

NASA
NAS 2-99048
CONF. PAPER
IN/06

APPENDIX A: Condition Monitoring for Helicopter Data

Presented at the IEEE SMC International Conference, Nashville, TN, Oct 8-11, 2000

Condition Monitoring for Helicopter Data

Fang Wen, Peter Willett
Electrical and Computer Engineering Department
U-157, University of Connecticut
Storrs, CT 06269
(860) 486-2195, willett@engr.uconn.edu

Somnath Deb
Qualtech Systems, Inc.
100 Great Meadow Rd, Ste. 501
Wethersfield, CT
(860) 257-8014, deb@teamqsi.com

Abstract

In this paper the classical “Westland” set of empirical accelerometer helicopter data is analyzed with the aim of condition monitoring for diagnostic purposes. The goal is to determine features for failure events from these data, via a proprietary signal processing toolbox, and to weigh these according to a variety of classification algorithms.

As regards signal processing, it appears that the autoregressive (AR) coefficients from a simple linear model encapsulate a great deal of information in a relatively few measurements; it has also been found that augmentation of these by harmonic and other parameters can improve classification significantly. As regards classification, several techniques have been explored, among these restricted Coulomb energy (RCE) networks, learning vector quantization (LVQ), Gaussian mixture classifiers and decision trees. A problem with these approaches, and in common with many classification paradigms, is that augmentation of the feature dimension can degrade classification ability. Thus, we also introduce the Bayesian data reduction algorithm (BDRA), which imposes a Dirichlet prior on training data and is thus able to quantify probability of error in an exact manner, such that features may be discarded or coarsened appropriately.

1 Introduction

Qualtech Systems has developed a suite of fault-isolation tools (TEAMS) which can, in real time and based on binary sensor data, isolate single and even multiple faults in complex systems. However, many sensors (for example, of vibration) are incapable of reliable decision-making on their own, and hence it has become necessary to develop a (real-time) signal processing “front-end” to the TEAMS inference engine whose goal is to render decisions as intelligent as possible. The signal processing system includes a wide menu of spectral and statistical manipulation primitives such as filters, harmonic analyzers, transient detectors, and multi-resolution decomposition.

The signal processing kit includes pattern classifica-

tion software, including techniques based on restricted Coulomb energy (RCE), decision trees (DT), learning vector quantization (LVQ), fuzzy logic, Bayesian data reduction (BDRA), Gaussian mixtures (GM) and multi-layer perceptrons (MLP). At present the former three are implemented within the SP toolkit, and the fifth and sixth are implemented off-line in MATLAB using features provided by the toolkit.

Recognition of faults can hence be automated provided there is sufficient training data. This paper thus includes analysis of no-fault and seeded-fault vibration data from a CH-46 (“SeaKnight”) helicopter aft gearbox as collected from a test-stand. This data is made freely available through the generosity of the Penn State ARL [8].

Results show promising fault detection accuracy, particularly when learning is based on auto-regressive (AR) coefficient features. The analysis presented in this paper is an outgrowth of that in [11]. In that paper, only a very abbreviated version of the Westland dataset was explored, and the RCE, LVQ, and DT schemes were discussed. In this paper the full dataset is used, and the set of classifiers is augmented by the GM and BDRA approaches.

In section 2 we go into detail about the toolbox classification techniques: LVQ, DT, RCE, Gaussian mixture, and BDRA classifiers. In section 3 we apply the signal processing and classifiers to the Westland helicopter dataset. Similar to results reported elsewhere, we find near-perfect fault-recognition accuracy, in our case with relatively small feature sets involving autoregressive coefficients.

2 The Classifiers

2.1 Restricted Coulomb Energy Classification

The RCE classifier [4, 9] relies on the approximation of a decision region via a union of hypersphere “cells”. Cells may overlap if they do not belong to the same class, and this may produce ambiguous outputs. Note that partition of the observation space into decision re-

gions is not exhaustive in the RCE approach. Training is iterative, and is described in [11]. After the network has become fixed classification is accomplished by interrogation of membership of the various cells: each cell is assigned a class, and the output corresponds to that class. For the cases that data is either a member of *no* cell, or of several which are of different classes, the RCE classifier gives an indeterminate output: such cases may be decided randomly or by heuristic.

