Investigating the Effect of Variable Mass Loading in Structural Health Monitoring from a Machine Learning Perspective

A thesis submitted to the University of Sheffield

for the degree of Doctor of Philosophy in the Faculty of Engineering

by

Sharafiz Abdul Rahim

Department of Mechanical Engineering

University of Sheffield

July 2018

ABSTRACT

Structural Health Monitoring (SHM) aims at monitoring bridges or other engineering structures by accessing their condition and alerting the user about new defects in the structure by means of autonomous sensing system. For an aircraft structure, the needs for a reliable SHM increases considerably due to its expensive value and the requirement of higher safety regulations. Due to these reasons and the airline's sustainable economy, Structural Health Monitoring (SHM) techniques seemingly beneficial as a solution for automated and continuous monitoring hence reducing the needs for a manual inspection and structures disassembly. The technology potentially brings higher revenues to the operator by reducing the costly manual inspections and avoiding the situation where the aircraft has to be suspended for inspection procedures.

This study is centered around investigating the effects of operational loading variations on damage detection, particularly of an aircraft wing, using Vibration Based Damaged Detection (VBDD) using a machine learning framework. It is comprised of two main parts, the first of which involves implementing SHM on a wing box pseudo-fuel-tank structure. To do this frequency response function (FRF) data is acquired and analyzed using several machine learning approaches.

Standard Principal Component Analysis (PCA) and nonlinear PCA through kernel Gaussian PCA are tested using the data as preliminary techniques of isolating damage effects from the loading effects on the selected features. A more concrete damage identification model is established using Artificial Neural Network (ANN) via nonlinear

regression analysis. To obtain a better insight of the data distribution model and characteristics, various models based on the Gaussian Mixture Model (GMM) are incorporated on the wing box data sets.

The second part covers the pinnacle of the current work, which involves performing similar SHM procedures using VBDD on a full-scale aircraft wing. The same machine learning approaches are used for this data acquired from the wing of a small Jabiru airplane. Decision trees via Random Forest algorithm is performed in an effort to model the data in a hierarchical manner by connecting the data variables of different variance at different levels.

There has been increasing research surrounding operational loading and environmental variability in the field of SHM [1]–[6]. However, there is no significant work found in the literature on investigating fuel tank loading effects on SHM, particularly on aircraft wings. This current study aims to fill this research area by highlighting the influence of fuel tank loading variables on damage detection and SHM.

ACKNOWLEDGEMENTS

It is a great blessing to be able to continue this journey in life and to reach this stage. Without doubt, the highest gratitude must go to the Almighty God, to Whom is worth the highest praise. Indeed, all living things in the earth and heavens shall be returned to Him. With great responsibility, each man will be asked for what he has done with his given abilities.

It is a huge blessing to have the author's parents, his mother Wan Enshah and his father Abdul Rahim, who never cease to pray for the author's success for this life and thereafter, and who always motivate the author to stay strong and optimistic in every rise and fall in his life for the last 42 years.

Indeed, it is a wonderful blessing to have been gifted with an amazing wife Dr Munirah, who has devoted all her life to live together with the author for the last 10 years with endless understanding and loyalty, and sacrificing her time and energy in nurturing our most beloved gifts in this world – Zahraa, Safiyaa, Umar and Aliyaa. Indeed, they are always the source of the author's enjoyment and warmth in his life.

Certainly, it is a blessing and gift to have Dr Graeme Manson as the main supervisor, whose understanding and support has ensured the author to complete his PhD, beginning the work from scratch to what it is now today.And importantly, Professor Keith Worden who always makes the time for the author in providing valuable guidance in the field of SHM and machine learning.

Lastly to all the author's friends and colleagues at the University of Sheffield. Thank you.

NOMENCLATURE

Acronyms

ANN: Artificial Neural Network

AANN: Auto-Associative Neural Network

CVM: Comparative Vacuum Monitoring

EM: Expectation Maximization

FAA: Federal Aviation Administration

FRF: Frequency Response Function

GMM: Gaussian Mixture Model

JW: Jabiru Wing

KPCA: Kernel Principal Component Analysis

MLE: Maximum Likelihood Estimation

OEV: Operational and Environmental Variability

PCA: Principal Component Analysis

PC: Principal Component

SHM: Structural Health Monitoring

SMS: Structural Monitoring System

SPR: Statistical Pattern Recognition

VBDD: Vibration Based Damaged Detection

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

This chapter introduces the motivation for undertaking the current research, namely damage detection under the confounding influence of operational loading conditions within Structural Health Monitoring (SHM). In this context, vibration based damage detection (VBDD) will be carried out in the interest of monitoring the structural health of an aircraft undergoing incremental changes in operational loading. The underlying concept behind the SHM framework is briefly described here, followed by a short note on the statistical pattern recognition in the field of SHM which is utilized in the current work. In the final part, an introduction to the layout of this thesis is presented.

1.1 Motivation

The primary interest of the current research is to investigate the effect of confounding influences arising from operational loading variations, particularly due to changes in the fuel tank loading, upon the ability to conduct damage detection on an aircraft wing structure. In reality, the changing of fuel load of an operating aircraft typically creates a similar effect to a damage present in the structure as illustrated in Figure 1.1. In the aircraft

industry, hot spot monitoring is a regulated practice where inspection and maintenance has to be carried out in a fixed schedule on certain structures including the wing box, due to their high susceptibility to damage. Consequently, this poses a significant challenge if one plans to detect any malfunctions in the structure.

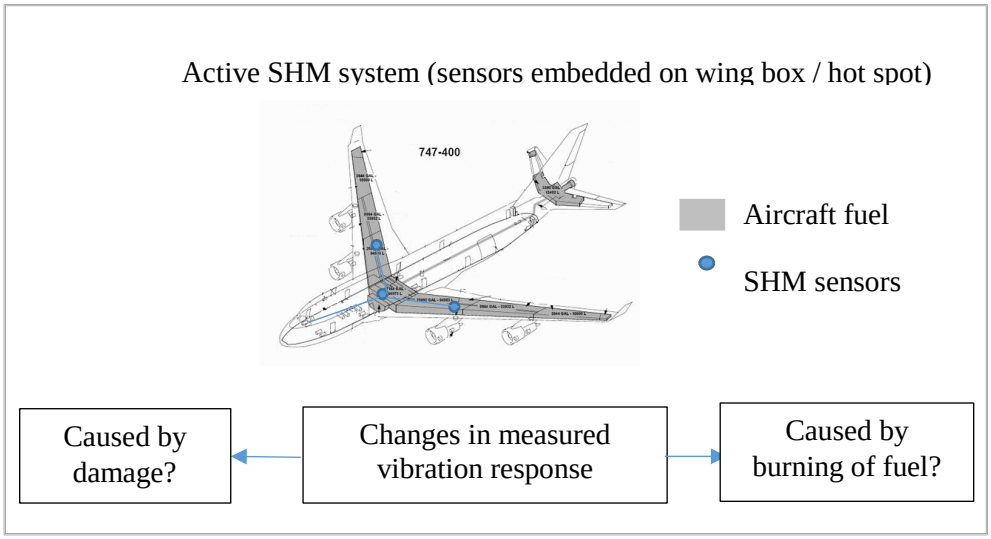


Figure 1.1: The changing of mass loading due to the burning of fuel as the operational loading variables of an aircraft in flight.

In the basic premise of Vibration Based Damage Detection (VBDD) theory, the effects of changing in mass will alter the natural frequency of the system, similar to the different levels of damage that will also alter the stiffness property of the system. Based on this principle, the study aims to investigate and explore various ways of unmasking the damage effects from the loading-sensitive features by means of Statistical Pattern Recognition (SPR) and probabilistic approach.

In this thesis, Statistical Pattern Recognition (SPR) using Principal Component Analysis (PCA) will be exploited and further extended to kernel Gaussian Principal Component Analysis (kPCA) to establish a nonlinear form of PCA. The foundation of SPR lies on the concept of learning representations from training data in order to distinguish between the damage and undamaged states of a new measured data set. This is the idea of machine leaning which is to learn the relationship from the data by means of a mathematical and statistical approach [1], [7], [8]. This concept is adopted throughout this thesis as the means

of accomplishing a pattern recognition in making the decision about the health condition of a structure.

Under the influence of operational loading variations, the feature caused by loading effects can be misleading and falsely interpreted as detection of damage. Such false alarms can result in costly measures of dismantling and repairing the structure. Therefore, it is crucial to differentiate the features caused by damage from those caused by loading, before concluding the health condition of the structure by means of SPR.

Small levels of damage can be challenging to detect due to its effects being nonsignificant and the resulting feature mostly remaining hidden in the high dimensional data space. To reveal this feature, an intelligent machine learning algorithm should be considered as a means to detect the damage. The work in this thesis attempts to address this concern by exploring kernel Gaussian PCA, which is a nonlinear form of PCA. Kernel PCA has been used very successfully in the fields of facial recognition and pattern recognition [9]–[11] following the wide application of linear PCA in these fields [12], [13]. Thus, it will be the scope of visualising the data pattern corresponding to various damage severities in the SHM paradigm. It is in the interest of the current work to resolve the issues pertaining to linear PCA in the scope of visualisation and damage detection.

Another motivation that drives the current work is the fact that SHM has been accepted and recognised as an economically competitive system for aircraft maintenance programmes. A significant aspect of considering SHM for structural maintenance is the result of successful project collaboration among the main SHM players in the aircraft industry, which are Sandia National Labs, Federal Aviation Administration (FAA), Boeing and Structural Monitoring Systems (SMS). They have started to work in partnership in collecting and verifying the data from SHM sensors mounted on operating aircrafts since February 2014 [14]. This development brought a new phase of implementing SHM on operating aircrafts using active inflight sensors. The programme implements local based damage detection using Comparative Vacuum Monitoring (CVM) sensors fitted on the wing box fittings. The approach of using CVM allows damage detection in a localised

region on a particular component. The main obstacle is that the sensors can be placed in locations where damage does not actually exist, which may result in cost inefficiency.

In view of SHM implementation into the maintenance of aerospace structures, the current study aims to present another approach of SHM, using Vibration Based Damage Detection (VBDD) as a global method of detecting damage. Using VBDD as an alternative to local inspection methods such as radiography, eddy-current method, acoustic and ultrasonic methods, it can provide an economic advantage through internal damage detection without the need to dismantle the component. Despite this advantage, there are many challenges arising from VBDD, one of which is the operational and environmental variability. A brief discussion about this method of detection is presented in Chapter 2.

SHM strategies that utilise Statistical Pattern Recognition (SPR) via machine learning algorithms can be implemented using condition based maintenance, which means that the structure is inspected when there is a signal from an offset damage present in the structure. Using vibration based SHM techniques, a component or full structure from a complicated system can be examined without the need to disassemble and reassemble the structure of various components, which can be costly and renders the structure unavailable for service. However, there are operational and environmental variability (OEV) factors that need to be addressed as they might affect the structural response of the structure while in service. These effects can arise from varying OEVs, for instance while the aircraft is on the ground for a short transit where fuel loading and payloads are mostly accounted for, compared to the aircraft under actual flight conditions. These factors can considerably influence the measured structural response by masking any possible effects of damage if the SHM procedures are to be implemented as condition based maintenance.

The changing trend of SHM implementation into new aircraft maintenance programmes promises a good prospect for SHM application, especially in its transition from laboratory-based conditions into real operational environments for commercial passenger aircrafts. Moving forward in that direction, the current work presents vibration based SHM using piezo electric sensors, focusing on the effects of varying operational loading on damage detectability in an aircraft wing. As discussed earlier, the main challenge arising from

vibration based SHM is that the measurements derived from the sensors are not only very sensitive to the presence of damage in the structure, but are also highly influenced by operational loading and environmental variations during the data measurements. To the author's knowledge, up to the present time, there is no research published in the field of vibration based SHM concerning the influence of operational fuel loading variations particularly on damage detection of aircraft wings.

The understanding of the dynamic behaviour of the wing structure under the influence of operational fuel tank loading is essential to develop a model for damage detection for SHM. Having knowledge on the data response under the influence of operational loading variations is paramount, and the current work specifically addresses the effects from varying fuel tank loading on the damage detection and the identification of damage severity.

The objective of the current work is to provide a SHM framework for detecting damage and damage severity classification of an aircraft wing under the effects of varying fuel tank loading. The approach is based on SPR and machine learning algorithms, which are firstly used to differentiate various damage severities from various loading conditions. Secondly, the features of identified loading variabilities are used to separate those features from the undamaged state from the damage states. It also intends to classify the features of various damage severities. All of these pattern recognition of the features can be obtained feasibly using machine learning paradigms which are encompassed mostly in this current work.

1.2 A brief introduction on SHM

Structural Health Monitoring (SHM) relates to a series of damage identification strategies that monitors the health states of aerospace structures, bridge infrastructures and mechanical engineering structures through the evaluation of responses acquired from sensors on the structure with minimal human intervention. The key processes involved in a SHM strategy includes operational evaluation, data acquisition, feature extraction and SPR. All of these components play an important role in reaching the final stage of SHM

where a decision can be made based on the valid hypotheses of whether the feature can reliably indicate damage. Only brief points on the SHM strategy shall be presented here and for further detail the reader can refer to [1] on general Structural Health Monitoring procedures.

The field of SHM has developed to a state where a number of fundamental axioms that represent SHM principles have emerged. Worden et al. suggest nine axioms for SHM, and four of the axioms which are particularly relevant to this work are presented as follows [1]:

Axiom II: Damage assessment requires a comparison between two system states.

Axiom III: Identifying the existence and location of damage can be done in an unsupervised learning mode, but identifying the type of damage present and the damage severity can generally be done in a supervised learning mode.

Axiom IVa: Sensors cannot measure damage. Feature extraction through signal processing and statistical classification are necessary to convert sensor data into damage information

Axiom IVb: Without intelligent feature extraction, the more sensitive a measurement is to damage, the more sensitive it is to changing operational and environmental conditions.

1.2.1 Operational evaluation

In SHM strategy, variabilities can arise from many external factors such as temperature fluctuations, operational loading variations and humidity. The first step is to set the limitations on what is to be monitored and how the monitoring could be performed based on these external factors. In the initial stage of SHM, it is important to justify the economical and safety benefits behind the SHM work.

1.2.2 Data acquisition

Data acquisition decisions that encompass selection of type, capability and number of sensors for acquiring the data response also depend on economic considerations and the application specific for that application. At this step, the sensor locations and the bandwidth of the acquisition hardware are again determined and justified based on the economic and application considerations. Data pre-processing begins after data sensing level where it prepares the data for feature extraction. Pre-processing includes data cleansing and data normalisation. Data cleansing relates to the removal of data samples with high noise or those seemingly deviating from most of the samples belonging to the same class label. Normalisation refers to the separation of changes in the system responses caused by benign operational and environmental variations from the changes caused by damage. In this context, the current work emphasises on the data normalisation in order to obtain the data response that can give better representation of the actual structure's damage condition.

1.2.3 Feature extraction

The task of feature extraction is to distinguish a selected measured response between an undamaged state and a damaged state with the intention to use it as a training or baseline set for SPR. A feature that contains the most significant data behaviour, for instance if using frequency spectrum, the natural frequencies that are distinct and with larger peaks, should be of priority. For damage detection strategy where damage is normally a local phenomenon, it is suggested to consider higher frequency modes where damage could be likely present.

With supervised data learning, where prior knowledge about the structural health states and environmental and operational loading variabilities are known, and equipped with good data pre-processing, the process is more guaranteed to produce a reliable and accurate feature extraction. The information from the measured response on the changes in frequency peaks due to damage and OEV can be used as a comparison to the measured response of the undamaged structure under the equivalent OEVs. From the frequency

spectrum response acquired, damage-sensitive features are extracted and selected for pattern recognition in the latter stages.

Features can be selected simply on the basis of engineering judgement. Feature extraction can generate a feature vector with less dimensions. Data feature with low dimensionality provides significant advantages in SPR as the number of data samples grows significantly with the dimensions of the data.

In this work, Chapter 4 is mostly related to operational evaluation, data acquisition and data pre-processing procedures. Chapter 5 and Chapter 6 relate mostly to feature extraction using factorised sampling and PCA for dimensional reductions. Chapter 7 presents the starting work of SPR using specific machine learning algorithms such as Gaussian Mixture Models (GMM) and Neural Network. SPR is briefly described in the next section.

1.2.4 Statistical pattern recognition (SPR) paradigm

Prior to acquiring samples from a structure that may already be exposed to damage, data samples from an undamaged condition should be recorded and be called as the training data set. This data from the undamaged structure should be inclusive of all operational or environmental conditions in order to establish a comprehensive dynamic behaviour of the structure. Under these circumstances, the training set should provide a good representation of the structure with no damage and operating in normal conditions. This step then allows a comparison between two different states of the structure, one of which is assumed to represent an undamaged state which is represented by the training set, and the other which possibly contains some damage flaws in the structure. This is the basis for SPR where a physical based model is adopted and compared to the changes in measured data from test data produced by the structure with an unknown health condition [1], [15].

