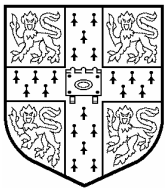
Fault Identification Using Neural Networks And Vibration Data



A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

by

Tshilidzi Marwala
St. John's College

August 2000

PREFACE

The work described in this thesis was carried out in the Cambridge University Engineering Department between October 1997 and June 2000.

I express my gratitude to Dr H.E.M. Hunt who acted as my supervisor, for his guidance and support throughout my years at Cambridge, to Professor J. Woodhouse, Professor R.S. Langley and Prof. M. Niranjan for their valuable advice on vibration analysis and neural networks. I thank Mr. D. Miller for assisting me with the experimental aspect of this work, Mr. J. Talbot, Ms. T.C. McMenemy and Dr. R. Blonbou for correcting the grammar and giving valuable comments on the presentation of this work. I thank Dr. Mulalo R. Doyoyo for much valuable advice and to Professor P.S. Heyns for introducing me to fault identification. I thank all the people who contributed to this work–their fruitful contributions are highly appreciated–particularly the anonymous reviewers of several journals. I am grateful to the Bradlow Foundation; Mandela and Malaysian Commonwealth Scholarship; St. John's College and the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom for the financial support.

I express my gratitude to my family, friends and community. I thank my parents Mr. S.A. Marwala and Mrs. R.K. Marwala for their unwavering support. This thesis is dedicated to all those good people whose voices have never been heard.

Except where reference is made to the work of others, this thesis is the result of my own original work. No part of this work has already been, or is currently being, submitted for any other degree, diploma or other qualification. This thesis is 192 pages in length, and contains approximately 42000 words.

T. Marwala

Cambridge, August 2000

SUMMARY

This dissertation introduces fault identification methods using vibration data. The pseudo modal energies, defined as the integrals of the real and imaginary components of the frequency response functions over various frequency ranges, are proposed for fault identification in structures. Equations that formulate the pseudo modal energies in the modal domain and their respective sensitivities are derived in receptance and inertance form.

When tested on a simulated cantilevered beam, the pseudo modal energies are found to be more resistant to noise than the mode shapes while natural frequencies are found to be more resistant to noise than the pseudo modal energies. The pseudo modal energies are found to be better indicators of faults than natural frequencies and equally as good indicators of faults as the mode shapes. They are also found to be faster to compute than the modal properties.

The committee-of-neural-networks technique, which employs both the pseudo modal energies and modal properties simultaneously to identify faults, is proposed. Neural networks are formulated in the maximum-likelihood framework and trained using the scaled-conjugate-gradient method. When tested on a simulated cantilevered beam, the committee-of-networks method gives more accurate classification of faults than the individual methods.

To test the proposed methods experimentally, vibration data from a population of 20 steel cylinders are measured. Holes of 10-15 mm diameter are introduced in each of the cylinders. The networks are subsequently formulated in the Bayesian framework and trained using the hybrid Monte Carlo method. The committee-of-networks is observed to provide more accurate classification of faults than the individual methods, when the maximum-likelihood and Bayesian methods are used. The Bayesian approach is, on average, found to give more accurate classification of faults than the maximum likelihood method and to allow the calculation of confidence levels.

CONTENTS

NOMENCLATURE

a_i	lower frequency bounds for i^{th} pseudo modal energy
b_i	upper frequency bounds for i^{th} pseudo modal energy
[C]	damping matrix
CF	committee factor
COMAC	coordinate modal assurance criterion
COPMEAC	coordinate pseudo-modal energy assurance criterion
CPU	computer process unit
d[•]	the differential operator
e(•)	error as a function of •
E	potential energy
E, [E]	objective function, error matrix
f_{inner}, f_{outer}	hidden and output activation function
F	force input vector
FE	finite element model
FN	false negative
FP	false positive
FRF(s)	frequency response function(s)
j	$\sqrt{-1}$
g_p	represents changes in p structural parameters
h(•)	approximated mapping function of as a function of •
H	Hamiltonian
$H_{kl}(\omega)$	frequency response function due to the excitation at k and measurement at l
IME(s)	inertance pseudo modal energy (energies)
K(p)	kinetic energy
[K]	stiffness matrix
L	number of leapfrog steps
[M]	mass matrix
MAC	modal assurance criterion

MDLAC	multiple damage location assurance criterion
ME	pseudo modal energy
MLP	multi-layer perceptron
MP(s)	modal propertie(s)
MS	mode shapes
MSEs	mean of square of errors
n	the iteration step
p[•]	the probability density function
PCA	principal component analysis
p	momentum
P(D|w)	the likelihood function
P(w|x)	probability distribution function of the weight space
P(w|D)	the posterior probability distribution
P(y|x,w)	the distribution of the noise on the neural network output
PMAC	partial modal assurance criterion
PME(s)	pseudo modal energy (energies)
R	constraint equation
RBF	radial basis function
RME(s)	receptance pseudo modal energy (energies)
SOF(s)	statistical overlap factor(s)
STD	standard deviation
STDSEs	standard deviation of the squares of errors
STFT	short time Fourier transform
STRETCH	structural translational and rotational error checking
TN	true negative
TP	true positive
X,X′,X″	displacement, velocity and acceleration
x	input to the neural network
y	output to the neural network
y_1, y_2	identity of fault when pseudo modal energy and modal-property data are used respectively

w	weights
WT(s)	wavelet transform(s)
WVD	Wigner-Ville distribution
Z_i	normalization factor of i
α	prior contribution to the error
β	data contribution to the error
χ	modal property vector
Δ	change of
∇	gradient of
ε	step size
$\varepsilon[\bullet]$	expected value of $\bullet$
$\{\phi\}_i, \{\bar{\phi}\}_i$	i^{th} mode shape and complex mode shape vector
λ	Lagrange's multipliers, direction of the trajectory
μ	mean of the input data
η	learning rate
ν	regularisation coefficient
Σ	covariance matrix
τ	fictitious time
$\omega, \omega_i, \bar{\omega}_i$	frequency, natural frequency i, complex natural frequency
ζ	damping ratio
ψ	inertance pseudo modal energy vector
{0}	null vector
{ }	vector
[]	matrix
$\Im$	imaginary part
$\Re$	real part

Subscripts

AVE	average
COM	committee

d	damaged properties
i,j,n,r,k,l	indices
u	undamaged
m	measured
MED	median

Superscripts

K	number of network output units
M	number of hidden units
N	number of (measured degrees of freedom, modes, training examples)
P	the number of vectors in the training set
Q	number of discarded states
R	number of retained states
T	transpose
W	number of weights
*	complex conjugate

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION TO FAULT IDENTIFICATION IN MECHANICAL SYSTEMS

1.1 INTRODUCTION

The process of monitoring and identifying faults in structures is of great importance in aerospace, civil and mechanical engineering. Aircraft operators must be sure that aircraft are free from cracks. Bridges nearing the end of their useful life must be assessed for load-bearing capacities. Cracks in turbine blades lead to catastrophic failure of aero-engines and must be detected early. Many techniques have been employed in the past to locate and identify faults. Some of these are visual (e.g. dye penetrant methods) and others use sensors to detect local faults (e.g. acoustics, magnetic field, eddy current, radiographs and thermal fields). These methods are time consuming and cannot indicate that a structure is fault-free without testing the entire structure in minute detail. Furthermore, if a fault is buried deep within the structure it may not be visible or detectable by these localised techniques. The need to detect faults in complicated structures has led to the development of global methods, which are able to utilise changes in the vibration characteristics of the structure as a basis of fault detection [Doebling *et al.*, 1996].

The vibration characteristics of any structure can be described in terms of its distributed mass, damping and stiffness properties. Through the use of vibration modelling and using differential equations, an eigenvalue problem can be formulated and the mode shapes and the

natural frequencies of the structure calculated. If a fault is introduced to the structure, the model changes. This results in changes in the vibration characteristics of the structure.

For a fault detection procedure to be useful, it should be able to detect small faults at an early stage of their development. Within the sensor resolution, it should be possible to estimate the magnitudes of faults and to predict the remaining life cycle of the structure. An ideal procedure would be automated, and would not rely on the engineering judgement of the user. Such procedures do not exist at present because sufficiently robust mathematical tools needed to solve complex fault identification problems have not been clearly identified. A more realistic goal is to develop a process that uses measured response of a faultless structure as a baseline for future comparisons of measured response. Two issues need to be considered in developing such a fault identification procedure. First, any fault identification method must be subject to operational constraints, such as those dictated by the size and location of the structure. For instance, it is not possible to isolate an offshore oil platform from the sea as well as the seabed and therefore there will be a great deal of background noise. In addition the procedure will need to be insensitive to the technique of individual operators and the results need to be repeatable. Second, any fault identification methods must be able to discriminate changes caused by faults from measurement inaccuracy.

The effects of faults in structures may be classified as either linear or non-linear. In the linear case, faults do not alter the linearity of the structure. The changes in vibration data are a result of changes in geometry and/or the material properties, but the structure may still be modelled by a linear equation of motion. Non-linear cases occur when introducing faults changes the linearity of the structure. One example is the opening and closing of a crack, which may be caused by fatigue damage. Other examples include loose connections and non-linear material behaviour. A robust fault detection method should be able to detect both types of faults. To date most research has concentrated on the problem of linear fault detection [Doebling *et al.*, 1996].

In this chapter, different ways of presenting and using vibration data for fault identification are reviewed. The applications of these methods to engineering scenarios are investigated. In conclusion, questions not previously tackled are addressed and the contributions of this work are discussed.

1.2 VIBRATION DATA TO BE USED

There are four main methods by which vibration data may be represented: time, modal, frequency and time-frequency domains. Raw data are measured in the time domain. In the time domain, Fourier transform techniques may be used to transform data into the frequency domain. From the frequency domain data, and sometimes directly from the time domain, the modal properties may be extracted. All of these domains theoretically contain similar information but in reality this is not necessarily the case. Because the time domain data are relatively difficult to interpret, they have not been used extensively for fault identification, and for this reason, this chapter reviews only the modal, frequency and time-frequency domains.

1.2.1 Modal domain data

The modal domain data are expressed as natural frequencies, damping ratios and mode shapes. The most popular technique of extracting the modal properties is through the use of modal analysis [Ewins, 1995]. This type of data has been used individually and concurrently for fault identification. In this study, methods that use this type of data individually and concurrently are reviewed.

(a) Natural frequencies

The analysis of shifts in natural frequencies caused by faults has been used to identify structural damage. As the changes in natural frequencies caused by average damage levels are of small magnitudes, an accurate measurement technique is required for this approach to be successful. This problem limits the level of damage that natural frequencies are able to identify to that of high magnitude.

Cawley and Adams (1979) used changes in natural frequencies to detect damage in composite materials. In order to calculate the ratio between frequency shifts for two modes, they considered a grid between possible damage points and constructed an error term that related measured frequency shifts to those predicted by a model based on a local stiffness reduction. Farrar *et al.* (1994) applied the shifts in natural frequencies to detect damage on an I-40 bridge and observed that shifts in the natural frequencies are not sufficient to be used for detecting

damage of small faults. To improve the accuracy of natural frequency method, it was found to be more feasible to conduct the experiment in controlled environments where the uncertainties of measurements are relatively low. One example of such a controlled environment used resonance ultrasound spectroscopy to measure the natural frequencies and determine the out-of-roundness of ball bearings [Migliori *et al.*, 1993].

Damage in different regions of a structure may result in different combinations of changes in the natural frequencies. As a result, multiple shifts in the natural frequencies can indicate the location of damage. Messina *et al.* (1996; 1997) successfully used the natural frequencies to locate single and multiple damage in a simulated 31-bar truss and tabular steel offshore platform. Damage was introduced to the two structures by reducing the stiffness of the individual bars by up to 30%. This technique was experimentally validated on an aluminium rod test structure, where damage was introduced by reducing the cross-sectional area of one of the members from 7.9 mm to 5.0 mm.

To improve the ability of the natural frequencies to detect faults of small magnitude, high-frequency modes, which are associated with local responses, may be used. There are two main problems with working with high frequency modes. Firstly, modal overlap is high; and secondly, high frequency modes are more sensitive to environmental conditions than low frequency modes.

(b) Damping ratios

The use of damping ratios to detect the presence of damage in structures has been applied most often to composite materials. Lifshitz and Rotem (1969) investigated changes caused by damage to dynamic moduli and damping of quartz particle filled resin specimens having either epoxy or polyester as the binder. They introduced damage by applying a static load and it was observed that damping was more sensitive to damage than to the dynamic moduli. Schultz and Warwick (1971) also observed that damping was more sensitive to faults than natural frequencies in glass-fiber-reinforced epoxy beams. Lee *et al.* (1987) studied the damping loss factors for various types of damage cases in Kevlar/epoxy composite cantilevered beams. It was found that damping changes were difficult to detect where damage was introduced by milling two notches of less than 5% of the cross-sectional area. However, it was also found that the damping factors

were sensitive when damage was introduced through the creation of delaminations by gluing together two pieces of glass/epoxy and leaving particular regions unglued.

Lai and Young (1995) observed that the delamination of graphite/epoxy composite materials increases the damping ratio of the specimen. They also observed that the damping ratios decrease significantly when the specimen is exposed to humid environments for a prolonged period.

(c) Mode shapes

West (1984) used the Modal Assurance Criterion (MAC) [Allemang and Brown, 1982], a criterion that is used to measure the degree of correlation between two mode shapes, to locate damage on a Space Shuttle Orbiter body flap. By applying the MAC, the mode shapes before damage were compared to those after damage. Damage was introduced using acoustic loading. The mode shapes were partitioned and changes in the mode shapes across various partitions were compared.

Kim *et al.* (1992) used the Partial MAC (PMAC) and the Co-ordinate Modal Assurance Criterion (COMAC) proposed by Lieven and Ewins (1988) to isolate the damaged area of a structure. Mayes (1992) used the mode shape changes for fault localisation by using the Structural Translational and Rotational Error Checking (STRECH). STRECH is obtained by taking the ratios of the relative modal displacements from undamaged and damaged structures and is a measure of the accuracy of the structural stiffness between two different structural degrees of freedom.

Salawu (1995) introduced a global damage integrity index, based on a weighted-ratio of the natural frequencies of damaged to undamaged structures. The weights were used to indicate the sensitivity of each mode to damage.

The main drawbacks of the modal properties are:

- they are computationally expensive to identify;
- they are susceptible to additional noise due to modal analysis;
- they do not take into account out-of-frequency-bandwidth modes;
- they are only applicable to lightly damped and linear structure.

However, the modal data have the following advantages:

- they are easy to implement for damage identification;
- they are most suitable for detecting large faults;
- they are directly linked to the topology of the structure;
- they highlight important features of the dynamics of the structure.

1.2.2. Frequency domain data

The measured excitation and response of the structure are transformed into the frequency domain using Fourier transforms [Ewins, 1995]. The ratio of the response to excitation in the frequency domain at each frequency is called the Frequency Response Function (FRF). Recently, the direct use of the FRFs to identify faults has become a subject of research [Sestieri and D'Ambrogio, 1989]. D'Ambrigio and Zobel (1994) directly applied the FRFs to identify the presence of faults in a truss-structure.

Imregun *et al.* (1995) observed that the direct use of FRFs to identify faults in simulated structures offers certain advantages over the use of modal properties. Lyon (1995) and Schultz *et al.* (1996) have advocated the use of measured FRFs for structural diagnostics.

FRFs are difficult to use in that they contain more information than is needed for damage identification. There is also no method to choose the frequency bandwidth of interest, and they are generally noisy in the anti-resonance regions. Yet, FRF-methods have the following advantages:

- measured data include the effects of out-of-frequency-bandwidth modes;
- one measurement provides abundant data;
- modal analysis is not required and therefore modal identification errors are avoided;
- FRFs are applicable to structures with high damping and modal density.

These techniques have shown some promise but extensive research still needs to be conducted on how FRFs could best be used for fault identification.

1.2.3. Time-frequency data

Some types of damage, such as cracks caused by fatigue failures, cause linear structures to become non-linear. In these cases, methods such as linear finite element analysis and modal analysis are no longer applicable and non-linear approaches are needed. Non-linear structures give vibration data that are non-stationary. A non-stationary signal is the one whose frequency components vary as a function of time.

Examples of non-stationary signal include noise and vibration from accelerating traffic. In order to analyse non-stationary signal the use of Fast Fourier Transform (FFT) method, which only shows the frequency components of the signal and is adequate for analysing stationary signals, is not sufficient. As a result, time-frequency methods that simultaneously show the time and frequency components of the signals are required. Some of the time-frequency methods that have been used for damage identification are: the Short-Time Fourier Transform (STFT) [Newland, 1993], Wavelet Transform (WT) [Daubechies, 1987] and Wigner-Ville Distribution (WVD) [Wigner, 1932].

The STFT essentially transforms a small time window into frequency domain. The time window is shifted to a new position and the transform is repeated again. By so doing a time-frequency spectrum is obtained. The shorter the time window, the better the time domain resolution and the worse the frequency resolution and vice versa. The time-frequency spectrum obtained from the STFT is therefore limited in that any increase in the frequency resolution is at the expense of the time resolution. This limitation describes a principle called the Uncertainty Principle, which is analogous to Heisenberg Uncertainty Principle [Wheeler and Zurek, 1983], and in the context of signal processing it may be understood as the consequence of generating a linear representation of a potentially non-linear signal. The STFT is said to be linear because when calculating it the integral includes a single, linear function of the signal and it is said to be time-invariant because the time shifted version of the signal results only in the time shifting of the time-frequency representation. Furthermore, the STFT is optimal for signals with a linearly increasing phase.

The WVD was developed by Wigner (1932) in the context of quantum mechanics and was brought to signal processing by Ville (1947). The WVD is based on the calculation of a

correlation of a signal with itself (autocorrelation) to give an energy density. The Fourier transform of the calculated energy density gives the WVD. The WVD is said to be bilinear because it uses two linear functions of the signal being analysed, as opposed to one for the STFT, when calculating it. It provides optimal representation of linear frequency modulation signals such as stationary frequency case. The advantages of the WVD are that it is optimised in the time and frequency domain and that non-stationary signals exhibit reduced distortion. The disadvantages of the WVD are that it does not cover the local behaviour of the data at a given time and introduces cross-terms when the signal being analysed has many frequency components. The other disadvantage, which is reported by Cohen [1989], is that this distribution propagates noise. It has been shown that if there is noise present in a small section of a signal, it appears again within the WVD spectrum and this is related to the interference caused by cross-terms. The other problem with the WVD is that negative amplitude values may be obtained in the results and this is physically irrelevant making the results obtained from the WVD difficult to interpret.

The WT breaks down the signal into a series of basis functions called wavelets located at different positions in the time axis the same way that the Fourier transform decomposes the signal into harmonic components. A given wavelet decays to zero at a distance away from its centre. Local features of a signal can be identified from the scale, which is similar to frequency, and the position in the time axis of the wavelets into which it is decomposed. The wavelet analysis permits the construction of orthonormal bases with good time-frequency resolution. Wavelets have an advantage that they are able to identify local features of a signal from the frequency and the position in the time axis of the wavelets while the WVD does not truly describe the character of a signal at a given time because it gives equal degree of importance between the far away times and the near times making it non-local. The limitation of the wavelet approach is that frequency is logarithmically scaled and, as a result, low resolution is obtained at high frequencies [Barschdorf and Femmer, 1995].

Surace and Ruotolo (1994) applied complex Morlet WTs to identify damage in a finite element simulation of a cantilevered beam. The authors found that for damage simulated by the

reduction of 20% to 45% of the beam thickness, the amplitude of the WTs exhibited modulations that were consistent with the opening and closing of the crack.

Prime and Shevitz (1996) studied experimental data of a cantilevered beam with fatigue cracks of various magnitudes and found that the 'harmonic mode shapes' are more sensitive to crack depth and location than conventional mode shapes. The harmonic mode shapes are calculated using the magnitudes of harmonic peaks in the cross-power spectra. They observed that the Wigner-Ville transforms are more sensitive to non-linearity than the Fourier transforms.

One weakness of the time-frequency approach is that there are many types, including WT, WVD and STFT, and there is no systematic method to choose the most appropriate type for damage identification. However, comparative studies have shown that wavelet transforms are better suited for a damage detection problem than the WVD and STFT [Newland, 1994]. Nevertheless, time-frequency approaches have the following advantages:

- one measurement provides abundant data;
- they are effective in identifying damage resulting in the loss of linearity of a structure.

In this section three different domains in which vibration data may be presented were reviewed. The main outstanding question is whether these data contain similar information, and if they do not contain the same information, how the information should be integrated such that it provides the best identification of faults.

1.3. STRATEGIES IMPLEMENTED

This section describes the most common strategies that have been used for damage identification in structures using vibration data in various domains. The three strategies reviewed are the non-model methods, finite element updating methods and neural network methods.

1.3.1. Non-model based methods

Non-model based methods use vibration data in the frequency or modal domains in order to identify faults without the use of mathematical models. They are computationally cheaper to implement than the methods that use mathematical models. The MAC [Allemang and Brown,

1982] and the COMAC [Lieven and Ewins, 1988], which are a measure of correlation between mode shapes, have been used to identify damage [West, 1984; Fox, 1992; Kim *et al.*, 1992; Ko *et al.*, 1994; Salawu and Williams, 1995; Lam *et al.*, 1995]. Pandey *et al.* (1991) used absolute changes in the displacement mode shape curvature to indicate faults in a finite element model of a beam. The curvature was calculated using a central difference approximation method. Messina *et al.* (1998) introduced the Multiple Damage Location Assurance Criterion (MDLAC), which used the correlation between the natural frequencies from undamaged and damaged structures to identify the location and size of damage.

A different method was employed by Maia *et al.* (1997) and Silva *et al.* (1998), whose Frequency-Response-Function-Curvature-Method used the difference between curvatures of damaged and undamaged structures to identify damage. The Response-Function-Quotient-Method used quotients between the FRFs at different locations for damage detection [Maia *et al.*, 1999]. Gawronski and Sawicki (2000) used modal norms to successfully locate damage in structures. The modal norms were calculated from the natural frequencies, modal damping and modal displacements at the actuator and sensor locations of healthy and damaged structures. Finally, Worden *et al.* (2000) applied outlier analysis to detect damage on various simulated structures and a carbon fibre plate by comparing the deviation of a transmissibility-function-signal from what is considered normal.

1.3.2. Finite element updating methods

Finite element (FE) model updating has been used to detect damage on structures [Friswell and Mottershead, 1995; Mottershead and Friswell, 1993, Maia and Silva, 1996]. When implementing FE updating methods for damage identification, it is assumed that the FE model is a true dynamic representation of the structure. This means that changing any physical parameter of an element in the FE model is equivalent to introducing damage in that region. There are two approaches used in FE updating: direct methods and iterative methods. Direct methods, which use the modal properties, are computationally efficient to implement and reproduce the measured modal data exactly. They do not take into account the physical parameters that are updated.

Consequently, even though the FE model is able to predict measured quantities, the updated model is limited in the following ways:

- it may lack the connectivity of nodes − connectivity of nodes is a phenomenon that occurs naturally in finite element modelling because of the physical reality that the structure is connected;
- the updated matrices are populated instead of banded − the fact that structural elements are only connected to their neighbours ensures that the mass and stiffness matrices are diagonally dominated with few couplings between elements that are far apart;
- there is a possible loss of symmetry of the matrices.

Iterative procedures use changes in physical parameters to update FE models and produce models that are physically realistic. In this section, direct and indirect methods that use the FRFs or modal properties for FE updating are examined.

(a) Matrix update methods

Matrix update methods are based on the modification of structural model matrices, for example the mass, stiffness and damping matrices, to identify damage in structures [Baruch, 1978]. They are implemented by minimising the distance between analytical and measured matrices. The difference between updated matrices and original matrices identifies damage. One way of implementing this procedure is to formulate the objective function to be minimised, place constraints on the problem such as retaining the orthogonal relations of the matrices [Ewins, 1995] and choose an optimisation routine. These methods are classified as iterative because they are implemented by iteratively modifying the relevant parameters until the error is minimised.

Ojalvo and Pilon (1988) minimised the Euclidean norm of the residual force for the i^{th} mode of the structure by using the modal properties. The residual force, E_i, is defined by the following equation of motion:

$$\{E_i\} = (-\omega_i^2[M^u] + j\omega_i[C^u] + [K^u])\{\phi_i^d\} \tag{1.1}$$

where subscripts u and d indicate undamaged and damaged respectively.

The residual force is the harmonic force with which the undamaged structure will have to be excited at a frequency of ω_i, so that the structure will respond with the mode shape $\{\phi_i^d\}$. In

Equation 1.1 $[M^u]$, $[C^u]$ and $[K^u]$ are the mass, damping and stiffness matrices of the undamaged structure. The Euclidean norm is minimised by updating physical parameters in the model.

D'Ambrogio and Zobel (1994) minimised the residual force in the equation of motion in the frequency domain as:

$$[E] = \left(-\omega^2[M^u] + j\omega[C^u] + [K^u]\right)[X^d] - [F^d] \tag{1.2}$$

where $[X^d]$ and $[F^d]$ are the Fourier transformed displacement and force matrices respectively from the damaged structure. Each column corresponds to a measured frequency point. The Euclidean norm of the error matrix $[E]$ is minimised by updating the physical parameters in the model.

The methods described in this subsection are computationally expensive. In addition, it is difficult to find a global minimum through the optimisation technique, due to the multiple stationary points, which are caused by its non-unique nature [Janter and Sas, 1990]. Techniques such as the use of genetic algorithms and multiple starting design variables have been applied to increase the probability of finding the global minimum [Mares and Surace 1996; Levin and Lieven, 1998[a]; Larson and Zimmerman, 1993; Friswell *et al.*, 1996; Dunn, 1998].

(b) Optimal matrix methods

Optimal matrix methods are classified as direct methods and employ analytical, rather than numerical, solutions to obtain matrices from the damaged systems. They are normally formulated in terms of Lagrange multipliers and perturbation matrices. The optimisation problem is posed as:

$$\text{Minimise } \{E(\Delta M, \Delta C, \Delta K) + \lambda R(\Delta M, \Delta C, \Delta K)\} \tag{1.3}$$

where E, λ and R are the objective functions, Lagrange multiplier and constraints equations, respectively; and Δ is the perturbation of system matrices. In equation 1.3, different combinations of perturbations are experimented with until the difference between the FE model and the measured results is minimised.

Baruch and Bar Itzhack (1978), Berman and Nagy (1983) and Kabe (1985) formulated equation 1.3 by minimising the Frobenius norm of the error while retaining the symmetry of the matrices. McGowan *et al.* (1990) added an additional constraint that maintained the connectivity

of the structure and used measured mode shapes to update the stiffness matrix to locate structural damage.

Zimmerman *et al.* (1995) used a partitioning technique in terms of matrix perturbations as sums of element or sub-structural perturbation matrices in order to reduce the rank of unknown perturbation matrices. The result was a reduction in the modes required to successfully locate damage. One limitation of these methods is that the updated model is not always physically realistic.

(c) Sensitivity based methods

Sensitivity based methods assume that experimental data are perturbations of design data about the original FE model. Due to this assumption, experimental data must be close to the FE data for these methods to be effective. This formulation only works if the structural modification is small (i.e. the magnitude of damage is small).

These methods are based on the calculation of the derivatives of either the modal properties or the frequency response functions. There are many methods that have been developed to calculate the derivative of the modal properties and frequency response functions. One such method was proposed by Fox and Kapoor (1968) who calculated the derivatives of the modal properties of an undamped system. Norris and Meirovitch (1989), Haug and Choi (1984) and Chen and Garba (1980) proposed other methods of computing the derivatives of the modal properties to ascertain parameter changes. They utilised orthogonal relations with respect to the mass and stiffness matrices to compute the derivatives of the natural frequencies and mode shapes with respect to parameter changes. Ben-Haim and Prells (1993) and Prells and Ben-Haim (1993) proposed selective FRF sensitivity to uncouple the FE updating problem. Lin *et al.* (1995) improved the modal sensitivity technique by ensuring that these methods were applicable to large magnitude faults.

Hemez (1993) proposed a method that assesses the sensitivity at an element level. The advantage of this method is its ability to identify local errors. In addition, it is computationally efficient. Alvin (1996) modified this approach to improve the convergence rate by utilising a

more realistic error indicator and by incorporating statistical confidence measurements for both initial model parameters and measured data.

(d) Eigenstructure assignment methods

Eigenstructure assignments methods are based on control system theory. The structure under investigation is forced to respond in a predetermined manner. During damage detection, the desired eigenstructure is the one that is measured in the test. Zimmerman and Kaouk (1992) applied these methods to identify the elastic modulus of a cantilevered beam using measured modal data. Schultz *et al.* (1996) improved this approach by utilising measured FRFs.

The one limitation of the methods outlined in this section is that the number of sensor locations is less than the degrees of freedom in the FE model. This is especially problematic since it renders the integration of the experimental data and FE model – the very basis of finite element updating fault identification methods – difficult. To compensate for this limitation, the mode shapes and FRFs are either expanded to the size of the FE model or the mass and stiffness matrices of the FE model are reduced to the size of the measured data. Among the reduction methods that have been applied are the static reduction [Guyan, 1965], dynamic reduction [Paz, 1984], improved reduced system [O'Callahan, 1989] and system-equivalent-reduction-process [O'Callahan *et al.*, 1989]. The system-equivalent-expansion-process proposed has been used to expand the measured mode shapes and FRFs. Techniques that expand the mass and stiffness matrices have also been employed [Gysin 1990; Imregun and Ewins, 1993; Friswell and Mottershead, 1995].

It has been shown that FE updating techniques have numerous limitations. Most importantly, they rely on an accurate FE model, which may not be available. Even if the model is available, the problem of the non-uniqueness of the updated model makes the problem of damage identification using FE updating non-unique. Non-uniqueness in fault identification is a phenomenon that describes a situation where more than one fault diagnosis is suggested.

1.3.3. Neural network methods

Recently, there has been increased interest in applying artificial neural networks to identify faults in structures. Neural networks approximate functions of arbitrary complexity using given data. Supervised neural networks are used to represent a mapping from an input vector onto an output vector, while unsupervised networks are used to classify the data without prior knowledge of the classes involved. The most common neural network architecture is the multilayer perceptron (MLP), trained using the back-propagation technique [Bishop, 1995]. An alternative network is the radial basis function (RBF) [Bishop, 1995].

Kudva *et al.* (1991) used MLP neural networks to identify damage on a plate. The inputs to the neural network were the readings from a strain gauge, obtained by applying a static uniaxial load to the structure, while the output was the location and size of a hole. Damage was modelled by cutting holes of diameters that varied from 12.7 mm to 63.5 mm. The authors found that neural network was able to predict the error location without failure, although difficulty was experienced in predicting the size of a hole. In cases where the neural network successfully identified the size of a hole, there was approximately 50% error.

Wu *et al.* (1992) used an MLP neural network to identify damage in a model of a three-story building. Damage was modelled by reducing member stiffness by between 50% to 75%. The input to the neural network was the Fourier transform of the acceleration data, while the output was the level of damage in each member. The network was able to diagnose damage within 25% accuracy.

Leath and Zimmerman (1993) applied MLP to identify damage on a four-element cantilevered beam, which was modelled by reducing the Young's modulus by up to 95%. The inputs to the neural network were the first two natural frequencies and the output was the Young's modulus. The neural network could identify damage to within 35% accuracy.

Worden *et al.* (1993) used an MLP neural network to identify damage in a twenty-member structure, which was modelled by removing each member. The input to the neural network was the strain in twelve members. The network was trained by using data from the FE model. When applied to experimental data, the network was usually able to detect the location of damage.

Levin and Lieven (1998[b]) applied a RBF neural network and modal properties to identify errors in the FE model of a cantilevered beam. The method was found to give good identification of faults even with a limited number of experimentally measured degrees of freedom and modes. Atalla and Inman (1998) trained a RBF neural network using FRFs in order to identify faults in structures.

The finite element updating methods discussed in Section 1.3.2 require the availability of an accurate finite element model to perform damage identification, which may not be available. Methods in section 1.3.1 avoid the need of a finite element model but are mostly only able to detect faults and do not seem to be able to locate and quantify faults well. The implementation of neural network methods does not necessarily require the availability of a finite element model but requires that the vibration data be available to train the network and are able to detect, locate and quantify faults. The use of neural network techniques circumvents the reduction of finite element degrees-of-freedom as well as the expansion of measured data, procedures that are usually implemented when finite element updating is used to identify faults. One problem that has not been covered widely in the damage identification literature is how to use neural networks to give the confidence levels on the fault identification results obtained.

1.4. ENGINEERING APPLICATIONS

The ultimate goal of a fault identification procedure is to apply the methods to real structures. These methods have been applied to beams, trusses, plates, frames, bridges, offshore structures and composite structures [Yeun, 1985; Ju and Mimovitch, 1986; Kashangaki, 1991; Kondo and Hamamoto, 1994; Wolff and Richardon, 1989; Friswell *et al.*, 1994; Saitoh and Takei, 1996; Biswas *et al.*, 1990; Mazurek and DeWolf, 1990; Vandiver, 1975, 1977; Loland and Dodds, 1976; Mayes, 1992; Grygier, 1994; Lifshitz and Rotem, 1969].

Srinivasan and Kot (1992) investigated the feasibility of assessing damage in structural systems by measuring changes in the dynamics of a freely suspended thin cylindrical shell. A notch was machined into the shell to simulate a small amount of damage. A comparison between modal parameters before and after damage had occurred was performed and it was found that the natural frequencies were not sensitive to the crack introduced. Some of the mode shapes showed

significant changes, suggesting that they were more sensitive parameters for damage detection than natural frequencies. Royton *et al.* (2000) proposed a damage localisation procedure based on the split-mode phenomenon that occurs when modes of an axisymmetric structure are separated due to faults on the structure. Further research needs to be conducted on how fault identification might be performed in axisymmetric structures. An important question to be answered is whether fault identification is more successful when symmetry is retained than when mass is added to the structure to break the symmetry.

1.5 CONTRIBUTION OF THIS WORK

The literature indicates that using the FRFs directly eliminates errors associated with modal analysis. It is not clear how the FRF data should be presented in order to enhance their ability to identify faults. In this study, a parameter called pseudo modal energy, defined as the integral of the FRF over a chosen frequency bandwidth, is proposed for damage identification. This parameter and its derivative with respect to parameter changes are mathematically derived.

The above literature review suggests that it is unclear which domain (frequency, modal or time-frequency) should be used to present the data in order to achieve successful fault identification. It is clear that the time-frequency domain is most suitable for faults that cause the structure to become non-linear. Furthermore, there has been little research into the effects when data, which theoretically should give identical fault identification when the data are presented in different domains, actually give conflicting results.

This study proposes a committee of neural networks, which employs the pseudo modal energies and modal properties to identify faults in structures. Conditions on how the individual networks should be weighted to give an optimum committee-of-networks are mathematically derived. First, the proposed approach is tested on simulated data from a cantilevered beam, which is substructured into three regions. Substructuring is applied to reduce the number of possible regions where faults could be located. The effect of adding noise to the simulated data on the performance of the proposed approach is investigated. Second, the committee-of-networks is used to identify damage in a population of cylindrical shells. Population of cylindrical shells are used because the literature suggests that there has been little research on

fault identification in a population of structures. This is important for identifying faults in an assembly line during the production of nominally identical structures.

All the neural network based fault identification techniques that have been reviewed above are deterministic in that they give the identity of faults only. This is because they have been formulated using the maximum-likelihood framework. This framework maximises the capacity of neural networks to predict the correct identification of faults when presented with data. This study proposes a neural network-based fault identification strategy, which formulates the neural network problem in a Bayesian framework. The Bayesian framework ensures that neural networks give the identities of faults and their respective confidence intervals. The Bayesian approach generates a population of feasible models and from the variance of these models the confidence level are determined.

1.6 STRUCTURE OF THIS THESIS

Chapter 2 introduces pseudo modal energies for fault identification in mechanical structures. The pseudo modal energies are mathematically derived and their sensitivities to changes in physical parameters are calculated. The pseudo modal energies are compared to the classical modal properties.

In Chapter 3 a committee-of-networks employing the pseudo modal energies and modal properties, for identification of faults, is proposed. Equations that show how each method should be weighted are derived. Neural networks are formulated using a maximum-likelihood framework and the committee-of-networks is tested on a simulated cantilevered beam. The neural networks are trained using the scaled conjugate gradient method.

Chapter 4 assesses the feasibility of using vibration data measured from a population of cylinders to identify faults in structures. This chapter outlines all the experiments conducted in this thesis and pertinent signal processing issues that are involved.

In Chapter 5 the experimental data measured in Chapter 4 are used to test the committee of neural networks formulated in the maximum-likelihood framework. The neural networks are trained using the scaled conjugate gradient method.

Chapter 6 proposes the Bayesian-formulated neural networks to perform probabilistic fault identification in a population of cylindrical shells. The neural networks are trained using the hybrid Monte Carlo method.

Chapter 7 concludes this study and recommends issues for further research. From this thesis, the papers listed in Appendix A have been published.

Chapter 2

FAULT IDENTIFICATION USING PSEUDO MODAL ENERGIES AND MODAL PROPERTIES

2.1 INTRODUCTION

As noted in Chapter 1, vibration data have been used with varying degrees of success to identify damage in structures. Three types of signals have been used to this end: modal domain e.g. the modal properties, frequency domain e.g. Frequency Response Functions (FRFs) and time-frequency domain e.g. the Wavelet Transforms (WTs). In this thesis, the FRFs and the modal properties are studied because of the class of faults considered, which does not change the linearity of the structure [Doebling *et al.*, 1996]. In this chapter, a parameter called *pseudo modal energy*, defined as the integral of the FRF over a chosen frequency bandwidth, is introduced for fault identification in structures. Previously, a parameter similar to this was used for finite element model updating [Atalla, 1996] and was treated as a numerical convenience rather than a physical parameter that is related mathematically to the modal properties. In this study, a formal use of this parameter is made in order to identify faults in structures.

The structure of this chapter is as follows: first, the analytical expressions defining the pseudo modal energies in receptance and inertance form are calculated as functions of the modal properties for the case where damping is low. Second, the expressions, which describe the sensitivities of the pseudo modal energies with respect to any structural modification, are

derived. Finally, the pseudo modal energies are used to identify faults on a simulated cantilevered beam and are compared to the modal properties. In addition the sensitivities of these parameters to noise and faults are investigated.

One issue raised in this study is how to choose the frequency bandwidths for calculating the pseudo modal energies. This issue is discussed in qualitative terms and suggestions are made on how to optimise the choice of these bandwidths. Faults are introduced to the cantilevered beam by reducing the cross-sectional area of one of the elements of the beam. Simulating faults by reducing the cross-sectional areas of the structure does not cover all possible types of faults that may occur in structures. Nonetheless, the literature suggests that it has been used successfully [Surace and Ruotolo, 1994; Manning, 1994; Rizos *et al*., 1990; Stubbs and Osegueda, 1990].

2.2 MODAL PROPERTIES

Before the pseudo modal energy is introduced, the modal properties, which have been used extensively in fault identification in mechanical systems, are reviewed [Doebling *et al*, 1996]. The modal properties are related to the physical properties of the structure. All elastic structures may be described in terms of their distributed mass, damping and stiffness matrices in the time domain through the following expression:

$$[M]\{X''\} + [C]\{X'\} + [K]\{X\} = \{F\} \tag{2.1}$$

where $[M]$, $[C]$ and $[K]$ are the mass, damping and stiffness matrices respectively, and X, X' and X'' are the displacement, velocity and acceleration vectors respectively. F is the applied force.

If equation 2.1 is transformed into the modal domain to form an eigenvalue equation for the i^{th} mode, then [Ewins, 1995]:

$$(-\bar{\omega}_i^2[M] + j\bar{\omega}_i[C] + [K])\{\bar{\phi}\}_i = \{0\} \tag{2.2}$$

where $j = \sqrt{-1}$, $\bar{\omega}_i$ is the i^{th} complex eigenvalue, with its imaginary part corresponding to the natural frequency ω_i, and $\{\bar{\phi}\}_i$ is the i^{th} complex mode shape vector with the real part corresponding to the normalized mode shape $\{\phi\}_i$.

Fox and Kapoor (1968) derived the sensitivities of the modal properties for undamped case to be:

$$\omega_{i,p} = \frac{1}{2\omega_i}\left[\{\phi\}_i^T\left[[K]_{,p} - \omega_i^2[M]_{,p}\right]\{\phi\}_i\right] \tag{2.3}$$

and

$$\{\phi\}_{i,p} = \sum_{r=1;r\neq i}^{N} \frac{\{\phi\}_r\{\phi\}_r^T}{\omega_i^2 - \omega_r^2}\left[[K]_{,p} - \omega_i^2[M]_{,p}\right]\{\phi\}_i - \frac{1}{2}\{\phi\}_i\{\phi\}_i^T[M]_{,p}\{\phi\}_i \tag{2.4}$$

In equations 2.3 and 2.4, N is the number of modes, $\omega_{i,p}=\partial\{\omega\}_i/\partial g_p$, $\{\phi\}_{i,p}=\partial\{\phi\}_i/\partial g_p$, $[K]_{,p}=\partial[K]/\partial g_p$, $[M]_{,p}=\partial[M]/\partial g_p$ and g_p represents changes in the p^{th} structural parameters. Adhikari (2000) has calculated the damped version of equations 2.3 and 2.4.

The introduction of damage in structures changes the mass and stiffness matrices. Equations 2.3 and 2.4 demonstrate that changes in the mass and stiffness matrices cause changes in the modal properties of the structure.

2.3 PSEUDO MODAL ENERGIES

The pseudo modal energies are defined as the integrals of the real and imaginary components of the FRFs over various frequency ranges that bracket the natural frequencies. The FRFs may be expressed in receptance and inertance form [Ewins, 1995]. On the one hand, receptance expression of the FRF is defined as the ratio of the Fourier transformed displacement to the Fourier transformed force. On the other hand, inertance expression of the FRF is defined as the ratio of the Fourier transformed acceleration to the Fourier transformed force. This section expresses the pseudo modal energies in terms of receptance and inertance forms in the same way as the FRFs are expressed in these forms.

2.3.1 Receptance and inertance pseudo modal energies

The FRFs may be expressed in terms of the modal properties by using the modal summation equation [Ewins, 1995]. From the FRFs expressed as a function of modal properties, pseudo modal energies may be calculated as a function of the modal properties. This is done in order to infer the capabilities of pseudo modal energies to identify faults from those of modal properties. The FRFs may be expressed in terms of the modal properties using the modal summation equation as follows [Ewins, 1995]:

$$H_{kl}(\omega) = \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} \tag{2.5}$$

Here H_{kl} is the FRF due to excitation at k and measurement at l and ζ_i is the damping ratio corresponding to the i^{th} mode. Here it is assumed that the system is proportionally damped. This assumption is valid if the structure being analysed is lightly damped. Proportional damping is defined as a situation where the viscous damping matrix [C] (see equation 2.2) is directly proportional to the stiffness [K] or mass [M] matrix or to the linear combination of both.

The Receptance pseudo Modal Energy (RME) is calculated by integrating the receptance FRF in equation 2.5 as follows:

$$RME_{kl}^{q} = \int_{a_q}^{b_q} H_{kl} d\omega = \int_{a_q}^{b_q} \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} d\omega \tag{2.6}$$

In equation 2.6, a_q and b_q represent respectively the lower and the upper frequency bounds for the q^{th} pseudo modal energy. The lower and upper frequency bounds bracket the q^{th} natural frequency. By assuming light damping ($\zeta_i << 1$), equation 2.6 is simplified to give [Gradshteyn *et al*., 1994]:

$$RME_{kl}^{q} \approx \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i j}{\omega_i} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \tag{2.7}$$

The most commonly used techniques to measure vibration data measure the acceleration response instead of the displacement response [Doebling *et al.*, 1996]. In such a situation, it is better to calculate the Inertance pseudo Modal Energies (IMEs) as opposed to the RMEs calculated in equation 2.7.

The inertance pseudo modal energy is derived by integrating the inertance FRF (see [Ewins, 1995] for the definition of inertance) written in terms of the modal properties by using the modal summation equation as follows:

$$IME_{kl}^{q} = \int_{a_q}^{b_q} \sum_{i=1}^{N} \frac{-\omega^2 \phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} d\omega \tag{2.8}$$

Assuming that damping is low, equation 2.8 becomes [Gradshteyn *et al*., 1994]:

$$IME_{kl}^{q} \approx \sum_{i=1}^{N} \left[\phi_k^i \phi_l^i (b_q - a_q) - \omega_i \phi_k^i \phi_l^i j \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \right] \tag{2.9}$$

Equation 2.9 reveals that the inertance pseudo modal energy may be expressed as a function of the modal properties. The inertance pseudo modal energies may be calculated directly from the

FRFs using any numerical integration scheme. This avoids going through the process of modal extraction and using equation 2.9.

The advantages of using the pseudo modal energies over the use of the modal properties are:

- all the modes in the structure are taken into account as opposed to using the modal properties, which are limited by the number of modes identified;
- integrating the FRFs to obtain the pseudo modal energies smoothes out the zero-mean noise present in the FRFs.

In this section the pseudo modal energies have been mathematically derived. The next step is to calculate their sensitivities to structural changes the same way Fox and Kapoor (1968) calculated the sensitivities of the modal properties with respect to parameter changes.

2.3.2 Sensitivities of pseudo modal energies

In this section the sensitivity of pseudo modal energies to parameter changes is assessed. This gives some insights into how these parameters are affected by the presence of faults in structures. Because the pseudo modal energies have been derived as functions of the modal properties, these sensitivities are calculated as functions of the sensitivities of the modal properties. The sensitivity of the RMEs are determined by calculating the derivative of equation 2.7 with respect to the p^{th} structural changes to give the following expression:

$$\mathrm{RME}^{q}_{kl,p} \approx \sum_{i=1}^{N} \left[\begin{array}{l} \left\{ \dfrac{j}{\omega_i} [\phi^i_{k,p}\phi^i_l + \phi^i_k \phi^i_{l,p}] - \dfrac{1}{\omega_i^2} j\phi^i_k \phi^i_l \omega_{i,p} \right\} \ldots \\ \left\{ \arctan\left(\dfrac{-\zeta_i\omega_i - jb_q}{\omega_i} \right) - \arctan\left(\dfrac{-\zeta_i\omega_i - ja_q}{\omega_i} \right) \right\} \ldots \\ -\left\{ \dfrac{\phi^i_k \phi^i_l}{\omega_i} \right\} \left\{ \dfrac{b_q \omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + jb_q)^2} - \dfrac{a_q \omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + ja_q)^2} \right\} \end{array} \right] \tag{2.10}$$

Equation 2.10 is obtained by assuming that $\partial\zeta_i/\partial g_p = 0$ and that $\zeta^2_i \approx 0$. In this chapter, faults are introduced by reducing the cross-sectional area of the beam and in later chapters by drilling holes in structures. Introducing faults this way has been found not to change the damping properties of

the structure, thereby justifying the assumption that damping is independent of faults, which is used to derive equation 2.10.

Equation 2.10 shows that the sensitivity of the RME is a function of the natural frequencies, the damping ratios, the mode shapes and the derivatives of the natural frequencies and mode shapes. Substituting the derivatives of the modal properties [Adhikari, 2000] into equation 2.10 gives the sensitivity of the pseudo modal energies in terms of the mass and stiffness matrices, which are directly related to the physical properties of the structure.

The derivative of the IME [equation 2.9] with respect to the p^{th} parameter changes may be written as follows as derived in Appendix B:

$$\mathrm{IME}^{q}_{kl,p} \approx \sum_{i=1}^{N} \left[\begin{array}{l} (b_q - a_q)(\phi^i_{k,p}\phi^i_l + \phi^i_k\phi^i_{l,p}) - j\omega_{i,p}\phi^i_k\phi^i_l \left\{ \arctan\left(\frac{-\zeta_i\omega_i - jb_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i\omega_i - ja_q}{\omega_i} \right) \right\} \ldots \\ -j\omega_i\,(\phi^i_{k,p}\phi^i_l + \phi^i_k\phi^i_{l,p}) \left\{ \arctan\left(\frac{-\zeta_i\omega_i - jb_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i\omega_i - ja_q}{\omega_i} \right) \right\} \ldots \\ +\omega_i\,\phi^i_k\phi^i_l \left\{ \frac{b_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + jb_q)^2} - \frac{a_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + ja_q)^2} \right\} \end{array} \right] \quad (2.11)$$

Similarly, equation 2.11 may be expressed in terms of the mass and stiffness matrices by substituting the derivatives of the modal properties [Adhikari, 200] into equation 2.11.

In this section the receptance pseudo modal energies and the inertance pseudo modal energies have been derived (see Appendix B) and their respective sensitivities have been calculated. It is shown how these parameters are related to the modal properties as well as the mass and stiffness matrices. It is found that the sensitivities of the receptance pseudo modal energies and the inertance pseudo modal energies depend upon the sensitivities of the modal properties. From here onwards the word *pseudo modal energy* (PME) will be used mainly to describe the inertance pseudo modal energy.

By analysing the pseudo modal energies it is observed that if the frequency bandwidth is too narrow, they are dominated by the behaviour of the peaks of the FRFs. This is undesirable because near the peaks, factors such as damping ratios, which show high degrees of uncertainty, dominate the dynamics of the pseudo modal energies. At the same time, if the bandwidth is too wide, the influence of the anti-resonances, which are sensitive to noise, dominates the

characteristics of the pseudo modal energies. An optimal bandwidth is that which is sufficiently narrow to capture the characteristics of the peaks but adequately wide to smooth out the zero-mean noise in the FRFs. It must not be so wide, however, that it includes the anti-resonances.

Equations 2.7, and 2.9-2.11 show that the pseudo modal energies depend on the modal properties and the frequency bounds chosen. This implies that as long as the FRF information contain the modal properties, then it does not matter how many frequency points are included in the calculation of the pseudo modal energies. Here it should be noted that the number of frequency points is a separate issue from the frequency bandwidth. On calculating the pseudo modal energies the smallest number of frequency points must be used and this minimises the errors in the FRFs that are propagated into the pseudo modal energies. In other words, for a given frequency bandwidth for calculating the pseudo modal energies, increasing the number of frequency points in the bandwidth beyond a certain threshold does not necessarily add any additional information about the dynamics of the system. It should be noted that the dynamics of the system is the source of information that indicate the presence or the absence of faults. The details of the calculations performed in this section are shown in Appendix B.

