Journal of Sound and Vibration (2003) **259**(2), 323–343
doi:10.1006/jsvi.2002.5168, available online at http://www.idealibrary.com on IDEAL®

EXPERIMENTAL VALIDATION OF A STRUCTURAL HEALTH MONITORING METHODOLOGY: PART I. NOVELTY DETECTION ON A LABORATORY STRUCTURE

K. Worden and G. Manson

*Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S*1 3*JD, England. E-mail: k.worden@sheffield.ac.uk*

AND

D. Allman

*Aero-Structures Department, Mechanical Sciences Sector, Glauert Building (A*9*), DERA*†*, Farnborough, Hampshire GU*14 0*LX, England*

(*Received* 9 *July* 2001, *and in final form* 5 *February* 2002)

This paper is concerned with the experimental validation of a structural health monitoring methodology, previously only investigated using synthetic data. The structure considered here is a simplified model of a metallic aircraft wingbox i.e., a plate incorporating stiffening elements. Damage is simulated by a saw-cut to one of the panel stringers (stiffeners). The analysis approach uses novelty detection based on measured transmissibilities from the structure. Three different novelty detection algorithms are considered here: outlier analysis, density estimation and an auto-associative neural network technique. All three methods are shown to be successful to an extent, although a critical comparison indicates reservations about the density estimation approach when used on sparse data sets.

1. INTRODUCTION

For many years, vibration analysis has been used for structural health monitoring and machine condition monitoring, although it is arguably in the latter field only that the approach has made the jump from academia to industry. Most vibration-based fault detection methods are based on empirical procedures which establish some correlation between the structure or machine behaviour and the vibration level. The object is to identify if and when the system departs from normal condition. At a slightly more sophisticated level, the problem of fault detection can be regarded as a hierarchy of levels [1].

Level 1 (*Detection*): The method gives a qualitative indication that damage might be present in the structure.

Level 2 (*Localization*): The method gives information about the probable position of the damage.

† Now QinetiQ.

0022-460X/02/$35.00

Level 3 (*Assessment*): The method gives an estimate of the extent of the damage.

Level 4 (*Prediction*): The method offers information about the safety of the structure e.g., estimates a residual life.

In terms of the detection problem (level 1), recent work has established a framework based on the so-called *novelty* or *anomaly detection*, which unifies approaches from a number of different disciplines [2–4]. The philosophy of the approach is simply to establish a description of normality using features representing the undamaged condition of the machine or structure and then test for abnormality or novelty when new data becomes available. This is clearly in line with the philosophy of classical machine condition monitoring.

There are numerous different methods of novelty detection. These include methods based on simple distance measures [5], probability density estimation [2, 3], artificial neural networks [4, 6, 7] and wavelet analysis [8]. Previous work by the authors has concentrated on the neural network and statistical procedures [4, 5, 9].

There is less success reported in the literature on the construction of higher level diagnostics, particularly in the experimental context. One notable exception to this is the *modal strain energy* approach of Stubbs and co-workers which has been validated on large-scale civil structures [10, 11]. A later paper in this sequence of three [12], will address the issue of damage location, but from the viewpoint of a set of novelty detectors.

Despite the recent proliferation of damage detection work, much of it has been of a theoretical or computational nature and experimentally validated studies remain the exception rather than the rule. The object of the current sequence of papers is to validate the novelty detection methods discussed in previous papers [4, 5, 9] on real structures. In this first paper, the structure of interest is a stiffened panel which is designed to simulate an aircraft wingbox. The second and third papers in this series will consider the problems of damage detection and location respectively on an aircraft wing [12, 13]. The features used for the analysis are transmissibility frequency response functions (FRFs) as used in reference [4]. The advantage of these features is that they use only responses of the structure and can therefore give a diagnostic even if the ambient excitation is unknown. Although this is highly speculative, in the context of aerospace structures, such features could potentially be used for in-flight monitoring.

The work for these three papers has been carried out over a period of 3 years. It has been, as all research should be, a learning experience and the papers have been written in such a way as to reflect this. The authors feel that descriptions of wrong turns and blind alleys can be as useful to other investigators as a straightforward presentation of successful work.

The layout of this paper is as follows: Section 2 is concerned with describing the panel, the instrumentation and the measurement strategy. Section 3 describes the three methods of novelty detection which are applied here. Section 4 contains the results of novelty analysis. The paper concludes with some discussion in Section 5.

2. THE EXPERIMENTAL PANEL AND DATA CAPTURE

The first stage of the work was to construct the panel to the following specifications. The upper surface is $750 \times 500 \times 3$ mm aluminium sheet (Figure 1: this is a schematic figure, it is not to scale. It is simply intended to show the construction of the panel). This is stiffened by the addition of two ribs composed of lengths of *C*-channel riveted to the short edges. Two stiffening stringers composed of angle section run along the length of the sheet. This

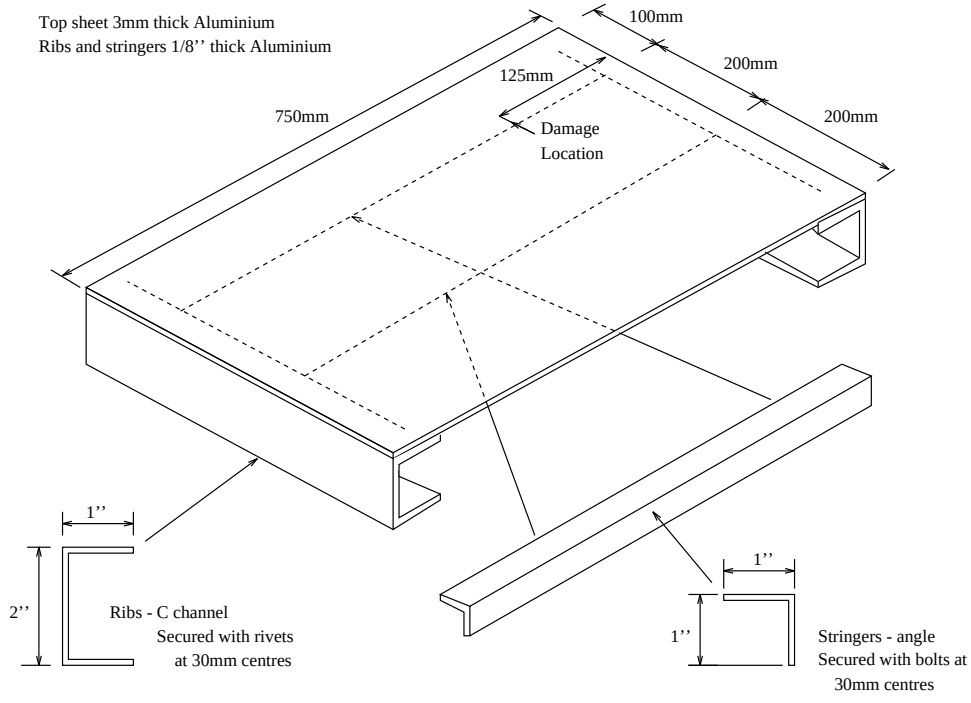

Figure 1. Schematic of simulated skin panel.

design was intended to be compatible with a finite element simulation to be carried out in parallel by DERA. The tests were all conducted with free–free boundary conditions for the panel, which was suspended from a substantial frame using springs and nylon line.

Damage was simulated here by the introduction of a saw-cut in the outside stringer 125 mm from the edge of the panel (Figure 1). Nine levels of damage were investigated from 10% depth to 90%. As the stringer is 1 in in extent, each level corresponds to 2.5 mm of damage.

The analysis approach was novelty detection based on transmissibility FRFs as in reference [4]. At each stage of damage, FRFs were taken from the panel. Three transmissibility paths were identified: AB, AC and DC, as seen in Figure 2. AB is along the line of the damaged stringer, while DC is offset by 100 mm (again note that Figure 2 is a schematic only).