Through machine learning algorithms, statistical pattern recognition (SPR) is performed by means of mathematical and statistical framework and associating measured data with given class labels of varying damage severities. SPR is exploited using the concept of learning

representations from the training data set, from which the basis of this current work is established.

1.2.5 A shift into condition-based maintenance

Since 1991, research efforts related to structural integrity and non-destructive inspection led by Sandia National Laboratory was initiated in order to improve aviation safety. This move was as a result from the FAA decision to reduce the number of aviation incidents that had occurred during that time. In 2011, Sandia National Laboratories, Delta Air Lines and an airline manufacturer company had reached a more serious decision and commitment to implement SHM in their maintenance programme and to have the SHM technology certified and approved by the FAA. This step would allow the SHM technology to be implemented throughout all commercial aircrafts to improve inspection, maintenance and repair processes without affecting the aircraft turnaround and reducing flying hours [14], [16].

SHM is beneficial by ensuring that maintenance is performed based on the actual condition of the aircraft when it is necessary, rather than fixed schedules and inspection routines. This condition-based maintenance can help reduce airplane downtime and improve profit margins of the commercial airliner. The work by Sandia and the team also aims to validate the sustainability and reliability of built-in sensors on aircrafts under real operating conditions and flying in real time. The signal was collected using the acquisition system after each flight. The programme also includes installation procedures and to oversee monitoring on the inflight test [14], [16]. The main benefits of such SHM applications into aircraft maintenance programmes are expected to help save the airline industry time and money, especially if sensors are mounted in hard to reach areas and located widely throughout the entire aircraft.

1.2.6 Global versus local damage detection

Current damage detection methods using localised experimental techniques such as acoustic or ultrasonic, radiography and eddy current methods are quite common in the structure's maintenance . These techniques require the locality of the damage to be known and the relevant part for inspection to be readily accessible. Usually, damage like micro-cracks exist in a complex structure hidden in many other structural materials that need to be removed to access the area of damage. Removal of these components involve a substantial cost.

On the other hand, vibration-based damage detection is considered as a global experimental method of detecting damage which can provide monitoring over a wide structural area and does not necessarily require dismantling of the structure. The basic premise of vibration-based damage detection is that damage would modify the stiffness, mass or dissipation energy properties of the material, which in turn would change the global dynamic response properties of the system. This is the basis of vibration-based SHM that examines changes in the vibration characteristics of the structure [1], [17], [18]. Intuitively, an indication of damage present in the system is based on the changes of the natural frequency peaks displayed in the measured response. This is often referred to as an engineering judgement where it is used to reflect a sign of damage in the system.

In terms of damage detection and SHM on aircraft structures however, there has been little work currently done on incorporating the changes in fuel tank loading into a damage detection strategy. In the aerospace industry, it is reported that one of the main components that is consistently under routine inspection is the aircraft wing box [19][14]. It is one of the vital structural components of an aircraft which is concealed inside the wings. It acts as a rigid support for the wings and is connected to the main fuselage. The wing box is inevitably exposed to fatigue cracks especially around the rivets due to mass loading variations from the fuel tank and payload. To the author's knowledge, to date there has been no published work reported on the effect of fuel loading variability on damage detection and its sensitivity on commercial aircraft structures, particularly related to the wing structure. Some literature reviews on previous research related to aerospace structures

particularly using vibration based damage detection will be summarised in Chapter 2, where the attention has been focused on static load and dynamic using solid weights and temperature variations [20].

Despite the capability of the VBDD technique to detect damage based on the change in global dynamic properties (natural frequencies, stiffness and dissipation energy) of a particular system, damage typically remains a local phenomenon and can normally only be recognised on a higher frequency mode of the measured vibration response.

Another challenge from the vibration test in SHM is the fact that it always has to be a trade-off between having a very sensitive sensing system to capture the smallest possible damage and being able to distinguish this from other significant factors that can affect the vibration response. These other significant factors are the influence of the operational and environmental variations and the curse of dimensionalities. In VBDD, the number, type and location of sensors to mount on the structure are also important to ensure feasibility and reliability of the SHM system. If the sensing system is highly sensitive to damage, it is likely that the system will also tend to be highly sensitive to operational loading and environmental variations.

One of the main challenges in data analysis is related to the phenomenon usually known as ‘the curse of dimensionalities’. This refers to the need for having more data observations due to the increased number of data dimensions. This is a typical problem faced during the data processing stage. Subject to these limitations, the current study intends to explore the techniques in overcoming the challenge arising from the operational loading variations through the use of machine learning algorithms and the SPR shall be presented in the later chapters of this thesis.

1.3 Brief outline of the thesis

Chapter 2 provides a brief literature review on the subject of damage detection under the influence of operational loading and environmental variations through the SPR and machine learning algorithms. The focus is on aircraft structures in addition to the present research on the civil and engineering structures.

Chapter 3 covers the underlying mathematical equations associated with the machine learning algorithms that are used in this work. Some of the concepts pertaining to the algorithms are illustrated in the later chapters where their application in the current work will also be covered.

Chapter 4 details the preliminary experimental work for a vibration SHM which involves a replicated wing-box structure. The feature selection is demonstrated based on the effects of operational loading variations on the measured Frequency Response Function (FRF). Four damage severity classes referring to four levels of incremental cross-sectional cuts are introduced into a stringer of the wing box. This chapter also illustrates the implementation of the operational loading variations on the laboratory structure.

Chapter 5 highlights the centre of the current work. It describes the dimensional reduction and feature extraction analysis of the selected feature based on Principal Component Analysis (PCA). The visualisation of the multivariate data is achieved using PCA through establishing a baseline data set (training set) encompassing all loading conditions of the undamaged structure (referred as the baseline set). In addition to that, three more models of visualisations are considered by computing each data matrices for those models named as Principal Component model B, C and D. Each model has its own eigenvalues decomposition in addition to the primary baseline model (PC model A as the leading model corresponding to the undamaged set). These PC models are determined by various data arrangements relating five loading conditions to five structural health conditions (one undamaged state and four damaged states). Kernel PCA is then introduced as an attempt to improve the separation of the multivariate data and identify the damage severities. Kernel PCA depends on the inverse variance (called here as the parameter sigma) and the distance

matrix between two structural states. This study proposes a unique approach of computing the parameters for the kernel Gaussian function using a coloured scale distance matrix to present the change of data distance with the aim to ease the selection process of the inverse variance (sigma).

Chapter 6 extends the work from Chapter 5 by using the data variables obtained from the PCA to explore the use of clustering technique based on Gaussian Mixture Model (GMM). The main interest is to predict each of the data points that is most likely to fit the Gaussian distribution based on the model parameters. Nonlinear regression in the use of Artificial Neural Network presents a damage detection framework by introducing a step-target benchmark comprising of different levels to find the most suitable method in separating the data variables of various damage classes from various loading conditions.

Chapter 7 highlights the pinnacle of the current work by using a full-scale aircraft wing from a Jabiru airplane. PCA is applied on the vibration data set followed by kernel PCA. Q-statistic is demonstrated here in order to identify outliers. ANN is also applied on this data set and some studies on the contribution of sample size, number of dimensions and size of data groups (damaged and undamaged states) on the network performance are presented. In the Jabiru Wing (JW) experiment, two different size of loading variations are produced, one by smaller incremental loading and the other has a larger incremental loading. Prediction trees using bag-classification trees are also demonstrated in this chapter with the aim of describing the data set in a hierarchical order.

Chapter 8 concludes the finding in this thesis and the scope of future work is discussed.

Chapter 2

LITERATURE REVIEW

In this chapter, recent SHM literature reviews on the subject of data normalisation and damage detection under the influence of operational loading and environmental variations are reviewed with the primary interest on investigations of loading effects on damage detection and identification. Prior to that, brief descriptions of the current SHM approach with respect to discriminating the operational and environmental variability (OEV) in engineering structures such as bridges, buildings and mechanical structures are described. A number of Statistical Pattern Recognition (SPR) techniques in the perspective of data normalisation are discussed. The highlight in this chapter is to discuss some approaches used in discriminating OEVs in aircraft structures.

2.1 Definition of damage according to SHM

In the context of SHM, engineers or maintenance personnel are interested in the ability of a diagnostic method that can conduct damage assessment of the structure, in operation, in the earliest time possible. According to [1], in the study of damage identification in structural and mechanical systems, damage can be defined as intentional or unintentional changes to the material and/ or geometric properties of the systems. This includes changes to the boundary conditions and system connectivity which can adversely affect the present

or future performance of these systems. Damage can present in all engineering systems to some degree, and usually the system can still continue to perform its intended function, but with some reduced performance. In aerospace structures and for systems with a high safetyfactor, the damage can progress rapidly and be exaggerated under operational loading and extreme temperature conditions, and this can lead to a catastrophic failure if prompt maintenance action is not implemented a priori[21].

2.2 Damage-sensitive features generated from vibration-based approaches

The main concern of this thesis centres on vibration-based approaches in SHM. In most cases, damage would manifest itself as changes to the mass, stiffness and energy dissipation of the system. These structural changes can be identified through the shift of the frequency response function (FRF) peaks that decrease as the damage intensifies or as the loading increases. A broad review of vibration-based condition monitoring with particular attention on measuring the changes in vibration modal parameters as an indicator of damage presence in structures are provided by [22][17]. In reality, under the OEV presence, the changes in frequency measurement due to damage are often masked or can become unpredictable [17]. In this case, the changes in structural response due to damage might not be in accordance with the engineering judgment. Previous studies related to OEV are described further in the following sections.

2.3 The effects of OEV on SHM

In view of separating structural changes resulting from the effects of operational and environmental variability (OEV) from changes caused by true damage, SHM procedures for civil structures such as bridges seem to be at the frontier of SHM as compared to the aerospace industry. This may be due the strict FAA regulations set on safety related

technology that must be subjected to numerous testing and validation processes under realoperating conditions of an inflight aircraft before it can be approved for standard use [16]. Having aforementioned, the significant benefits accumulated from implementing condition-based maintenance using SHM drives Sandia and its team to push for SHM implementation in the real operating environment [14], [16].

It is recognised that sources of variability present in measured data from the dynamic response of a system can be associated with OEV conditions [3], [6]. The effects from the OEVs are often described as nonstationary due to the benign operational changes and environmental variations in the measured structural response of the structure. This nonstationary nature is generally revealed in the structural response as slowly varying trends or abrupt changes between regimes [3]. The varying operational conditions can arise from variable mass loading including changing fuel levels and changing payloads. Operational conditions also include changing operational speeds and changing of excitation sources. Varying environmental conditions can consist of thermal effects, wind-loading and moisture content. In this current work, operational loading conditions infer to varying fuel tank loading.

In the context of civil engineering structures, many authors have investigated the variability of modal properties with consideration to environmental and traffic loading effects [6][23]. It is found that the effects of the varying load on the dynamic properties are dependent on the type and magnitude of the bridges, whereby for middle and long-span bridges, the changes in measured natural frequencies are hardly detectable [24]. In a separate work by [25]Soyoz and Feng (2009) which is related to the mass loading and environmental effects on a concrete bridge structure, the findings revealed that the structure's first natural frequency varies in the order of ±10%. The significant variations are attributed by the changes in its structural mass caused by traffic loading and environmental effects [6].

It is also reported that the damping ratio of the structure can increase because of energy dissipation due to higher traffic loading [24]. A test performed on the I-40 Bridge revealed that variations of mass loading from traffic and varying environmental conditions due to thermal variations can alter the natural frequencies of the structure by a similar magnitude

that is caused by severe damage [1]. Temperature variations may also alter the material stiffness and the boundary conditions of a structure. Using experimental data obtained from Alamosa Canyon Bridge in New Mexico, Farrar et al. noted that the significant change in natural frequencies in a 24-hour time period is correlated to surface temperature differences across the structure but it is not correlated to the change in ambient temperature. Another vibration test on the I-40 Bridge, also in New Mexico, indicated that an increase in damage levels caused less effect on the natural frequency compared to the ambient temperature of the bridge [6].

If these OEV influences are neglected when developing SHM technology, it can falsely indicate that the structure is damaged and may result in unnecessary and costly inspection. This error is known as false positive indication of damage or Type I error. Similarly, variability arising from OEV can also result in false negative indication of damage or Type II error [6].

In the aerospace industry, this false information and consequently decision on structural health condition can potentially cause a catastrophe, posing a greater risk to loss of human lives and increasing expenditure for aircraft repair if the faults are misinterpreted as normal due to type II errors [6].

2.4 The OEV in aerospace structures

To deploy a robust SHM system for aircraft structures requires incorporation of changes caused by operational and environmental variability into the features selection. This is a vital aspect when adapting SHM for practical application, especially in the aerospace industry. At present, most of SHM applications for aircraft structures involve localised detection such as differential pressure-based method, removing the OEVs effects from the consideration. However, its main disadvantage as described earlier, is that it is often unable to detect the presence of damage over a large area of a structural component [14], [16].

Sierra-Pérez et al. reported a damage detection method on a wing section of an unmanned air vehicle (UAV) under variable load conditions by using a fibre-based detection as strain sensors. The authors proposed strain field pattern recognition based on principal component analysis (PCA) and its statistical damage indices using T^2 index and Q index to classify various damage severities under various loading conditions. The work effectively incorporated principal component analysis (PCA) when the structure is under variable variable operational load conditions in developing a pattern recognition technique using strain sensors [20].

Another SHM work related to environmental and loading variability is performed by Lim et al. (2011) on a metal fitting lug of an aircraft using impedance signals. This system is to detect bolt loosening conditions under varying temperature and external loading conditions using mounted piezoelectric materials. The performance of damage detection technique using kernel Principal Component Analysis (kPCA) has shown improvement on damage detectability especially those under temperature variability compared to linear PCA. However for detections of damage under static and dynamic loading, the damage is successfully detected with the use of linear PCA [26].

A damage propagation monitoring method based on Gaussian Mixture Model (GMM) using a guided wave and piezoelectric sensor based SHM technique is presented by [27]. GMM is constructed based on the GW features acquired under time-varying conditions when the structure is in a healthy state. The results indicated that the crack propagation under changing structural boundary conditions of an aircraft wing spar can be monitored reliably using the optimised GMM.

One might realise that damage detection via SHM alone will not be feasible without suitable machine learning algorithms that are capable of separating the effects of operational and environmental variations in the measured response. In the next section, some previous work on SHM data normalisation will be presented, with a bias towards the machine learning algorithms currently used in this work. Principal Component Analysis (PCA), kernel PCA, Gaussian Mixture Model (GMM), Artificial Neural Network (ANN), Bagged

Classification Tree (BGT) and Mahalanobis Squared Distance (MSD) are the selected machine learning algorithms relevant in the use of machine learning for this work.

2.5 Statistical Pattern Recognition via machine learning algorithms

This section highlights the techniques that can be used in pattern recognition through machine learning algorithms. One of the main challenges in pattern recognition and feature processing in SHM is the pattern representation and classification of data classes. Some of the techniques often used in SHM includes dimensional reduction tool such as PCA, factor analysis, singular value decomposition and the nonlinear approach, which includes kernel PCA, neural network etc. The taxonomy of dimensional reduction and classifiers technique are described in detail in [23].

Robust regression has been used in [28] as a means to discriminate the environmental and operational conditons from the experimental data acquired from Z24 and Tamar Bridges. Dervilis et al. (2015) has used the combination of minimum covariance determinant (MCD) via the FAST-MCD algorithm and least trimmed square (LTS) estimator via FAST-LTS algorithm for removing outliers and revealing the leverage points. This is done in order to establish a normal condition set clear from external outliers before applying a normalisation technique for damage detection under the influences of operational and environmental variations.

Tibaduiza et al. have purposed statistical reference model based on PCA, damage index (T^2 and Q-statistic) and a self-organizing map (SOM) as a classification tool by grouping in clusters data of similar characteristics to establish a baseline pattern for healthy and damage states. The vibration-based experiment is integrated with a multi-actuator system whereby the piezoelectric transducers (PZTs) reversibly function as sensors and actuators within an aluminium plate. The graphical technique has successfully classified all various damage in the baseline pattern model with better clustering accuracy by using Q-statistic index than

T^2 index and selected number of the first principal component score as input features for SOM [29].

Gharibnezhad et al. performs a robust PCA technique by developing a statistical principal component that is not sensitive to outliers by substituting the classical covariance with a robust covariance matrix estimator called ROBPCA. It uses a combined Projection Pursuit approach for dimensional reduction and a robust minimum covariance determinant (MCD) to provide accurate estimates in high dimensional data. It is observed that the damage clusters in the robust PCA are kept more distant from the healthy pattern which suggests better cluster separation and detection [30].

