Journal of Sound and Vibration (2003) **259**(2), 345–363
doi:10.1006/jsvi.2002.5167, available online at http://www.idealibrary.com on IDEAL®

EXPERIMENTAL VALIDATION OF A STRUCTURAL HEALTH MONITORING METHODOLOGY. PART II. NOVELTY DETECTION ON A GNAT AIRCRAFT

G. Manson and K. Worden

*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 concerns the second phase of an experimental validation programme for a structural health monitoring methodology based on novelty detection. This phase seeks to apply one of the methods considered in the first stage of the work on a more realistic structure, namely the wing of a Gnat aircraft, as opposed to the previously investigated laboratory structure. The novelty detection algorithm used is that of outlier analysis and damage is introduced by making several copies of an inspection panel, each with a different controlled fault. All of these faults were detectable, a single feature was highlighted which proved capable of separating all the fault conditions from the unfaulted.

1. INTRODUCTION

The safety gains and financial rewards to be made from developing a robust on-line structural health monitoring (SHM) system far outweigh the financial and other costs associated with obtaining such a system. A large proportion of the cost of aircraft ownership is due to current lengthy inspection procedures. Integrated SHM technology would significantly reduce inspection times whilst delivering an objective criterion stating whether or not an aircraft was damaged.

This paper details work carried out as part of an on-going programme of SHM research at the University of Sheffield, funded by DERA Farnborough. The first phase of the work [1] looked at the experimental verification of a novelty detection method, carried out using a model wingbox which consisted of an aluminium panel stiffened with ribs and stringers. The damage was induced by making a progressive saw-cut through one of the stringers. The novelty detection method was successful in signaling all the damage states where the depth of the cut exceeded 5 mm. The features used for the analysis were transmissibilities

† Now QinetiQ

0022-460X/02/$35.00

measured along the stringer. Three novelty detection methods were investigated: kernel density estimation (KDE), artificial neural network (ANN) and outlier analysis. The last method proved to be the most robust of the three.

In this paper, the work is extended to diagnosis of local damage on a full-scale structure - in this case a Gnat aircraft. The damage was effectively introduced into an inspection panel and as before, transmissibility measurements (across the panel) were used to try and detect the damage. Outlier analysis was the chosen method of novelty detection throughout the study.

The layout of this paper is as follows: section 2 describes the experimental layout and strategy and section 3 shows, visually, how the damage affected the measured transmissibilities. section 4 discusses the issue of selecting the best features for novelty detection followed by the results of the outlier analysis in section 5. The paper is completed with a discussion and some conclusions.

2. THE EXPERIMENTAL PANEL AND DATA CAPTURE

The structure to be tested was a Gnat trainer aircraft. It was not possible to damage the aircraft, so it was decided to simulate damage to an inspection panel on the starboard wing. This was accomplished by making 10 copies of the panel; one was left intact and the remaining nine received controlled damage. One panel was reserved as normal condition because it proved impossible to exactly match the aircraft aluminium of the true panel. As close a match as possible was made. The geometry of the panel is shown in Figure 1. (This figure is schematic and not to be regarded as a precise engineering drawing.)

The panel was fixed to the wing by 23 screws. There were originally 26 threaded holes but a previous removal of the panel had involved drilling out some of the screws resulting in damage to three of the threaded holes.

Problems were anticipated during the test as a result of the variability in the attachment of the panel. Because a different panel was used for each damage state, it had to be fixed onto the wing prior to the test and then removed. This produced inevitable variations in the fixing/boundary conditions. An attempt was made to minimize the problem by using a constant torque electric screwdriver. A more rigorous approach would have been to use a screwdriver head in a torque wrench and to make sure that the same torque was applied to

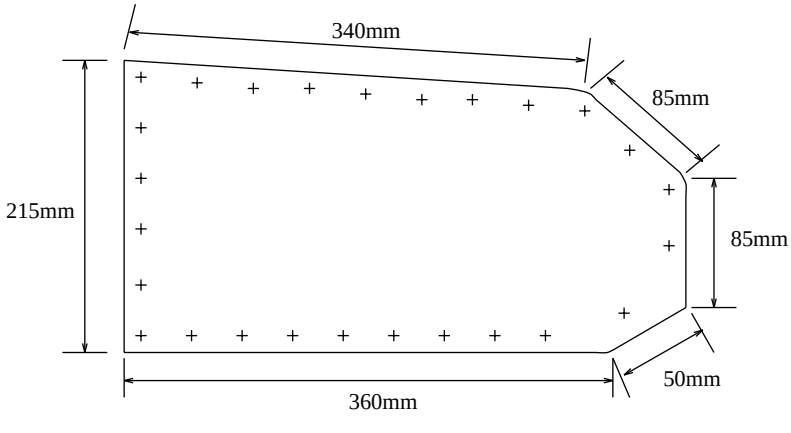

Figure 1. Geometry of Gnat inspection panel.

each screw in each test. This was not carried out because: (a) it is simply unrealistic to assume that degree of care is used in removing and replacing inspection panels and; (b) it would have been far too time consuming.

As in the previous experimental study [1], it was decided to use transmissibility data in order to locally monitor the region of interest. The sensors used were piezoelectric accelerometers of PCB type. Four sensors were used in all: one pair to measure the transmissibility across the panel in the length direction (T1 and T2 in Figure 2) and one pair across the width (T3 and T4 in Figure 2). An additional pair were fixed across a panel diagonal as shown in the photograph in Figure 3; however, the data capture proved too time consuming to use this pair. The accelerometers were fixed with beeswax.

The wing was excited using a Ling electrodynamic shaker attached directly below the inspection panel on the bottom surface of the wing. A white Gaussian excitation was generated within the acquisition system and amplified using a Gearing and Watson power amplifier.

The transmissibilities were measured using a DIFA Scadas 24-channel acquisition system controlled by LMS software running on a HP workstation as shown in Figure 4. Both real and imaginary parts of the functions were obtained. In order to resolve the defects in the inspection panel, it was assumed necessary to excite modes with appropriately short wavelengths and hence high frequencies. In order to select a suitable excitation band, a transmissibility between transducers 1 and 2 was measured in the range 0–2000 Hz with the undamaged inspection panel attached.

The panel was then completely removed (in order to give a worst-case damage state) and the measurement repeated. Comparison between the two transmissibilities confirmed that the lower frequency modes were insensitive to the damage, so the excitation band for the main body of tests was selected as 1000–2000 Hz. In all cases 2048 spectral lines were measured.