2.2 Learning Vector Quantization Classification

The LVQ classifier [5] is a variation on the traditional cluster-classifier based on K-means training [10]. In essence, each class is assigned sub-clusters defined by their centroids, and data are classified based on the membership of the centroid to which they are nearest. Training is iterative, and is described fully in [11]. An LVQ classifier may be considered a development on the earlier K-means based cluster classifiers in that non-separable classes cause no intrinsic, and in that there is an intelligent means of “pruning” clusters.

2.3 Decision Tree Classifier

At core the DT classifier [10] produces its output by asking a series of questions which must have binary answers. The “path” taken may be thought of as traversal of a logical tree; but the form of the resultant decision regions must be as hyper-rectangles. In principle it is possible and easy to separate the training data precisely via a sufficiently-rich question set. In practice there are too many “questions” (*parameters*), and the DT classifier is found not to have a particularly good generalization ability. There are means to limit the number of questions, and these generally amount to the choice of a cost to be placed on a question’s posing. In our implementation we use an information-theoretic cost function, although admittedly its basis is empirical rather than true prior statistics.

2.4 Gaussian Mixture Classifier

This classification technique has a greater statistical grounding than the previous, in that a probability density function (pdf) is sought for each class. The specific pdf used is a mixture of multivariate Gaussians:

$$f(\mathbf{x}) = \sum_{i=1}^M \pi_i \frac{1}{\sqrt{|2\pi\mathbf{R}|}} e^{-\frac{1}{2}[\mathbf{x}-\mu_i]^T \mathbf{R}^{-1}[\mathbf{x}-\mu_i]} \quad (1)$$

There are M elements to the mixture, and each has a different mean μ_i and prior probability π_i . Decisions are made according to the maximum posterior probability of each class (in fact, classes are assumed to

be equally-likely a-priori). Note that if $M = 1$ this is identical to the quadratic discriminant classifier.

Training is via the expectation/maximization (EM) algorithm [10]. The correlation matrix \mathbf{R} is common to all elements of the mixture within a class of fault – this is known as a “homoscedastic” mixture – and the ideas behind this are that the number of elements to be estimated can be reduced and that there is little concern about unboundedness of the likelihood function. A variant of the above restricts \mathbf{R} to be diagonal; this reduces the number of parameters to estimate considerably, but in this particular case (see, for example, figure 1) the ability to “tilt” the pdf level curves arising from the use of a full \mathbf{R} is valuable.

2.5 BDRA Classification

The Bayesian data reduction approach is perhaps the most statistically defensible of the classifiers used. It begins with a quantized version of the data, and assumes a Dirichlet prior (of complete ignorance) on this a prior, for each fault class. From that prior distribution classification is relatively simple; the key is that the prior enables an explicit (and correct) probability of error to be calculated, and thence features may be pruned in an optimal way. The BDRA is discussed in detail in [6], among other places. Generally the BDRA works very well when there are too many features for the training data to support, and/or when the classes are not easily separable.

The BDRA requires that the data be pre-quantized. To some extent this is not a concern, since the quantization may be as fine as desired – the BDRA coarsens the quantization as part of its feature/level selection. For practical reasons, the quantization cannot be too fine, and hence it is not expected that this dataset will be kind to the BDRA. In fact, the BDRA results are reasonable, but what is interesting is its ability to select features and its prediction of its own performance.

3 Results on CH-46 Data

3.1 The Data

In the early 1990’s the US Navy contracted with Westland, a British helicopter manufacturer, to develop and study vibration signatures for the CH-46 (SeaKnight) aft gearbox. Essentially this is “test-stand” (not in-flight) data; this is a disadvantage from the perspective of result reliability, but offers a distinct advantage in that the vibration signatures are *labeled*. The data is as follows:

- There are 68 files each containing data traces of 100,000 samples.

- For each case there is data available from eight accelerometers.
- There are a total of nine fault conditions, ranging in severity from mild to severe. Faults were “seeded” (by electronic discharge milling) in the sense that parts with known defects were installed and de-installed.
- There is data from no-fault (normal) operating conditions.
- Data was observed at nine different torque levels (since this is a rotorcraft, angular velocities are relatively constant), ranging from 27% to 100%.