Now that expressions for the pseudo modal energies and their respective sensitivities have been derived, as shown in Appendix B, the next objective is to verify these equations. This is done by comparing their sensitivities with respect to the presence of faults in structures calculated directly from the FRFs to those calculated from the modal properties.

2.4 CANTILEVERED BEAM

A simulated aluminium beam, studied in this section, is shown in Figure 2.1. The beam is fixed on one end and free on the other end. The motion of the beam is restricted to displacement in the y-axis and rotation about the z-axis. A finite element model with 9 standard Bernoulli-Euler elements based on linear interpolations for traction and torsion and cubic interpolations for flexion is constructed using the Structural Dynamics Toolbox [Balmès, 1997[a]] which runs in MATLAB [Mathworks, 1992]. Using this toolbox, the geometry of the beam is specified by defining the nodes of the beam and an arbitrary reference node to allow a distinction between the principal directions of the beam cross-section. Since the beam has been divided into 9 elements,

11 nodes are defined (this includes the reference and clamped nodes). The model description indicating the connections between the nodes as well as the material description and section properties between the nodal connections are specified. From these properties the mass [M] and stiffness [K] matrices are assembled. Since there are 9 active nodes (node 1 is clamped and node 11 is a reference point), each with 2 degrees of freedom (displacement in the y-axis and rotation about the z-axis), the mass [M] and stiffness [K] matrices are of size 18 by 18. The MATLAB *eig* function is used to calculate the eigenvalues and eigenvectors of equation 2.2. It should be noted that on calculating these eigenproperties in equation 2.2 the damping matrix [C] is set to a null matrix. This assumption is valid for structures where damping is low, which covers quite a great deal of structures. If [M] is positive definite and [K] is positive semi-definite, then solving the eigenvalue problem (equation 2.2 with [C]=[0]) gives N independent eigenvectors $\{\bar{\phi}\}_i$ and eigenvalues ω_i^2 (where ω_i^2 forms a diagonal matrix). Here N = 18 because the structure under consideration has been constructed such that it has 18 degrees of freedom. The eigenvalues are mass normalised such that $[\bar{\phi}]^T[M][\bar{\phi}] = I$ and $[\bar{\phi}]^T[K][\bar{\phi}] = \begin{bmatrix} \ddots & & \\ & \omega_i^2 & \\ & & \ddots \end{bmatrix}$, where I is the identity matrix and $\begin{bmatrix} \ddots & & \\ & \omega_i^2 & \\ & & \ddots \end{bmatrix}$ is a diagonal matrix.

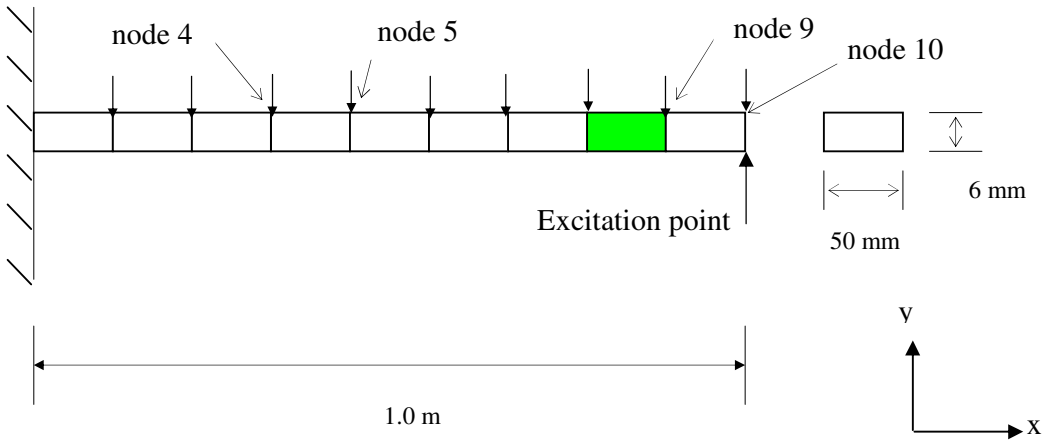


Figure 2.1. A cantilevered beam modelled with 9 Euler beam elements.

The first 5 of the 18 natural frequencies calculated by solving an eigenvalue equation are 31.4, 197.1, 551.8, 1081.4 and 1787.6 Hz. Using the 18 calculated modal properties the FRFs are computed using the modal summation equation, which is an expression inside the integral in equation 2.8. The total number of FRFs that may be calculated from the modal properties is 324 (18 by 18). In this study the FRFs corresponding to excitation at node 10 and acceleration measurements at nodes 2 to 10 are used. It should be noted that because of the principle of reciprocity [Ewins, 1995], the acceleration point and excitation point are interchangeable.

The calculated FRFs are transformed into the time domain, in which the data are contaminated with noise. The data are then transformed back to the frequency domain and the modal properties are extracted using the frequency domain modal extraction technique described in Appendix C. In summary, the FRFs are calculated from the finite element model by following this procedure:

(1) Using the mass and stiffness matrices from the finite element model, calculate the first 18 modal properties by solving the eigenvalue problem in equation 2.2 while assuming that [C]=[0].

(2) Assuming the damping ratios (ζ_i) of 0.001, use the modal summation equation [Ewins, 1995] to generate the FRFs of frequency bandwidth [0.8545 7000] Hz (the size of the FRFs is 8192).

(3) Generate the mirror image of the FRFs centered on 0Hz (new frequency bandwidth is [–7000 7000] Hz).

(4) On the FRFs generated in 3, perform the inverse Fast Fourier Transform [Cooley and Tukey, 1965] to obtain the impulse response function (of size 16384).

(5) Add Gaussian noise levels (±0%, ±1%, ±2%, ±3% and ±4 %) to the impulse response data.

(6) Perform the Fast Fourier Transform to the impulse response function to obtain the FRFs (of size 8192).

From the FRFs calculated in step 6, the modal properties are extracted as follows [See Appendix C]:

(1) Choose the frequency bandwidth for mode extraction ([17 1880] Hz) (first five modes).

(2) Obtain the initial estimates of the natural frequencies and damping ratios. Here it is assumed that the natural frequencies and damping ratios are known (the natural frequencies are calculated from the finite element model and damping ratios are set to be 0.001).

(3) Using the natural frequencies and the damping ratios in step 2, calculate the mode shapes.

Various fault cases are simulated by reducing the cross-sectional area of element 8 [Figure 2.1] by 5%, 10%, and 15%. For each fault case and the undamaged case the FRFs are generated. From these FRFs, the modal properties are extracted as described in the previous paragraph, and the pseudo modal energies are calculated by integrating over the FRFs using the trapezoidal rule technique. On calculating the pseudo modal energies, the following bandwidths are chosen in order to bracket the natural frequencies of the beam: 18-44, 155-240, 484-620, 1014-1151 and 1726-1863 Hz. It should be noted that these frequency ranges are chosen following the general guidelines outlined in Section 2.3.

Finite element model is run 1000 times with a fault case and the maximum noise level chosen and varying the cross-sectional area of all nine elements by ±1% about a mean following a Gaussian distribution. This introduces uncertainties to the data, which are typically observed in real measurements [Maia and Silva, 1997]. The same procedure is performed when element 8 is reduced by 5%, 10% and 15%. In other words, the abilities of the modal properties and the pseudo modal energies to detect faults of 5%, 10% and 15% reduction in the cross-sectional area of element 8, [see Figure 2.1] are assessed despite the presence of ±1% perturbations in the cross-sectional areas of all nine elements.

The sensitivities of the pseudo modal energies and the modal properties to damage are investigated in the presence of various noise levels by using the *statistical overlap factor* (SOF). The SOF between two distributions is defined as the ratio of the distance between the averages of the two distributions, to the mean of the two standard deviations. The SOF may be written mathematically as follows:

$$\mathrm{SOF} = \left| \frac{\bar{x}_1 - \bar{x}_2}{(\sigma_1 + \sigma_2)/2} \right| \tag{2.12}$$

Here $\bar{x}_1$ and $\bar{x}_2$ are the means of distributions; σ_1 and σ_2 are their respective standard deviations; and $|\bullet|$ stands for the absolute value of $\bullet$.

The statistical overlap factor was inspired by the modal overlap factor, which is widely used in modal analysis to assess the degree of modal overlap between two modes [Langley, 1999]. The higher the SOF, the better is the degree of separation between the two distributions. The influence of noise on the pseudo modal energies and modal properties, as well as the effect of excluding high-frequency modes when identifying these parameters, is investigated.

2.5 RESULTS AND DISCUSSIONS

The sample of real and imaginary parts of the FRF from the cantilevered beam, and the frequency bandwidths of integration are shown in Figure 2.2. The next subsection verifies equations 2.9 and 2.11.

2.5.1 Confirmation of equations 2.9 and 2.11

This section verifies the accuracy of equations 2.9 and 2.11. In order to do this, the pseudo modal energies are calculated from the equations using the modal properties extracted from the FRFs and from numerical integration of the FRFs. It should be noted that the FRFs used in this section are not contaminated with noise. The results of the real and imaginary parts of the pseudo modal energies are shown in Figure 2.3. These graphs show that the real and imaginary parts of the pseudo modal energies from equation 2.9 are similar to those from numerical integration. The absolute percentage errors of the real part of pseudo modal energies shown in Figure 2 is 8% for the first mode, 1% for the second mode, 0% for the third mode, 3% for the fourth mode, and 30% for the fifth mode. These errors are obtained by calculating the difference between the results obtained from direct integration and those estimated from equation 2.10 and dividing the difference by the direct integration results. The imaginary part shows the absolute percentage differences of 4% for the first mode and 0% for the second to fifth mode.

Likewise, to confirm the accuracy of equation 2.11, the sensitivities of the pseudo modal energies obtained through the use of equation 2.11 are compared to those from numerical

integration of the FRFs and the results are shown in Figure 2.4. Figure 2.4 indicates that the two procedures give similar results of the real and imaginary parts of these sensitivities.

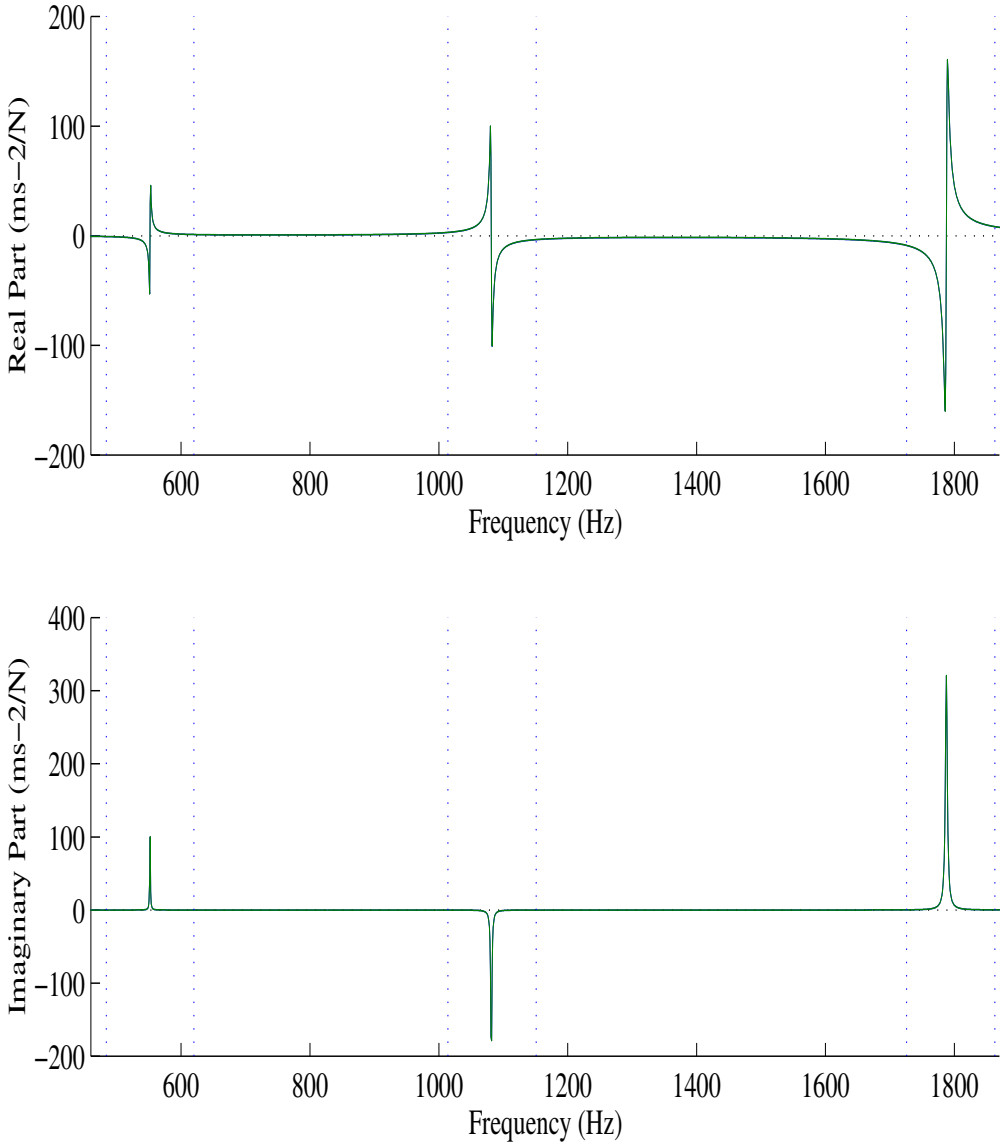


Figure 2.2. Real and imaginary parts of the FRF with no faults introduced and no noise added (the dotted lines indicate the bounds of integration). This figure shows the third to the 5^{th} modes.

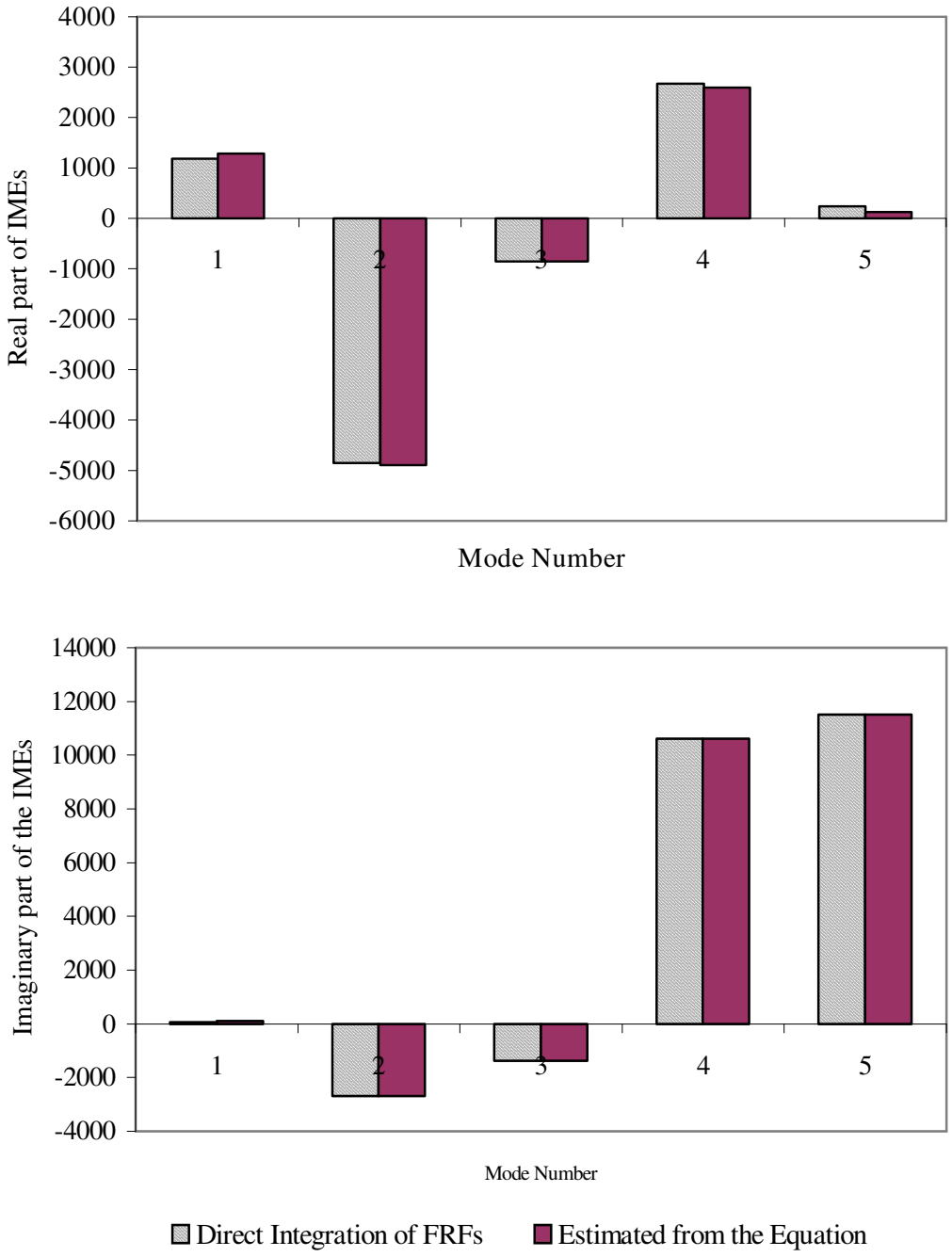


Figure 2.3. Real and imaginary parts of the pseudo modal energies obtained using direct integration of the FRF and calculated from the modal properties using equation 2.9 (the FRF is noise-free).

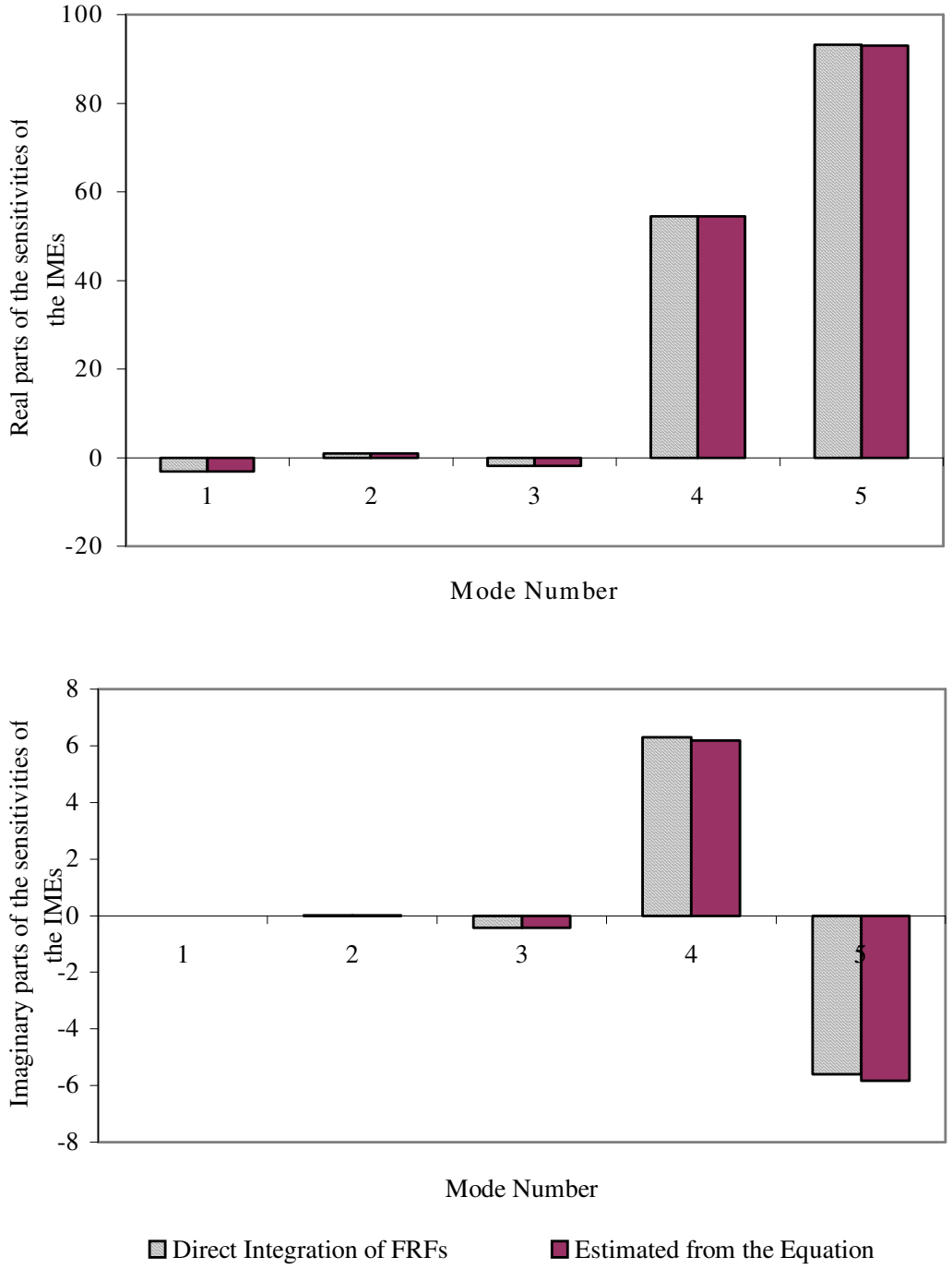


Figure 2.4. Real and imaginary parts of the sensitivities of the pseudo modal energies due to 1% reduction in the cross-sectional area of element 8, calculated from direct integration of the FRF and from equation 2.11 (the FRF is noise-free).

The absolute percentage errors between the real parts of the sensitivities of the pseudo modal energies are approximately 0% for all modes. For the imaginary part the difference is approximately 0% for the first three modes as well as 2% and 4% for the fourth and fifth modes. The reasons why some modes have higher differences than others is because of the numerical errors encountered when calculating the sensitivities of the pseudo modal energies and the assumptions made on deriving these equations. In this section, equations 2.9 and 2.11 were proven to give similar results to those obtained through direct integration of the FRFs.

Equations 2.10 and 2.11 may be used to calculate the uncertainties of the pseudo modal energies from the uncertainties of the modal properties, the extracted modal properties and the chosen frequency bandwidths. This is shown in Section B.2 in Appendix B and the results obtained show that for the chosen frequency bandwidths, the pseudo modal energies give lower uncertainties than the mode shapes and higher uncertainties than the natural frequencies. Choosing the frequency bandwidth of integration carefully when calculating the pseudo modal energies may reduce the level of uncertainties on the pseudo modal energies calculated.

2.5.2 Influence of noise on the pseudo modal energies and the modal properties

One issue that is important in fault identification is that the inevitable presence of noise in the data must be such that their effects are less influential than those of faults. It has been observed in the literature [Doebling *et al*., 1996], that for some fault types, the changes in the modal properties due to faults are more visible than the variations in measured data due to noise. This makes the modal properties viable data for fault identification. The modal properties and the pseudo modal energies before damage are calculated with various noise levels added to the time domain data as explained in Section 2.4. It should be noted that the first five modes are used for this analysis. Figure 2.5 shows the sample FRF with various noise levels added to the time domain data. Figure 2.5 indicates that the effects of noise on the FRFs are mostly observable in the anti-resonances.

The other issue to be resolved is to assess the effect of the presence of noise in the data on the modal properties and pseudo modal energies. To achieve this goal, the finite element model is

run 1000 times and each time introducing noise in the vibration data. From these data, the modal properties and the pseudo modal energies are calculated. The noise levels added to the vibration data are 1-4% Gaussian noise.

The statistical overlap factors between the distributions of the simulated data with zero noise added and those from the data simulated with 1-4% noise added are shown in Figure 2.6.

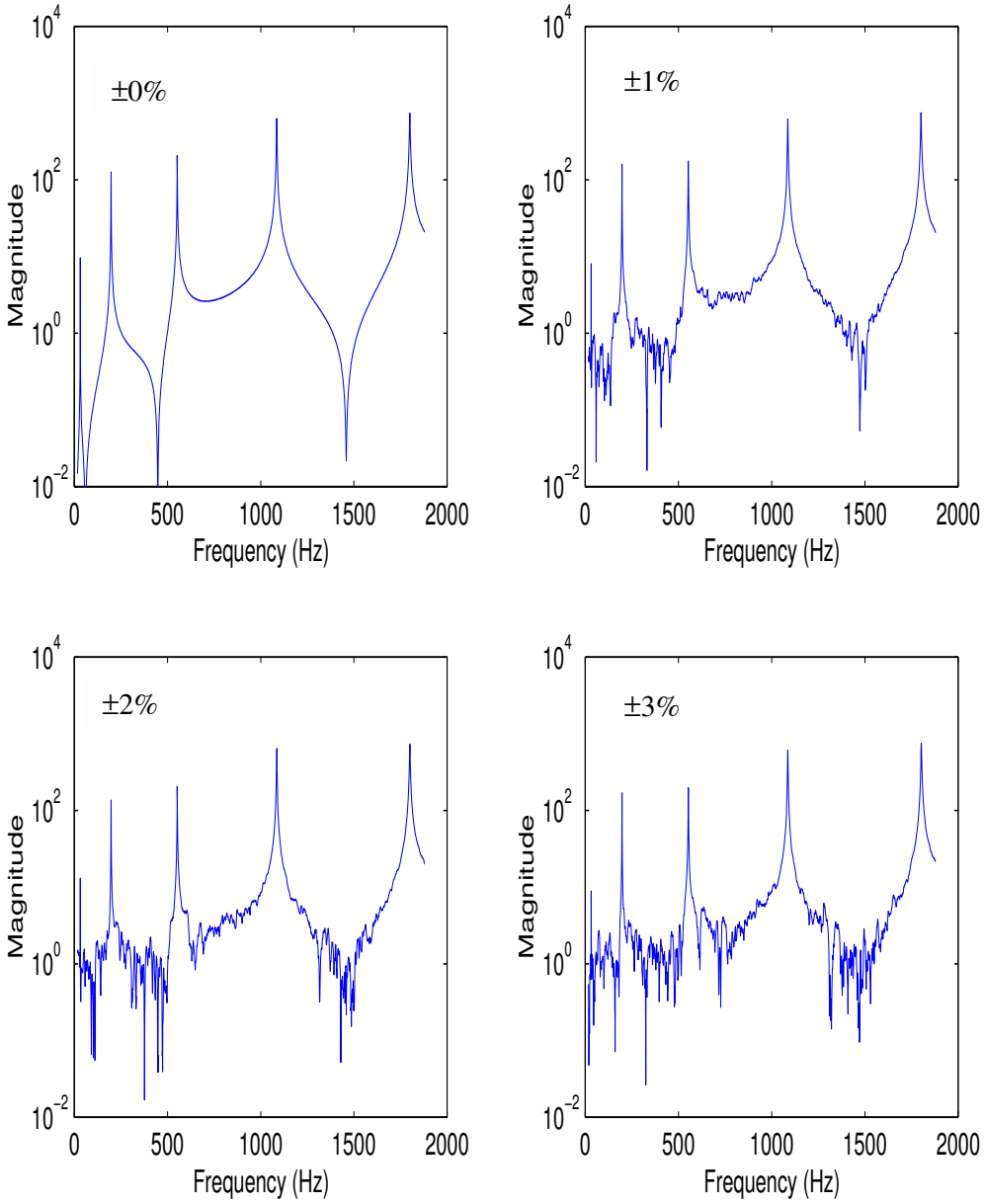


Figure 2.5. Magnitudes of the FRFs obtained using the time domain data contaminated with ±0%, ±1%, ±2% and ±3% noise.

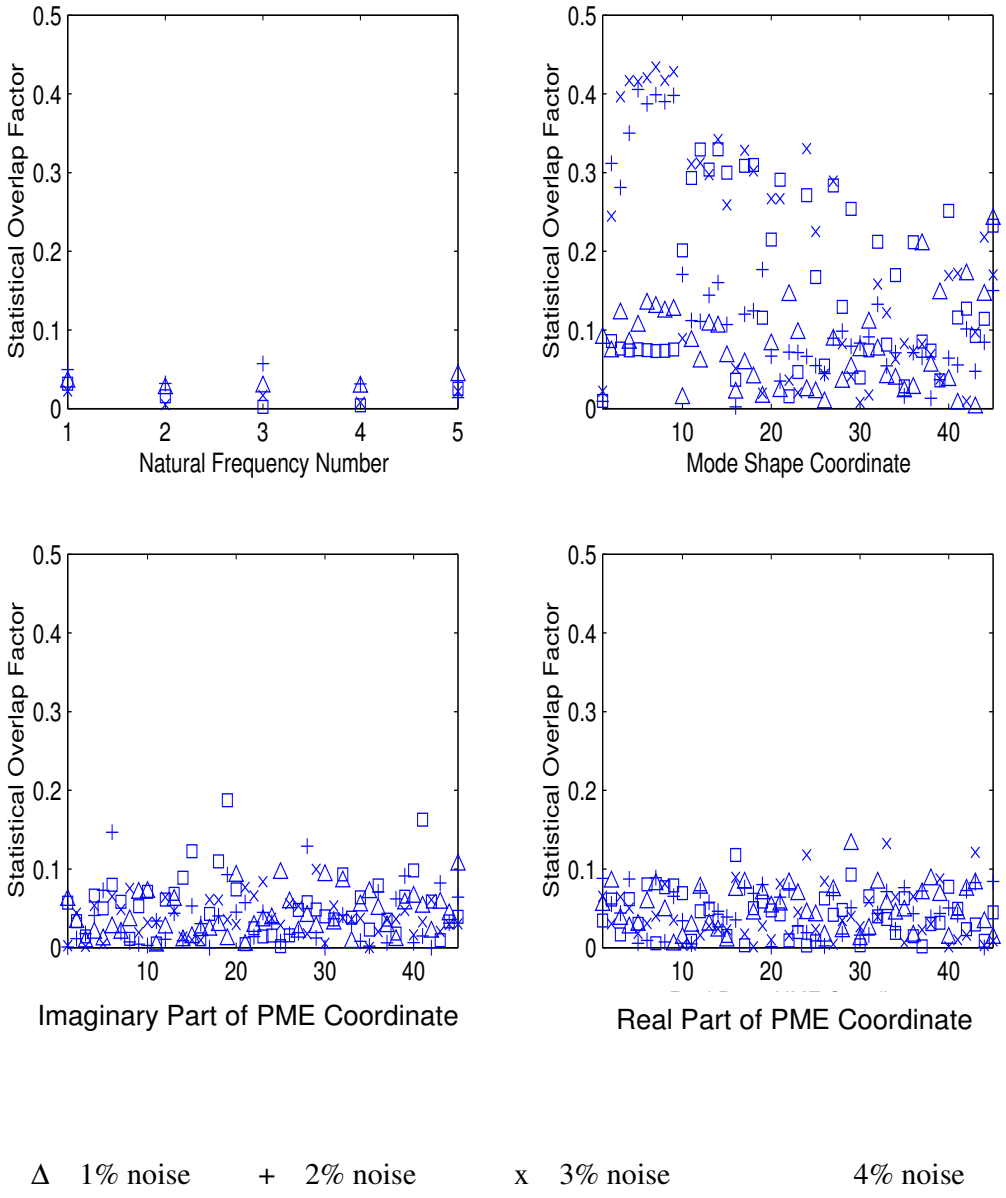


Figure 2.6. Statistical overlap factors for undamaged cases between 0% noise and various noise levels for the natural frequencies, mode shapes, and imaginary and real part of the pseudo modal energies.

Figure 2.6 indicates that the pseudo modal energies are, on average, at least two times more resistant to noise than the mode shapes. Natural frequencies are found to be the most resistant to noise than the other three parameters. It is noted below that the natural frequencies are not only resistant to noise as demonstrated in Figure 2.6, but are also insensitive to faults [Figure 2.7].

The pseudo modal energies are more resistant to noise than the mode shapes because:

- performing numerical integration on the FRFs to obtain the pseudo modal energies smoothes out the zero-mean-noise;
- the optimisation nature of modal analysis introduces additional uncertainties.

In addition, the mode shapes show high susceptibility to noise because of the inclusion of mode 1, which is noisier than other modes [Figures 2.5]. This illustrates the main limitation of modal analysis where the inclusion of a mode with high noise level compromises the identification of other modes [Balmès, 1997[b]]. This problem does not arise for pseudo modal energies because they are calculated independent of one another.

The noise in the measured data may be minimised by averaging the FRFs before the pseudo modal energies and modal properties are extracted. The averaging process improves the accuracy of the measured FRFs. The main drawback with this process is that several tens of samples need to be acquired before a smooth FRF is obtained. This may be economical if only a handful of sets of measurements are needed, but for most applications of modal analysis a great deal of measurements are required.

In this section the pseudo modal energies are found to be, on average, more resistant to noise than the mode shapes. The natural frequencies are observed to be more resistant to noise than the other three parameters. The resistance to noise of the measured data is one of the essential requirements in fault identification even though it is not a sufficient condition for successful fault identification in structures. The most essential feature is the sensitivity of faults to parameter changes, which is the subject of the next section.

2.5.3 Changes in pseudo modal energies and modal properties due to damage

In this section the pseudo modal energies are compared to the modal properties using data simulated from the cantilevered beam. It is demonstrated in equation 2.11, that the sensitivity of the pseudo modal energies, are functions of the sensitivities of the modal properties implying that these parameters are not necessarily independent of the modal properties. The cross-sectional area of element 8 in Figure 2.1 is reduced by 0%, 5%, 10% and 15%. In the presence of these reductions in the cross-sectional area of element 8, a simulation with 1000 samples is performed by perturbing the cross-sectional area of all the elements by 1% Gaussian scatter. In this chapter the word "scatter" is used to refer to the perturbation of cross-sectional area while "noise" refers to the noise added directly to the vibration data.

The four distributions of the pseudo modal energies and modal properties from the simulation are used to calculate the statistical overlap factors between 0% and 5%, 0% and 10% as well as 0% and 15% reductions in the cross- sectional areas. The pseudo modal energies and modal properties correspond to the first five modes. The results are shown in Figure 2.7. Figure 2.7 shows that the pseudo modal energies and mode shapes are sensitive to faults and that the pseudo modal energies about equally indicate faults as the mode shapes. However, the pseudo modal energies are more resistant to noise that the mode shapes (see Figure 2.6). The natural frequencies are found to be less sensitive to faults than the other three parameters. Figure 2.7 shows that, on average, the more severe the faults are the higher the statistical overlap factors.

The sample distributions of the modal properties and pseudo modal energies used to calculate the statistical overlap factors shown in Figure 2.7 are shown in Figure 2.8. Figure 2.8 indicates that the reduction in cross-sectional area of element 8 does not change the natural frequencies significantly enough to show separate distributions. Figure 2.8 shows that the mode shapes, as well as the real and imaginary parts of the PMEs, show four different distributions, which indicates that these fault cases are separable. Figure 2.8 shows that the reduction in the cross-sectional area of element 8 increases the mode shapes co-ordinate and decreases the real and imaginary part of the pseudo modal energies. The distributions of the imaginary part of the PMEs for various fault cases are the most separable followed by that of the real part of the PMEs, then the mode shapes, and finally the natural frequencies.

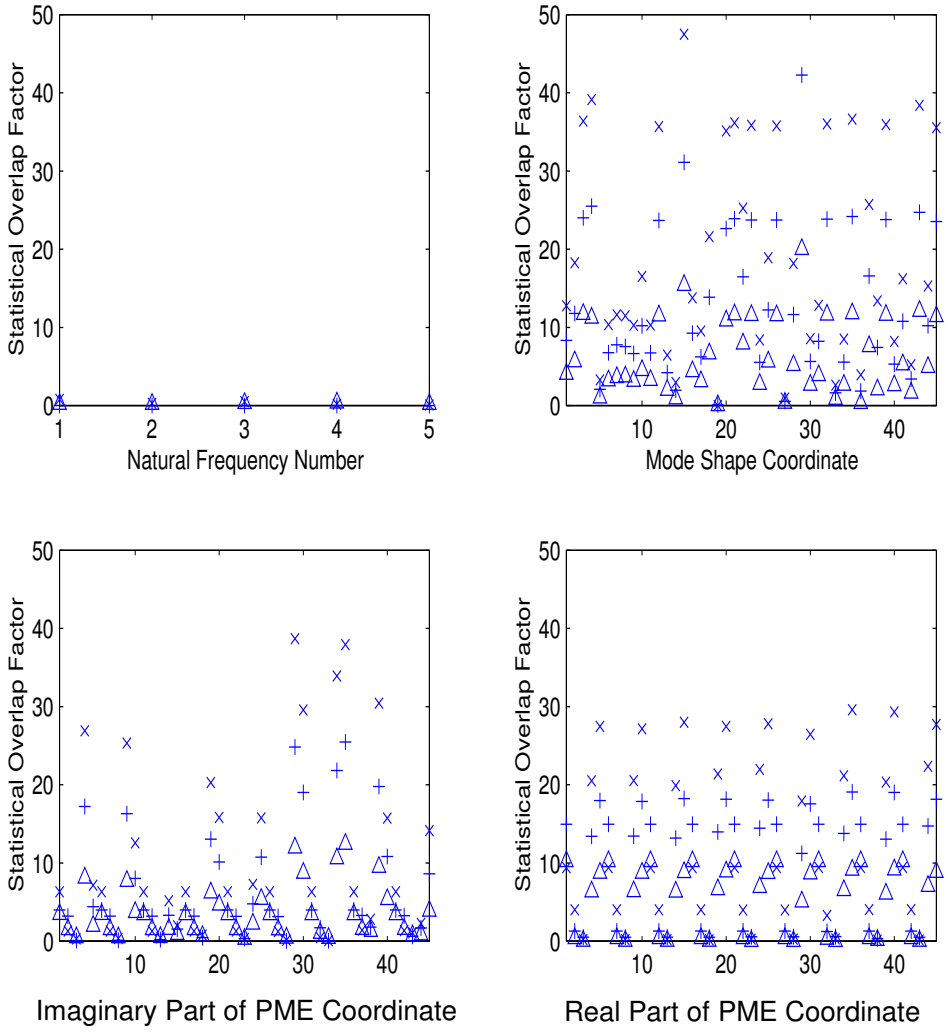


Figure 2.7. Statistical overlap factors between 0% reduction and various reductions (5, 10 and 15%) in cross-sectional area of element 8 for the natural frequencies, mode shapes, and imaginary and real parts of the pseudo modal energies due to damage. Key: PME = pseudo modal energy

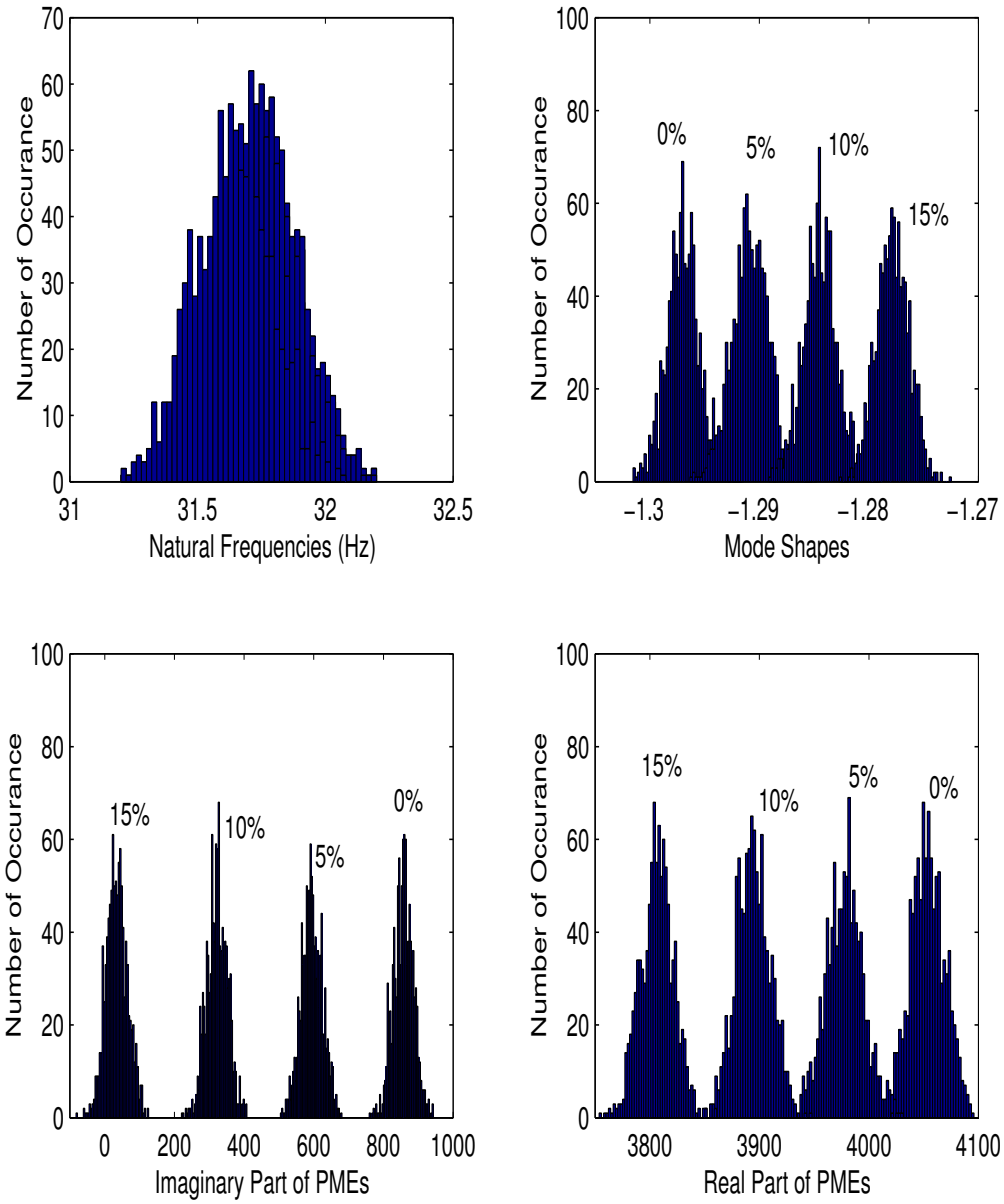


Figure 2.8. Distributions of the first natural frequency, mode shape co-ordinate, and imaginary and real parts of the pseudo modal energies for various reductions in the cross-sectional area of element 8. Key: PMEs = pseudo modal energies

The mode shapes and the pseudo modal energies in Figure 2.8 correspond to mode 4 and co-ordinate corresponding to node 3. Note that in real measurements natural frequencies are measured to the accuracy of 0.125% and mode shapes to the accuracy of 10% [Maia and Silva, 1997]. Natural frequency shown in Figure 2.8 will not be able to diagnose even 15% reduction in cross-sectional area of element 8. However, Figure 2.8 is shown to highlight the differences between various types of data.

This section has demonstrated that the pseudo modal energies are on average better indicators of faults than the natural frequencies and are equally as good indicators of faults as the mode shapes.

2.5.4 Effects of excluding high frequency modes

This section assesses the ability of the pseudo modal energies and modal properties to capture the information from the frequencies that lie outside the frequency bandwidth of interest. In this study the FRFs are calculated from the modal properties. In physical structures the FRFs are constructed from all excited modes in the structure. For simulated beam used in this work, constructing the FRFs using 18 modes is assumed to represent the real structure. Because high frequency modes are required, the FE model is re-meshed so that it contains 50 elements. This is done to improve the accuracy of high frequency modes.

The results showing the effect of excluding high frequency modes on the identification of pseudo modal energies and modal properties are shown in Figure 2.9. Figure 2.9 shows the average absolute percentage difference between data obtained using FRFs that are calculated using the first 18 modes and those obtained using the first 5 to 17 modes. The natural frequencies are not shown here because they are not sensitive to the inclusion of high frequency modes. Figure 2.9 shows that the exclusion of high frequency modes does not affect the identified mode shapes and imaginary parts of the pseudo modal energies. However, the real parts of the pseudo modal energies are found to be sensitive to the exclusion of high frequency modes. This is because as the frequency of excitation of the structure approaches infinity the real part of inertance FRF dominates the imaginary part. Due to the fact that using inertance pseudo

modal energies takes into account high frequency modes it is tempting to think that using mobility would reduce the effect of high frequency modes. However, in this thesis an anti-aliasing filter, which eliminates the effect of high frequency modes such as those observed on the real part of the pseudo modal energies, is used and therefore using mobility PMEs instead of inertance PMEs will not make any difference. The details on the use of an anti-aliasing filter are in Chapter 4.

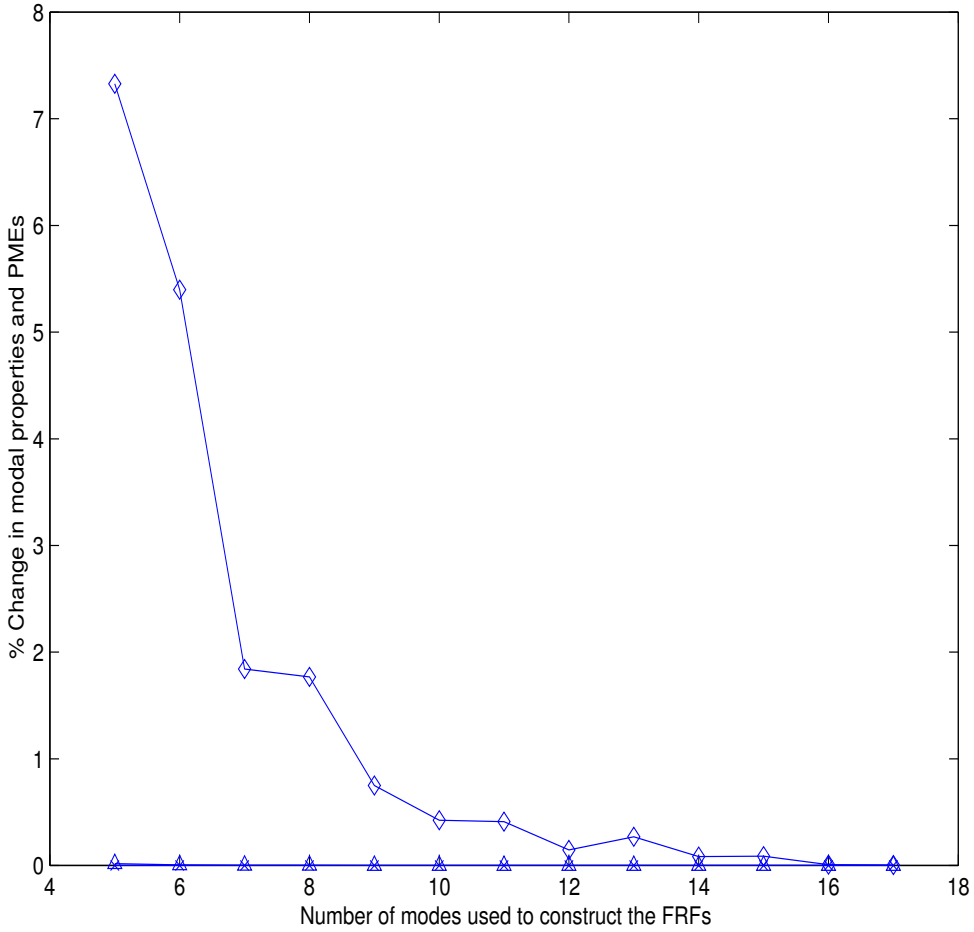


Figure 2.9. Percentage differences between the mode shapes and pseudo modal energies identified from the FRFs calculated using the first 5-17 modes and from the first 18 modes. Key: PMEs = pseudo modal energies.

2.5.5 Computational load

The time taken to compute the mode shapes using the Structural Dynamics Toolbox is 3.19 (computer process unit) CPU seconds (this does not include the time taken to obtain the first estimates of the natural frequencies) while the time taken to compute the pseudo modal energies is 0.11 CPU seconds. Running the finite element model 1000 times requires 270 CPU minutes. All the computations were performed on a Pentium 200MHz Personal Computer with 32RAM.

In this chapter the following issues are addressed:

- The pseudo modal energies are mathematically derived in terms of the modal properties.
- The sensitivities of the pseudo modal energies are derived in terms of the sensitivities of the modal properties.
- The expressions of the pseudo modal energies and their respective sensitivities are numerically verified.
- The inertance pseudo modal energies are found to be more resistant to noise than the mode shapes and more sensitive to faults than the natural frequencies.
- The pseudo modal energies are found to be computationally cheaper to calculate than the modal properties.

2.6 CONCLUSION

In this chapter, the pseudo modal energies are proposed and used to detect faults in a simulated cantilevered beam. They are then compared to the modal properties by using the statistical overlap factors and various noise levels added to the data. The pseudo modal energies are, on average, found to be more resistant to noise than the mode shapes and more sensitive to faults than the natural frequencies. Furthermore, the pseudo modal energies are, on average, found to be about equally sensitive to faults as the mode shapes. The pseudo modal energies are found to be at least 30 times faster to compute than the modal properties and are found to be a viable alternative, for fault identification, to the most widely used modal properties. The next

question to be answered is how these data could be used simultaneously in a way that combines the advantages of both these data which is the subject of the next chapter.

Chapter 3

FAULT IDENTIFICATION USING A COMMITTEE OF NEURAL NETWORKS

3.1 INTRODUCTION

Damage in structures often leads to failure. The detection, identification and quantification of faults may prevent such failures and the detrimental consequences thereof. As discussed in Chapter 1, D'Ambrogio and Zobel (1994) used the Frequency Response Functions (FRFs) directly to update a finite element model and subsequently detect faults in structures by minimising the error in the equation of motion. Baruch (1978) applied the modal property approach to identify faults in a finite element model. Marwala and Heyns (1998) introduced the multiple criterion approach, which employs both the modal properties and frequency response functions simultaneously in a finite element framework for the identification of faults. The implementation of the multiple criterion method was found to give results that are more accurate than the results obtained when either the frequency response function approach or the modal property approach is utilised in isolation.