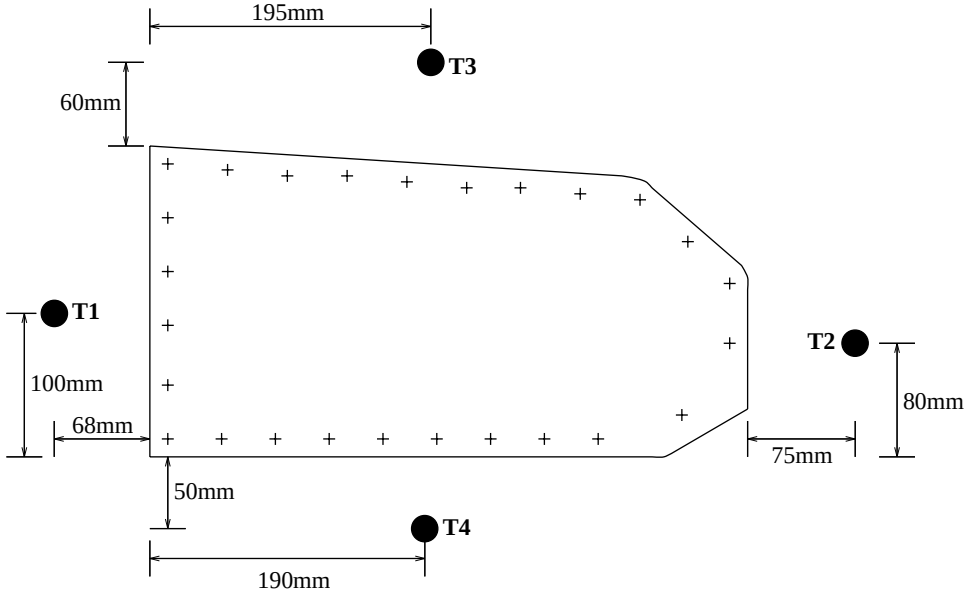

Figure 2. Sensor positions for transmissibility measurements.

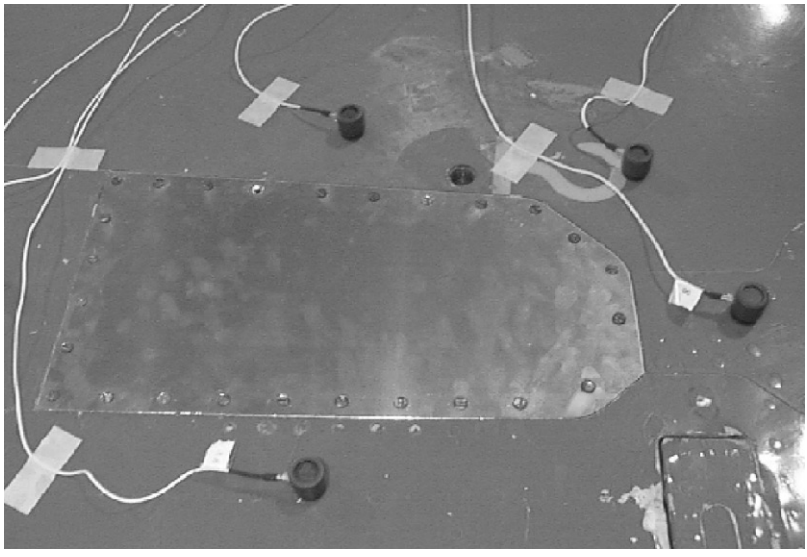

Figure 3. The wing in the vicinity of the inspection panel showing the positions of the transducers.

Figure 4. The Gnat aircraft with the acquisition system.

The measurement strategy for each transmissibility was as follows. First, for the undamaged panel the function was obtained using 128 averages. This was done to provide a clean reference signal to help with feature selection. Next, 110 measurements were taken sequentially using only a single average. Of these, 100 would be used to establish the statistics of the patterns for the novelty detection and 10 would be used for testing. The single average data was very fast to acquire and the most likely candidates for an on-line

system. The use of 8-average data was investigated but was rejected on the grounds of acquisition time.

The second series of tests worked through the damaged panels in the order shown in Figure 5. Damage states (a), (b) and (c) were holes of diameter 20, 38 and 58 mm respectively. States (d), (e) and (f) were saw-cuts across the width of the panel with (d) an edge cut of 50 mm and (e) and (f) central cuts of extent 50 and 100 mm respectively. States (g), (h) and (i) were saw-cuts along the long axis of the panel with (g) a 100 mm edge cut