For details on the faults, etc., please see [1, 8]. Note that if all fault levels and torques were represented there would be 90 files; in fact, a number of conditions are unrepresented in the data. As regards training versus testing, the entire dataset is split randomly into two parts, which are used separately.

The data has been analyzed previously (e.g. [1, 7, 12]) using a variety of classification techniques such as multi-layer perceptrons and fuzzy reasoning. Indeed, this is apparently an “easy” (or separable) dataset for classification, as the reported accuracies approach 100%. Thus, our goal here is not really to *beat* previous (unbeatable!) approaches, but to attempt to match them using the SP toolbox classifiers. Further, it appears that past approaches have often used a rather dense feature set (several hundred features, such as FFT outputs), and we attempt here to use a much sparser arsenal.

3.2 The Features

3.2.1 AR Coefficients

It is possible to use periodogram outputs explicitly as features for classification; however, in general this implies a great many features, and the usual “curse of dimensionality” may ensue. Since it is clear that spectral features do indeed yield much relevant information, we propose to use a concise way of representing the spectrum: the autoregressive (AR) parameters [2]. These are estimated on blocks of various sizes, from $N = 256$ to $N = 16384$.

Examples of AR coefficients are given in figures 1 and 2. It is clear that there is a reasonable amount of structure to these, but also that certain conditions cannot be separated reliably using only such data. In fact, there are 8 accelerometers from which to choose, and a further two AR coefficients.

3.2.2 FFT Features

AR coefficients are able to digest much *global* spectral information into a small dimension. There is some in-

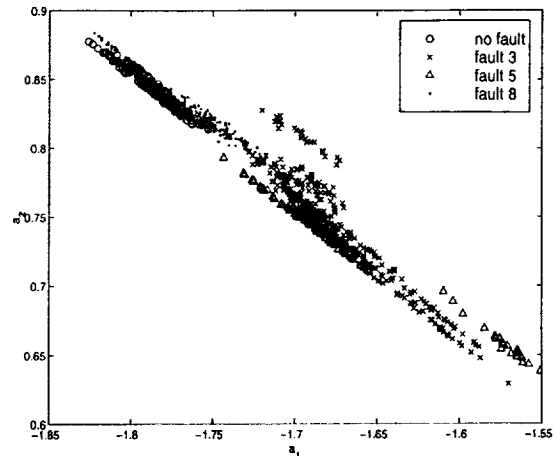


Figure 1: Scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels, estimated on blocks of length 4096, for faults 3, 5, 8 and no-fault conditions.

dication that faults may manifest in *specific* frequency behavior, and hence we additionally investigate the use of relative harmonic power (RHP). The i^{th} RHP is the ratio of the i^{th} -highest spectral peak (measured via FFT) to the average power. In the sequel we use 4 RHP’s. Examples are given in figures 3 and 4, for the same conditions as figures 1 and 2. It is clear that these features are less a direct indication of fault class.

acc	RCE	LVQ	DT
1	54	29	52
2	49	76	72
3	78	27	73
4	78	38	72
5	60	33	53
6	57	32	48
7	84	34	81
8	30	14	31

Table 1: Percentage of correct classification for three classifiers, versus accelerometer number – features are 4th-order AR coefficients from individual accelerometers.

3.3 Results for RCE, LVQ and DT

We first examine the results for the case that accelerometers are used individually. The features used are AR coefficients of order $p = 4$, each estimated on a block of length $N = 4096$. Results are reported in table 1. None of these performances is acceptable, although accelerometers 3, 4, and 7 appear to be the most promising. Motivated by this, we attempt to classify by combining accelerometers. Example results

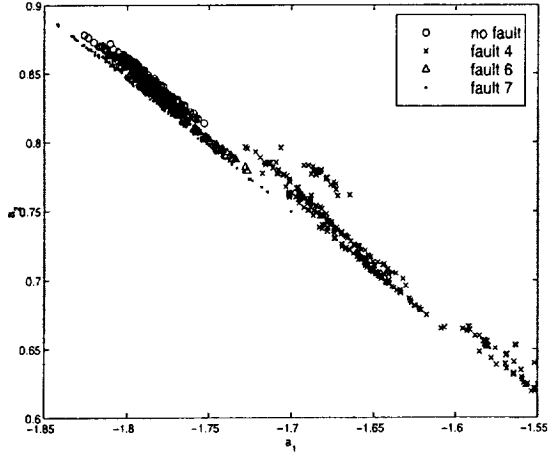


Figure 2: Scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels, estimated on blocks of length 4096, for faults 4, 6, 7 and no-fault conditions.