The system was excited using a Gearing and Watson electrodynamic shaker driven by broadband white noise amplified by a Gearing and Watson power amplifier. The responses were measured using PCB resonant piezoelectric accelerometers and sampled using a DIFA/Scadas acquisition system running LMS software under the control of a HP computer. The DIFA system was also used to form the random excitation.

The frequency range over which the transmissibilities were taken was 0–250 Hz and in all cases 2048 spectral lines were taken. In order to have clean data to identify which modes were sensitive to the damage, an averaged transmissibility was taken for each path. One hundred and twenty eight averages were taken in each case. In order to accumulate a reasonable size of normal condition set, 128 transmissibilities (not averaged) were taken for each path. This was in order to validate the damage detection methods which require reasonably sized data sets without resorting to *a priori* assumptions regarding the extent

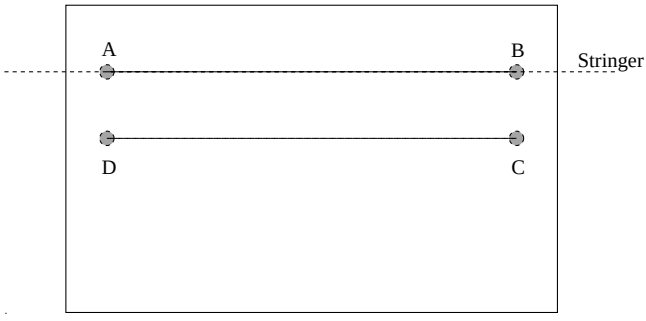

Figure 2. Transmissibility measurement paths.

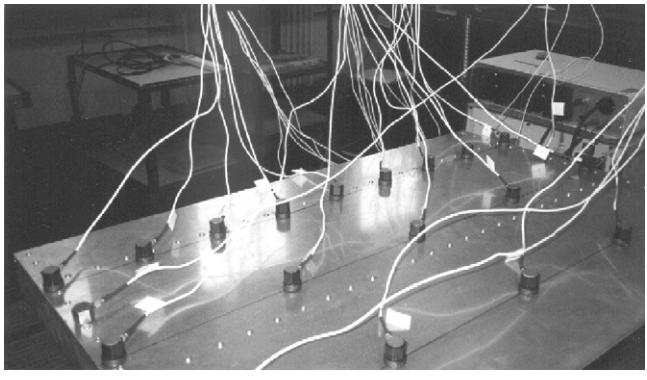

Figure 3. Top of plate showing accelerometer placement along the stringer.

and colour of measurement noise. For each of the damage cases, an average over 128 transmissibilities was taken as well as 10 unaveraged patterns to form the test set. The analysis of this data forms the basis of section 4 of this paper.

Figures 3 and 4 show the test facility. (Note the extra accelerometers, they were used to perform a top surface modal analysis for purposes reported elsewhere [14].)

3. METHODS OF NOVELTY DETECTION

The following sections will use the terminology of neural networks as follows: the normal condition data used to construct the novelty detector will be referred to as the *training set* and the process of extracting the detector as *training*. The subsequent data which the detector checks for anomalies will be termed the *testing set*.

3.1. OUTLIER ANALYSIS

An outlier in a data set is an observation that appears inconsistent with the rest of the data and therefore is believed to be generated by an alternate mechanism to the other data. The *discordancy* of the outlier is some quantitative measure of the extent of this inconsistency. The standard reference on outlier detection is reference [15].

In the case of univariate data, the detection of outliers is a relatively straightforward process in that the outliers protrude from one or other end of the data set. The most

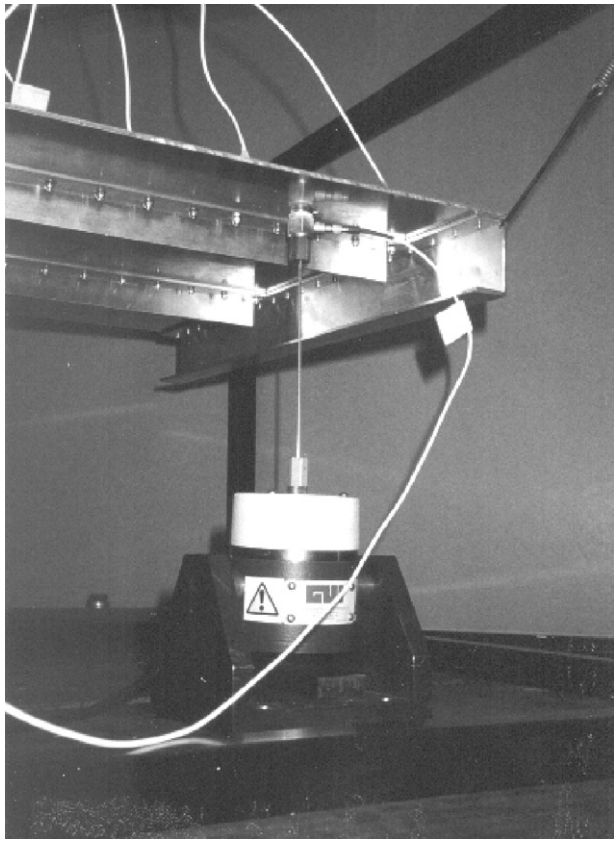

Figure 4. Base of plate showing shaker attachment point.

common discordancy test which exploits this, and the one whose extension to multivariate data will be employed later is given by

$$z_\zeta = \frac{|x_\zeta - \bar{x}|}{s}, \tag{1}$$

where x_ζ is the measurement corresponding to the potential outlier and $\bar{x}$ and s the mean and standard deviation of the sample respectively. The latter two values may be calculated with or without the potential outlier in the sample depending upon whether inclusive or exclusive measures are preferred. This discordancy value is then compared to some threshold value and the observation declared, or not, to be an outlier. There are numerous ways of defining and computing the thresholds for the univariate and multivariate cases. The method used here is based on extreme values and is defined in reference [5]. The threshold is dependent on both the number of observations and the number of dimensions of the problem being studied. The value also depends upon whether an inclusive or exclusive threshold is required.

A multivariate data set consists of n observations in p variables and can be represented as n points in a p-dimensional space. Detection of outliers in multivariate data is more difficult than the univariate situation due to the potential outlier having "more room to hide" in the data space. However, many of the ideas and techniques associated with the

detection of outliers in multivariate data follow on from those applicable to univariate problems.

The discordancy test which is the multivariate equivalent of equation (1) is the Mahalanobis squared distance measure given by

$$D_\zeta = (\underline{\mathbf{x}}_\zeta - \bar{\underline{\mathbf{x}}})^{\mathrm{T}} \mathbf{S}^{-1} (\underline{\mathbf{x}}_\zeta - \bar{\underline{\mathbf{x}}}), \tag{2}$$

where $\underline{\mathbf{x}}_\zeta$ is the potential outlier, $\bar{\underline{\mathbf{x}}}$ is the mean of the sample observations and $\mathbf{S}$ the sample covariance matrix. $^{\mathrm{T}}$ indicates transpose.

As with the univariate discordancy test, the mean and covariance may be inclusive or exclusive measures. In many practical situations the outlier is not known beforehand and so the test would necessarily be conducted inclusively. In the experiments discussed here however, the candidate outlier is always known beforehand and so it is more sensible to calculate a value for the Mahalanobis squared distance without this observation "contaminating" the statistics of the normal data.

Note that the training data has to be composed of many patterns representing normal condition in order to produce a diagnostic robust in the presence of measurement noise. If only a few normal patterns are available (which is the situation that arises later for density estimation), these can be artificially corrupted with noise to multiply the training set. Clearly, the latter exercise requires assumptions about the level and colour of the noise process. For simplicity, and to facilitate comparison with theory, the noise is often assumed to be Gaussian.

