Journal of Sound and Vibration (2003) **259**(2), 365–385
doi:10.1006/jsvi.2002.5169, available online at http://www.idealibrary.com on IDEAL®

EXPERIMENTAL VALIDATION OF A STRUCTURAL HEALTH MONITORING METHODOLOGY: PART III. DAMAGE LOCATION ON AN AIRCRAFT WING

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 documents the third phase of a programme of experimental work which validates a structural health monitoring methodology based on novelty detection. In this paper, an extension of the detection method for damage location is proposed and demonstrated. The structure of interest is a Gnat aircraft wing. Although it was not possible to damage the aircraft, the method was demonstrated by determining which of a set of inspection panels had been removed.

1. INTRODUCTION

The problem of on-line structural health monitoring (SHM) of aircraft is a very difficult one. However, the motivation for pursuing research in the field is strong; integrated SHM technology may well allow significant reductions in the cost of ownership of aircraft, both military and civil. There is no shortage of research activity in the field; however, progress is slow and incremental and the effort tends to be concentrated on computer simulations and relatively simple laboratory structures.

At the risk of repeating material from earlier papers [1], it is useful to think of the SHM problem in terms of a hierarchical structure, perhaps the most well-known framework is that of Rytter [2].

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.

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.

† Now QinetiQ.

0022-460X/02/$35.00

Each level is more difficult than the last. Although results have been presented supporting the existence of methods working up to level 3, they are usually based on systems and structures of low complexity and are far from convincing aerospace industry of their applicability in the field. Although level 4 is most difficult of all, it is paradoxically the case that some prediction methods have actually been implemented. For example, operational load monitoring (OLM) has been used as a tool for prognosis; measured strains from an airframe can be used to estimate the fatigue life consumption for a given mission and the residual life of the structure can be estimated. These prediction methods are rather blunt instruments however, as they do not—cannot in fact—use diagnostic information from lower levels. A true level 4 system would use information about the location and extent of the damage in forming its prognosis.

The philosophy of the programme of work to which this paper belongs is to develop methods of addressing the SHM problem at each level which are robust enough to survive experimental validation on real aircraft structures. This has entailed a reevaluation of damage identification methods and a return to basic detection (level 1) methods.

The first stage involved the development and benchmarking of methods of *novelty detection* [3, 4], i.e., the problem of distinguishing between the normal operation of a system and any anomalous conditions which may be symptomatic of damage. The methods investigated included a kernel density estimation (KDE) approach and an artificial neural network (ANN).

The first phase of experimental work [1] began the verification of the novelty detection method. The question of what to measure and how to convert data into features suitable for diagnosis arose at this point, so a simplified structure was adopted. Tests were carried out on 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 methods were successful in signalling all the damage states where the depth of the cut exceeded 5 mm. The features used for the analysis were transmissibilities measured along the stringer. A further analysis using modal strain energy methods proved capable of locating the damage along the stringer [5]. However, it was recognized that the location method was rather insensitive and would not generalize well to location in a two- or three-dimensional structure. Also as part of this work, a new novelty detection approach based on multivariate statistics was introduced. The method—outlier analysis—proved to be the most robust of the three methods used and was carried forward to the next stage of the work.

The second phase of the experimental research was a successful attempt to diagnose local damage on a full-scale structure in this case a Gnat aircraft [6]. Damages of various forms were introduced into copies of a wing inspection panel and, as before, transmissibility measurements across the panels were used to try and detect the damages using outlier analysis. In all cases, it proved possible to detect the flaws. However, the exercise was non-trivial and raised important points concerning the problem of detection in varying operational or environmental conditions.

The conclusion drawn from the above work was that damage detection based on novelty analysis is a feasible means of detecting damage on real structures and to a large extent addresses the problem of SHM up to level 1. Note that this is true only for the context addressed i.e., for vibration-based features it is possible to detect damage of a certain type and extent. This extent is large compared with, for example, a subcritical fatigue crack, and the methods will almost certainly fail if this degree of resolution is required. However, it is unlikely that any global monitoring system will be capable of such high resolution. It is almost certain that a working SHM system for large structures must

be a fusion of a global system looking for moderate to severe damage in largely damage tolerant areas, with a set of local monitoring systems which aim to diagnose subcritical faults in significant structural items. The vibration-based system adopted in this study should be considered as a candidate for the former.

The objective of the latest phase of the programme was to move from detection to location (level 2). This was to be achieved by extending the vibration-based approach. A network of sensors was used to establish a set of novelty detectors, each sensitive to different regions of the wing. Once the relevant features for each detector had been identified and extracted, a neural network was used to interpret the resulting set of novelty indices. Because it was not possible to damage the wing, faults were simulated by sequentially removing inspection plates—of which there were nine. Providing a means of simulating damage, this strategy also converted the location problem naturally into a discrete classification problem. Because the panels were different sizes, the analysis gave some insight into the sensitivity of the method, i.e., what was likely to be the minimum detectable damage.

The layout of this paper is as follows. Section 2 describes the experimental layout and strategy and section 3 describes how features were chosen for novelty detection. Section 4 discusses how the novelty indices are interpreted by a neural network in order to locate the damage; the results of the analysis are presented and the report is completed with discussion and conclusions.

2. TEST SET-UP AND DATA CAPTURE

As in the previous paper in this series [6], the structure under investigation was a Gnat trainer aircraft; the object of the exercise being to locate damage in the starboard wing (Figure 1). The first problem encountered was that it was not possible to damage the aircraft. This problem was overcome by noting the presence of numerous inspection panels distributed over the wing. It was decided to simulate damage by sequentially removing panels; this also had the distinct advantage that each damage scenario was reversible and it would be possible therefore to monitor the repeatability of the measurements.

Of the various panels available, nine were chosen, mainly for their ease of removal and also to cover a range of sizes. These were distributed as shown in Figure 2. Figure 2 is not drawn to scale, it is a schematic showing a rough estimate of the relative sizes of the panels.

The areas of the panels P1–P9 are given in Table 1. Removal of any of these panels actually constitutes a rather large damage. Panels P3 and P6 are likely to give trouble as they are by far the smallest.

Each panel was fixed to the wing by a number of screws, the numbers varying between 8 and 26. On some of the panels, screws were missing as a result of damaged threads in the holes. In fact during the repeated removal of the plates during testing, further holes were damaged. This meant that there was some variation throughout the test even for normal condition (all panels attached). However, given experience gained during the previous experimental phase of the programme [6], it was assumed that this was unlikely to be a source of major variation, compared to the uncertainty in the fixing conditions of the remaining screws. The screws were secured and removed with an electric screwdriver with a controllable torque; the same torque setting was used throughout.