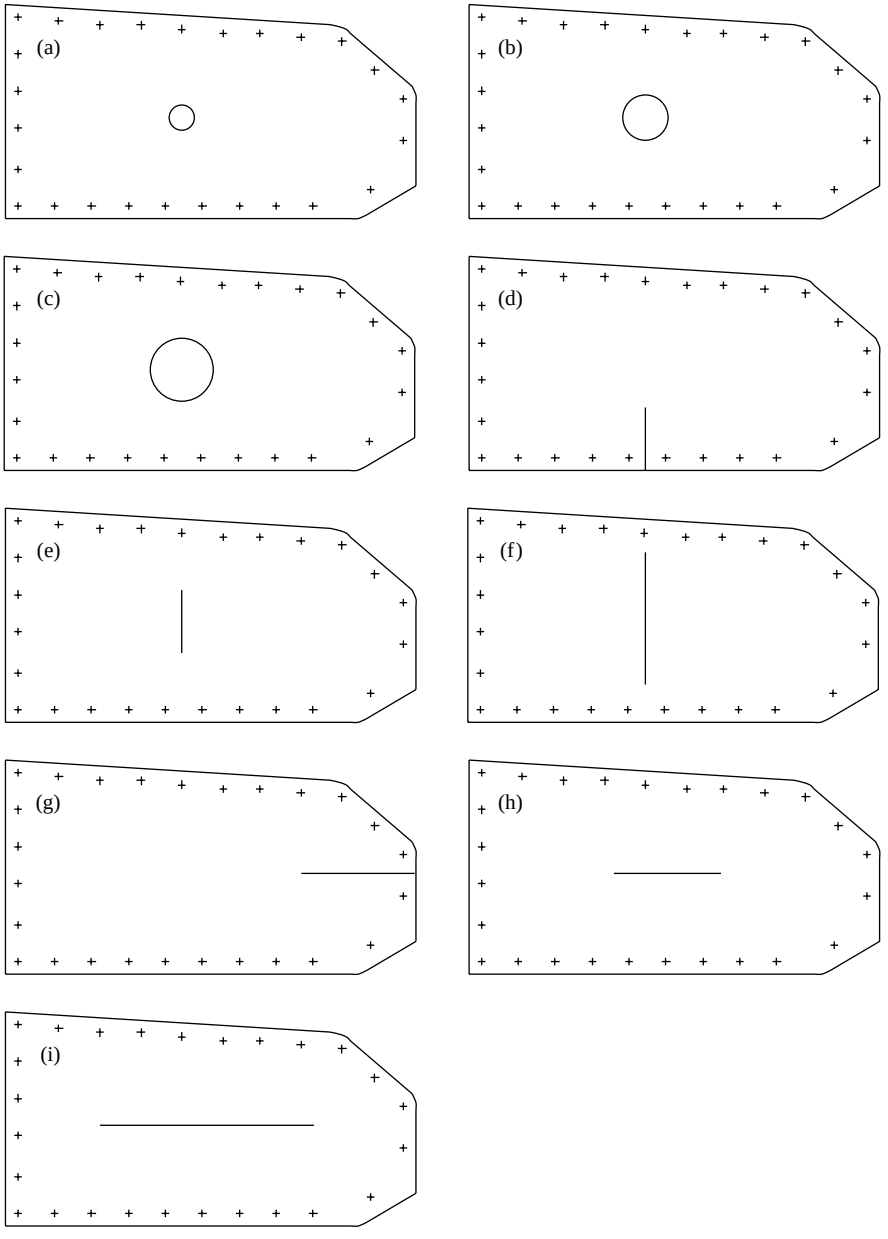

Figure 5. Schematic of damage states.

and (h) and (i) central cuts 100 and 200 mm long respectively. For each panel, the first function measured took 128 averages, again this was for reference and feature selection. The second set of tests on each panel recorded 10 single-shot functions sequentially, all for testing the novelty detector.

After all the damaged panels had been tested, another set of measurements were taken for the situation with the undamaged panel to check repeatability. These consisted the same 11 functions as taken for the damage states: one 128-average measurement and 10 raw measurements. The latter 10 were all for testing the novelty detector.

The next sequence of tests took the same 11 measurements for the situation with the panel removed completely.

The penultimate series of tests also addressed the problem of repeatability. The undamaged panel was fixed and tested and then removed four times in order to investigate the effect of the fixing conditions on the measured features. Only 128-average patterns were taken. Another important reason for this portion of the programme was to obtain a range of normal condition patterns which characterized the variation in the fixing conditions. These could be used to construct the normal condition data for the novelty detector.

Finally, a 128-average transmissibility was taken with the panel completely removed again. The object of this test was to investigate test variability which could not be associated with the fixing conditions, i.e., variability as a result of environmental changes and instrument drift. This is the least variability which could be expected.

3. PRELIMINARY INSPECTION OF THE TRANSMISSIBILITIES

After the measurements were taken, an inspection of the 128-average transmissibilities was made in order to assess the repeatability of the measurements and to form an opinion of the likely success of the novelty detection in the face of the variability in the fixing conditions. The following discussion will concentrate on the transmissibility between transducers 1 and 2, i.e., along the long axis of the inspection panel. Over all the tests, six normal condition transmissibilities were measured. As stated previously, 2048 spectral lines were recorded over the frequency range 1000–2000 Hz. Figure 6 shows the six normal condition transmissibilities between spectral lines 1250 and 1750. There is a high degree of variability in the results. In order to investigate if this is the result of the fixing conditions, the two transmissibilities obtained when the panel was completely removed were compared and are shown in Figure 7, again between spectral lines 1250–1750. There is far less variability in the plots with the panel removed. This shows that most of the variations in normal condition will be expected from changes in the fixing conditions when the panels are screwed down. The results for the "panel removed" state give an estimate of the sort of changes in the patterns which might be expected purely on the basis of changes in the environmental conditions of the test.

In order to have an effective novelty detection procedure, it is necessary to select features from the transmissibility patterns which distinguish between the faulted and unfaulted states, yet are insensitive to the variations due to the fixing conditions. This will be the aim of the next section: to highlight possible features which, when presented to the outlier analysis algorithm, will result in the various damage states being detected.

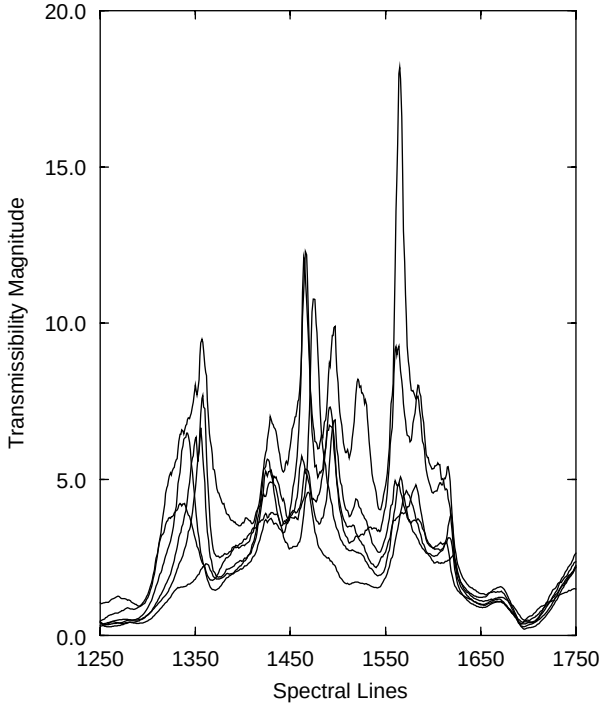

Figure 6. Zoomed transmissibilities T_{12} for the six normal condition states.

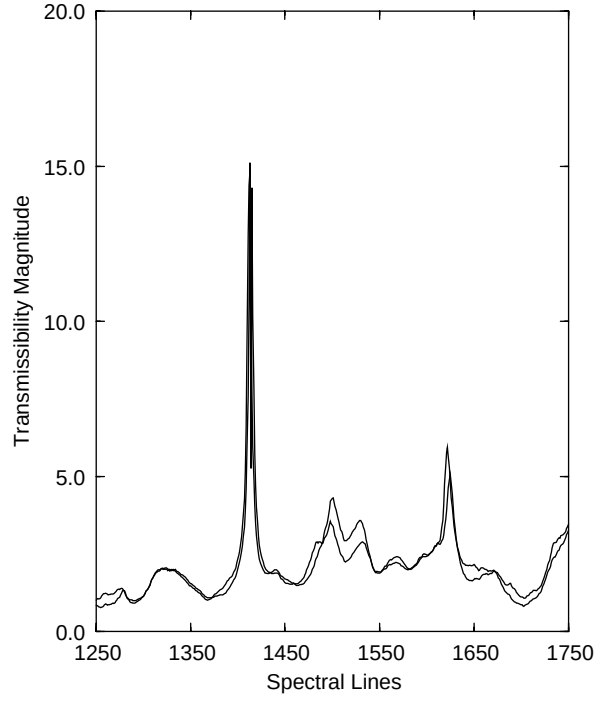

Figure 7. Zoomed transmissibilities T_{12} for the two states when the panels were removed.