acc	RCE	LVQ	DT
2,7	96.4	89.4	94.7
3,7	99.1	96.7	95.5
4,7	98.3	89.6	94.9
all	98.5	96.6	91

Table 2: Percentage of correct classification for three classifiers, with combined accelerometer AR coefficients ($p = 4$) from individual accelerometers as features. (In the 1st row $p = 2$.)

are shown in table 2. We find that the combination of accelerometers 3 and 7 is the most propitious. There is apparently little benefit from using all accelerometers. In table 3 we explore the choice of AR order. The results indicate that $p = 4$ is a good compromise between sensitivity and dimensionality. With this choice we consider adding the RHP features. In table 4 we do, and additionally compare the results for different block lengths. The results become quite outstanding in the cases $N = 4096$ and $N = 16384$, particularly for the RCE classifier; the LVQ classifier is somewhat less satisfying, and the DT scheme has been overcome. Finally, we note that we have chosen to ignore the torque level in our classification feature set. That is, we have trained using combined data from all torque levels, and results to this point are given in terms of combined probability of correctness. It could be argued that this is dangerous, in that poor performance may lurk at some torque level; in fact, as seen in table 5, this is not the case.

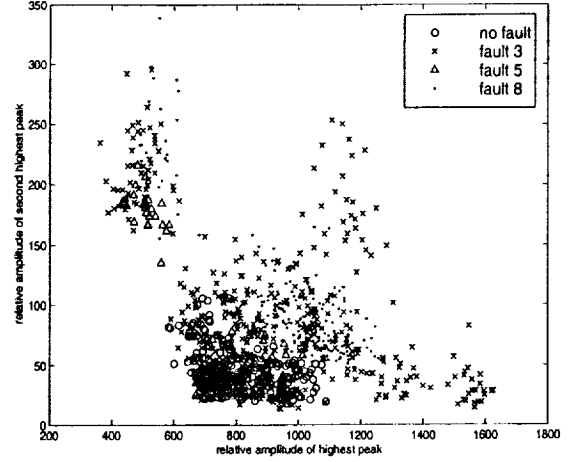


Figure 3: Scatter plot of RHP values, corresponding to figure 1.

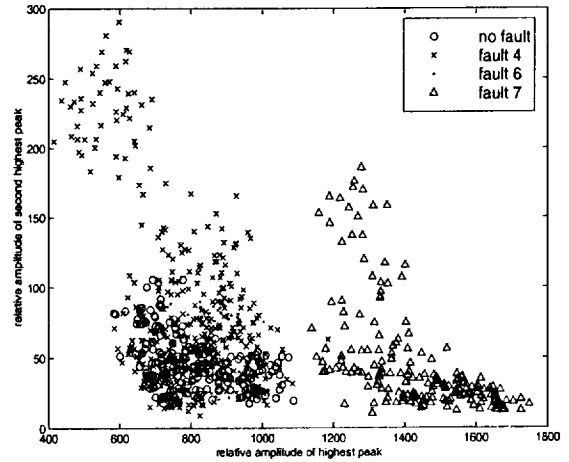


Figure 4: Scatter plot of RHP values, corresponding to figure 2.

3.4 Results for GM

We show example results for the two homoscedastic GM classifier variants in figure 5. Apparently the GM classifier is not as good as the RCE scheme in this situation; GM classifiers are often more useful when the data is less separable and when confidence information is desired, so it is perhaps interesting that the performance is as good as it is. Of particular note is the $M = 1$ GM classifier – this is essentially a quadratic discriminant, and its probability of error is very low. As regard the second GM classifier variant – that with a diagonal covariance matrix – it is interesting to observe from figure 5 that the performance improves markedly as the number of mixture elements M is increased. There is some explanation of this in figure 6,

p	RCE	LVQ	DT
2	83.2	88.7	87.7
3	94.0	85.4	97.4
4	99.1	96.7	95.5
6	99.0	92.8	95.0
8	98.9	91.1	95.9
12	97.6	96.1	93.8

Table 3: Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders p , estimated on data blocks of length $N = 1024$.