As in the previous studies, it was decided to use transmissibilities for the base measurements. This had proved effective in the novelty detection phase of the work,

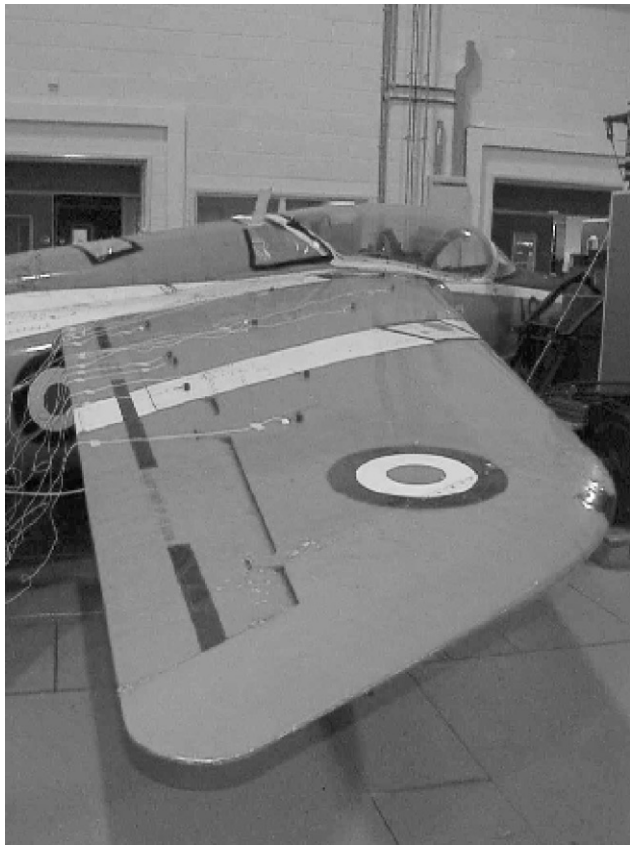

Figure 1. Starboard wing of Gnat trainer.

and was expected to extend naturally to the location problem. This decision was supported by a study of damage location in a plate based on measured transmissibilities [7]. The initial decision was to measure a transmissibility across each plate. In order to use transducers effectively, the panels were split into three groups A, B and C. Each group was allocated a centrally placed reference transducer, together with three other transducers, each associated with a specific plate. The transducer layout is shown in Figure 3.

Although this network of sensors offers the possibility of forming many transmissibilities, only those measured directly across the plates were used in this study. The transmissibility across plate P*n* was denoted T*n*. Later when feature selection is discussed, it will be necessary to label the reciprocals of these transmissibilities and these will be denoted T*n**. Table 2 summarizes how the measurements were formed. The sensors used were standard piezoelectric accelerometers of the PCB type.

The wing was excited using a Ling electrodynamic shaker attached directly below the panel P4 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. In all cases 1024 spectral lines were measured; both real and imaginary parts of the functions were

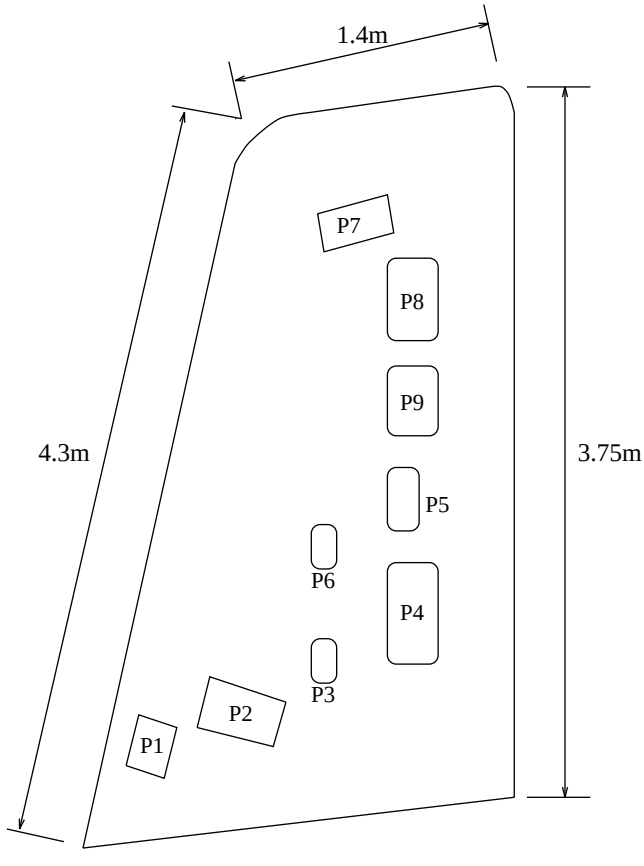

Figure 2. Schematic of starboard wing inspection panels.

TABLE 1

Area of inspection panels

Panel	Area (m^2)
1	0·0221
2	0·0496
3	0·00825
4	0·08
5	0·0176
6	0·00825
7	0·0392
8	0·0468
9	0·0234

obtained. As the smaller panels were of the same order of size as the major defects in the previous novelty detection study [6], it was assumed necessary to excite modes with the same order of wavelengths and hence frequencies. The transmissibilities were therefore measured between 1024 and 2048 Hz.

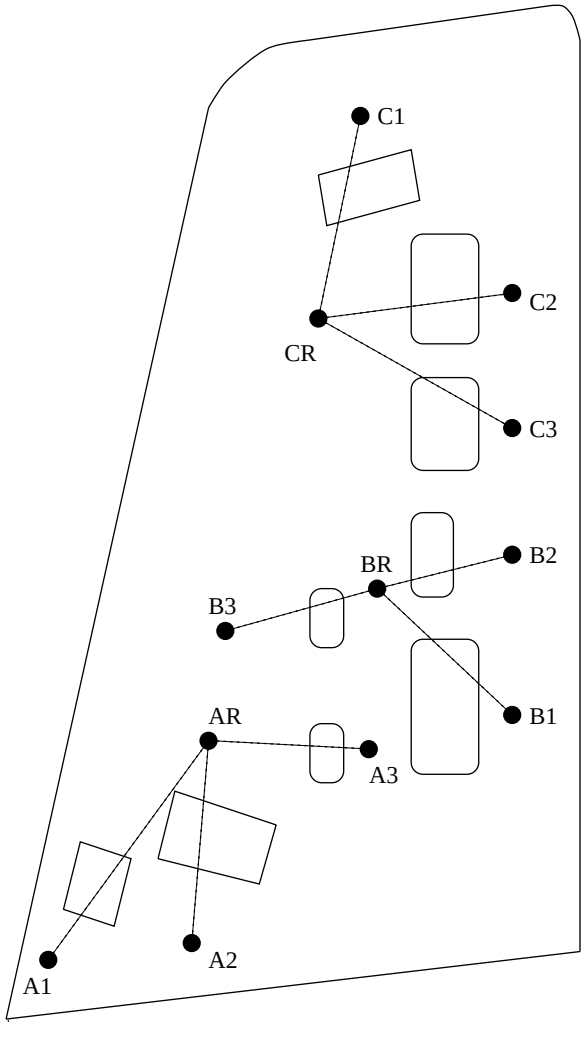

Figure 3. Schematic of transducer locations.

TABLE 2

Transducers used to form transmissibilities

Plate	Associated transmissibility	Reference transducer	Response transducer
P1	T1	AR	A1
P2	T2	AR	A2
P3	T3	AR	A3
P4	T4	BR	B1
P5	T5	BR	B2
P6	T6	BR	B3
P7	T7	CR	C1
P8	T8	CR	C2
P9	T9	CR	C3

In order to make use of the Single-input-multiple-output (SIMO) mode of the software, each configuration of the wing was tested three times, once each for transducer groups A, B and C. In all, 25 configurations were taken as follows:

1. Normal condition (all plates in place).
2. Plate P1 removed.
3. Plate P2 removed.
4. Plate P3 removed.
5. Normal condition.
6. Plate P1 removed.
7. Plate P2 removed.
8. Plate P3 removed.
9. Normal condition.
10. Plate P4 removed.
11. Plate P5 removed.
12. Plate P6 removed.
13. Normal condition.
14. Plate P4 removed.
15. Plate P5 removed.
16. Plate P6 removed.
17. Normal condition.
18. Plate P7 removed.
19. Plate P8 removed.
20. Plate P9 removed.
21. Normal condition.
22. Plate P7 removed.
23. Plate P8 removed.
24. Plate P9 removed.
25. Normal condition.