2.5.1 Data normalisation techniques in SHM

Data normalisation is an approach implemented to separate changes in damage-sensitive features caused by changing operational and environmental conditions from those truly caused by damage [2], [6] [24]. At present, there are two popular approaches to perform this task. The first approach involves measuring the parameters related to the OEVs. Then the features caused by the OEVs that correspond to the normal condition are parameterised as a function of the measured OEVs. The second approach, which is the one adopted in this study, involves a machine learning algorithm to develop models that assess the effects of changing OEVs on damage-sensitive features when the measure of OEV parameters are not available. This approach leads the way for data-based models where training data sets are obtained from the structure when it is undamaged under the influence of OEVs.

Data normalisation has been widely used in SHM when the effects of operational and environmental variations on damage detection and assessment are considered in the SHM process. Some of the data normalisation methods include neural network, novelty detection, dimensional reduction, regression modelling and principal component application [1], [2], [6], [24], [26], [31] [23].

Recently, switching of latent based model using Gaussian Tree Process (GTP) has been proposed by Worden et al. as a technique to remove measurement variations arising from

wide range of variabilities including from the environmental and operational variables before a damage can be inferred. This is performed by fitting a Bayesian GTP regression model on the second natural frequency of the structure. The results of the latent variables models which is named as the response surface model that shows a slow trend of changing (switching) data variation between data regimes [32] gives better data characterisation in the interest of understanding the data behaviour under environmental and loading variables.

There are numerous reported work on principal component analysis (PCA) based damage detection method under the presence of varying environmental and operational conditions. PCA has been addressed in great detail by [33] and in the perspective of pattern recognition and machine learning by [34]. PCA has been applied extensively in SHM in many ways such as being integrated with the Automatic Clustering Techniques based on Self-Organising Maps (SOM) [35], as a direct method to identify and distinguish changes due to damage from those due to environmental and operational loading conditions, and to localize damage through visualization of projected data. PCA also serves as a significant primary step for other methods or integrated with other algorithms to improve data classification and separation [27][36] [29]. A more recent application is to generate a statistical data-driven model using PCA to represent the structure and apply the model to detect damage in the structure [37].

PCA has been proven to solve many multivariate data analysis problems related to dimensional reduction and visualisation, damage detection and correlation analysis. The primary concern with PCA is that its restriction is on mapping only linear correlation among variables. Nonlinear PCA including Auto-Associative Neural Network (AANN) and kernel PCA (KPCA) can reveal the nonlinear correlations presented in data [36], [38] [4].

A comparative study using four machine learning algorithms which are PCA, AANN, KPCA and greedy KPCA (KPCA with reduced numbers of training set) been done by [39]. The goal is to detect structural damage of three-storey frame structure using VBDD under the presence of operational and environmental conditions. AN autoregressive model is used to extract damage-sensitive features from the time-series domain. The Mahalanobis Squared Distance (MSD) is computed and the result shows that KPCA as overall

outperforms AANN and PCA. KPCA and greedy KPCA (GKPCA) give the best performance in Type I error (false-positive indication of damage) whereas AANN gives the best performance in Type II error (false-negative indication of damage). KPCA shows its capability to balance between the type I and Type II errors.

Oh and Sohn have proposed nonlinear principal component analysis by employing kernel functions and solving simple eigenvalues in order to characterize the nonlinear relationship between extracted damage sensitive features and unmeasured environmental and operational parameters. The proposed method is based on unsupervised KPCA which is compared with the authors' previous work on AANN using the same time series data. Both analyses are performed to detect damage in the presence of temperature variation on a simplified model of a computer hard disk using a numerical model [4]. The authors conclude that proposed KPCA method resulted in an unsupervised support vector machine PCA results with similar performance in damage diagnosis to those obtained by AANN [40]. The finding also indicated that the advantages of KPCA over AANN are as follow:

- KPCA involves a simple formation of an eigenvalue problem instead of solving complex nonlinear optimization problem
- KPCA can avoid the overfitting problems by employing regularization
- the flexibility in computing multiple principal components without redesigning KPCA. With AANN, the number of nodes in the bottleneck layer should be known beforehand and the network parameters need to be optimized.

Sierra-Pérez et al. presented a new unfolding technique and scaling method to deal with variable load conditions of an unmanned aircraft wing section using strain measurement technique. The unfolding technique is performed by treating a group of data consisting of different load cases as a new single variable, so that the information acquired from all sensors are analysed simultaneously. The single data matrix can then be projected into the PCA model of a normal baseline model. Finally, a number of principal components are retained in order to calculate T^2 statistical index (measure the variation of a sample in the PCA) and Q statistical index, which measures the smaller change of the variation that are

not detected in the PCA. The T^2 and Q indices are able to detect some deviations between the baseline and the damage cases with the Q index showing higher sensitivity to damage [41].

Torres-Arrendondo et al. have compared three data-driven multivariate algorithms based on PCA, independent component analysis (ICA) and hierarchical non-linear PCA (h-NLPCA) to build a baseline pattern for damage detection and identification using graphical topology known as self-organizing map (SOM) [42]. The test data from healthy and damaged states are projected into the different models in order to produce input feature vectors of SOM including a cluster map and the U-matrix. Damage identification and cluster classification between damage and undamaged states are performed using three scores and squared prediction measures. The results show that PCA is able to classify and identify damage on the same performance as ICA and h-NLPCA algorithms but with lower computational cost and requires less time.

Sohn (2007) has presented a good review of data normalization methods using regression analysis and interpolation analysis. He reported that if the operational and environmental variations are unmeasurable, a modelling of the underlying relationship between damage-sensitive features and OEV sensitive features can be implemented if the signal changes caused by damage is orthogonal to the changes produced by OEVs. Other methods reported in his work are subspace-based identification method and novelty detection using Mahalanobis Squared Distance (MSD) [24].

Farrar & Worden (2013) [1] have summarised comprehensive main tehniques explored in data normalization as highlighted in the literature:

Experimental/ conventional method

Through experiment, the characteristics of normal condition and damaged condition features independent from OEV are identified before the features under the influence of OEV are examined and gathered for comparison.

Regression modelling

A modelling technique can be developed to predict the influence of OEV parameters on the measured response if the OEV parameters can be meassured [28].

Look-up tables

A method of recording features under relevant and complete OEV conditions while the structure is in an undamaged condition which is referred to as training phase. When a feature from a potential damage state is available, it can be compared to the data in the look-up table by using Euclidean distance metric or Mahalanobis squared-distance to calculate the feature distance from the undamaged condition.

Machine learning algorithm

Involves machine learning algorithm which is trained using the same feature vectors from an undamaged structure under influence of OEVs. In the test phase, each input vector is transformed into scalar feature known as damage index (DI) using the trained machine learning algorithm. Damage can be classified using the novelty detection approach applied to the DI. Auto-Associative Neural Networks (AANN), factor analysis, Mahalanobis squared-distance (MSD) and Singular value decomposition (SVD) algorithms can be implemented using this approach to perform data normalization.

AANN architecture contains three hidden layers comprising of the mapping, the bottleneck layer and the mapping layer. It basically performs the mapping and de-mapping, where it simply reproduces the network input. In the case where the measurement of OEV is not presented, the number of nodes in the bottleneck layer is chosen such that it represents the number of OEVs that cause the variability in the data feature. Firstly, the network is trained using the features from the undamaged condition under all OEVs. Then the network is tested with features arising from damaged conditions also under the exposure of the

equilvalent OEVs. It is anticipated that the prediction error of the network will grow when it is tested with the features that come from damaged conditions [1], [6].

Projection method

It is an intelligent feature selection by selecting damage-sensitive features but insensitive to OEV. Principal component analysis (PCA) is one of the approach based on the projection method that reduces the feature dimensions and retain the most variance present in the data set. Features containing evidence of EOVs but insensitive to damage can also be projected in lower dimensional subspace with most variance present in the data set.

The idea of cointegration comes when it is realized that damage sensitivity and damage evidence may be lost in different principal components, as thought damage manifest better in longer time scales i.e nonstationary time series than dynamics of the system.

Cross et al. (2012) has compared three different approaches in finding damage-sensitive features that are insensitive to environmetal variations:

- using original features that show sensitivity to damage but none to environmental variations with outlier analysis is the potential technique,
- using minor principal component analysis by projecting Lamb wave data, the temperature dependancy data can be removed which is initilaly proposed by Manson [39] and
- by using cointegration for creating damage detector which is insensitive to environmental variations.

All these approaches have shown encouraging results especially for cointegration and PCA that are able to create features that remain unchanged to temperature variations but very sensitive to damage [5].

Manson [43] used PCA and MSD novelty detection for damage identification by using a subset of features which are more sensitive to damage yet insensitive to temperatature variations. His approach focuses on isolating those features that are more sensitive to damage than the environmental variations. Using Lamb wave propogation data on a

composite plate, the feature projection produced by minor prinicipal components corresponding to the smallest singular values, has successfully separated the primary environmental variations from the desirable damage features [5].

There are numerous studies introducing Gaussian Mixture Model (GMM) to the field of SHM to model the uncertainty and nonlinearity of data signals under time-varying condition [44]–[49]. Qiu et all [27] proposed a damage propagation monitoring method based on an improved Gaussian Mixture Model (GMM) using a Guided Wave (GW) and piezoelectric sensors. With this method, a baseline GMM is first constructed based on GW features using PCA when the structure is at normal state. When a new PCA feature is obtained during the on-line monitoring, the GMM is updated using dynamic learning and split-merge of Gaussian components. The results indicate that crack propagation on a wing spar can be monitored reliabily without a structural model and prior knowledge of damage and time varying conditions.

Probability decision tree model is used to build ittertively subdividing the surface of the structure in such a way that it takes into account the actual geometry structure. The results show an advantage in terms of hierarchical classification and considers the true geometry of the structure [50]. A multi-class support vector machine classifier utilising binary decision tree is presented by Madzarov et al. [51].

Chapter 3

FEATURES PROCESSING TECHNIQUES

To distinguish between normal condition and damage states, features processing techniques are performed on measured data. This may be achieved using different machine learning algorithms, the choice of which may depends in which its usage depends on the goal of the analysis, either classification, regression or novelty damage detection. Features can be thought of as data derived from the acquired measurements which the judgement is in view of their capability to distinguish between different structural conditions.

The foundation of this work relies on the statistical analysis and probabilistic model that are used to analyse the Frequency Response Function (FRF) data. The FRF data constitutes a number of spectral lines which in this work are referred to as spectral variables. In the current context, the differences in the set of spectral variables are as a result of varying operational loading and damage state parameters.

In statistical analysis, the mean and variance-covariance values are the essential parameters if a statistical model is to be established for distinguishing between the undamaged and damaged state. This chapter describes mathematical algorithms for Principal Component Analysis (PCA), kernel PCA as a generalised nonlinear form of PCA, novelty detection indices and Gaussian Mixture Model (GMM). The algorithms described here are for the purposes of data visualisation and classification, data separation, damage detection and clustering technique respectively.

3.1 Principal Component Analysis (PCA)

In many scientific applications, PCA is a well known statistical tool for being an effective method for data dimensionality reduction and the best known technique for multivariate analysis. Pearson (1901) first introduced it and later by Hotelling (1933) who developed it further. Only after the advent of electronic computers, PCA became more common in many statistical packages.

PCA is a linear projection technique used for feature extraction, data compression, visualization and interpretation. PCA finds combinations of variables or components that describe fundamental trends and patterns in a data set. PCA is concerned with describing the variance-covariance structure through several linear combinations of the original variables. The formulation of PCA is relatively straightforward. Consider a data set, $X_{n\,x\,m}$, with n observations and m measured variables or dimensions. The first step in applying PCA is to standardise the data matrix X as PCA is scale sensitive. The mean values of each dimensions removed and all variables are made to have equal variance. As a result, the trends or patterns on the observations and their standard deviations are removed from the data set. The covariance matrix C_X is calculated using the standardised variables as:

$$C_X = \frac{1}{n-1} X^T X \tag{3.1}$$

This covariance matrix measures the degree of linear relationship between the data variables. PCA is computed by finding the eigenvectors and eigenvalues of the covariance matrix which is defined as follows:

$$C_X \mathrm{v} = \mathrm{v}\Lambda \tag{3.2}$$

Eigenvectors of C_X are the columns of v and the eigenvalues of C_X are the diagonal elements of Λ (the off-diagonal elements of Λ are zero). The eigenvector with the highest eigenvalues becomes the first principal component which has the largest variance in the dataset. The vectors with significant variance contain the most important underlying pattern in the data and consequently contain the largest quantity of information. Therefore, these vectors are called the principal components of the data set. Ordering the eigenvectors by eigenvalues from the highest to the lowest, arranges the components in order of significance and produce a new matrix P. The columns of matrix P are named as loading vectors and describe the linear combinations of the original variables that represent each principal component.

The aim is to reduce the number of dimensions and consequently, the components with smaller eigenvalues can be eliminated. This results in information loss but because the information is insignificant, it can typically be disregarded. If only k-first eigenvectors are selected, the final data set will therefore be k-dimensional. The projected matrix T or the transformed matrix in the new space is defined by equation (3.3), and the transformed matrix T (or score matrix) can be projected back into k-dimensional observation by equation (3.4). The difference between the full dimension and reduced k-dimension between X and loss of information or named as the residual matrix E is given by equation (3.5). The formulation of PCA is stated by equation (3.6).

$$\mathrm{T} = \mathrm{XP} \tag{3.3}$$

$$\hat{\mathrm{X}}=\mathrm{TP}^{\mathrm{T}} \tag{3.4}$$

$$\mathrm{X}=\hat{\mathrm{X}}+\mathrm{E} \tag{3.5}$$

$$X=TP^{T}+E \tag{3.6}$$

3.2 Kernel PCA for nonlinear feature extraction

Kernel PCA is a nonlinear generalisation of PCA in such a way that it is performing PCA in feature space of arbitrary large dimensionality. The key idea of performing kernel PCA (kPCA) on the data is to extract some features related to some nonlinearity that often hide in high dimensionality. The general idea is that this can be performed by mapping the input feature into the higher dimensional feature space via a nonlinear mapping function. Principal components analysis can then be implicitly performed in the new high dimensional feature space by using a substitution by the integral operator of kernel function [52].

Kernel PCA is a relatively efficient method for performing nonlinear PCA simply by substituting a nonlinear kernel function that uses dot products in feature space. This method works in a sense that it only requires an eigenvalue solution, without having to work implicitly in the higher dimensional feature space.

The mapping of input data into the feature space is performed by simply solving a dot product of input data *k(x, y)* via nonlinear kernel Gaussian function in an eigenvalue solution [52].

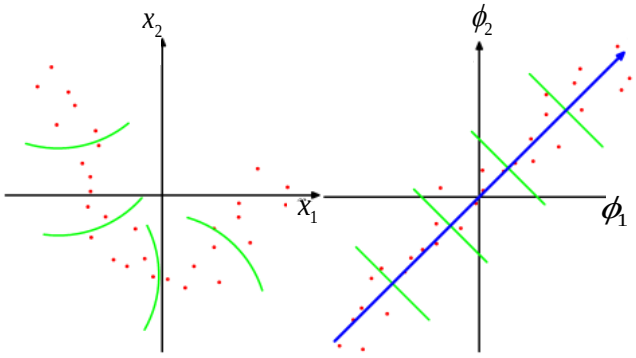


Figure 3.1: The basic idea of kernel PCA by performing PCA in the feature space (right-hand plot) instead of the input space (left-hand plot) [34]

Figure 3.1 illustrates the basic idea of kernel PCA. First, a data set in the original data space (left-hand plot) is projected by a nonlinear transformation $\phi(x)$ into a feature space (right-hand plot). PCA is then performed in the feature space to obtain the first principal component (PC) denoted by v_1 (shown in blue line). The green lines in feature space represent the linear projections onto the first PC which correspond to the nonlinear projections in the original input space. As an advantage, by performing kernel PCA, there is no need to carry out the transformation or mapping into the feature space. The complicated mapping is substituted by computing only the kernel function which consists of a dot product k(x,y) in input space.

3.2.1 The principle

One of the main concerns about a linear (standard) PCA is the capability to extract nonlinear features especially for high dimensional data. In practical case, the data normally do not follow linear relationships due to measurement variation and OEV. To overcome these issues, a nonlinear form of PCA via kernel function is utilised in this work.

KPCA uses an efficient method to extract nonlinear features by first mapping the input features into higher dimensional feature space by performing a dot product in the input space, and then solving the eigenvalues problem without having to work explicitly in the feature space [34]. Essentially the principle follows Cover's theory, which states that given training data that is not linearly separable, it can vary linearly and be made linearly separable by projecting it into a higher dimensional space via some nonlinear transformation using a high dimensional space mapping [52][53][54].