3.2. AUTO-ASSOCIATIVE NETWORK

This approach employs a standard feed-forward multi-layer perceptron (MLP) neural network trained by backpropagation of errors [16]. The network is trained to reproduce at the output layer those patterns which are presented at the input. In order to ensure that this is a non-trivial process the patterns are passed through hidden layers which have fewer nodes than the input layer. This "bottleneck" structure, presented in Figure 5, forces the network to learn the significant features of the patterns. The activation of the smallest, central layer, corresponds to a compressed representation of the input. According to the interpretation of reference [17], the network is carrying out a process of non-linear principal component analysis, the *eigenstructure* of the normal condition data is learned and the network can signal departures from this condition. The full five-layer network is needed for this exercise (although the full non-linear structure can be relaxed to one in which the third and fifth layer nodes can have linear activation functions; this was not carried out here). In principal, the activations of the hidden nodes could have been used to provide a reduced-dimensional representation of the data for KDE or outlier analysis. However, because the network provides a diagnostic based on the full 50-dimensional data vectors, they were retained to allow a fairer comparison with the other methods.

The novelty index or measure $v(\underline{\mathbf{z}})$, corresponding to a pattern $\underline{\mathbf{z}} = z_i,\ i = 1, 2, \ldots, N$, is defined as the Euclidean distance between the pattern $\underline{\mathbf{z}}$ and the result of presenting it to the network $\hat{\underline{\mathbf{z}}}$,

$$v(\underline{\mathbf{z}}) = ||\underline{\mathbf{z}} - \hat{\underline{\mathbf{z}}}||. \tag{3}$$

If the new data pattern is typical of normal condition and therefore falls within the limits of the data set used for training, it is reproduced accurately at the output of the network, thus $v(\underline{\mathbf{z}}) \approx 0$; otherwise a non-zero $v(\underline{\mathbf{z}})$ indicates abnormality which could correspond to a fault condition. The advantage of the neural network over distance

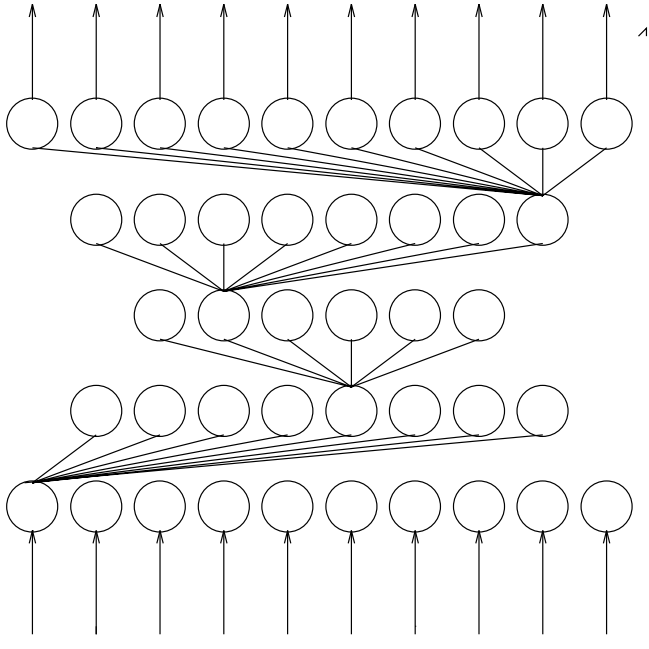

Figure 5. Auto-associative neural network.

measures like the outlier metric discussed above is that it can, in principle, learn complex normal condition sets including those which are non-convex or even disconnected.

Clearly, there is still a need to establish a threshold on the index above which the pattern is considered to differ significantly from normal condition. In this work, the threshold is estimated from the training data. It is set to $\bar{\nu} + \alpha\sigma_\nu$, where $\bar{\nu}$ and σ_ν are respectively the mean and standard deviation of the novelty index ν over the training set. The factor α controls the degree of confidence in normality. In this study, confidence levels of 99 and 99·99% are used and these have α values of 2·576 and 3·891 respectively. Note that there is an implicit assumption here that the statistics of ν are Gaussian or near-Gaussian.

3.3. KERNEL DENSITY ESTIMATION

A more direct approach to novelty detection is to estimate the probability density function (PDF) for the feature vectors over the normal condition set. Once the PDF is known, new data can be accepted or rejected as normal on the basis of the PDF magnitude for the pattern. Patterns with very low values are considered unlikely to have come from the unfaulted distribution and are diagnosed as damaged. There are numerous methods of estimating densities for multivariate data [18], the approach discussed here is the standard *kernel* method as applied previously in reference [9]. The basic form of the estimate is

$$\hat{p}(\underline{\mathbf{z}}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\underline{\mathbf{z}} - \underline{\mathbf{Z}}_i}{h}\right), \tag{4}$$

where $\underline{\mathbf{Z}}_i$ is the ith data point, n is the number of points in the training set and h is the smoothing parameter which controls the width of the individual kernels. $\hat{p}(\underline{\mathbf{z}})$ is the estimate of the true density $p(\underline{\mathbf{z}})$. The *kernel function* can be any localized function satisfying appropriate constraints [18].

The most common choice of kernel function, and the one adopted here is the multivariate Gaussian,

$$K(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}||\underline{\mathbf{x}}||^2), \tag{5}$$

where d is the dimension of the data space.

Once the estimate $\hat{p}(\underline{\mathbf{z}})$ is established, the PDF values at any new measurement points are trivially evaluated.

The kernels above have equal radii in all directions. In order to allow elliptical atoms with more effective coverage of the data, the data here are transformed using the method of Fukunuga [5].

The quality of the estimate depends critically on two factors. The first is the size of the training set; this has been discussed in previous papers and will arise again later. The other factor of importance is the value of h. If h is too small, the PDF will contain a lot of spurious local structure. If h is too large, the estimate will be oversmoothed and its decay rate with $\underline{\mathbf{z}}$ will be underestimated. The method used to establish the "correct" h here is *least-squares cross-validation* [18].

4. NOVELTY DETECTION RESULTS

The first results given here are based on the transmissibility data for the path AB in Figure 2, i.e., directly along the stringer. This analysis was carried out first as it was anticipated that it would be most sensitive to damage.

4.1. PRE-PROCESSING

Pre-processing of the patterns was kept to a minimum, it was decided to use the raw transmissibility functions for the basic patterns. Figure 6 shows the magnitude of the normal condition transmissibility for the path AB (Figure 2). (As the magnitude proved sensitive enough to the damage in previous studies i.e., reference [4], the phase was discarded.) As one might expect, the functions only proved sensitive to the damage in the immediate vicinity of the peaks. There are four dominant peaks.

The next step in the pre-processing was to reduce the dimension of the patterns as far as possible; 2048 spectral lines is excessive. It was decided to single out the four dominant peaks and examine which ones showed highest sensitivity to the damage. The first peak (*first of the dominant four*) proved almost insensitive to the damage and was discarded. Figure 7 shows that the second peak showed significant variations as the damage level increased (the four lowest levels of damage are shown). The third peak appeared to be insensitive to all but the highest levels of damage and was also rather noisier than the other peaks. The fourth peak (Figure 8) showed significant variations with damage level; furthermore, the variations have a systematic nature, the peak frequency shifts downward with increasing damage and the peak size increases. Because peak 4 showed marked *and* systematic variation in its form as a function of the damage, it was selected for the basic feature. Spectral lines 1886–1935 were selected to form the 50-dimensional feature vector.