This programme meant that seven sets of measurements for normal condition were made and two each for each damage state. This was done in order to investigate variability of the normal and damaged condition data between plate removals.

The measurement strategy for each transmissibility was as follows:

First, each function was obtained using 16 averages. This was done to provide a clean reference signal to help with feature selection. In the previous paper, 128 averages were used, however, given the number of tests to be performed here this was regarded as too time-consuming. A brief experiment showed that the 16-average transmissibilities were adequately smooth. Next, 100 measurements were taken sequentially using only 1 average. Over the full sequence of 25 configurations, this gave 700 one-shot measurements for the normal condition and 200 for each damage condition.

All the data were archived on the HP workstation and transferred back to Sheffield. Each set of test results was extracted into an LMS universal file and then the transmissibilities were separated into individual ASCII files using a C program. At this stage the real and imaginary parts of the transmissibilities were converted to magnitudes and the phases were discarded.

3. FEATURE SELECTION AND NOVELTY DETECTION

The first stage in the analysis was to establish which features could be used to individually detect damage in the plates. The detection algorithm used was the outlier

analysis procedure described and validated in references [1, 6]. Here it suffices to say that a model of normality is constructed from features or patterns taken from data when the wing is undamaged. Subsequent data are evaluated with respect to this model and any significant departures are assumed to flag damage. As before, the source of the features/patterns are the transmissibility functions measured in the experimental phase of this work. For the purposes of this work, a feature is a region of the transmissibility which separates unambiguously the normal condition data from the damage data. Examples will follow shortly, however first, a little terminology will be introduced. This phase of t he work entailed the examination of many transmissibilities and the number of candidate features was very large. In order to separate out the best, in as objective a manner as possible, features were classified as *weak*, *fair* or *strong* according to the following criteria:

- A *strong* feature is a region of the frequency range on which the normal data and damage data appear to be structurally different. Also, the damage data should be strongly separated compared to the spread of the normal data.
- A *fair* feature is a region of the frequency range on which the normal data and damage data appear to be structurally different *or* the damage data are strongly separated compared to the spread of the normal data.
- A *weak* feature is a region of the frequency range on which the normal data and damage data are separated.

Figure 4 shows an example of a feature which was judged strong. Note that all the solid lines—which represent samples from the seven normal condition transmissibilities are tightly clustered. The damage data is not only well separated from the normal cluster, but shows a peaked structure which is distinct from the normal data.

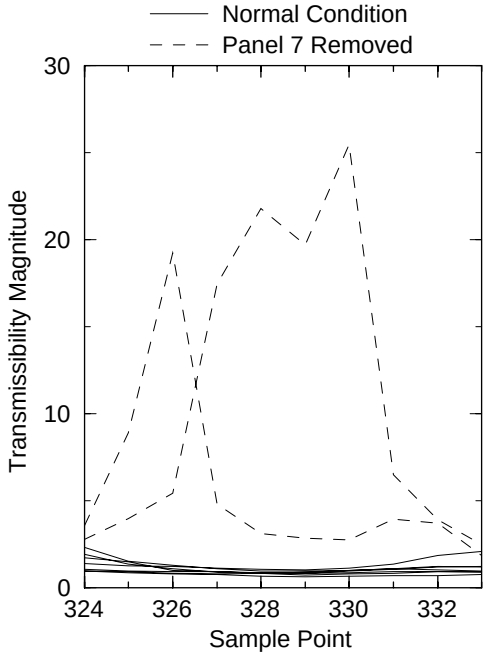

Figure 4. An example of a strong feature from one of the transmissibility functions.

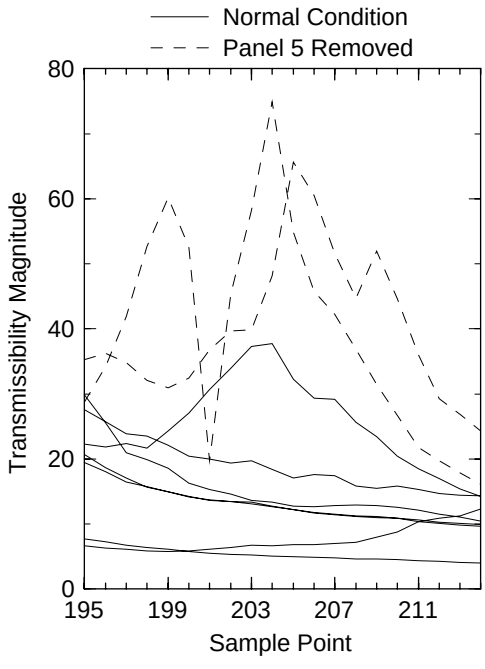

Figure 5. An example of a fair feature from one of the transmissibility functions.

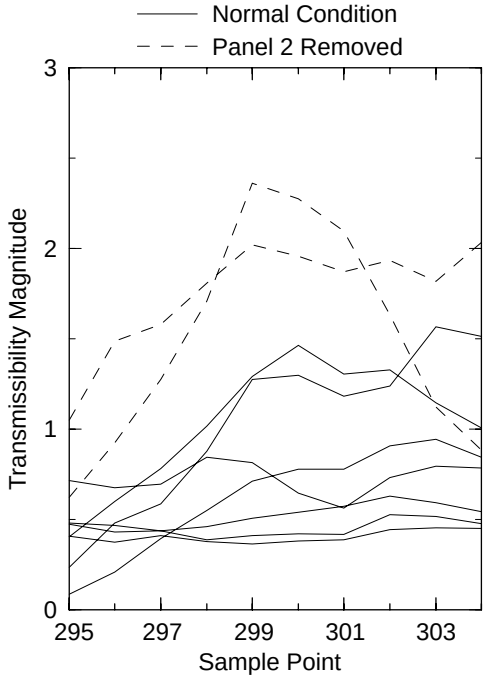

Figure 6. An example of a weak feature from one of the transmissibility functions.

Figure 5 shows a pattern which was judged to be a fair feature. Apart from one of the normal condition patterns the damaged patterns show a distinct peaked structure. However, the damaged patterns are not separated well from the normal patterns by the standards of their own internal spacing. Figure 6 shows a weak feature; the normal and damaged patterns are not structurally distinct, both have a peaked nature. Also, the damaged patterns are not separated strongly from the normals by the standards of the variation in the normal patterns. However, this still has potential as a feature, because the damage patterns are at least outlying the normal cluster (over most of the range).

Having established criteria for less subjective judgement of feature fitness, the next stage of the analysis was an exhaustive visual search through the transmissibility database for features for novelty detection. In order to simplify matters, the transducer groupings were used as follows. Only transmissibilities T1–T3 were examined for sensitivity to the removal of Panels P1–P3. Similarly, T4–T6 were examined for features to detect removal of P4–P6. T7–T9 were examined for features for P7–P9 removal. The selection of features was carried out by eye on the 16-averaged transmissibilities. The inverses of the transmissibilities T1*–T9* were also checked.

Once the candidate features had been identified, they were each evaluated for effectiveness in the outlier analysis. Two main criteria were applied. In the first case, each point in the normal condition data was tested for discordancy using an inclusive outlier statistic [1]. The number of points falsely identified as outliers was recorded. Secondly, the normal condition data were used to form statistics for an exclusive outlier analysis on the damaged data points (one for each repetition of a panel removal). The number of points falsely recorded as inliers was noted. The results of this analysis are given in Tables 3–5 for a representative sample of the panels. For each feature the following data are given: the transmissibilty of origin, the range of spectral lines, the number of inliers on the normal data. In what follows, for each repetition of a plate removal: number of outliers on the damage set, average discordancy normalized according to the threshold (exclusive). Finally, the strength of the feature is given. Features marked with a star are taken from the inverse transmissibility.