4. FEATURE SELECTION FOR NOVELTY DETECTION

The procedure for selecting candidate features for detection of one or more of the damage states was straightforward: each of the 128-average transmissibilities measured from the various damage conditions was compared to the six 128-average transmissibilities measured from the unfaulted structure which, using knowledge gained during the wingbox experiment concerning feature selection, resulted in between one and three areas of interest being highlighted for each of the four main damage types (namely: no panel, holes, width-spanning cuts and length-spanning cuts) for each of the transmissibilities.

Figure 8 shows an example of the features selected from one of the transmissibilities. In this case, the areas of interest used to construct features capable of detecting length-spanning cuts from the T_{12} transmissibilities are shown. These features were used to construct training and testing sets for the outlier analysis algorithm in order to ascertain whether detection of the various damage cases was possible. Figure 9 shows a zoomed version of one of the features to give some idea of the differences that were observed between unfaulted and faulted patterns.

In total, 10 areas of interest were highlighted from the T_{12} transmissibilities and eight from the T_{34} transmissibilities. In the third paper in this series [2], features will be visually classified as being strong, fair or weak due to the large number of potential features highlighted: this is not the case in this paper, where the more limited analysis means that all candidate features can be examined.

It is a simple matter to convert these areas of interest into feature patterns: the transmissibility function is simply sub-sampled over the required region to give an array of 50 sample points or a 50-dimensional pattern in multivariate statistics terminology.

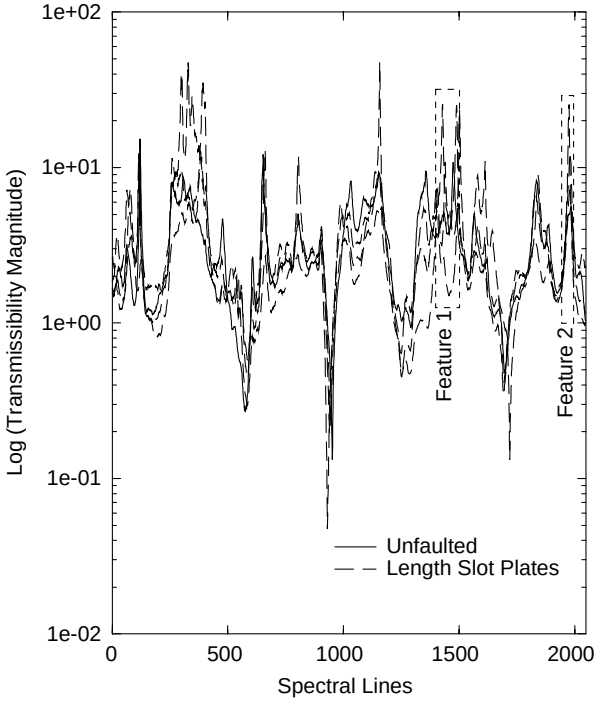

Figure 8. Areas of interest to be used to construct features from the measured transmissibilities T_{12} to detect the length-spanning cuts.

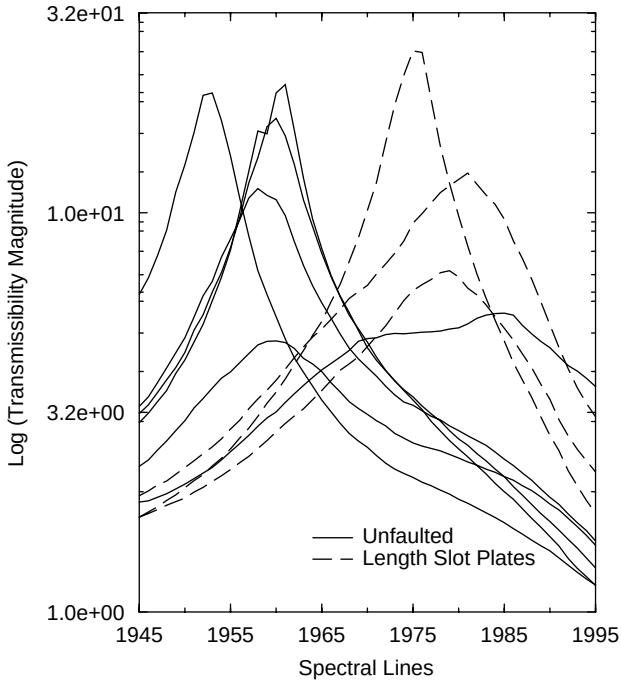

Figure 9. Novelty detection feature from T_{12} for the six undamaged states and the three length slotted panels states (feature 2 from Figure 8).

5. NOVELTY DETECTION RESULTS

All the results presented in this section have been obtained using an outlier analysis algorithm. Details of the method and a description of how threshold levels are calculated are given in reference [3]. For a more in-depth analysis, readers are referred to reference [4].

Basically, a discordant outlier in a data set is an observation that is surprisingly different to the rest of the data set and is therefore believed to be generated by some alternate mechanism; in the present work, this is expected to be a damage mechanism. The main requirement for performing an outlier analysis is a sufficiently large set of training data representing the normal condition of the structure. Once this condition is satisfied, a set of testing data, which may or may not have been recorded from a damaged structure, can be compared with the training data and judged to be statistically likely or unlikely to have been generated by the same mechanism.

In section 4 the procedure for feature selection was discussed and 10 features from the transmissibility T_{12} were highlighted as candidates for detecting one or more of the main damage types. Eight features were highlighted from the transmissibility T_{34}. Individual training sets were constructed for each of these 18 features.

Initially, it was hoped that these training sets would consist of 100 50-dimensional patterns taken from the first 100 of the 110 1-average transmissibilities recorded using the undamaged panel during the first series of tests detailed in section 2. However, upon comparison of the six 128-average transmissibilities recorded using the unfaulted panel, as discussed in section 3, it became clear that repeatability of measurements due to fixing conditions was a major issue. It was decided that, in order to obtain a more complete picture of the unfaulted condition of the structure and also, to avoid many false positives

in the analysis, it was necessary to use the four extra 128-average unfaulted transmissibilities to strengthen the training set.

Due to time limitations and the fact that the issue of repeatability was not investigated until the end of the experimental testing, it was not possible to record another 100 1-average transmissibilities for each of these four extra tests. An alternative procedure was adopted: 100 noise patterns were calculated by subtracting the original 'mean' 128-average transmissibility from each of the first 100 original 1-average transmissibilities. These noise patterns were then added to the four extra 128-average transmissibilities thus allowing the possibility of training sets consisting of 500 50-dimensional patterns. This decision was vindicated upon comparison of the results obtained when using the 100 observation training set and those obtained when using the 500 observation set which more completely represents normal condition. Only the 500 observation results will be given here.