3.2.2 Constructing the kernel matrix

In this section, the steps of performing standard PCA using kernel methods are presented. Note again that this method avoids one from performing standard PCA in the new feature space which can be extremely costly and inefficient [52], [55]. Generally, for a nonlinear

transformation $\phi(x)$ from the original-dimensional feature space D to an M-dimensional feature space, the case $M \gg D$ must be satisfied. First, assume that the projected new features have zero mean, defined as

$$\frac{1}{N}\sum_{i=1}^{N}\phi(x_i) = 0 \tag{3.7}$$

The covariance matrix of the projected features with dimensional size $M \times M$ is

$$C = \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^T \tag{3.8}$$

Its eigenvalues and eigenvectors are indicated by

$$\mathrm{C}v_k = \lambda_k v_k \tag{3.9}$$

where k=1, 2, …, M. From equation (3.8) and equation (3.9), it gives

$$\sum_{i=1}^{N}\phi(\mathrm{x_i})\left\{\phi(\mathrm{x_i})^T \mathrm{v_k}\right\} = \lambda_\mathrm{k}\mathrm{v_k} \tag{3.10}$$

which can be restated as

$$\mathrm{v_k} = \sum_{i=1}^{N} a_{ki}\phi(\mathrm{x_i}) \tag{3.11}$$

Now by substituting $\mathrm{v_k}$ in equation (3.10) with equation (3.11), it yields

$$\frac{1}{N}\sum_{i=1}^{N}\phi(\mathrm{x}_i)\phi(\mathrm{x}_i)^T\sum_{i=1}^{N}a_{kj}\phi(\mathrm{x}_j) = \lambda_\mathrm{k}\sum_{i=1}^{N}a_{ki}\phi(\mathrm{x}_i) \tag{3.12}$$

Kernel function can be defined as

$$\kappa(\mathrm{x}_i, \mathrm{x}_j) = \phi(\mathrm{x}_i)^T\phi(\mathrm{x}_j) \tag{3.13}$$

Multiply both sides of equation (3.12) by $\phi(\mathrm{x}_l)^T$, gives

$$\frac{1}{N}\sum_{i=1}^{N}\kappa(\mathrm{x}_l,\mathrm{x}_i)\sum_{j=1}^{N}a_{kj}\kappa(\mathrm{x}_i,\mathrm{x}_j)=\lambda_k\sum_{i=1}^{N}a_{ki}\kappa(\mathrm{x}_l,\mathrm{x}_i) \tag{3.14}$$

Using the matrix notation

$$\mathrm{K}^2 a_k = \lambda_k N\mathrm{K}a_k \tag{3.15}$$

where

$$\mathrm{K}_{i,j}=\kappa(\mathrm{x}_i,\mathrm{x}_j), \tag{3.16}$$

and a_k is the N-dimensional column vector of a_{ki}

$$a_k=\left[a_{k1}a_{k2}...a_{kN}\right]^T. \tag{3.17}$$

solving a_k by

$$\mathrm{K}a_k=\lambda_k N a_k \tag{3.18}$$

then the resulting kernel principal components can be solved using

$$\mathrm{y}_k(\mathrm{x})=\phi(\mathrm{x})^T\mathrm{v}_k=\sum_{i=1}^{N}\mathrm{a}_{ki}\kappa(\mathrm{x},\mathrm{x}_i). \tag{3.19}$$

In general, the projected data set $\{\phi(\mathrm{x}_i)\}$ usually does not have zero mean. As the technique tends to avoid working directly in the feature space, the data cannot simply be subtracted off the mean. The algorithm can be formulated in terms of the kernel function using the Gram matrix representation, given by

$$\tilde{\mathrm{K}}=\mathrm{K}-1_N\mathrm{K}-\mathrm{K}1_N+1_N\mathrm{K}1_N \tag{3.20}$$

where 1_N equals to N-by-N matrix with all elements equal to $1/N$ [34].

As stated earlier, the power of kernel methods is that the computation is not explicitly done in the feature space but by directly constructing kernel matrix from the training data set $\{x_i\}$. A kernel PCA is a nonlinear generalisation of PCA which can recover the standard PCA if using the kernel $\kappa(x, y) = (x \cdot y)$. Some commonly used kernels are the polynomial kernel and sigmoid (hyperbolic tangent) kernel. For this study, Gaussian kernel has shown to be most effective in separating the variables of different damage severities and therefore is used to solve the current problem. Gaussian kernel is given by

$$k(\mathrm{x},\mathrm{y}) = \exp\left(\frac{-\|\mathrm{x}-\mathrm{y}\|^2}{2\sigma^2}\right) \tag{3.21}$$

where σ^2 is the inverse variance in the kernel of which value depends on the data set variation.

3.3 Damage Identification Indices

The transformed matrix T obtained from the Equation 3.3 is also known as score matrix in which its columns consist of score vectors t_i. Each of the vectors t_i is associated with the corresponding principal components PC_i. PCA can also be used to indicate damage severities or abnormalities in a system. Two well known statistics are commonly used to measure the indices: the T^2-statistic (also known as T^2- Hotelling's or D-statistic) and Q-statistic (square prediction error (SPE)-statistic). These indices can be used to represent variability of the projection in the new space (with T^2-statistic) or in the residual subspace (using Q-statistic index). These methods are based on the assumption that its underlying behaviour follows approximately a multivariate normal distribution which originates from the central limit theorem [34].

3.3.1 T^2-statistic

The Hotelling's T^2-statistic is a generalisation of Student's t-statistic commonly used in multivariate hypothesis testing. It analyses the variation of each sample within the space of PCA model. High T^2 indices indicate high variation within the normal and damage data sets of which is desired where the aim is to distinguishing the data produced by normal condition from those by damage conditions. It can be obtained by using the squared score value of the test data that has been normalised using matrix Λ of the test data as illustrated in the following equation

$$T_i^2 = t_i \Lambda^{-1} t_i^T \tag{3.22}$$

Where t_i is the i th row vector of the score matrix T_i, which represents the new principal component projections. Both score matrix and principal components are correlated as defined by equation (3.3). Note that T^2 in the reduced space corresponds to the Mahalanobis distance in the reduced space.

To compute the T^2 index in the reduced space corresponding to the principal component projections in which its focus is to compare the degree of variability between the undamaged and damaged states, the following equation is applied by computing the score matrix obtained from the PCA

$$T_i^2 = \left|t_i - \bar{T}_N\right| \times \left|\sum T_N\right|^{-1} \times \left|t_i - \bar{T}_N\right|^T \tag{3.23}$$

$\bar{T}_N$ is the mean of the score matrix associated to the undamaged condition and $\sum T_N$ is the covariance of the similar matrix. Each of the score vectors t_i corresponds to a test data set characterised by different levels of damage severity. To study the damage severities variability, each score vector of a damage state is compared with the score values from the undamaged condition using Equation (3.25). Equation 3.25 is equivalent to Mahalanobis distance equation calculated in terms of the score values in the PCA setting. It is used in the present study to calculate and compare the data variability with the focus of examining

damage severities separation in the reduced dimensional space. In other words, T^2-statistic is a measure of the variation of each sample within the PCA model.

3.3.2 Q-statistic

Q-statistic reflects the combination value of unused/ unselected principal components in regard to the selected principal components used in the T^2-statistic. It is desirable to achieve a very small Q-statistic index to reflect the optimal projection of the PCA. It is a measure of the difference or residual between an original sample space and its projection in a reduced space. The Q-statistic of the *i* th observation vector x_i is stated as

$$Q_i = \tilde{x}_i \tilde{x}_i^{\mathrm{T}} = x_i(\mathrm{I\text{-}PP}^{\mathrm{T}})x_i^{\mathrm{T}} \tag{3.24}$$

$\tilde{x}_i$ denotes its projection into the residual subspace. In relation to the PCA, Q-statistic is more sensitive than T^2-statistic because Q represents a very small change in the system characteristic and T^2-statistic has larger values and higher variance in the feature space. It is noted that any change in the T^2-statistic can be detected on the score plots for the relevant significant principal components. However, for Q-statistic, the values cannot be recognised from the principal components projections plots. This is because Q index relates to the residuals that are not included in the score plot. It can be obtained by plotting Q-statistic plot separately from T^2-statistic.

3.4 Outlier analysis

In the statistical literature, it is common to find the novelty detection being considered in the perspective of an outlier analysis. The fundamental idea is to compute the discordancy values associated to the data and then compare the discordancy with a predetermined threshold value. The data is simply inferred as discordant or novel if the measure exceeds the threshold.

The current study involves a multivariate data consisting of multi-dimensional variables. Multi-dimensional variables refer to the high number of spectral variables corresponding to the changing of frequency series represented as the system's dynamic characteristics. The discordancy test of *n* observations in *m* variables can be represented as *n* points in an *m*-dimensional feature space. In data with higher number of variables such as this study, the detection of outliers becomes more challenging as the outliers have 'more room' to hide [1].

In this work, the discordancy test is computed using T^2-statistic based on score matrix T (as stated in Equation 3.3) obtained from the linear PCA. The discordancy test for multivariate data is related to the Mahalanobis squared distance *D* as defined by

$$D_{\varsigma} = (\{x\}_{\varsigma} - \{\bar{x}\})[\Sigma]^{-1}(\{x\}_{\varsigma} - \{\bar{x}\})^{T} \tag{3.25}$$

Where $\{x\}_{\varsigma}$ is the potential outlier, $\{\bar{x}\}$ is the mean of the sample and Σ is the sample covariance matrix. In the context of pattern recognition, the mean and covariance matrix are often computed from the training data set (undamaged state).

3.5 Calculation of Thresholds

A threshold value is necessary to conclude an observation as an outlier or inlier by comparing the discordancy value against the threshold value. The threshold value depends on both the number of observations and the number of dimensions of the problem. For univariate case, the calculation is based on a Gaussian distribution applied on a normal condition case with only 5% of the measurement has 1.96 standard deviations from the mean. This study involves multivariate analysis where the Gaussian distribution model is assumed for the data density. A Monte Carlo method is used to generate the threshold value which follows the steps given by [56]. The procedure for this numerical method is summarised as below:

- Randomly generate a number from a zero mean and a unity standard deviation in which it constructs an $n \times p$ matrix (*n* is number of observations and *p* is number of dimensions)
- The Mahalanobis squared-distance (MSD) using equation (10) is then computed for all the rows with the mean and the covariance matrix are either inclusive or exclusive measures and the largest value is stored.
- The process is repeated for a minimum of 1000 trials in which the array consisting of all the largest MSD values are then ordered in terms of magnitude and stored.
- The critical values are calculated based on 5% and 1% test of discordancy values for a *p*-dimensional sample of *n* observations using MSD in the array above in which 5% and 1% of the trials occur.

The exclusive measures mean that the data produced by damaged condition is not present in the the training data which is typically consisting data set from undamaged condition.

3.6 Gaussian Mixture Models (GMM)

GMM is a probabilistic model that adopts all of its data points that are assumed to be generated by combinations of Gaussian distributions [34]. In GMM, all data points are assumed as Gaussian distributed. The goal is to find the probability of each Gaussian generating each particular data point. Mixture of Gaussian models result in a useful probability distribution model that can approximate any densities models with arbitrary accuracy by specifying a sufficient number of components or clusters and using initial means and covariances.

The method begins with a mixture of several components or clusters indexed by *c*, each cluster *c* is described by a Gaussian density $\mathcal{N}(\mathrm{x}|\mu_c, \Sigma_c)$ known as a component of the mixture. Each cluster is described by the parameters mean μ_c , covariance Σ_c and a mixing coefficient π_c. Under the mixture of Gaussian distribution, the superposition of *C* Gaussians

densities is then defined by the mixing coefficients π_c of these individual components. The mixture of Gaussians can be described as

$$p(\mathrm{x}) = \sum_{c=1}^{C} \pi_c \mathcal{N}(\mathrm{x} | \mu_c, \Sigma_c) \tag{3.26}$$

$$\sum_{c=1}^{C} \pi_c = 1 \tag{3.27}$$

The second equation is a constraint for the weights that ensures all the Gaussians weights sum up to 1, in order to satisfy the requirement for them to be considered as probabilities.

Next, discrete latent variables *z* are introduced into the GMM formulation to provide a deeper insight into this Gaussians mixture. This will motivate a very useful and essential Expectation Maximisation algorithm to find maximum likelihood solutions for the Gaussians generating the data point by introducing discrete latent variables *z* into formulation of Gaussian mixtures.

The joint distribution *p(*x,*z)* is then defined in terms of a marginal distribution *p(z)* and conditional distribution *p(*x|*z),* which is illustrated in the graphical model in Figure 3.2. The marginal distribution over *z* can be specified in terms of the mixing coefficients π_c, such that

$$p(z_c = 1) = \pi_c \tag{3.28}$$

where the parameters { π_c }must satisfy

$$0 \leq \pi_c \leq 1 \tag{3.29}$$

To be valid probabilities, it needs to satisfy

$$\sum_{c=1}^{C} \pi_c = 1 \tag{3.30}$$

For the conditional distribution of x given a particular *z* is a Gaussian, it can be written as

$$p(\mathrm{x}|z_c = 1) = \mathcal{N}(\mathrm{x}|\mu_c, \Sigma_c) \tag{3.31}$$

The data x is modelled jointly with an additional variable *z* that cannot be observed (hidden). The unknown value of *z* helps explain patterns in the values of x.

The joint distribution is provided $p(z)p(\mathrm{x}|z)$ and the marginal distribution of x is then found by summing the joint distribution over all possible states of *z* that gives

$$p(\mathrm{x}) = \sum_z p(z)\, p(\mathrm{x}|z) = \sum_{c=1}^{C} \pi_c \mathcal{N}(\mathrm{x}|\mu_c, \Sigma_c) \tag{3.32}$$

The marginal distribution of x is a Gaussian mixture equivalent to the form in equation (3.26). Noticed that in equation (3.32), the latent variable *z* is described in terms of the observed observations $\mathrm{x}_1, \ldots, \mathrm{x}_N$.

The advantage of describing the GMM in terms of the joint distribution $p(\mathrm{x},z)$ instead of only $p(\mathrm{x})$ is that it simplifies and motivates the EM algorithm in the process to obtain the maximum likelihood solution for the parameters.

Another quantity that plays a significant role is the conditional probability of *z* given x, $p(z|\mathrm{x})$. $r(z_c)$ will be used to denote $\mathrm{p}(z_c=1|\mathrm{x})$ of which value can be obtained using Bayes' theorem:

$$\begin{aligned} r(z_c) \equiv p(z_c = 1|\mathrm{x}) &= \frac{p(z_c = 1)\, p(\mathrm{x}|z_c = 1)}{\sum_{j=1}^{C} p(z_j = 1)\, p(\mathrm{x}|z_j = 1)} \\ &= \frac{\pi_c \mathcal{N}(\mathrm{x}|\mu_c, \Sigma_c)}{\sum_{j=1}^{C} \pi_j \mathcal{N}(\mathrm{x}|\mu_j, \Sigma_j)} \end{aligned} \tag{3.33}$$

Initially, π_c is viewed as the prior probability of $z_c = 1$ and the quantity $r(z_c)$ becomes the corresponding posterior probability once x is observed. $r(z_c)$ is viewed as the responsibility that component *c* takes for 'explaining' the observation of x. The responsibility also denotes

the posterior probability for each component in the mixture distribution from which the data set is generated by evaluating the responsibility for each data point.

Suppose there is a data set of observations $\{x_1, \ldots, x_N\}$ and the objective is to model the data using GMM. If the data points are drawn independently from the distribution, the GMM can be described using the graphical representation in Figure 3.2. From equation (3.26), the log of the likelihood function can be stated by

$$\ln p(\mathrm{X}|\pi,\mu,\Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{c=1}^{C} \pi_c \mathcal{N}(\mathrm{x}_n|\mu_c,\Sigma_c) \right\} \tag{3.34}$$

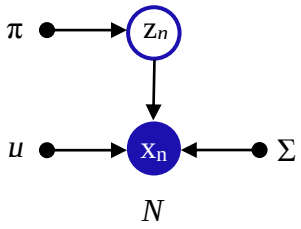


Figure 3.2: Graphical representation of a GMM for an independently drawn N data points $\{x_n\}$ corresponding to latent points $\{z_n\}$ where n=1,…,N

3.7 Artificial Neural Network functions

An alternative approach for models that use linear combinations of fixed basis functions is the Artificial Neural Network (ANN). ANN enables those basis functions to be adaptive and adjustable during the network training. In the context of pattern recognition paradigm, feed-forward neural network is also commonly known as multilayer perceptron (MLP). By its name, it illustrates that the model consists of multiple layers of models with discontinuous nonlinearities.

In Chapter 6, the wing box data applied with the ANN is presented. In this section, the brief principles of ANNs will be described. The functional form of the network model, including the parameters of the network comprises a linear combination of adjustable bias functions $\phi_j(\mathrm{x})$. It can be given as

$$y(\mathrm{x},\mathrm{w}) = f\left(\sum_{j=1}^{M} w_j \phi_j(\mathrm{x})\right) \tag{3.35}$$

where f(·) is an activation function. The parameter weight w_j and the basis function are made adjustable during the training via the differentiable f function.