N	RCE	LVQ	DT
1024	96.8	95.1	93.1
4096	99.6	97.9	93.9
16384	99.3	96.7	91.9

Table 4: Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders $p = 4$ estimated on data blocks of length N . The feature set is augmented by the RHP spectral peak clues.

in which the “coverage” of one class’s data by the mixture elements is illustrated. It is clear that the more elements, the more complete the coverage.

3.5 Results for BDRA

In table 6 we show the results for the BDRA in terms of correct detection of a fault condition – no attempt is made here to isolate the fault, but testing is simply binary. (The BDRA is capable of multi-class operation, but the version used does not support that.) Despite the fact that the BDRA is not particularly well-suited to the problem, the results are quite good. It is particularly notable that the algorithm is able to predict its own performance with reasonable fidelity.

As indicated earlier, a strength of the BDRA is that it is able to determine *for itself* a feature set. In fact, it is originally “given” a the entire set of features, quantized to whatever fineness is desired – in table 6 this is 5 or 10 levels per feature, thresholded for equal probability, meaning in the case of 10 levels and $p = 6$, there are initially $8 \times (6 + 4) \times 10 = 800$ possible observations. In table 7 the final quantization from the BDRA is shown, and the dominance of accelerometers 3 and 7 is clear. Table 7 deals only with AR coefficients: if RHP features are also presented to the BDRA, it turns out that these are often chosen.

4 Summary

Here we have reported on a signal processing toolbox specially matched to Qualtech Systems TEAMS

torque	RCE	LVQ	DT
27%	97.6	94.1	92.9
40%	100	100	98.8
45%	100	100	100
50%	100	100	97.2
60%	100	98.6	88.9
70%	99.1	82.4	83.3
75%	97.9	99.0	84.8
80%	99.1	98.2	91.7
100%	100	98.3	89.2

Table 5: Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders $p = 4$ estimated on data blocks of length $N = 4096$. The feature set is augmented by the RHP spectral peak clues. Training data is combined over all torque levels, and testing is done individually at each torque level.

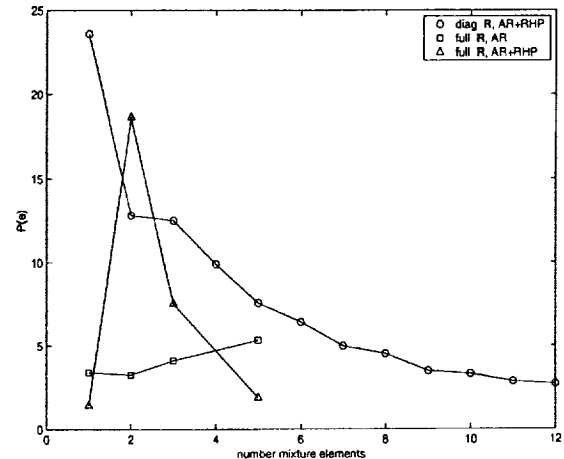


Figure 5: Probability of error performance for Gaussian mixture classifiers. Here $p = 4$ and $N = 1024$.

diagnostic inference engine, and in particular on its classification capability as applied to the “Westland” data set. We have found that essentially perfect diagnostic performance is achievable via the use of AR coefficient features augmented by harmonic peak information. The best classification performance appears to come from the RCE learning/classification scheme. The approach works well across all torque levels, so there is no need to supply engine load information to the classifier. We have also found that the Bayesian data reduction (BDRA) approach, despite not being well-matched to the problem, works surprisingly well, and indeed that its ability to select features (perhaps for other classifiers?) is particularly promising.