Note that in each case above, the features were constructed only on the basis of detecting the removal of *a specific panel*. Following the analysis, the best feature for each panel was selected using the following criteria. If possible a feature was chosen which had no false negatives for damage detection, i.e., had the number 100 in columns 5 and 7 above. If more than one feature satisfied this condition, the one with the least false positives on the normal condition was chosen, i.e., the feature with the highest score in column 3. If it was not possible to find a feature with no false negatives for damage, the selection process was a little more subjective and the average discordancy figures for the feature were evaluated also. The features marked with *** in the tables are those

TABLE 3

Results of outlier analysis of potential features to diagnose Panel 2 *missing*

Trans. number	Spectral line range	Normal inliers	Rep. 1 ratio	Rep. 1 outliers	Rep. 2 ratio	Rep. 2 outliers	Feature strength
1	39–48*	677	13·908183	98	5·058033	62	W
1	845–874*	645	84·220006	100	35·977528	100	S***
2	295–304	666	2·047300	66	1·428106	44	W+

TABLE 4

Results of outlier analysis of potential features to diagnose Panel 5 missing

Trans. number	Spectral line range	Normal inliers	Rep. 1 ratio	Rep. 1 outliers	Rep. 2 ratio	Rep. 2 outliers	Feature strength
4	194–213	640	369·207091	100	81·124033	100	F/W
4	586–605	659	21·880078	100	4·484882	100	F/S
4	504–513*	659	13·500478	96	21·519737	100	W
5	195–214	639	375·436938	100	92·861504	100	F++
5	263–272	673	0·840937	6	0·361309	0	F
5	590–609	662	6·881952	100	3·270125	100	F/S
5	38–57*	662	75·757640	100	59·564716	100	S
5	352–361*	658	2·314719	53	2·837032	70	F/W
5	419–428*	677	0·506041	7	0·616805	9	W
5	438–447*	642	2·968112	52	2·829786	63	F/W
5	770–789*	664	9·740520	100	52·485135	100	F/S
5	825–834*	679	130·349070	100	38·344141	100	S***
6	198–207	662	597·018268	100	22·167515	98	F
6	0–9*	680	5·150848	99	23·862631	100	F/S
6	26–35	666*	17·577650	100	10·596527	100	F/S
6	52–71	654*	5·643725	100	3·231036	100	S
6	275–294*	656	53·172082	100	25·132314	100	F/S
6	622–631*	681	5·137724	88	5·208337	98	F
6	698–727*	633	6·413629	100	4·027143	100	F/S
6	745–754*	676	5·767129	99	10·016291	98	F
6	866–875*	671	1·324489	31	0·711700	12	W

TABLE 5

Results of outlier analysis of potential features to diagnose Panel 7 missing

Trans. number	Spectral line range	Normal inliers	Rep. 1 ratio	Rep. 1 outliers	Rep. 2 ratio	Rep. 2 outliers	Feature strength
7	14–23	680	1·179119	82	1·795337	99	F/W
7	108–117	671	3·209194	100	7·122971	100	F
7	37–56*	666	30·177145	100	14·435880	100	F
7	324–333*	673	99·250716	100	559·516509	100	S+++
7	454–463*	681	17·107345	100	2·723250	100	F
7	613–642*	661	40·775171	100	47·583174	100	F
7	757–766*	681	637·000692	100	1357·539606	100	S***

selected for the novelty detectors. The features marked with +, ++ and +++ in the tables are those which were selected earlier to illustrate the idea of weak, fair and strong features.

Once the features for each novelty detector had been selected, each was analyzed using principal component analysis. In order to visualize the feature and see how well separated the normal and damaged data were, each normal and damage set were projected onto the first two principal components [8]. Figure 7 shows the separation of the normal and damage clusters for the feature chosen to signal the removal of panel 1. This was a strong feature and this is reflected in the excellent separation of the clusters. In contrast, Figure 8 shows the corresponding plot for the panel 3 feature. This was one of the two smallest

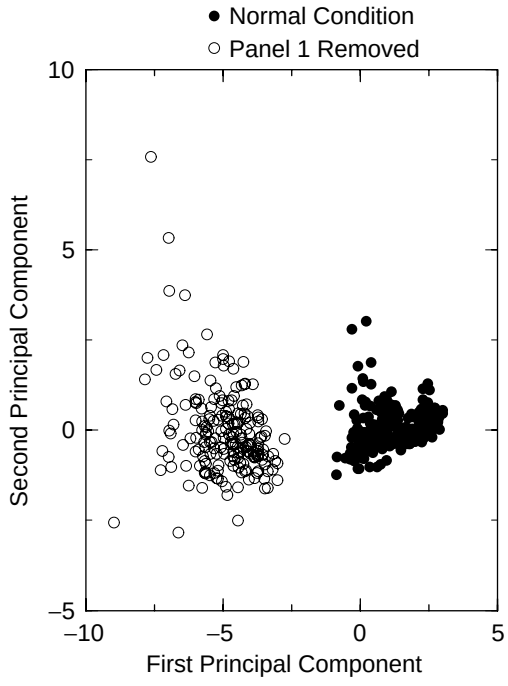

Figure 7. PCA visualization of feature for panel 1 removal.

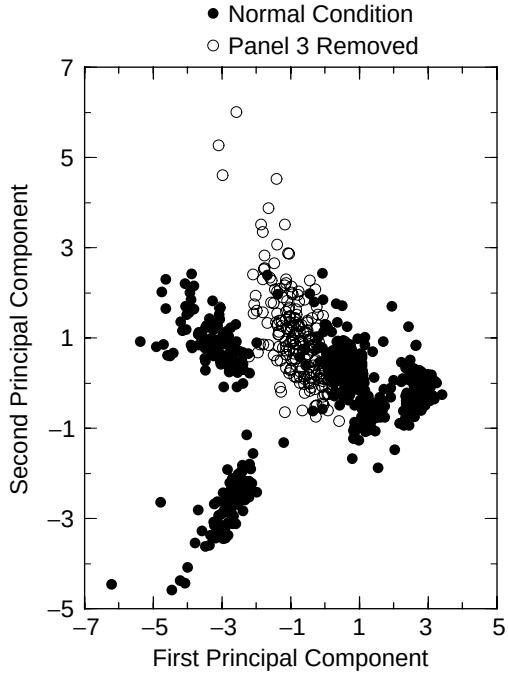

Figure 8. PCA visualization of feature for panel 3 removal.

panels and it was only possible to find a fair feature which picked up 31% of the damage cases. This uninspiring performance is reflected in Figure 8, where it can be seen that the normal and damage clusters overlap. In fact, all the PCA decompositions except two show a good separation between the normal and damage data. The exceptions are those for the removal of panels 3 and 6—the smallest two. Panel 6 gave somewhat better results than