Most recently, several researchers have focused their attention on the application of neural networks and finite element models to identify faults in structures. Levin and Lieven (1998[b]) applied a neural network employing modal property data to identify errors in a finite element model of a cantilevered beam. The method was found to work well even with limited numbers

of experimentally measured degrees of freedom and modes. This suggests that neural network approach does not require the implementation of reduction or expansion techniques such as the Guyan reduction method (1965). In addition, they observed that when data used to train a neural network are contaminated with noise, the method is robust and therefore resistant to experimental noise.

Atalla (1996) applied neural network successfully for the identification of faults in mathematical models of structures. In one experiment, the data similar to the measured pseudo modal energies studied in Chapter 2, rather than the modal data, were used to identify faults in a three-degree-of-freedom truck suspension. The results obtained suggest that the use of neural networks is an accurate method of identifying faults in finite element models.

In this chapter, a multiple criterion method [Marwala and Heyns, 1998] is extended to a neural network framework by applying the concept of a *committee-of-networks.* A committee-of-networks is a strategy, which integrates information from independently trained networks. Perrone and Cooper (1993) observed that a committee method usually performs with more accuracy than the networks used in isolation. Perrone and Cooper's as well as Marwala and Heyns' observations provide the motivation for combining both these approaches to identify faults in structures by using vibration data.

When a finite element model is applied to detect faults in structures, the ideal approach is to use all the physical parameters of every element as design variables and perturb the design variables until the model predicts the response of the faulty structure. The changes in the design variables are then utilised to infer faults in the structure. There are usually more than one set of design variables that make the model predict the measured data and this makes the problem of identifying faults from the measured vibration response non-unique [Janter and Sas, 1990]. In the context of neural networks, the problem of uniqueness has not been widely addressed, but some preliminary studies by Williamson and Helmke (1995) and Coetzee and Stonick (1996) demonstrate that for some limited classes of neural network architectures uniqueness can be mathematically proven. Williamson and Helmke (1995) studied a single input, single output and one hidden layer feedforward neural networks by reformulating this class of network in terms of rational functions of a single variable. They managed to derive conditions for the existence of a

unique best mapping between the input and the output. Coetzee and Stonick (1996) used the geometric formulation of a single layer perceptron weight optimization to describe sufficient conditions for uniqueness of weight solutions. In the present work, measures such as regularisation techniques, described later in the chapter, are implemented to improve the quality of the solutions given by the networks, which is the objective of seeking a unique solution.

Essentially, the use of finite element models to identify faults is an optimisation problem [Wei and Janter, 1988]. In optimisation, generally the greater the number of design variables, the greater the number of local optimum points. In order to avoid the high incidence of local optima making the determination of the global optimum more difficult, it is often desirable to reduce the number of design variables. The method that is proposed here to achieve this objective is the method of global substructuring.

This chapter unifies the approach formulated by Levin and Lieven (1998[b]) with that of Atalla and Inman (1998) by utilising a committee of neural networks. The proposed committee approach implements the method of global substructuring and is tested on the data from a simulated cantilevered beam. In subsequent chapters, this technique is tested on experimentally measured data from a population of cylindrical shells. The input data to the neural networks are both the pseudo modal energies and modal properties, while the output from the neural networks consists of the parameters that correspond to the level of fault in each substructure. The parameters that correspond to each substructure form a vector space, also known as the identity of faults, and this information is defined as substructure space. In other words, the information from the pseudo modal energies and modal properties are transformed into substructure space using two independent neural networks. The new procedure transforms the information from both the pseudo modal energies and modal properties into substructure space using a committee of neural networks.

The mathematical proof for applying the pseudo modal energies and modal properties [Levin and Lieven, 1998[b]; Atalla and Inman, 1998] as input basis simultaneously is demonstrated. The method is then generalised for more than two sets of input data. The effect of adding noise on simulated vibration data is subsequently examined and the behaviour of the mean- and variance-of-square-of-errors in predicting the simulated faults is investigated.

3.2 NEURAL NETWORK

In this thesis, neural networks are viewed as parameterised graphs that make probabilistic assumptions about data. Learning algorithms are viewed as methods for finding parameter values that look probable in the light of the data. Learning processes may occur by training the network through either supervised or unsupervised learning. Unsupervised learning is used when only the input data are available. To illustrate the unsupervised learning route in the context of structural dynamics, one may consider two kinds of failures in structures: failure due to loosening of joints or due to cracks in the structure. If the responses of the two failures are inherently different, an unsupervised learning scheme may be employed to distinguish these types of failures as either belonging to class 1 or 2. Supervised learning is the case where the input (x) and the output (y) are both available and neural networks are used to approximate the functional mapping between the two. In this thesis, supervised learning is applied.

There are several types of neural network procedures, some of which will be considered later, for example, multi-layer perceptron (MLP) and radial basis function (RBF) [Bishop, 1995; Tarassenko, 1998]. In this chapter, the MLP is used because it provides a distributed representation with respect to the input space due to cross-coupling between hidden units, while the RBF provides only local representation. In this study, the MLP architecture contains a hyperbolic tangent basis function in the hidden units and logistic basis functions in the output units [Bishop, 1995]. A schematic illustration of the MLP is shown in Figure 3.1 [Jordan and Bishop, 1996].

This network architecture contains hidden units and output units and has one hidden layer. The bias parameters in the first layer are shown as weights from an extra input having a fixed value of $x_0=1$. The bias parameters in the second layer are shown as weights from an extra hidden unit, with the activation fixed at $z_0=1$. The model in Figure 3.1 is able to take into account the intrinsic dimensionality of the data. Models of this form can approximate any continuous function to arbitrary accuracy if the number of hidden units M is sufficiently large. The MLP may be expanded by considering several layers but it has been demonstrated by the Universal Approximation Theorem [Haykin, 1999] that a two-layered architecture is adequate for

the multi-layer perceptron. As a result of this theorem, in this study a two-layered network shown in Figure 3.1 is chosen.

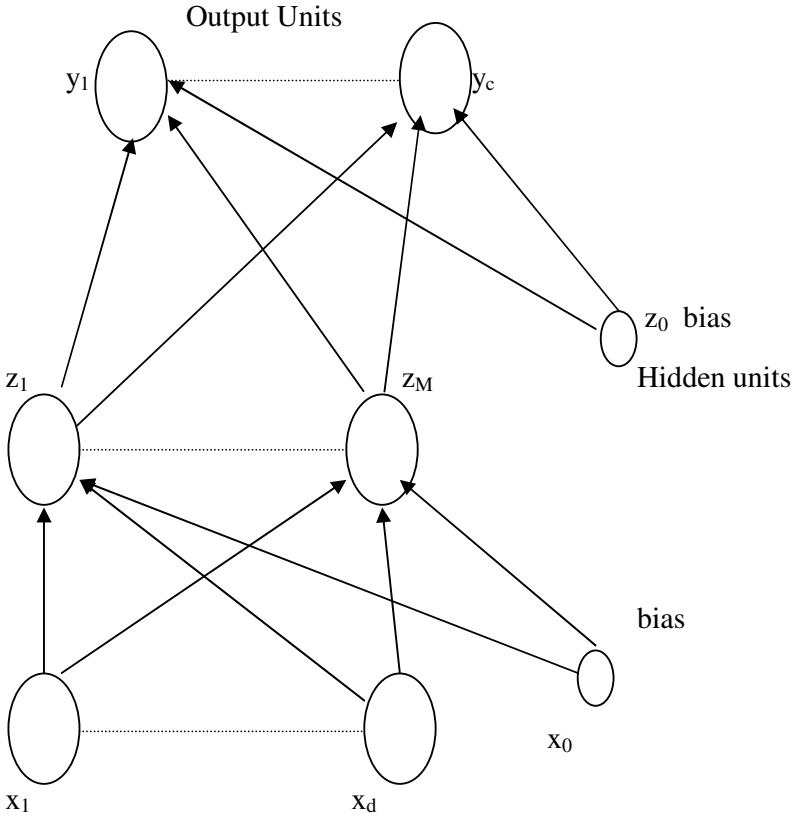


Figure 3.1 Feed-forward network having two layers of adaptive weights.

In Figure 3.1, the output of the j^{th} hidden unit is obtained by calculating the weighted linear combination of the d input values to give:

$$a_j = \sum_{i=1}^{d} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \tag{3.1}$$

Here, $w_{ji}^{(1)}$ indicates weight in the first layer, going from input i to hidden unit j while $w_{j0}^{(1)}$ indicates the bias for the hidden unit j. The activation of the hidden unit j is obtained by transforming the output a_j in equation 3.1 into z_j, which is shown in Figure 3.1, as follows:

$$z_j = f_{inner}(a_j) \tag{3.2}$$

The output of the second layer is obtained by transforming the activation of the second hidden layer using the second layer weights. Given the output of the hidden layer z_j in equation 3.2, the output of unit k may be written as:

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \tag{3.3}$$

Similarly, equation 3.3 may be transformed into the output units by using some activation function as follows:

$$y_k = f_{outer}(a_k) \tag{3.4}$$

If equations 3.1, 3.2, 3.3 and 3.4 are combined, it is possible to relate the input x to the output y by a two-layered non-linear mathematical expression that may be written as follows:

$$y_k = f_{outer}\left(\sum_{j=1}^{M} w_{kj}^{(2)} f_{inner}\left(\sum_{i=1}^{d} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right) \tag{3.5}$$

The biases in equation 3.5 may be absorbed into the weights by including extra input variables set permanently to 1 making $x_0 = 1$ and $z_0 = 1$, to give:

$$y_k = f_{outer}\left(\sum_{j=0}^{M} w_{kj}^{(2)} f_{inner}\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i \right) \right) \tag{3.6}$$

In this thesis, the function $f_{outer}(\bullet)$ is logistic while f_{inner} is a hyperbolic tangent function. The logistic function is defined as follows:

$$f_{outer}(\nu) = \frac{1}{1 + e^{-\nu}} \tag{3.7}$$

The logistic activation function maps the interval $(-\infty,\infty)$ onto a $(0,1)$ interval and can be approximated by a linear function provided the magnitude of ν is small. In other words, the linear activation function may be viewed as a special case of a logistic function. The hyperbolic tangent function is:

$$f_{inner}(\nu) = \tanh(\nu) \tag{3.8}$$

Training the neural network identifies the weights in equations 3.6. There are several different ways in which these neural networks may be trained. For example, the maximum-likelihood approach or the Bayesian methods [Bishop, 1995] may be used. In maximum-likelihood training, optimisation methods are used to identify a set of weights that maximises the ability of a network to predict the output whenever presented with the input data. The Bayesian method uses Bayes's theorem [Bishop, 1995] to identify the probability distribution of weights in the light of the training data. The maximum-likelihood method may be treated as a special case of the Bayesian method. In this chapter the maximum-likelihood method is implemented.

3.2.1 Maximum-likelihood-based cost function

An optimisation procedure is used to identify the weights and biases of the neural networks in equation 3.6 using the maximum-likelihood framework. A cost function must be chosen in order to use the optimisation technique. A cost function is a mathematical representation of the overall objective of the problem. In this thesis, the main objective, this is used to construct a cost function, is to identify a set of neural network weights given vibration data and identity of faults. If the training set $D = \{x_k, t_k\}_{k=1}^{N}$ is used and assuming that the targets t are sampled independently given the inputs x_k and the weight parameters, w_{kj}, the cost function, E, may be written as follows:

$$E = \sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk} - y_{nk}\}^2 + \frac{\alpha}{2}\sum_{j=1}^{W} w_j^2 \tag{3.9}$$

In equation 3.9, n is the index for the training pattern, k is the index for the output units, N is the number of training patterns, K is the number of output units and W is the number of weights. The sum-of-square-of-errors cost function, which is the first term in equation 3.9, tends to give similar absolute error for each pattern. As a result it performs poorly on target values of small magnitude. The second term in equation 3.9 is the regularisation parameter and it penalises weights of large magnitudes [Vapnik, 1995]. This regularisation parameter is called the weight decay and its coefficient, α, determines the relative contribution of the regularisation term to the training error. This regularisation parameter ensures that the mapping function is smooth.

The other cost function that has been used is the cross-entropy function [Hopfield, 1987; Hinton, 1989]. Minimisation of the cross-entropy function tends to give the same relative errors for small and large targets. The cross-entropy cost function plus the weight decay regularisation parameters may be written as follows:

$$E = -\sum_{n=1}^{N}\sum_{k=1}^{K}\{t_{nk}\ln(y_{nk}) + (1-t_{nk})\ln(1-y_{nk})\} + \frac{\alpha}{2}\sum_{j=1}^{W} w_j^2 \tag{3.10}$$

The cost function in 3.10 has been found to be suited for classification problems while the one in equation 3.9 has been found to be suited for regression problems [Bishop, 1995].

Including the regularisation parameter has been found to give significant improvements in network generalisation [Hinton, 1987]. In neural networks, to produce an over-fitted mapping

with regions of large curvature requires large weights. The weight decay regularisation penalises large weights thereby encouraging the weights to be small and avoiding an over-fitted mapping between the inputs and the outputs. If α is too high then the regularisation parameter over-smoothes the network weights giving inaccurate results. If α is too small then the effect of the regularisation parameter is negligible and unless other measures that control the complexity of the model, such as the early stopping method [Bishop, 1995], are implemented then the trained network becomes too complex and thus performs poorly on the validation set.

Before minimisation of the cost function is performed, the network architecture needs to be constructed by choosing the number of hidden units, M. If M is too small, the neural network will be insufficiently flexible and will give poor generalisation of the data because of high bias. However, if M is too large, the neural network will be unnecessarily flexible and will give poor generalisation due to a phenomenon known as overfitting caused by high variance [Geman et al., 1992].

The weights (w_i) and biases (with subscripts 0 in Figure 3.1) in the hidden layers are varied using optimisation methods until the cost function is minimised. Gradient descent methods are implemented and the gradient of the cost function is calculated using the back-propagation method [Bishop, 1995]. The details of the back-propagation method are found in Appendix D. Both the conjugate gradient [Shanno, 1978] and the scaled conjugate gradient methods [Møller, 1993] were implemented at the preliminary stage of this research. It was decided to pursue the scaled conjugate gradient method because it was found to be computationally efficient and yet retains the essential advantages of the conjugate gradient technique. The reason behind higher computational efficiency of the scaled conjugate gradient method over the conjugate gradient method are not the subject of this thesis but may be obtained in [Haykin, 1999]. The details of these optimisation techniques are explained in Appendix E.

3.3 THEORETICAL FORMULATION

This section is composed of the following:

(1) A brief examination of the pseudo modal energy and modal property based fault identification methods.

(2) An outline of how the substructuring method is used to reduce the order of the fault identification problem.

(3) An explanation of the relevance of a committee of neural networks for structural dynamics.

(4) An adaptation of the committee-of-networks used previously in situations where the inputs to each of the networks are the same, to a multiple criterion idea, where the inputs to the networks, the pseudo modal energies and modal properties are different.

(5) A mathematical proof that a committee-of-networks gives more accurate results than using either the modal-property-network or the pseudo-modal-energy-network individually.

(6) A derivation of the optimal weighting conditions for the two methods.

3.3.1 Pseudo modal energy method

When a fault is introduced to a structure, the dynamic properties of the structure will change. These changes may be observed on the pseudo modal energies, and were mathematically derived in Chapter 2 [see equation 2.11]. Sufficient data that defines the relationship between changes in physical parameters and changes in the pseudo modal energies may be generated from a mathematical model or from experiment. From this set of data, a functional mapping between the identity of fault y_1 and the pseudo modal energy vector ψ may be represented in the following form:

$$y_1 = h(\psi) \tag{3.11}$$

Atalla and Inman (1998) employed neural networks to quantify the relationship between the identity of fault and the parameter that is similar to the pseudo modal energies and used it for finite element updating. In this thesis, this relationship is used to identify mechanical faults in structures.

3.3.2 Modal property method

There is an implicit relationship between the physical properties of a structure and its modal properties. This relationship may be used to identify faults in structures and was calculated by Fox and Kapoor (1968). Baruch (1978) applied the modal properties to identify the correct physical properties for given measured modal data. When faults are introduced to the structure,

the physical parameters of the elements in the finite element model will change. This will result in changes in the natural frequencies and mode shapes. If a finite element model is available then sufficient data defining the relationship between changes in physical parameters and changes in the modal properties may be generated. A functional mapping between the identity of fault y_2 and the modal property vector χ may therefore be quantified by the following equation:

$$y_2 = h(\chi) \tag{3.12}$$

where $\chi=\{\omega_j \{\phi_j\}\}$.

Levin and Lieven (1998[b]) applied neural network to identify the mapping function that gives the relationship between the identity of the faults and the modal properties for finite element updating. In this thesis, this relationship is used to identify faults in structures.

3.3.3 Method of substructuring

In this section, the method of substructuring is outlined and its advantages are discussed. Throughout this chapter it is assumed that the initial finite element model has been updated subsequent to the introduction of faults and that it is able to reproduce the measured dynamic properties. Therefore, perturbing the finite element model is equivalent to introducing faults in the structure.

For notational convenience a model with two substructures is considered. If one was interested in locating damage in substructure 1 or 2, the equation of motion might be partitioned into two super elements as follows:

$$\left(-\omega_n^2\begin{bmatrix}[M_{11}] & [M_{12}]\\ [M_{21}] & [M_{22}]\end{bmatrix}+\begin{bmatrix}[K_{11}] & [K_{12}]\\ [K_{21}] & [K_{22}]\end{bmatrix}\right)\begin{Bmatrix}\phi_{n1}\\ \phi_{n2}\end{Bmatrix}=\begin{Bmatrix}0\\ 0\end{Bmatrix} \tag{3.13}$$

The substructure framework could be easily extended to the case where equation 3.13 contains the damping matrix. If substructure 1 was faulty, this would be reflected by changes in the n^{th} natural frequency and mode shape vectors $\{\phi_{n1}\}$ and $\{\phi_{n2}\}$. If a fault was in substructure 1, vector $\{\phi_{n1}\}$ would be affected more than vector $\{\phi_{n2}\}$. By comparing the relative changes in these two vectors as a result of a fault, it is possible to deduce that a fault is in substructure 1. Equation 3.13 might be decomposed into the following equations:

$$-\omega_n^2([M_{11}]\{\phi_{n1}\}+[M_{12}]\{\phi_{n2}\})+([K_{11}]\{\phi_{n1}\}+[K_{12}]\{\phi_{n2}\})=\{0\}$$
and (3.14)
$$-\omega_n^2([M_{21}]\{\phi_{n1}\}+[M_{22}]\{\phi_{n2}\})+([K_{21}]\{\phi_{n1}\}+[K_{22}]\{\phi_{n2}\})=\{0\}$$

If the changes in the natural frequencies as a result of damage are negligible, then equation 3.14 ensures that the changes are primarily in the partition of the mode shape vector containing a fault. This occurs because the presence of faults affect the mode shapes more than the natural frequencies. The analysis in equation 3.14 may be similarly performed for the frequency response functions, and thus the pseudo modal energies and should be taken into account before training the network.

Methods that use only finite element updating to identify faults are not effective in locating faults in a small area of the structure. In this section it is demonstrated that the substructuring method might be used as a first step in fault diagnostics by pointing to a large location before localised methods such as ultrasonic methods are applied. The purpose of this thesis is not to study substructuring per se, but to illustrate its role in the effective use of a committee-of-networks method.

3.4 COMMITTEE-OF-NETWORKS

The committee-of-networks is a strategy in which more than one network is used to tackle a problem. This strategy is inspired by the rule of divide-and-conquer, which has found many applications in different areas. Examples of the practical implementation of this strategy include car assembly line where individuals are trained to perform a small task towards the overall assembly of a complete car. From this example it could be deduced that there are two main components of this strategy: (1) training the individuals to become experts in their individual tasks and (2) integrating the individual tasks together to form a solid structure. In a car example the sum of these individual tasks form a complete car. In this study, the strategy of using a collection of networks to solve a particular task is applied to fault identification using neural networks. In this thesis, instead of dividing the tasks, neural networks are trained with data that have been processed differently and their outputs are combined.

The foundation of the idea of a committee of neural networks may be traced back to Nilsson (1965) and a literature review on this subject can be found in Haykin (1999) and Sharkey (1999). There are various ways in which a committee-of-networks can be framed. Here two types are discussed: (1) static and (2) dynamic structures. In a static committee structure the output are combined without any direct participation of the input data. The dynamic structure is the case where the input data are involved in the decision on how the outputs of the networks are combined. An important question in the area of a committee-of-networks is how to combine the output of the members of the committee in an optimal way. Some of the methods that have been implemented to combine the outputs of networks to form a committee are the averaged or weighted-averaged and non-linear combination methods. In the averaged approach, the outputs of the members of the committee are averaged to form a single output. This method may be modified to take into account the relative accuracy of the individual networks by calculating the weighted average. In the non-linear combination method, a non-linear function is used to combine the outputs of the networks.

An example of the framework that uses a static structure is the ensemble averaging method where the output of the committee-of-networks is an average or weighted average of the outputs of the individual networks [Perrone and Cooper, 1993; Wolpert, 1992 and Hashem, 1997]. This procedure has been found to reduce the output error, and this is explained in detail later in the chapter. Another example of a static committee is the boosting method [Schapire, 1990; Freund, 1995; Breiman, 1996] where a network that generalises poorly is converted into one that generalises well. In this chapter the two types that are discussed are boosting by filtering and adaptive boosting. Schapire (1990) first proposed boosting by filtering. The implementation of this technique involves training a network with the data that have been filtered by previously trained members of the committee. The filtered data are those that the previously trained networks do not generalise well. This method requires an abundance of training data, making it inaccessible to many engineering applications where there is no abundance of training data. To compensate for the limited number of the training data adaptive boosting (AdaBoost) technique, proposed by Freund and Schapire (1996), may be used. In this technique, the training sets are

adaptively resampled according to a probability distribution such that the contribution of the part of the training data that was most often misclassified is increased.

As mentioned earlier the other framework for creating a committee is the dynamic structure. In this structure, the input data are directly involved in deciding how the outputs of the members of the committee are combined. An example of a dynamic committee is a mixture of experts [Nowlan, 1990; Jacobs, 1990]. In this committee framework a gating network, that uses the input data and has as many outputs as the number of networks that form a committee, mediates the outputs of the networks.

In this section a static committee framework illustrated schematically in Figure 3.2 is adopted and developed. The reason for choosing such a framework is because of its simplicity yet containing powerful properties such as the ability to reduce the mean squared errors and variance. In this section it is proven that a committee-of-networks that uses both the pseudo modal energies and modal properties, gives results that are more reliable than networks that use the pseudo modal energies and modal properties separately. This section presents an adaptation and extension of the work by Perrone and Cooper (1993). In this section, three advantages of the committee method as compared to the individual methods are outlined and mathematically proven, namely:

(1) the committee gives lower mean-of-square-of-error than the individual method;

(2) the committee gives lower variance-of-square-of-error than the individual method;

(3) the committee is a more reliable fault identification procedure than the individual method.

3.4.1 The committee gives the least mean-of-square-of-error

The mapping between the identity of faults and the pseudo modal energies, as indicated by equations 3.11 and 3.12, is written as the desired function plus an error as follows:

$$\begin{aligned} y_1(\psi) &= h(\psi) + e_1(\psi) \\ y_2(\chi) &= h(\chi) + e_2(\chi) \end{aligned} \tag{3.15}$$

Here y_1 is the identity of the fault when the pseudo modal energy data are used, y_2 is the identity of the fault when the modal property data are utilised, $h(\bullet)$ is the approximated mapping function and $e(\bullet)$ is the error. For notational convenience the mapping functions are assumed to

have single outputs y_1 and y_2. Equations 3.15 can be easily adapted to multiple outputs. The average of squares of errors for model $y_1(\psi)$ and $y_2(\chi)$ may be written as follows:

$$E_1 = \varepsilon[\{y_1(\psi) - h(\psi)\}^2] = \varepsilon[e_1^2]$$
$$E_2 = \varepsilon[\{y_2(\chi) - h(\chi)\}^2] = \varepsilon[e_2^2] \quad (3.16)$$

where $\varepsilon[\bullet]$ indicates the expected value and corresponds to an integration over the input data, defined as follows:

$$\varepsilon[e_1^2] \equiv \int e_1^2(\psi)p(\psi)d\psi$$
$$\varepsilon[e_2^2] \equiv \int e_2^2(\chi)p(\chi)d\chi \quad (3.17)$$

and where $p[\bullet]$ indicates the probability density function and $d[\bullet]$ is the differential operator.

The average expected squared errors made by the two networks [Atalla and Inman, 1998; Levin and Lieven, 1998[b]] acting individually may be written as follows:

$$E_{AV} = \frac{E_1(\psi) + E_2(\chi)}{2} = \frac{1}{2}\left(\varepsilon[e_1^2] + \varepsilon[e_2^2]\right) \quad (3.18)$$

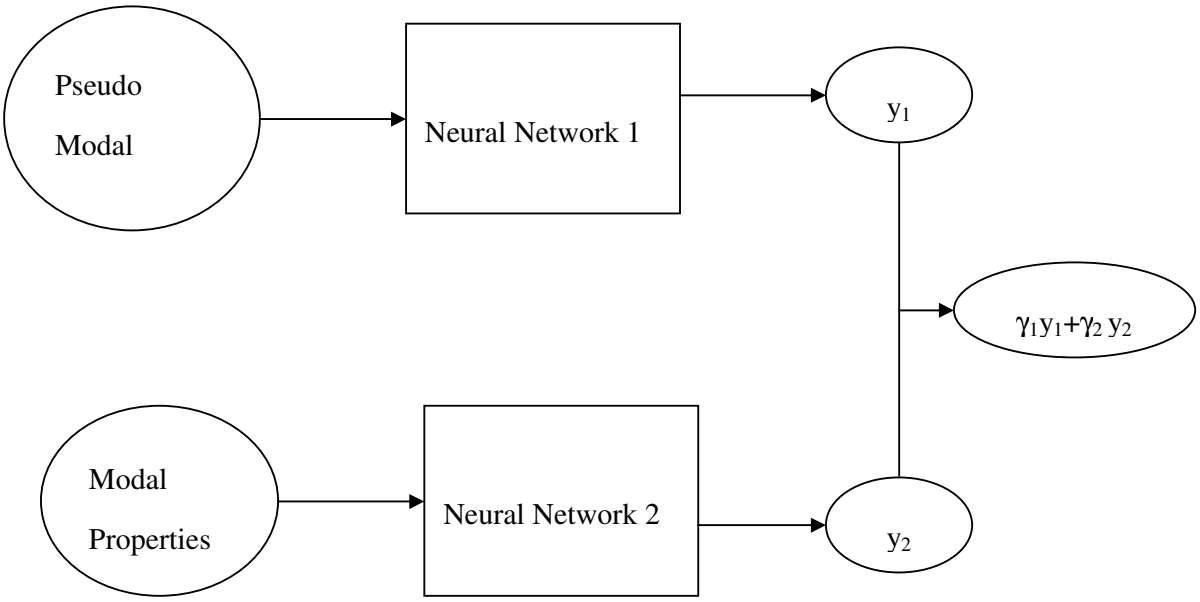


Figure 3.2 Illustration of a committee-of-networks. Here γ_1 and γ_2 are the committee-weights given to the pseudo-modal-energy-network and modal-property-network, respectively.

In the next subsection some aspects of the committee obtained by giving the two approaches in Figure 3.2 equal weights are examined.

(a) Equal weights

Now the relationship between the mean-of-square-of-error (MSE) given by the committee and given by the individual networks is investigated. The output of the committee is the average of the outputs of the individual networks. The committee prediction may be written in the following form by giving equal weighting functions:

$$y_{COM} = \frac{1}{2}\left(y_1(\psi) + y_2(\chi)\right) \tag{3.19}$$

The error due to the committee can be written as follows:

$$\begin{aligned} E_{COM} &= \varepsilon\left[\left(\frac{1}{2}\{y_1(\psi) + y_2(\chi)\} - h(\psi,\chi)\right)^2\right] \\ &= \varepsilon\left[\left(\frac{1}{2}(e_1 + e_2)\right)^2\right] \\ &= \frac{1}{4}\left(\varepsilon[e_1^2] + 2\varepsilon[e_1 e_2] + \varepsilon[e_2^2]\right) \end{aligned} \tag{3.20}$$

If it is assumed that the errors e_1 and e_2 have a zero mean and are uncorrelated then:

$$\varepsilon[e_1 e_2] = 0 \tag{3.21}$$

The error of the committee [equation 3.20 assuming equation 3.21] can be related to the average error of the networks acting individually [equation 3.18] as follows:

$$\begin{aligned} E_{COM} &= \frac{1}{4}\left(\varepsilon[e_1^2] + \varepsilon[e_2^2]\right) \\ &= \frac{1}{2}E_{AV} \end{aligned} \tag{3.22}$$

Equation 3.22 indicates that the average expected squares of errors is reduced by a factor of 2 by employing the committee of the two approaches [Atalla and Inman, 1998; Levin and Lieven, 1998[b]]. In equation 3.22 and 3.18, it is observed that the committee error is always less than or equal to the averaged error and this is written as follows:

$$E_{COM} \leq E_{AV} \tag{3.23}$$

In reality the assumption in equations 3.21 does not usually hold because the errors are usually highly correlated and therefore the error in the committee is higher than the one predicted by equation 3.22.

The analysis conducted above leads to a need for a criterion that could be used to assess the effectiveness of the committee. In this thesis, a committee factor defined as a measure of the effectiveness of the committee method is introduced. This factor is defined as the ratio of the average of the expected errors of the two independent methods acting individually to the expected error of the committee. By assuming that the two networks have equal weights, the Committee factor (CF) may be written as follows:

$$\begin{aligned} CF &= \frac{E_{AVERAGE}}{E_{COM}} \\ &= \frac{1/2(\varepsilon[e_1^2] + \varepsilon[e_2^2])}{1/4(\varepsilon[e_1^2] + \varepsilon[e_1 e_2] + \varepsilon[e_2^2])} \\ &= \frac{2}{1 + \dfrac{2\varepsilon[e_1 e_2]}{\varepsilon[e_1^2] + \varepsilon[e_2^2]}} \end{aligned} \tag{3.24}$$

In equation 3.24 it is observed that if the assumption that the two methods are uncorrelated [equation 3.21] is applied then the committee factor is equal to 2. When the individual methods are correlated then the committee factor is less than 2. Equation 3.24 may be generalised for n-trained network to give:

$$CF = \frac{n}{1 + \dfrac{2\sum_{i,j=1,i\neq j}^{n} \varepsilon[e_i e_j]}{\sum_{i=1}^{n} \varepsilon[e_i^2]}} \tag{3.25}$$

(b) Variable weights

In realistic situations, the two networks might not necessarily have the same predictive capacity, for example, the modal properties are extracted from the frequency response functions using an optimisation-based modal analysis technique while the pseudo modal energies are calculated using a direct integration scheme. As a result, the pseudo modal energy data may give

more accurate results than the modal properties. Consequently, each network should be given appropriate weighting functions as follows:

$$y_{COM} = \gamma_1 y_1(\psi) + \gamma_2 y_2(\chi) \tag{3.26}$$

where γ_1 and γ_2 are the weighting functions and $\gamma_1+\gamma_2=1$.

The error due to the committee can be written as follows:

$$\begin{aligned} E_{COM} &= \varepsilon[(\gamma_1 y_1(\psi) + \gamma_2 y_2(\chi) - h(\psi,\chi))^2] \\ &= \varepsilon[(\gamma_1 e_1 + \gamma_2 e_2)^2] \end{aligned} \tag{3.27}$$

Equation 3.27 may be rewritten in Lagrangian form as:

$$E_{COM} = \varepsilon[(\gamma_1 e_1 + \gamma_2 e_2)^2] + \lambda(1 - \gamma_1 - \gamma_2) \tag{3.28}$$

where λ = Lagrangian multiplier.

The derivative of the error in equation 3.28 with respect to γ_1 may be calculated and equated to zero as:

$$\begin{aligned} \frac{dE_{COM}}{d\gamma_1} &= \varepsilon[2(\gamma_1 e_1 + \gamma_2 e_2)e_1] - \lambda = 0 \\ \frac{dE_{COM}}{d\gamma_2} &= \varepsilon[2(\gamma_1 e_1 + \gamma_2 e_2)e_2] - \lambda = 0 \\ \frac{dE_{COM}}{d\lambda} &= (1 - \gamma_1 - \gamma_2) = 0 \end{aligned} \tag{3.29}$$

Solving equation 3.29 and assuming equation 3.21, the minimum errors are obtained when the weights are:

$$\gamma_1 = \frac{1}{\frac{\varepsilon[e_1^2]}{\varepsilon[e_2^2]} + 1}; \gamma_2 = \frac{1}{\frac{\varepsilon[e_2^2]}{\varepsilon[e_1^2]} + 1} \tag{3.30}$$

The second derivative of the error in equation 3.27 [assuming equation 3.21] with respect to γ_1 is:

$$\frac{d^2 E_{COM}}{d\gamma_1^2} = 2\left(\varepsilon[e_1^2] + \varepsilon[e_2^2]\right) > 0 \tag{3.31}$$

Equation 3.31 shows that this stationary point is a minimum point. In other words the optimal committee is always lower than the individual methods. Equation 3.31 may be generalised for a committee of n-trained networks. Thus equation 3.31 for network i may be rewritten as follows:

$$\gamma_i = \frac{1}{\sum_{j=1}^{n} \frac{\varepsilon[e_i^2]}{\varepsilon[e_j^2]}} \tag{3.32}$$

From equation 3.32, the following conditions may be derived:

$$\begin{aligned} &\varepsilon[e_1^2] = \varepsilon[e_2^2] \Rightarrow \gamma_1 = 0.5 \\ &\varepsilon[e_1^2] > \varepsilon[e_2^2] \Rightarrow \gamma_1 < 0.5 \\ &\varepsilon[e_1^2] < \varepsilon[e_2^2] \Rightarrow \gamma_1 > 0.5 \end{aligned} \tag{3.33}$$

Conditions in equations 3.33 show that if the predictive capacity of the pseudo modal energy approach is equal to that of the modal property approach then each method should be given equal weights. If the pseudo modal energy approach gives smaller errors than the modal property approach, then the pseudo modal energy method should be given more weight and vice versa. These conditions are trivial but have been derived in this study to confirm the effectiveness of the method. The strength of each method may be obtained by assessing how robust the methods are in the presence of noise. From this section the following theorem is derived:

Theorem 3.1: *If n uncorrelated networks are combined to form a committee, there exists an optimal committee which gives the least mean-of-square-of-error, lower than any individual network and this is obtained by assigning to the i^{th} network a weight of:*

$$\gamma_i = 1 / \sum_{j=1}^{n} \frac{\varepsilon[e_i^2]}{\varepsilon[e_j^2]}.$$

3.4.2 Committee gives the least variance-of-square-of-error

The variance of the committee method may be derived as follows:

$$\begin{aligned} V_{COM}[(\gamma_1 e_1 + \gamma_2 e_2)^2] &= V[\gamma_1^2 e_1^2 + 2\gamma_1\gamma_2 e_1 e_2 + \gamma_2^2 e_2^2] \\ &= \gamma_1^2 V[e_1^2] + 2\gamma_1\gamma_2 V[e_1 e_2] + \gamma_2^2 V[e_2^2] \\ &= \gamma_1^2 V[e_1^2] + 2\gamma_1(1-\gamma_1) V[e_1 e_2] + (1-\gamma_1)^2 V[e_2^2] \end{aligned} \tag{3.34}$$

Finding the derivative of the variance in equation 3.34 by substituting $\gamma_2 = 1-\gamma_1$, the following is obtained:

$$\frac{dV_{COM}}{d\gamma_1} = 2\gamma_1 V[e_1^2] + 2(1-2\gamma_1) V[e_1 e_2] - 2(1-\gamma_1) V[e_2^2] = 0 \tag{3.35}$$

Solving equation 3.35 in the same manner that equation 3.27 is solved gives the optimum weights of:

$$\gamma_1 = \frac{V[e_2^2] - V[e_1 e_2]}{V[e_1^2] - 2V[e_1 e_2] + V[e_2^2]} \quad \text{and} \quad \gamma_2 = \frac{V[e_1^2] - V[e_1 e_2]}{V[e_1^2] - 2V[e_1 e_2] + V[e_2^2]} \tag{3.36}$$

If it is assumed that $V[e_1e_2] = 0$, then equation 3.36 may be simplified to become:

$$\gamma_1 = \frac{1}{\frac{V[e_1^2]}{V[e_2^2]}+1} \quad \text{and} \quad \gamma_2 = \frac{1}{\frac{V[e_2^2]}{V[e_1^2]}+1} \tag{3.37}$$

Similarly, equations in 3.37 may be generalised for n-trained networks, and the i^{th} network is given the weight:

$$\gamma_i = \frac{1}{\sum_{j=1}^{n} \frac{V[e_i^2]}{V[e_j^2]}} \tag{3.38}$$

Differentiating 3.35 and assuming that $V[e_1e_2] = 0$, the following is obtained:

$$\frac{d^2 V_{COM}}{d\gamma_1^2} = 2_1 V[e_1^2] + 2V[e_2^2] > 0 \tag{3.39}$$

Similarly, equation 3.39 shows that the stationary point is a minimum point.

From equation 3.38, the following conditions may be derived:

$$\begin{aligned} &V[e_1^2] = V[e_2^2] \Rightarrow \gamma_1 = 0.5 \\ &V[e_1^2] > V[e_2^2] \Rightarrow \gamma_1 < 0.5 \\ &V[e_1^2] < V[e_2^2] \Rightarrow \gamma_1 > 0.5 \end{aligned} \tag{3.40}$$

From this section the following theorem may be deduced:

Theorem 3.2: *If n independent networks are combined to form a committee, there exists an optimal committee which gives a lower variance-of-square-of-error than any individual network and this is achieved by assigning to the i^{th} network a weight of:*

$$\gamma_i = 1 / \sum_{j=1}^{n} \frac{V[e_i^2]}{V[e_j^2]}.$$

From equations 3.32 and 3.38 the following theorem may be deduced:

Theorem 3.3: *The weight parameters of the n-trained networks that give the optimum mean-of-square-of-error are equal to those that give the optimum variance-of-square-of-error if* $\sum_{j=1}^{n} \frac{\varepsilon[e_i^2]}{\varepsilon[e_j^2]} = \sum_{j=1}^{n} \frac{V[e_i^2]}{V[e_j^2]}$.

This theorem suggests that the weights that give the least mean-of-square-of-error do not necessarily give the least variance-of-square-of-error. This implies that a decision has to be made as to which is the more important attribute. In this chapter, the average of the weights that give the least mean-of-square-of-error and those that give the least variance-of-square-of-error are used as the optimal committee. The average weight given to the individual methods to give

the optimal committee that takes into account of the mean- and variance-of-square-of-error may be written from equations 3.30 and 3.37 as follows:

$$\gamma_1 = \frac{0.5}{\frac{\varepsilon[e_1^2]}{\varepsilon[e_2^2]}+1} + \frac{0.5}{\frac{V[e_1^2]}{V[e_2^2]}+1}; \gamma_2 = \frac{0.5}{\frac{\varepsilon[e_2^2]}{\varepsilon[e_1^2]}+1} + \frac{0.5}{\frac{V[e_2^2]}{V[e_1^2]}+1} \tag{3.41}$$

Substituting equation 3.41 into equation 3.26 the optimal committee output is obtained to be:

$$y_{COM} = \left(\frac{0.5}{\frac{\varepsilon[e_1^2]}{\varepsilon[e_2^2]}+1} + \frac{0.5}{\frac{V[e_1^2]}{V[e_2^2]}+1} \right) y_1 + \left(\frac{0.5}{\frac{\varepsilon[e_2^2]}{\varepsilon[e_1^2]}+1} + \frac{0.5}{\frac{V[e_2^2]}{V[e_1^2]}+1} \right) y_2 \tag{3.42}$$

The properties of averaging the weights that give the least MSEs and variance of errors are illustrated by Figure 3.3.

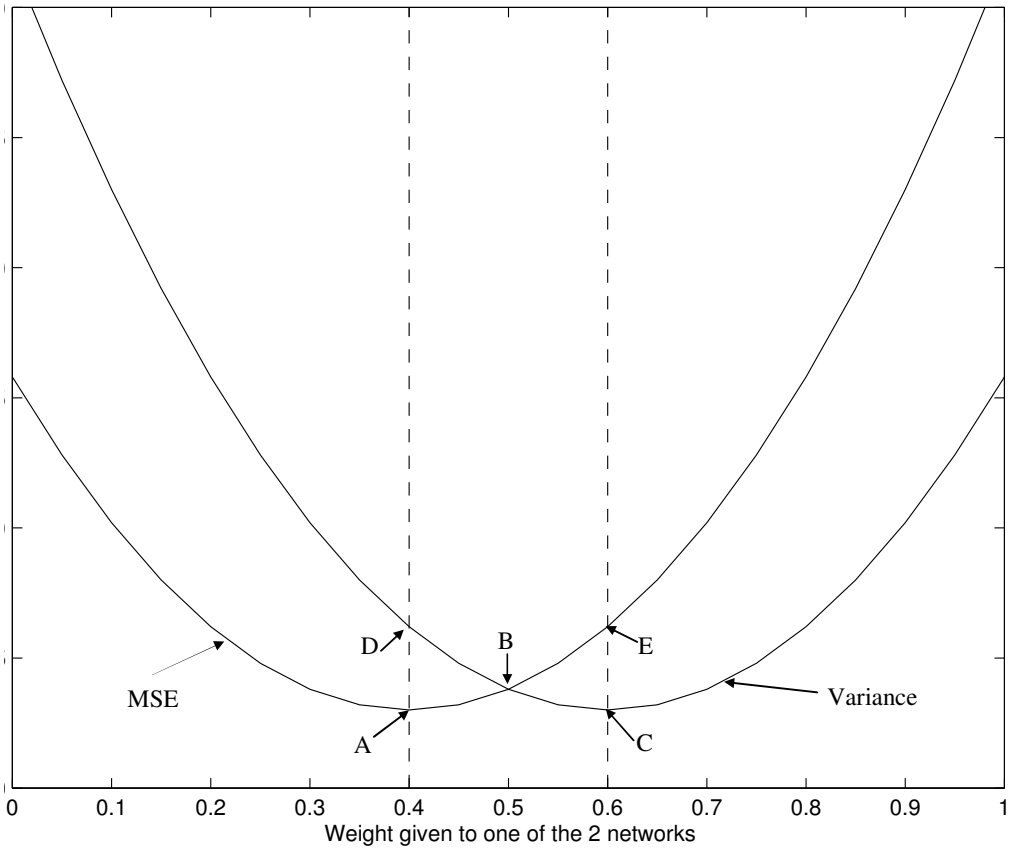


Figure 3.3. Illustration of the effect of averaging the committee-weights that give the least MSEs and variance of squared errors

In equation 3.27 it may be deduced that provided that the errors of the two methods are uncorrelated, the committee error versus the committee-weight is a parabola with a minimum point. The same can be deduced for variance of square of errors from equation 3.34. If these two parabolas are plotted together, Figure 3.3 is obtained. This figure is for illustrative purposes, and therefore the values in the y-axis are not essential for demonstrating the properties of equation 3.42. In this figure the minimum MSE occurs at A. This corresponds to a variance shown in D, which is higher than that of B where the two committee-weights are averaged. By looking at the same figure, the minimum variance is at C. This corresponds to the MSE shown in E, which is higher than that of B where the two committee-weights are averaged. From this figure it is clear that the average of the two methods is the best compromise between the two minimum values. Additionally, B is still better than when the methods are used in isolation.

3.4.3 Committee gives a more reliable solution

One of the main advantages of the committee method is its reliability. Reliability is a criterion that is used in industry, together with economic factors, to choose the appropriate method. This section seeks to demonstrate that the committee method is more reliable than using individual networks.

Suppose that the probabilities of success for the pseudo-modal-energy-network and the modal-property-network are $P(x_1)$ and $P(x_2)$ respectively. The reliability of the two methods acting in parallel is:

$$\begin{aligned} P(x_1 \cup x_2) &= P(x_1) + P(x_2) - P(x_1 \cap x_2) \\ &= P(x_1) + P(x_2) - P(x_1)P(x_2) \end{aligned} \tag{3.43}$$

From equation 3.43, it can be deduced that the reliability of the committee, is at least, as good as either of the individual networks [McColl, 1995]. Note that equation 3.43 has been formulated to accommodate two neural networks even though it can be generalised to any number of networks, in order to put this equation in the context of this thesis. From the analysis above the following theorem is deduced:

Theorem 3.4: *If n independent (uncorrelated) networks are used simultaneously, the reliability of the committee is at least as good as when the networks are used individually.*

3.5 CASE STUDY: SIMULATED CANTILEVERED BEAM

The committee approach outlined in the previous sections is applied to identify faults in a cantilevered beam shown in Figure 3.4. The beam is the same as the one used in Chapter 2 and is made of aluminium with the following dimensions: the length of the beam of 1.0 m, the width of 50 mm and the thickness of 6 mm. The beam is free on one end and clamped on the other end and is restricted to translation in the y-axis and rotation about the z-axis. It is partitioned into three substructures and modelled with 50 elements (51 nodes) using the Structural Dynamics Toolbox [Balmès, 1997[a]] in MATLAB [Mathworks, 1992]. Node 51 is located at the free end of the beam and node 1 is located at the clamped part of the beam. Since the beam is restricted to translation in the y-axis and rotation about the z-axis, the beam has 50 active nodes (because node 1 is clamped).

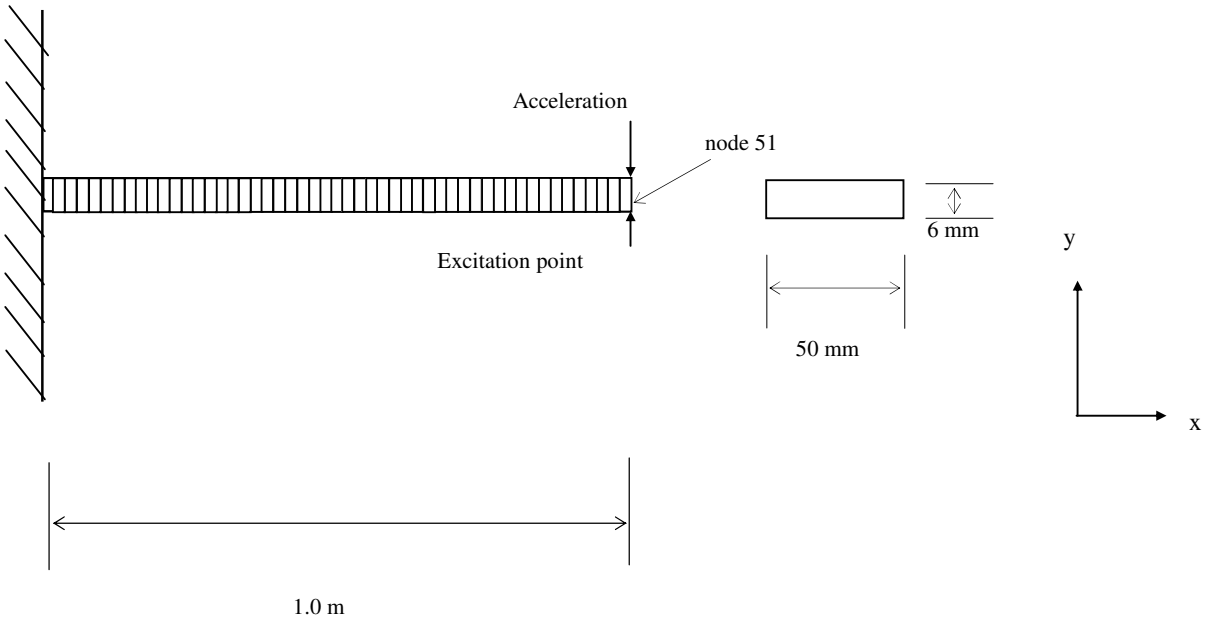


Figure 3.4 The cantilevered beam modelled with 50 elements and divided into 5 substructures

Using the Structural Dynamics Toolbox the mass and stiffness matrices of size 100 by 100 are assembled. Here 50 active nodes, each with two degrees of freedom corresponding to translation in the y-plane and rotation about the z-axis, give 100 degrees of freedom and thus the mass and stiffness matrices of size 100 by 100. From the mass and stiffness matrices, 100 modal properties are calculated. From the calculated modal properties, inertance FRFs are calculated

using the modal summation equation (an expression inside the integral in equation 2.6), by assuming that the beam is lightly damped and fixing the damping ratios to 0.001. The FRFs calculated correspond to excitation at node 51, as shown in Figure 3.4, and acceleration measurements in the direction shown in the same figure for all 50 active degrees of freedom. From the 50 FRFs calculated, 9 FRFs that correspond to nodes 3, 7, 13, 17, 23, 27, 33, 43 and 51 are selected. The FRFs that are selected exclude those corresponding to rotational degrees of freedom. This is due to the fact that in practical situations measuring the rotational degrees of freedom is difficult.

The 9 FRFs are converted into the time domain and Gaussian noise levels of ±0%, ±1%, ±2%, ±3% and ±4 % are added to the time domain data. Using the Fourier transform technique, as described in Section 2.4, the FRFs are calculated from the time domain data. The effect of these noise levels on the FRFs was shown in Chapter 2, Figure 2.5. From the FRFs calculated from the time domain data, 5 modal properties are extracted using the procedure outlined in Appendix C and the pseudo modal energies are calculated by integrating the FRFs at various frequency bandwidths. On calculating the pseudo modal energies, the following bandwidths are chosen in order to bracket each of the first 5 natural frequencies of the beam: 18-44, 155-240, 484-620, 1014-1151 and 1726-1863 Hz. These bandwidths are the same as those chosen in Chapter 2. The procedure implemented in this paragraph was described in detail in Chapter 2, Section 2.4.

Ninety pseudo modal energies are extracted from one set of simulation (45 imaginary and 45 real parts of the pseudo modal energies). Here it should be noted that 45 conform to 9 co-ordinates corresponding to translation in the y-direction and 5 modes. The total number of modal properties calculated is 50 (45 mode shape co-ordinates and 5 natural frequencies).