The averaged transmissibilities took over 20 min to acquire. In order to allow a fast on-line diagnostic, it was decided to use unaveraged data to train and test the novelty

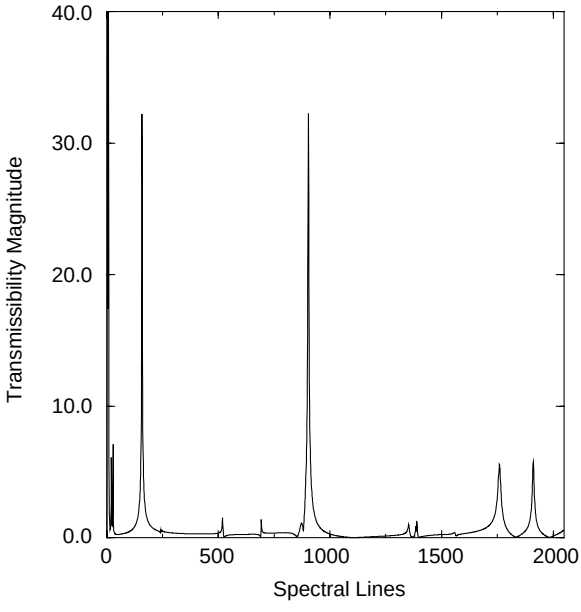

Figure 6. Measured transmissibility magnitude along path AB.

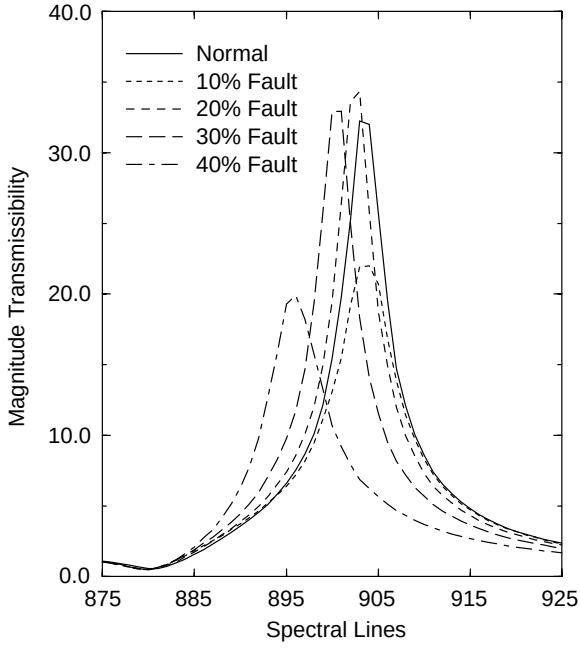

Figure 7. Variation in transmissibility peak 2 with damage.

detectors. Figure 9 shows three examples of unaveraged patterns from peak 4 of the transmissibility, the level of noise is clearly substantial in the region of the maximum.

In order to train the various novelty detectors, 118 of the raw patterns were used to form the training set. The remaining 10 patterns were taken, with each of the 10 patterns from the nine damage levels, to form the testing set.

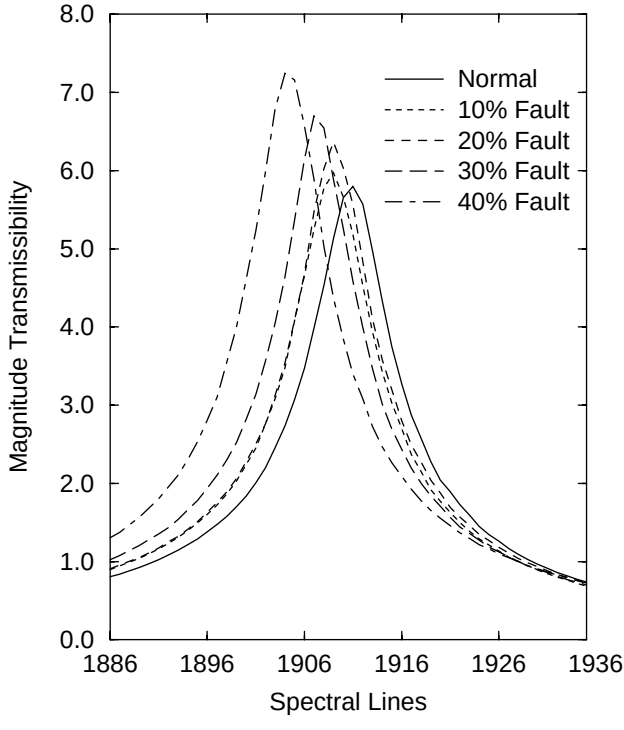

Figure 8. Variation in transmissibility peak 4 with damage.

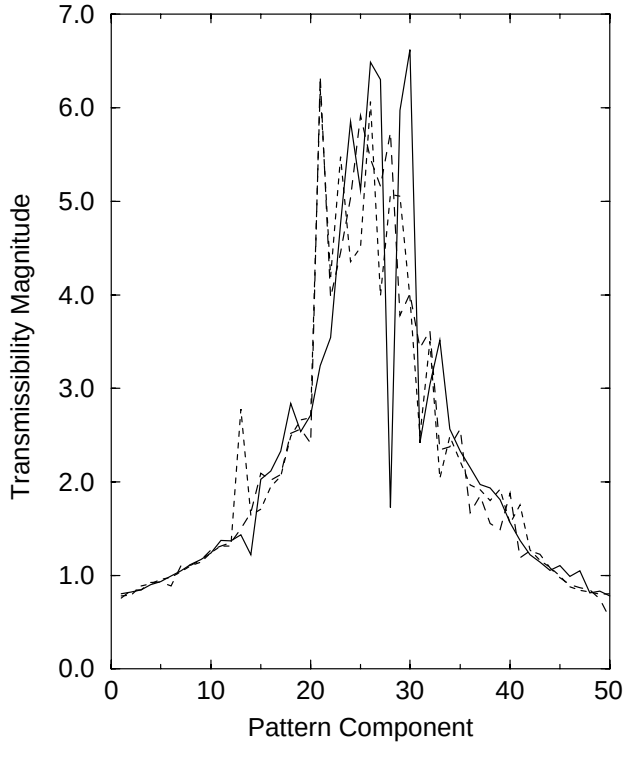

Figure 9. Three samples of the normal condition data.

4.2. OUTLIER ANALYSIS

The training set was used to estimate a mean vector and covariance matrix for the patterns. Equation (2) was then used to obtain the Mahalanobis distances for all points in the training and testing sets. The results are shown in Figures 10 and 11 respectively. The 99·99% (dotted) and 99% (dashed) confidence threshold are shown, these were estimated using the Monte Carlo method and found to equal 285·0 and 227·8 respectively. All points in the training set fall well below the threshold as required. In the case of the testing set, apart from one isolated excursion, the analysis fails to flag the damage until the cut is 30% of the way through the stringer (7·5 mm long). (The vertical dashed lines in Figure 11 separate the normal data and the nine damage states.) Beyond this point, damage is signalled unambiguously. Note that the distance increases to a maximum and then decreases again. This is due to the fact that at high levels of damage, the maximum for peak 4 actually shifts out of the pattern window. When the resonance is in the wrong place, there are two sources of novelty, one is the displaced peak and the other is the missing peak at the location corresponding to normal condition. Once the resonance has moved from the window, there is only the latter source of novelty and the Mahalanobis distance comes down as a result. Note that for the 99% threshold, there is one false positive (no real cause for alarm), while for the 99·99% there are no false alarms. However, for the higher threshold the unequivocal damage level is 40%.