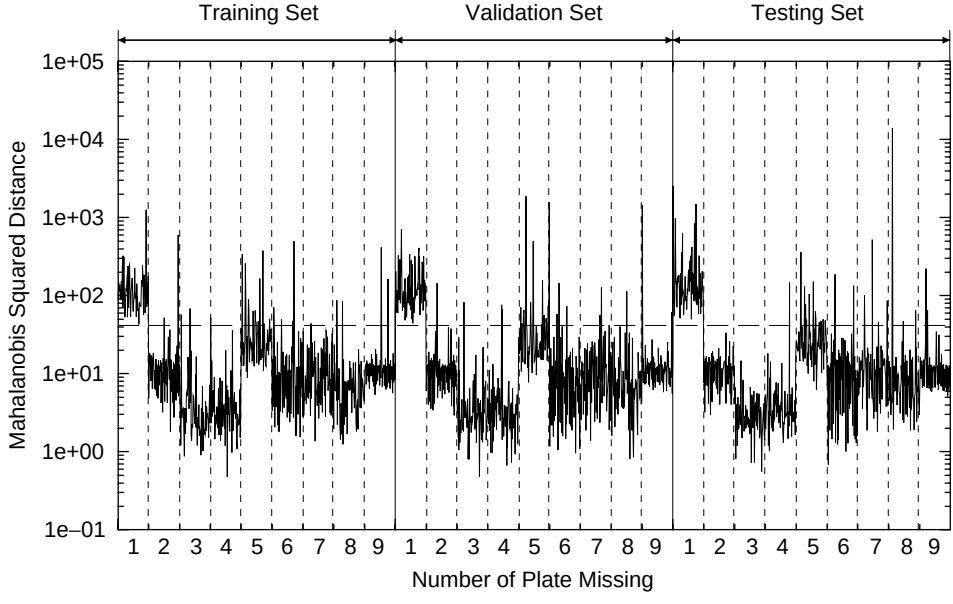

Figure 9. Outlier statistic for all damage states for the novelty detector trained to recognize panel 1 removal.

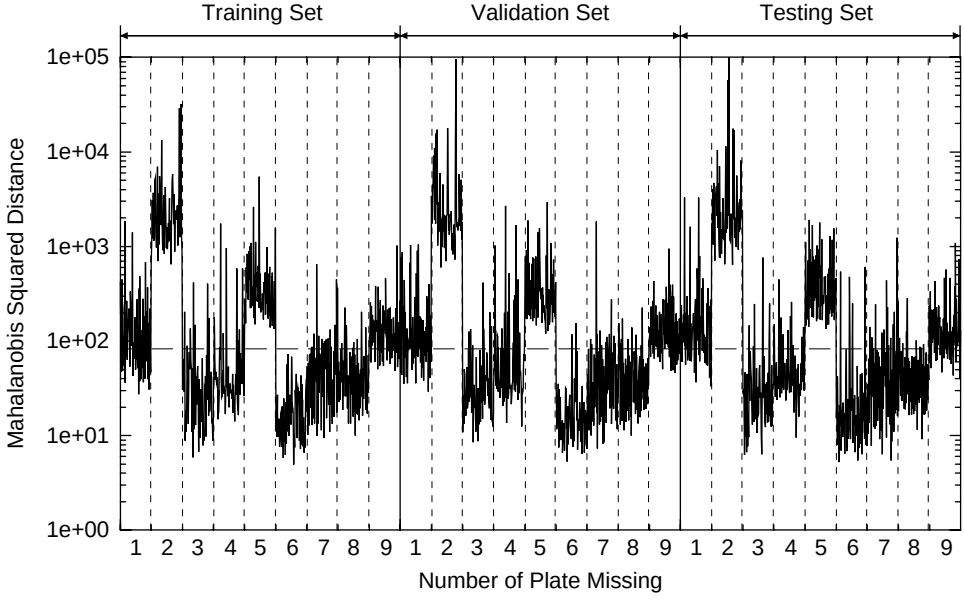

Figure 10. Outlier statistic for all damage states for the novelty detector trained to recognize panel 2 removal.

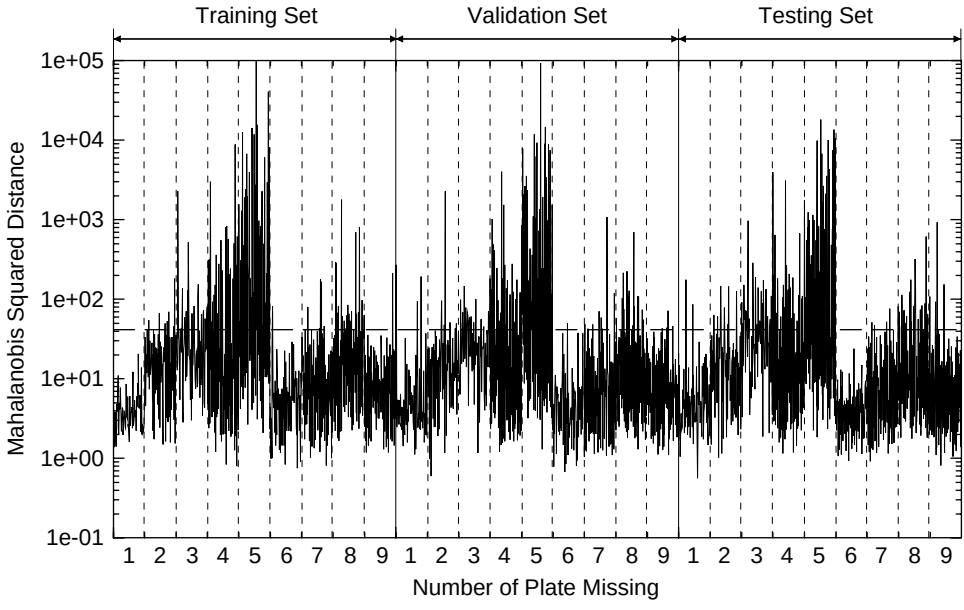

Figure 11. Outlier statistic for all damage states for the novelty detector trained to recognize panel 3 removal.

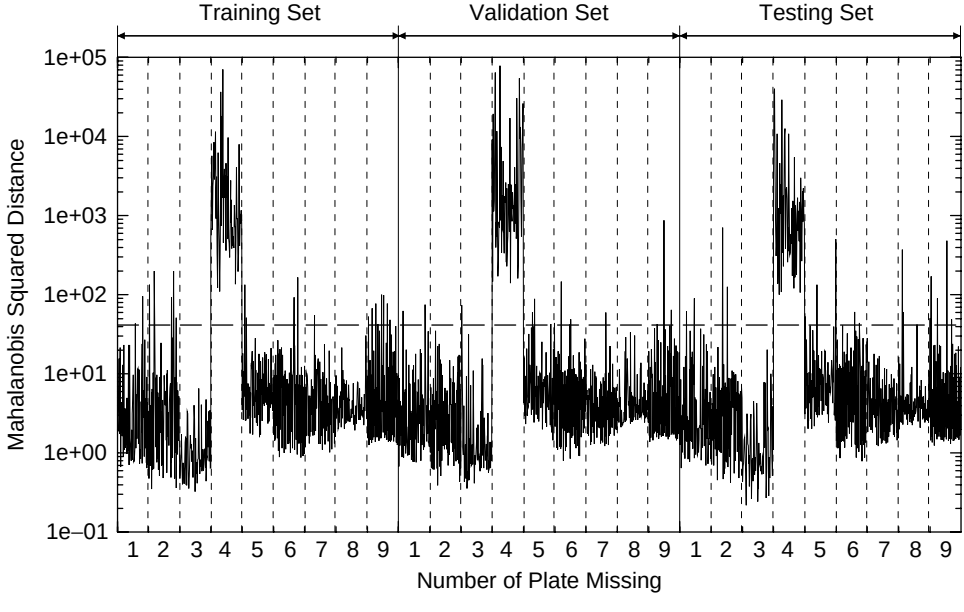

Figure 12. Outlier statistic for all damage states for the novelty detector trained to recognize panel 4 removal.

panel 3, there was no overlap, but the normal and damage clusters were so close that the probability of detection was only 60·5%.