The procedure for constructing testing sets was less problematic than the training sets. For each of the same 18 features as described previously, a set of 120 50-dimensional patterns was stored. The order of these 120 patterns was as follows:

Patterns 1–10 (uf): From the unfaulted panel. Last 10 of the 110 signals.

Patterns 11–20 (f1): Ten signals from the panel with the 20 mm diameter hole.

Patterns 21–30 (f2): Ten signals from the panel with the 38 mm diameter hole.

Patterns 31–40 (f3): Ten signals from the panel with the 58 mm diameter hole.

Patterns 41–50 (f4): Ten signals from the panel with the 50 mm width-spanning edge cut.

Patterns 51–60 (f5): Ten signals from the panel with the 50 mm width-spanning internal cut.

Patterns 61–70 (f6): Ten signals from the panel with the 100 mm width-spanning internal cut.

Patterns 71–80 (f7): Ten signals from the panel with the 100 mm length-spanning edge cut.

Patterns 81–90 (f8): Ten signals from the panel with the 100 mm length-spanning internal cut.

Patterns 91–100 (f9): Ten signals from the panel with the 200 mm length-spanning internal cut.

Patterns 101–110 (np): Ten signals with the panel removed.

Patterns 111–120 (uf2): Ten signals with the unfaulted panel reattached.

The terms given in brackets will be used in the tables of results to refer to these sets of patterns.

The ideal feature would be one which, upon application of the novelty detection algorithm to the data, would result in all of patterns 1–10 and 111–120 being classified as inliers and the other 100 patterns being classified as outliers. However, it is accepted that the more likely situation is that certain features will be sensitive to certain types of damage yet relatively insensitive to other types. That said, whichever feature is being used, there is a requirement that all 20 patterns recorded from the undamaged panel be classified as such.

Rather than plot the results of the outlier analysis for all 18 features, the results will be tabulated and only certain interesting results will be plotted. Table 1 shows the results when considering the 10 features relating to transmissibilities T_{12}, while Table 2 shows the results from the eight features relating to the transmissibilities T_{34}. The first column gives the spectral line range associated with each feature whilst the second column gives the type of damage that particular feature was *intended* to detect (where NP means no panel, H signifies panels with holes and W and L mean width and length-slotted panels respectively). The pattern types follow the same order as stated in the above list and use the previously given nomenclature. The Monte Carlo method discussed in reference [3]

TABLE 1

Outlier analysis results for transmissibility T_{12} features using the critical value of 1% *test of discordancy as a threshold value*

Spectral range	Chosen damage type	Number of detections per damage type											
		uf	f1	f2	f3	f4	f5	f6	f7	f8	f9	np	uf2
200–350	NP	2	1	2	10	1	6	0	7	0	10	10	10
1325–1425	NP	3	10	10	10	10	10	10	10	10	10	10	8
1900–2000	NP	0	10	1	10	0	1	10	10	10	10	10	7
950–1050	H	1	9	10	10	10	9	10	0	10	10	1	0
1410–1460	H	2	10	10	10	0	10	10	10	1	10	10	1
1800–1900	H	0	10	10	10	10	10	10	0	10	1	10	0
440–490	W	5	1	1	1	10	0	1	1	0	0	10	0
950–1100	W	0	10	10	10	10	10	10	1	10	10	1	0
1400–1500	L	3	10	10	10	10	10	10	10	10	10	10	1
1945–1995	L	1	10	0	0	1	1	8	10	10	10	0	8

TABLE 2

Outlier analysis results for transmissibility T_{34} features using the critical value of 1% *test of discordancy as a threshold value*

Spectral range	Chosen damage type	Number of detections per damage type											
		uf	f1	f2	f3	f4	f5	f6	f7	f8	f9	np	uf2
600–750	NP	1	6	2	10	1	2	0	8	2	10	10	6
1900–2000	NP	0	9	10	7	10	0	10	10	10	10	10	0
250–350	H	0	2	1	0	5	2	10	4	4	2	10	0
1070–1120	H	1	10	9	10	0	0	8	10	10	10	5	3
1920–1970	W	0	0	8	0	10	0	10	0	0	10	0	1
450–500	L	1	10	10	10	10	10	10	10	10	10	10	10
1050–1100	L	0	10	10	10	0	4	10	10	10	10	6	5
1150–1300	L	0	10	3	0	2	10	10	10	0	10	10	2

was used to calculate threshold values based upon the critical values for the 1% test of discordancy: for a 500 observation, 50-dimensional problem this was found to be 120·36.

The first issue which should be mentioned concerns the size of the training set. Earlier in this section, the problem of repeatability was mentioned and, due to this problem, it was explained that it was necessary to use a 500 observation training set as opposed to a 100 observation one in order to better represent the range of normal condition data and also to avoid the occurrence of a large number of false positives. The 100 observation results are not given here, due to space considerations. However, for the 100-observation training sets in all 18 cases, the outlier analysis incorrectly classified all 10 of the uf2 patterns as outlying the normal condition. This is clearly undesirable: a novelty detection scheme incapable of recognizing unfaulted patterns is of no value. This vindicates the use of the more representative 500-observation training set.

As stated earlier, the ideal feature would be one which would classify all f1–f9 patterns and all np patterns as outliers whilst classifying all uf and uf2 patterns as inliers. This

would give a row in Table 1 or 2 showing 0 10 10 10 10 10 10 10 10 10 10 0; however examination of these tables shows no such row. The next best thing, is to find a number of features which at least identify all 10 patterns of one or more fault type while observing the minimum requirement of inlier classification for all 20 of the unfaulted uf and uf2 patterns. There are only 4 of the 18 features which satisfy these requirements. Two are features constructed from the transmissibilities T_{12}, two are based on the transmissibility T_{34}.

Figures 10–13 show plots of the results of the outlier analysis for these four cases.

Figure 10 gives the results using the feature using the spectral range 1800–1900 of transmissibility T_{12}. This feature was chosen as being likely to identify faults f1–f3, the three holes. Figure 10 shows that this was the case, with all 10 patterns from each of these faults resulting in discordancies well in excess of the 120·36 threshold value (shown as a horizontal dotted line). The correct classification of the uf and uf2 patterns as inliers is also observed with all discordancy values below threshold. Finally, as well as being able to detect the faults for which the feature was chosen, the plot shows that the only faults which were not classified correctly were f7, the length-spanning edge cut, and f9, the longer of the two length-spanning internal cuts.

Figure 11 concerns the feature constructed from the spectral range 950–1100 of transmissibility T_{12} which was chosen as likely to identify faults f4–f6, the width-spanning cuts. These faults were all identified correctly along with the three holes (f1–f3), and the two length-spanning internal cuts (f8 and f9). The only faults incorrectly classified using this feature were again f7, the length-spanning edge cut, and np, the most severe case where the inspection panel is completely removed.