To construct the network model, firstly the activation a_j is formed from M linear combinations of the input variables $x_1, \ldots, x_D$ specified as

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \tag{3.36}$$

where j=1, …, M and the superscript (1) denotes the corresponding parameters weights and biases located in the first layer. Each of the activations is transformed using a differentiable function $h(\cdot)$ which is nonlinear to produce

$$z_j = h\left(a_j\right) \tag{3.37}$$

This transformation follows within hidden units *z*. Generally, the nonlinear function $h(\cdot)$ is sigmoidal functions such as logistic sigmoid or the 'tanh' function. The transformation, which is corresponding to the hidden units, is illustrated in Figure 3.3. In the next layer, these values are again linearly combined to produce the activations for the outputs. The inputs are now z_j's, which previously are the output obtained from the hidden units transformation as given by

$$a_k = \sum_{i=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \tag{3.38}$$

where k=1, …, K are the series of outputs. This activation occurs in the second layer of the network. Finally, the output unit activations are transformed using a suitable function and produce the set of the network outputs y_k.

The choice of the activation function within the output units depends whether the problem is regression or classification based solution. In case of regression (as for the study), the function is simply the identity, that is $y_k = a_k$.

$$y_k = \sigma(a_k) \tag{3.39}$$

For binary classification problems, a logistic sigmoid function as given below is used as to transform each output unit activation.

$$\sigma(a) = \frac{1}{1+\exp(-a)} \tag{3.40}$$

Finally, the overall stages, including all network function can be combined to give

$$y_k(\mathrm{x},\mathrm{w}) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right) \tag{3.41}$$

In this way, the neural network simply comprises a nonlinear function, h with corresponding input variables $\{x_i\}$ to a set of output variables $\{y_k\}$ in respect to adjustable parameters w. The process of evaluating the overall network outputs can be represented by Figure 3.3 which is a forward neural network, whereby the information propagates in forward direction.

It is noted that the bias parameters can be absorbed into the set of weight parameters by defining x_0=1, which results in simplified form, to give

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} x_i \tag{3.42}$$

$$y_k(\mathrm{x},\mathrm{w}) = \sigma\left(\sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i\right)\right) \tag{3.43}$$

It is also noted that the neural network functions that are differentiable with respect to the network parameters *w* are the central role in neural network training [34].

As shown in Figure 3.3, a neuron in a hidden layer (hidden node) computes a weighted summation over all inputs in which the result's threshold is computed to produce a binary output *y* which is either -1 or +1. [57].

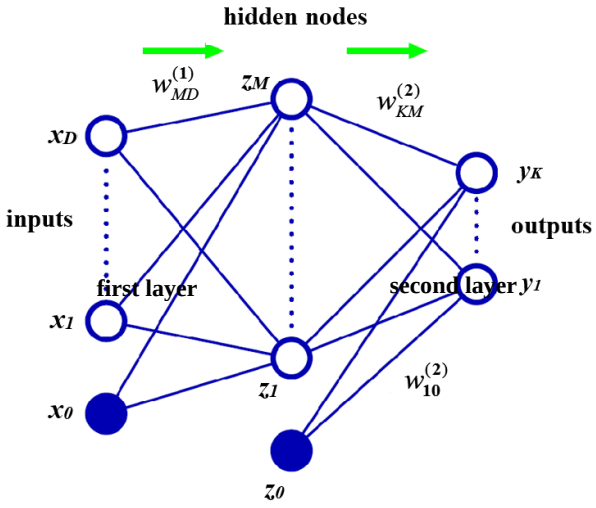


Figure 3.2: A Network diagram consisting of two-layer network with one hidden layer and one output layer. The white circles denote the nodes while the blue circles are the bias [34].

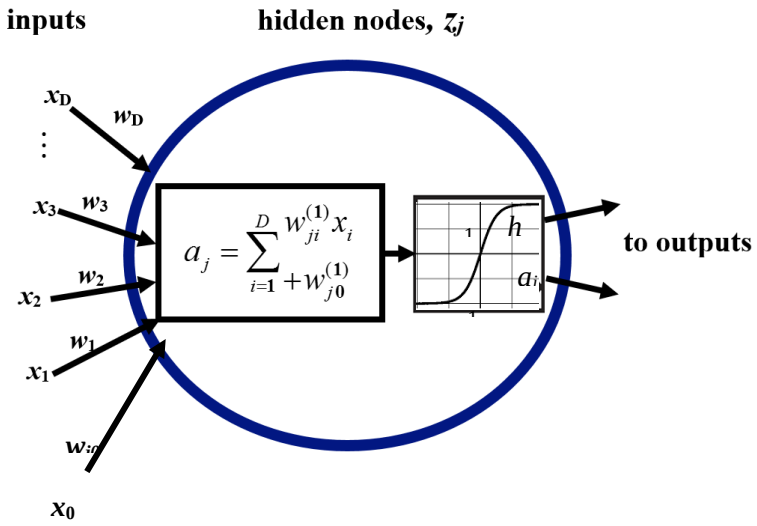


Figure 3.3: Schematic diagram of an artificial neuron (node) in a hidden layer.

Chapter 4

THE EXPERIMENTAL TEST OF PSEUDO FUEL TANK-WING BOX STRUCTURE

The work in this chapter presents the first part of the study that centres around an experimental work of a replicated wing box structure based on a vibration based damage detection (VBDD). This experiment involves a systematic procedure of filling the wing box tank with some amount of liquid loading in sequential order in order to introduce some operational loading conditions on the wing box structure.

The aim of the experiment is to develop a structural system of different structural conditions that undergoes some incremental loading before performing VBDD. The operational loading variations in this experiment is measured in terms of the FRF domain. The measured FRF data will be used to investigate the dynamic behaviour of the wing box as it is subjected to various operational loading variations. Therefore, the aim is to acquire some understanding on the dynamic characteristics of this wing-box structure under the influence of operational loading variations for the purpose of damage detection.

4.1 Overview

The core findings of this experimental study are to provide some understandings on the structural vibration of an aircraft structure under the controlled environment when variable mass loading is introduced in the structure in sequential order under different structural conditions. Structural conditions here refer to the undamaged and several damaged conditions of the structure when a damage in a form of saw-cut is introduced in the wing box stiffener.

The changing of liquid loading inside the tank is referred as the operational loading variability. This is done in the experiment procedure in an attempt to represent a scenario of a changing fuel load or a pseudo fuel load experienced by an aircraft structure.

The first part of this chapter highlights the experimental tasks regarding the procedures and configurations of the filling and extraction of liquid water into/from the wing box tank. The second part presents the analysis of the FRF data obtained from the experiment in the interest of finding the possible damage-sensitive features. By using visualisation plot obtained in the FRF, the study attempts to identify the influence of the operational loading variables on the fundamental frequency changes.

4.2 The experimental highlights

The main idea of implementing this experiment is to study the correlation between mass loadings and damage variations as produced in the Frequency Response Function (FRF). To achieve this goal, a novel experimental work is developed in the aim to simulate a scenario of an aircraft wing box loaded with various mass of liquid water that is done in systematic and under supervised manner. This is important and interesting in order to get a first view and understanding of the vibration data characteristics due to this water mass loading onto the wing box structure. The motivation behind this experiment, is to acquire

a vibration data set generated by means of sensible operational loading conditions, motivated by the aircraft fuel loading onto the aircraft wing-box structure.

As stated previously, damage variations are introduced in the structure and at the same time, the loadings are consistently changing systematically. In this initial section of the chapter, the objective is to highlight the experimental strategy in order to produce a good and useful vibration signal with minimum noise. With its good practise, the generated data set can provide sufficient information and indirectly can improve the structural damage assessment task undertaken in future chapters. Some preliminary work related to signal processing is attached in the appendix section selection procedure for signal processing aspects undertaken in the experimental test are presented in the appendix chapter.

To acquire the data set, a vibrational test using a replicated wing-box structure attached with two liquid tanks is performed by using a random signal excitation test. The experiment is performed after a comprehensive study on the suitable signal processing aspect of the structure with the additional of water loading effects.

4.2.1 Description of structure

The structure used is a stiffened aluminium panel to represent an aircraft wing box and it is a similar structure used in [58] and [38] indicated in Figure 4.1. The top sheet of the wing box is a 750 X 500 X 3 mm aluminium sheet. The structure is stiffened by two ribs of length of C-channel riveted to the shorter edges and two stiffening stiffeners composed of angle section which are bolted along the length of the sheet. Free-free boundary conditions are approximated by suspending the wing box from a substantial frame using springs and nylon lines of heavy-duty type attached at the corners of the top sheet. The structural weight of the wing box is around 6.464 kg.

The tanks that filled up with water successively is made of two transparent rectangular prisms. They are made from clear acrylic Perspex sheet, bonded onto the wing-box top plate using strong adhesive (see Figure 9). The tanks are 604 mm X 155 mm X 25 mm in size. Each tank is attached with small inlet for filling of liquid and one open/close outlet

for discharging the liquid. The inlets and outlets are 5 mm in diameter size. The tank can be filled up to maximum volume of 2.34 L.

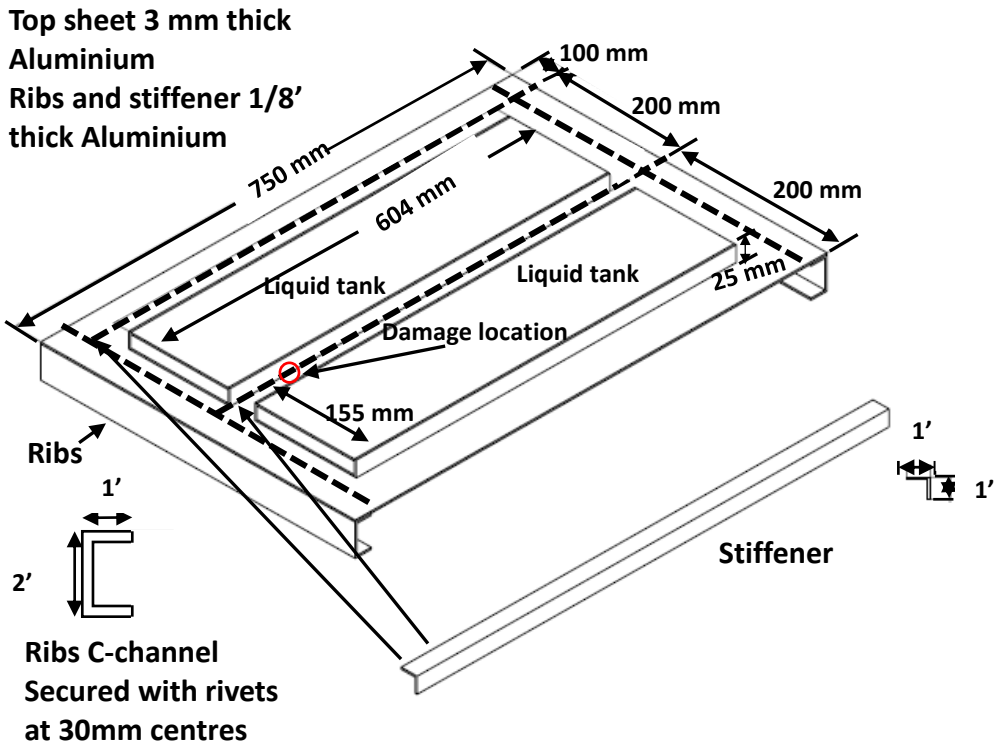


Figure 4.1: Physical model of the wing box-water tank bonded with the indication location of damage

4.2.2 Sensor/ transducers mounting locations

Deciding where to locate sensors largely depends on the type of damage and type of response to be captured. These issues are very important to ensure observable and significant features are achievable, ultimately would ease the pattern recognition in the later stage [1].

Prior performing the experiment, a procedure of locating suitable place to mount the accelerometers is briefly implemented. The procedure of a driving point is suggested by

LMS software where an impact hammer is moved from one indicated point to another called as 'Roving method' and triggered on locations throughout the top surface of the structure [59].

The ideal place for placing the sensors/ transducers is at driving point where most and apparent resonance peaks are indicated in the selected frequency bandwidth. This 'Roving method' uses the principle of reciprocal, which is one of property in linear system. It states that the ratio of input force X_A applied to the structure at point A to the output or response Y_B at point B, the ratio will be equal to the applied force X_B applied at location B and the measured response Y_A measured at location A. This property can presented as

$$\frac{X_A}{Y_B} = \frac{X_B}{Y_A} \tag{2.1}$$

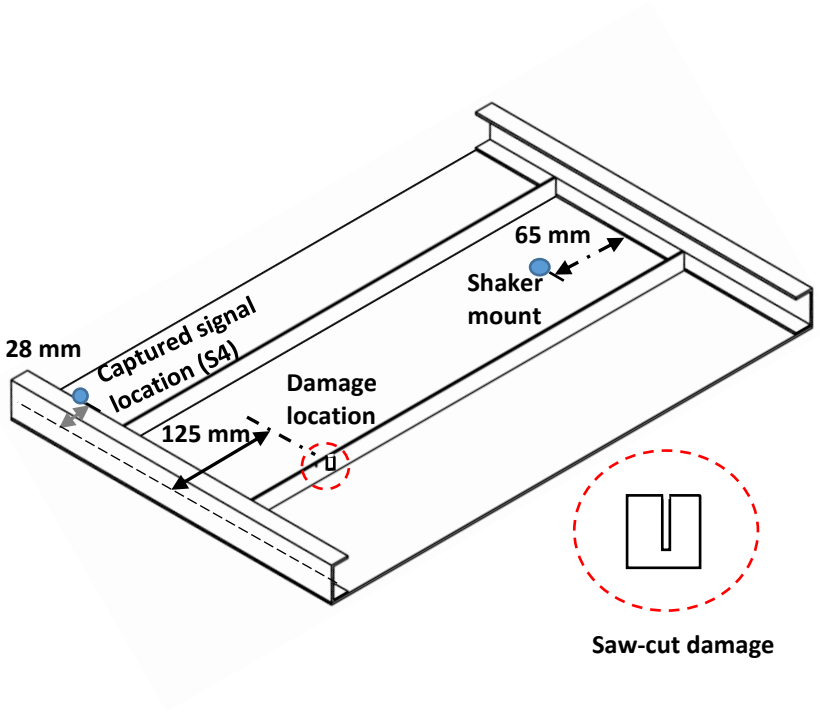


Figure 4.2: Bottom view- the location of damage, mount of shaker and the accelerometer.

Figure 4.2 illustrates the location of selected accelerometers in comparison to damage and shaker situated below the wing box. Observe that all FRF measurements are taken for analysis and it is indicated that reading from accelerometer S4 provides better FRF data signal in terms of FF definition and higher FRF amplitude compared to other sensors as shown in Figure 4.3.

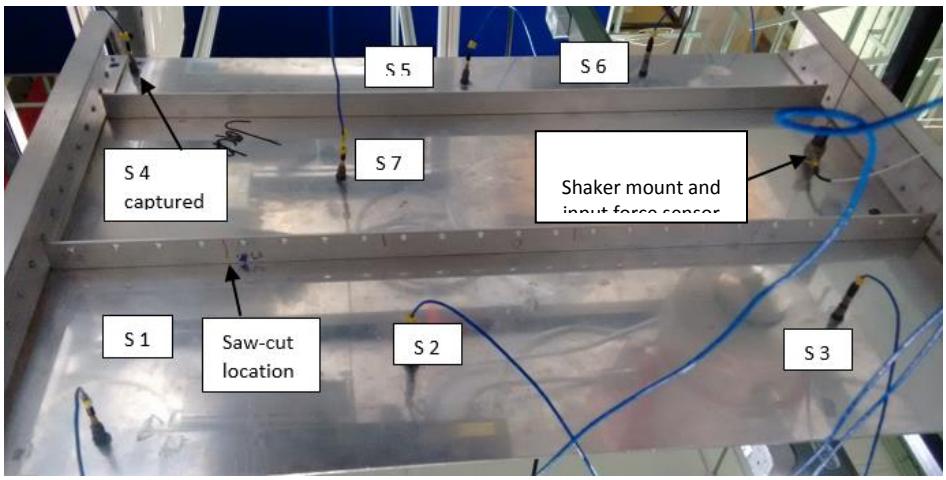


Figure 4.3: The actual bottom view of the wing box with sensor placements and shaker mount.

S1 indicates accelerometer number 1 and other corresponding sensors located below the wing box as, indicated in Figure 4.3. Intuitively, the sensors should be placed near to the damage and not to overlap with the nodes of resonance modes[38]. As in this test, the accelerometer labelled as S4 located near the edge corner of the plate, results in the most desired FRF signal. The signal also shows apparent shifting of FRF peaks as a result to load changes which will be discussed in details in Section 4.4.