In order to generate different fault cases, the structure is divided into three substructures. Substructure 1 contains elements 1 to 16; substructure 2 contains elements 17 to 32 and substructure 3 elements 33 to 50. A fault is introduced to a substructure by simultaneously reducing the cross-sectional areas of four elements in a substructure by 5 to 10%. This is defined as a presence of a fault in a substructure. These reductions are sampled from a uniform distribution. The absence of a fault in a substructure is defined as the simultaneous reductions of four elements in a substructure by 0 to 2% sampled from a uniform distribution. A fault present

in substructure 1 is restricted to elements 7 to 10; for substructure 2 elements 24 to 27; and substructure 3 elements 42 to 45. For a given fault in a substructure the maximum reduction in cross-sectional area corresponds to a reduction of 0.8% of the total volume of the beam.

Each fault case is assigned a fault identity, corresponding to a location of a fault in a substructure. For example, a fault existing in substructure 1 indicating a reduction of 8% of the cross-sectional areas of four elements in substructure 1 yields an identity of [1 0 0]. In this study the presence of faults in one substructure e.g. [1 0 0] is defined as a single fault and the presence of faults in more than one substructures e.g. [1 1 0] as a multiple fault case. The types of fault cases simulated are a zero-fault case [0 0 0]; one-fault cases [1 0 0], [0 1 0] and [0 0 1]; two-fault cases [1 1 0], [1 0 1] and [0 1 1]; and three fault case [1 1 1]. Thus 8 different fault cases are generated. For each noise contamination of the data 2400 data are simulated, 300 for each of the 8 fault cases.

3.5.1 Normalisation of the measured data

Since the modal properties and pseudo modal energies extracted are to be used for neural network training, it is important to normalise them so that those inputs that have higher magnitudes do not dominate the training. Here the scaling technique that is implemented ensures that all parameters fall within the interval [0,1]. To achieve this, the following scaling method is used:

$$x_m^{new} = \frac{x_m^{old} - \min(x_m^{old})}{\max(x_m^{old}) - \min(x_m^{old})} \tag{3.44}$$

where x_m is a row of the input parameters.

3.5.2 Statistical overlap factor and principal component analysis

In the previous section a simulated beam was found to give more vibration data that needed for fault identification, and all these could not be possibly used for neural network training. These data must therefore be reduced and the reason for this reduction is discussed in this section.

In statistics there is a phenomenon called the *curse of dimensionality* [Bellman, 1961], which refers to the difficulties associated with the feasibility of density estimation in many dimensions.

It is therefore a good practice to reduce the dimension of the data, hopefully without the loss of essential information. This section deals with the techniques implemented in this thesis to reduce the input space. The techniques implemented in this thesis reduce the dimension of the input data by removing the parts of the data that do not contribute significantly to the dynamics of the system being analysed or those that are too sensitive to irrelevant parameters such as slight changes in temperatures. To achieve this, the statistical overlap factor introduced in Chapter 2 is implemented to select those data that show high sensitivity to faults. Here the statistical overlap factors are calculated between undamaged and damaged beam and those input values that show higher statistical overlap factors are selected. It should be noted than on calculating the statistical overlap factors, one-fault cases are used as data from a damaged structure. There is a possibility that the data selected using the statistical overlap factor may be correlated and therefore a procedure that would reduce the data already reduced using the statistical overlap factor into independent input space is required.

In this thesis the principal component analysis (PCA) [Jolliffe, 1986] is implemented to reduce the input data into independent components. The PCA orthogonalizes the components of the input vector so that they are uncorrelated with each other. When implementing the PCA for data reduction, correlations and interactions among variables in the data are summarised in terms of a small number of underlying factors. The PCA was introduced by Pearson (1901) to recast linear regression analysis in a new framework and was developed further by Hotelling (1933) who applied it to Psychometry and was subsequently generalised by Loéve (1963). The PCA has been successfully used to reduce the dimension of the data [Bishop, 1995]. Some of the researchers who have successfully applied this technique include Partridge and Calvo (1998) who used the PCA to reduce the dimensions of two high-dimensional image databases, one of handwritten digits and one of handwritten Japanese characters.

The variant of the PCA implemented in this thesis finds the directions in which the data points have the most variance. These directions are called principal directions. The data are then projected onto these principal directions without the loss of significant information of the data. Here a brief outline of the implementation of the PCA adopted in this thesis is described. The

first step in the implementation of the principal component analysis is to construct a covariance matrix defined as follows [Tarassenko, 1998]:

$$\Sigma = \sum_{p=1}^{P} (x^p - \mu)(x^p - \mu)^T \tag{3.45}$$

Here Σ is the covariance matrix, T is for transpose, P is the number of vectors in the training set, μ is the mean vector of the data set taken over the number of training set and x is the input data. The second step is to calculate the eigenvalues and eigenvectors of the covariance matrix and arrange them from the biggest eigenvalue to the smallest. The first *N* biggest eigenvalues are chosen. In this thesis the first *N* eigenvalues are chosen in such a way that their sum constitutes at least 85% of the total sum of all the eigenvalues and by so doing at least 85% of the variance of the data are retained. The data are then projected onto the eigenvectors corresponding to *N* most dominant eigenvalues. A schematic illustration of how the PCA works is shown in Figure 3.5.

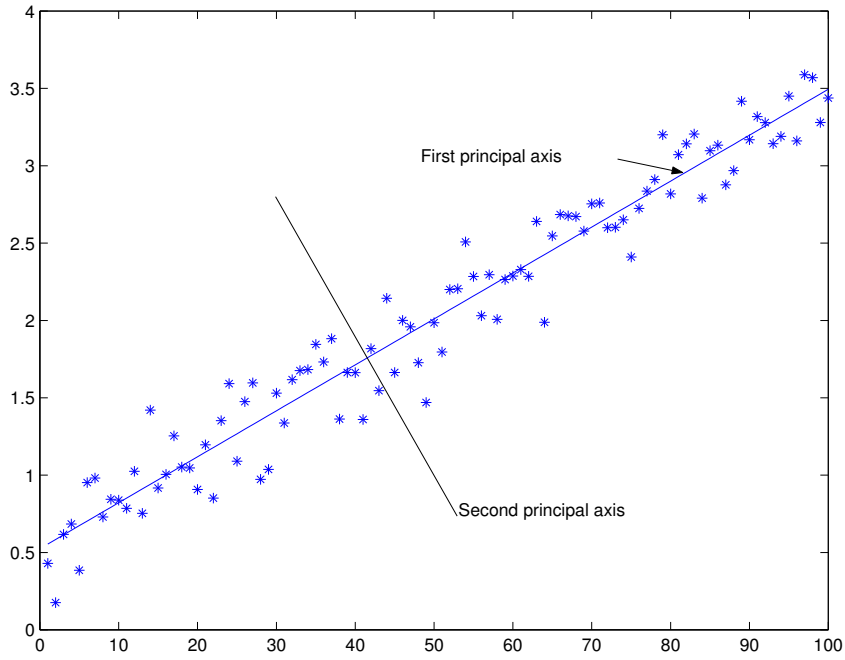


Figure 3.5. A schematic illustration of how a two-dimensional data can be reduced to a one-dimensional data. The data set is decomposed into eigenvalues and eigenvectors. Since this is a two-dimensional data set, there are two principal axes corresponding to two eigenvectors. In the illustration above, the data are projected onto the first principal axis corresponding to a larger eigenvalue. The principal axes are 90^o to each other.

If all the simulated degrees of freedom are used for training the networks with the pseudo modal energies, then the size of the input units will be 90 (2×45 for the real and imaginary parts of the pseudo modal energies). For training the modal-property-network, the network will have 50 (45 mode shapes and 5 natural frequency) input units. In this thesis both these input data sets are reduced from 90 and 50, respectively, for the pseudo modal energies and modal properties, to 10. The pseudo modal energies are reduced from 90 to 40 using the statistical overlap factor by choosing 40 input data that show the highest statistical overlap factor between data from the damaged and undamaged simulated beam. The remaining 40 pseudo modal energies are reduced to 10 using the principal component analysis. The modal properties are reduced from 50 to 30 using the statistical overlap factor by choosing 30 modal properties that show the highest statistical overlap factors. The remaining 30 modal properties are reduced to 10 using the principal component analysis.

3.5.3 Training, validation and testing

The procedure followed for training the MLP networks, in this chapter, uses the training, validation and testing stages [Bishop, 1995; Haykin, 1999; Tarassenko, 1998]. Each stage has a data set assigned to it and ideally these three data sets should be of equal size and must be independent of one another. The network-weights that map the input to output data identified by minimising the network error and using the training data, may over-fit the training data. The neural network that over-fits the training data may not necessarily perform well on the validation and test data sets. Over-fitting the training data set is a situation where a network stops learning how to approximate the hidden dynamics of the system and learns the noise in the data. In order to combat this problem more networks than required are trained and the network that gives the least mean squared errors, and in this chapter, classification errors on the validation data set is chosen. The chosen network may also over-fit the validation data in addition to the training data, and so the test data set is used to evaluate the performance of the trained network.

By using the finite element model, 800 vibration data are generated by perturbing the cross-sectional areas of the beam, as it has been described at the beginning of Section 3.5, and used as a training set. This data set contains 8 fault cases and 100 examples for each fault case. The

validation data set with 800 examples is generated and used to select the neural network architectures. The test data set is also generated and contains 800 examples. The validation and testing sets each contain 8 fault cases and 100 examples for each fault case.

Fifty pseudo-modal-energy networks and fifty modal-property-networks are trained using the training data set with zero-noise contamination and with the number of hidden units randomly chosen to fall from 8 to 16. On training these networks the cross-entropy cost function shown in equation 3.10 is used as opposed to the sum-of-square-of-errors cost function shown in equation 3.9 because it has been found to be better suited for classification problems than the sum-of-squares-of-errors [Bishop, 1995]. On training all these networks, 100 iterations are used and are found to be sufficient for convergence of the training errors. These networks have 10 input data, which are chosen using the statistical overlap factor and the principal component analysis, and 3 output units corresponding to 3 substructures. The logistic function, described by equation 3.7, is chosen as the output activation function, and the hyperbolic tangent function, described by equation 3.8, is chosen as the activation function in the hidden layer. The network is trained using the scaled conjugate gradient method [See Appendix E]. On training these networks the coefficient of the contribution of the regularisation parameter (α), shown in equation 3.10, to the training error is set to 15. This value is chosen because it is found that it sufficiently smoothes out the network weights without compromising the abilities of the networks to generalise the data. Here smooth weight vectors are defined as vectors with components of the same order of magnitudes. Of the two sets of 50 trained networks, the two sets of networks that give the least classification errors on the validation data set are chosen. These classification errors are calculated by rounding-off the fault identities given by the networks and calculating the proportion of fault cases classified correctly. From these two chosen networks the optimal sizes of the hidden units are 11 for the pseudo-modal-energy-network and 10 for the modal-property-network.

Using the data contaminated with 1% Gaussian noise, ten pseudo-modal-energy-networks and ten modal-property-networks are trained with different network-weights initialisations and with the number of hidden units set to 11 and 10, respectively. The two sets of networks that

give the least classification errors on the validation data set are chosen. The same process is repeated for ±2%, ±3% and ±4% Gaussian noise contamination of the data.

3.6 RESULTS AND DISCUSSIONS

The results showing the classification errors between the training and validation data sets are shown in Table 3.1. Each row in Table 3.1 shows a given noise contamination of the vibration data while the columns show the training and validation classification errors of the pseudo-modal-energy-network and modal-property-network. For a given noise level and a given network, the training and validation data sets give classification errors of the same order of magnitudes implying that the networks have not over-fitted the data. Table 1 shows that the higher the noise contamination of the data the higher the classification error. It also shows that, on average, the pseudo-modal-energy-network gives more accurate classification results than the modal-property-network.

Table 3.1 Classification errors when using two individual approaches and various noise levels added to the data. Key: PME-N: pseudo-modal-energy-network; MP-N: modal-property-network

	Training (%)		Validation (%)	
Noise (%)	PME-N	MP-N	PME-N	MP-N
0	1.75	4.70	1.78	4.98
1	2.95	7.30	3.15	7.65
2	4.05	9.97	4.21	10.83
3	8.92	11.84	9.67	11.91
4	15.22	21.71	16.33	23.10

The graph of error versus number of iterations when training the network using the pseudo modal energies is illustrated in Figure 3.6. Figure 3.6 indicates that convergence is not dependent on the noise level in the data. Figure 3.7 shows the graph of error versus number of iterations for various noise levels when the modal data are used for training. Again, there is no consistent relationship between the convergence and noise levels in the data. It is therefore

concluded that the presence of noise in the data does not influence the convergence rate during training. Figure 3.8 shows the graph of the mean squared error versus the weight assigned to the pseudo-modal-energy-network and Figure 3.9 is a graph showing the standard-deviation-of-squares-of-error versus the pseudo-modal-energy-network for various noise levels.

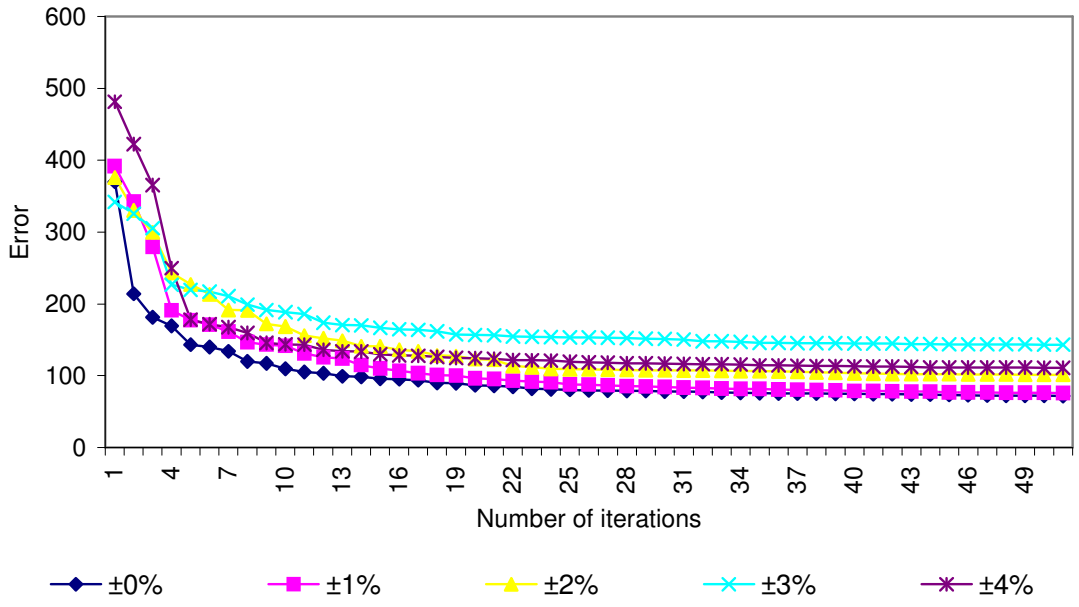


Figure 3.6. Convergence histories when training the networks using the pseudo modal energies for noise levels of 0%, ±1%, ±2%, ±3% and ±4%.

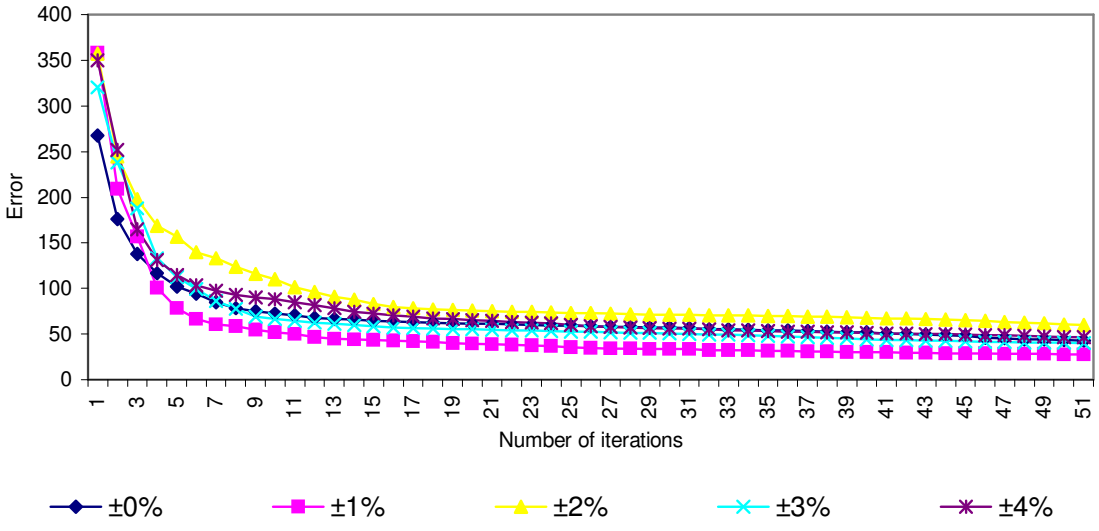


Figure 3.7. Convergence histories when training the network using the modal properties for noise levels of 0%, ±1%, ±2%, ±3% and ±4%.

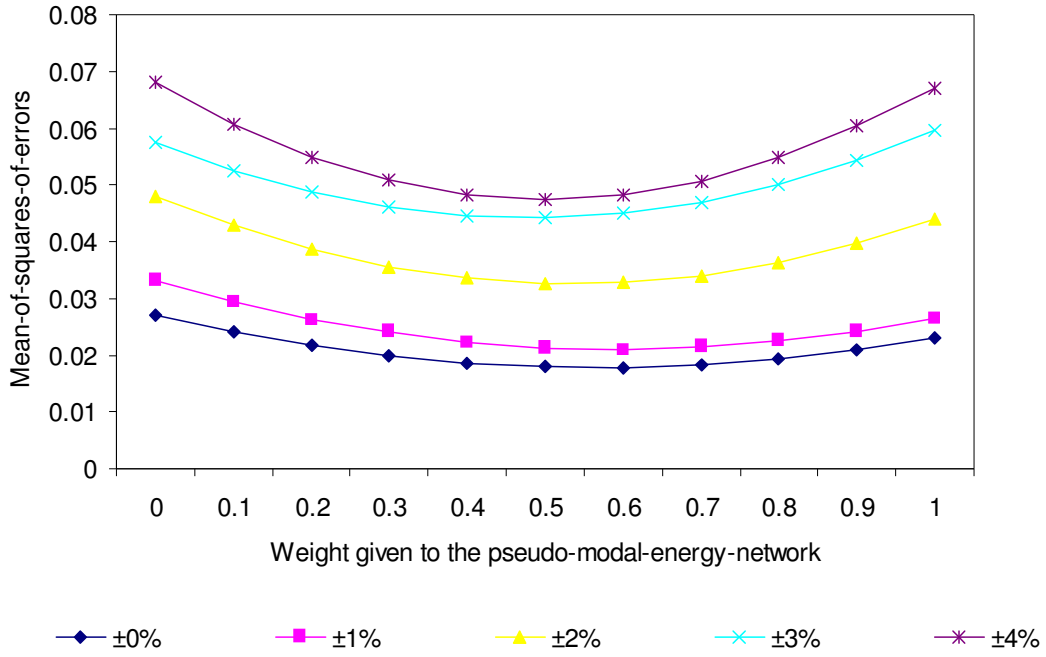


Figure 3.8 Mean squared errors versus weighting function given to the pseudo-modal-energy-network for noise levels of 0%, ±1%, ±2%, ±3% and ±4%. These results are obtained when the networks are assessed on the validation data set.

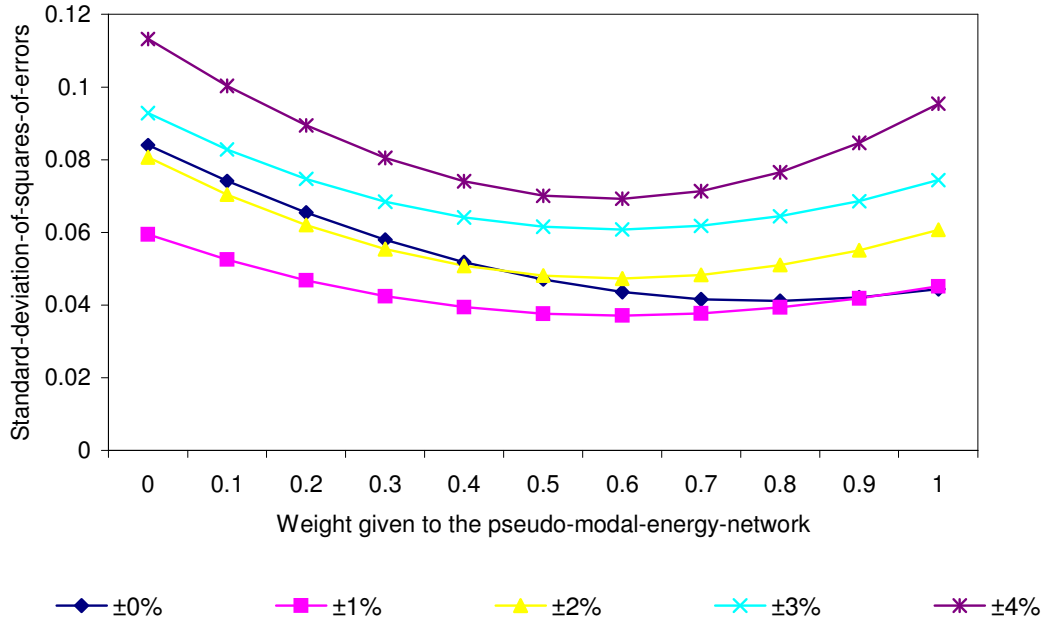


Figure 3.9 Standard-deviation-of-square-of-error versus weighting function given to the pseudo-modal-energy-network for noise levels of 0%, ±1%, ±2%, ±3% and ±4%. These results are obtained when the networks are assessed on the validation data set.

These figures are obtained when the trained networks are evaluated on the validation data set. From Figures 3.8 and 3.9, it is found that giving the two sets of networks approximately equal committee-weights is an optimal way of constructing a committee of networks trained with data contaminated with ±1% to ±4% noise. Furthermore, assigning 70% committee-weight to the pseudo-modal-energy-network is an optimal way of constructing a committee of networks trained with no noise contamination. The optimal committees are calculated using equation 3.41. Here a difference should be noted between committee-weights (sometimes called weighting functions), which are the weights given to the pseudo-modal-energy-network and modal-property-network on calculating the committee and network-weights, which are the weights that map the vibration data to the identity of faults. Figure 3.9 shows that the standard-deviation-of-squares-of-errors obtained when the data is contaminated with 0% noise shows the optimal committee occurring when the pseudo-modal-energy-network is given 80% weighting function. The reason for this is because the assumption made in equation 3.39 that $V[e_1 e_2] \approx 0$ does not strictly hold for this case.

The average committee factor (CF) over all noise levels shown in Figure 3.8 is 1.40. The CF is defined by equation 3.24 as the ratio of the mean of the expected square of errors of the individual methods to the expected squared errors of the optimal committee. The higher the CF the more effective is the committee over the individual methods. Similarly, the CF of the results shown in Figure 3.9 is 1.48. The CF for Figure 3.9 is calculated by substituting for the mean squared errors in equation 3.24 by the standard deviation.

The trained networks are used to classify faults from the test data set into their respective classes. The fault cases given by networks are rounded off to the nearest whole number, i.e. 0 and 1. To assess the predictive capacities of the trained networks, a confusion matrix is used [Gopinath, 1999]. Tables 3.1 to 3.4 show the confusion matrices obtained when using the networks trained with data contaminated with ±2% Gaussian noise to classify fault cases in the test data set. The confusion matrix consists of the predicted fault cases, which are shown vertically and the actual fault cases, which are shown horizontally. The diagonal entries of the confusion matrix represents the fault cases that have been classified correctly, while the off-diagonal entries of this matrix represent fault cases that have been classified incorrectly. A

perfect fault identification procedure gives a diagonal matrix with all off-diagonal entries equal to zero.

Table 3.2. Confusion matrix from the classification of fault cases using the pseudo-modal-energy-network extracted from the FRFs calculated from the time domain data with ±2% Gaussian noise. These results are obtained when the trained networks are assessed on the test data set.

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	95	5	0	0	0	0	0	0
	[100]	11	89	0	0	0	0	0	0
	[010]	2	0	98	0	0	0	0	0
	[001]	0	0	0	100	0	0	0	0
	[110]	0	1	3	0	96	0	0	0
	[101]	0	0	0	7	0	93	0	0
	[011]	0	0	0	0	0	0	98	2
	[111]	0	0	0	0	0	2	2	96

From the confusion matrices in Tables 3.2 to 3.4, the abilities of the networks to detect faults is evaluated by comparing the proportion of fault cases that are classified correctly into fault and no-fault classes. Table 3.2 shows that the pseudo-modal-energy-network detects 97.8% fault cases correctly; Table 3.3 shows that the modal-property-network detects 95.4% fault cases correctly; and Table 3.4 shows that the committee method detects 98.3% fault cases correctly. Here detection is defined as classifying fault cases into no-fault and fault classes.

When the networks are used to classify fault cases in the test data set contaminated with ±2% Gaussian noise into 8 fault cases, the pseudo-modal-energy-network, as shown in Table 3.2, classifies 95.6% cases correctly; the modal-property-network classifies 88.5% cases correctly; and the committee method classifies 97.4% of cases correctly.

When the networks are used to classify faults in the test data set into faults and no-fault classes with various noise levels added to the vibration data, the results in Table 3.5 are obtained. This table shows that the committee method gives the best results followed by the pseudo-modal-

energy-network and then the modal property network. Table 3.5 shows that in general the higher the levels of noise contamination of the vibration data, the lower the accuracy of the trained networks.

Table 3.3. Confusion matrix from the classification of fault cases using the modal-property-network extracted from the FRFs calculated from the time domain data with ±2% Gaussian noise. These results are obtained when the trained networks are assessed on the test data set.

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	87	6	0	7	0	0	0	0
	[100]	6	94	0	0	0	0	0	0
	[010]	13	0	86	1	0	0	0	0
	[001]	12	0	0	88	0	0	0	0
	[110]	0	6	3	0	84	0	0	7
	[101]	0	9	0	3	0	81	0	7
	[011]	0	0	9	1	0	0	90	0
	[111]	0	0	0	0	0	0	2	98

Table 3.4. Confusion matrix from the classification of fault cases using the committee formed by the data extracted from the FRFs calculated from the time domain data with ±2% Gaussian noise. These results are obtained when the trained networks are assessed on the test data set.

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	98	2	0	0	0	0	0	0
	[100]	7	93	0	0	0	0	0	0
	[010]	4	0	96	0	0	0	0	0
	[001]	1	0	0	99	0	0	0	0
	[110]	0	1	3	0	96	0	0	0
	[101]	0	0	0	2	0	98	0	0
	[011]	0	0	0	1	0	0	99	0
	[111]	0	0	0	0	0	0	0	100

Table 3.5. The accuracy results obtained when the networks are used to classify faults into 2 classes, i.e. fault and no-fault. These results are obtained when the trained networks are assessed on the test data set. Key: PME-N: pseudo-modal-energy-network; MP-N: modal-property-network

Noise level (%)	PME-N	MP-N	Committee
±0	98.6	96.3	99.5
±1	96.5	96.1	97.9
±2	97.8	95.4	98.3
±3	95.3	93.8	97.3
±4	93.4	91.3	96.0

Table 3.6 shows the results obtained when the networks are used to classify faults in the test data into 8 fault cases. This table shows that the committee method gives more accurate classification results than the individual networks. The pseudo-modal-energy-network gives, on average, more accurate results than the modal-property-network. Table 3.6 shows that in general the accuracy of the methods decreases with the increase in the levels of noise contamination of the vibration data.

Table 3.6. The accuracy results obtained when the networks are used to classify faults into 8 fault cases. These results are obtained when the trained networks are assessed on the test data set. Key: PME-N: pseudo-modal-energy-network; MP-N: modal-property-network

Noise level (%)	PME-N	MP-N	Committee
±0	98.1	94.9	99.3
±1	96.5	92.0	97.1
±2	95.6	88.5	97.4
±3	89.8	87.1	95.5
±4	83.1	75.6	92.4

In this chapter, it is observed that using the committee-weights in equation 3.41 to construct a committee gives marginally better results than using committee-weights in equation 3.30 for data contaminated with 0% noise level. For this noise level, the committee constructed using equation 3.30 gives 97.8% classification accuracy while that obtained using equation 3.41 gives 99.3% classification accuracy. For other noise levels the classification accuracy between the two

methods are similar due to the fact that the two equations give approximately equal committee-weights.

In this chapter the following issues were observed:

1. The neural networks are able to classify fault cases including multiple faults accurately. Here multiple faults are defined as faults occurring in more than one substructure, e.g. [1 1 0] fault case.
2. The committee method gives better detection and classification of faults than the individual methods.
3. The pseudo-modal-energy-network gives, on average, better classification of faults than the modal-property-network.
4. The committee gives lower mean squared errors and variance than the individual methods. The advantage of low variance is that it decreases the level of uncertainty on the solution offered by the committee.

The average computer processing unit time taken to train 50 networks is 48 CPU minutes on a Dell DIMENSIONXPS600 Desktop Computer with 260MB of random access memory. When training all the neural networks, the maximum number of iterations is set to 100. The disadvantage of the proposed method is that it is computationally more expensive than the existing methods because it requires at least two trained networks.

3.7 CONCLUSION

In this chapter a committee-of-networks is introduced and applied to structural diagnostics. The committee method is constructed by processing the vibration data using two different methods; the most widely used modal properties and pseudo modal energies introduced in Chapter 2. It is shown that the committee method gives lower mean squared errors, standard-deviation-of-squares-of-errors and classification errors than the two existing methods acting individually. The pseudo-modal-energy-network is observed to give, on average, more accurate classification of faults than the modal-property-network.

Chapter 4

EXPERIMENTAL MEASUREMENTS ON A POPULATION OF CYLINDERS

4.1 INTRODUCTION

In the literature, there is little coverage of fault identification using vibration data from a population of structures. This is because it is expensive to damage a population of cylinders and measure vibration data. Srinivasan and Kot (1992) studied the feasibility of using vibration data to identify faults in cylinders. These authors introduced faults by machining a notch on a cylinder suspended by relatively soft springs to simulate free boundary conditions. Vibration data before and after the introduction of faults were measured. From these measurements, the modal properties before and after damage were compared and it was found that modal properties changed as a result of faults.

The primary motivation of this chapter is to extend the work of Srinivasan and Kot (1992) by studying the feasibility of the use of experimentally measured vibration data from a population of cylinders to identify faults. In addition to using the conventional modal properties to conduct this feasibility study, a new parameter called pseudo modal energy, which was introduced for fault identification in Chapter 2, is also used. Cylinders are chosen because cylindrical shapes are used in many structures, including the fuselage, engine casing and nacelle of aircraft.

Vibration data from a population of 20 seam-welded cylindrical shells made of steel are measured by exciting the cylinders at various locations using an impulse hammer and measuring vibration responses using an accelerometer located at a fixed position. Each cylinder is divided into three equal substructures and holes of 10-15 mm are introduced at the centers of each substructure. From the measured vibration data, the Frequency Response Functions (FRFs) are calculated. The modal properties, i.e. natural frequencies and the mode shapes, are then extracted from the FRFs using the procedure outlined in Appendix B. Mode shapes are transformed into the Coordinate-Modal-Assurance-Criterion (COMAC) [Lieven and Ewins, 1988] by calculating the correlation of each set of the measured modal properties to the median modal properties calculated from a population of undamaged cylinders. The pseudo modal energies introduced in Chapter 2 and applied on a simulated cantilevered beam in Chapter 3 are calculated from the FRFs by determining the integrals of the real and imaginary components of the FRFs. Similarly, the pseudo modal energies are transformed into the Coordinate-Pseudo-Modal-Energy-Assurance-Criterion (COPMEAC), a criterion that is introduced in this chapter.

By comparing the modal properties, the pseudo modal energies, the COMAC and the COPMEAC between undamaged and damaged populations the feasibility of using vibration data for fault identification in a population of cylinders is assessed. The changes of these parameters as a result of faults are investigated while taking into account the following issues:

(1) Changes in modal properties or pseudo modal energies resulting from the presence of faults compared with those resulting from variation in measurements and in physical properties of the population of cylindrical shells.
(2) Uncertainty in measurement positions.
(3) Errors introduced during modal analysis.
(4) Changes in support conditions and environmental conditions.

4.2 FINITE ELEMENT MODEL OF A CYLINDRICAL SHELL

Before experimental measurements are conducted on a population of cylindrical shells, a finite element model is constructed using ABAQUS [Anonymous, 1994] to study the dynamics of the cylinders. The cylinder has a diameter of 100mm, a height of 100mm and a thickness of 1.75mm. The finite element model consists of 1001 8-noded-shell-elements and 4100 nodes. This size of elements is chosen because it is found that increasing the mesh size does not improve the results any further, implying that the finite element model has converged. The diagram showing a finite element model of a cylinder is shown in Figure 4.1.

This figure shows the mode shape of the first natural frequency occurring at 433Hz. Further natural frequencies calculated from the finite element model are shown later in this chapter. From the finite element model and taking into account the limitation of the equipment used to measure vibration data, it is decided that the frequency range of 0-5000Hz would be used for fault identification. This is because this frequency bandwidth contains enough modes to be used for fault identification. The specifications of the equipment used are given later in the chapter. Adding a mass of 5g at various positions in the finite element model is used to study the dynamics of the cylinder and it is observed that adding this mass to a cylinder, which is symmetrical, breaks down the symmetry, thereby eliminating the incidence of repeated modes. The mass of the cylinder is 0.43kg. In the next section, details on the experimental measurements are described.

4.3 EXPERIMENTAL MEASUREMENTS

In this section, the experimental measurements conducted on a population of cylinders are described. Details about the equipment used in the experiments are provided, the experimental procedure is outlined and the relevant signal processing tools implemented in this study are discussed.

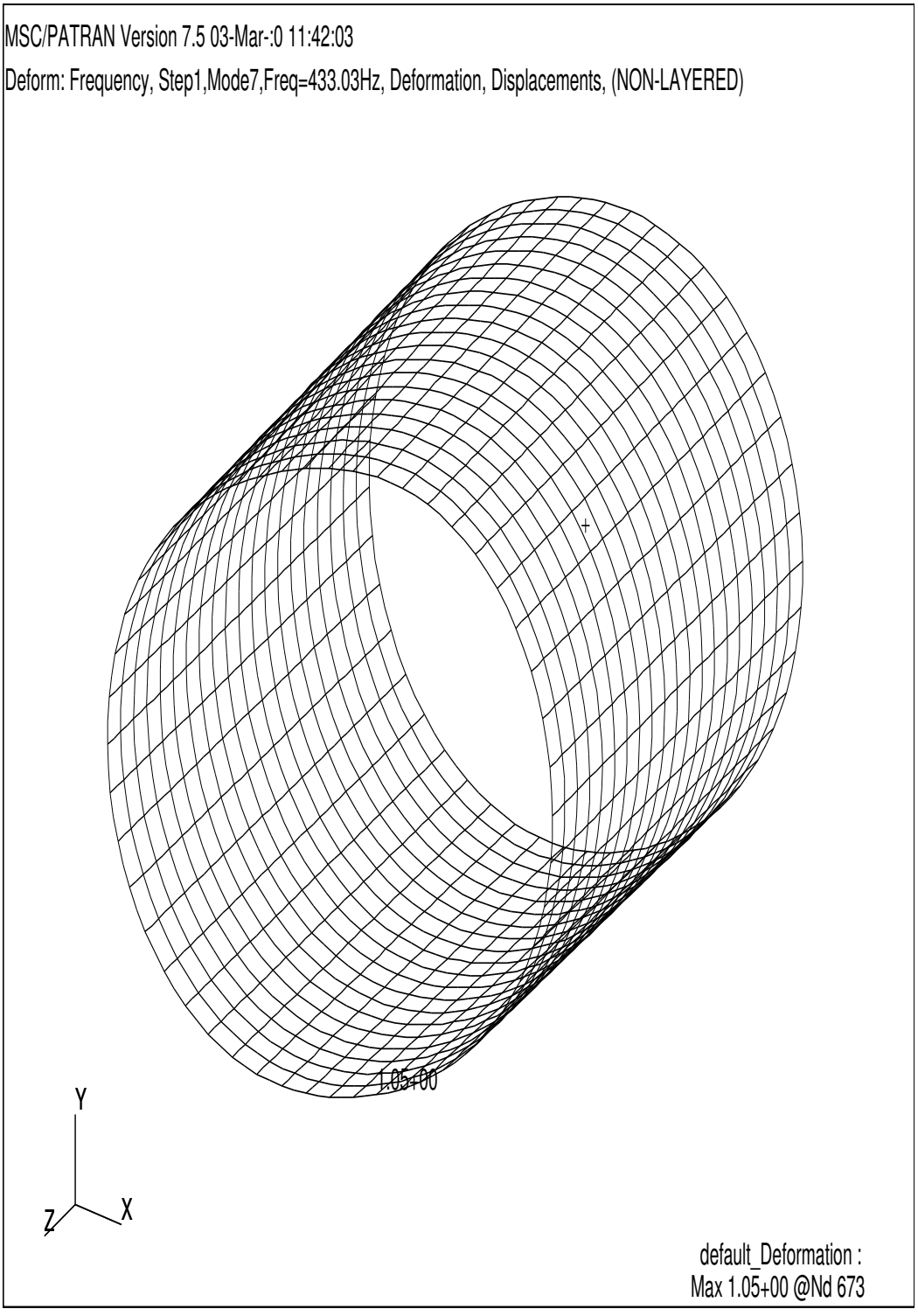


Figure 4.1 A finite element model of a cylindrical shell

4.3.1 Experimental equipment

A schematic diagram of the experimental set-up is shown in Figure 4.2, while details of the measuring equipment are given in Table 4.1. Figure 4.2 shows three main components of the measurement technique implemented:

1. The excitation of the structure: a modal hammer is used to excite the cylinders.
2. The sensing of the response: an accelerometer is used to measure the acceleration response.
3. The data acquisition and processing: the data is amplified, filtered, converted from analogue to digital format (i.e. A/D converter) and finally stored in the computer. The procedure to process the data is explained later in this chapter.

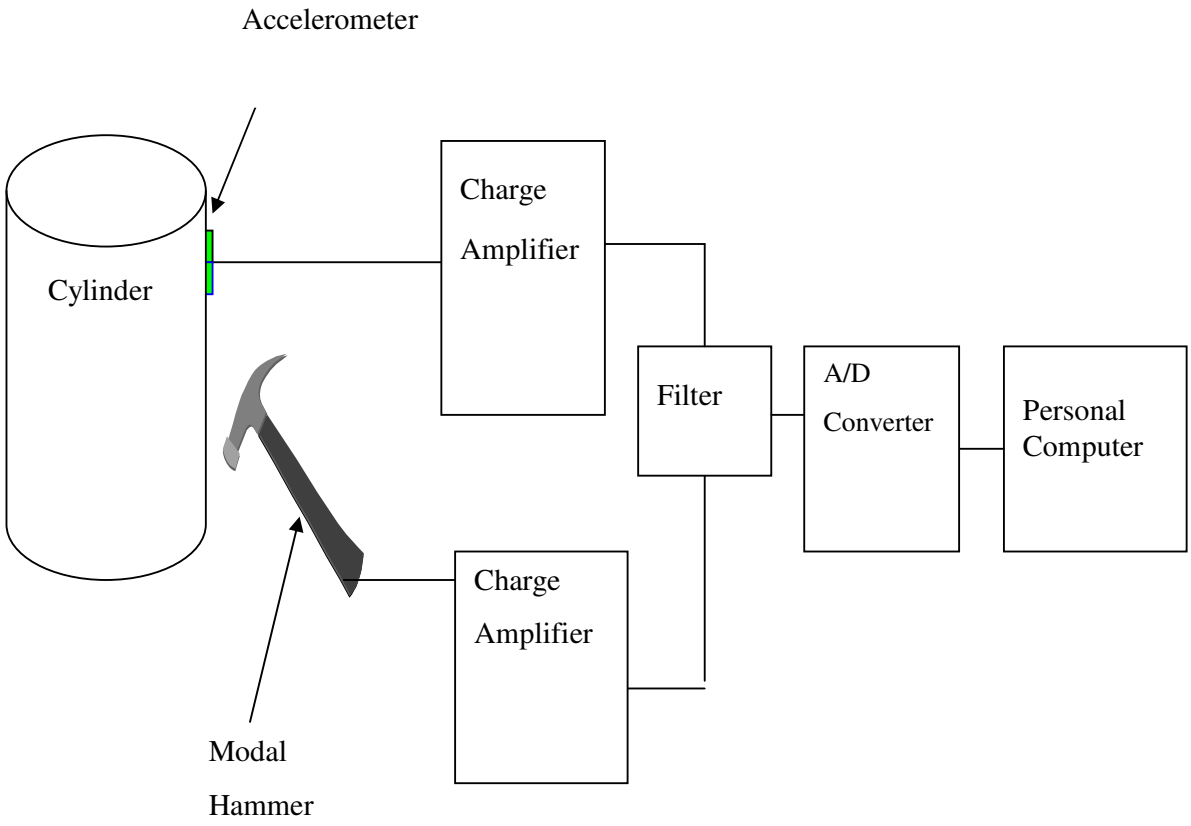


Figure 4.2 Schematic representation of the test set-up.

Table 4.1. Details of the equipment used to measure the cylinders.

Item	Specification	Use
DJB Accelerometers	Sensitivity: 26.6pC/g Mass: 19.8g	Measuring the acceleration response of the cylinders
PCB modal hammer	Sensitivity: 4pC/N Mass of the head: 6.6g	Measuring the force applied using the impulse hammer
Impulse charge amplifier	Sensitivity: 2.0mV/pC Cut-off frequency: 0.44-10kHz	Amplifying the impulse force exerted on the structure
Response charge amplifier	Sensitivity: 14mV/pC Cut-off frequency: 0.44-10kHz	Amplifying the acceleration response
National Instruments DAQCard 1200 data acquisition personal computer card	8-channel 12-bit A/D Gain set to 1000	Sampling transducer time signals
VBF/3 Kemo filter	Cut-off frequency: 0.1Hz-5kHz	Low pass filter
Daytek Desktop Computer	300MHz Pentium microprocessor 64MB RAM 3.99GB	Control of DAQCard 1200, data storage and analysis

(a) Accelerometers and impulse hammer

The cylinders are excited using a hand-held modal hammer. The modal hammer essentially consists of three main components: a handle, a force transducer and a hammer tip. The impact force of the hammer depends on the mass of the hammer and the velocity of the impact. When a modal hammer is used to impact the structure, usually the operator controls the velocity of impact rather than the force itself. The most appropriate way of adjusting the force of the impact

is by adjusting the mass of the hammer. The frequency range, which is excited by the hammer, depends on the mass of the hammer tip and its stiffness. The hammer tip setup has a resonance frequency above which it is difficult to deliver energy into the structure and this resonance frequency may be calculated as (contact stiffness/mass of the tip)$^{0.5}$. In this study the force transducer used is a PCB A218 and a plastic hammer tip is chosen because it is found to deliver adequate energy to excite frequencies within the bounds of our interest. The sensitivity of the transducer is 4pC/N, with a mass of the head of 6.6g.

The responses are measured using a DJB piezoelectric accelerometer with a sensitivity of 2.6pC/ms^{-2} and a mass of 19.8g. A small hole of size of 3mm is drilled into the cylinder and the accelerometer is attached by screwing it through the hole.

(b) Amplifiers

Signals from the impulse hammer and the accelerometer give small charges. As a result the signals need to be amplified by using a charge amplifier. For this purpose in-house charge amplifiers were designed. The acceleration signal is amplified by using a charge amplifier with a sensitivity of 14mV/pC while the impulse signal is amplified by using a charge amplifier with a sensitivity of 2.0mV/pC. These amplifiers have a frequency range of 0.44-10kHz.

(c) Filter

One problem associated with modal testing is a problem of *aliasing*. When vibration signal is measured, it must be converted from analogue into digital form, as it is sampled by an analogue to digital (A/D) converter. This requires that a sampling frequency be chosen. If the signal has significant variation over a short time then the sampling frequency must be high enough to provide an accurate approximation of a signal that is being sampled. Significant variation of a signal over a short period of time usually indicates that high frequency components are present in the signal. If the sampling frequency is not high enough, then high frequency components are not sampled correctly resulting in the problem called aliasing, which is a phenomenon that arises as a result of discretising a signal that was originally continuous. The discretisation process may misinterpret high frequency components of the signal if the sampling rate is too slow, and this may result in high frequency components appearing as low frequency

components. During data acquisition, the data are sampled at a rate at least twice the signal frequency to prevent the problem of aliasing and this technique is due to the Nyquist-Shannon theorem [Maia and Silva, 1997]. In addition, an anti-aliasing filter may be used before the analogue signal is converted into digital format to avoid the aliasing problem. Anti-aliasing filter is a low-pass filter and only allows low frequencies to pass through. This filter essentially cuts off frequencies higher than about half of the sampling frequency.

In this study, the experiment is performed with a sampling frequency of 10kHz and the number of samples taken is 8192. The impulse and the response signals are filtered using a VBF/3 Kemo filter with a gain of 1 and a cut-off frequency of 5kHz.

(d) Data-logging system

The National Instruments DAQCard 1200 of 12-bit over ±5V analogue-digital conversion is used to log the impulse force and the acceleration response. The sampling rate is set to 10kHz, which is sufficient for the frequency bandwidth of interest (i.e. 0-5kHz).

A Visual Basic program running on a Daytek desktop computer controls the DAQCard. This program is used to start data logging, set sampling frequencies, check sample saturation and save the data. After the raw data are measured and saved, they are then opened using MATLAB and checked as to whether they are acceptable or not by calculating the FRFs.

4.3.2 Experimental procedure

Having chosen the frequency bandwidth of interest from a finite element model and the specification of the equipment, and discussed the experimental setup, a description of the experiment is now provided. The experiment is performed on a population of cylinders that are supported by inserting a sponge rested on a *bubble-wrap* to simulate a 'free-free' environment [see Figure 4.3 and for more details see Appendix F]. The sponge is inserted inside the cylinders to control boundary conditions. This will be further discussed below. Conventionally, a 'free-free' environment is achieved by suspending a structure usually with light elastic bands. A 'free-free' environment is implemented so that rigid body modes, which do not exhibit bending or flexing, can be identified. These modes occur at frequency of 0Hz and they can be used to calculate the mass and inertia properties, but are not essential for a successful fault identification

procedure. In this thesis, a 'free-free' environment is approximated by using a bubble-wrap. Testing the cylinders suspended is approximately the same as testing it while resting on a bubble-wrap, because the frequency of cylinder-on-wrap is below 100Hz. The first natural frequency of cylinders being analysed is over 400Hz and this value is several order of magnitudes above the natural frequency of a cylinder on a bubble-wrap. Therefore the cylinder on the wrap is effectively decoupled from the ground. It should be noted that the use of a bubble-wrap adds some damping to the structure but the damping added does not compromise the identification of modes.

The impulse hammer test is performed on each of the 20 steel seam-welded cylindrical shells (1.75 ± 0.02mm thickness, 101.86 ± 0.29mm diameter and of height 101.50 ± 0.20mm). The impulse is applied at 19 different locations as indicated in Figure 4.3: 9 on the upper half of the cylinder and 10 on the lower half of the cylinder. Problems encountered during impulse testing include difficulty of exciting the structure at an exact position especially for an ensemble of structures and in a repeatable direction. Each cylinder is divided into three equal substructures and holes of 10-15 mm in diameter are introduced at the centers of the substructures to simulate faults. Simulating faults by drilling holes has been implemented before. One example of simulating faults this way was conducted by [Kudva *et al*., 1991] where the authors introduced faults by drilling holes in plates. Cylinders are idealised into substructures so that the method of substructuring introduced in Chapter 3 [see Section 3.3.3] can be experimentally verified.

For one cylinder the first type of a fault is a zero-fault scenario. This type of fault is given the identity [0 0 0], indicating that there are no faults in any of the three substructures. The second type of fault is a one-fault-scenario, where a hole may be located in any one of the three substructures. Three possible one-fault-scenarios are [1 0 0], [0 1 0], and [0 0 1] indicating one hole in substructures 1, 2 or 3 respectively. The third type of fault is a two-fault scenario, where one hole is located in two of the three substructures. Three possible two-fault-scenarios are [1 1 0], [1 0 1], and [0 1 1]. The final type of fault is a three-fault-scenario, where a hole is located in all three substructures, and the identity of this fault is [1 1 1]. There are 8 different types of fault cases considered.

The zero-fault and three-fault scenarios are over-represented. This is because each of the 20 undamaged cylinders is progressively damaged by drilling first one, then two and finally three holes. As there is one type of zero- and three-fault cases; and three types of one- and two-fault cases, all cylinders exhibit zero- and three-fault cases but do not all exhibit all one- and two-fault cases.