4.3. THE NEURAL NETWORK

The neural network was given a structure 50:40:30:40:50 as this had proved useful with 50-dimensional patterns in the past [4]. A rigorous approach could have been followed

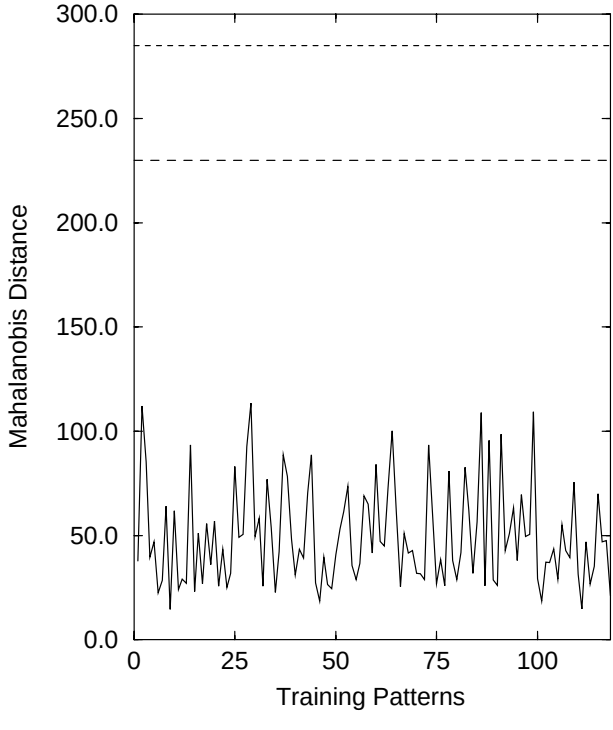

Figure 10. Mahalanobis distances on training set (normal condition data).

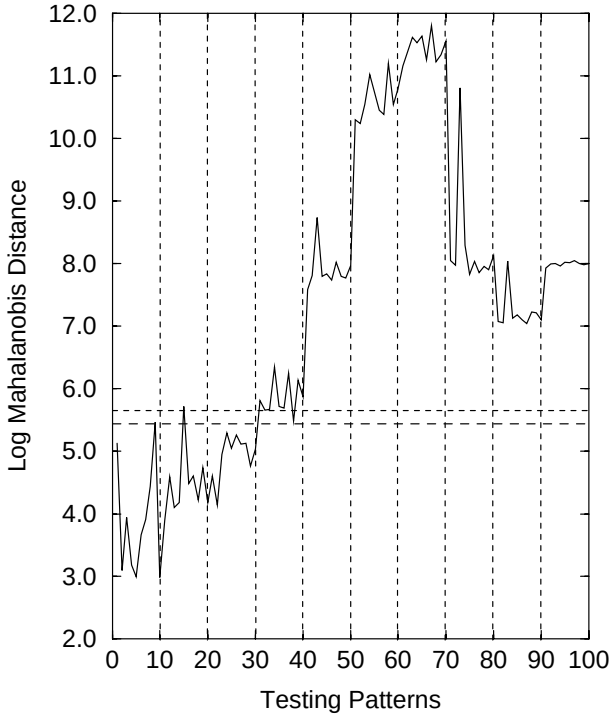

Figure 11. Mahalanobis distance on testing set (damage data).

which used a separate validation data set to optimize the structure. However, the application of the structure above proved acceptable. Note that there are no claims here to optimality. Because of the small training set and the large number of connections in the network, Gaussian white noise was added during training in order to provide regularization. The noise-to-signal ratio was approximately 10%. After training, the value of the index in equation (3) was computed for all points in the training and testing sets and the results are shown in Figures 12 and 13 respectively. As before the thresholds correspond to 99 and 99·99% confidence. There is one excursion above the 99% threshold on the training set—again this is no cause for alarm. On the testing set, at the 99% confidence level the index is consistently high from the 20% damage level onwards; however, it only signals damage unequivocally from 40% onwards. This corresponds to detection of a 10 mm cut in the stringer. At the 99·99% level, the network only responds with better than 50% accuracy from the 40% damage level onwards. Overall, based on the 99% confidence level, it might be argued that the network is a slightly more sensitive diagnostic than the outlier analysis.

4.4. KERNEL DENSITY ESTIMATION

The final method of analysis used KDE. As this is known to be more sensitive to small data sets, a pseudo-synthetic training set of 1000 points was obtained by computing the mean and covariance matrix of the training set and then generating 1000 points around the

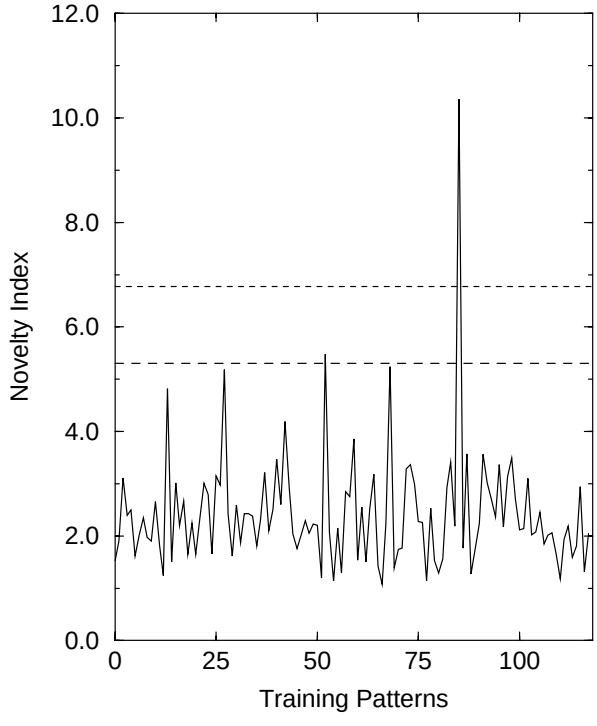

Figure 12. ANN novelty index on training set (normal condition data).

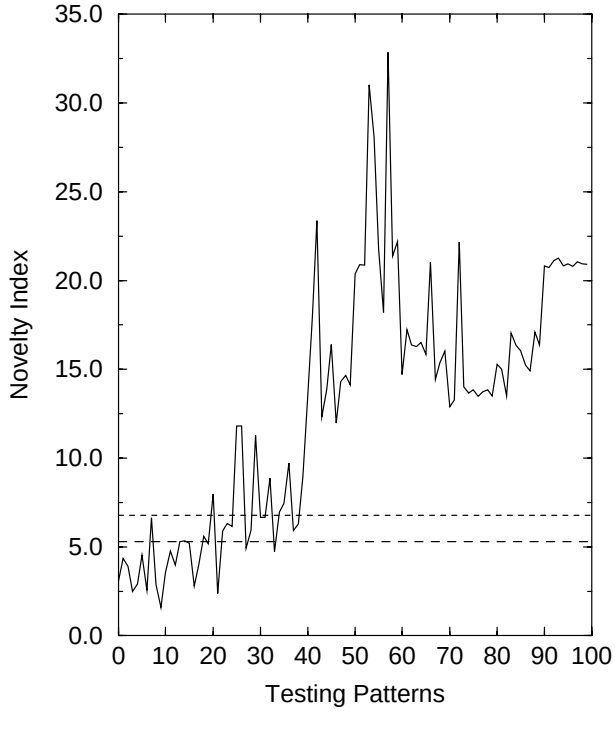

Figure 13. ANN novelty index on testing set (damage data).