4.2.3 Program configuration

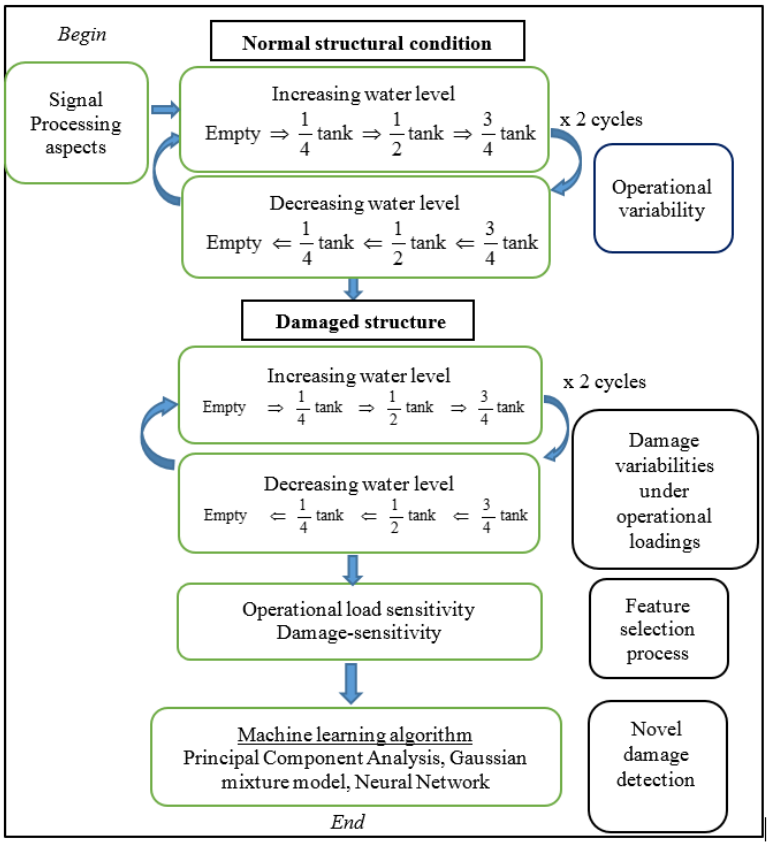


Figure 4.4: A flow diagram of the experimental configuration

Figure 4.4 shows the work-flow of this experiment and the feature selection that will be emphasised later in this chapter. The loading process is performed in a cyclic and a repetition manner as illustrated in Figure 4.4 where each particular structural conditions are divided into undamaged and damage conditions. Figure 4.4 illustrates the steps taken with regard to the current experimental work. It started with determining the suitable selection of signal processing aspects before implementing the experiment. As described previously, it assumed 5 loading conditions in each cycles of increasing and decreasing of mass loading for each structural conditions (undamaged and damaged conditions). Future selection will be addressed later when completing the data acquisitions and data organising process.

Loading of liquid into the tank was performed in 2 cycles; the first cycle assumed increase of loading from empty to full tank (E-Q-H-TQ-F) and decrease of loading from full to empty (F-TQ-H-Q-E). The second cycle repeated the same process. All cycles of mass loading classes consists of a particular loading condition. This strategy of cyclic loading is done in order to validate the data consistency and repeatability later in machine learning with regards to the variabilities that may occur due slight variation of loading amount during the filling the tank and emptying liquid mass from the tank (shown in Figure 4.5).

The experiment was began by measuring the vibration response of the wing box in normal condition in series of loading cycles from empty tank to full tank and continue the cycle from full tank to empty tank to complete one cycle. This filing up to full and reducing to empty was repeated in two cycles (as shown in Figure 4.5).

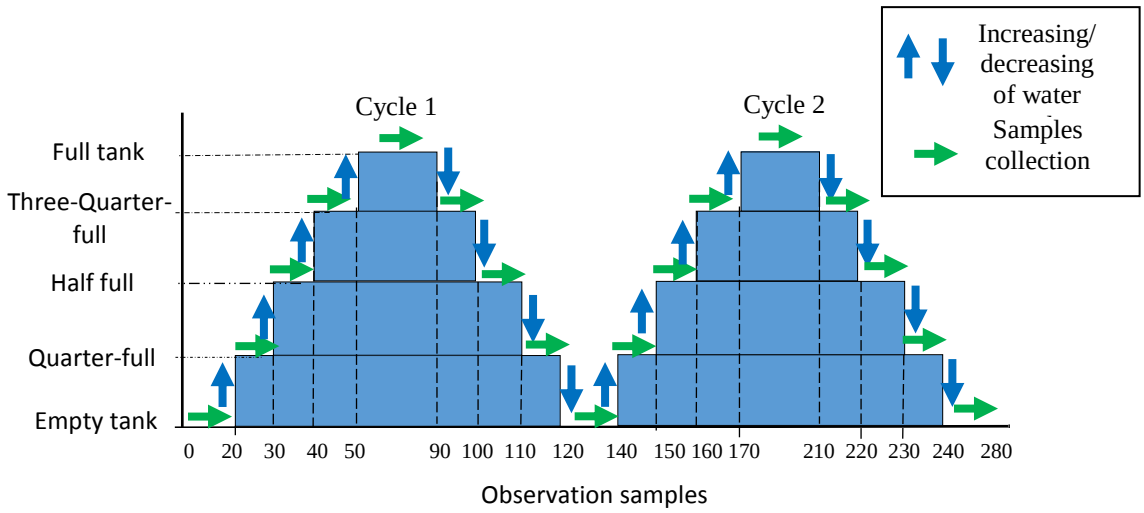


Figure 4.5: The strategy of loading/ unloading and then data measurements in sequence.

This is a practical approach in introducing the loading in changing mode with respect to the practical scenario where the loading changes are more eminent and more abrupt compared to slower changes of damage severities in a real situation. In this case, the loading conditions undergo more changes in a repetitive manner while the structural damage conditions change only after completing all the loading conditions.

FRF data was measured after completing each loading sequence corresponding to the specified tank level associated with one particular structural conditions which process was indicated in Figure 4.5. The data measured from normal condition structure is categorized as a baseline model encompasses all five the loading conditions. There are five different loading conditions performed in each case of five structural conditions. Therefore, there are 25 loading classes all together which each structural conditions comprises five loading conditions.

The loading process and unloading of liquid water were performed conventionally as shown in Figure 4.6. The process configuration is in accordance to the test program (shown in Figure 4.4). The wing box plate is excited using the shaker placed underneath the plate. The experimental procedure performed in systematic manner, in fact is an important strategy of introducing operational loading variation consistently with a minimal measurement variations.

- Water loading and offloading procedure

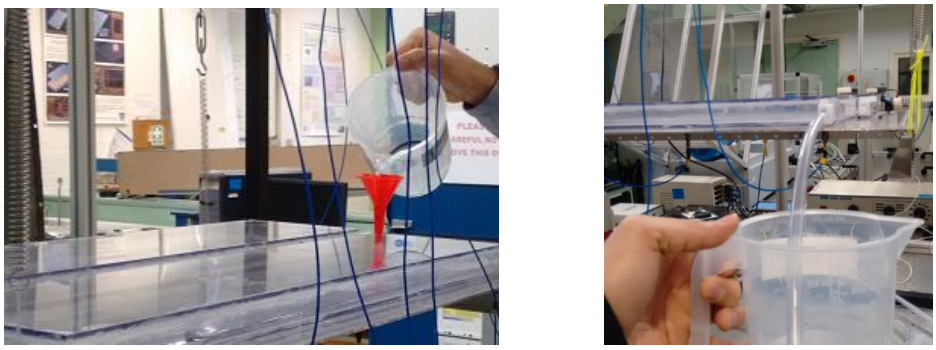


Figure 4.6: (Right) The manual filling of water into the wing-box tank. (Left)The extraction of liquid out of the tank using a manual switch.

Before performing the experiment, some essential aspects of signal processing are examined and tested before the appropriate techniques are chosen according to the guidelines provided by the LMS manufacturer [60]. This includes type of excitation signal and its frequency range together with windowing selection. For suitable techniques and

procedures in signal processing where data pre-processing is concerned, a comprehensive reference can be found from [60], [61].

4.2.4 Introducing load and damage variables

- Damage initiation

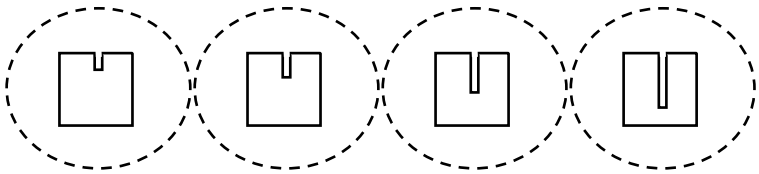


Figure 4.7: Quantification of damage severities introduced into one of the stiffeners

A successive increase of damage severity is introduced into an inboard stiffener of the wing box by 16%, 31.5%, 47%, and 63% cut of the stiffener depth, corresponding to D1, D2, D3 and D4 respectively as shown in Figure 4.7. These cuts quantified to the length of 4 mm, 8 mm, 12 mm and 16 mm respectively of the 25.4 mm stiffener's height. The saw-cuts are introduced successively directly into one of the stiffener by using a hacksaw device after completing all operational loading conditions corresponding to the damage condition of the saw-cut level. The procedure is done without removing the stiffener as to avoid disruption to the signal response in the effects of boundary condition. These issues changing structural boundary conditions such as tightening of bolts have been described in the previous research by [62].

Once completing the loading process for all five loading conditions separately for each undamaged condition, data measurements were recorded. The process was repeated again for each damage conditions as described in Figure 4.4. The measurements of all data samples can be gathered as loading class based or structural conditions based. Figure 4.8 shows the overall data acquisitions grouped based on loading conditions. For the sake of convenience in machine learning and data processing, all loading class and damage

conditions will have 200 number of data measurements respectively. So that, there are 1000 data measurements to be used in data pre-processing.

Wing-box tank level	Structural conditions	Total acquisitions	
Empty	Normal	80	
	Damage 1	60	
	Damage 2	60	320
	Damage 3	60	
	Damage 4	60	
Quarter-full	Normal	40	
	Damage 1	40	
	Damage 2	40	200
	Damage 3	40	
	Damage 4	40	
Half full	Normal	40	
	Damage 1	40	
	Damage 2	40	200
	Damage 3	40	
	Damage 4	40	
Three-quarter-full	Normal	40	
	Damage 1	40	
	Damage 2	40	200
	Damage 3	40	
	Damage 4	40	
Full tank	Normal	80	
	Damage 1	80	
	Damage 2	80	400
	Damage 3	80	
	Damage 4	80	
Total observations samples			1320

Figure 4.8: Summary of total observations samples collected during the wing-box experiment (all data set has been arranged according to the mass loading class)

4.2.5 Data acquisition

The data acquisition system used in the test is a DIFA SCADAS III of 16-channel and high speed data acquisition system, controlled by LMS software running on a Dell desktop PC.

The measurements were recorded using frequency range of 0-2048 Hz with a resolution of 0.25 Hz. With this resolution and sampling rate, the samples were presented in terms of spectral lines in total 8192 spectral lines. The wing box was excited with a white Gaussian signal through LDS shaker powered by an amplifier of similar brand. The response was measured using PCB piezoelectric accelerometers mounted vertically on top of the wing box (Figure 4.3). The excitation signal was measured by a standard PCB force transducer.

The base measurements used in the test are FRFs acquired using input sensor which is located at shaker mount on the plate. Prior locating the best place to attach the sensors, an impact test is done to detect the location where significant energy amplitude present using a hover method.

In this test, the structure is excited using random signal. The signal is applied with Hanning window to improve the signal continuity and the measurements are performed with 8-averages. It is essential to note that Gaussian signal is selected due to its better generalisation of data measurements compared to periodic signal. In view of variabilities that may present during the test, periodic signal will be sensitive to the unmeasured variability such as the fitting conditions other than the variability introduced and specified in the experiment. Prior the test, a preliminary test on the suitable signal processing properties to use including the comparison between periodic chirp and Gaussian signal in regard to Hanning and Uniform window with different averages are studied and undertaken before reaching to a decision on the signal processing.

4.3 Data decomposition

In the signal post-processing phase where assessment of the structural health is done, it is crucial to select beforehand, an appropriate arrangement for the multivariate data in the matrix form for the more convenient implementation of machine learning process. How the data samples are arranged and organised depending on the variables of interest and whether the data should be centralised and normalised prior to the use machine learning algorithm are some important aspects to consider. In this case, the variable of main interest

is the damage and loading is the operational condition that is expected to be the main driving force of the system dynamic characteristics. It is desirable to detect damage accurately and free from any masking effects of the loading variables.

In this experiment, the signals collected from the LMS acquisition system and were gathered in a three-dimensional (3D) matrix (Figure 4.9). In the 3D matrix, K represents the number of observations or experimental trials, L denotes the loading class, N represents the number of spectral lines selected and J is the series of sensors used. Because this experiment is performed in an independent test for each mass loading and is discontinued before the next mass loading, the 3D data matrix consists of one operational loading.

To extract the data, the 3D matrix shall be reduced into a bi-dimensional matrix by selecting the particular accelerometer that produces better definition and clearer shifting of the FRF peaks in regard to the changes in damage sensitivities. It is acknowledged that accelerometer number 4 is the most suitable choice in view of this criterion as suggested in many SHM references [1], [63].

Once the data signal is unfolded into the 2D matrix (Figure 4.9) in which each rows and columns represent the number of samples and the number of spectral dimensions respectively. Then, every loading conditions are join successively into another loading class in seceding order as one matrix corresponding to one type of structural condition (Figure 4.10).

All data samples from undamaged conditions are grouped into one matrix, named as a baseline set and becomes a reference set in this study. Establishing a baseline set that incorporates all operational loading variables has been reported in some earlier work in novelty detection in changing environment and concluded that it is a very effective strategy in normalizing the changing operational and environmental variations in the data set [2].

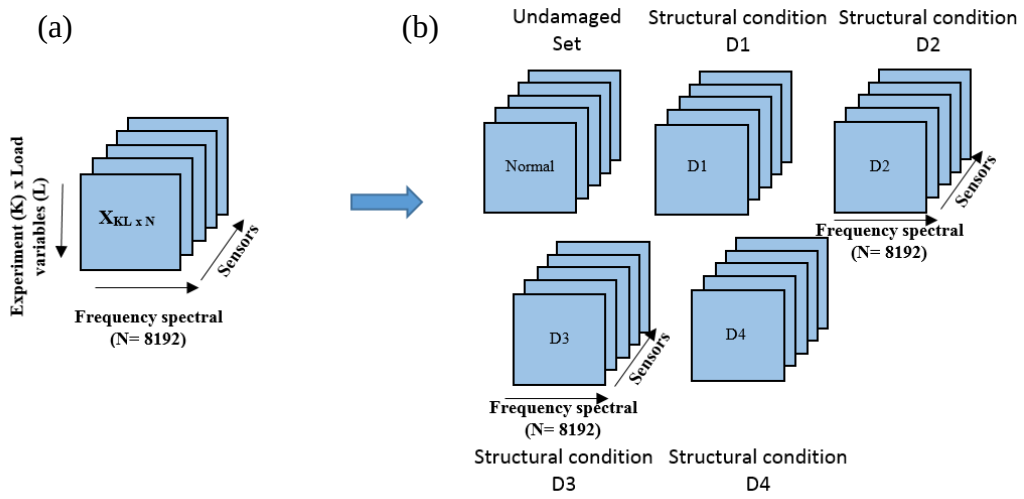


Figure 4.9: Collected data (a) initial 3D-matrix and (b) repeated for different structural conditions.

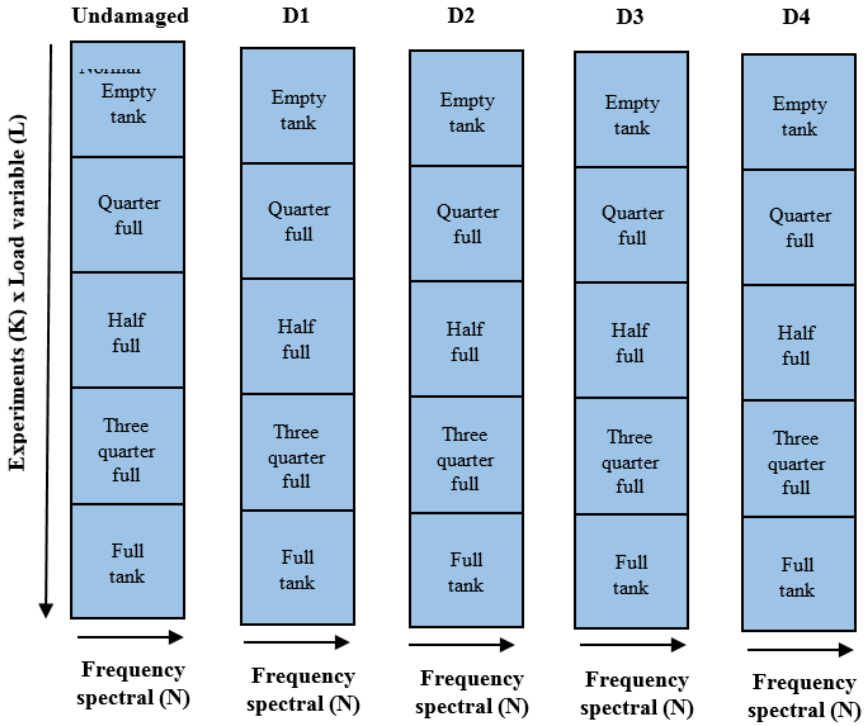


Figure 4.10: General scheme for data organization in the multivariate loading case after selecting the best sensor for data decomposition.