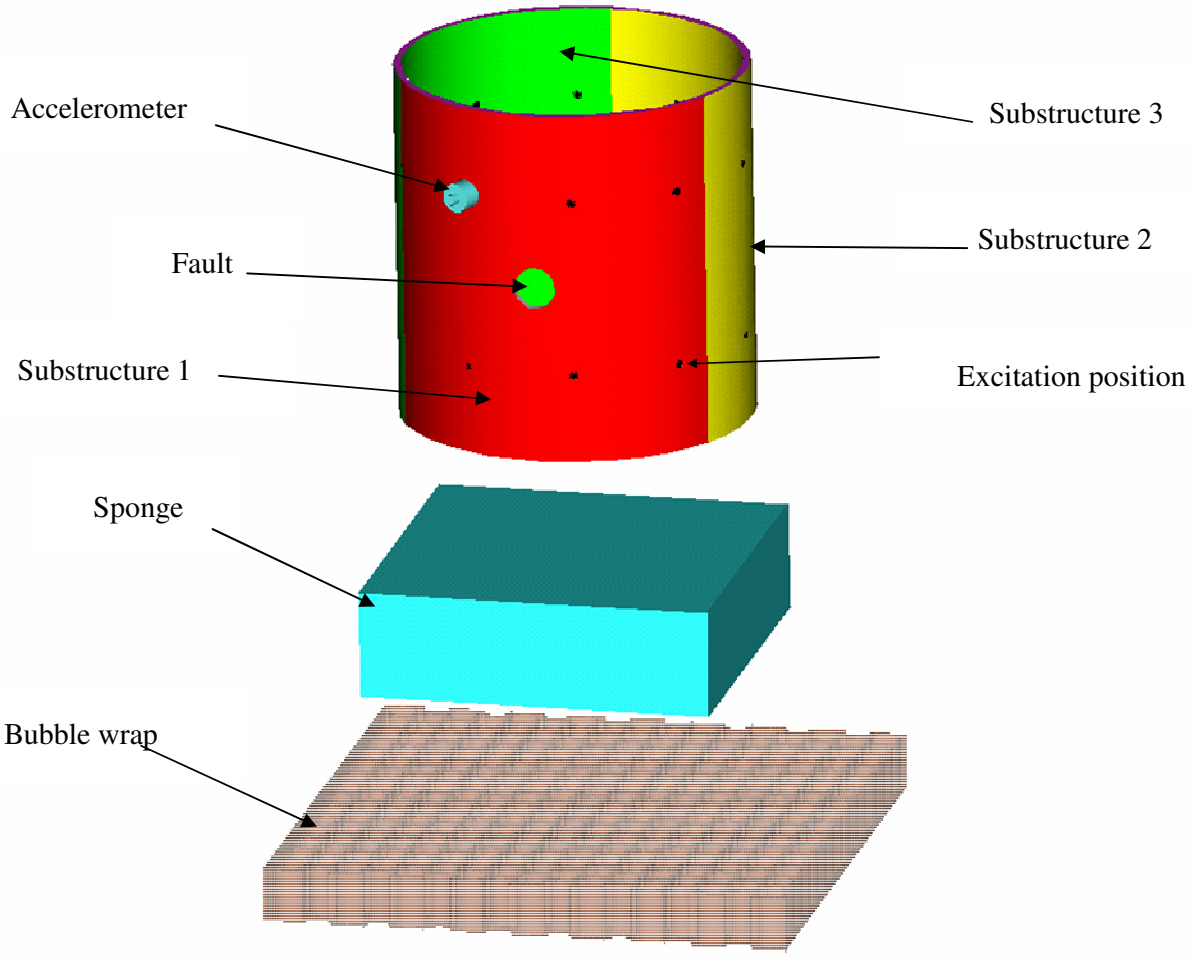


Figure 4.3 Illustration of cylindrical shell showing the positions of the impulse, accelerometer, substructure, fault position and supporting sponge. The sponge is inserted inside the cylinder to control boundary conditions by rotating it every time a measurement is taken. The bubble wrap simulates the free-free environment. The top impulse positions are located 25mm from the top edge and the bottom impulse positions are located 25mm from the bottom edge of the cylinder. The angle between two adjacent impulse positions is 36°. The holes are located at the centre of the substructures with diameter randomly chosen to fall in the interval [10 15] mm.

Because the zero-fault scenarios and the three-fault scenarios are over-represented, some fault cases are measured after increasing the magnitudes of the holes. This is done before the next fault case is introduced to the cylinders. Only a few fault cases are selected because of the limited computational storage space available. In addition, it is the intention of this thesis to devise a fault identification procedure that works well even when there is limited data available because this situation is common in industry [Tarassenko, 1998]. For each fault case, acceleration and impulse measurements are taken. The types of faults that are introduced (i.e. drilled holes) do not influence damping. Each cylinder is measured three times under different boundary conditions by changing the orientation of a rectangular sponge inserted inside the cylinder. Measurements are taken by following these steps:

1. For all 20 cylinders measure vibration data with each cylinder measured three times and this gives 60 (20×3) sets of vibration data.
2. On the 20 cylinders, introduce [1 0 0] fault case to 7 cylinders, [0 1 0] fault case to 7 cylinders; and [0 0 1] fault case to 6 cylinders and measure each cylinder 3 times. This gives 21 [1 0 0] fault case, 21 [0 1 0] fault case and 18 [0 0 1] fault case.
3. Pick 3 cylinders with [1 0 0] fault case and increase the sizes of holes; 3 cylinders with [0 1 0] fault case and increase the sizes of holes; and all 6 cylinders with [0 0 1] case and increase the size of holes and take the measurements. Steps 2 and 3 give 24 measurements for each one-fault case.
4. Introduce additional faults to; 7 cylinders with [1 0 0] case making each a [1 0 1] case; to 7 cylinders with [0 1 0] case making each a [1 1 0] case; and to 6 cylinders of [0 0 1] case making each a [0 1 1] case. Take measurements 3 times and this gives 21 [1 0 1] cases, 21 [1 1 0] and 18 [0 1 1] cases.
5. Pick 3 cylinders with [1 0 1] case and increase the sizes of holes; 3 cylinders with [1 1 0] case and increase the sizes of holes; and all 6 cylinders with [0 1 1] case and increase the sizes of holes and take the vibration measurements. Steps 4 and 5 give 24 measurements for each two-fault case.
6. Introduce additional holes to all the 20 cylinders making all of them a [1 1 1] case and take the vibration measurements three times giving 60 fault cases.

The number of sets of measurements taken for undamaged population is 60 (20 cylinders × 3 for different boundary conditions). All the possible fault types and their respective number of occurrences are listed in Table 4.2. It Table 4.2 it should be noted that the numbers of one- and two-fault cases are each 72. This is because as mentioned above, increasing the sizes of holes in the substructures and taking vibration measurements was used to generate additional one- and two-fault cases.

Table 4.2. The number of different types of fault cases generated

Fault	[0 0 0]	[1 0 0]	[0 1 0]	[0 0 1]	[1 1 0]	[1 0 1]	[0 1 1]	[1 1 1]
Number	60	24	24	24	24	24	24	60

Sample results from the measurements are shown in Figures 4.4 and 4.5. Figure 4.4 shows the impulse force as a function of time and Figure 4.5 shows the acceleration response as a function of time. When measuring these data it is ensured that the impulse force history does not have multiple impulses (i.e. double hits). In addition, a check is made that the acceleration response has decayed to zero by viewing the data before saving.

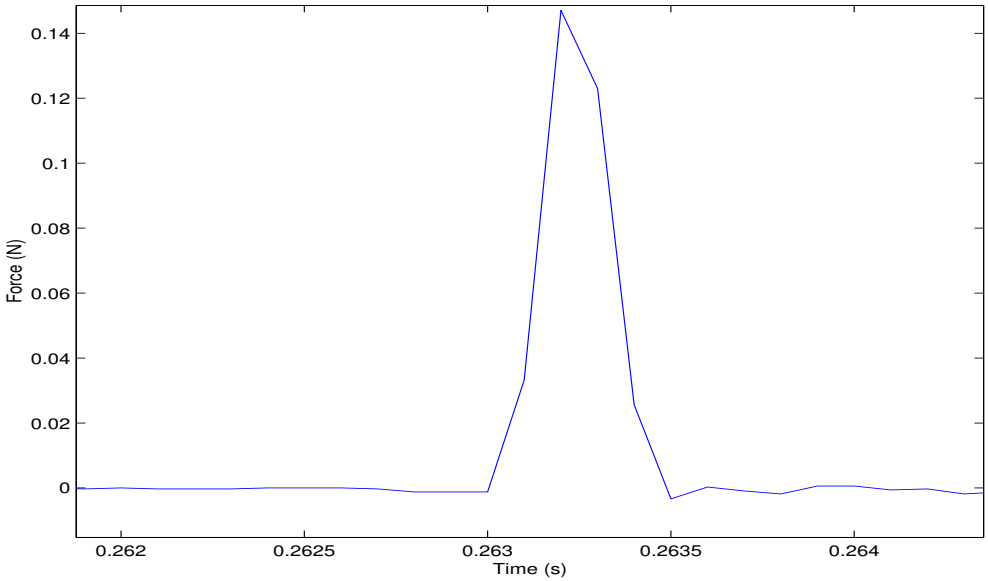


Figure 4.4. The sample impulse force response history.

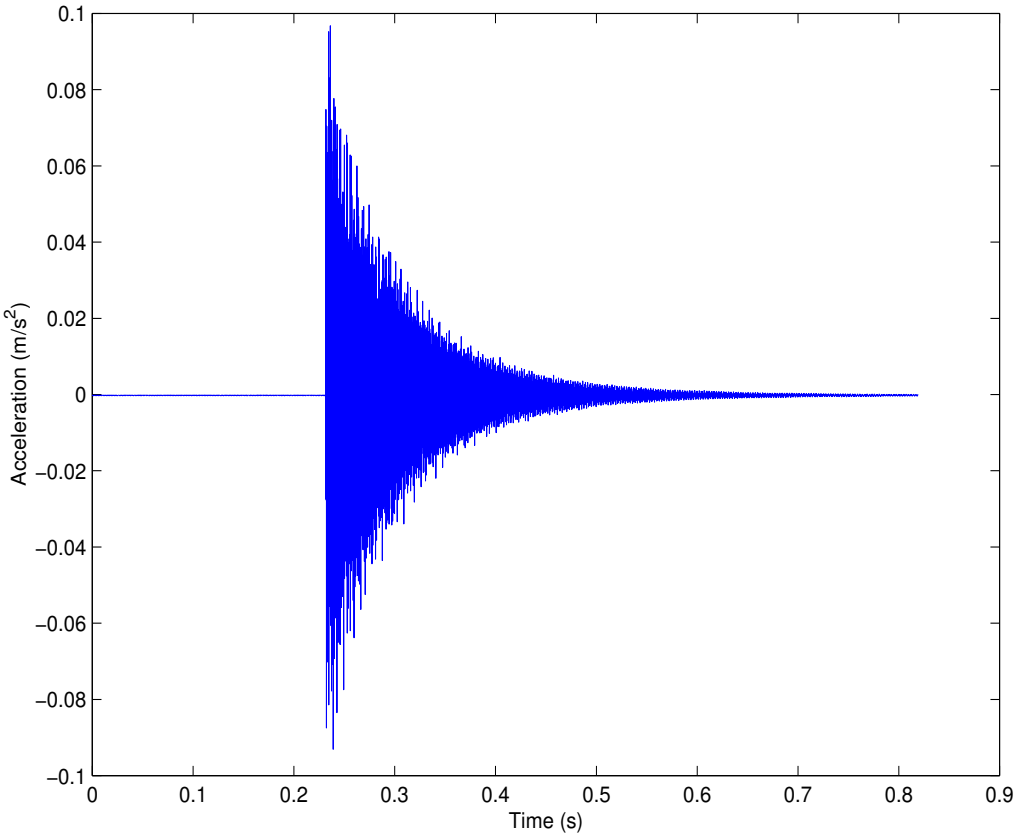


Figure 4.5 The sample acceleration response history.

4.3.3 Signal processing

Now that all the measurements have been taken, the next step is to process the data. Recall from Chapters 2 and 3 that the signals identified for fault identification are the pseudo modal energies and modal properties. The aim of performing signal processing in this study is to calculate these two parameters. However, before this can occur, the problem of the DC offset in the raw data has to be addressed. To do so, two points one before another after the impulse are selected on the impulse history. The mean of the signal before the first point and after the second point is calculated and subtracted from the impulse data to ensure that the force history has a value close to zero before and after the impulse. On the acceleration response, the mean of the signal is calculated and subtracted from the signal.

The impulse and response data that are conditioned in the manner described in this section are processed using the Fast Fourier Transform (FFT) [Cooley and Tukey, 1965] to convert the time domain impulse history and response data into the frequency domain. The data in the frequency domain are used to calculate the FRFs. The sample FRF results from an ensemble of 20 undamaged cylinders are shown in Figure 4.6. This figure indicates that the measurements are generally repeatable at low frequencies and are not as repeatable at high frequencies. Axisymmetric structures such as cylinders have repeated modes due to their symmetry [Royton *et al*., 2000]. The presence of an accelerometer and the imperfection of cylinders destroy the axisymmetry of the structures. The incidence of repeated natural frequencies is destroyed making the process of modal analysis easier to perform [Maia and Silva, 1997].

From the FRFs, the modal properties are extracted using modal analysis [see Appendix C] and the pseudo modal energies are calculated using the integrals under the peaks for given frequency bandwidths using the trapezoidal technique. When the pseudo modal energies are calculated, frequency ranges spanning over 6% of the natural frequencies are chosen. These bandwidths are as follows in Hz: 393-418, 418-443, 536-570, 570-604, 1110-1180, 1183-1254, 1355-1440, 1450-1538, 2146-2280, 2300-2440, 2450-2601, 2500-2656, 3140-3340, 3450-3665, 3800-4039, 4200-4458 and 4640-4928.

The guidelines outlined in Chapter 2 are taken into consideration when choosing these frequency ranges. These guidelines state that the frequency bandwidth must be:

(1) sufficiently narrow to capture the resonance behavior;

(2) sufficiently wide to capture the smoothing out of zero-mean noise;

(3) must not include the regions of the anti-resonance, which are generally noisy.

The sample results from modal analysis are shown in Figure 4.7. This figure illustrates the magnitude and phase of the measured data and the reconstruction from the identified modal properties.

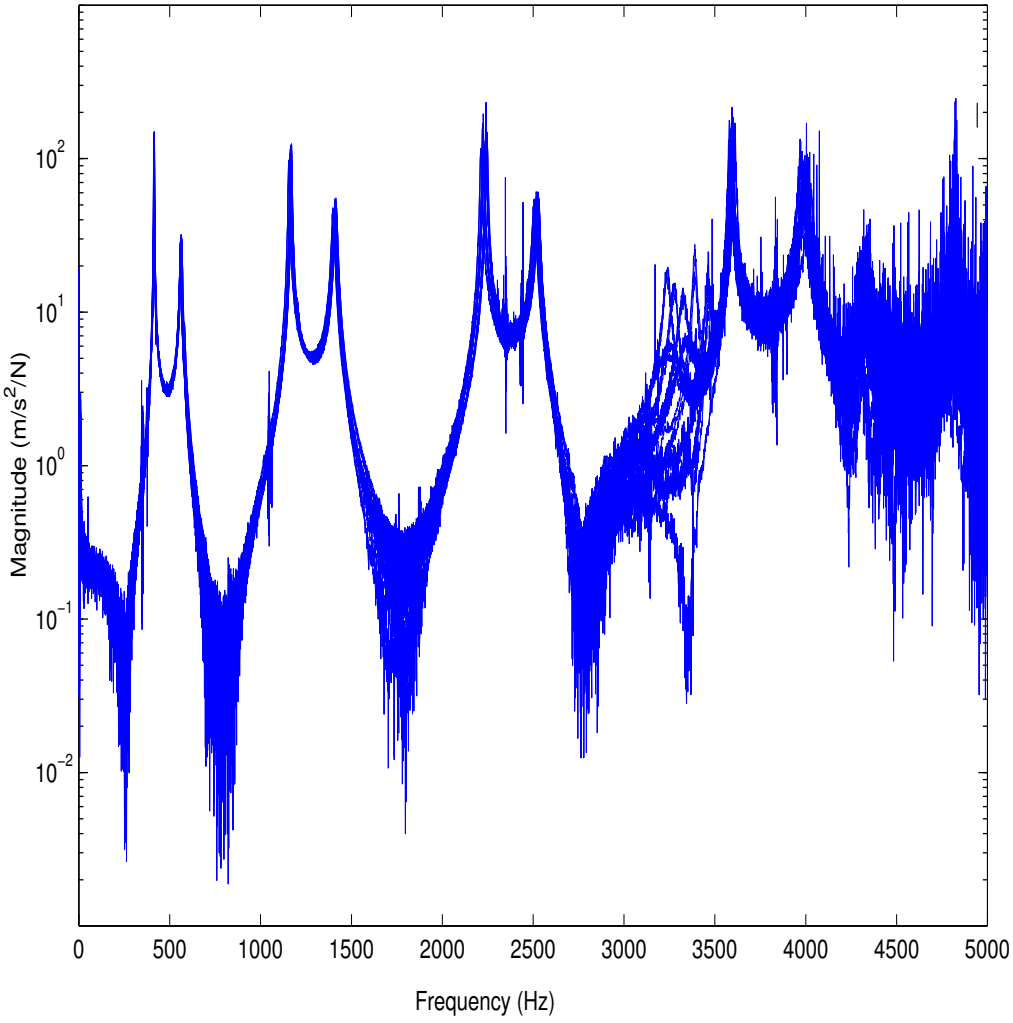


Figure 4.6 The measured FRFs from a population of undamaged cylinders. The different curves correspond to different measurements.

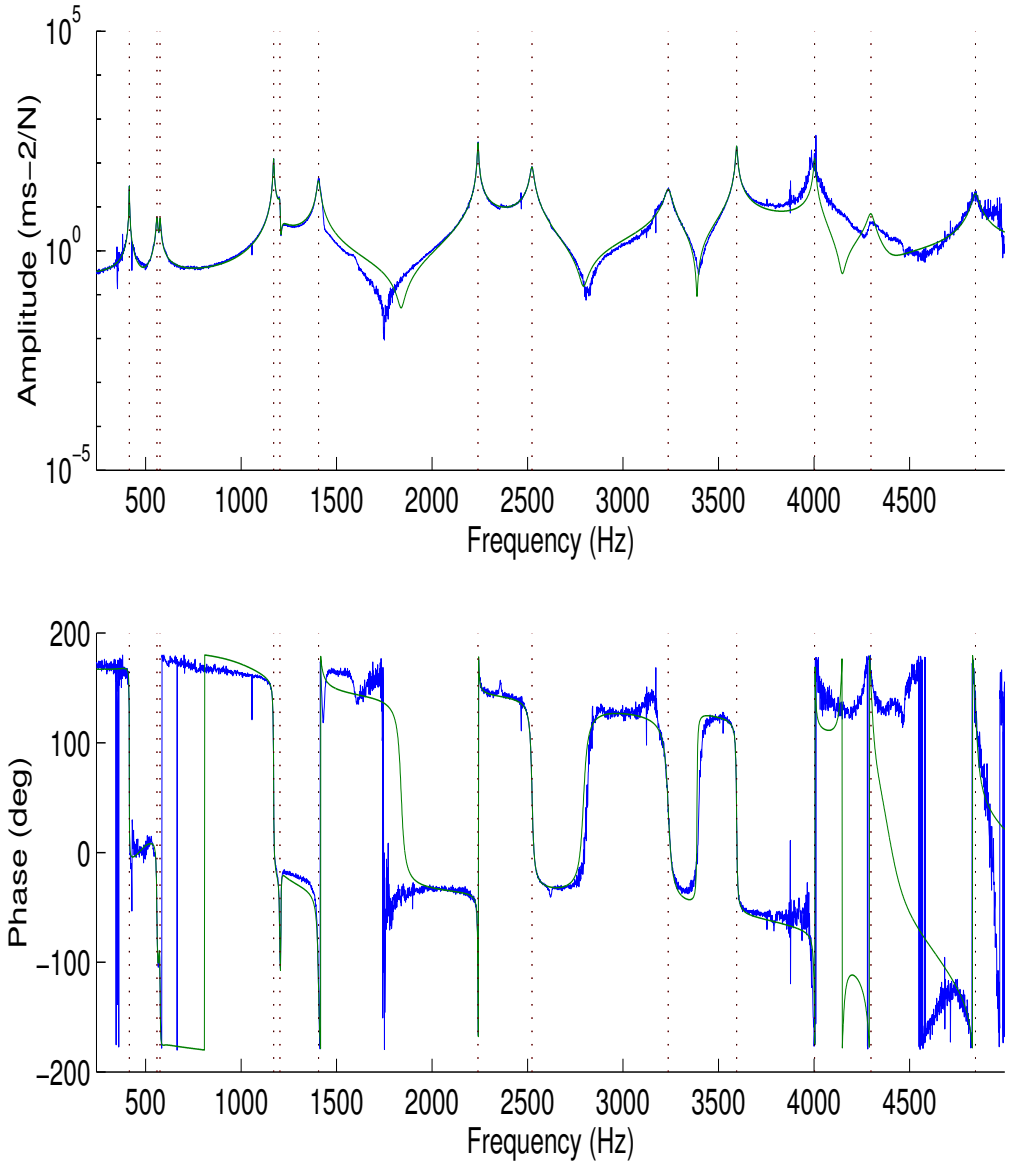


Figure 4.7 A graph showing the amplitude and phase and their respective reconstruction obtained from the identified modal properties.

4.4 PARAMETERS TO BE STUDIED

Having gathered the data and performed the pertinent signal processing, the next stage is to assess the feasibility of using these data for fault identification. As explained at the beginning of this chapter, the literature does not address this situation [Doebling *et al*., 1996]. The most reliable and sensitive pseudo modal energies and modal properties, as well as the COMAC and COPMEAC are used to make the assessment. The most reliable data are those that do not change significantly over a population of undamaged cylinders, while the most sensitive data are those that change significantly when a fault is introduced. In order to select the most reliable and sensitive data, the most sensitive data are chosen from the most reliable data. The details of this procedure are explained later in the chapter. In this section, the method of calculating and comparing the data is described.

4.4.1 Modal properties and pseudo modal energies

The modal properties and pseudo modal energies are chosen by employing the following five steps:

1. Find the means and the standard deviations of the modal properties and pseudo modal energies at each index for data from undamaged and damaged cylinders (e.g. mode 5 coordinate 3 is assigned its own index number). Here an index is any whole number.
2. Calculate the difference between the means of the data from the undamaged and damaged cylinders at each index.
3. Calculate the average of the standard deviations from the undamaged and damaged cylinders while keeping track of the indices.
4. Calculate the *Statistical Overlap Factor,* defined as the ratio between the average of the magnitude of the difference in step 2 to the average-standard-deviations in step 3 at each data index. Statistical overlap factor is analogous to the modal overlap factor which has been used successfully in Statistical Energy Analysis [Langley, 1999], and was defined in Chapter 2.

5. From the statistical overlap factors, select 19 indices with the highest ratios and assess their corresponding data.

4.4.2 Coordinate-modal-assurance-criteria and coordinate-pseudo-modal-energy-assurance-criteria

In this section the method of calculating the COMAC and COPMEAC is discussed. In Section 4.4.1, the method of choosing the most sensitive and reliable pseudo modal energies and modal properties is described. One weakness of this method is that a great deal of information is not used because only those data that are deemed sensitive and reliable are selected. In this chapter it is assessed whether using all the available data by employing the COMAC and the COPMEAC is better than choosing the data using the procedure outlined in Section 4.4.1. The COMAC [Lieven and Ewins, 1988] is a measure of the correlation between two sets of data of the same dimension. The COMAC reduces the dimension of the modal properties to be equal to the number of mode shape coordinates. The COMAC for coordinate i, between the measured data ϕ_m and the median data calculated from a population of undamaged structures ϕ_{MED} may be written as follows:

$$\mathrm{COMAC}(i) = \frac{\left(\sum_{j=1}^{n} \left| \phi_{MED}(i,j) \phi_m^*(i,j) \right| \right)^2}{\sum_{j=1}^{n} \left| \phi_{MED}(i,j) \right|^2 \sum_{j=1}^{n} \left| \phi_m(i,j) \right|^2} \tag{4.1}$$

Here, * stands for complex conjugate. In equation 4.1, when ϕ_m and ϕ_{MED} are perfectly correlated then the COMAC for all degrees of freedom is 1. Otherwise, when perfectly uncorrelated, the COMAC for all degrees of freedom is 0. In this chapter, the natural frequency vector is taken as an additional degree of freedom.

The COPMEAC is a criterion that measures the correlation between two sets of the pseudo modal energies of the same dimension. The COPMEAC for coordinate i, between the measured pseudo modal energies (PME_m) and the median pseudo modal energies (PME_{MED}) calculated from a population of undamaged structures is:

$$\mathrm{COPMEAC}(i) = \frac{\left(\sum_{j=1}^{n}\left|\mathrm{PME}_{\mathrm{MED}}(i,j)\mathrm{PME}^{*}_{m}(i,j)\right|\right)^{2}}{\sum_{j=1}^{n}\left|\mathrm{PME}_{\mathrm{MED}}(i,j)\right|^{2}\sum_{j=1}^{n}\left|\mathrm{PME}_{m}(i,j)\right|^{2}} \tag{4.2}$$

Essentially, equation 4.2 is similar to equation 4.1. The only difference is that equation 4.2 uses the pseudo modal energies while equation 4.1 uses the modal properties. Similarly, when the PME_m and the PME_{MED} are perfectly correlated the COPMEAC for all degrees of freedom is 1. Otherwise, when perfectly uncorrelated, the COPMEAC for all degrees of freedom is 0. It should be noted that the pseudo modal energies are complex.

4.5 RESULTS AND DISCUSSION

The average and sample standard deviations of the natural frequencies for both the damaged and undamaged cylinders are shown in Table 4.2 and compared to the results from the finite element model (which includes the mass of the accelerometer). The data from the damaged cylinders are for the one-fault cases only. The inclusion of the accelerometer breaks the symmetry of the structure thereby making the process of modal analysis easier to perform.

In Table 4.3, it is shown that, for the undamaged cases, the average natural frequency of mode 1 is the most repeatable and, in order of repeatability, followed by 3, 4, 2, 5, 7, 6, 11, 9, 16, 10, 17, 19, 18, 12, 8 and mode 15. For the damaged cases the average natural frequency for mode 1 is the most repeatable, followed by modes 2, 3, 4, 5, 7, 6, 12, 9, 17, 11, 10, 18, 16, 8, 15 and mode 19. These trends are observed by comparing the magnitudes of the sample standard deviations.

In Table 4.3, the natural frequencies corresponding to modes 13 and 14 could not be identified. When the average natural frequencies from a population of undamaged cylindrical shells are compared to those of damaged cylinders, it is observed that none of the natural frequencies decreases by more than 2% except mode 15. Modes 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 15, 16 and 17 are used to calculate the COMAC while the first seven peaks are used to calculate the COPMEAC.

The statistical overlap factors between the data from the undamaged and damaged cylinders (only a one-fault case) are shown in Figure 4.8. This figure indicates that, on average, faults are

mostly observed on the pseudo modal energies, followed by the COPMEAC, then the modal properties followed by the COMAC.

Table 4.3 Natural frequencies for both damaged and undamaged cylinders. The damage cases represent the one-fault cases only.

Mode Number	Finite Element Results (Hz)	Average (f_n) Undamaged (Hz)	Standard deviation (σ) Undamaged (Hz)	Average (f_n) Damaged (Hz)	Standard deviation (σ) Damaged (Hz)
1	433.3	413.7	1.5	412.9	1.9
2	445.5	425.3	3.2	425.1	1.9
3	587.5	561.0	2.4	558.7	2.8
4	599.0	576.6	3.0	576.9	2.8
5	1218.3	1165.0	5.5	1164.6	6.0
6	1262.9	1196.8	6.9	1196.8	7.2
7	1480.0	1408.1	5.7	1404.4	6.3
8	1510.0	1483.4	73.5	1463.4	52.5
9	2273.5	2229.3	11.0	2224.7	11.5
10	2323.6	2346.2	12.6	2360.4	17.5
11	2422.3	2520.1	9.6	2511.4	13.8
12	2657.4	2612.1	39.9	2630.1	10.5
13	2711.3	-	-	-	-
14	2778.4	-	-	-	-
15	3713.7	3330.2	96.5	3239.7	113.3
16	3914.3	3585.8	12.1	3580.7	22.4
17	4138.5	3990.6	16.6	3983.8	13.5
18	4222.8	4309.5	21.2	4316.8	21.0
19	4634.2	4814.2	20.9	4724.7	181.0

In the simulation study conducted in Chapter 2, the pseudo modal energies were found to be more sensitive to faults than the natural frequencies and as sensitive to faults as the mode shapes. In Chapter 2 when identifying the mode shapes it was assumed that the natural frequencies and the damping ratios are known. In this chapter where the experimental data are used, the natural frequencies and damping ratios are not known with absolute certainty, thereby compromising the identification of mode shapes. So the reason why modal properties are poorer indicators of faults than the pseudo modal energies is partly due to the additional uncertainty introduced when identifying the modal properties and that on calculating the pseudo modal energies the zero-mean noise is smoothed out. The use of direct data (the pseudo modal energies and the modal properties) is better than the use of the correlation criteria (the COPMEAC and the COMAC) because the latter average the changes as a result of faults while the former use exact changes.

Figure 4.9 shows the statistical overlap factors obtained when comparing the modal properties between various fault cases.

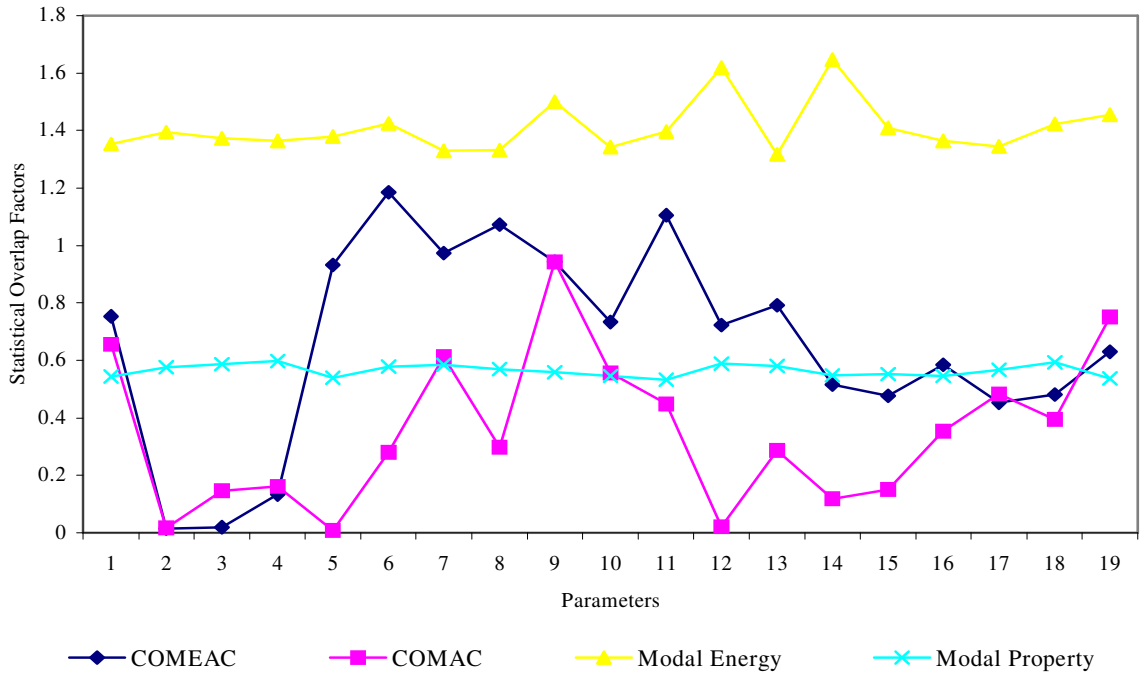


Figure 4.8 Graph of statistical overlap factors between data from undamaged and damaged cylinders (one fault case).

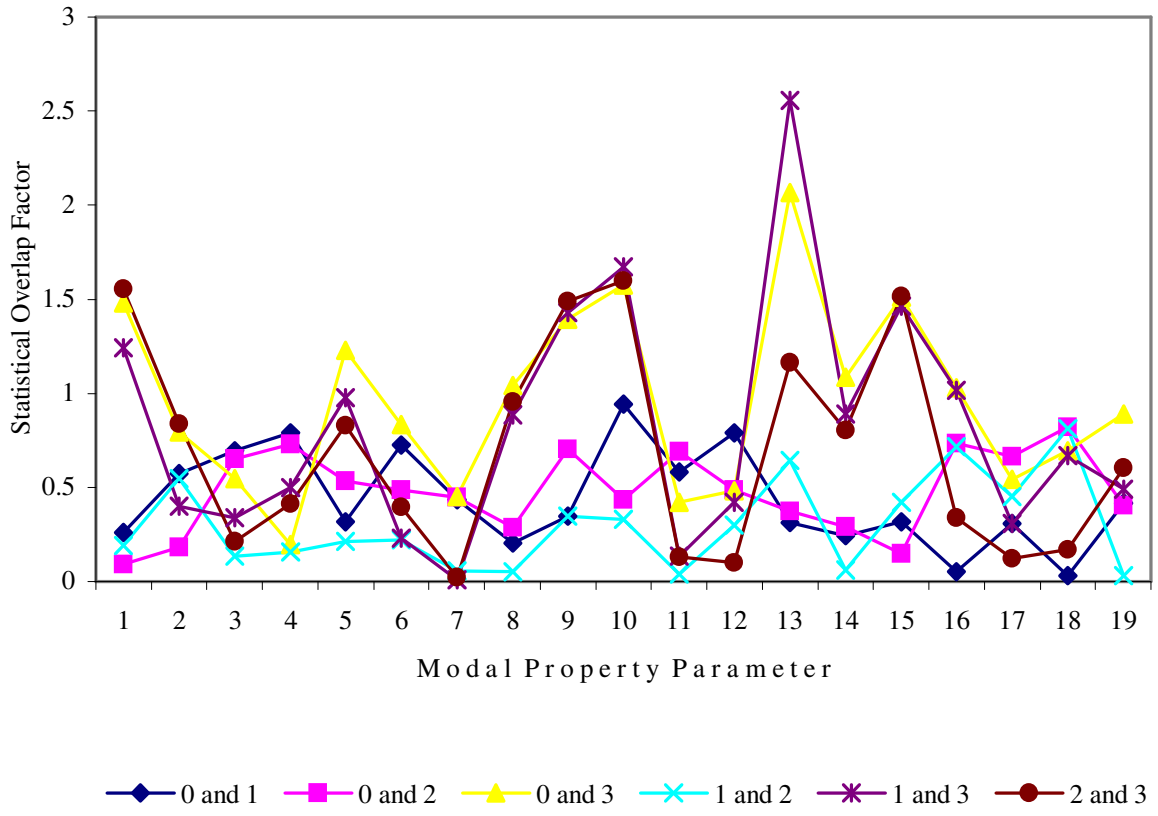


Figure 4.9 Graph of statistical overlap factors between modal properties from undamaged and damaged cylindrical shells. The '0 and 1' shown at the bottom of this figure, indicates the statistical overlap factors between the distribution of all the zero- and one-fault cases over the entire population of cylinders.

Figure 4.9 shows that the first 19 most reliable and sensitive modal properties indicate that the statistical overlap factors vary from 0.0073 to 2.5577 with an average of 0.63. A similar figure to this, but using the pseudo modal energies, is shown in Figure 4.10. The statistical overlap factors for Figure 4.10 are found to vary from 0.0243 to 2.286 with an average of 0.91. The fact that the average statistical overlap factor for the pseudo modal energies is higher than that of the modal properties implies that the distributions of the pseudo modal energies for various fault cases are relatively easier to distinguish than that of the modal properties.

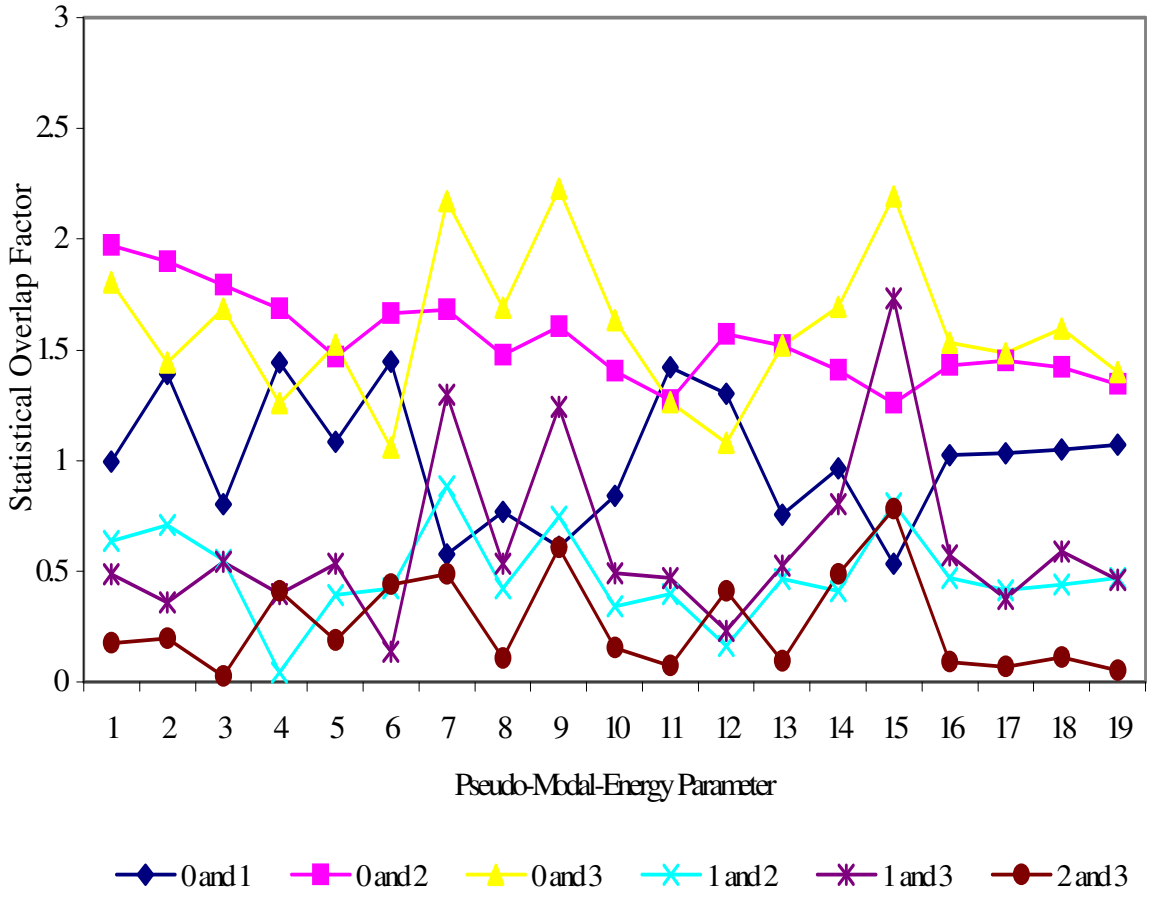


Figure 4.10 Graph of statistical overlap factors between pseudo modal energies from undamaged and damaged cylindrical shells. The '0 and 1' shown at the bottom of this figure indicates the statistical overlap factors between the distribution of all the zero- and one-fault cases over the entire population of cylinders.

The statistical overlap factors for the COMAC data between various fault cases are shown in Figure 4.11. Figure 4.11 shows that the statistical overlap factors vary from 0.0006 to 1.6668 with an average of 0.43. The degree-of-separations of these factors between various fault cases are less than those observed in Figures 4.9 and 4.10.

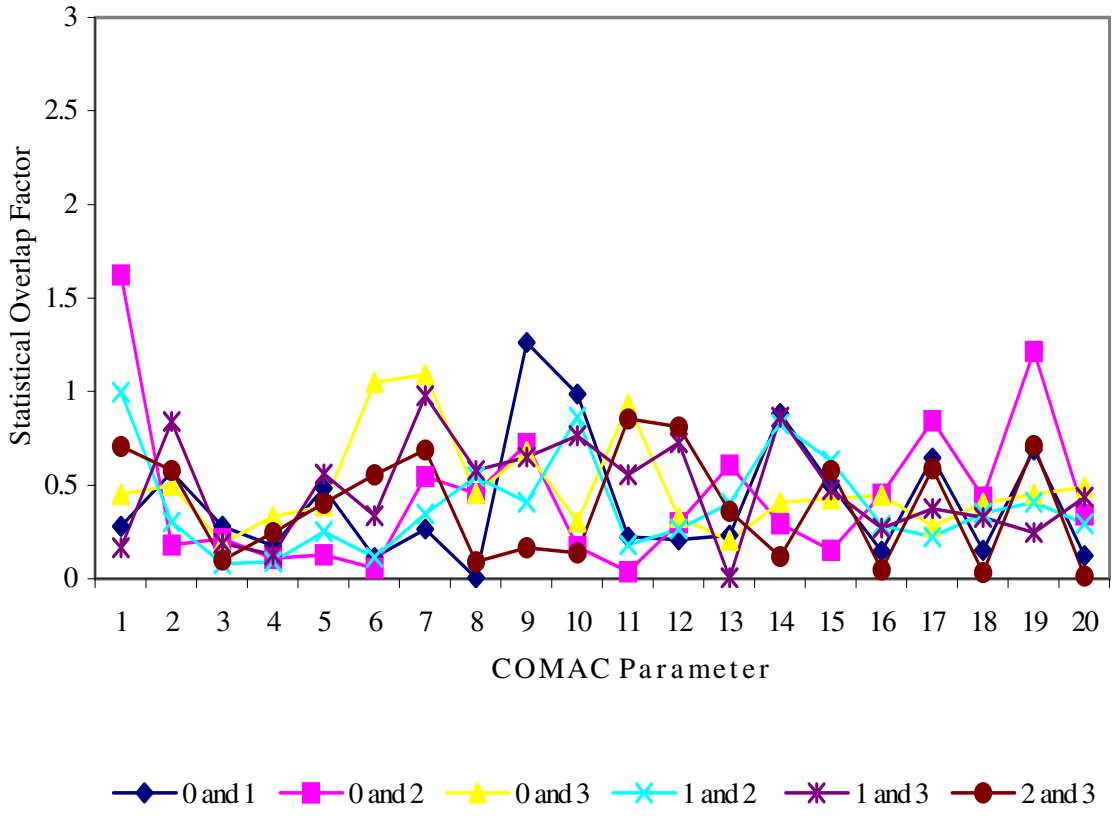


Figure 4.11 Graph of statistical overlap factors between COMAC from undamaged and damaged cylindrical shells. The legend of this figure, for example '0 and 2' represents the statistical overlap factors between the distributions of all the zero- and two-fault cases.

Statistical overlap factors calculated from the COPMEAC are shown in Figure 4.12. Figure 4.12 shows the statistical overlap factors varying from 0.00059 to 1.539 with an average of 0.52. The degree-of-separation between various fault cases when the COPMEAC data are used to calculate the statistical overlap factors, is higher than when the COMAC data are used (Figure 4.11). Furthermore, the degree-of-separation for this case is observed to be less than that obtained when the pseudo modal energies and the modal properties (Figures 4.9 and 4.10) are used directly.

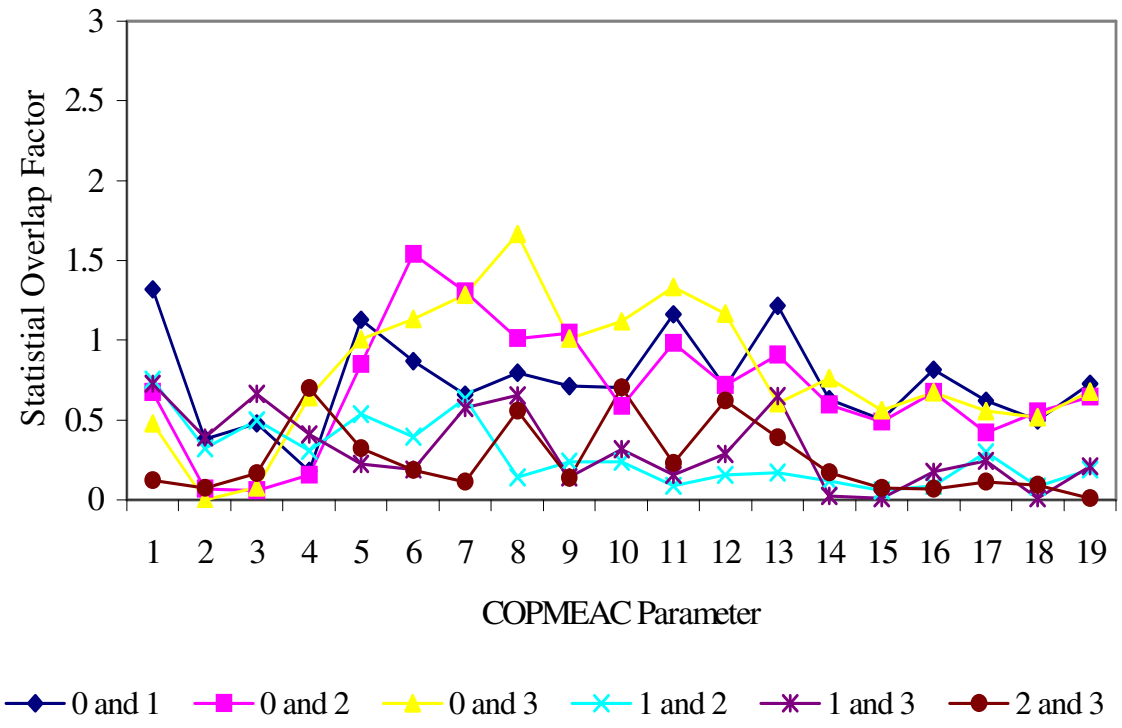


Figure 4.12 Graph of statistical overlap factors between COPMEAC from undamaged and damaged cylindrical shells. The legend of this figure, for example '0 and 2' represents the statistical overlap factors between the distributions of all the zero-fault and two-fault cases.

The results showing the averages of the statistical overlap factors when the pseudo modal energies, the modal properties, the COPMEAC and the COMAC data are used and between various fault cases are listed in Table 4.4. Table 4.4 gives the average results shown in Figures 4.9 to 4.12. From this table, it can be seen that the statistical overlap factors between no-fault- and one-fault cases indicate that the pseudo modal energies are, on average, most sensitive to faults followed by the COPMEAC and the modal properties. The statistical overlap factors between the no-fault- and the two-fault cases indicate that the pseudo modal energies are, on average, better indicators of faults followed by the COPMEAC, then the modal properties. The degree-of-separation between no-fault- and three-fault cases is highest when the pseudo modal energies are used followed by the modal properties and the COPMEAC. When one-fault cases are compared to two-fault cases, the statistical overlap factors indicate that the pseudo modal energies are, on average, the best separators of these faults, followed by the COMAC and then the modal properties.

Table 4.4 Average statistical overlap factors between various fault cases. Here the '0 and 1' indicate the statistical overlap factors between the distribution of zero-fault and one-fault cases.

Parameter	0 and 1	0 and 2	0 and 3	1 and 2	1 and 3	2 and 3
Modal properties	0.44	0.48	0.96	0.36	0.82	0.70
Pseudo modal energies	1.01	1.54	1.59	0.48	0.62	0.26
COMAC	0.41	0.44	0.49	0.39	0.47	0.39
COPMEAC	0.74	0.70	0.80	0.28	0.32	0.25

The statistical overlap factors between one-fault- and three-fault cases show that the modal properties are, on average, better separators of these cases than the pseudo modal energies and COMAC. The degree-of-separation between the two-fault- and the three-fault cases are higher when the modal properties are used than when the COMAC and pseudo modal energies are used.

Figures 4.9 to 4.12 and Table 4.4 show that the pseudo modal energies are the best indicators of faults, followed by the modal properties, the COPMEAC and the COMAC. As the use of the COMAC and COPMEAC, is found to provide a weaker indication of faults than the use of the pseudo modal energies and modal properties directly, the next chapters do not consider them for fault identification.

4.6 CONCLUSION

In this chapter vibration data from a population of cylinders are measured and modal analysis is employed to obtain the natural frequencies and mode shapes. Modal properties are transformed into the coordinate-modal-assurance-criterion. The pseudo modal energies are calculated by integrating the frequency response functions over various frequency bandwidths. The pseudo modal energies are then transformed into the coordinate-pseudo-modal-energy-assurance-criterion. It is observed that using the pseudo modal energies and modal properties directly is better than using the coordinate-modal-assurance-criterion and coordinate-pseudo-modal-energy-assurance-criterion respectively.

Chapter 5

VALIDATION OF THE COMMITTEE-OF-NETWORKS USING EXPERIMENTAL DATA

5.1 INTRODUCTION

Most fault identification methods in mechanical systems, which use vibration data, update the mathematical models until their predictions match the measured data and use the updated parameters to infer the identity of damage. The main problem with this approach is that it relies on the availability of an accurate mathematical model of the structure. In reality, it is very rare that a mathematical model exists which describes the structure accurately enough to be used for fault identification. The use of experimental data and approximation methods, in this thesis neural networks, rather than the exact mathematical model is the subject of this chapter. In the literature, there is little work that is experimentally based that identifies faults in a population of structures [Doebling *et al.*, 1996]. The reason for this is the relatively high expense of acquiring experimental data in which a variety of combinations of faults are introduced to the structure.

This chapter seeks to achieve the following goals:

1. Implement the pseudo modal energies proposed in Chapter 2 and modal properties (MPs) to identify faults using experimentally measured data from a population of cylinders described in Chapter 4.

2. Experimentally validate the proposition that the committee-of-networks, introduced in Chapter 3, gives more accurate results than the individual methods used separately.
3. Implement a fault identification method based on the work of Srinivasan and Kot (1992), who assessed the feasibility of using vibration data to detect faults in cylinders. In this chapter, the method is implemented on the population of cylinders studied in Chapter 4.
4. Apply the multifold cross-validation method [Stone, 1974], to deal with the problem of limited amount of vibration data, for fault identification in a population of cylinders.

5.2 EXPERIMENTAL DATA

In this section, the experiments performed on a population of cylinders and discussed in Chapter 4, are briefly reviewed. The data from the experiments are used for fault identification. To exploit the advantages of the substructuring technique introduced in Chapter 3, each cylinder is divided into three substructures, as discussed in Chapter 4, and holes with diameters varying from 10 to 15mm are drilled in each substructure. Hole sizes are different because this best approximates reality, where faults in structures are not necessarily of the same magnitudes.

As explained in the previous chapter, 8 different fault cases are introduced. Each fault case is given an identity vector, which corresponds to the presence or absence of a fault in a substructure. Since there are three substructures, the identity of a fault is defined as a 3 by 1 vector with the first, second and third components indicating substructures 1, 2 and 3, respectively. For example, assigning a 1 to the first component of the substructure vector indicates the presence of a fault in substructure 1. The total number of zero-fault cases measured is 60, one-fault cases 72, two-fault cases 72 and three-fault cases 60. These fault cases are described in Chapter 4. It was originally intended that each measurement would be taken three times under different boundary conditions. However, to increase the number of one-fault and two-fault cases some of these fault cases were measured more than three times with increased sizes of holes. The reason behind this is explained in Section 4.3.2.