Figure 12 shows the first of two results taken from the transmissibility T_{34}, using the transducers across the width of the inspection panel. The feature used is that from the

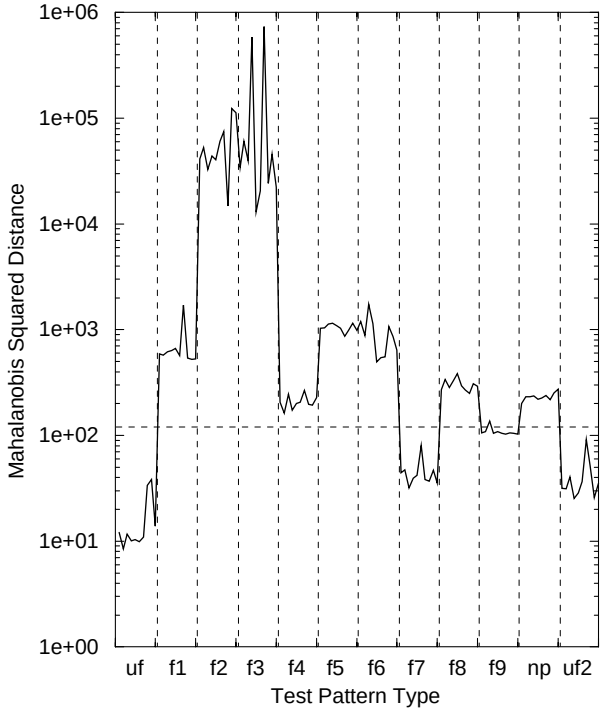

Figure 10. Outlier analysis results for feature from spectral lines 1800 to 1900 of transmissibility T_{12}.

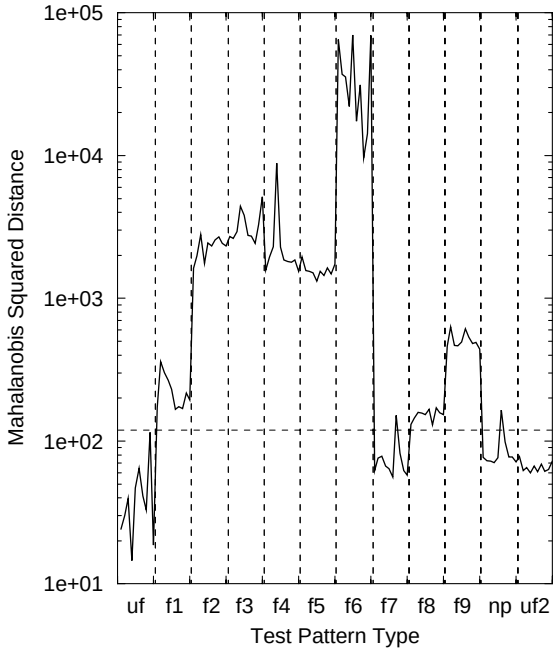

Figure 11. Outlier analysis results for feature from spectral lines 950 to 1100 of T_{12}.

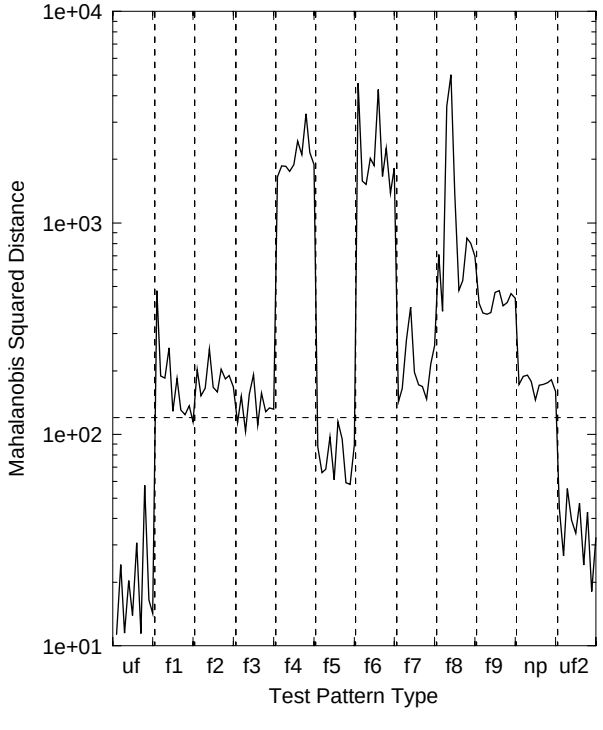

Figure 12. Outlier analysis results for feature from spectral lines 1900 to 2000 of T_{34}.

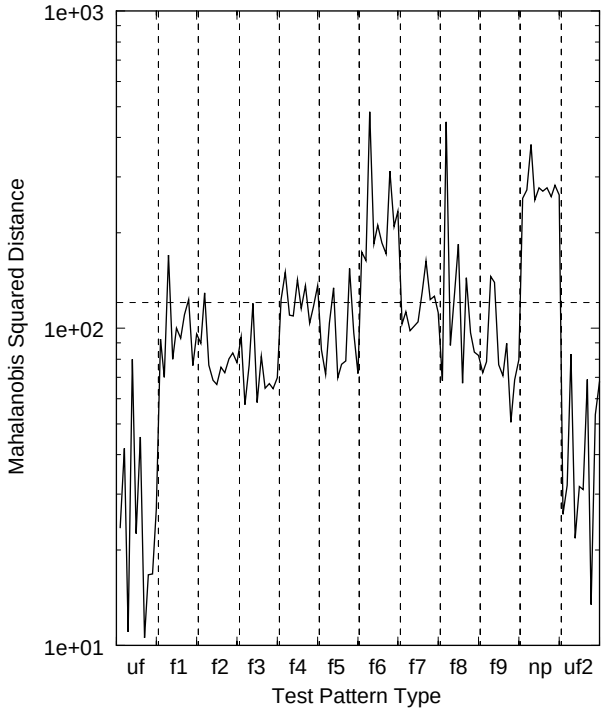

Figure 13. Outlier analysis results for feature from spectral lines 250 to 350 of T_{34}.

spectral range 1900–2000, which was chosen for identification of the np, panel removed, condition. Again, the feature detects the fault for which it was intended, perhaps with slightly lower discordancy values than expected. This minor disappointment is offset by the fact that the feature has correctly classified faults f6–f9, the length-spanning cuts, which caused the majority of problems when using features constructed from transmissibility T_{12}. This makes sense; it might be expected that the transmissibility across the panel width would be more affected by length-spanning cuts than the transmissibility over the panel length and *vice versa*. The analysis also detected faults f2, f4 and f6 correctly but not faults f1, f3 and f5.