Figure (4.10) illustrates a decomposition of the experimental signal reduced from the three-dimensional matrix into two-dimensional $(K \times N)_{JL}$ matrix corresponding to sensor J and tank load L. The data acquisition composed of FRF data from one loading condition (L_i) where i can assume empty, quarter-full, half-full, three-quarter full and full tank. The data matrix $X_{KL \times N}$ comprised of the observations for all loading classes (in the rows) in ascending order and selected number of spectral dimensions before applying PCA on the data matrix. The same procedure data decomposition is repeated for structural conditions D1, D2, D3 and D4 and they are categorised as test sets.

4.4 The system's dynamic characteristic

One of the key aspects identified in the dynamics response of the structure when loaded with mass of water is that the lower frequencies modes showing very distinctive shifting and high definition of the FRF peaks compared to those peaks in the highest frequency range (Figure 4.11).

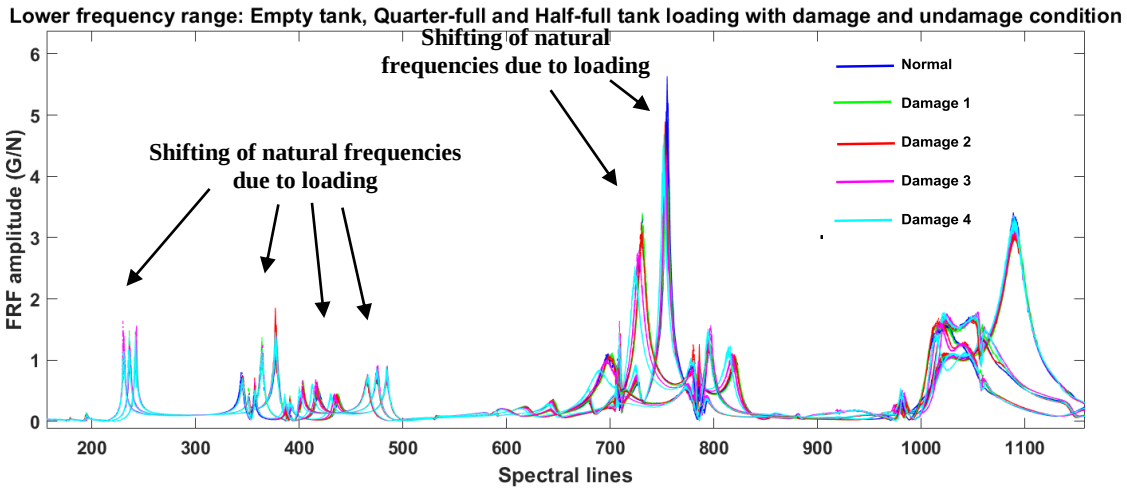


Figure 4.11: Comparisons of lower frequencies modes to the higher frequencies modes for operational loading variables under undamaged and damage structural condition

One obvious indication from the overall plot is that the mass loading plays a significant effect in the global dynamic response compared to damage phenomenon. This requires more detailed analysis focusing on the smaller region of the frequency range.

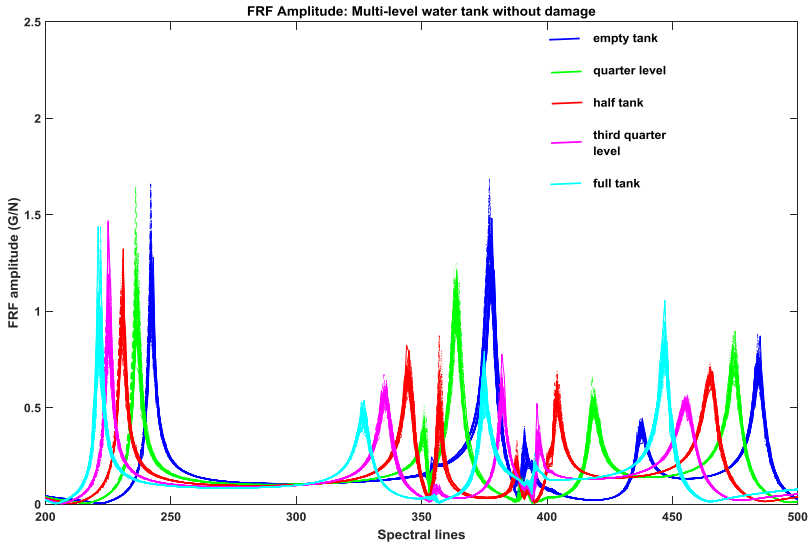


Figure 4.12: Showing distinctive shifting of the FRF peaks (lower frequency range).

Zooming in the lower-frequency response between spectral lines 200 and 500 of the previous plot, Figure 4.12 indicates very clear shifting of the FRF peaks. Note that, this FRF results are presented before the response from damage variables are plotted along the loading variables.

Intuitively, these global dynamic modes effects give desirable conditions in SHM as they indicate clear shifting in the natural frequencies. However, many studies related to vibration based damaged detection has revealed that damage is a local phenomenon and the effects can be found typically on higher frequency mode [1], [4]. According to the basic principle of vibration based damage detection, any changes to the mass and stiffness properties can alter the global dynamic response of a system. In this case, extra care should be taken to verify if these shifting are damage-sensitive or only loading-sensitive especially if the shifting lies on the lower-frequency global mode.

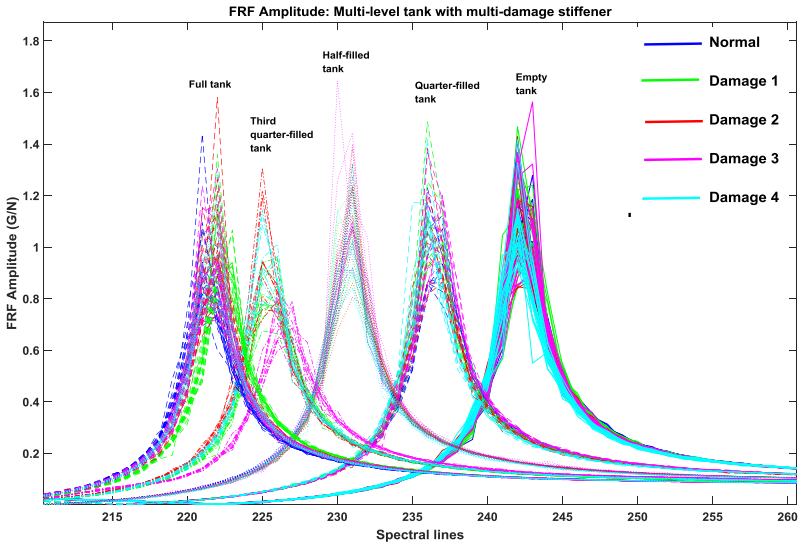


Figure 4.13: A zooming into the above FRF shows no damage-sensitive about damage data signal

Consequently, the FRF plots associated with damage variables are plotted on the same plot previously shown in Figure 4.12. By focusing on the range from 200 to 250, Figure 4.13 reveals that the sensitivities of the FRF peaks are in fact initiated by loading changes and not by damage. These damage effects also include the most severed case (D4) but still it does not influence the sensitivity of the FRF peaks. It is shown that these global modes of natural frequencies are conclusively dominated by the changing of mass loading.

These are the main challenge in SHM considering operational loading variability, which requires one to differentiate the effects of damage from the more dominant effects of changing of mass loading illustrated on these FRF peaks. On a positive note, under a supervised learning, data labels can provide basic information about the underlying data that cause changes to the dynamic properties of the system.

In this test, the mass loading is introduced successively under several quantified damage conditions. A selection of frequency range between the spectral lines between 200 and 500 displays mostly the global frequency modes of the system (Figure 4.14). The result reveals that damage effects cause insignificant shifting in FRF peaks in comparison to the variables of mass loading. The damage-sensitive FRF peaks begin to emerge on the higher

spectral lines, in this case around the spectral lines 400 to 500 as indicated by the dash-box in Figure 4.14.

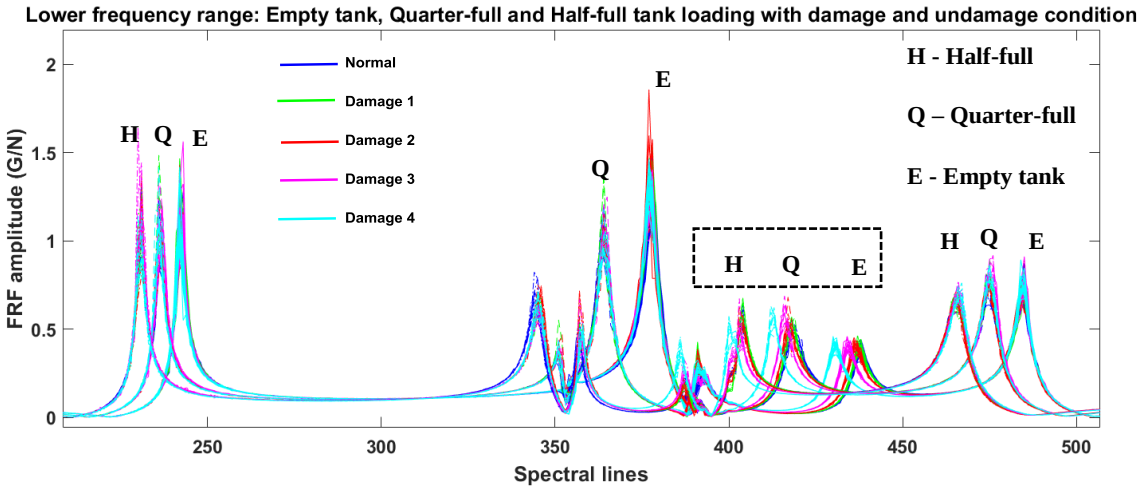


Figure 4.14: Operational loading variations with all damage variables present- the box with dash-line indicates presence of damage sensitivity.

The results in Figure 4.13 and 4.14 affirm the dominant effects of operational loadings that overshadow the effects of the damage variables. Further examination on the changes of these FRF peaks, confirms that there are small damage sensitivity on the peaks as indicated by a dash-line box, H Q E in Figure 4.14. These findings conclude the dominance effects of operational loading variables in regards to the structural damage.

With respect to higher spectral range where a possibility having peaks with higher damage sensitivity is higher, following plots show such characteristics.

In the spectral range dimensions of between 1800 and 2200 (425 Hz and 550 Hz), where local modes of natural frequencies having higher influence in that higher frequency range, it presumably the damage identification and detection is better (Figure 4.15). They are true in the sense of damage identification and separation. The FRF peaks are higher in magnitude but shown to be less definitive. Despite good separation of various damage severities in these higher frequency range, the peaks in lower frequency range is more defined and the changes of FRF peaks are more evident. In fact, it can briefly provide general judgement on the progression of damage under the operational loading variations.

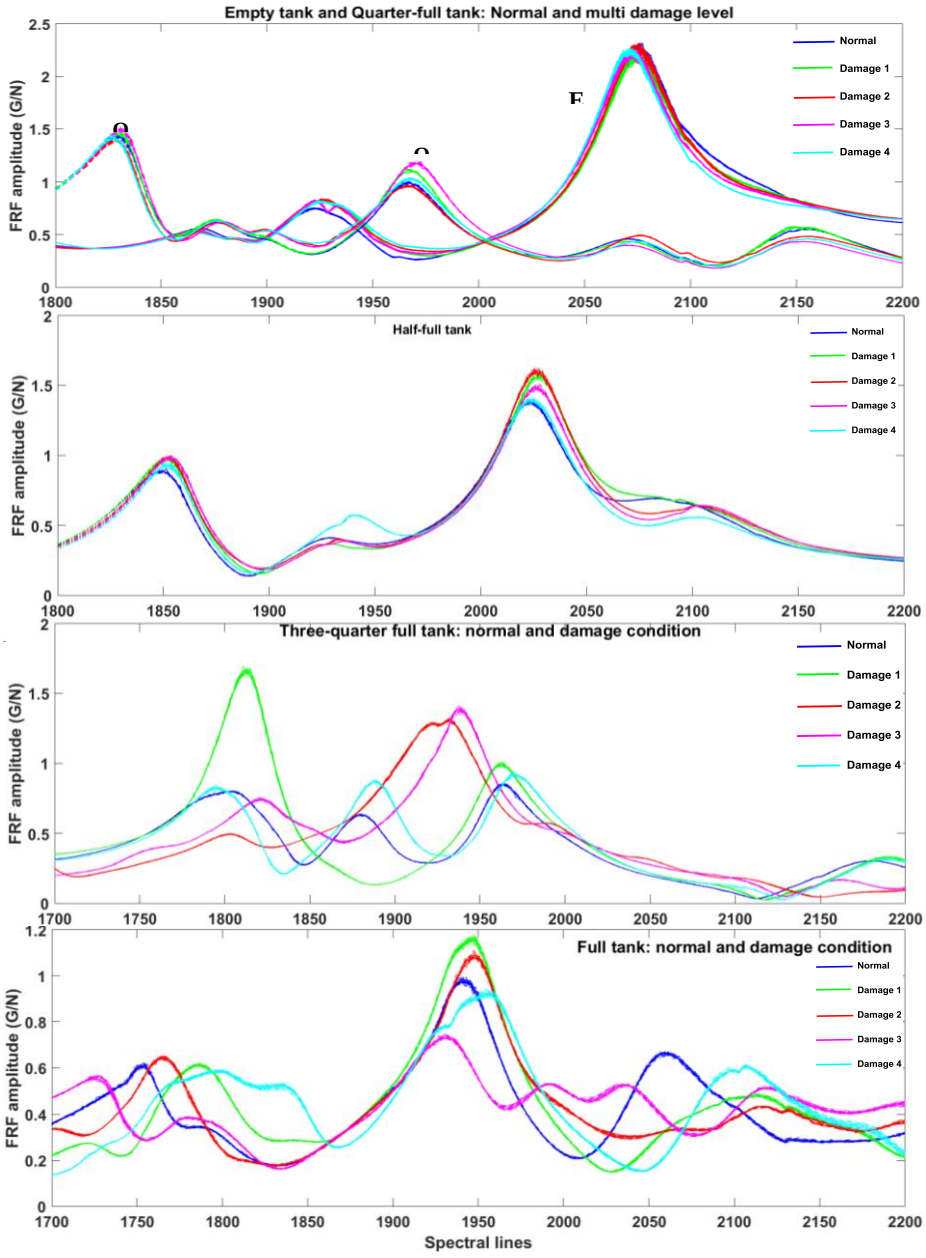


Figure 4.15: Individual operational tank loading plots of various structural conditions analysed in the higher frequency range.

In the aspect of damage detection and SHM, practically, the idea of using the shifting of natural frequencies is not convenience and inaccurate if the whole purpose of SHM is to be realised. There have been long-standing work concerning the detection and

characterisation of structural damage by examining the changes in structural vibration response [17], [22], [64].

An intelligent way is advisable by using the appropriate machine learning tool in order to discriminate the effects of operational loading variations from the damage features rather than using visualisation method as shown previously as the means to achieve the purpose of SHM. This concern leads to the next work that focuses on feature extraction in bringing out the damage-sensitive feature out of the dominance load variables effects. In fact, this is the key challenge in this work in the perspective of damage detection. A standard linear PCA is to be applied for this purpose and it is highlighted after making some strategies (in data organization) to ensure a more efficient use of PCA.

Despite the minor effects of the damage variables on the FRF changes, the good news is that the damage-sensitivity are observable and identifiable by increasing the plot resolution (zooming the plot) as highlighted in Figure 4.16. Using a range of selected spectral range which the focuses on damage-sensitive peaks and spans across all loading conditions, it can be a representation of one of the important characteristics of the system. Such representation is called as a feature in terms of data pre-processing prior to its application in the machine learning which will discussed in the next section.