The performance of the novelty detectors on all damage states *including* those they had been trained for is summarized in Figures 9–17. (The repetition of similar figures here is

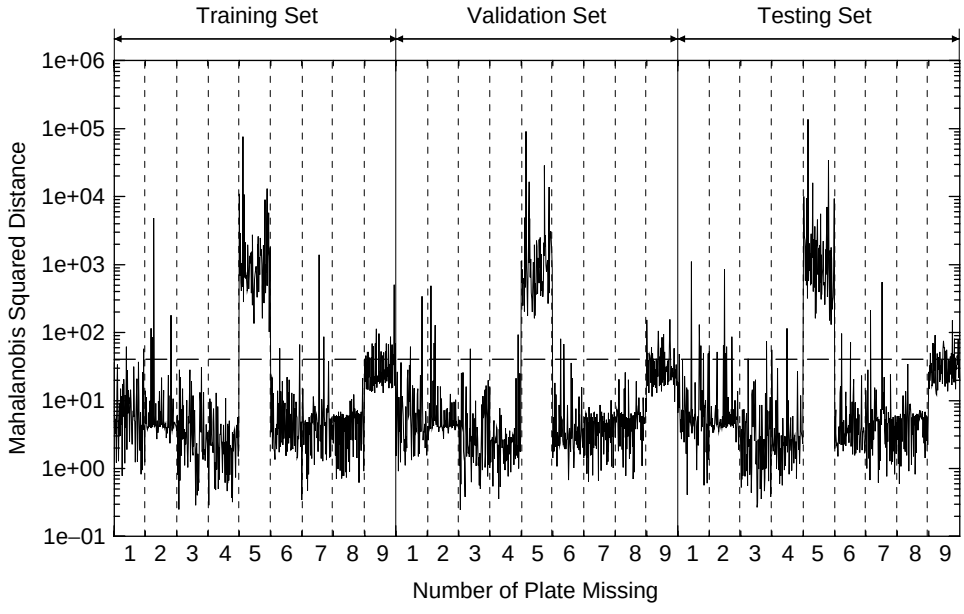

Figure 13. Outlier statistic for all damage states for the novelty detector trained to recognize panel 5 removal.

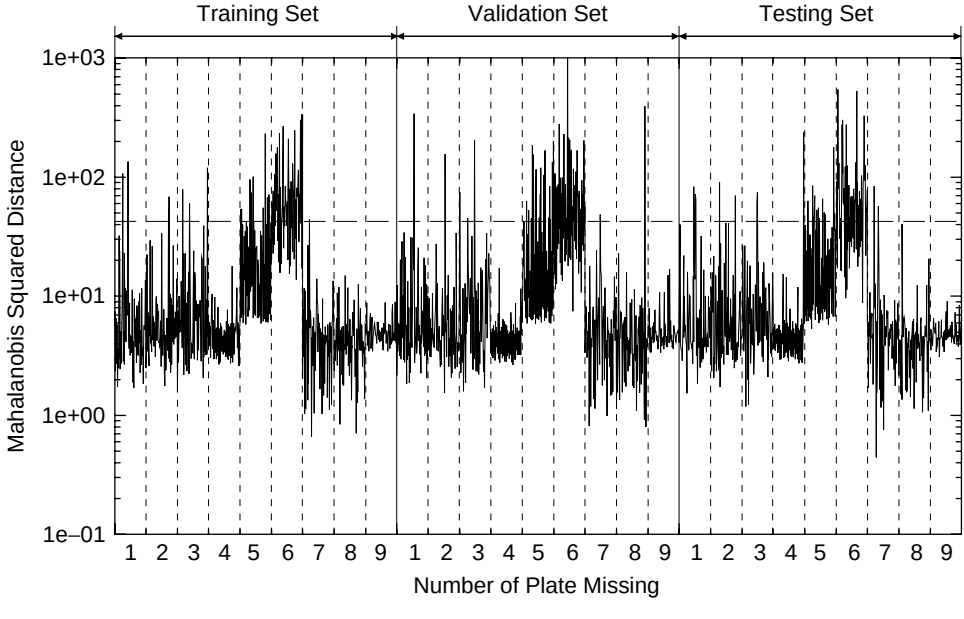

Figure 14. Outlier statistic for all damage states for the novelty detector trained to recognize panel 6 removal.

not needless, they shed considerable light on the performance of the eventual damage locator.) Each of these plots shows the Mahalanobis distance for each novelty detector evaluated over the whole set of damage states. In anticipation of the neural network

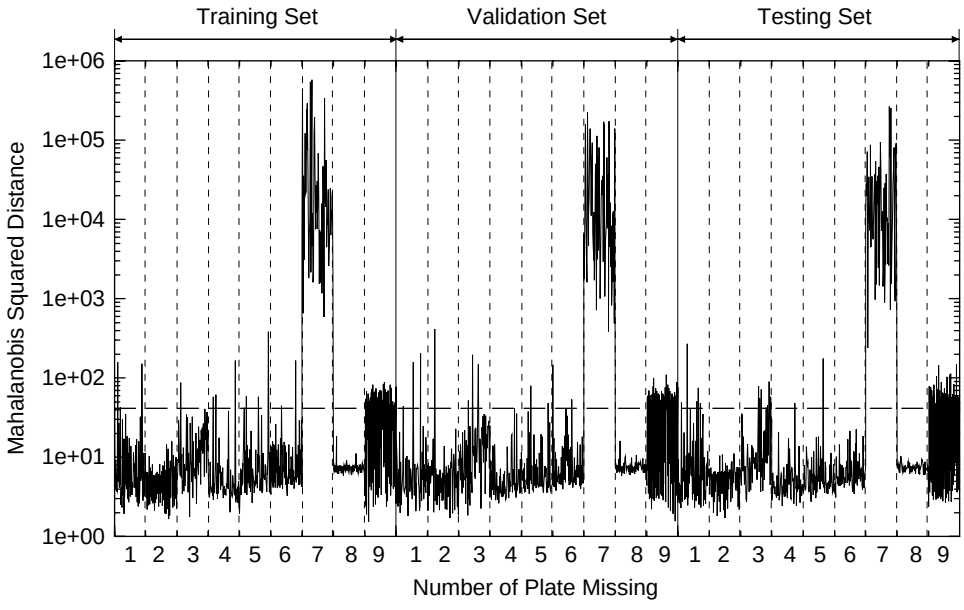

Figure 15. Outlier statistic for all damage states for the novelty detector trained to recognize panel 7 removal.