From the measured data, the frequency response functions (FRFs) are calculated using the Fast Fourier Transform [Cooley and Tukey, 1965] as explained in Chapter 4. From the calculated FRFs, the modal properties are identified using the procedure outlined in Appendix C.

The FRFs are also transformed into the pseudo modal energies. The pseudo modal energies are used to train the pseudo-modal-energy-network, and modal properties are used to train the modal-property-network. To train a multi-layer perceptron (MLP) neural network, conventional training, validation and testing data sets are used [Bishop, 1995] as in Chapter 3. Due to the fact that there is a limited amount of data available for the present chapter, there is no validation data set put aside. The networks are validated using the multifold cross-validation technique [Bishop, 1995], which is described later in the chapter. Table 5.1 shows the training and testing data sets for all fault cases that are shown in Chapter 4 (see Table 4.2). For each fault case, e.g. [1 0 1] case, 21 measured examples are randomly selected from the total number of examples shown in Table 4.2. For example, the [0 0 0] fault case has 60 examples and from these 21 examples are randomly selected.

Table 5.1. Fault cases used to train, cross-validate and test the networks. The multifold cross-validation technique [Bishop, 1995] is used because of the lack of availability of data

Fault	[0 0 0]	[1 0 0]	[0 1 0]	[0 0 1]	[1 1 0]	[1 0 1]	[0 1 1]	[1 1 1]
Training set	21	21	21	21	21	21	21	21
Test set	39	3	3	3	3	3	3	39

5.3 NETWORKS ARCHITECTURE AND OTHER ATTRIBUTES

In this section the neural network architecture is described. The following issues are followed to train the neural networks:

1. The type of neural network is chosen. Here an MLP network is chosen because it has been found in the past to be able to solve a fault identification problem using vibration data [Leath and Zimmerman, 1993].
2. The number of hidden layers is chosen. The Universal Approximation Theorem [Haykin, 1999; Lippmann, 1987] states that a two-layered network is sufficient for mapping data of arbitrary complexity.

3. The number of input units is chosen. Here 10 input units are selected using the statistical overlap factor introduced in Chapter 2 and the principal component analysis used in Chapter 3.
4. The number of hidden units is chosen. Twenty pseudo-modal-energy-networks and twenty modal-property-networks with the number of hidden units randomly chosen to fall from 7 to 11 are trained. The pseudo-modal-energy-network and modal-property-network, which give the least mean squared errors during cross-validation, are chosen. This process of selecting the number of hidden units is explained later in Section 5.4.
5. The types of activation functions are chosen. In this chapter the output activation function chosen is a logistic function, described in equation 3.7, while the activation function in the hidden layer is a hyperbolic tangent function. Logistic output function is chosen over linear output function because it has been found to be more suitable for classification problems [Bishop, 1995].
6. The cost function is chosen. The cross-entropy cost function shown in equation 3.10 is used. The reason why this cost function is chosen over the sum-of-square-of-errors is because it has been found to be better suited for classification problems than the sum-of-square-of-errors [Bishop, 1995]
7. The optimisation technique to be used for training is chosen. In this chapter the scaled conjugate gradient method [Møller, 1993] is chosen because it has been found to solve the optimisation problem encountered when training an MLP network more efficiently than the gradient descent and conjugate gradient methods [Bishop, 1995]. The details on these optimisation methods are described in Appendix E.
8. The coefficient of prior distribution, α, of the weight decay regularisation method is chosen (see equation 3.10). This coefficient determines how much the regularisation parameter contributes to the overall error during training. Here by trial and error, α of 0.001 is found to be the most appropriate value for both networks.

5.4 CHOOSING THE NUMBER OF INPUT AND HIDDEN LAYERS

5.4.1. Input data

This section describes the input data that are used to train the neural networks and the type of pre-processing techniques that are performed to reduce the input space. The networks are trained using the modal properties and pseudo modal energies. When training these networks, it is advantageous to make them as small as possible without compromising their abilities to generalise the data because of the phenomenon called the curse of dimensionality [Bellman, 1961] described in Chapter 3. One way of reducing the size of the networks is to reduce the input data through eliminating those inputs that do not contribute significantly to the dynamics of the data that is being modelled. In this study, the dimension of the input space is reduced by employing the statistical overlap factors coupled with the principal component analysis (PCA), and this procedure is described in detail in Chapter 3. The PCA is applied after the data have been normalised using a procedure described in Chapter 3, Section 3.5.1.

Any input reduction scheme is accompanied by the loss of some information. The statistical overlap factors reduce the data by choosing the data that are repeatable for a population of cylinders that are undamaged yet are still sensitive to faults. The PCA reduces the input data by retaining only the data that are independent of one another, yet are sufficiently rich in information about the dynamics of the data. To implement the PCA, first, the covariance matrix of the data is calculated using equation 3.45 shown in Chapter 3. Second, the eigenvalues and eigenvectors of the covariance matrix are computed. Finally, the input data are projected onto the eigenvectors that correspond to the dominant eigenvalues, thereby reducing the input data to the dimension of the eigenvalues chosen. The details of the implementation of the PCA have been described in Chapter 3, Section 3.5.2.

The numbers of pseudo-modal-energies and modal properties identified are 646 (corresponding to 17 natural frequencies×19 measured mode-shape-co-ordinates × 2 for real and imaginary parts of the pseudo modal energy) and 340 (17 modes×19 measured mode-shape-co-ordinates+17 natural frequencies), respectively. The statistical overlap factor and the PCA are used to reduce the dimension of the input data from 646×264 pseudo-modal-energies and

340×264 modal properties to 10×264 (264 fault cases are measured) for both these data types. The first stage of the input dimension reduction scheme is to use the statistical overlap factor. In this chapter, 50×264 pseudo-modal-energies and 50×264 modal properties are selected from 646×264 and 340×264 identified, respectively. The 50×264 pseudo modal energies and modal properties that are chosen are sufficiently repeatable for a population of faultless cylinders yet sufficiently sensitive to the introduction of faults. The statistical overlap factor is implemented by following the steps outlined below:

1. Find the means and the standard deviations of the modal properties and pseudo modal energies at each index for data from undamaged and damaged cylinders (e.g. mode 5 coordinate 3 is assigned its own index number). Here an index is any whole number.
2. Calculate the difference between the means of the data from the undamaged and damaged cylinders at each index.
3. Calculate the average of the standard deviations from the undamaged and damaged cylinders while keeping track of the indices.
4. Calculate the *Statistical Overlap Factor,* defined as the ratio between the average of the magnitude of the difference in step 2 to the average-standard-deviations in step 3 at each data index. Statistical overlap factor is analogous to the modal overlap factor, which has been used successfully in Statistical Energy Analysis [Langley, 1999], and was defined in Chapter 2 by equation 2.12.
5. From the statistical overlap factors, select 50 indices with the highest ratios and assess their corresponding data.

The main problem with using statistical overlap factors is that the data selected may be correlated. To ensure that the data chosen are uncorrelated, the PCA is employed to reduce 50×264 pseudo modal energies and 50×264 MPs selected using statistical overlap factors to 10×264 (here 10 rows are independent). From the 50×264 pseudo modal energies and 50×264 MPs the covariance matrix is calculated using equation 3.45. The eigenvalues and eigenvectors of the covariance matrix are then calculated. Ten eigenvectors, corresponding to the 10 largest eigenvalues, are retained. The variance of the data retained when truncating 50×264 data to

10×264 is 90% for the pseudo-modal-energy-network and 85% for the modal-property-network. This is calculated by dividing the sum of the first 10 dominant eigenvalues by the sum of 50 eigenvalues. The input data (50×264) are then projected onto the corresponding eigenvectors. The new input data are of dimension 10×264.

5.4.2. Number of hidden units

In this section the procedure followed to choose the number of hidden units is described. The output vector to the neural networks is of dimension 3×1. For example, a fault in substructure *n* has a fault identity vector with the n^{th} component containing a 1. Given the input data of size 10×168, and the output vector of size 3×168 for the training set, a network with 157 weights (11 hidden units) is the largest network that could be constructed. Here a value 168 corresponds to the number of training examples, as indicated in Table 5.1.

In this chapter, twenty pseudo-modal-energy-networks and twenty modal-property-networks are trained by randomly choosing the number of hidden units to fall from 7 and 11. The networks (one pseudo-modal-energy-network and one modal-property-network) that give the least mean squared errors, during cross-validation, are selected. The next section details how the networks are trained, cross-validated and tested.

5.5. TRAINING, CROSS-VALIDATION AND TESTING

In this chapter, a multi-layer perceptron network is trained by minimizing the cross-entropy-errors between its prediction and the target data. This procedure of identifying a set of weights by minimising the distance between the predicted and the target data is called the *maximum likelihood approach*.

One practice that has been adopted for training a muti-layer perceptron using back-propagation algorithm is to divide the process into three stages: (1) training, (2) validation and (3) testing. Each stage has a data set assigned to it and ideally these three data sets should be of equal size and must be independent of one another. The network-weights that map the input to output data identified by minimising the network error and using the training data, may over-fit the training data. The neural network that over-fits the training data may not necessarily perform

well on the validation and test data sets. Over-fitting the training data set is a situation where a network stops learning how to approximate the hidden dynamics of the system and learns the noise in the data. In order to combat this problem more networks than required are trained and the network that gives the least mean squared errors on the validation data set is chosen. The chosen network may also over-fit the validation data in addition to the training data, and so the test data set is used to evaluate the performance of the trained network.

In this study there is a limited amount of data available, a situation that is encountered in industry [Tarassenko, 1998]. In the present work, the limitation of the amount of data available is due to the fact that acquiring the data on a population of cylinders is an expensive process. For this reason, the technique of dividing the data into training, validation and testing data sets is not pursued in this chapter but the cross-validation method [Bishop, 1995] that uses a training data set as a validation set is used to estimate the validation error and to choose the number of hidden units in the network. The cross-validation method has been around for some time but was given prominence by Stone (1974) and Geisser (1975). The cross-validation method has been found to give good results for small data sets [Goutte, 1997; Zhu and Rohwer, 1996] and this renders it useful for many practical applications where there is a limited amount of data available. A literature review of the cross-validation method can be found in Stone (1974; 1978). Section 5.5.1 describes the multifold cross-validation procedure, which is one of the variants of the cross-validation methods. The multifold cross-validation method has been used successfully in the past to deal with the problem of limitation of the amount of data available for neural network modelling [Haykin, 1999; Tarassenko, 1998].

5.5.1 The multifold cross-validation method

In this thesis, due to the fact that there is a limited amount of data available, the training data set is also used as a validation data set by employing a multifold cross-validation method [Stone, 1978]. The multifold technique implemented in the present study is illustrated in Figure 5.1.

Each column in Figure 5.1 shows a partition of the training data set and each row represents a training case. The shaded box for a given training case is the partition that is used for validation purposes while the rest of the boxes in the same row are used to train the network.

When the multifold cross-validation technique is implemented, the training data set with N examples is divided into K partitions. Here it is assumed that N is divisible by K and that K > 1. For each training case (see Figure 5.1), the network is trained with the data from all partitions except one and the validation set is the subset that is left out. In Figure 5.1 the partition that is left out for each training case is a shaded one. For example, in Figure 5.1 for Training case 1 the network is trained using Partitions 2 to K and Partition 1 is used as a validation set. The process is repeated for K training cases, by leaving a shaded partition for validation and using the remaining partitions for training. It should be noted that the variant of the multifold cross-validation method implemented in this chapter initialises the network once in Training case 1 (see Figure 5.1). The network-weights obtained after Training case 1 become initial network-weights for Training case 2 and so on. The performance of the resulting network is evaluated by averaging the mean squared errors or classification error under validation over all the training cases.

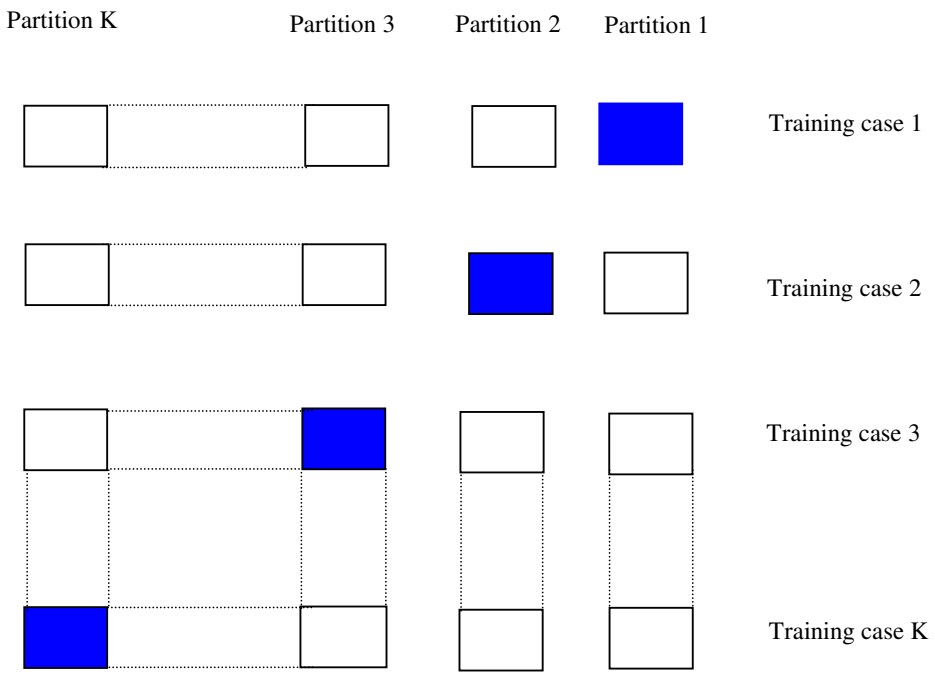


Figure 5.1. An illustration of the multifold cross-validation method used in this chapter. The network is trained K times each time leaving out the data indicated by the shaded area and using the omitted data for validation purposes. The validation error is obtained by averaging the squared error under validation over all the trials of the experiment.

If the limitation of the amount of data is severe, then a method called the leave-one-out method, which is a special case of a multifold cross-validation method is used, where all examples but one are used to train the network and the model is validated on the remaining one. The study conducted by Shao and Tu (1995) suggests that the multifold cross-validation method performs better than the leave-one-out method, for estimating generalisation errors. This is because the leave-one-out method over-fits the data.

For each training session there needs to be a stopping criterion and in this thesis training is stopped after 50 scaled conjugate gradient iterations have elapsed. It should be noted that there are other stopping criteria such as the early stopping method [Bishop, 1995], which are used to prevent over-training of the network. However, the early stopping method is not used in this thesis because, as already mentioned, the weight decay regularisation factor is used to prevent over-training.

5.5.2 Implementation of the multifold cross-validation method

In this chapter the multifold cross-validation method is used to train and validate the pseudo-modal-energy-network and modal-property-network. The fault cases used to train and test the networks are shown in Table 5.1. In Table 5.1 the training data set, with 168 fault cases, has equal number of fault cases indicating that the probabilities of occurrence of eight fault cases are equal. The remaining 96 fault cases are used to test the networks.

The training data set with 168 fault cases is partitioned into 21 subsets. Each partition has 8 different fault cases. This ensures that the training set is balanced in terms of the proportion of fault cases present. The first sets of networks, i.e. pseudo-modal-energy-network and modal-property-network (20 for each method), are trained with 160 fault cases (from Partitions 2 to 21) and the networks are validated on the remaining 8 fault cases (from Partition 1). The network-weights identified in the previous sentence are used as initial weights for Training case 2. The training for this case is conducted using all partitions except Partition 2, which is used to validate the trained networks. The complete training and validation of the networks is conducted 21 times until all the validation partitions have been used.

As already mentioned, 20 pseudo-modal-energies with the number of hidden units randomly chosen to fall from 7 and 11 are trained and validated using the multifold cross-validation technique. The same procedure is used to train 20 modal-property-networks. From these two sets of 20 trained networks, the pseudo-modal-energy-network and modal-property-network that give the least mean squared errors over the validation partitions, are chosen. Each validation partition (see Figure 5.1) gives a mean squared error. The average of the mean squared errors of all the partitions is the validation error used to select the networks.

The features of the pseudo-modal-energy-network and modal-property-network that have the least mean squared errors are listed in Table 5.2. This table also shows that the training, validation and testing mean squared errors are of similar order of magnitudes. This is an indication that the networks have not been over-trained. The classification accuracy rates, defined as the proportion of fault cases that are classified correctly into 8 fault cases, when using the pseudo-modal-energy-network on the training data set is 84.1%. The classification accuracy rate, when using the modal-property-network on the training data set, is 83.5%.

Table 5.2. The properties of the trained networks and the results obtained. Key: PME-N= pseudo-modal-energy-network; MP-N=modal-property-network; MSE=mean squared error; α = coefficient of the regularisation contribution to the cost function

Network	Number of input	Number of hidden nodes	Number of weights	α	MSE (Training)	MSE (validation)	MSE (test data)
PME-N	10	8	115	0.001	0.0240	0.0403	0.0490
MP-N	10	9	129	0.001	0.0307	0.0437	0.0611

5.6 RESULTS AND DISCUSSIONS

In the previous section, two networks (pseudo-modal-energy and modal-property-networks) were trained using data listed in Table 5.1 and cross-validated using the multifold cross-validation method and the training data. As described in the previous section, the multifold cross-validation method uses the training data to evaluate the validation errors. The multifold cross-validation error was used to select the number of hidden units that give the least mean squared error during cross-validation from the two sets of 20 pseudo-modal-energy-network and

modal-property-networks and the resulting networks are listed in Table 5.2. In this section the performances of the two selected networks are evaluated on the test data consisting of 96 fault cases.

From the pseudo-modal-energy-network and modal-property-network, the committee-of-networks is constructed. On calculating the committee, it is assumed that the pseudo-modal-energy-network and the modal-property-network have the same predictive capacity. As a result, the two approaches are given the same weighting functions.

In Chapter 3, it was demonstrated that if the error of the committee method is plotted against the weighting function assigned to one of the networks that form a committee-of-networks, a parabola with a minimum point is obtained. The situation is the same if the standard deviation of squared errors is plotted against the weighting function assigned to one of the two methods that form a committee-of-networks.

Figure 5.2 shows the mean squared errors versus the weighting function assigned to the pseudo-modal-energy-network. In this figure, the y-axis represents the mean squared errors on the 96 fault cases from the test data. The x-axis shows the weighting function assigned to the pseudo-modal-energy-network to form a committee method. A position on the x-axis corresponding to a weighting function of zero assigned to the pseudo-modal-energy-network indicates that corresponding mean squared errors are all due to the modal-property-network. A position on the x-axis corresponding to a weighting function of 1 assigned to the pseudo-modal-energy-network indicates that the corresponding mean squared errors are all due to the pseudo-modal-energy-network. From the performance of the networks on the test data results shown in Figure 5.2, the most optimal committee-of-networks is realised when 0.55-weighting function is assigned to the pseudo-modal-energy-network. The assumed weighting function of 0.5 assigned to each network is sufficiently close to the observed 0.55 because it gives the mean squared error that is close to that given by the optimal committee-of-networks. Furthermore, the assumed committee-weights, i.e. with equal weighting functions assigned to the individual methods, give lower mean squared errors than the two individual methods. It is also observed in this figure that the pseudo-modal-energy-network gives lower mean squared errors than the modal-property-network.

The graph showing the standard deviation of squared errors versus the weighting function assigned to the pseudo-modal-energy-network is shown in Figure 5.3. This figure is obtained by using the test data set. In this figure, the y-axis represents the standard deviation of squared errors. To obtain this figure, 96 fault cases from the test data set are used. The x-axis shows the weighting function assigned to the pseudo-modal-energy-network to form a committee method. From Figure 5.3, the optimal committee-weight is realised when 0.52-weighting function is assigned to the pseudo-modal-energy-network. The assumed committee-weights, i.e. with equal weighting functions assigned to the individual methods, gives a lower standard deviation of squared errors than the two individual methods.

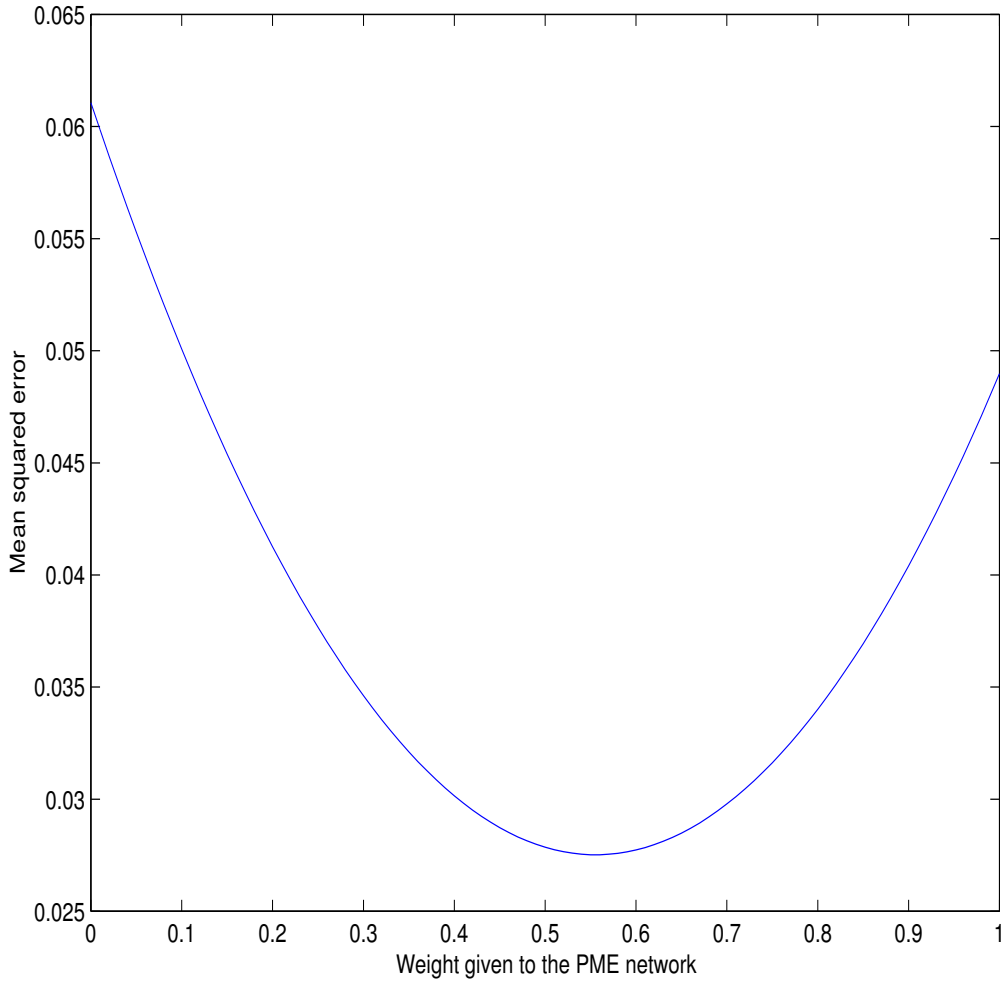


Figure 5.2. Mean squared errors versus weighting function given to the pseudo-modal-energy-network obtained from the test data set.

Another issue that is investigated is the effectiveness of the committee method over the individual methods. This is achieved by using a committee factor defined in equation 3.24. This factor calculates the ratio of the average of the mean squared errors of the two individual methods to the mean squared errors of the optimum committee-of-networks. The committee factor may also be calculated by using the standard deviation of squared errors rather than the mean squared errors. The mean squared errors used to calculate the committee factors are obtained by testing the networks on the test data set. Here it should be noted that the higher the committee factors the more effective is the committee method over the individual methods. Figure 5.2 shows a committee factor of 1.72 while Figure 5.3 shows a committee factor of 2.14. The committee factors, calculated in this chapter, are found to be higher than those calculated in Chapter 3, i.e. 1.40 and 1.48 for mean squared errors and standard deviation of squared errors, respectively. This is because the degree-of-separation between the pseudo-modal-energies and modal properties is higher for experimental data measured in Chapter 4 than for simulated data conducted in Chapter 3. As a result, the assumption made in equation 3.21 becomes more apparent for experimental data than for measured data.

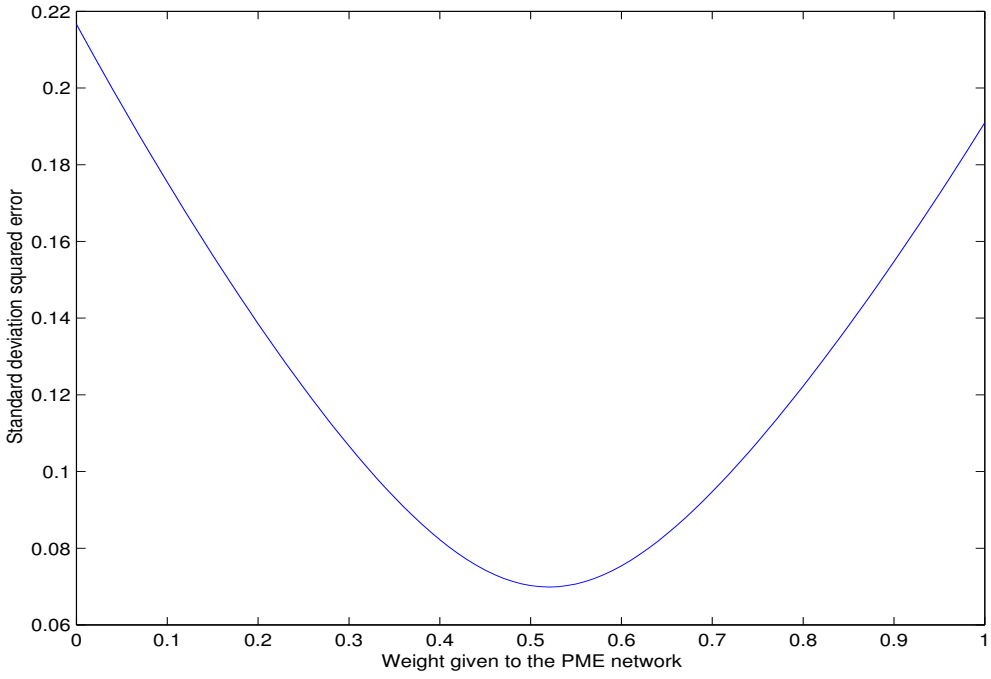


Figure 5.3. Standard deviation of squared errors versus weighting function given to the pseudo-modal-energy-network obtained from the test data set.

Figure 5.4 shows a fault case of identity [1 1 1] and how the three networks predicted this fault case. This example seeks to illustrate how a committee-method can do better than the individual methods. In this figure, the pseudo-modal-energy-network identifies this fault case as a [0.65 1.00 0.49], which when rounded off becomes [1 1 0]. Therefore, the pseudo-modal-energy-network fails to identify the presence of a fault in substructure 1. It is, however, able to detect that the pseudo modal energies measured are from a cylinder with a fault. On the same figure, the modal-property-network identifies this fault case as a [0.39 1.00 1.00], which when rounded off becomes [0 1 1]. The modal-property-network, therefore, correctly detects the presence of faults in the cylinder but fails to detect its exact fault identity. The committee method predicts a fault identity vector of [0.52 1.00 0.75], which when rounded-off becomes a [1 1 1] fault case and this is the correct fault identity. This paragraph has demonstrated how a committee method can perform better than the individual methods.

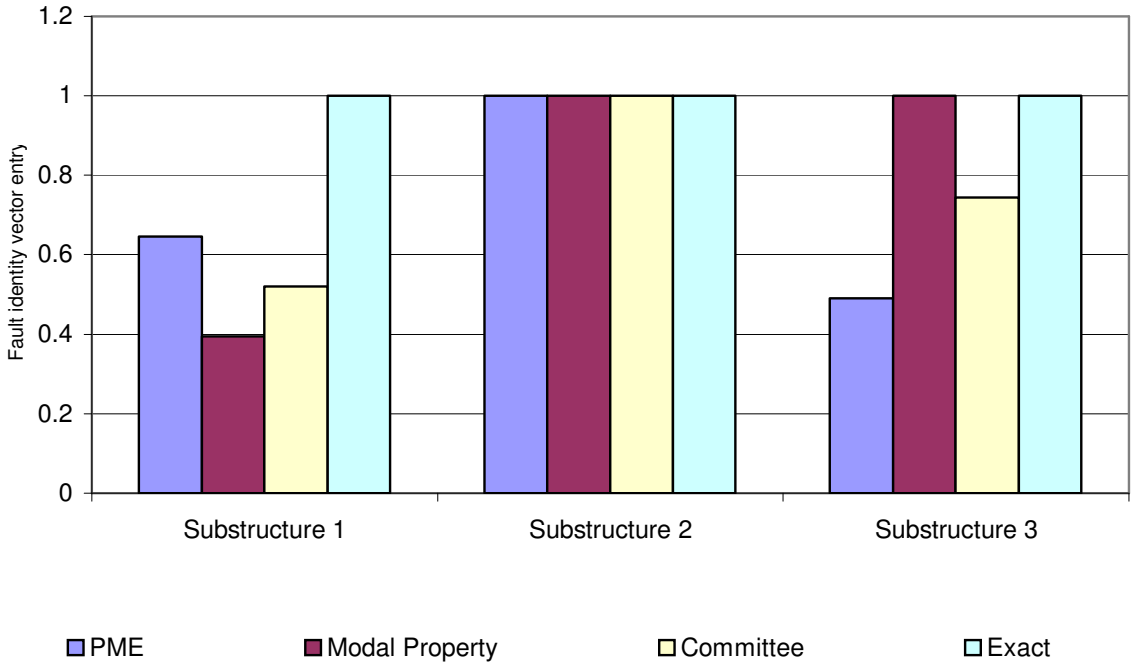


Figure 5.4. An illustration of fault identification results obtained using a modal-property-network, a pseudo-modal-energy-network and a committee method. These results are plotted together with the exact fault case. Key: PME-pseudo modal energy

5.7 CLASSIFICATION OF FAULTS

In the previous sections the pseudo-modal-energy-network and modal-property-network were identified. The two individual networks were given equal weighting functions when combined to form a committee-of-networks. In this section, the pseudo-modal-energy-network, modal-property-network and committee method are used to classify faults in the test data set listed in Table 5.1. This is achieved by classifying the presence or the absence of faults and by classifying fault cases into 8 classes. A fault identity vector is a 3 by 1 vector where component *n* represents the presence or absence of a fault in substructure *n.* Here the maximum value of *n* is 3 because there are three substructures. If a fault is present in substructure *n*, then the n^{th} component of that fault identity vector is 1, otherwise if a fault is absent in that particular substructure then the corresponding fault identity vector component has a value of 0.

One question that needs to be answered is: "*what then constitutes a fault?"* In this thesis the presence of a fault in a substructure is defined as a case where the component of a fault identity vector corresponding to a substructure has a value of 0.5 or more. If a fault identity vector component has a value of less than 0.5 then fault is absent in the substructure corresponding to that vector component.

5.7.1 Detection of faults

In this thesis, detecting faults is defined as classifying faults into two classes i.e. there is no-fault or there is a fault. When detecting faults the following definitions are used:

(1) True positive rate (TP) is the number of positive fault cases that are correctly identified.

(2) False positive rate (FP) is the number of negative fault cases that are incorrectly classified as positive.

(3) True negative rate (TN) is the number of negative fault cases that are classified correctly.

(4) False negative rate (FN) is the number of positive fault cases that are incorrectly classified as negative.

(5) The accuracy is the proportion of the total number of predictions that are classified correctly.

In this section, factors that are viewed as being vital for a fault detection method are discussed. Here the implication of classifying faults into TN, FP, TP and FN is discussed. In this section, the way fault classification impacts on society, which is the main motivation of a fault identification method, is discussed. For any fault identification procedure, the best possible outcome is to have all fault cases classified correctly, or rather, in the semantics that have been adopted in this chapter, TN or TP outcomes. These two cases serve society well. The next level of outcome of a fault identification procedure that is hoped for is to detect wrongly the presence of faults in the structure (FP). The FP case has economic implications and an aircraft example can be used to illustrate this point. Suppose an aircraft component is undamaged but is classified wrongly as being damaged. In this case, the worst that could happen is for that component to be discarded and replaced, which would have economic consequences but would not impact on the safety of passengers. The worst possible scenario is FN, where the presence of faults is wrongly classified as the absence of faults. One example of FN is an aircraft with a fault being predicted as safe. This situation puts the passengers of that aircraft at risk and therefore impacts negatively on society. In this paragraph, the impact on society of certain fault classifications is discussed. The next paragraph describes how well the trained neural networks perform on identifying faults in the test data.

The results showing the detection of faults are given in Table 5.3. In this table, three networks and their corresponding numbers of TN, FP, TP and FN classifications are shown. For example, the pseudo-modal-energy-network gives 36 TN cases.

A successful fault identification procedure gives as many true negatives and true positives as possible and as fewer false positives and false negatives as possible. In Table 5.3 the committee method gives more true negatives than the two approaches acting individually and also gives equally as many true positives as the two individual methods. From this table it is observed that the committee method gives better results than the two individual methods and that the pseudo-modal-energy-network is marginally better than the modal-property-network.

From Table 5.3, Table 5.4 is obtained by calculating the proportion of fault cases that were classified correctly. This assesses the accuracy of the three networks in classifying vibration data into faults/no faults classes. Table 5.4 shows that the committee method gives higher accuracy

than the pseudo-modal-property-network and modal-property-network. The pseudo-modal-energy- and modal-property-networks give similar accuracy levels.

Table 5.3. Classification of faults on the test data set into fault/no-fault classes

Network	True Negative	False Positive	True Positive	False Negative
Pseudo modal energy	36	3	57	0
Modal Property	35	4	57	0
Committee	39	0	57	0

Table 5.4. Accuracy of the classification of faults results in Table 5.3

Network	% Accuracy
Pseudo modal energy	96.9
Modal Property	95.8
Committee	100

5.7.2 Classification of fault cases

This sub-section deals with the classification of faults from the test-data into eight fault cases. As indicated at the beginning of Section 5.7, fault cases given by a network are rounded off to the nearest whole number, i.e. 0 and 1. To assess the predictive capacities of the trained set of networks, a confusion matrix is used and it may be viewed in Table 5.5. In this table the predicted fault cases are shown vertically and the actual fault cases are shown horizontally. A row of this matrix represents all fault cases present in the test data for that particular fault case. For example, a row with a fault case [0 0 0] in Table 5.5 represents the number of [0 0 0] fault cases used in the test data set. From the confusion matrix certain information may be extracted. The diagonal components of this matrix represent fault cases classified correctly, while the off-diagonal components of this matrix represent fault cases classified incorrectly. A perfect fault identification procedure gives a diagonal matrix with all off-diagonal components equal to zero. A completely imperfect confusion matrix gives zero diagonal components and non-zero off-diagonal components.

The results showing the confusion matrices when the pseudo-modal-energy-network, modal-property-network and committee-of-networks are used, are given in Tables 5.5 to 5.7 respectively. In Table 5.5, 92.3% of [0 0 0] cases; all the one- and two-fault cases; and 64.1% of [1 1 1] cases are correctly classified. Of the three [0 0 0] fault cases that are classified incorrectly using the pseudo-modal-energy-network, two are classified as [0 1 0] cases and one as a [0 0 1] case. Of the fourteen [1 1 1] cases that are classified incorrectly by the pseudo-modal-energy-network, four are classified as [1 1 0] cases, three as [1 0 1] cases, six as [0 1 1] cases, and one as a [0 1 0] case.

Table 5.5. Confusion matrix from the classification of fault cases in the test data using the pseudo-modal-energy-network

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	36	0	2	1	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	0	0	3	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	1	0	4	3	6	25

The confusion matrix obtained when the modal-property-network is used is shown in Table 5.6. This table shows that this network classifies 89.7% of [0 0 0] fault cases correctly; all one- and two-fault cases with, the exception of three [1 0 1] cases correctly; and 71.8% of [1 1 1] fault cases correctly. Of the four [0 0 0] cases that are classified incorrectly by the modal-property-network, one is classified as a [0 1 0] case and three as [0 0 1] cases. Of the eleven [1 1 1] cases that are classified incorrectly by the modal-property-network, eight are classified as [1 1 0] cases, one as a [1 0 1] case and two as [0 1 1] cases. The three [1 0 1] cases that are misclassified by the modal-property-network are all classified incorrectly as [0 0 1] cases.

The results obtained when the committee method is constructed by assigning equal weighting functions to the two individual methods, are shown in Table 5.7. This table shows that all the [0 0 0] cases; all the one- and two-fault cases; and 89.7% of the [1 1 1] fault cases, are correctly classified. Of the four [1 1 1] cases that are classified incorrectly by the committee method, three are classified as [1 1 0] cases and one as a [0 1 1] case.

Table 5.6. Confusion matrix from the classification of fault cases using the modal-property-network

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	35	0	1	3	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	3	0	0	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	0	0	8	1	2	28

The committee method classifies all the [0 0 0] cases correctly while the pseudo-modal-energy-network misclassifies three cases and the modal-property-network misclassifies four [0 0 0] cases. The committee method and pseudo-modal-energy-network classify all the one- and two-fault cases correctly, while the modal-property-network misclassifies all [1 0 1] cases. The committee method misclassifies four [1 1 1] cases while the modal-property-network misclassifies eleven [1 1 1] cases and the pseudo-modal-energy-network misclassifies fourteen [1 1 1] cases.

A summary of the classification results in Table 5.5 to 5.7 is shown in Table 5.8. This table is obtained by calculating the proportion of fault cases that are classified correctly into 8 fault cases. This table shows that in classifying all fault cases, the committee method gives the best results followed by the pseudo-modal-energy-network and then the modal property network. It

should be noted that the pseudo-modal-energy-network is only marginally better than the modal-property-network. If, however, account is taken of the fact that the modal-property-network is unable to correctly classify an entire fault case, i.e. [1 0 1], whereas this is never the case for the pseudo-modal-energy-network, then it could be concluded that the pseudo-modal-energy-network is better than the modal-property-network.

Table 5.7. Confusion matrix from the classification of fault cases in the test data using the committee approach

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	39	0	0	0	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	0	0	3	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	0	0	3	0	1	35

Table 5.8. Percentage of fault cases in the test data classified correctly into 8 fault cases.

Network	%Classified correctly
Pseudo modal energy	82.29
Modal Property	81.25
Committee	95.8

The average time taken to train the 20 pseudo-modal-energy-networks and 20 modal-property-networks, is 5.61 and 5.45 CPU minutes, respectively, on a Dell DIMENSIONXPS600 Desktop Computer with 260MB RAM. The procedure used to train the networks is found to be computationally affordable for the industry.

5.8 CONCLUSIONS

In this chapter, a committee-of-networks is used to classify faults in a population of cylindrical shells and thus, is experimentally validated. The statistical overlap factors and the principal component analysis are used to reduce the dimensions of the input data. The multifold cross validation method is implemented to select the optimal number of hidden units amongst 20 trained pseudo-modal-energy-networks and 20 trained modal-property-networks. The results obtained show that the committee method gives more accurate classification of faults than two networks acting individually. The pseudo-modal-energy-network and the modal-property-network are observed to give similar levels of accuracy on classifying faults.

Chapter 6

PROBABILISTIC FAULT IDENTIFICATION USING A BAYESIAN METHOD

6.1 INTRODUCTION

In this chapter, the committee of networks, which employs pseudo modal energies and modal properties simultaneously for fault identification, is extended to a probabilistic framework and experimentally validated. Two Bayesian-formulated neural networks, trained using pseudo modal energies and modal properties, are weight-averaged and used to identify faults in the population of cylindrical shells described in Chapter 4. In Chapters 3 and 5, the committee method was found to give better results than the two individual methods. The Bayesian approach is used because it allows easier determination of the confidence intervals of the fault identities than when using the maximum-likelihood approach implemented in Chapter 5 [Bishop, 1995]. It also automatically penalises highly complex models and is therefore able to select an optimal model without applying independent methods such as cross-validation (as is the case when using the maximum-likelihood approach).

As explained in earlier chapters, each of the 20 cylinders measured in Chapter 4 is divided into three substructures and faults are located within these substructures. The information from the pseudo modal energies and modal properties are transformed into substructure space using the weighted average of the two independent neural networks. The Bayesian approach performs

fault identification by using changes in vibration data resulting from the presence of faults. This is done despite the inherent presence of the changes such as those due to uncertainties in measured data due to variation in physical properties of a population of cylinders, uncertain measurement positions, and changes in support conditions.

6.2 BAYESIAN APPROACH

In Chapter 3 [see Figure 3.1], it is shown that the main goal for supervised neural networks is to find the functional mapping between the input vector $\{x\}$ and the output vector $\{y\}$. In this chapter, the input vectors are the pseudo modal energies and modal properties and the output vector is the indication of the presence or absence of faults in the three substructures. If more than one measurement of the one fault-cases is taken for a population of cylinders under different boundary conditions, there exist different input vectors $\{x\}$ for the same output vector $\{y\}$. If enough measurements are available, the input vector $\{x\}$ for a given fault-case will be random (it has a mean and variance). In other words, the variation in the physical properties of cylinders and (the variation in) their boundary conditions for the one fault-case give rise to the randomness of vector $\{x\}$ for a given fault-case. This randomness of the input vector causes the fault identity vector $\{y\}$ to be probabilistic. This requires the weight-space to be assigned a probability distribution representing the relative degrees of confidence in different values for the weight vector. This implies that the mapping function between the input and output vector has a probability distribution. The weight-space vector is initially assigned some prior distribution. Once the data, in this case the pseudo modal energies or modal properties, and fault identities have been observed, the weight vector can be transformed into a posterior distribution using Bayes' theorem [Bishop, 1995]. In this chapter the posterior distribution is the distribution of the weight vector after the data has been seen. The posterior distribution can then be used to evaluate the predictions of the trained network for data not used during training [Neal, 1992]. Bayes' theorem may be written as follows:

$$P(w \mid D, x) = \frac{P(D \mid w, x)P(w \mid x)}{P(D \mid x)} \tag{6.1}$$

where $P(w|x)$ is the probability distribution function of the weight-space in the absence of any data (also known as the prior distribution) and $D \equiv (y_1,\ldots,y_N)$ is a matrix containing the identity of damage data. The quantity $P(w|D,x)$ is the posterior probability distribution after the data have been seen and $P(D|w,x)$ is the likelihood function. The multi-layer perceptron (MLP) network trained by supervised learning does not model the distribution of the input data. This x is a conditioning variable that always appears on the right-hand side of the probabilities [Neal, 1993]. For the remaining part of this study, x will be omitted to simplify the notation.

The posterior probability, in equation 6.1, of the network weights given the data may be obtained using various techniques such as Monte Carlo based methods or by assuming that the network weights have a Gaussian distribution. MacKay [1991, 1992] approximated the posterior probabilities of the network weights by a multivariate Gaussian distribution around one of its many modes identified by an optimisation procedure during neural network training. Monte Carlo methods do not make the assumption made by MacKay and therefore are more general than the Mackay approach. Some of Monte Carlo methods used to solve this problem include Metropolis algorithm [Metropolis, *et al.,* 1953]. However, this approach is very slow and does not make use of the gradient information given by the backpropagation algorithm. An algorithm that has proven to be successful, the hybrid Monte Carlo method [Duane, *et al.,* 1987], is implemented in this chapter. This approach makes use of the gradient information within the general framework of Metropolis algorithm.

6.2.1 Likelihood function

The likelihood function may be written as:

$$P(D \mid w) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \tag{6.2}$$

Here β is the coefficient of the data contribution to the error. The probability in equation 6.2 is called canonical distribution [Haykin, 1999]. The function E_D is the cost-function (in this chapter the cross-entropy cost function described in equation 3.10 is used) and the function $Z_D(\beta)$ is a normalisation factor given by:

$$\begin{aligned} Z_D(\beta) &= \int \exp(-\beta E_D) dD \\ &= \int \exp(-\beta E_D) dy_i ... dy_N \end{aligned} \tag{6.3}$$

If the identity of the damage data is a smooth function with zero-mean Gaussian noise, the probability of observing the identity of the damage data D for a given input vector may be written by using the cross-entropy cost function as follows:

$$P(D \mid w) = \frac{1}{Z_D(\beta)} \exp\left(\beta \sum_n \sum_k \{ t_{nk} \ln(y_{nk}) + (1 - t_{nk}) \ln(1 - y_{nk}) \} \right) \tag{6.4}$$

where n is the index of the training pattern and k is index of the number of output units. The integral in equation 6.3 is the normalisation factor that can be calculated to give [Bishop, 1995]:

$$Z_D(\beta) = \left(\frac{2\pi}{\beta} \right)^{N/2} \tag{6.5}$$

6.2.2 Prior probability function of weights

The prior probability function for weights may be written in canonical form as:

$$P(w \mid x) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W) \tag{6.6}$$

Here α is the coefficient of the prior distribution. In neural networks, it is known that the best weights that map the input vector to the output vector should be as small in magnitude as possible. To achieve this, the function E_W in equation 6.6 ensures the magnitude of the weight-vector is small. This is done using the regularisation parameter [Vapnik, 1995]. The function $Z_W(\alpha)$ is the normalisation factor given by:

$$Z_W(\alpha) = \int \exp(-\alpha E_W) dw \tag{6.7}$$

If E_W is substituted by the regularisation parameter following Vapnik (1995), the probability of vector w may be written as:

$$\begin{aligned} P(w) &= \frac{1}{Z_W(\alpha)} \exp\left(-\frac{\alpha}{2} \|w\|^2 \right) \\ &= \frac{1}{Z_W(\alpha)} \exp\left(-\frac{\alpha}{2} \sum_{i=1}^{W} w_i^2 \right) \end{aligned} \tag{6.8}$$

The integral in equation 6.7 gives the same form of expression as equation 6.5 and is as follows:

$$Z_W(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{W/2} \tag{6.9}$$

6.2.3 Posterior distribution of weight vector

The distribution of the weights $P(w|D,x)$ after the data have been seen is calculated by substituting equations 6.8 and 6.4 into equation 6.1 to give:

$$P(w \mid D) = \frac{1}{Z_s} \exp\left(\beta \sum_n \sum_k \{t_{nk} \ln(y_{nk}) + (1 - t_{nk}) \ln(1 - y_{nk})\} - \frac{\alpha}{2} \sum_j^W w_j^2\right) \tag{6.10}$$

where

$$Z_S(\alpha, \beta) = \int \exp\left(-\beta E_D - \alpha E_W\right) dw \tag{6.11}$$

E_D and E_W are shown in equations 6.4 and 6.8. The optimal weight vector corresponds to the maximum of the posterior distribution. The distribution in equation 6.10 is a canonical distribution. Equation 6.10 shows that the probability of the existence of states with low errors is higher than the probability of states with higher errors. This is analogous to a physical system where states of low energies are more stable and therefore more prevalent than states with high energies. For the physical system the temperature will come into equation 6.10 by being the denominator of the term in the exponent.

6.3 SAMPLING THE DISTRIBUTION

In this chapter we sample through a distribution described in equation 6.10. Distributions of this nature have been studied extensively in Statistical Mechanics. In Statistical Mechanics macroscopic thermodynamic properties are derived from the state space, i.e. position and momentum, of microscopic objects such as molecules. The number of degrees of freedom that these microscopic objects have is enormous, so the only way to solve this problem is to formulate it in a probabilistic framework.

In the present chapter we borrow some of the techniques that have been used in Statistical Mechanics to sample through a canonical distribution and use them to sample through the weight space in order to identify the distribution given in equation 6.10. Some of the methods that have been implemented in the past to simulate the canonical distributions are Gibbs sampling [Geman and Geman, 1984; Gelfand and Smith, 1990] and Metropolis algorithm [Metropolis, *et al.*, 1953]. In this study we choose to use the hybrid Monte Carlo method to identify the posterior probability of the weight because it makes use of the gradient information which is available via back-propagation. This use of the gradient ensures that we sample through the regions of higher probabilities.

In the present chapter we use the terminology of Statistical Mechanics such as sometimes referring the network weights *w* as a 'position' vector and the network error as 'potential energy'. In order to give the simulation a more physical realism we also include the 'momentum' vector, *p*, of a system and this is described in detail later in the chapter. In Statistical Mechanics the thermodynamic properties a system with many degrees of freedom are described by specifying the position (potential energy) and the momentum (kinetic energy) of every molecule in the system. The sum of the potential energy and the kinetic energy is called the Hamiltonian of the system. The 'position' and 'momentum' vector at a given time is called a state space. There are many states that molecules in a system may exhibit. Some of these states are stable and some are unstable. The main motivation of our work is to sample through the state space and accepting all the states that are stable and accepting with low probabilities those states that are unstable. This is achieved by considering states that are linked together by some transition probability to form a distribution that we seek to sample. A Markov process can be used to simulate this chain of linked states [Ash, 1950; Feller, 1968; Fill, 1991]. In the next section we formally define Markov chains and the properties that we desire in order to sample through the distribution in equation 6.10.

6.3.1. Markov chains

Here we consider a system whose evolution is described by a stochastic process $\{X_1, X_2, ..., X_n\}$ consisting of random variables. A random variable X_n occupies a state x_n at

discrete time n. The list of all possible states that all random variables can possibly occupy is called state space. If the probability that the system is in state x_{n+1} at time $n+1$ depends completely on the fact that it was in state x_n at time n, then the random variables $\{X_1, X_2, ..., X_n\}$ form a Markov chain. Markov chain that we intend to simulate are ergodic [Kemeny and Snell, 1960] — meaning that the probabilities, $P_n(x)$, at time n, converges to a single stationary distribution as n goes to infinity regardless of the initial probabilities. Here this stationary is described by equation 6.11. If a Markov chain is not ergodic, then sampling at different times results with convergence to different distributions, and in the context of neural networks — different distributions of network-weights.

In the present chapter, we sample through state space using the hybrid Monte Carlo method [Duane et. al., 1987]. This technique is viewed as a form of a Markov chain with transition between states achieved by alternating the 'stochastic' and 'dynamic moves'. In this chapter 'stochastic' moves are achieved by unconditionally replacing the momentum vector, p, by one drawn from a Gaussian distribution [Duane et. al., 1987]. The 'stochastic' moves allow the algorithm to explore states with different total energy. The 'dynamics' moves are achieved by using a Hamiltonian dynamics and allow the algorithm to explore states with the total energy approximately constant. Sections 6.3.2 describe in more detail the stochastic dynamics model.