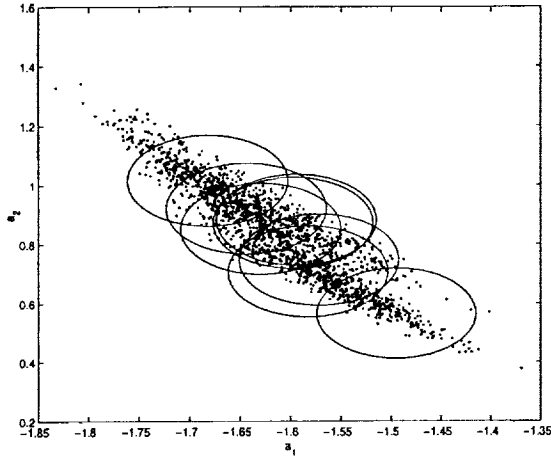


Figure 6: Dots show scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels. Ellipses are probability contours for elements of diagonal-R Gaussian mixture fit with 8 elements.

N	p	C_5^t	C_5^a	C_{10}^t	C_{10}^a
1024	2	93.0	92.3	95.8	96.9
1024	4	92.0	93.8	94.9	95.7
1024	6	93.3	93.3	94.4	96.9
4096	2	95.1	96.6	96.0	98.5
4096	4	95.6	98.8	95.6	98.8
4096	6	92.0	95.2	94.1	96.4

Table 6: Percentage of correct fault detection for BDRA, using $AR(p)$ coefficients and RHP clues. Subscript of C denotes number of initial quantization levels per feature; superscript a means actual, and t is theoretical.

5 Acknowledgments

This research has been supported by ARO STTR contract DAAG55-98-C-0057, with prime contractor Qualtech Systems, Inc., and a subcontract to the University of Connecticut. The authors also wish to acknowledge the help of Vivek Rajan for his work on the classification package, and the support of NASA. The authors are grateful to Penn State ARL for making this data freely available on the world wide web.

References

- [1] B. Cameron, "Final Report on CH-46 Aft Transmission Seeded Fault Testing", *Westland Helicopters Research Paper RP907*, Westland Helicopters, 1993.
- [2] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall, 1996.

N	p	f_1	f_2	f_3	f_4
1024	2	a_3^1	$2 \times a_7^2$	—	—
1024	4	a_3^4	a_3^4	a_7^3	$2 \times a_7^4$
1024	6	a_2^6	a_4^5	a_7^1	$2 \times a_7^3$
4096	2	a_3^4	$2 \times a_7^2$	—	—
4096	4	a_3^4	a_4^4	—	—
4096	6	a_2^6	a_3^4	$2 \times a_7^1$	—

Table 7: Features used by BDRA; only AR features present, initially quantized to 10 levels. The notation a_i^j means that the j^{th} AR coefficient of accelerometer i is active, and $2 \times a_i^j$ means that three levels (two thresholds) are kept.

- [3] M. Horsey, "WHL Vibration Analysis Results from CH46 Aft Gearbox Testing", *Westland Helicopters Mechanical Research Report MRR 20168*, 1993.
- [4] M. Hudak, "RCE Classifiers: Theory and Practice", *Cybernetics and Systems: An International Journal*, vol.23, pp483-515, 1992.
- [5] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, "LVQ-PAK: A Program Package for the Correct Application of Learning Vector Quantization Algorithms", *Proceedings of the IJCNN*, pp725-730, 1992.
- [6] R. Lynch and P. Willett, "Performance Considerations for a Combined Information Classification Test Using Dirichlet Priors", *IEEE Transactions on Signal Processing*, pp1711-1714, June 1999.
- [7] P. Monson, M. Dwonczyk, and E. Manolakos, "Analog Neural Networks Based Helicopter Gearbox Health Monitoring System", *Journal of the Acoustical Society of America*, vol.96(6), pp3235-3249, 1995.
- [8] Pennsylvania State University Applied Research Laboratory, <http://wisdom.arl.psu.edu/Westland/>.
- [9] D. Reilly, "The RCE Neural Network", *CRC Press Industrial Handbook*, pp.1025-1037, 1997.
- [10] B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1997.
- [11] F. Wen, P. Willett, and S. Deb, "Signal Processing and Fault Detection with Application to CH46 Helicopter Data", *Proceedings of the 2000 Aerospace Conference*, Big Sky MT, March 2000.
- [12] G. Yen, "Health Monitoring of Vibration Signatures in Rotorcraft Wings", *Neural Processing Letters*, vol.4(3), pp127-137, 1996.