Chapman and Hall, 1988.
- [6] L. H. Chiang, E. L. Russell, R. D. Braatz. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243–252, 2000. DOI: 10.1016/S0169-7439(99)00061-1.
- [7] Y. Zhang, C. M. Bingham, M. Gallimore. Fault detection and diagnosis based on extensions of PCA. *Advances in Military Technology*, vol. 8, no. 2, pp. 27–41, 2013.
- [8] B. R. Bakshi. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, 1998. DOI: 10.1002/aic.690440712.
- [9] R. P. Shao, W. T. Hu, Y. Y. Wang, X. K. Qi. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement*, vol. 54, pp. 118–132, 2014. DOI: 10.1016/j.measurement.2014.04.016.
- [10] Y. W. Zhang, S. Li, Z. Y. Hu. Improved multi-scale kernel principal component analysis and its application for fault detection. *Chemical Engineering Research and Design*, vol. 90, no. 9, pp. 1271–1280, 2012. DOI: 10.1016/j.cherd.2011.11.015.
- [11] V. T. Tran, F. Al Thobiani, A. Ball. An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014. DOI: 10.1016/j.eswa.2013.12.026.
- [12] Q. Qin, Z. N. Jiang, K. N. Feng, W. He. A novel scheme for fault detection of reciprocating compressor valves based on basis pursuit, wave matching and support vector machine. *Measurement*, vol. 45, no. 5, pp. 897–908, 2012. DOI: 10.1016/j.measurement.2012.02.005.
- [13] A. Smith, F. S. Gu, A. Ball. Maintaining model efficiency, avoiding bias and reducing input parameter volume in compressor fault classification. In *Proceedings of the 7th International Conference on Mechanical and Aerospace Engineering*, London, UK, pp. 196–201, 2016. DOI: 10.1109/ICMAE.2016.7549534.
- [14] R. Hassan, B. Cohanin, O. De Weck, G. Vente. A comparison of particle swarm optimization and the genetic algorithm. In *Proceedings of the 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Structures, Structural Dynamics, and Materials and Co-located Conferences*, Austin, USA, pp. 18–21, 2005. DOI: 10.2514/6.2005-1897.
- [15] Y. H. Lin, W. S. Lee, C. Y. Wu. Automated fault classification of reciprocating compressors from vibration data: a case study on optimization using genetic algorithm. *Procedia Engineering*, vol. 79, pp. 355–361, 2014. DOI: 10.1016/j.proeng.2014.06.355.
- [16] V. R. Rao, V. D. Kalyankar. Parameter optimization of modern machining processes using teaching–learning-based optimization algorithm. *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 524–531, 2013. DOI: 10.1016/j.engappai.2012.06.007.
- [17] B. Samanta, C. Nataraj. Use of particle swarm optimization for machinery fault detection. *Engineering Applications of Artificial Intelligence*, vol. 22, no. 2, pp. 308–316, 2009. DOI: 10.1016/j.engappai.2008.07.006.
- [18] M. Ahmed, A. Smith, F. Gu, A. D. Ball. Fault diagnosis of reciprocating compressors using relevance vector machines with a genetic algorithm based on vibration data. In *Proceedings of the 20th International Conference on Automation and Computing*, Cranfield, UK, 2014. DOI: 10.1109/ICAC.2014.6935480.
- [19] M. E. Tipping. Sparse Bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001. DOI: 10.1162/15324430152748236.
- [20] A. C. Faul, M. E. Tipping. Analysis of sparse Bayesian learning. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, MIT Press, Vancouver, Canada, pp. 383–389, 2001.
- [21] S. X. Ding. *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools*, London, UK: Springer, 2008.
- [22] F. S. Gu, B. Payne, A. D. Ball. Instantaneous Angular Speed Signature Extraction through Hilbert Transform and Fourier Transform, Technical Report: MERG-0402, Manchester University, UK, 2002.