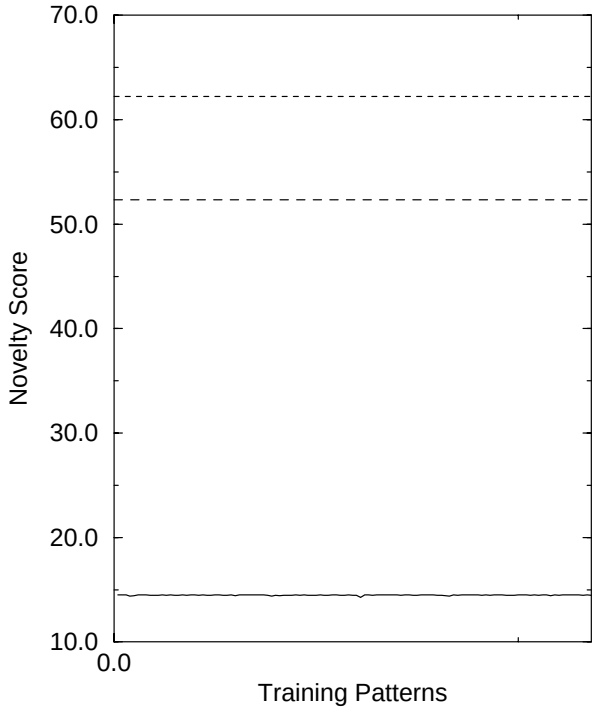

Figure 14. KDE novelty score on training set (normal condition data).

mean with the corresponding Gaussian distribution.[‡] A first attempt using the smoothing parameter from least-squares cross validation gave results which were clearly undersmoothed. This is understandable, as the cross-validation calculation itself may suffer if the data set is sparse. The value of 0·8 was increased to 2·0 in order to give more smoothing. The results are shown in Figures 14 and 15 in the form of a *novelty score*, which is $-\log \hat{p}(\underline{\mathbf{z}})$, on the training and testing sets. The results are probably still undersmoothed as witnessed by the small variation in the score over the training set. The damage is only detected unambiguously beyond the 40% stage. Note that the novelty score appears to saturate at an upper bound. At these points, the density was returned as zero (to machine precision) and an arbitrary value of $\hat{p}(\underline{\mathbf{z}}) = 10^{-100}$ was assigned.

The thresholds shown in Figures 14 and 15 correspond to 99 and 99·99% confidence as before. They were computed as follows. Given a 50-dimensional Gaussian density, the radius was calculated which would bound 99 or 99·99% of the probability mass. The value of the density at these radii were found and used to scale the estimated density i.e., to find the value of density corresponding to 99% or 99·99% confidence.

[‡]The authors are aware of the apparent contradiction here. If the distribution is assumed to be Gaussian, then it is characterized completely by the mean and standard deviation. The object of the exercise is a little different here. A preliminary analysis showed that the density estimation did not work with 118 data points. The pseudo-synthetic data set was constructed in order to see if 1000 points would suffice. It was assumed that up to 1000 normal condition measurements would be available. This was not the case here because the authors were constrained by the testing schedule. In practice, if a structural health monitoring system based on this approach were deemed desirable, testing might be scheduled for the appropriate number of normal conditions.

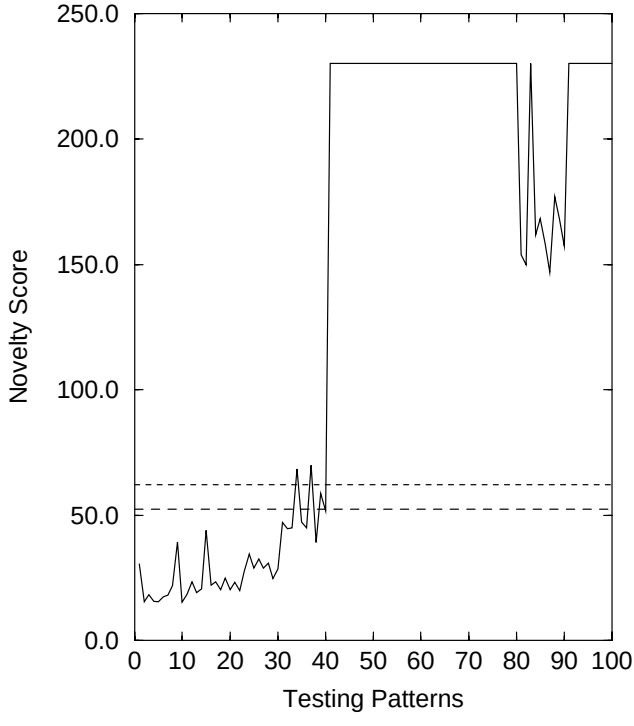

Figure 15. KDE novelty score on testing set (damage data).

4.5. DISCUSSION

The results of the three methods of analysis are summarized in Table 1.

Considered in terms of the two thresholds, the table argues that the 99% threshold is better in terms of sensitivity for detection. The 99·99% threshold misses 14 damage states disclosed by the lower threshold. However, the lower threshold gives two false positives for detection, while the higher threshold gives none. Statistically, two false positives from 20 cases at the 99% confidence level is somewhat unlikely and is probably an indication that the assumption of Gaussianity used in constructing the thresholds is not quite holding. On the whole, the 99% threshold appears to be preferable because of the heightened sensitivity.

In terms of the three methods, at the 99% confidence level, the outlier analysis is the only method to unequivocally signal damage at the 30% level. However, the neural net only misses one of the 30% damage cases, and in addition signals 8 from 10 of the 20% cases. The neural net appears to be a little more sensitive. KDE can be seen from the table to lag in sensitivity.

In fact, the results from KDE give some cause for concern. Of the three methods, it is arguably the most sensitive to using a sparse training set. Applying the guidelines in reference [18], 1000 training points would appear to be a ridiculously small number for a 50-dimensional data set. In fact, Silverman's guidelines are rather conservative and sparse data sets can still be acceptable if the intrinsic dimensionality of the data is much less than 50. Reducing the dimension of the data by some means is always a possibility; however, there is a caveat here. Principal component analysis, for example, discards those dimensions in the data which contribute least to the overall power in the signal. In reality,

TABLE 1

Summary of novelty detector results

Method	Confidence	Detections at damage percentage									
		0	10	20	30	40	50	60	70	80	90
OA	99·00	1	1	0	10	10	10	10	10	10	10
OA	99·99	0	1	0	9	10	10	10	10	10	10
ANN	99·00	1	2	8	9	10	10	10	10	10	10
ANN	99·99	0	0	4	5	10	10	10	10	10	10
KDE	99·00	0	0	0	3	10	10	10	10	10	10
KDE	99·99	0	0	0	2	10	10	10	10	10	10

one might discard precisely those components which distinguish between normal condition and damage simply because they are relatively small. In support of this, it is shown in reference [19] that a high-dimensional data set allows a kernel-based method to substantially outperform a neural classifier. When the data is transformed to a lower-dimensional representation with PCA, the performance of the kernel method is reduced to that of the neural network. In many ways one should adopt a pragmatic approach to statistics, if a method appears to be working well and is supported by a sensible validation strategy, it seems uneconomical to discard it on the basis of a possibly highly conservative theory. Having said this, the present study has two methods of analysis which are outperforming the density estimation method and do not give such substantial reservations concerning the high dimensions. The results above indicate that KDE may not be appropriate here.

Of the other two methods, both the outlier statistic and the neural network give acceptable results, with the neural network marginally more sensitive. The deciding argument in favour of the outlier statistic is, however, its simplicity. Occam's razor demands that it be given priority. There is also an important practical point here. Whether or not statistical techniques will find acceptance for SHM of aerospace structures or not will be decided by the question of certification. If the safety and reliability of the methods can be established, there is a chance they will pass into use. In these terms, the outlier analysis, which is a simple linear algebraic procedure will lend itself to reliability analysis far easier than the highly non-linear complex neural network. This is not a reason to discard neural networks, it is a reason to explore methods of certification for such systems. With the simplicity of the outlier analysis come a number of restrictions, the assumption of Gaussianity being one of the foremost. If the normal condition sets for some real structures have complex topographies, for example if they are non-convex, the outlier analysis will fail. However, it is suggested here as the first choice for an application.