The final outlier analysis concerns the feature constructed from the spectral range 250–350 of the T_{34} transmissibilities. This was chosen to identify the panels with holes, f1–f3, but Figure 13 shows that it proved unable to detect these faults. The only faults which this analysis managed to correctly classify were f6, the largest width-spanning cut, and the panel-removed condition.

Examination of these four analyses shows that it is actually possible to completely detect all 10 fault conditions (nine faulted panels and missing panel) by using only two features. These features are those constructed from lines 1800 to 1900 of T_{12} as shown in Figure 10, and those constructed from lines 1900 to 2000 of T_{34} as depicted in Figure 12. This implies that both transmissibilities are required to give detection of all faults.

All threshold values were calculated using the Monte Carlo method outlined in reference [3] based upon the critical value of 1% test of the discordancy. There is however a school of thought [5] that argues that it is perfectly acceptable practice to set a threshold value as being slightly greater than the highest discordancy value of all the testing patterns

taken from the unfaulted structure. Accepting this argument allows Tables 1 and 2 to be adjusted, with the results given in Tables 3 and 4. The new, manually set, threshold values are given in the tables for each feature. The four previously discussed cases keep the same threshold values as it is only deemed correct to *raise* threshold values thus giving a more stringent test than the 1% test.

Tables 3 and 4 show that the most obvious effect of this alteration of the threshold values is a column of zeros in the uf and uf2 columns indicating that the unfaulted patterns would be correctly classified as inliers. An equally obvious effect of this increase in the threshold values is that the method is less sensitive to damage. All numbers in the f1–f9 columns and the np column decrease or stay the same compared to the entries in Table 1 or 2. In some cases, the raising of the threshold level eliminates much of the effect of what initially appeared to be a potentially useful feature. For example, consider the feature constructed from the spectral range 1900–2000 of transmissibility T_{12}. Examination of the relevant row in Table 1 shows that, with a threshold value of 120·36, the analysis could correctly classify faults f1, f3, f6–f9 and np but 7 of the 10 uf2 patterns were wrongly classified as outliers. Raising the threshold level to 600 is sufficient to correctly

TABLE 3

Outlier analysis results for transmissibility T_{12} features using manually set threshold values

Spectral range	Threshold value	Number of detections per damage type											
		uf	f1	f2	f3	f4	f5	f6	f7	f8	f9	np	uf2
200–350	1000	0	0	0	0	0	0	0	0	0	0	10	0
1325–1425	600	0	8	10	10	10	10	1	10	10	6	10	0
1900–2000	600	0	4	0	0	0	0	0	0	3	10	10	0
950–1050	250	0	2	10	10	10	0	10	0	1	10	0	0
1410–1460	160	0	10	9	10	0	10	10	10	0	10	10	0
1800–1900	120	0	10	10	10	10	10	10	0	10	1	10	0
440–490	300	0	0	0	0	10	0	0	0	0	0	10	0
950–1100	120	0	10	10	10	10	10	10	1	10	10	1	0
1400–1500	200	0	10	10	10	10	10	10	10	10	10	10	0
1945–1995	300	0	3	0	0	0	0	0	0	0	10	0	0

TABLE 4

Outlier analysis results for transmissibility T_{34} features using manually set threshold values

Spectral range	Threshold value	Number of detections per damage type											
		uf	f1	f2	f3	f4	f5	f6	f7	f8	f9	np	uf2
600–750	220	0	0	0	4	1	1	0	0	0	10	10	0
1900–2000	120	0	9	10	7	10	0	10	10	10	10	10	0
250–350	120	0	2	1	0	5	2	10	4	4	2	10	0
1070–1120	180	0	10	6	10	0	0	3	10	10	10	2	0
1920–1970	185	0	0	2	0	10	0	10	0	0	9	0	0
450–500	2500	0	10	0	1	1	0	1	10	10	0	0	0
1050–1100	170	0	10	10	10	0	0	9	10	10	10	5	0
1150–1300	145	0	10	1	0	2	8	10	10	0	10	10	0

classify all uf2 patterns as inliers; however, examination of the relevant row of Table 3 shows the effect of this action upon fault classification. With the raised threshold, only the patterns from faults f9 and np are now completely classified as outliers with those from faults f1, f3, f6–f8 being only partially classified as damage.

The detrimental effect on the features is not universal. There are three cases where the raising of the threshold value does not have a very destructive effect on the fault detection. Consider the feature constructed in the range 1410–1460 of transmissibility T_{12} with the threshold value raised to 160 (sufficient to cause all unfaulted patterns to be classed as inliers). Even with the higher threshold, the analysis has managed to correctly classify all patterns from faults f1, f3, f5, f6, f7, f9 and np. The only degradation in performance is that only nine patterns of fault f2 are correctly identified, compared to 10 with the Monte Carlo threshold.

Figure 14 shows the outlier analysis results for the feature constructed from 1400 to 1500 range of transmissibility T_{12} with the threshold value raised to 200. This has produced a *perfect* result; the feature can correctly classify each of the 120 patterns as either inlier or outlier.

The final outlier analysis of interest concerns the T_{34}. The threshold level was raised to 170 for the feature constructed from 1050 to 1100 spectral range. The new analysis correctly classifies all patterns from faults f1, f2, f3, f7–f9 (the holes and the length-spanning cuts) whilst detecting nine of the 10 f6 patterns. (All of the f6 patterns were correctly classified using the lower threshold.)

The final plot, Figure 15, shows a visualization via principal component analysis of the 500-observation normal condition set from one of the features (in this case, the feature constructed from the 1900 to 2000 spectral range from T_{34}). Note that the data form five

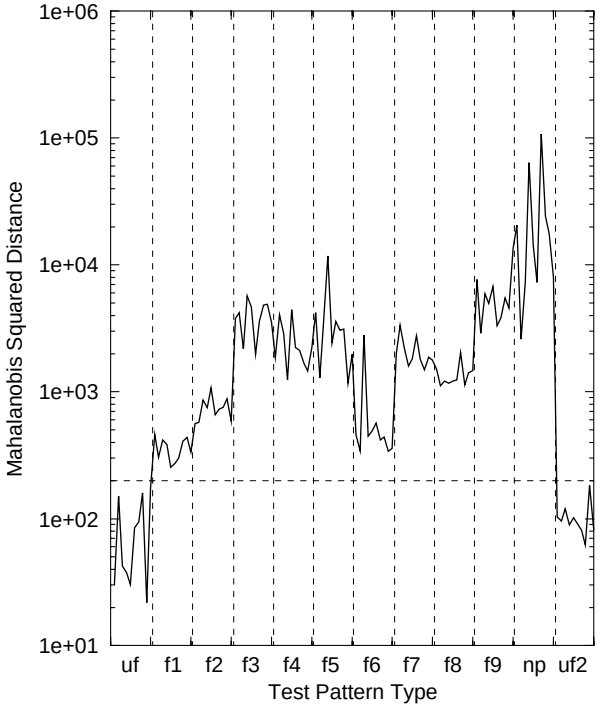

Figure 14. Outlier analysis results for feature from spectral lines 1400 to 1500 of T_{12} with manually set threshold value of 200.