6.3.2. Stochastic dynamics model

As mentioned before, in Statistical Mechanics the positions and the momentum of all molecules at a given time in a physical system define the state space of the system at that time. The positions of the molecules define the potential energy of a system and the momentum defines the kinetic energy of the system. In this chapter, what is referred to in Statistical Mechanics as the canonical distribution of the 'potential energy' is the posterior distribution in equation 6.11. The canonical distribution of the system's kinetic energy is:

$$\begin{aligned} P(p) &= \frac{1}{Z_K} \exp(-K(p)) \\ &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_i p_i^2\right) \end{aligned} \tag{6.13}$$

In molecular dynamics p_i is the momentum of the p^{th} molecule. Here p is not to be mistaken with, P, which indicate probability. In neural network, it is a fictitious parameter that is used to give the procedure a molecular dynamics framework. It should be noted that the weight vector w and momentum vector p, are of the same size. The combined kinetic and potential energy is called the Hamiltonian of the system and can be written as follows:

$$H(w,p) = -\beta\sum_{n}\sum_{k}\{t_{nk}\ln(y_{nk}) + (1-t_{nk})\ln(1-y_{nk})\} + \frac{\alpha}{2}\sum_{j=1}^{W} w_j^2 + \frac{1}{2}\sum_{i} p_i^2 \tag{6.14}$$

In equation 6.14 the first two terms are the potential energy of the system, which is the exponent of the posterior distribution of equation 6.14, and the last term is the kinetic energy. The canonical distribution over the phase space, i.e. position and momentum, can simply be written as follows:

$$P(w,p) = \frac{1}{Z}\exp(-H(w,p)) = P(w \mid D)P(p) \tag{6.15}$$

If we can find a way of sampling through the distribution in equation 6.15, the posterior distribution of weight vector w is obtained by simply ignoring the distribution of the momentum vector p.

The dynamics in the phase space may be specified in terms of Hamiltonian dynamics by expressing the derivative of the 'position' and 'momentum' in terms of fictitious time τ. The word 'position' used here is synonymous to network weights. The dynamics of the system may thus be written by using Hamiltonian dynamics as follows:

$$\frac{dw_i}{d\tau} = +\frac{\partial H}{\partial p_i} = p_i \tag{6.16}$$

$$\frac{dp_i}{d\tau} = +\frac{\partial H}{\partial w_i} = -\frac{\partial E}{\partial p_i} \tag{6.17}$$

The dynamics specified in equations 6.16 and 6.17 cannot be followed exactly and as a result these equations may be discretised using a 'leapfrog' method. The leapfrog discretisation of equations 6.16 to 6.17 may be written as follows:

$$\hat{p}_i(\tau + \frac{\varepsilon}{2}) = \hat{p}_i(\tau) - \frac{\varepsilon}{2}\frac{\partial E}{\partial w_i}(\hat{w}(\tau)) \tag{6.18}$$

$$\hat{w}_i(\tau+\varepsilon) = \hat{w}_i(\tau) + \varepsilon \hat{p}_i(\tau+\frac{\varepsilon}{2}) \tag{6.19}$$

$$\hat{p}_i(\tau+\varepsilon) = \hat{p}_i(\tau+\frac{\varepsilon}{2}) - \frac{\varepsilon}{2}\frac{\partial E}{\partial w_i}(\hat{w}(\tau+\varepsilon)) \tag{6.20}$$

Using equation 6.18, the leapfrog takes a little half step for the momentum, p, and using equation 6.19 takes a full step for the 'position' or rather the network weights w and using equation 6.20 takes a half step for the momentum, p. The combination of these three steps form a single leapfrog iteration and calculates the 'position' and 'momentum' of a system at time $\tau+\varepsilon$ from the network weight vector and 'momentum' at time τ. The above discretisation is reversible in time, it almost conserve the Hamiltonian (representing the total energy) and preserves the volume in the phase space, as required by Liouville's theorem [Neal, 1993]. The volume preservation is achieved because the moves the leapfrog steps take are shear transformations

One issue that should be noted is that following Hamiltonian dynamics does not sample through the canonical distribution ergodically because the total energy remains constant, but rather at most samples through the microcanonical distribution for a given energy. One way used to ensure that the simulation is ergodic, is by changing the Hamiltonian, H, during the simulation. To achieve ergodicity, the 'momentum' p is replaced after before the following leapfrog iteration. In this chapter a normally distributed vector with a zero-mean replaces the 'momentum' vector.

The dynamic steps introduced in this section make use of the gradient of the error with respect to the 'position' (network weights). In this chapter this gradient is evaluated by using the back-propagation algorithm outlined in Appendix D.

6.3.3. Metropolis algorithm

An algorithm due to Metropolis, *et al*. (1953) has been used extensively to solve problems of Statistical Mechanics. In Metropolis algorithm if we wish to sample for a stochastic process $\{X_1, X_2, ..., X_n\}$ consisting of random variables, we randomly consider changes to X accepting or rejecting the new state according to the following criterion:

$$\begin{aligned} &\text{if } H_{new} < H_{old} \text{ accept state } (w_{new}, p_{new}) \\ &\text{else} \\ &\text{accept } w_{new} \text{ with probability} \\ &\exp\{-(H_{new} - H_{old})\} \end{aligned} \tag{6.21}$$

In this chapter we view this procedure as a way of generating a Markov chain with the transition from one state to another conducted using the criterion in equation 6.21. By investigating carefully equation 6.21, it may be observed that states with high probability will form the majority of the Markov chain, and those with low probability will form the minority of the Markov chain. However, simulating a distribution by perturbing a single vector X and in the context of neural networks, a weight vector w is infeasible due to high dimensional nature of the space and the variation of the posterior probability of weight vector w. A method that exploits the gradient of the Hamiltonian with respect to the weight vector, w, is used to improve the Metropolis algorithm described in this section and is the subject of the next section.

6.3.4. Hybrid Monte Carlo

Hybrid Monte Carlo method works by taking a series of trajectories from the initial states, i.e. 'positions' and 'momentum', and moving in some direction in the state space for a given length of time and accepting the final states using Metropolis algorithm. The system moves from one state to another by combining the Hamiltonian dynamics described in Section 6.3.2 and the Metropolis algorithm described in Section 6.3.3. The validity of the hybrid Monte Carlo rests on three properties of Hamiltonian dynamics and these are:

1. Time reversibility: it is invariant under $t \rightarrow -t$, $p \rightarrow -p$.
2. Conservation of energy: the H(w,p) is the same at all times.
3. Conservation of space-state volumes due to Liouville's theorem [Neal, 1993].

For a given leapfrog step size, ε_0, the number of leapfrog steps, L, the dynamic transition of the hybrid Monte Carlo procedure is conducted as follows:

1. Randomly choose the direction of the trajectory λ to be either –1 for backward trajectory and +1 for forward trajectory.

2. Starting from the initial state (w,p) perform L leapfrog steps with the actual step size $\varepsilon = \varepsilon_0(1 + 0.1k)$ resulting in state (w^*,p^*). Here ε_0 is a chosen fixed step size and k is the number chosen from a uniform distribution and lies between 0 and 1. The reason why this step size is used is explained later.
3. Reject or accept (w^*,p^*) using Metropolis criterion. If the state is accepted then the new state becomes $(w,p) = (w^*,p^*)$. If rejected the old state is retained as a new state.

After implementing step (3) the momentum vector is reinitialised before moving on to generate the subsequent state. In this study, we choose to randomly generate the momentum vector from a Gaussian distribution before starting to generate the subsequent state.

One remark that could be made about the hybrid Monte Carlo method is that it makes use of the gradient information in step (2) above via the leapfrog steps described in Section 6.3.2. The advantages of using this gradient information is that the hybrid Monte Carlo trajectories move in the direction of high probabilities resulting in the improved probability that the resulting state is accepted and that they are not highly correlated. In neural networks, this gradient is calculated using back-propagation, which is described in Appendix D. In step (2) the leapfrog actual step size varies from ε_0 to $1.1\varepsilon_0$. The number of leapfrog steps, L, must be significantly higher than one to allow a faster exploration of the state space. If L is too large this could lead to a higher rejection rate at step (3) above. The choice of ε_0 and L only affects the speed at which the simulation converges to a stationary distribution and the correlation between the samples accepted. The leapfrog discretisation does not introduce systematic errors due to the occasional rejection of states, which result, with the increase of the Hamiltonian.

In step (2) of the implementation of the hybrid Monte Carlo method, the actual step size $\varepsilon = \varepsilon_0(1 + 0.1k)$ where k is uniformly distributed between 0 and 1, is not fixed. This in effect ensures that the actual step size for each trajectory is varied and so that the accepted states do not have a high correlation [Mackenzie, 1989]. The same effect can be achieved by varying the trajectory length. In this chapter we choose to vary the step size.

One problem with the hybrid Monte Carlo method is that the simulation may spend a great deal of time in the region of relatively high Hamiltonian corresponding to local minimum in the

regularised error function. A technique that could be implemented to deal with this problem is simulated annealing [Kirkpatrick, *et al.*, 1983]. However, preliminary investigation of the use of this technique has found that this method is not essential for a problem we are dealing with in this thesis.

6.3.5 Distribution of network outputs

The application of the Bayesian approach to neural networks results in weight vectors that have a mean and standard deviation and thus have a probability distribution. As a result, the output parameters have a probability distribution. Following the rules of probability theory, the distribution of the output vector y for a given input vector x may be written in the following form:

$$p(y \mid x, D) = \int p(y \mid x, w) p(w \mid D) dw \tag{6.22}$$

In this chapter, the hybrid Monte Carlo method is employed to determine the distribution of the weight vectors, and subsequently, of the output parameters. The integral in Equation 6.22 may be approximated as follows:

$$I \cong \frac{1}{L} \sum_{i=1}^{K+L-1} F(w_i) \tag{6.23}$$

Here K is the number of initial states that are discarded in the hope of reaching a stationary distribution and L is the number of retained states.

6.3.5 Committee of neural networks

The committee procedure, introduced in Chapter 3, is adapted to the Bayesian framework by generating I_1 and I_2 [in equation 6.23] representing the distribution of the fault identities for the given set of modal properties and pseudo modal energies. The overall distribution of the committee is then evaluated by calculating the weighted-average of I_1 and I_2. In this chapter, on calculating the committee method, it is assumed that the pseudo modal energies and the modal properties are equally reliable on identifying faults. As a result, the two individual networks are given equal weighting functions when forming a committee.

6.4 EXPERIMENTAL IMPLEMENTATION OF BAYESIAN NETWORKS

In Chapter 5, maximum-likelihood formulated networks are used to identify faults in a population of cylindrical shells. Similarly, in this section, the Bayesian-formulated networks are used to identify faults in cylindrical shells. The networks are trained and tested using the data described Table 5.1. The architectures of the MLP networks (i.e. the number of input and output vectors, activation functions, cost function and the number of hidden units) are the same as those in Chapter 5 and are shown in Table 6.1. Making the neural network architectures of the present chapter and Chapter 5 the same ensures that we are able to consistently compare the maximum-likelihood formulation of Chapter 5 to the Bayesian formulation of the present chapter. The information on how the number of input units, the number of hidden units and the coefficient of prior distribution are chosen is described in Chapter 5.

On implementing the hybrid Monte Carlo method the parameters used are listed in Table 6.2. This table shows the following:

(1) The number of initial states discarded, K, in the hope of reaching a stationary distribution.

(2) The number of steps in each hybrid Monte Carlo trajectory; the fixed step size that is used as a basis of choosing the actual step size via equation $\varepsilon = \varepsilon_0(1 + 0.1k)$.

(3) The number of samples retained to form a distribution in equation 6.22.

The number of states that are omitted is chosen to be 100 and the number of states retained, L, is 500. This is sufficient to give a stationary posterior probability of network weights because the acceptance rate of states was over 98%. The number of inputs and hidden units used are the same as those that were identified in Chapter 5. The number of output units corresponds to the number of substructures. The data contribution to the error function was chosen arbitrarily. The coefficient of prior used is the same as the one used in Chapter 5. The fixed step size was chosen through trial and error by investigating how a chosen step size influence the acceptance rate of the states (network-weights and momentum) visited. This step size should be as close to zero as possible. If the step size is too low, then the dynamics of the hybrid Monte Carlo method through the state space takes a long time to converge to a stationary posterior distribution. If it is

too large, then the process can potentially miss the stationary distribution. Here the number of steps in each trajectory is 100.

Table 6.1. The architectures of the neural networks trained. Key: α = coefficient of prior distribution; β = data contribution to the error; input = number of input; output = number of output; weights = number of weights; hidden = number of hidden units; tanh = hyperbolic tangent function.

Network	α	β	Input	Output	Weights	Hidden	Output activation	Input activation
PME	0.01	30	10	3	115	8	logistic	tanh
MP	0.01	30	10	3	129	9	logistic	tanh

Table 6.2. The list of parameters used for implementing the hybrid Monte Carlo method. This table shows, the number of steps omitted to allow the simulation to converge to a stationary distribution; the number of steps in the hybrid Monte Carlo trajectories; the fixed step size that is used to calculate the real step size via the formula $\varepsilon = \varepsilon_0(1 + 0.1k)$ and the number of samples retained to calculate the distribution of the output units. Key: PME = pseudo-modal-energy-network; MP-N=modal-property-network

Network	Number of steps omitted	Number of steps in a trajectory	Fixed step size (ε_0)	Number of samples retained (L)
PME-N	100	100	0.001	500
MP-N	100	100	0.001	500

6.4 RESULTS AND DISCUSSIONS

In this chapter, the pseudo-modal-energy-network and modal-property-network are trained and tested using the data in Table 5.1. The mean squared errors and the classification errors of these two networks on the training data set are shown in Table 6.3. From this table it is observed that the lower MSE gives lower the classification errors.

Table 6.3. The mean squared errors (MSE) and classification errors from the training sets.

Network	MSE	Classification errors (%)
PME	0.0216	4.8
MP	0.0325	5.2

Figure 6.1 shows the mean squared errors on the test data set versus the weighting function assigned to the pseudo-modal-energy-network. This figure shows that the pseudo-modal-energy-network gives lower mean squared errors than the modal-property-network. It also shows that the optimal committee method gives lower mean squared errors than the two individual methods. Furthermore, this figure shows that the optimal committee that gives lower mean squared errors than the individual methods is realised when 60% weighting function is assigned to the pseudo-modal-energy-network. In this chapter since we do not know this optimal weighting function in advance, we assumed that the two networks be given equal weighting functions. The calculated optimal weighting function observed on the test data is higher than the assumed one (with two equal weighting functions), however the assumed committee-weight of 50% still gives lower mean squared errors than the individual methods. The committee factor (CF) defined in equation 3.24 is in equation 3.24 is calculated to be 1.83, which is higher than the CF of 1.72 calculated in Chapter 5. Here it should be noted that the higher the CF the more effective is the committee over the individual methods.

The results of the standard deviation of squared errors versus the weighting function given to the pseudo-modal-energy-network are shown in Figure 6.2. In this figure the pseudo-modal-energy-network is found to give lower standard deviation of squared errors than the modal-property-network while the optimal committee method gives the least standard deviation of squared errors. The optimal committee is achieved when 55% weighting function is assigned to the pseudo-modal-energy-network. The assumed committee-weight of 50% assigned to each method gives lower standard-deviation-of-squared errors than the individual methods. The committee factor for standard deviation of squared errors is 2.52, which is higher than the 2.14 observed in Chapter 5.

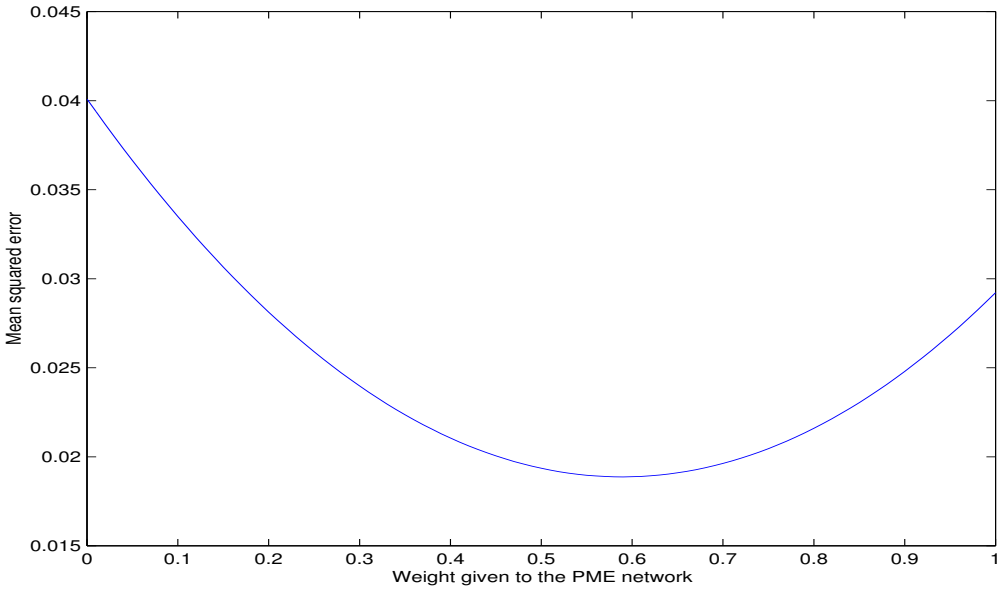


Figure 6.1. Mean squared errors versus weighting function given to the pseudo-modal-energy-network. The mean squared errors are calculated using the following formula: $\mathrm{MSE} = \frac{1}{\mathrm{NM}} \sum_{j=1}^{M} \sum_{i=1}^{N} (y_{ij} - t_{ij})^2$ Here N = 96 representing the total number of fault-cases in the test data; M=3 representing the number of substructures; y_{ij} and t_{ij} are the predicted and correct outputs, respectively, for substructure *i* and fault-case *j*.

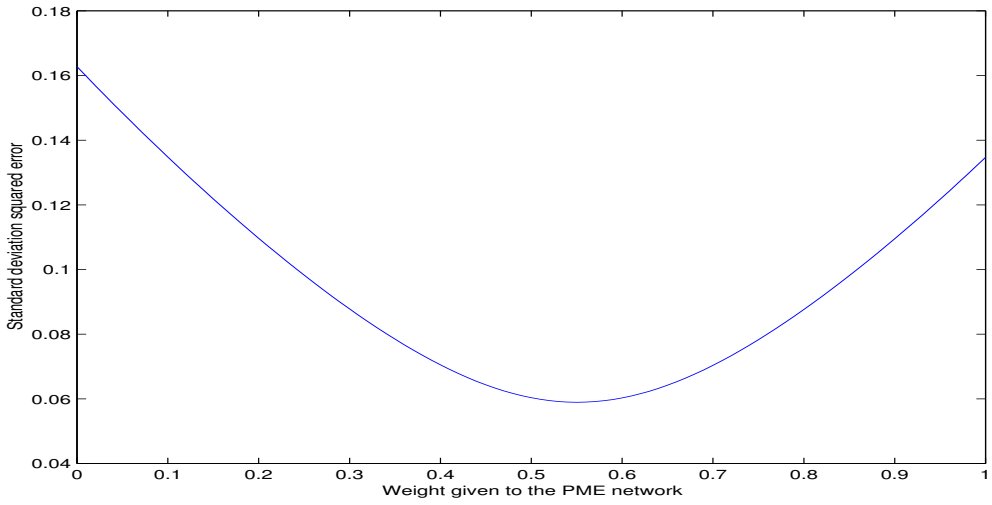


Figure 6.2. Standard deviation of squared errors versus weighting function given to the pseudo-modal-energy-network. The standard deviation was calculated as follows: $\mathrm{STD} = \sqrt{\mathrm{Var}[(y_{ij} - t_{ij})^2]}$. Here y_{ij} and t_{ij} are the predicted and correct outputs, respectively, for substructure *i* and fault-case *j*. Key: STD = standard deviation; Var = variance.

6.5 DETECTION AND CLASSIFICATION OF FAULTS

In this section, the ability of the networks to detect the presence of faults and classify fault cases from the test data set is studied. This section presents the results in the same fashion as it was done in Section 5.6. In this section we obtain the committee method by assigning equal weighting functions to the two individual methods.

When the trained networks are used to detect the presence of fault in the test data set in Table 5.1, the results in Table 6.5 are obtained. This table shows that none of the three procedures give any false negatives, which as described in Section 5.6 is the worse possible outcome of a fault identification procedure. The summary of the data in Table 6.5 is shown in Table 6.6. This table shows the accuracy of the three methods to detect the presence faults in the cylinders. This accuracy is calculated by using the proportion of the total number of predictions of fault-cases that were detected correctly. Table 6.5 shows that the committee method gives the best detection of faults followed by the pseudo-modal-property-network.

Table 6.4. Detection of the presence of faults in the test data set

Network	True Negative	False Positive	True Positive	False Negative
Pseudo modal energy	38	1	57	0
Modal Property	37	2	57	0
Committee	39	0	57	0

Table 6.5. Accuracy of the detection of fault results in Table 6.4

Network	% Accuracy
Pseudo modal energy	99.0
Modal Property	97.9
Committee	100

A step further than just detecting the presence or absence of faults is to classify fault-cases by indicating the parts of the cylinders that contain faults. This however, is not viewed in this thesis as a primary goal of a fault identification method but as an important information that could be

potentially useful in industry. In order to analyse the results a confusion matrix is used. The diagonal entries of a confusion matrix represent the number of fault-cases that are classified correctly, while the off-diagonal entries represent the fault-cases that are classified incorrectly.

The confusion matrix obtained when the pseudo-modal-energy-network is used is illustrated in Table 6.6. In this table 94.9% of [0 0 0]; all the one- and two-fault-cases; and 82.1% of [1 1 1] cases were correctly classified.

The confusion matrix obtained when the modal-property-network is used is shown in Table 6.7. This table shows that this network classified 97.4% [0 0 0] fault-cases correctly; all the one- and two-fault-cases; and 66.7% of [1 1 1] fault-cases were classified correctly.

The results obtained when the committee method constructed by assigning equal weighting functions to the two individual methods are shown in Table 6.8. This table shows that 100% of the [0 0 0]; 100% of the one- and two-fault-cases; and 84.6% of the [1 1 1] fault-cases were correctly classified.

The summary of Tables 6.6-6.8 is shown in Table 6.9 and shows that the committee method classified 8 fault-cases the best followed by the pseudo-modal-energy-network.

In this section several issues were observed:

(1) The committee method classifies faults better than the individual methods.

(2) The pseudo-modal-energy-network classifies fault-cases more accurately than the modal-property-network.

The time taken to train the pseudo-modal-energy-network and the modal-property-network is 5.59 and 5.64 CPU minutes, respectively. All the computations were performed on a Dell DIMENSIONXPS600 Desktop Computer with 261MB RAM. The Bayesian fault identification method trained using the hybrid Monte Carlo method is computationally efficient to be used in industry.

Table 6.6. Confusion matrix from the classification of fault-cases using the pseudo-modal-energy-network

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	37	2	0	0	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	0	0	3	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	0	0	5	1	1	32

Table 6.7. Confusion matrix from the classification of fault-cases using the modal-property-network

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	38	0	0	1	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	0	0	3	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	0	0	5	2	6	26

Table 6.8. Confusion matrix from the classification of fault cases using the committee approach

		Predicted							
		[000]	[100]	[010]	[001]	[110]	[101]	[011]	[111]
Actual	[000]	39	0	0	0	0	0	0	0
	[100]	0	3	0	0	0	0	0	0
	[010]	0	0	3	0	0	0	0	0
	[001]	0	0	0	3	0	0	0	0
	[110]	0	0	0	0	3	0	0	0
	[101]	0	0	0	0	0	3	0	0
	[011]	0	0	0	0	0	0	3	0
	[111]	0	0	0	0	5	1	0	33

Table 6.9. Percentage of fault-cases classified correctly

Network	%Classified correctly
Pseudo modal energy	90.6
Modal Property	85.4
Committee	93.8

6.6 MAXIMUM-LIKELIHOOD VERSUS BAYESIAN APPROACH

In this section, a comparison is made between the results obtained from training the network using the scaled conjugate gradient method (maximum-likelihood approach) in Chapter 5, and those obtained using the hybrid Monte Carlo method (Bayesian approach) in this chapter.

In this chapter the samples retained when the hybrid Monte Carlo simulation is implemented is 500. When the trained neural networks are presented with the input data from the test data set [Tables 5.1], for each fault-case 500 fault identities are obtained. The distributions of the output of the networks may be used to assess the integrity of the fault identification results. Figure 6.3 shows the distribution of the predicted fault-case that was correctly diagnosed by the pseudo-modal-energy-network. The top graph is the identification of substructure 1, the middle graph the identification of substructure 2 and the bottom graph is the identification of substructure 3.

This figure shows the output distributions for all three substructures that are conclusive, meaning that it is possible to infer the identity of this fault-case and construct error bars for this fault-identity vector. When these distributions are averaged it is observed that pseudo-modal-energy-network gives a fault identity vector of [0.82 1.00 0.80] which when rounded-off becomes [1 1 1] which is the correct fault identity. The maximum likelihood method of Chapter 5 predicted the correct fault identity of [0.59 0.66 0.94] which when rounded-off becomes a [1 1 1] case.

The distribution of the prediction of the modal-property-network for a [1 1 1] case is shown in Table 6.4. This figure shows that the network failed to predict this fault-case conclusively because the distribution of the output corresponding to substructure 1 spans on both sides of the 0.5-mark, which is a point of separation between the presence and the absence of faults. The average fault identity vector for these distributions is [0.31 1.00 1.00] which when rounded-off becomes [0 1 1] which is a incorrect fault-identity vector. It should be noted however that it is able to correctly identify the presence of faults in the cylinder but only failed to detect the presence of faults in substructure 1. The maximum likelihood method (see Chapter 5) gives the correct diagnosis but does not give a user an opportunity to view the distributions of the network output. For all fault-cases that were diagnosed incorrectly in this chapter, it was observed that inconclusive distributions were obtained. Therefore the implementation of the Bayesian framework offers a user an opportunity to know in advance the level of confidence the network has on the fault identity it gives. This option is not available when the maximum-likelihood as implemented in Chapter 5 is used.

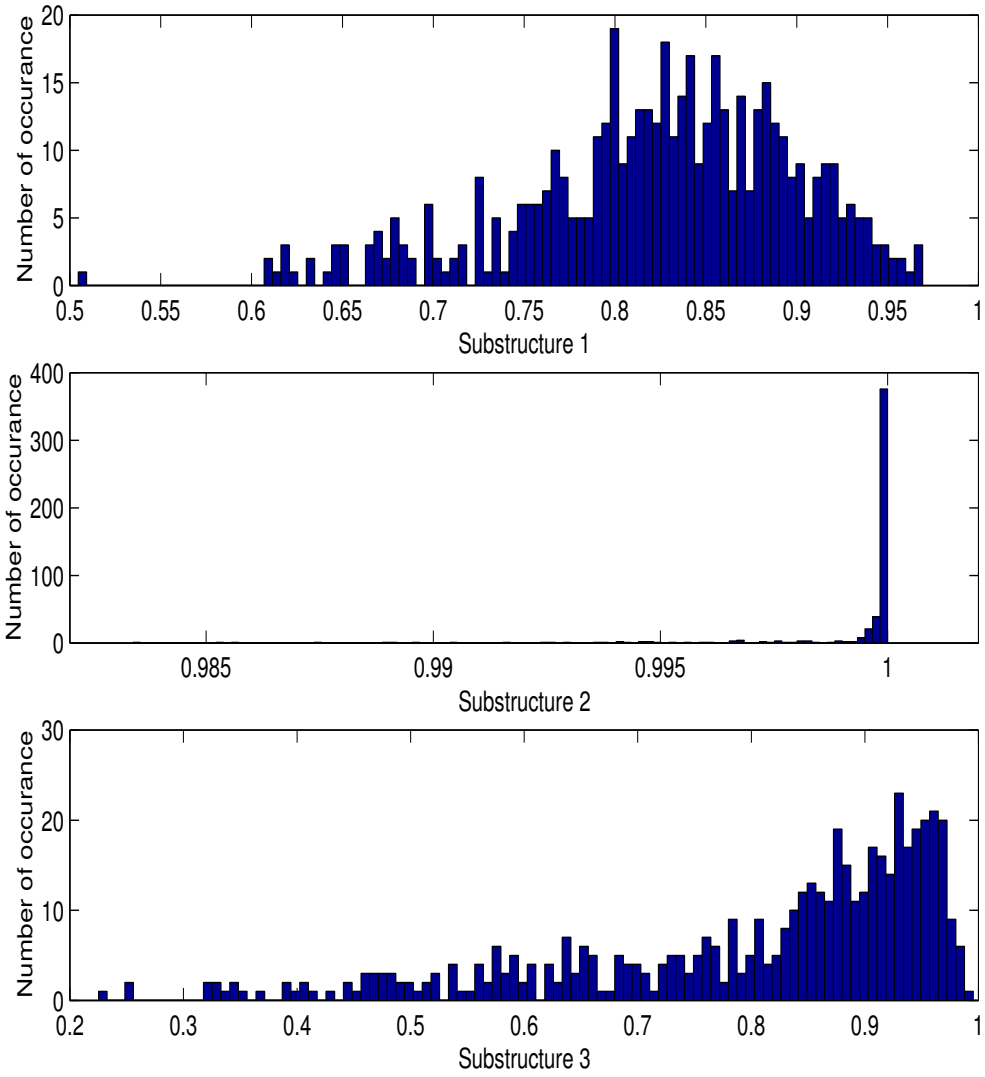


Figure 6.3. The graphs representing the distribution of the network output using the pseudo-modal-energy-network. The top graph is for substructure 1, the middle graph for substructure 2 and the bottom graph for substructure 3. The true identity of this fault-case is [1 1 1]. The average fault-identity vector of the distribution given in this figure is [0.32 1.00 1.00]. The maximum-likelihood method in Chapter 5 predicted this case to be [0.59 0.73 0.94] without giving any confidence interval.

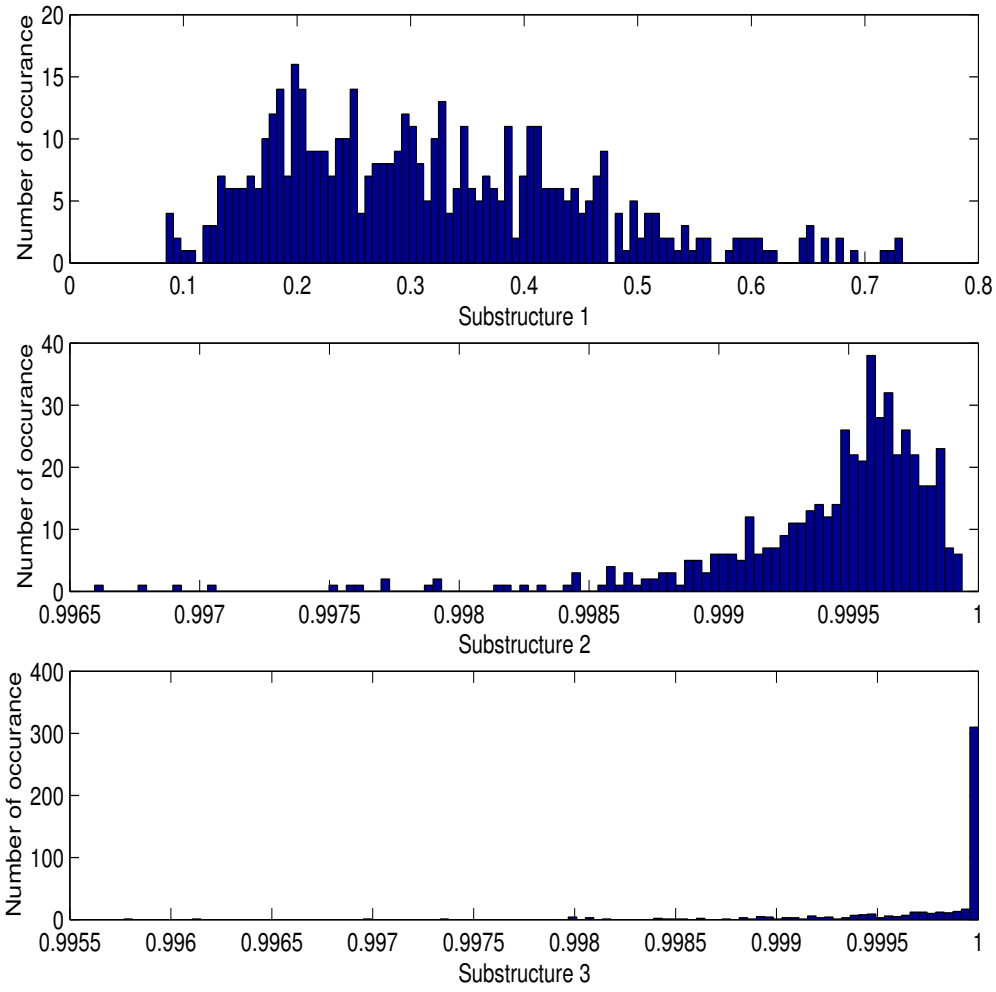


Figure 6.4. The graphs representing the distributions of the output of the network trained using the modal properties. The top graph is for substructure 1, the middle graph for substructure 2 and the bottom graph for substructure 3. The true identity of this case is [1 1 1] and the average of these distributions of the output gives [0.9721 0.6067 0.0966]. The maximum likelihood approach gives a fault identity of [0.59 0.58 1.00]

Table 6.10 compares the mean squared errors for the networks trained using the maximum likelihood method and Bayesian method. This figure shows that the mean squared errors of the training and testing sets are the same order of magnitude. This indicates that the training process did not over-fit the training data set. However, it should be noted that the Bayesian network manages to achieve this without the use of validation methods whereas a cross-validation method was implemented to ensure that the maximum likelihood training does not over-fit the data.

Table 6.10. Mean squared errors of the maximum-likelihood and Bayesian methods using fault-cases listed in Sections 5.2.1-5.2.3. Key: PME≡Pseudo-modal-energy; MP≡Modal-property

	Maximum Likelihood		Bayesian	
Network	PME	MP	PME	MP
Training	0.0240	0.0307	0.0216	0.0325
Testing	0.0490	0.0611	0.0292	0.0401

The results of comparison of the ability of the maximum-likelihood method and Bayesian method to detect and classify faults are shown in Table 6.11. From these table it is observed that, on average, the networks trained using the Bayesian methods are found to be better than those trained using the maximum likelihood method.

Table 6.11. Classification of the presence or absence of faults in the test-data listed in Sections 5.2.1-5.2.3 using the maximum-likelihood and Bayesian methods. Key: PME-N≡Pseudo-modal-energy-network; MP-N≡Modal-property-network; C-N≡Committee method

	Maximum-likelihood approach			Bayesian approach		
	PME-N (%)	MP-N (%)	C-N (%)	PME-N (%)	MP-N (%)	C-N (%)
Detection	95.8	94.8	99.0	99.0	97.9	100
Classification	82.2	81.2	95.8	90.6	85.4	93.8

6.7 CONCLUSIONS

A committee of two Bayesian formulated neural networks trained using pseudo modal energies and modal properties is successfully used to perform probabilistic fault identification in a population of cylindrical shells. The Bayesian networks are identified using the hybrid Monte Carlo method. The results indicate that the committee method gives lower mean- and standard-deviation-of-square-of-errors than the methods used individually. On average, it is found that the pseudo modal energies detect and classify faults more reliably than the modal properties. Bayesian formulation is observed to offer the advantage of providing the distributions of the

network outputs, which may be used to assess the degree of confidence the network has, on the solution it gives. The maximum-likelihood method as implemented in Chapter 5 is unable to offer this option. The Bayesian method is found to give more accurate results than the maximum likelihood method.

Chapter 7

CONCLUSIONS AND FURTHER WORK

This chapter summarises the results that have been presented in this thesis and, based on these results, further work is recommended.

7.1 CONCLUSIONS

Pseudo modal energies, defined as the integrals of the frequency response functions (FRFs) over given frequency ranges, were proposed for fault identification in structures. The pseudo modal energies were compared to the modal properties by using statistical overlap factors. Mathematical expressions defining the pseudo modal energies in terms of the modal properties were derived. Subsequently, their sensitivities were derived in terms of the sensitivities of the modal properties.

The pseudo modal energies are found to be more sensitive to faults than the natural frequencies and equally as sensitive to faults as the mode shapes. In addition, the pseudo modal energies are more resistant to noise than the mode shapes. The natural frequencies are found to be more resistant to noise than the mode shapes and pseudo modal energies. Furthermore, the pseudo modal energies are found to be faster to compute than the modal properties.

A committee-of-networks, that uses both the pseudo modal energies and the modal properties, was introduced for structural diagnostics. The optimal weights that give the best committee were derived in terms of the expected errors and variances of the individual methods. When tested on a simulated cantilevered beam, the committee method formulated using the

maximum-likelihood approach and trained using the scaled conjugate gradient method gives more accurate classification of faults than the methods acting individually. It is also observed that, on average, the pseudo-modal-energy-network gives more accurate classification of faults than the modal-property-network.

Vibration data from a population of 20 cylinders were measured and the modal properties as well as the pseudo modal energies were calculated. The modal properties and pseudo modal energies were transformed into the coordinate-modal-assurance-criterion and the coordinate-pseudo-modal-energy-assurance-criterion, respectively. The pseudo modal energies and modal properties are found to be better indicators of faults than the coordinate-modal-assurance-criterion and the coordinate-pseudo-modal-energy-assurance-criterion, respectively.

A committee-of-networks was used classify faults in a population of cylinders. Firstly, the networks were formulated using the maximum-likelihood framework and trained using the scaled conjugate gradient method. The committee method is found to give more accurate classification of faults than the individual methods. The pseudo-modal-energy-network and the modal-property-network are observed to give similar levels of accuracy on classifying faults.

The committee approach was extended to a probabilistic framework using the Bayesian approach and the networks were trained using the hybrid Monte Carlo method. This extension offers the possibility of obtaining the confidence intervals of the fault identities given by neural networks and was successfully implemented for fault identification in a population of cylinders. It is observed that the committee method gives more accurate classification of faults than the individual networks. Furthermore, it is observed that the pseudo-modal-energy-network classifies faults more accurately than the modal-property-network.

When the Bayesian formulation is compared to the maximum likelihood method, it is observed that the Bayesian approach, on average, gives more accurate classification of faults than the maximum-likelihood method. Furthermore, the Bayesian approach gives distributions of network outputs, which may be used to deduce confidence levels of the fault identification results. The maximum likelihood method does not offer this option. The Bayesian approach is found to give good results without the need of applying validation methods, which are usually required when a maximum likelihood method is used.

7.2 RECOMMENDATIONS FOR FURTHER WORK

One of the areas closely related to fault identification is the area of finite element model updating. If an accurate finite element model of a structure is identified then it could be used as a basis for fault identification in that structure. To date, the problem of identifying a set of finite element model parameters that best match model predictions to measured data has been solved using the maximum likelihood approach. As shown in the present thesis, the Bayesian approach has certain advantages that the maximum likelihood method does not have, such as the ease with which confidence intervals can be determined. When the maximum-likelihood approach is implemented there is a need to perform cross-validation while this is not the case when the Bayesian approach is used. As an extension to the present work, the finite element model updating problem may be formulated in a Bayesian framework. The hybrid Monte Carlo method, which has proven to be successful in damage identification problems, may be used to solve the Bayesian formulated finite-element-model-updating problem.

In this thesis the vibration data were mapped to the identity of faults using neural networks. The identity of a fault was a vector with binary components. The neural networks implemented in this thesis had continuous activation functions. As an extension to the present work, neural networks with binary activation functions could be implemented and their performance compared to those obtained when continuous activation functions are used.

The committee-of-networks implemented here used two types of data as inputs to the neural networks. A choice was made to restrict this study to the types of faults that do not change the linearity of structures. Hence, the data decomposed into the time-frequency domain was not examined in this thesis. To ensure that the committee-of-networks is able to deal with the types of faults that change the linearity of a structure, the data decomposed into the time-frequency domain should be incorporated into the committee in future.

Neural networks implemented in this study used a supervised learning technique, namely, the networks had to be given the input (the pseudo modal energies and modal properties) and the corresponding output (identities of faults) data. However, in real life situations, the output data are not necessarily available. As an extension to the present work, unsupervised neural networks

may be used for fault identification and a comparison made between using the pseudo modal energies and modal properties for fault identification.

REFERENCES

1. Adhikari, S., (1999) "Rates of change of eigenvalues and eigenvectors in damped system" American Institute of Aeronautics and Astronautics Journal, Vol. 37, 1452-1458.
2. Alder, B.J. and Wainwright, T.E., (1959) "Studies in molecular dynamics. I. General method" Journal of Chemical Physics, Vol. 31, 459-466.
3. Allemang, R.J. and Brown, D.L., (1982) "A correlation coefficient for modal vector analysis" Proceedings of the 1st International Modal Analysis Conference, 1-18.
4. Alvin, K.F., (1996) "Finite element model updating via Bayesian estimation and minimisation of dynamic residuals" Proceedings of the 14th International Modal Analysis Conference, 428-431.
5. Anderson, H.C., (1980) "Molecular dynamics simulations at constant pressure and/or temperature" Journal of Chemical Physics, Vol. 72, 2384-2393.
6. Anonymous, (1994) *ABAQUS Users Manual*, Version 5.4, Hibbitt, Karlsson and Sorensen, Inc, Pawtucket, Rhode Island, USA.
7. Ash, R.E., (1965) *Information theory*, New York, Wiley, USA.
8. Atalla, M.J., (1996) *Model updating using neural networks*, Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.
9. Atalla, M.J. and Inman, D.J., (1998) "On model updating using neural networks" Mechanical Systems and Signal Processing, Vol. 12, No. 1, 135-161.
10. Balmès, E., (1997[a]) *Structural Dynamics Toolbox User's Manual* Version 2.1, Scientific Software Group, Sèvres, France.

11. Balmès, E., (1997[b]) “New results on the identification of normal modes from experimental complex modes” Mechanical Systems and Signal Processing, Vol. 11, 135-161.
12. Barschdorf, D. and Femmer, U., (1995) "Signal processing and pattern recognition methods for biomedical sound analysis", Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques, 2nd International Symposium, CETIM, , 279-290.
13. Baruch, M., (1978) “Optimisation procedure to correct stiffness and flexibility matrices using vibration data” American Institute of Aeronautics and Astronautics Journal, Vol. 16, 1208-1210.
14. Baruch, M. and Bar Itzhack, I.Y., (1978) “Optimum Weighted Orthogonalization of Measured Modes” American Institute of Aeronautics and Astronautics Journal, Vol. 20, 1623-1626.
15. Bellman, R., (1961) *Adaptive control processes: A guided tour* Princeton, Princeton University Press, New Jesrsey, USA.
16. Ben-Haim, Y. and Prells, U., (1993) “Selective sensitivity in the frequency doman, part I: theory” Mechanical Systems and Signal Processing, Vol. 7, 461-475.
17. Berman, A. and Nagy, E.J., (1983) “Improvement of large analytical model using test data” American Institute of Aeronautics and Astronautics Journal, Vol. 21, 1168-1173.
18. Bertsekas, D.P., (1995) *Nonlinear programming,* Athenas Scientific, Belmont, Massachusetts, USA.
19. Bishop, C.M., (1995) *Neural Networks for Pattern Recognition.* Oxford University Press, Oxford, UK.
20. Biswas, M.A., Pandey, A.K. and Samman, M.M., (1990) “Diagnostic experimental spectral/modal analysis of a highway bridge” Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis, Vol. 5, 33-42.

21. Breiman, L., (1996) “Bagging predictors” Machine Learning, Vol. 24, 123-140.
22. Cawley, P. and Adams, R.D., (1979) “The location of defects from measurements of natural frequencies” Journal of Strain Analysis, Vol. 14, 49-57
23. Chen, J.C. and Garba, J.A., (1980) “Analytical model improvement using modal test results” American Institute of Aeronautics and Astronautics Journal, Vol. 18, 684-690.
24. Coetzee, F.M., and Stonick, V.L., (1996) “On the uniqueness of weights in a single-layer perceptrons’ IEEE Transactions on Neural Networks, Vol. 7, No. 2, 318-325.
25. Cohen, L., (1989) "Time-frequency distributions - a review", Proceedings of the IEEE, Vol.77, No. 7, 941-981.
26. Cooley, J.W. and Tukey, J.W., (1965) “An algorithm for the machine calculation of complex Fourier series” Mathematics of Computation, Vol. 19, 297-301.
27. D’Ambrogio, W. and Zobel, P.B., (1994) “Damage detection in truss structures using a direct updating technique” Proceedings of the 19th International Seminar for Modal Analysis, 657-667.
28. Daubechies, I., (1987) “Orthogonal bases of wavelets with finite support connection with discrete filters” Proceedings of the International Conference on Wavelets, Marseille, published as pp. 38-66 of “Wavelets, Time-Frequency Methods and Phase Space” J.M. Combes, A. Grossmann and Ph. Tchamitchian, eds., Springer-Verlag, Berlin, 1989.
29. Doebling, S.W., Farrar, C.R., Prime, M.B., and Shevitz, D.W., (1996) *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review* Los Alamos National Laboratory Report LA-13070-MS.
30. Duane, S., Kennedy, A.D., Pendleton, B.J., and Roweth, D., (1987) “Hybrid Monte Carlo”, Physics Letters. Vol. 195, 216-222.

31. Dunn, S.A. (1998) “The use of genetic algorithms and stochastic hill-climbing in dynamic finite-element model identification” Computers and Structures, Vol. 66, 489-497.
32. Ewins, D.J., (1995) *Modal Testing: Theory and Practice.* Research Studies Press, Letchworth, U.K.
33. Farrar, C.R., Baker, W.E., Bell, T.M., Cone, K.M., Darling, T.W., Duffey, T.A., Eklund, A., and Migliori, A., (1994) *Dynamic characteritics and damage detection in the I-40 bridge over the Rio Grande* Los Alamos National Laboratory report LA-12767-MS.
34. Feller, W., (1968) *An introduction to probability theory and its applications,* Third Edition, New York, John Wiley, USA.
35. Fill, J.A., (1991) “Eigenvalue bounds on convergence to stationarity for non-reversible Markov chains, with an application to the exclusion process” Annals of Applied Probability, Vol. 1, 62-87.
36. Fitzgerald, W.J., (1996) *Numerical Bayesian methods applied to signal processing.* Springer-Verlag, Berlin, Germany.
37. Fletcher, R., (1987) *Practical methods of optimization.* 2nd edition, Wiley, New York, USA.
38. Fox, C.H.J., (1992) “The location of defects in structures: a comparison of the use of natural frequency and mode shape data” Proceedings of the 10th International Modal Analysis Conference, 522-528.
39. Fox, R.L. and Kapoor, M.P., (1968) “Rates of change of eigenvalues and eigenvectors” American Institute of Aeronautics and Astronautics Journal, Vol. 6, 2426-2429.
40. Freund, Y., (1995) “Boosting a weal learning algorithm by majority” Information Computation, Vol. 121, 256-285.

41. Fruend, Y. and Schapire, R., (1996) "Experiments with a new boosting algorithm" Proceedings of the Thirteenth International Conference on Machine Learning, Morgan Kaufmann, 149-156.
42. Friswell, M.I. and Mottershead, J.E., (1995) *Finite Element Model Updating in Structural Dynamics.* Kluwer Academic Publishers Group, Dordrecht, The Netherlands.
43. Friswell, M.I., Penny, J.E.T., and Wilson, D.A.L., (1994) "Using vibration data and statistical measures to locate damage in structures" Modal Analysis: The International Journal of Analytical and Experimental Modal Analysis, Vol. 9, 239-254.
44. Friswell, M.I., Penny, J.E.T., and Garvey, S.D., (1996) "A combined genetic and eigensensitivity algorithm for location of damage in structures" Proceedings of the 1st International Conference on Identification in Engineering Systems, 41-52.
45. Gawronski, W. and Sawicki, J.T., (2000) "Structural damage detection using modal norms" Journal of Sound and Vibration, Vol. 229, 194-198.
46. Geisser, S., (1975) "The predictive sample reuse method with applications" Journal of the American Statistical Association, Vol. 70, 320-328.
47. Gelfand, A.E. and Smith, A.F.M., (1990) "Sampling-based approaches to calculating marginal densities" Journal of American Statistical Association, Vol. 85, 398-409.
48. Geman, S. and Geman, D., (1984) "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images" IEEE Transactions on Pattern Analysis and Machine Inteligence, Vol. 6, 721-741.
49. Geman, S., Bienenstock, E., and Doursat, R., (1992) "Neural networks and the bias/variance dilema" Neural Computation, Vol. 4, 1-58.
50. Gopinath, O.C., (1999) "Diagnostic classifier ensembles: enforcing diversity for reliability in combination" Ph.D. Dissertation, University of Sheffield, Sheffield, UK.

51. Goutte, C., (1997) "Note on free lunches and cross-validation," Neural Computation, Vol. 9, 1211-1215.

52. Gradshteyn, I.S. and Yyzhik, I.M., (1994), *Tables of integrals, series, and products.* Academic Press, London, U.K.

53. Grygier, M.S., (1994) *Modal test technology as non-destructive evaluation of space shuttle structures*, NASA Conference Publication 3263.

54. Guyan, R.J., (1965) "Reduction of stiffness and mass matrices" American Institute of Aeronautics and Astronautics Journal, Vol. 3, 380.

55. Gysin, H., (1990) "Comparison of expansion methods for FE model localization" Proceeding of the 8th International Modal Analysis Conference, 195-204.

56. Hashem, S., (1997) "Optimal linear combinations of neural networks" Neural Networks, Vol. 10, 599-614.

57. Haug, E.F. and Choi, K.K., (1984) "Structural design sensitivity with generalized global stiffness and mass matrices" American Institute of Aeronautics and Astronautics Journal, Vol. 22, 1299-1303.