4.6. OUT-OF-LINE TRANSMISSIBILITIES

The next stage in the analysis concerned the off-stringer transmissibilities i.e., those along lines AC and CD (Figure 2). In both cases, the dominant peaks were examined in order to identify those with the most marked systematic variation with damage. In both cases, the same peak as for transmissibility AB was selected; data from the same 50-point window was used. The analysis for case AB indicated that the outlier analysis was the method of choice in terms of sensitivity to damage *and* computational simplicity, so only

this method was applied to the off-stringer data. The analysis was carried out in an identical fashion to the *AB* case.

Figure 16 shows the results of outlier analysis on the testing set for the *AC* transmissibility. All points in the training set were sub-threshold as in Figure 10. On the testing set, there are no excursions above threshold over the normal condition data (first 10 points of Figure 16). There are two points signalling damage on the 10% damage set; however, there are none on the 20% damage data and it is only when the damage level reaches 30% that the distance measure is consistently above threshold. One concludes that the analysis flags 30% damage, i.e. the 7·5 mm cut. The damage measure is not monotonically increasing with damage for the same reason as discussed in the transmissibility *AB* case.

The results for outlier analysis on the *CD* transmissibility testing set are given in Figure 17. The results are broadly similar to those for paths *AB* and *AC*: the analysis detects damage consistently when the level is above 30%.

The conclusion from the analysis of different transmissibility paths is that the damage detection capability does not seem to be overly sensitive to the transmissibility path within the restricted range considered here.

4.7. REDUCED DIMENSION

At this stage, it was decided to investigate the effect of reducing the dimension of the data set. As discussed above, the usual arguments based on the "curse of dimensionality" in statistical procedures would suggest that 50 dimensions may be excessive. It is shown above, by the performance of the outlier analysis that the information for diagnosis is

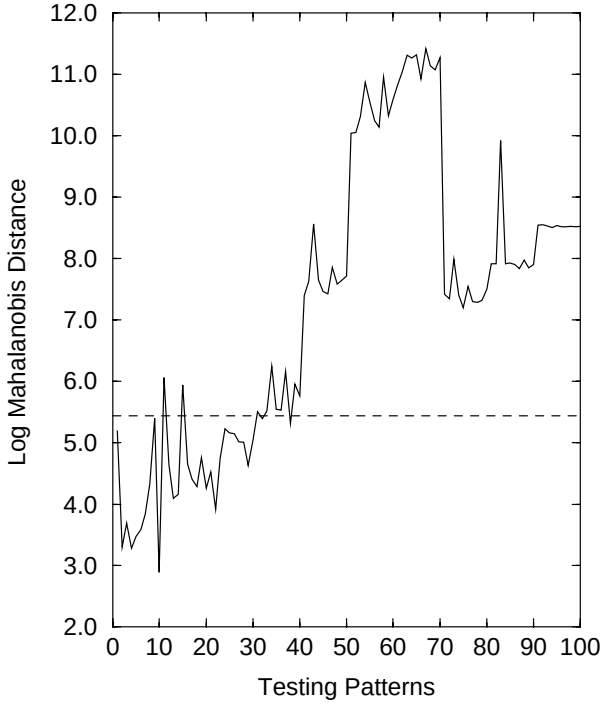

Figure 16. Mahalanobis distances on testing set—transmissibility *AC*.

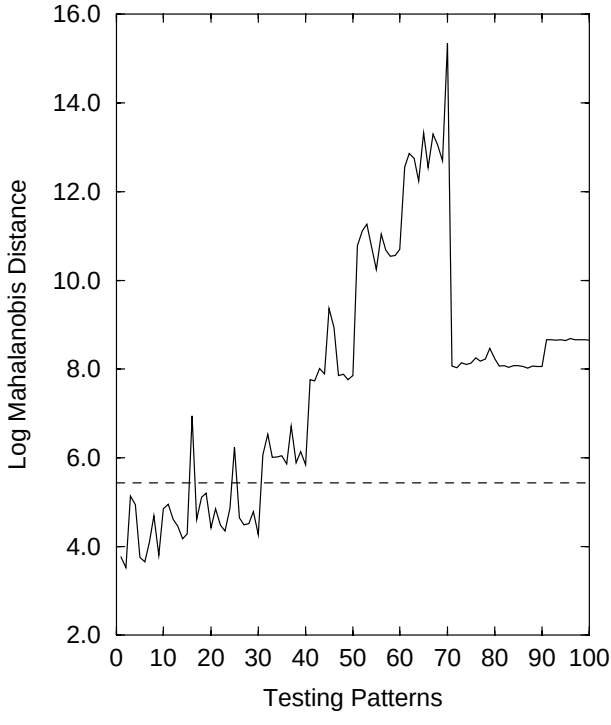

Figure 17. Mahalanobis distances on training set—transmissibility *CD*.

present in the 50 dimensions. However, one might argue that a similar performance might be obtained with a lower-dimensional feature set.

In order to examine the effect of reducing the dimension, the first strategy was simple sub-sampling. 10 equally spaced *AC* transmissibility lines were taken from the 50-dimensional feature vector and used for an outlier analysis. The results are shown in Figure 18 for the testing set. Although the results are superficially very close to the results for the full feature set (Figure 11), there is an overall drop in the peak discordancy from just below 12·0 (in the units of the figure) to 9.5. Also the threshold for unequivocal detection moves from 30 to 40% damage. Dimensions between 10 and 50 give results which interpolate smoothly between the levels of performance shown. As the outlier statistic is not showing any signs of pathology as a result of the higher dimension, it is considered acceptable for further use—this will be described in the later papers in this sequence.

Figure 19 shows the performance of a 10-dimensional feature set extracted by PCA. The caveat discussed in an early section is illustrated. Not only is the level of discordancy vastly reduced, but it appears that PCA has made a qualitative change in the data. There is no fall in the discordancy at high levels as the peak moves out of the window.

5. DISCUSSION AND CONCLUSIONS

This report describes an experimental verification of an approach to structural health monitoring. The structure of interest is a simulated aircraft wingbox constructed from aluminium and incorporating features found in true aircraft panels, namely ribs and

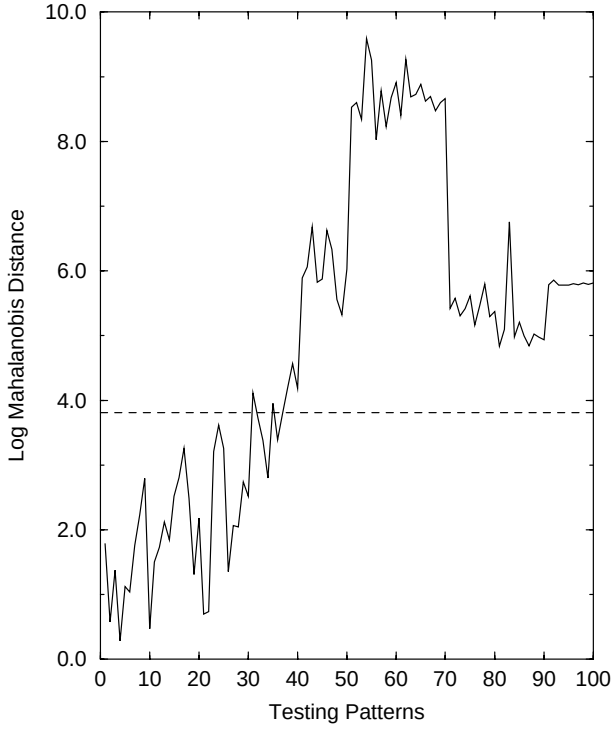

Figure 18. Mahalanobis distances on testing set—transmissibility AC reduced to 10 dimensions by sub-sampling.