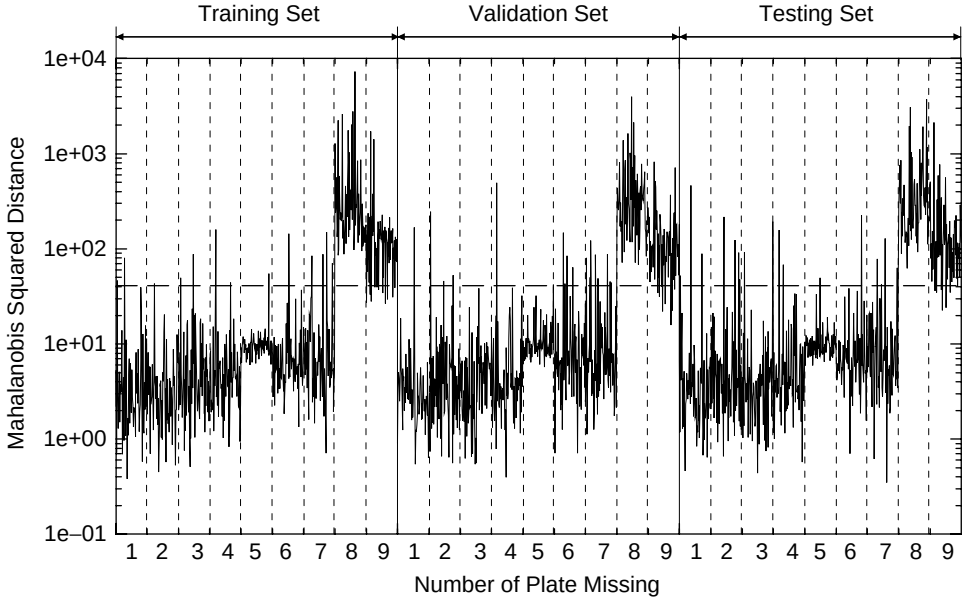

Figure 16. Outlier statistic for all damage states for the novelty detector trained to recognize panel 8 removal.

analysis, this set has been split into three subsets for training, validation and testing. Each contains 66 points from each damage state.

The horizontal dashed lines in the figures are the thresholds for 99% confidence in identifying an outlier, they are calculated according to the Monte Carlo scheme described in reference [9]. The thresholds are a function of the number of sample points and the

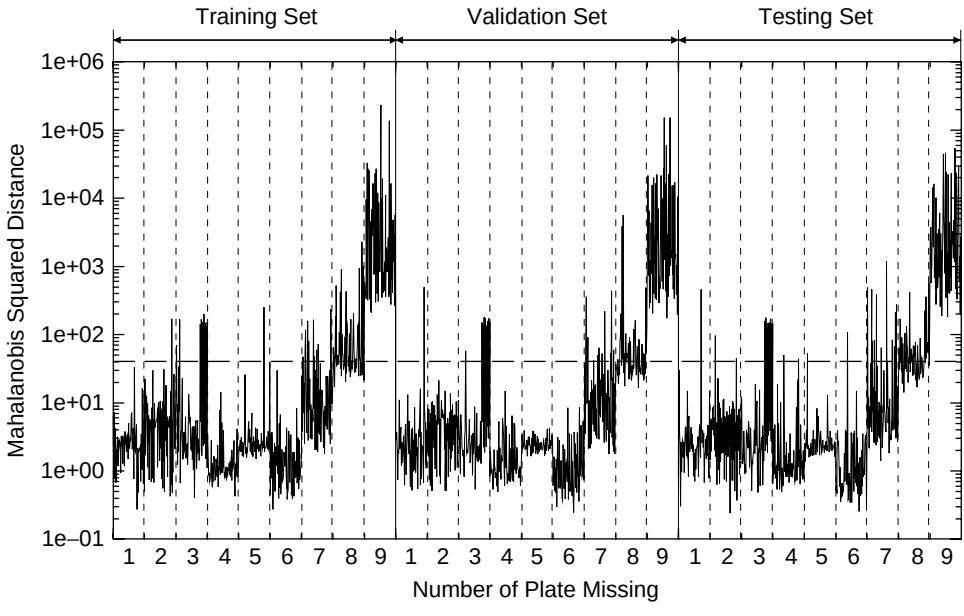

Figure 17. Outlier statistic for all damage states for the novelty detector trained to recognize panel 9 removal.

TABLE 6

99% detection thresholds for outlier analysis

Dimension	Inclusive	Exclusive
10	39·37	41·78
20	56·53	61·59
30	64·62	79·96
40	78·37	97·76

dimension of the data set. In this case the number of normal points was always 700; however, the dimension varied as different spectral line ranges were selected for features. The relevant thresholds are given in Table 6.

The results shown in Figures 9–17 are very encouraging for classification purposes. Each novelty detector substantially fires only for the panel removal for which it has been trained. There are exceptions e.g., the features for panels 3 and 8 (Figures 11 and 16). In the first case, there are excursions above threshold when panel 3 is removed, but the main excursions are for panel 5. This was expected as the feature in question was not classified as strong. In the case of Figure 16, it can be seen that the feature is almost as sensitive to the removal of panel 9 as for panel 8 for which it was trained.

4. NEURAL NETWORK RESULTS FOR LOCALIZATION

The final stage of the analysis was to produce a damage location system. The algorithm chosen here was a standard multi-layer perceptron (MLP) neural network. The idea is

simply to map the novelty indices obtained from the transmissibilities to the damage location, or in this case to which panel was removed. The neural network is supplied with the values of the nine novelty indices at the input layer and required to predict the damage class at the output layer.

Note that there are now two layers of feature extraction. At the first level, certain ranges of the transmissibilities were selected for sensitivity to the various damage classes. These were used to construct novelty detectors for the classes. At the second level of extraction, the nine indices themselves are used as features for the damage localization problem. This depends critically on the fact that the various damage detectors are *local* in some sense, i.e., they do not *all* fire over *all* damage classes. That this is true can be seen from Figures 9–17. In fact, the features are almost "orthogonal" in the sense that each largely only flags damage for one panel.

The procedure followed the guidelines in reference [10]. The data were divided into a training set, a validation set and a testing set. For the network structure, the input layer necessarily had nine nodes, one for each novelty index, and the output layer had nine nodes, one for each class. The number of hidden nodes was determined during training. In order to estimate the classification accuracy of each network the 1 *of M* strategy was used. This means that each class was assigned a specific network output. During training the network was required to respond with unity at the output corresponding to the presented class and zero elsewhere. The rationale for this training scheme is as follows. It can be shown [11], that after training with the 1 *of M* strategy, if the network is presented with a test vector it will respond at the outputs with the Bayesian *a posteriori* probability that the input vector belongs to that class. In order to assign the class then, one simply finds the highest output, i.e., the class with highest posterior probability. There are a number of conditions which must apply for this to work correctly, one is that the prior probabilities of each class must be specified. In this case they were assumed equal, i.e., all damage types are equally likely. In order to enforce the condition of equal priors, a balanced training set was used with equal numbers of examples from each class, 66 from each.

In terms of pseudo-code, the training strategy was:

```
for number of hidden layer neurons=1 to 50
{
  for different random initial conditions=1 to 10
    {
    train network on training data
    evaluate on validation data
    terminate training at minimum in validation set error
    }
  }
```

This means that the training set is used to establish the weights. The structure and training time, etc. are optimized by selecting the conditions which give the lowest validation set error. At the end of this procedure, the network has in a sense been tuned to both the training and validation sets and therefore an independent testing set is required for proper verification of the network.

When the networks were trained, the one that gave the lowest validation error had 10 hidden units and gave a misclassification error of 0·155, corresponding to a training error of 0·158. The best results were obtained after 150 000 presentations of data. The network weights were updated after each presentation, i.e., the training epoch was 1. When the best

TABLE 7

Confusion matrix from best neural network: testing set

Prediction	1	2	3	4	5	6	7	8	9
True class 1	62	1	0	0	2	0	0	1	0
True class 2	0	61	0	0	5	0	0	0	0
True class 3	0	1	52	0	7	4	0	2	0
True class 4	1	0	3	60	0	1	0	1	0
True class 5	2	1	0	0	60	3	0	0	0
True class 6	2	0	6	0	8	52	0	0	0
True class 7	1	0	4	0	1	1	58	1	0
True class 8	0	0	0	0	1	1	0	62	2
True class 9	2	1	1	0	0	0	0	15	47

TABLE 8

Confusion matrix from neural network with seven hidden neurons: testing set

Prediction	1	2	3	4	5	6	7	8	9
True class 1	62	1	0	0	0	2	0	1	0
True class 2	0	62	0	0	3	1	0	0	0
True class 3	0	1	52	0	0	7	4	1	1
True class 4	1	0	6	57	0	1	0	1	0
True class 5	3	4	2	0	47	10	0	0	0
True class 6	0	0	8	0	0	58	0	0	0
True class 7	1	0	4	0	0	2	57	0	2
True class 8	1	0	0	0	0	1	0	53	11
True class 9	1	1	0	0	0	0	0	8	56

network was tested, it gave a generalization error of 0·135, i.e., 86·5% of the patterns were classified correctly. The confusion matrix for the test is given in Table 7.