- [23] Y. G. Lei, J. Lin, M. J. Zuo, Z. J. He. Condition monitoring and fault diagnosis of planetary gearboxes: A review. *Measurement*, vol. 48, pp. 292–305, 2014. DOI: 10.1016/j.measurement.2013.11.012.
- [24] G. J. Feng, J. Gu, D. Zhen, M. Aliwan, F. S. Gu, A. D. Ball. Implementation of envelope analysis on a wireless condition monitoring system for bearing fault diagnosis. *International Journal of Automation and Computing*, vol. 12, no. 1, pp. 14–24, 2015. DOI: 10.1007/s11633-014-0862-x.
- [25] M. Elhaj, F. Gu, J. Shi, A. A. Ball. A comparison of the condition monitoring of reciprocating compressor valves using vibration, acoustic, temperature and pressure measurements. In *Proceedings of the 6th Annual Maintenance and Reliability Conference*, Gatlinburg, Tennessee, USA, 2001.
- [26] M. Elhaj, F. Gu, A. Ball. Numerical simulation study of a two stage reciprocating compressor for condition monitoring. In *Proceedings of the 17th International Congress on Condition Monitoring and Diagnostic Engineering Management*, Cambridge, UK, pp. 602–611, 2004.
- [27] C. Chatfield, A. J. Collins. *Introduction to Multivariate Analysis*, London, UK: Chapman and Hall, 1980.
- [28] B. S. Everitt, G. Dunn. *Applied Multivariate Data Analysis*, 2nd ed., London, UK: Arnold, 2001.
- [29] R. A. Johnson, D. W. Wichern. *Applied Multivariate Statistical Analysis*, 5th ed., Upper Saddle River, USA: Prentice Hall, 2002.
- [30] B. F. J. Manly. *Multivariate Statistical Methods: A Primer*, 3rd ed., Chatfield and Collins, London, UK, 1986.
- [31] S. J. Chapman. *Essentials of Matlab Programming*, 2nd ed., Stamford, USA: Cengage Learning, 2009.
- [32] S. J. Chapman. *Matlab Programming with Applications for Engineers*, Stamford, USA: Cengage Learning, 2013.
- [33] L. Breiman, J. H. Friedman, C. J. Stone, R. A. Olshen. *Classification and Regression Trees*, New York, USA: Chapman and Hall, 1993.
- [34] SAS Institute. SAS user's guide: statistics (vol.1 and 2). Cary, USA: SAS Institute, 1985.
- [35] H. P. Bloch, J. J. Hoefner. *Reciprocating Compressors: Operation and Maintenance*, Houston, USA: Gulf Pub. Co., 1996.
- [36] R. Isermann. *Fault-diagnosis Applications: Model-based condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-tolerant Systems*, Berlin Heidelberg: Germany: Springer, 2011. DOI: 10.1007/978-3-642-12767-0.
- [37] R. Isermann. *Fault-diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Berlin Heidelberg, Germany: Springer, 2006. DOI: 10.1007/3-540-30368-5.
- [38] G. deBotton, J. Ben-Ari, E. Sher. Vibration monitoring as a predictive maintenance tool for reciprocating engines. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 214, no. 8, pp. 895–903, 2000. DOI: 10.1177/095440700021400808.
- [39] R. R. Bond. *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*, Wiley, London, UK, 2010.
- [40] D. J. Murray-Smith. *Modelling and Simulation of Integrated Systems in Engineering: Issues of Methodology, Quality, Testing and Application*, Cambridge, UK: Woodhead, 2012.
- [41] J. Osarenren. *Integrated Reliability: Condition Monitoring and Maintenance of Equipment*, Boca Raton, USA: CRC Press, Taylor and Francis Group, 2015.
- [42] J. S. Rao. *Vibratory Condition Monitoring of Machines*, Alpha Science International Ltd, Oxford, UK, 2000.
- [43] J. R. Kolodziej, J. N. Trout. An image-based pattern recognition approach to condition monitoring of reciprocating compressor valves. *Journal of Vibration and Control*, vol. 24, no. 19, pp. 4433–4448, 2018. DOI: 10.1177/1077546317726453.

Ann Smith received the M.Sc. degree in applied statistics from Sheffield Hallam University, UK in 1997, the Ph.D. degree in engineering from University of Huddersfield, UK in 2017. She began lecturing in theoretical and applied mathematics and statistics to a wide range of students across the University in 1992 following her award of PGCE(FE) (postgraduate certificate in education, higher education). She is a Fellow of the HEA (the Higher Education Academy) and IMA (the Institute of Mathematics and its Applications).

Her research interests include condition monitoring, modelling system behaviours, predictive analytics for applied process monitoring, non-linear systems approaches to detection of deviant events, autonomous abnormality assessment and evidence based health care.

E-mail: a.smith@hud.ac.uk (Corresponding author)
ORCID iD: 0000-0002-4154-7868

Fengshou Gu received the B.Sc. and M.Sc. degrees in mechanical engineering from Taiyuan University of Technology, China in 1985, and the Ph.D. degree in mechanical engineering from University of Manchester, UK in 2009. He is a professor in diagnostic engineering, working as the head of Measurement and Data Analytics Research Group (MDARG) and the deputy director of the Centre for Efficiency and Performance Engineering (CEPE), University of Huddersfield, UK.

His research interests include machine dynamics, tribology dynamics, advanced signal processing, measurement system and sensor development, artificial intelligence and related fields.

E-mail: f.gu@hud.ac.uk
ORCID iD: 0000-0003-4907-525X

Andrew D. Ball received the B.Sc. degree in mechanical engineering from University of Leeds, UK in 1987, his degree having been sponsored by BICC Electronic Cables. He went on to work for Ruston Gas Turbines and then gained a sponsorship from WM Engineering and the Royal Navy, enabling him to join the Total Technology Scheme at University of

Manchester, UK, from which he graduated in 1991 with a Ph.D. in Machinery Condition Monitoring. He took the shell sponsored lectureship in maintenance engineering at University of Manchester, UK in 1991 and was promoted to professor of main-

tenance engineering in 1999. He was the head of School of the Manchester School of Engineering from 2003 to 2004, and he became dean of Graduate Education in 2005. In late 2007, he moved to University of Huddersfield, UK as professor of diagnostic engineering and pro-vice-chancellor for research and enterprise. His personal research expertise is in the detection and diagnosis of faults in mechanical, electrical and electro-hydraulic machines, in data analysis and signal processing, and in measurement systems and sensor development. He is the author of over 300 technical and professional publications, and he has spent a large amount of time lecturing and consulting to industry in all parts of the world. Andrew has to date graduated

almost 100 research degrees, in the fields of mechanical, electrical and diagnostic engineering. He has acted as external examiner at over 30 overseas institutions, he holds visiting and honorary positions at 4 overseas universities, he sits on 3 large corporate scientific advisory boards, and he is also a registered expert witness in 3 countries.

His research interests include machinery condition monitoring, vibration analysis, signal processing, fault detection and acoustics.

E-mail: a.ball@hud.ac.uk

ORCID iD: 0000-0001-7540-8965