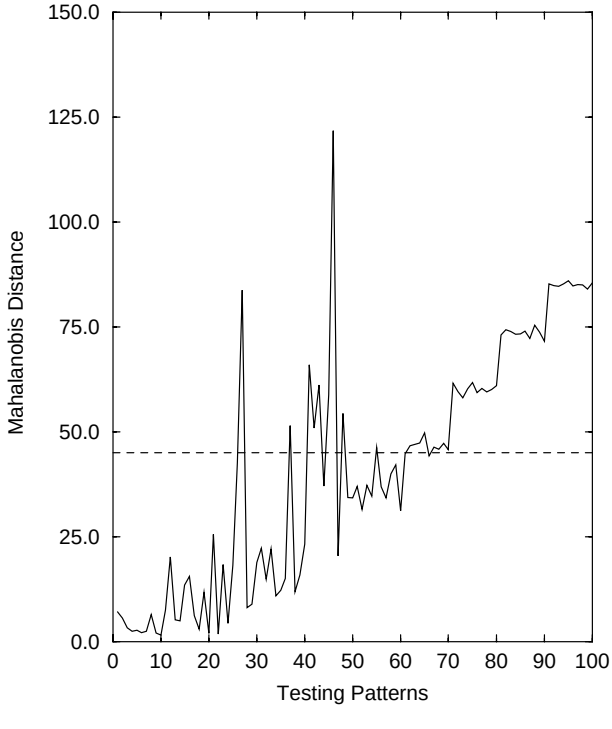

Figure 19. Mahalanobis distances on testing set—transmissibility AC reduced to 10 dimensions PCA.

stringers. The object of the exercise was to detect damage induced in the stringer using vibration measurements.

The method considered was a level one diagnostic based on the idea of novelty detection. The patterns used for detection were measured transmissibilities centred on a particular peak which proved sensitive to the damage. Three different algorithms were applied: outlier analysis, a neural network approach and kernel density estimation. Each proved successful in detecting damage to an acceptable extent. Outlier analysis and the neural network proved most sensitive, giving consistent indications of damage corresponding to a 7·5 mm saw-cut. The approach based on density estimation gave some cause for concern about the size of the training set for such high-dimensional patterns. In fact, all of these approaches suffer to some extent from the "curse of dimensionality", however, for the outlier analysis and neural network there are mitigating circumstances. In the case of the outlier analysis, the damage threshold is determined by a Monte Carlo procedure which implicitly takes account of the size and dimension of the training set and this may well contribute to the apparent robustness of the method. In the case of the neural network, noise was added during training as a simple means of regularization, and thus the generalization properties of the network may well have been adequately determined. Overall the outlier analysis was selected as the means of moving forward as it offered a much simpler solution than the neural network with no apparent deterioration in performance. It was noted, however, that the very simplicity of the approach could preclude its use in some applications and that this should be monitored.

The method was shown to be partly insensitive to the transmissibility path used. This is important as the requirement that detectable damage be located directly between sensors is too restrictive to be of any practical use. As the transmissibility estimation requires no specification of the input excitation, it can potentially be used in an in-flight context.

One important issue which the analysis raised was the question of averaging the data. When averaged transmissibility data were examined, even the lowest level of damage (2·5 mm cut) gave a pattern distinct from normal condition. However, this required 128 averages which took 20 min to acquire. The results discussed above were for single measurements which took 10 s to obtain and the sensitivity, detection threshold of 5·0–7·5 mm damage, reflects this. There is a clear trade-off between speed of acquisition of the patterns and sensitivity of the diagnostic. If an on-line monitoring system is needed, the degree of averaging which will be possible will depend on the desired frequency of the health reports. It is strongly advised that the maximum averaging consistent with speedy reportage should be used.

A final remark concerns the nature of the simulated damage. The saw-cut was sufficiently wide that the cut surfaces never met, at the levels of vibration experienced here. This means that the damaged system is still linear, but simply with a local reduction in stiffness. In the case of a true fatigue crack, the crack would be likely to "breathe" i.e., open and close. This induces a local non-linearity in the stiffness with a reduction only when the crack is open. Further non-linearity in the form of friction may appear if the crack faces rub. The transition from a linear to a non-linear system with damage may actually help in the design of a diagnostic system as energy will be transferred between frequency bands and this may lead to sensitive features for damage detection.

ACKNOWLEDGMENTS

The foremost acknowledgement must be to DERA, Air Systems Sector, Structures Department for funding this project through a DTI CARAD scheme. The authors would

also like to thank Mr David Webster for constructing the experimental rig and lending indispensable technical support. Finally, the authors are grateful to the referees for their constructive criticism.

REFERENCES

1. A. Rytter 1993 *Ph.D. Thesis, Department of Building Technology and Structural Engineering, University of Aalborg, Denmark*. Vibration based inspection of civil engineering structures.
2. C. M. Bishop 1994 *IEE Proceedings on Vision and Image Signal Processing* **141**, 217–222. Novelty detection and neural network validation.
3. L. Tarassenko, P. Hayton, N. Cerneaz and M. Brady 1995 in *Proceedings of 4th IEE International Conference on Artificial Neural Networks, Cambridge, UK* IEE Conference Publication No. 409, 442–447. Novelty detection for the identification of masses in mammograms.
4. K. Worden 1997 *Journal of Sound and Vibration* **201**, 85–101. Structural damage detection using a novelty measure.
5. K. Worden, G. Manson and N. R. Fieller 1999 *Journal of Sound and Vibration* **229**, 647–667. Damage detection using outlier analysis.
6. S. Roberts and L. Tarassenko 1994 *Neural Computation* **6**, 270–284. A probabilistic resource allocating network for novelty detection.
7. D. A. Pomerleau 1993 in *Advances in Neural Information Processing Systems*, Vol. 5 (S. J. Hanson, J. D. Cowan and C. L. Giles, editors). Los Altos, CA: Morgan Kaufman Publishers Input reconstruction reliability estimation.
8. W. J. Staszewski, S. G. Pierce, K. Worden, W. R. Philp, G. R. Tomlinson and B. R. Culshaw 1997 *Optical Engineering* **36**, 1877–1888. Wavelet signal processing for enhanced Lamb wave defect detection in composite plates using optical fibre detection.
9. K. Worden 1998 in *Proceedings of ISMA 23, Noise and Vibration Engineering Conference, Leuven*, 81–89. Damage detection using multivariate statistics: Part II: kernel density estimation.
10. N. Stubbs, S. Park, C. Sikorsky and S. Choi 2000 *International Journal of System Science* **31**, 1361–1374. A level IV global nondestructive damage assessment methodology for civil engineering structures.
11. S. W. Doebling, C. R. Farrar, M. B. Prime and D. W. Shevitz 1996 *Los Alamos National Laboratories Report LA*-13070-*MS*. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review.
12. G. Manson, K. Worden and D. J. Allman 2003 *Journal of Sound and Vibration* **259**, 345–363. Experimental validation of structural health monitoring methodology: Part II. Novelty detection on an aircraft wing.
13. G. Manson, K. Worden and D. J. Allman 2003 *Journal of Sound and Vibration* **259**, 365–385. Experimental validation of structural health monitoring methodology: Part III. Damage location on an aircraft wing.
14. K. Worden and G. Manson 2001 *in Proceedings of International Workshop on Damage Assessment in Structures — DAMAS 2001, Cardiff*, 35–46. An experimental appraisal of the strain energy damage location method.
15. V. Barnett and T. Lewis 1994 *Outliers in Statistical Data*, Chichester: John Wiley and Sons; Third edition.
16. S. Haykin 1994 *Neural Networks. A Comprehensive Foundation*. New York: Macmillan College Publishing Company.
17. M. A. Kramers 1991 *AIChE Journal* **37**, 233–243. Nonlinear principal component analysis using autoassociative neural networks.
18. B. W. Silverman 1986 *Chapman & Hall Monographs on Statistics and Applied Probability*, No. 26. Density estimation for statistics and data analysis.
19. K. Worden and G. Manson 2000 *Inverse Problems in Engineering* **8**, 25–46. Damage identification using multivariate statistics: kernel discriminant analysis.