As one might expect the main errors are associated with panels 3, 6 and 9. The reasons are as follows. Panel 3 had a weaker feature which gave a novelty detector which only fired on 31% of the damaged patterns. In fact, as can be seen from Figure 11, the detector fired more often when panel 5 was removed. This is reflected in the confusion matrix, as most erroneous classifications for panel 3 are to panel 5. Similarly, panel 6 had a weak feature. More interesting is the situation with panel 9. This confusion is due to the fact that the novelty detectors for panels 8 and 9 often both fired when either panel was removed. There is therefore scope here for reselecting features which are more "orthogonal".

The network selected had 190 weights. As there were 660 training patterns, there were 3.5 patterns per weight. This falls a little short of the accepted rule of thumb which suggests that for proper generalization, 10 patterns per weight are needed [10]. This is not considered as a cause for concern here as the error on the validation and test sets are both actually lower than the error on the training set. If a more parsimonious network were required, one of the trained nets with seven hidden units actually gave a validation error of 0·158 which is barely higher than the minimum. The testing error was however 0·151 and the ratio of patterns to weights was 4·6 which is not a vast improvement. For comparison, the confusion matrix for this network is given in Table 8.

In summary, a network has been found which localizes the damage with an 86·5% success rate. This is excellent, but could probably be improved by a more extended feature selection strategy.

The results above have been left simply as numbers of correct classifications. In order to convert them to probabilities it is necessary to factor in the probabilities of actually detecting each fault in the first place. The effect of this is as follows: the probability of detecting damage and locating it correctly is 0·77; the probability of detecting damage and mislocating it is 0·11; the probability of failing to detect is 0·12. Note that *all* failures to detect are associated with panels 3 and 6, the smallest. Given the features selected, the probability here of a false alarm is found to be 0·04.

5. DISCUSSION AND CONCLUSIONS

The results above show that it is possible to establish a level 2 damage detection system, i.e., detect and locate, on a real structure. The success rate of the classifier is 86·5% which is comparable with results in the past for simulated or laboratory data. As discussed in the Introduction, it is important to recognize that this is a success only in a certain context, that of locating sizable damage in large structures using vibration-based data. The method is not at this stage proposed as a means of locating subcritical fatigue damage, and is frankly unlikely to be. This approach should be used in conjunction with local high-resolution monitoring of significant structural items.

What is new here is in the damage localization algorithm which uses a two-level feature extraction procedure in order to process the results of a network of novelty detectors, each sensitive to a different region of the aircraft. The idea here is that the novelty detectors integrate and thus magnify and smooth the difference between normal and damage condition data.

The signal processing proved to be fairly straightforward but relied on an exhaustive search through the transmissibility database for appropriate features for the novelty detectors. This is an area which could be automated. Another desirable effect of automation would be to remove the element of subjectivity from the feature selection process. The procedure chosen here was made as rigorous as possible by adopting a principled division into weak, fair and strong features. The strongest features possible were selected.

One second-level feature or novelty index per damage class was taken here. These proved to be largely independent and gave a good classification rate. It may well be possible to select a smaller set of features with a lower error rate by using an optimization procedure. Future work will include the use of genetic algorithms to select an optimum second-level feature set. This is related to the issue of sensor optimization. Because the only allowed locations for damage were the inspection panels, it was possible to make an (arguably) sensible choice for the locations of the sensors. A network was used which crossed each plate with a sensor pair. This network would not necessarily be optimal for a general location system. However, if a model were available which allowed the simulation of arbitrary damage states and locations, it could be used in an optimization procedure to determine the best positions for the sensors. This in turn raises the issue of modelling complex engineering structures and systems with a fidelity appropriate for training a damage identification system. This is a very difficult problem and progress is slow and incremental. What was achieved here was a demonstration that a location system based on a network of novelty detectors can work in an, albeit limited, large-scale environment.

As a last word on the subject of generalization, the neural network here has been trained with the limited aim of telling which of a set of panels has been removed. It cannot diagnose any other type of damage. The location problem here has been converted to a finite-class classification problem by fixing damage locations at the panels. This does not invalidate the approach. A general classification system could be designed to locate damage to within a given substructure. Despite the fact that the location system is powerless to provide information about damage off the panels, the individual features used in the network training are themselves novelty detectors, and can potentially signal any damage type. Thus, in a worst-case analysis, the system degrades to a level 1 diagnostic, which is still valuable.

Another problem for further work which should be addressed is the extension of the method to level 3 in Rytter's scheme, i.e., definition of a diagnostic which can detect, locate and size damage. One can imagine a situation in which level 2 is sufficient, i.e., when the diagnosed location of the damage is used to focus an NDT inspection which can size the fault. However, if a completely automated estimate of residual life or safety is required (level 4), the information about severity will be critical and level 3 would have to be automated. This might be the case for an in-flight (or generally on-line) diagnostic system. However, such systems are not within reach of current technology.

ACKNOWLEDGMENTS

The main acknowledgment must be to DERA, Air Systems Sector, Structures Department for funding this project as part of the DTI CARAD scheme. The authors are grateful to Mrs Sue Copley for her patience and help in administering the project. Dr Graham Skingle gave valuable technical support during the testing phase at Farnborough.

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. A. Rytter 1993 *Ph.D. Thesis, Department of Building Technology and Structural Engineering, University of Aalborg, Denmark*. Vibration-based inspection of civil engineering structures.
3. K. Worden and W. J. Staszewski 1998 *Technical Report to DERA Farnborough*. Novelty detection using kernel density estimates.
4. K. Worden 1998 *Technical Report, Department of Mechanical Engineering, University of Sheffield*. KDE—kernel density estimator, version 1.1—a users manual.
5. K. Worden and G. Manson 2001 In *Proceedings of the International Workshop on Damage Assessment in Structures—DAMAS* 2001, *Cardiff*, 35–46. An experimental appraisal of the strain energy damage location method.
6. G. Manson, K. Worden and D. J. Allman 2003 *Journal of Sound and Vibration*. **259**, 345–363. Experimental validation of a structural health monitoring methodology: Part II. Novelty detection on an aircraft wing.
7. G. Manson, C. Remillat and K. Worden 2000 *Proceedings of the ISMA* 25, *Noise and Vibration Engineering Conference, Leuven*. Damage localisation in a plate structure using novelty indices.
8. S. Sharma 1996 *Applied Multivariate Techniques*, New York: John Wiley and Son.
9. K. Worden G. Manson and N. R. Fieller 1999 *Journal of Sound and Vibration* **229**, 647–667. Damage detection using outlier analysis.
10. L. Tarassenko 1998 *A Guide to Neural Computing Applications*, Paris: Arnold.
11. C. M. Bishop 1995 *Neural Networks for Pattern Recognition*, Oxford: Oxford University Prescs.