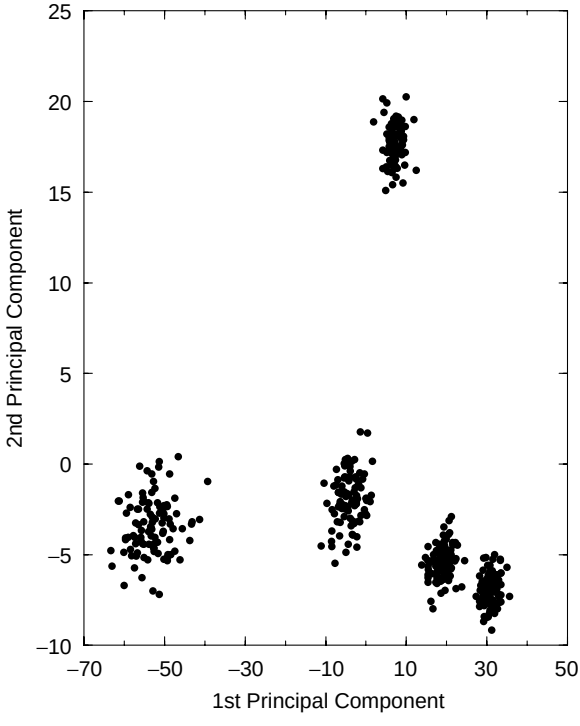

Figure 15. PCA visualization of normal condition set for feature constructed from spectral lines 1900 to 2000 of T_{34}.

distinct clusters as opposed to one uniform set; the consequences of this fact will be discussed in the following section.

6. DISCUSSION AND CONCLUSIONS

The overall conclusions from this piece of work are very positive. A single measurement feature was obtained which served to separate all the fault cases from normal condition. This is particularly significant given the wide-ranging normal condition data obtained in the, necessarily restricted, testing programme. However, the feature selection process was not trivial and a number of points worthy of discussion were raised.

First, the feature selection procedure made use of fault data in order to separate out intervals from the transmissibility functions which distinguished damaged states from normal. This is inconsistent with a true novelty detection procedure which would *only* use normal condition data to train the diagnostic. This is not necessarily a cause for concern here, the main problem is that the diagnostic has been optimized to detect damage only of the type inflicted in the test programme. It is not guaranteed to detect flaws which are at different positions or even have different orientations from the training sample, although some generalization is likely. The main problem is that, in general, it will not be possible to damage the structure in order to obtain training data as was done here. This means that either true novelty detection is needed or some means of simulating damage is found. If the former approach is used, it may be possible to form a set of features centred on *all* the significant peaks in the transmissibility (and all troughs which are peaks in the reciprocal transmissibility). In this case, a vector of novelty indices will be obtained and monitored,

any deviation from normality on any index will be taken to signal damage. If the latter approach is selected there are again two possible schemes. A model-based method could be pursued in which a FE model of the structure is used. This would be updated and validated on the normal state of the structure and the damage could then be simulated within the model. The problem with this would be the accurate modelling of a potentially complex structure. In contrast, an experimental approach could be adopted. This would be based on the assumption that damage will normally manifest itself by a local reduction in stiffness with a consequent change in resonant frequencies. One might hypothesize that similar reductions in frequency could be obtained via local *increases* in mass. Feature selection could then be performed by (non-destructively) adding mass to the structure and observing which resonances in the transmissibility are sensitive. An initial study [6] based on the experimental wingbox used in reference [1] has provided some justification for this approach.

A second issue raised by the study concerns the selection of the novelty threshold. The high degree of variability in the fixing conditions of the panel here meant that the coverage of the normal condition set obtained during the testing programme was rather inadequate; this is testified by the principal component visualization of Figure 19 which showed that the five normal condition clusters were actually disjoint. (A further assumption of the approach was that the noise colour and size from one normal cluster could be imposed with impunity on the other clusters.) The immediate consequence of this is that the index values on the normal *testing* set were not guaranteed to be sub-threshold. In fact, when the threshold was chosen by the usual assumption of normality of the index distribution, many supra-threshold excursions were observed over the various features measured. This problem was overcome by changing the means of prescribing the threshold. Tarassenko's approach to threshold setting was modified as follows: instead of setting the threshold as the highest value observed on the training set, it was fixed at the highest value observed on a (normal) validation set. (Because of the limited test programme, the normal testing set was used here; however, with the benefit of hindsight this is no real cause for concern.)

As always, further work is required. The issues discussed above will be addressed via further laboratory studies. An important consideration which was ignored here concerns variability in the normal condition due to the *loading* and *environmental* conditions. The consequences of variability in boundary conditions are marked and one would expect the other sources of variation to be similarly significant.

ACKNOWLEDGMENTS

The foremost acknowledgement is to DERA, Mechanical Sciences Sector, Aero-Structures Department for funding this project. Special thanks are due to Drs Graham Skingle and Terry Griffiths who gave extremely valuable technical support during the testing phase at Farnborough.

The authors would also like to thank Mr David Webster, Mr Chris Lewis, Mr David Nuttall and Mr Les Morton for constructing the inspection panel copies and lending indispensable technical support.

REFERENCES

1. K. Worden, G. Manson and D. J. ALLMAN 2003 *Journal of Sound and Vibration* **259**, 323–343. Experimental validation of a structural health monitoring methodology. Part I: novelty detection on a laboratory structure.

2. G. Manson, K. Worden and D. J. Allman 2003 *Journal of Sound and Vibration* **259**, 365–385. Experimental validation of a structural health monitoring methodology. Part III: damage location on an aircraft wing.
3. K. Worden, G. Manson and N. R. Fieller 1999 *Journal of Sound and Vibration* **229**, 647–667. Damage detection using outlier analysis.
4. V. Barnett and T. Lewis 1994 *Outliers in Statistical Data*, Chichester: John Wiley and Sons; third edition.
5. A. Nairac, T. A. Corbett-clark, R. Ripley N. W. Townsend and L. Tarassenko 1997 *Proceedings of the Fifth International Conference on Artificial Neural Networks, Cambridge, U.K.* Choosing an appropriate model for novelty detection.
6. K. Worden, L. Y. Cheung and J. A. Rongong 2001 *Proceedings of* 19*th International Modal Analysis Conference*, *Orlando*, *FL*, 1234–1241. Damage detection in an aircraft component model.