58. Haykin, S., (1999) *Neural networks*, 2nd edition, Prentice-Hall, Inc, New Jersey, USA.

59. Hemez, F.M., (1993) *Theoretical and experimental correlation between finite element models and modal tests in the context of large flexible structures* Ph.D. Dissertation, University of Colorado, Boulder, Colorado, USA.

60. Hansen, L.K. and Salamon, P., (1990) "Neural network ensembles" IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 12, No. 10, 993-1000.

61. Hestenes, M.R. and Stiefel, E., (1952) "Methods of conjugate gradients for solving linear systems" Journal of Research of the National Bureau of Standards, Vol. 6, 409-436.

62. Hinton, G.E., (1987) "Learning translation invariant recognition in massively parallel networks" In J.W. de Bakker, A.J. Nijman, and P.C. Treleaven (Eds.), Proceedings PARLE Conference on Parallel Architectures and Languages Europe, pp. 1-13. Berlin: Springer-Verlag.

63. Hinton, G.E., (1989) "Connectionist learning procedures" Artificial Intelligence, Vol. 40, 185-234.

64. Hopfield, J.J., (1987) "Learning algorithms and probability distributions in feed-forward and feed-back networks" Proceeding of the National Academy of Science, Vol. 84, 8429-8433.

65. Hotelling, H., (1933) "Analysis of a complex of statistical variables into principal components2" Journal of Educational Psychology, Vol. 24, 417-441.

66. Imregun, M. and Ewins, D.J., (1993) "An investigation into mode shape expansion techniques" Proceeding of the 11th International Modal Analysis Conference, 168-175.

67. Imregun, M., Visser, W.J., and Ewins, D.J., (1995) "Finite element model updating using frequency response function data-I. Theory and initial investigation" Mechanical Systems and Signal Processing, Vol. 9, 187-202.

68. Jacobs, R.A., (1990) "Tasks decomposition through computation in a modular connectionist architecture" Ph.D. Thesis, University of Massachusetts, Massachusetts, USA.

69. Jacobs, R.A., (1995) "Methods of combining experts' probability assessments" Neural Computation, Vol. 7, 867-888.

70. Jacobs, R.A., Jordan, M.I., and Hinton, G.E., (1991) "Adaptive mixtures of local experts" Neural Computation, Vol. 3, 79-87.

71. Janter, T. and Sas, P., (1990) "Uniqueness aspects of model-updating procedure" American Institute of Aeronautics and Astronautics Journal, Vol. 28, 538-543.

72. Jollife, I.T., (1986) *Principal component analysis* New York: Springer-Verlag, U.S.A.

73. Jordan, M.I. and Bishop, C.M., (1996) *Neural networks* Massachusetts Institute of Technology Technical Report A.I. Memo No. 1562.

74. Ju, F. and Mimovitch, M., (1986) "Modal frequency method in diagnosis of fracture damage in structures" Proceedings of the 4th International Modal Analysis Conference, 1423-1429.

75. Kabe, A.M., (1985) "Stiffness matrix adjustment using mode data" American Institute of Aeronautics and Astronautics Journal, Vol. 23, 1431-1436.

76. Kalos, M.H. and Whitlock, P.A., (1986) *Monte Carlo methods* Vol. 1: Basics, Wiley, New York.

77. Kashangaki, T.A.L, (1991) *On-orbit damage detection and health monitoring of large space trusses-status and critical issues*, NASA report TM-104045.

78. Kemeny, J.G. and Snell, J.L., (1960) *Finite Markov chains*, (reprinted, 1976), New York: Springer-Verlag, U.S.A.

79. Kim, J.H., Jeon, H.S., and Lee, S.W., (1992) "Application of modal assurance criteria for detecting and locating structural faults" Proceedings of the 10th International Modal Analysis Conference, 536-540.

80. Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., (1983) "Optimization by simulated annealing," Science, Vol. 220, 671-680.

81. Kondo, I. and Hamamoto, T., (1994) "Local damage detection of flexible offshore platforms using ambient vibration measurements" Proceedings of the 4th International Offshore and Polar Engineering, Vol. 4, 400-407.

82. Kudva, J., Munir, N., and Tan P., (1991) "Damage detection in smart structures using neural networks and finite element analysis" Proceedings of ADPA/AIAA/ASME/SPIE Conference on Active Materials and Adaptive Structures, 559-562.

83. Lai, J.Y. and Young, K.F., (1995) "Dynamics of graphite/epoxy composite under delamination fracture and environmental effects" Journal of Composite Structures, Vol. 30, 25-32.

84. Lam, H.F., Ko, J.M., and Wong, C.W., (1995) "Detection of damage location based on sensitivity analysis" Proceeding of the 13th International Modal Analysis Conference, 1499-1505.

85. Langley, R.S., (1999) "A non-Poisson model for vibration analysis of uncertain dynamic systems" Proceedings of the Royal Society of London Series A-Mathematical, Physical and Engineering Sciences, Vol. 455, 3325-3349.

86. Larson, C.B. and Zimmerman, D.C., (1993) "Structural model refinement using a genetic algorithm approach" Proceeding of the 11th International Modal Analysis Conference, 1095-1101.

87. Leath, W.J. and Zimmerman, D.C., (1993) "Analysis of neural network supervised training with application to structural damage detection" Damage and Control of Large Structures, Proceeding of the 9th Virginia Polytechnic Institute and State University Symposium, 583-594.

88. Lee, B.T., Sun, C.T., and Liu, D., (1987) "An assessment of damping measurement in the evaluation of integrity of composite beams" Journal of Reinforced Plastics and Composites, Vol. 6, 114-125.

89. Levin, R.I. and Lieven, N.A.J., (1998[a]) "Dynamic finite element updating using simulated annealing and genetic algorithms" Mechanical Systems and Signal Processing, Vol. 12, 91-120.

90. Levin, R.I., and Lieven, N.A.J., (1998[b]) "Dynamic finite element updating using neural networks" Journal of Sound and Vibration, Vol. 210, 593-608.

91. Lieven, N.A.J. and Ewins, D.J., (1988) "Spatial correlation of mode shapes, the co-ordinate modal assurance criterion" Proceedings of the 6th International Modal Analysis Conference, 690-695.

92. Lifshitz, J.M. and Rotem, A., (1969) "Determination of reinforcement unbonding of composites by a vibration technique" Journal of Composite Materials, Vol. 3, 412-423.

93. Lin, R.M., Lim, M.K., Du, H., (1995) "Improved inverse eigensensitivity method for structural analytical model updating" Transaction of the American Society of Mechanical Engineering, Journal of Vibration and Acoustics, Vol. 117, 192-198.

94. Lippmann, R.P., (1987) "An introduction to computing with neural networks" IEEE ASSP Magazine, April issue, 4-12.

95. Loéve, M., (1963) *Probability theory,* 3rd Edition, Van Nostrand, New York, USA.

96. Loland, O. and Dodds, J.C., (1976) "Experience in developing and opening integrity monitoring system in north sea" Proceeding of the 8th Annual Offshore Technology Conference, 313-319.

97. Lyon, R., (1995) "Structural diagnostics using vibration transfer functions" Sound and Vibration, Vol. 29, 28-31.

98. Luenberger, D.G., (1984) *Linear and nonlinear programming.* 2nd edition, Addison-Wesley, Reading, Massachusetts, USA.

99. MacKay, D.J.C., (1991) *Bayesian methods for adaptive models*. Ph.D. Dissertation, California Institute of Technology, CA, USA.

100. MacKay, D.J.C., (1992) "A practical Bayesian framework for backpropagation networks" Neural Computation, Vol. 4, No. 3, 448-472.

101. Mackenzie, P.B., (1989) "An improved hybrid Monte Carlo method" Physics Letters B, Vol. 226, 369-371.

102. Maia, N.M.M. and Silva, J.M.M., (1997) *Theoretical and experimental modal analysis* Research Studies Press, Letchworth, U.K.

103. Maia, N.M.M., Silva, J.M.M., and Sampaio, R.P.C., (1997) "Localization of damage using curvature of the frequency-response-functions" Proceedings of the 15th International Modal Analysis Conference, 942-946.

104. Maia, N.M.M., Silva, J.M.M., and Sampaio, R.P.C., (1999) "On the use of frequency-response-functions for damage detection" Proceedings of the 2nd International Conference on Identification in Engineering Systems, 460-471.

105. Manning, R., (1994) "Damage detection in adaptive structures using neural networks" Proceedings of the 35th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 160-172.

106. Mares, C. and Surace, C., (1996) "An application of genetic algorithms to identify damage in elastic structures" Journal of Sound and Vibration, Vol. 195, 195-215.

107. Marwala, T., and Heyns, P.S., (1998) "Multiple-criterion method for determining structural damage" American Institute of Aeronautics and Astronautics Journal, Vol. 36, 1494-1501.

108. Mathworks, (1992) *MATLAB Reference Guide.* The Mathworks, Natick, Massachusetts, USA.

109. Mayes, R.L., (1992) "Error localization using mode shapes-an application to a two link robot arm" Proceeding of the 10th International Modal Analysis Conference, 886-891.

110. Mazurek, D.F. and DeWolf, J.T., (1990) "Experimental study of bridge monitoring technique" American Society of Civil Engineering, Journal of Engineering Mechanics, Vol. 116, 2532-2549.

111. McColl, J.H., (1995) *Probability* Edward Arnold, London, UK.

112. McGowan, P.E., Smith, S.W., and Javeed, M., (1990) "Experiments for locating damage members in a truss structure" Proceedings of the 2nd USAF/NASA Workshop on System Identification and Health Monitoring of Precession Space Structures, 571-615.

113. Messina, A., Contursi, T., and Williams, E.J., (1997) "Multiple damage evaluation using natural frequency changes" Proceedings of the 15th International Modal Analysis Conference, 658-664.

114. Messina, A., Jones, I.A., and Williams, E.J., (1996) "Damage detection and localisation using natural frequency changes" Proceedings of the 1st International Conference on Identification in Engineering Systems, 67-76.

115. Messina, A., Williams, E.J., and Contursi, T., (1998) "Structural damage detection by a sensitivity and statistical-based method" Journal of Sound and Vibration, Vol. 216, 791-808.

116. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E., (1953) "Equations of state calculations by fast computing machines," Journal of Chemical Physics, Vol. 21, 1087-1092.

117. Migliori, A., Bell, T.M., Dixon, R.D., and Strong, R., (1983) *Resonant ultrasound non-destructive inspection* Los Alamos National Laboratory report LS-UR-93-225.

118. Møller, M., (1993) "A scaled conjugate gradient algorithm for fast supervised learning" Neural Networks, Vol. 6, 525-533.

119. Mottershead, J.E. and Friswell, M.I., (1993) "Model updating in structural dynamics: a survey" Journal of Sound and Vibration, Vol. 167, 347-375.

120. Neal, R.M., (1992) "Bayesian training of backpropagation networks by hybrid Monte Carlo method," University of Toronto Technical Report CRG-TR-92-1.

121. Neal, R.M., (1993) "Probabilistic inference using Markov chain Monte Carlo methods," University of Toronto Technical Report CRG-TR-93-1.

122. Neal, R.M., (1994) *Bayesian learning for neural networks* Ph.D. thesis, University of Toronto, Canada.

123. Newland, D.E., (1993) *An introduction to random vibration, spectral and wavelet analysis* Longman, Harlow and John Wiley New York, USA.

124. Newland, D.E., (1994) "Wavelet analysis of vibration: part I: theory" Journal of Vibration and Acoustics, Vol. 116, 409-416.

125. Nilsson, N.J., (1965) *Learning machines: foundations of trainable pattern-classifying systems* McGraw-Hill, New York, USA.

126. Norris, M.A. and Meirovitch, L., (1989) "On the problem of modelling for parameter identification in distributed structures" International Journal for Numerical Methods in Engineering, Vol. 28, 2451-2463.

127. Nowlan, S.J., 1990 "Maximum likelihood competitive learning" Advances in Neural Information Processing Systems, Vol. 2, 574-582, San Mateo, CA: Morgan Kaufmann.

128. O'Callahan, J.C., (1989) "A procedure for improved reduced system (IRS) model" Proceeding of the 7th International Modal Analysis Conference, 17-21.

129. O'Callahan, J.C., Avitabile, P., and Riemer, R., (1989) "System equivalent reduction expansion process' Proceeding of the 7th International Modal Analysis Conference, 17-21.

130. Ojalvo, I.U. and Pilon, D., (1988) "Diagnosis for geometrically locating structural mathematical model errors from modal test data" Proceeding of the 29th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 1174-1186.

131. Paz, M., (1984) "Dynamic condensation" American Institute of Aeronautics and Astronautics Journal, Vol. 22, 724-727.

132. Partridge, M.G., and Calvo, R.A., (1998) "Fast dimensionality reduction and simple PCA" Intelligent Data Analysis, Vol. 2, No.3, 724-727.

133. Pearson, K., (1901) "On lines and planes of closest fit to systems of points in space" Philosophical Magazine, Vol. 2, 559-572.

134. Perrone, M.P., (1993) "General averaging results for convex optimization." Eds. M.C. Mozer *et al*., Lawrence Erlbaum, Proceedings of Connectionist Models Summer School, Hillsdale, New Jersey, USA, 364-371.

135. Perrone, M.P. and Cooper, L.N., (1993) *When networks disagree: ensemble methods for hybrid neural networks* Ed. R.J. Mammone, Artificial Neural Networks for Speech and Vision, Chapman and Hall, London, UK, 126-142.

136. Polak, E. and Ribiére, G., (1969) "Note sur la convergence de methods de directions conjuguees" Revue Francaise Information Recherche Operationnelle, Vol. 16, 35-43.

137. Prells, U. and Ben-Haim, Y., (1993) "Selective sensitivity in the frequency doman, part II: application" Mechanical Systems and Signal Processing, Vol. 7, 551-574.

138. Prime, M.B and Shevitz D.W., (1996) "Damage detection of street frame by modal testing" Proceedings of the 11th International Modal Analysis Conference, 1437-1443

139. Rizos, P.F., Aspragathos, N., and Dimarogonas, A.D., (1990) "Identification of crack location and magnitude in a cantilever beam from the vibration modes" Journal of Sound and Vibration, Vol. 138, 381-388.

140. Robbins, H. and Monro, S., (1951) "A stochastic approximation method" Annals of Mathematical Statistics, Vol. 22, 400-407.

141. Royton, T.J., Spohnholtz, T., and Ellington, W.A., (2000) "Use of non-degeneracy in nominally axisymmetric structures for fault detection with application to cylindrical geometries" Journal of Sound and Vibration, Vol. 230, 791-808.

142. Salawu, O.S. and Williams, C., (1993) "Damage location using vibration mode shapes" Proceedings of the 11th International Modal Analysis Conference, 933-939.

143. Salawu, O.S., (1995) "Non-destructive Assessment of Structures Using Integrity Index Method Applied to a concrete highway bridge" Insight, Vol. 37, 875-878.

144. Saitoh, M. and Takei, B.T., (1996) "Damage estimation and identification of structural faults using modal parameters" Proceedings of the 14th International Modal Analysis Conference, 1159-1164.

145. Schaphire, R.E., (1997) "Using output codes and to boost multiclass learning problem" Machines Learning: Proceedings of the Fourteeth International Conference, Nashville, Tennessee.

146. Schultz, M.J. and Warwick, D.N., (1971) "Vibration response: a non-destructive test for fatigue crack damage in filament-reinforced composites" Journal of Composite Materials, Vol. 5, 394-404.

147. Schultz, M.J., Pai, P.F., and Abdelnaser, A.S., (1996) "Frequency response function assignment technique for structural damage identification" Proceedings of the 14th International Modal Analysis Conference, 105-111.

148. Sestieri, A. and D'Ambrogio, W., (1989) "Why be modal: how to avoid the use of modes in the modification of vibrating systems" Proceedings of the 7th International Modal Analysis Conference, 25-30.

149. Shanno, D.F., (1978) "Conjugate gradient methods with inexact searches" Mathematics of Operations Research, Vol. 3, 244-256.

150. Shao, J. and Tu, D.,(1995) *The Jackknife and Bootstrap* Springer-Verlag, New York, USA.

151. Sharkey, A.J.C., (ed.) (1999*) Combining artificial neural nets: ensemble and modular multi-net systems, Springer-Verlag* London, U.K.

152. Silva, J.M.M., Maia, N.M.M. and Sampaio, R.P.C., (1998) "Localization of damage using FRF's curvatures: assessment and discussion in experimental cases" Proceedings of the 15th International Modal Analysis Conference, 1587-1590.

153. Stone, M., (1978) "Cross-validation: a review" Mathematische Operationsforschung Statistischen, Serie Statistics, Vol. 9, 127-139.

154. Stone, M., (1974) "Cross-validatory choice and assessment of statistical predictions" Journal of the Royal Statistical Society, Vol. B36, 111-113.

155. Srinivasan, M.G. and Kot, C.A., (1992) "Effect of damage on the modal parameters of a cylindrical shell" Proceedings of the 11th International Modal Analysis Conference, 529-535.

156. Stone, M., (1974) "Cross-validatory choice and assessment of statistical predictions" Journal of Statistical Society, vol.B36, 111-113.

157. Stubbs, N.J., Kim, J.T., and and Topole, K., (1992) "An efficient and robust algorithm for damage localization in offshore platforms" Proceeding of American Society of Civil Engineers 10th Structures Congress, 543-546.

158. Stubbs, N., and Osegueda, R., (1990) "Global damage detection in solids − experimental verification" Modal Analysis: The Journal of Analytical and Experimental Modal Analysis, Vol. 5, 67-79.

159. Surace, C. and Ruotolo, R., (1994) "Crack detection of a beam using the wavelet transform" Proceedings of the 12th International Modal Analysis Conference, 1141-1148.

160. Tarassenko, L., (1998) *A guide to neural computing applications* John Wiley & Sons Inc., New York, USA.

161. Vandiver, J.K., (1975) "Detection of structural failure on fixed platforms by measuring of dynamic response" Proceedings of the 7th Annual Offshore Technology Conference, 243-252.

162. Vandiver, J.K., (1977) "Detection of structural failure on fixed platforms by measurement of dynamic response" Journal of Petroleum Technology, 305-310.

163. Vapnik, V.N., (1995) *The nature of statistical learning theory* Springer-Verlag, New York, USA.

164. Ville, J., (1948) "Théorie ef applications de la notion de signal analytique" Cables et Transmission, Vol. 2A, 61-74.

165. Wei, M.L. and Janter, T., (1988) "Optimization of mathematical model via selected physical parameters" Proceedings of the 6th International Modal Analysis Conference, 340-345.

166. West, W.M., (1982) "Single point random modal test technology application to failure detection" The Shock and Vibration Bulletin, Vol. 52, 25-31.

167. West, W.M., (1984) "Illustration of the use of modal assurance criterion to detect structural changes in an orbiter test specimen" Proceedings of Air Force Conference on Aircraft Structural Integrity, 1-6.

168. Wheeler, J.A. and Zurek, H., (1983) *Quantum theory and measurement,* Princeton University Press, USA.

169. Wigner, E.P., (1932) "On quantum correction for thermodynamic equilibrium" Physical Review, Vol. 40, 749-759.

170. Williamson, R.C., and Helmke, U., (1995) "Existence and uniqueness results for neural network approximations", Vol. 6, No. 1, 2-13.

171. Wolff, T. and Richardon, M., (1989) "Fault detection in structures from changes in their modal parameters" Proceedings of the 7th International Modal Analysis Conference, 87-94.

172. Wolpert, D.H., (1992) "Stacked generalization" Neural Networks, Vol. 5, 241-259.

173. Worden, K.A., Ball, A., and Tomilinson, G., (1993) "Neural networks for fault location" Proceedings of the 11th International Modal Analysis Conference, 47-54.

174. Worden, K., Manson, G., and Fieler, N.R.J., "Damage detection using outlier analysis" Journal of Sound and Vibration, Vol. 229, 647-667.

175. Wu, X., Ghaboussi, J. and Garret, J.H., (1992) "Use of neural networks in detection of structural damage" Computers and Structures, Vol. 42, 649-659.

176. Yeun, M.M.F., (1985) "A numerical study of the eigenparameters of a damaged cantilever" Journal of Sound and Vibration, Vol. 103, 301-310.

177. Zimmerman, D.C and Kaouk, M., (1992) "Eigenstructure assignment approach for structural damage detection" American Institute of Aeronautics and Astronautics Journal, Vol. 30, 1848-1855.

178. Zhu, H., and Rohwer, R., (1996) "No free lunch for cross-validation" Neural Computation, Vol. 8, 1421-1426

179. Zimmerman, D.C., Kaouk, M. and Simmermacher, T., (1995) "Structural damage using frequency response functions" Proceedings of the 13th International Modal Analysis Conference, 179-184.

Appendix A
PUBLISHED WORK

REFEREED JOURNAL PAPERS: Published

1. Marwala, T., and Hunt, H.E.M., (1999) "Fault identification using finite element models and neural networks" Mechanical Systems and Signal Processing, 13(3), 475-490.
2. Marwala, T., (2000) "On damage identification using a committee of neural networks" American Society of Civil Engineering, Journal of Engineering Mechanics, 126(1), 43-50.
3. Marwala, T. and Hunt, H.E.M., (2000) "Is damage detection using vibration data in a population of cylinders feasible?" Journal of Sound and Vibration, 237(4), 727-732.
4. Marwala, T., (2001) "Probabilistic fault identification using a committee of neural networks and vibration data" American Institute of Aeronautics and Astronautics, Journal of Aircraft, 38(1), 138-146.
5. Marwala, T., and Hunt, H.E.M., (2000) "Fault identification using pseudo modal energies" International Journal of Simulation in Engineering, 1(2), 4-10.

CONFERENCE PAPERS

6. Marwala, T., and Hunt, H.E.M., (1999) "Fault identification using a committee of neural networks" Proceedings of the Conference on Identification in Engineering Systems, 102-111.
7. Marwala, T., and Hunt, H.E.M., (2000) "Feasibility of damage detection using vibration data in a population of cylinders" Proceedings of the 18th International Modal Analysis Conference, 1577-1583.
8. Marwala, T. and Hunt, H.E.M., (2000) "Probabilistic fault identification using vibration data and neural networks" Proceedings of the 18th International Modal Analysis Conference, 674-680.
9. Marwala, T. and Hunt, H.E.M., (2000) "Comparison between maximum-likelihood and Bayesian networks for fault identification" Proceedings of International Conference on Advanced Problems in Vibration Theory and Applications, 255-261.

Appendix B

DERIVING PSEUDO-MODAL-ENERGIES AND CALCULATING THEIR UNCERTAINTIES

B.1 Deriving pseudo modal energies and their derivatives

In this appendix expressions for the pseudo modal energies are derived in terms of the modal properties. Furthermore, the sensitivities of the pseudo modal energies are expressed in terms of the sensitivities of modal properties studied by Fox and Kapoor (1968).

In vibration analysis, mechanical structures can be excited using the modal hammer and vibration response measured. If the discrete Fourier Transform is applied to the displacement response X(t) and the excitation F(t), X(ω) and F(ω), respectively, [Ewins, 1995] are obtained. The X(ω) and F(ω) are the displacement and force histories in the frequency domain. The receptance frequency response function (FRF) $H_{kl}(\omega)$ caused by the excitation at k and measurement at l is defined as the ratio of the transformed displacement response X(ω) to the transformed excitation F(ω) by the following equation:

$$H_{kl}(\omega_i) = \frac{X_l(\omega_i)}{F_k(\omega_i)} \tag{B.1}$$

The receptance FRF in equation B.1 is related to the mass and stiffness matrices [Ewins, 1995] through the following equation:

$$[H(\omega)] = [-\omega^2[M] + j\omega[C] + [K]]^{-1} \tag{B.2}$$

The inertance FRF is defined as:

$$\ddot{H}_{kl}(\omega_i) = \frac{-\omega_i^2 X_l(\omega_i)}{F_k(\omega_i)} \tag{B.3}$$

The receptance FRF in equation B.1 may be written in terms of the modal properties (natural frequencies, damping ratios and mode shapes) through the use of the modal summation equation [Maia and Silva, 1997] and by assuming that damping is light as follows:

$$H_{kl}(\omega) = \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} \tag{B.4}$$

where the contribution of each mode is given by the natural frequency ω_i, the damping ratio ζ_i and ϕ_k^i is the k^{th} entry of the i^{th} normalized mode shape vector.

These frequency response functions (FRFs) are transformed into the pseudo modal energies defined as the integrals of the real and imaginary components of the FRFs over various frequency ranges. The receptance pseudo modal energies (RME) are calculated as follows [Gradshteyn *et al*., 1994]:

$$\begin{aligned} RME_{kl}^{q} &= \int_{a_q}^{b_q} \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} d\omega \\ &= \left| \sum_{i=1}^{N} \frac{2\phi_k^i \phi_l^i j}{\sqrt{4\omega_i^2 - 4\zeta_i^2 \omega_i^2}} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j\omega}{\omega_i \sqrt{1-\zeta_i^2}} \right) \right\} \right|_{a_q}^{b_q} \\ &= \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i j}{\omega_i \sqrt{1-\zeta_i^2}} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - jb_q}{\omega_i \sqrt{1-\zeta_i^2}} \right) - \arctan\left(\frac{-\zeta_i \omega_i - ja_q}{\omega_i \sqrt{1-\zeta_i^2}} \right) \right\} \end{aligned} \tag{B.5}$$

In equations B.5, a_q and b_q are the lower and upper frequency bounds for the i^{th} pseudo modal energy calculated from the FRF caused by the excitation at k and measurement at l. The advantage of this equation is that the pseudo modal energies may be calculated directly using any numerical integration scheme and without having to go through the process of modal extraction. By assuming light damping ($\zeta_j << 1$) equation B.5 becomes:

$$RME_{kl}^{q} \approx \sum_{i=1}^{N} \frac{\phi_k^i \phi_l^i j}{\omega_i} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - jb_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - ja_q}{\omega_i} \right) \right\} \tag{B.6}$$

The sensitivity of the RME in equation B.6 with respect to any parameter change may be written in the following form:

$$RME_{kl,p}^{q} = \sum_{i=1}^{N} \left[\left(\frac{j\phi_k^i \phi_l^i}{\omega_i} \right)_{i,p} B_{kl}^{q} + \left(\frac{j\phi_k^i \phi_l^i}{\omega_i} B_{kl,p}^{q} \right) \right] \tag{B.7}$$

The results in equation B.7 assume that $\partial\zeta_i/\partial g_p = 0$ and the term $B^q_{kl,p}$ stands for $\partial B^q_{kl} / \partial g_i$ where:

$$B^q_{kl} = \arctan\left(\frac{-\zeta_i\omega_i - jb_q}{\omega_i}\right) - \arctan\left(\frac{-\zeta_i\omega_i - ja_q}{\omega_i}\right) \tag{B.8}$$

The derivative $B^q_{kl,p}$ is calculated to be:

$$\begin{aligned} B^q_{kl,p} &= \frac{1}{1+\left(\frac{-\zeta_i\omega_i - jb_q}{\omega_i}\right)^2}\left(\frac{-\zeta_i\omega_{i,p} - (-\zeta_i\omega_i - jb_q)\omega_{i,p}}{\omega_i^2}\right) - .... \\ &\quad \frac{1}{1+\left(\frac{-\zeta_i\omega_i - ja_q}{\omega_i}\right)^2}\left(\frac{-\zeta_i\omega_i\omega_{i,p} - (-\zeta_i\omega_i - ja_q)\omega_{i,p}}{\omega_i^2}\right) \\ &= \frac{jb_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + jb_q)^2} - \frac{ja_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + ja_q)^2} \end{aligned} \tag{B.9}$$

and

$$\begin{aligned} \left(\frac{j\phi^i_k\phi^i_l}{\omega_i}\right)_{,p} &= \frac{j\omega_i(\phi^i_k\phi^i_l)_{,p} - j\phi^i_k\phi^i_l\omega_{i,p}}{\omega_i^2} \\ &= \frac{j}{\omega_i}[\phi^i_{k,p}\phi^i_l + \phi^i_k\phi^i_{l,p}] - \frac{1}{\omega_i^2}j\phi^i_k\phi^i_l\omega_{i,p} \end{aligned} \tag{B.10}$$

Substituting equations B.8-B.10 into equation B.7, the derivatives of the receptance pseudo modal energies (RMEs) with respect to parameter changes may be written as follows:

$$\begin{aligned} &RME^q_{kl,p} \\ &\approx \sum_{i=1}^{N}\left[\begin{array}{l} \left\{\frac{j}{\omega_i}[\phi^i_{k,p}\phi^i_l + \phi^i_k\phi^i_{l,p}] - \frac{1}{\omega_i^2}j\phi^i_k\phi^i_l\omega_{i,p}\right\} \\ \left\{\arctan\left(\frac{-\zeta_i\omega_i - jb_q}{\omega_i}\right) - \arctan\left(\frac{-\zeta_i\omega_i - ja_q}{\omega_i}\right)\right\} \\ -\left\{\frac{\phi^i_k\phi^i_l}{\omega_i}\right\}\left\{\frac{b_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + jb_q)^2} - \frac{a_q\omega_{i,p}}{\omega_i^2 + (\zeta_i\omega_i + ja_q)^2}\right\} \end{array}\right] \end{aligned} \tag{B.11}$$

Equation B.11 shows that the sensitivity of the pseudo modal energies is a function of the natural frequencies and mode shapes and their respective derivatives. By substituting the derivatives of the mass, damping and stiffness matrices into equation B.11 using expressions given by Fox and Kapoor (1968) or Adhikari (2000) gives the sensitivity of the RMEs in terms of the mass and stiffness matrices, which are directly related to the physical properties. This shows that the pseudo modal energies may be related directly to the physical properties of the structure.

Some of the techniques that are used to measure the vibration data measure acceleration instead of displacement. In such a case it is relatively more efficient to calculate the inertance pseudo modal energies (IMEs) than the RMEs. The IMEs are derived as follows:

$$
\begin{aligned}
&IME_{kl}^{q} \\
&= \int_{a_q}^{b_q} \sum_{i=1}^{N} \frac{-\omega^2 \phi_k^i \phi_l^i}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} d\omega \\
&= \left| \sum_{i=1}^{N} \phi_k^i \phi_l^i \omega + \phi_k^i \phi_l^i \zeta_i \omega_i j \ln(-\omega^2 + 2\zeta_i \omega j + \omega_i^2) + \left(2\zeta_i^2 \omega_i^2 - \omega_i^2\right) \int \frac{\phi_k^i \phi_l^i d\omega}{-\omega^2 + 2\zeta_i \omega_i \omega j + \omega_i^2} \right|_{a_q}^{b_q} \qquad \text{(B.12)} \\
&= \sum_{i=1}^{N} \left[\begin{array}{l} \phi_k^i \phi_l^i (b_q - a_q) + \zeta_i \omega_i \, \phi_k^i \phi_l^i j \left\{ \ln\left(\frac{-b_q^2 + 2\zeta_i \omega_i b_q j + \omega_i^2}{-a_q^2 + 2\zeta_i \omega_i a_q j + \omega_i^2} \right) \right\} \\ + \left(2\zeta_i^2 \omega_i^2 - \omega_i^2\right) \frac{\phi_k^i \phi_l^i j}{\omega_i} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \end{array} \right]
\end{aligned}
$$

If it is assumed that damping is negligible, then $(b_q\text{-}a_q)/\omega_I \ll 1$ and equation B.12 reduces to:

$$
IME_{kl}^{q} \approx \sum_{i=1}^{N} \left[\phi_k^i \phi_l^i (b_q - a_q) - \omega_i^2 \frac{\phi_k^i \phi_l^i j}{\omega_i} \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \right] \qquad \text{(B.13)}
$$

Assuming that $\zeta_{i,j} = 0$ and $\zeta^2_i \ll 1$, the derivative of IME is:

$$
\begin{aligned}
&IME_{kl,p}^{q} \\
&\approx \sum_{i=1}^{N} \left[\begin{array}{l} (b_q - a_q)(\phi_{k,p}^i \phi_l^i + \phi_k^i \phi_{l,p}^i) - j\omega_{i,p} \phi_k^i \phi_l^i \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \ldots \\ - j\omega_i \left(\phi_{k,p}^i \phi_l^i + \phi_k^i \phi_{l,p}^i\right) \left\{ \arctan\left(\frac{-\zeta_i \omega_i - j b_q}{\omega_i} \right) - \arctan\left(\frac{-\zeta_i \omega_i - j a_q}{\omega_i} \right) \right\} \ldots \\ + \omega_i \, \phi_k^i \phi_l^i \left\{ \frac{b_q \omega_{i,p}}{\omega_i^2 + (\zeta_i \omega_i + j b_q)^2} - \frac{a_q \omega_{i,p}}{\omega_i^2 + (\zeta_i \omega_i + j a_q)^2} \right\} \end{array} \right] \qquad \text{(B.14)}
\end{aligned}
$$

B.2. Calculations of uncertainties of the pseudo modal energies

In this section the uncertainties of pseudo modal energies are estimated from the known uncertainties of modal properties using equations B.11 and B.14, modal properties extracted from the FRFs contaminated with ±1% noise, and frequency bandwidths defined in Section 2.4. The FRFs used here were calculated from the simulated beam in Chapter 2. The frequency bandwidths are 18-14, 155-240, 484-620, 1014-1151 and 1726-1863 Hz. The uncertainty of natural frequency is 0.125% and uncertainty of mode shapes is 10% [Maia and Silva, 1997].

Using these uncertainties of the modal properties, the given frequency bandwidths and the identified modal properties, the uncertainties of the pseudo modal energies are calculated and the results shown in Tables B.1 and B.2 are obtained. The pseudo modal energies are calculated from the FRF due to excitation in node 10 and acceleration measurement in node 5 (see Figure 2.1).

Table B.1 shows the average uncertainties of the real part of the receptance pseudo modal energies to be 7.9% and imaginary part to be 9.0%. Table B.2 shows the average uncertainties for the real part of the inertance pseudo modal energies of 5.0% and the imaginary part of 8.4%. In general, the average uncertainty of the inertance pseudo modal energies is lower than that of the receptance pseudo modal energies. The average uncertainty of the imaginary part of the pseudo modal energies for mode 1 is quite high which is consistent with Figure 2.5 where the first mode was observed to be sensitive to noise. Choosing the frequency bandwidth of integration carefully when calculating the pseudo modal energies may reduce the level of uncertainties on the pseudo modal energies calculated.

Table B.1. The uncertainties of receptance pseudo modal energies calculated from the uncertainties of the modal properties. Key: $\Re$ = Real part; $\Im$ = Imaginary part

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
$\Re$(RME) (%)	1.3	10.0	7.2	7.5	13.7
$\Im$(RME) (%)	16.9	14.8	5.8	0.8	7.0

Table B.2. The uncertainties of inertance pseudo modal energies calculated from the uncertainties of the modal properties. Key: $\Re$ = Real part; $\Im$ = Imaginary part

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
$\Re$(IME) (%)	3.6	0.9	8.5	3.0	8.9
$\Im$(IME) (%)	15.1	12.9	5.9	0.8	7.1

Appendix C

MODAL PROPERTIES EXTRACTION

This section outlines the procedure implemented by Balmès (1997[a,b]) in the Structural Dynamics Toolbox to extract the modal properties from the measured frequency response functions (FRFs). The modal properties are extracted by minimising the distance between the measured FRFs and those predicted by a theoretical model. The theoretical model usually contains the FRFs expressed in terms of the modal properties. In this thesis, a quadratic cost function is used and may be written as follows:

$$J_{Quad} = \sum_{\left\{\begin{array}{l}\text{accelerometer j,}\\ \text{impulse k,}\\ \text{frequency l}\end{array}\right\}} \left| H_{jk}^{model} - H_{jk}^{measured} \right| \tag{C.1}$$

The FRFs of a linear and diagonalisable model can always be written as:

$$[H(j\omega)] = \sum_{i=1}^{2N} \left(\frac{[R_i]}{j\omega - \lambda_i} \right) = \sum_{j=1}^{2N} \left(\frac{\{\bar{\phi}_i\}\{\bar{\phi}_i\}^T}{j\omega - \lambda_i} \right) \tag{C.2}$$

where $\{\bar{\phi}_i\}$ is the i^{th} complex mode shape vector.

In an experiment, only the modes in the frequency range of interest are measured. The contribution of modes outside of this range can only be approximated. Balmès (1997[b]) approximated the contribution of these modes by adding a factor $[E(i\omega)]$ into equation C.2 to form a second-order model. Therefore, equation C.2 can be written in the following form:

$$[H(j\omega)] = \sum_{i \in \left\{\begin{array}{l}\text{number of modes}\\ \text{in the frequency bandwidth}\end{array}\right\}}^{N} \left(\frac{\{\bar{\phi}_i\}\{\bar{\phi}_i\}^T}{j\omega - \lambda_j} + \frac{\left[\{\bar{\phi}_i\}\{\bar{\phi}_i\}^T\right]^*}{j\omega - \lambda_j} \right) + [E(j\omega)] \tag{C.3}$$

where N is the number of identified modes, λ_i is the i^{th} pole and superscript * stands for complex conjugate.

Equation C.3 may be rewritten and the resulting expression is:

$$[H(\lambda_j, R, s)] = [\Phi(\lambda_j, s)][R] \tag{C.4}$$

where $[R] = \{\bar{\phi}_i\}\{\bar{\phi}_i\}^T$ is called the residue matrix.

In equation C.4, a set of poles which minimises the cost function in equation C.1 also corresponds to the resolution of the linear least square regression in R. The residue may thus be calculated to give:

$$[R(\lambda_j)] = \arg\min_R \left(\sum_{i \in \left\{ \begin{array}{l} \text{number of modes} \\ \text{in the frequency bandwidth} \end{array} \right\}} \left| \Phi(\lambda_m, \omega_l) R_{mjk} - H_{jk}^{Test}(\omega_l) \right| \right) \tag{C.5}$$

Equation C.5 can be solved explicitly to give:

$$[R(\lambda_j)] = \left[\Phi(\lambda_j)^T \Phi(\lambda_j) \right]^{-1} [\Phi(\lambda_j)]^T [H_{Test}] \tag{C.6}$$

Equation C.6 gives the explicit expression of the quadratic cost in equation C.1 and may be written as:

$$\begin{aligned} J_{Quad}(\lambda_j) &= \text{trace}\left((H - H_{Test})^T (H - H_{Test}) \right) \\ &= \text{trace}\left(H_{Test}^T \left(I - \Phi(\lambda_j) \left(\Phi(\lambda_j)^T \Phi(\lambda_j) \right)^{-1} \Phi(\lambda_j)^T \right)^2 H_{Test} \right) \end{aligned} \tag{C.7}$$

The derivative of the quadratic cost J in equation C.7 may be written as follows:

$$\frac{\partial J}{\partial \theta} = 2(H - H_{Test})^T \frac{\partial \Phi}{\partial \theta} R \tag{C.8}$$

The modal properties are identified by estimating the poles λ_i, and re-computing the residues until convergence is obtained.

Appendix D

BACK-PROPAGATION METHOD

In this study, the output units represent the identity of the faults. In Figure 3.1 the output of the j^{th} hidden unit is obtained by calculating the weighted linear combination of the d input values to give:

$$a_j = \sum_{i=1}^{d} w_{ji}^{(1)} x_i \tag{D.1}$$

Here $w_{ji}^{(1)}$ indicates weight in the first layer, going from input i to hidden unit j while $w_{j0}^{(1)}$ indicates the bias for the hidden unit j. The activation of the hidden unit j is obtained by transforming the output a_j in equation D.1 into z_j, which is shown in Figure 3.1, as follows:

$$z_j = f_{inner}(a_j) \tag{D.2}$$

The output of the second layer is obtained by transforming the activation of the second hidden layer using the second layer weights. Given the output of the hidden layer z_j in equation D.2, the output of unit k may be written as:

$$a_k = \sum_{j=0}^{M} w_{kj}^{(2)} y_j \tag{D.3}$$

Similarly equation D.3 may be transformed into the output units by using some activation function as follows:

$$y_k = f_{outer}(a_k) \tag{D.4}$$

Combining equations D.1, D.2, D.3 and D.4 the input x to the output y can be related by a two-layered non-linear mathematical expression, which may be written as follows:

$$y_k = f_{outer}\left(\sum_{j=0}^{M} w_{kj}^{(2)} f_{inner}\left(\sum_{i=0}^{d} w_{ji}^{(1)} x_i \right) \right) \tag{D.5}$$

Here d is the number of input units, M is the number of hidden units, w_{ij} is the weight-vector, the function $f_{outer}(\bullet)$ is sigmoid and f_{inner} is a hyperbolic tangent function. These functions are defined as:

$$f_{outer}(\nu) = \frac{1}{1+e^{-\nu}} \tag{D.6}$$

and

$$f_{inner}(\nu) = \tanh(\nu) \tag{D.7}$$

The weights (w_i) and biases (weights with subscripts 0) in the hidden layers are varied until the error between the network prediction and the output from the training data is minimised.

Given the training set $D = \{x_k, t_k\}_{k=1}^{N}$ and assuming that the targets t_k are sampled independently given the inputs x_k and the weight parameters w_{kj} the sum of square of error cost function E may be written as:

$$E = \frac{1}{2}\sum_n \sum_k (y_{nk} - t_{nk})^2 \tag{D.8}$$

Here n is the index for the training pattern and k is the index for the output units. Alternatively, the cross-entropy cost function [Hopfield, 1987; Hinton, 1989] may be written as:

$$E = -\sum_n \sum_k \{t_{nk} \ln(y_{nk}) + (1-t_{nk})\ln(1-y_{nk})\} \tag{D.9}$$

The minimisation of E is achieved by calculating the derivative of the errors in equations D.8 or D.9 with respect to the weight. The derivative of the error is calculated with respect to the weight which connects the hidden to the output layer and may be written using the chain rule as follows:

$$\begin{aligned} \frac{\partial E}{\partial w_{kj}} &= \frac{\partial E}{\partial a_k}\frac{\partial a_k}{\partial w_{kj}} \\ &= \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial a_k}\frac{\partial a_k}{\partial w_{kj}} \\ &= \sum_n f'_{outer}(a_k)\frac{\partial E}{\partial y_{nk}} z_j \end{aligned} \tag{D.10}$$

where z_j is given in equation D.2. The derivative of the error with respect to weight which connects the hidden to the output layer may be written using the chain rule as follows:

$$\begin{aligned} \frac{\partial E}{\partial w_{kj}} &= \frac{\partial E}{\partial a_k}\frac{\partial a_k}{\partial w_{kj}} \\ &= \sum_n f'_{inner}(a_j)\sum_k w_{kj} f'_{outer}(a_k)\frac{\partial E}{\partial y_{nk}} \end{aligned} \tag{D.11}$$

The derivative of the sum of square cost function in equation D.8 may be written as:

$$\frac{\partial E}{\partial y_{nk}} = t_{nk} - y_{nk} \tag{D.12}$$

The derivative of the cross-entropy cost function in equation D.9 may be written as:

$$\frac{\partial E}{\partial y_{nk}} = \frac{(y_{nk} - t_{nk})}{y_{nk}(1 - y_{nk})} \tag{D.13}$$

The derivatives of the logistic activation function in equation D.6 is:

$$f'_{outer}(a_k) = f(a_k)(1 - f(a_k)) \tag{D.14}$$

while that of the hyperbolic tangent function is:

$$f'_{inner}(a_j) = \sec h^2(a_j) \tag{D.15}$$

This appendix shows the derivatives of the errors with respect to weights. Equations D.12 and D.13 show the two possible derivatives of the cost functions that could be incorporated into equations D.10 and D.11. Equations D.14 and D.15 show the derivatives of the two possible activation functions.

Appendix E

SCALED CONJUGATE GRADIENT METHOD

In this appendix the scaled conjugate optimisation method is described. Before introducing the scaled conjugate gradient method the conjugate gradient method is introduced. In supervised neural network training, the main goal is to identify weights that give the best prediction of the output whenever presented with the input. The scaled conjugate gradient method is used to sample through the weight space until the weight vector that minimises the distance between the neural network prediction and the target data is obtained.

E.1. CONJUGATE GRADIENT METHOD

The weight vector which gives the minimum error is achieved by taking successive steps through the weight space as follows:

$$w(n+1) = w(n) + \Delta w(n) \tag{E.1}$$

where n is the iteration step and Δ represents change. Different algorithms choose this step size differently. In this section, gradient descent method [Robbins and Monro, 1951] will be discussed, followed by how it is extended to the conjugate gradient method [Hestenes and Stiefel, 1952]. For the gradient descent method, the step size in equation E.1 is defined as:

$$\Delta w^n = -\eta \nabla E(w(n)) \tag{E.2}$$

where the parameter η is the learning rate and the gradient of the error is calculated using the back-propagation technique described in Appendix D. If the learning rate is sufficiently small, the value of error will decrease at each successive step until a minimum is obtained. The

disadvantage with this approach is that it is computationally expensive compared to other techniques.

For the conjugate gradient method the quadratic function of error is minimised at each iteration over a progressively expanding linear vector space that includes the global minimum of the error [Luenberger, 1984; Fletcher, 1987; Bertsekas, 1995]. For the conjugate gradient procedure, the following steps are followed [Haykin, 1999]:

1. Choose the initial weight $\mathbf{w}(0)$.
2. Calculate the gradient vector $\nabla E(\mathbf{w}(0))$.
3. At each step n use the line search to find $\eta(n)$ that minimises $E(\eta)$ representing the cost function expressed in terms of η for fixed values of $\mathbf{w}$ and $-\nabla E(\mathbf{w}(0))$.
4. Check that the Euclidean norm of the vector $-\nabla E(\mathbf{w}(n))$ is sufficiently less than that of $-\nabla E(\mathbf{w}(0))$.
5. Update the weight vector $\mathbf{w}(n+1) = \mathbf{w}(n) - \eta(n)\nabla E(\mathbf{w}(n))$.
6. For $\mathbf{w}(n+1)$ compute the updated gradient $\nabla E(\mathbf{w}(n+1))$.
7. Use Polak-Ribiére method to calculate $\beta(n+1)$

$$\beta(n+1) = \frac{\nabla E(\mathbf{w}(n+1))^T \left(\nabla E(\mathbf{w}(n+1)) - \nabla E(\mathbf{w}(n))\right)}{\nabla E(\mathbf{w}(n))^T \nabla E(\mathbf{w}(n))}$$

8. Update the direction vector $\nabla E(\mathbf{w}(n+1)) = \nabla E(\mathbf{w}(n+1)) - \beta(n+1)\nabla E(\mathbf{w}(n))$.
9. Set $n=n+1$ and go back to step 3.
10. Stop when the following condition is satisfied

$\left\|\nabla E(\mathbf{w}(n))\right\| = \varepsilon \left\|\nabla E(\mathbf{w}(0))\right\|$

where ε is a small number.

E.2. SCALED CONJUGATE GRADIENT METHOD

The scaled conjugate gradient method differs from conjugate gradient method in that it does not involve the line search described in step 3 in the previous section.

The step-size (see step 3) is calculated by using the following formula (Møller):

$$\bar{\eta}(n) = 2\left(\eta(n) - \frac{\nabla E(n)^T H(n) \nabla E + \eta(n) \left\| \nabla E(n) \right\|^2}{\left\| \nabla E(n) \right\|^2} \right) \tag{E.3}$$

where H is the Hessian matrix of the gradient.

APPENDIX F
A DETAILED FIGURE OF A CYLINDER

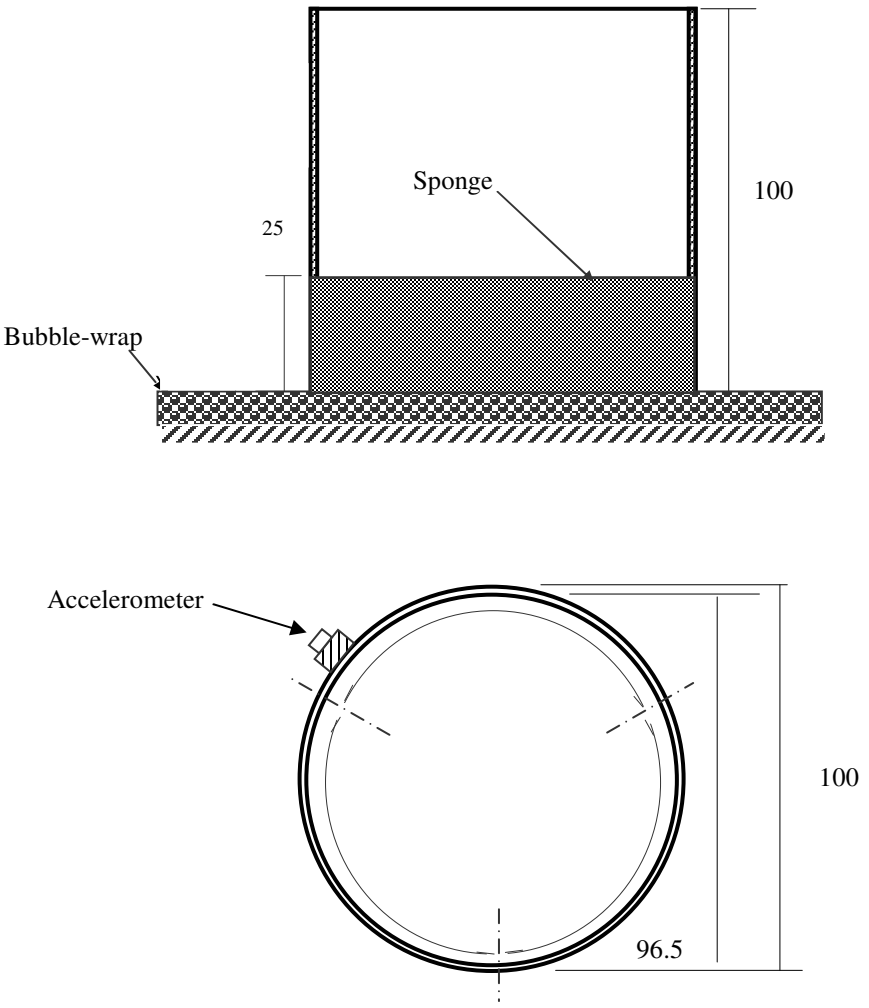


Figure E.1. The diagram of a cylindrical shell showing dimensions in millimetres.