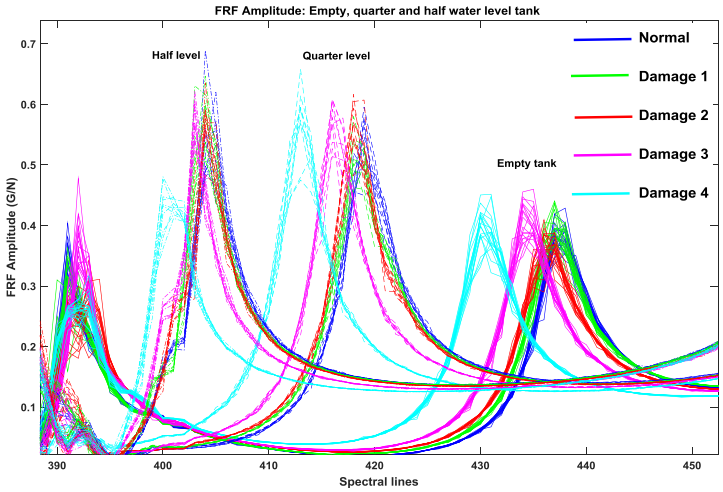


Figure 4.16: A zooming from the dash-box in Figure 4.14 highlighting the damage sensitive feature

From the finding based on Figure 4.16 and Figure 4.17, it illustrates that the vibration response of the wing box can provide some brief indications on the changing of damage severity to the extent of large damage quantification (Damage 4). This damage sensitive feature lies in the lower range of frequency range of between 350 and 450 spectral lines (87.5 Hz and 112.5 Hz).

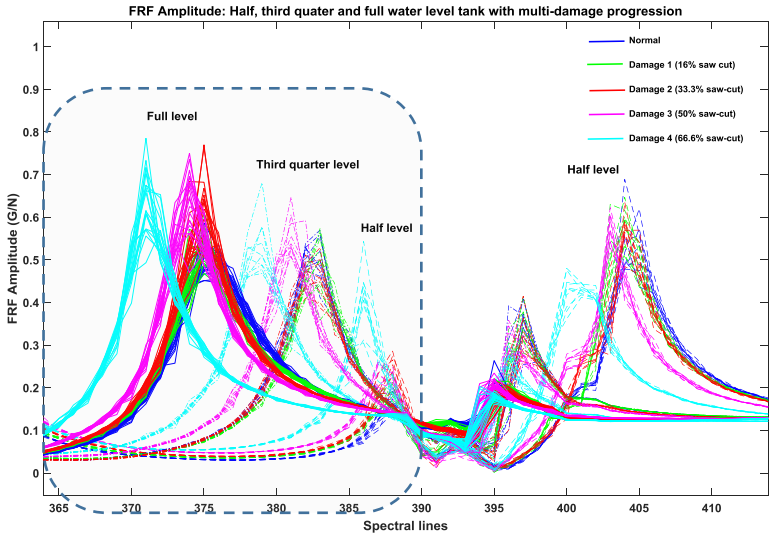


Figure 4.17: A close-up view of higher operational load with all damage variables

4.5 Features selection and extraction

In the SHM perspective, the performance of a damage detection system heavily depends on the how much data information represented by the selected features prior to the machine learning applications. The selected features should contain most information about the system dynamic behaviour especially for damage variations (damage-sensitive) as highlighted by the FRF changes [1].

The purpose of establishing good feature is for the use in the machine learning algorithm which its performance results heavily dependent on this feature. It has been demonstrated in the previous section of System's dynamic characteristic how the potential candidate of damage sensitive feature to choose. The potential feature at the selected frequency range displays detectable changes in the natural frequency due to varying load and damage severity changes. The clear advantage dealing with data in FRF domain, the criteria of frequency peaks shift due to load and damage change plays a key role in selecting good sensitive feature.

It has been indicated the selection process of the feature after examining all range of frequency spectrum in spectral dimensions. As a result, the spectral range of between spectral lines 350 to 450 is concluded as the feature candidate. They are selected based on the distinct changes of FRF peaks due to damage variations surrounded by the dominant effects of operational loadings and minimum noise around the signal.

In the current study, the selected feature serves as the input data for the machine learning algorithms that will used later in the study in Chapter 5 and Chapter 6. A good feature selection enables the extraction of underlying characteristics of the structure and subsequently allow sufficient pattern recognition to provide reasonable assessments on the system health condition.

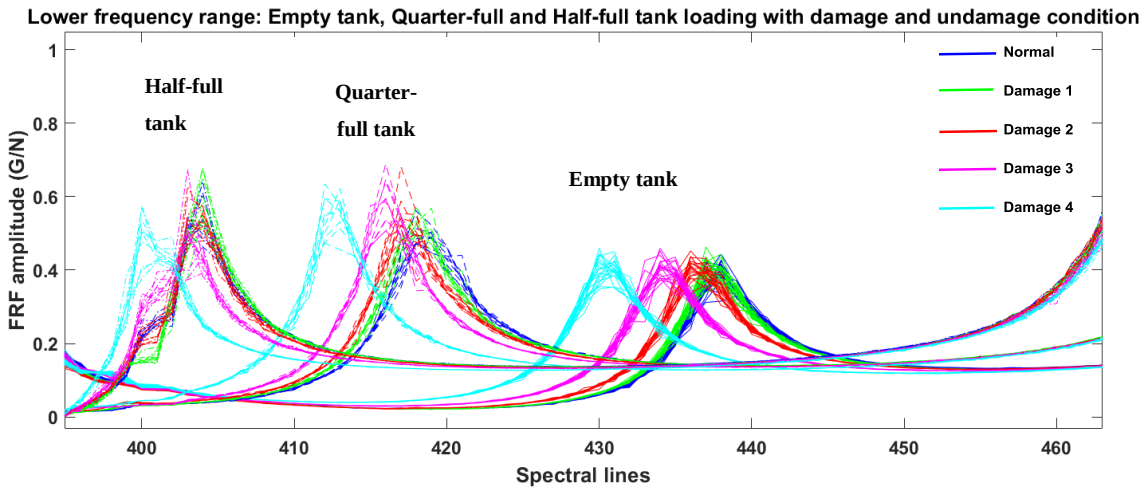


Figure 4.18: Selected feature showing damage sensitivity.

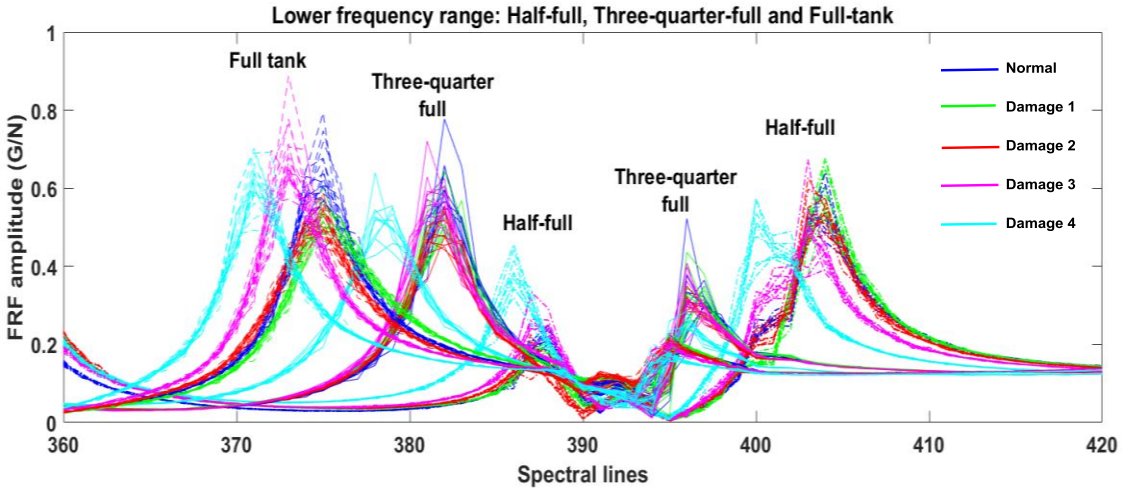


Figure 4.19: The extent of a selected feature encompassing Full-tank, Three-quarter-full and Half-full tank

As indicated in the plots of Figure 4.18 and Figure 4.19, detectable damage sensitive is observed from the spectrum ranged from 350 to 450 spectral lines which are now being selected as the damage-sensitive feature under loading variables for this work.

4.6 Data pre-processing prior to machine learning

In this stage, the data is manipulated before processing it with a machine learning algorithm in order to make it more convenient use with machine learning algorithm. Here, the selected damage-sensitive feature, which comprised of the spectral lines from 350 to 450 spectral lines (the total of 101 dimensions) is already established as the data feature. The essential part to remember that, this selected feature of 101 dimensions of frequency spectral lines consists of five different mass loadings with each mass loading. By each mass loading, it involves of four different damage conditions and one undamaged (normal) condition. There are number of interesting methods for different data organization that been reported for the last 5 years, mostly involved with the use of the PCA[65][66] [19]

In order to establish a more practical way in dealing of multi variate data of varying mass loading consisting with different structural health conditions, a baseline or reference set

is established. This imply that the whole normal (undamaged) structural condition of mass loadings are gathered and combined into a group. This data group will be the reference data points for other damaged conditions of any operational loads. The baseline set consists of 200 data samples of healthy condition comprising 40 samples in each loading class in empty tank, quarter-full, half-full, three-quarter-full and full tank loading conditions.

The same configuration is used for test sets which each set describes one particular damage class spanning across all loading conditions. Hence, there are 200 data samples for each structural conditions where in each structural conditions consist of 5 loading conditions. There are 1000 data samples that will be used for the purpose of machine learning. The details of the data configuration will be described in Chapter 5.

4.7 Summary

The experimental work performed in this chapter has addressed how the experiment is configured and implemented systematically concerning the operational loading variables and damage variables.

The underlying dynamic characteristics of the system under the influence of mass loading has been outlined. The findings reveal that the loading variations produce more significant effects on the shifting of the FRF peaks and more dominant compared to the damage effects. It is demonstrated that there is a higher sensitivity to loading variation detected in the lower frequency range whereas damage sensitivity is more subtle and can be recognised in few of the FRF peaks after the global frequency mode in slightly higher frequency range. The damage-sensitive feature is shown to be masked by the loading-sensitive feature which are more dominant over the damage-sensitive feature through all the frequency response.

It has been demonstrated how feature selection under the effects of operational loading variables has to be exercised with extra concern due to the masking effects of the loading

variables. Damage sensitivity features are present within each peaks corresponding to loading variables. Zooming has to be used to highlight these damage-sensitive features.

The data feature selected in this chapter will be used in the machine learning algorithms in Chapter 5 and 6. The performance of this machine learning approach critically depends on the quality of the selected feature performed in this chapter.

The procedure of feature selection is carried out with the availability of data labels (supervised learning). In real case scenarios, this privilege is not always available. This case will be explored in Chapter 7 where examinations of FRF shifting are not permissible due to high of data variables and complexity

Chapter 5

MULTIVARIATE DATA VISUALISATION AND DAMAGE DETECTION UNDER OPERATIONAL LOADING

This chapter highlights primarily visualisations approach using Principal Component Analysis (PCA) and novelty detection based on some principal component (PC) models. PCA technique is used in this chapter using the data set obtained from the wing-box experiment described previously in Chapter 4. It is identified that the multi-variate data sets presented in FRF series obtained from the VVBD test reveals a challenging task in interpreting and examining the damage severity changes under the effects of changing load in the purpose of damage detection and SHM. A technique of visualisation is proposed as opposed to the examination of the shifting of the FRF peaks. PCA is performed to produce some new representations of the data set that can reveal the underlying characteristics and isolate the effects of loading variables from the damage variables in a basis of the highest

data variance found in the data set and projected the new variables in direction of the highest variance. This chapter first describes the application using a standard PCA follow by a kernel PCA performed as a generalised nonlinear form of feature extraction using the given data set. PCA is a well-known linear feature extraction technique whereas kernel PCA is used here with concern to extract some nonlinearity and hidden features that typically 'live' in the high dimensional data space.

5.1 The scope of the chapter

The motivation for choosing PCA in this work is mainly for dimensional reduction and extracting data features from the high dimensional space to a more compact and useful data representation. Under varying operational loading and structural conditions, the main challenge underlying the work is will be to identify if the effects from the structural damage on the feature derived from the PCA can be distinguished from the effects arise from pseudo-fuel loading.

For this purpose, the standard PCA is performed and projected onto a two-dimensional plot by the first and second principal components. A kernel PCA is implemented to acquire a hidden feature that may lies inside the high dimensional data space. Both PCA projection methods are compared and utilised as visualisation techniques in the context of SHM.

The study explores some different data arrangements, named as Principal Component (PC) models A, B, C and D by configuring the multivariate data sets corresponding to loading, damage or loading/damage variations. One of these arrangement sets is then incorporated into a covariance matrix before eigenvalues decomposition is performed.

As a supervised learning, the multivariate data comprises five loading conditions: empty, quarter-full, half-full, three-quarter full and full tank load and classified corresponding to the structural conditions. It has five different structural conditions to be considered: undamaged, D1, D2, D3 and D4. The PC model depends on the loading matrix or coefficient matrix calculated from the PC model covariance matrix.

In the first part of this chapter, it presents visualisations based on the PC models corresponding to specific data set arrangements. In the second part, novelty test for detecting damage based on the PC models are computed using Mahalanobis Squared Distance (MSD) function. The results of the discordancy test are compared between the standard PCA and kernel PCA.

5.2 Overview

The main challenge in interpreting the effects of various loading conditions using the original FRF plot is that the multivariate data sets seemed to overlap and redundant. The issue escalates especially when there is damage present and the changes of the FRF become more subtle. Previously in the Chapter 4, the selected feature also shows that the FRF shift caused by loading are more dominant compared to the FRF shift initiated by the presence of damage. This kind of data characteristic demands a better way of re-expressing the data set in the most meaningful vector basis in the hope that this new representation of the original data set can reveal the underlying feature better.

In another perspective, it is widely known that the standard PCA is a powerful technique for linear dimensionality and feature extraction [52], [67]. Emphasising on its description, PCA attempts to find a linear subspace of lower dimensionality than the original feature space, where the new features have the largest variance [34]. Hence, the main objective in this chapter is to project a new representation of a feature from the original FRF data set into a lower dimensional space using a new basis of a linear combination of the feature such that the variance of the projected data is maximised.

The primary aim work in this chapter is to ascertain if the standard PCA is able to extract and unmask the damage feature from the more dominant loading features and transform them accordingly based on their loading and damage class.

Due to the limitation of the standard PCA that projects the data set only onto a linear dimensional space, a nonlinear PCA shall be incorporated into the work to compare with

the standard PCA to determine if it can improves the detectability of the damage feature. The nonlinear PCA utilised using kernel Gaussian function is applied and the work here addresses a novel way in establishing the kernel parameter by searching the average value of the smallest distance between two vectors using dissimilarity matrix plots.

This chapter also describes various plots for visualising purposes based on different Principal Component (PC) models. The PC models are developed by projecting the data set using different transformation or coefficient matrix. These PC models are produced by orthogonal transformations with maximal variance of the feature data set based on different data configurations. The work embodied in this chapter can be outlined based on the following tasks:

- The effects of standardising a dataset in the PCA application
- Highlighting the novel way of selecting parameters when applying a kernel Gaussian function into PCA via a similarity matrix
- Presenting different strategies of organising a reference set or a baseline model prior to PCA application in the aim to reveal the features caused by damage in the overall data set
- Evaluating damage detectability of each baseline model by applying statistical T-squared analysis based on the Mahalanobis Squared Distance (MSD). A threshold by Monto Carlo computation is used as a benchmark between undamaged and test data sets in the reduced dimensional space.

5.3 PCA implementation

Generally, before implementing a data analysis using any machine learning techniques, data pre-processing is vital in terms of reducing high dimensional data and extracting the features sensitive to damage [6], [31], [58]. This is important in order to get the summarised information described by the multivariate data. Concerning this factor, PCA

is selected to transform the original variable into a more predictable representation of important principal components that retains most of the important information.

PCA is a widely used non-parametric tool for dimensional reduction of unsupervised learning problems. In the current work, taking the advantage of its supervised learning mode, it will be able to determine if PCA can classify and distinguish between the features caused by loading and damage. The primary objective and benefits driven from the PCA implementation particularly for this work are briefly described here.

- Feature extraction and pattern recognition

The main concern of this chapter is to determine whether the effects of damage and load are separable by using PCA. As described previously in chapter 3 in equation (3.1), the transformation basis is established based on the maximum variance of the data set. The covariance matrix C_x measures the covariance between all possible pairs of spectral variables denoted by X. Each row of matrix X corresponds to all spectral variable measurements (FRF amplitude) belonged to an observation and each column corresponds to the FRF amplitude for each spectral dimension measured for all observations in the matrix group.

The off-diagonal terms of the covariance values reflect the redundancies and noise in the system. Large off-diagonal values correspond to high redundancies and noise in the data set. In other words, there is certain amount of high correlation between variables. The variance values in the diagonal terms corresponding to interesting behaviour of the structure [65], [68]. The goal of the PCA suggestively is to maximize the variance and to to minimise the redundancy of the covariance in the off diagonal terms by making the off-diagonal elements close to zero. It means that by making zero correlation between pairs of the spectral variables which can be achieved through diagonalising the covariance matrix.

Figure 5.1 shows the original data feature as acquired from the wing box experiment that shows the response looks almost similar in behaviour trend and very difficult to interpret

without proper data group labels. In this context, PCA is used to improve the interpretation by focusing on the maximal variance of the few first principal components which often realise the most important characteristics of the system.

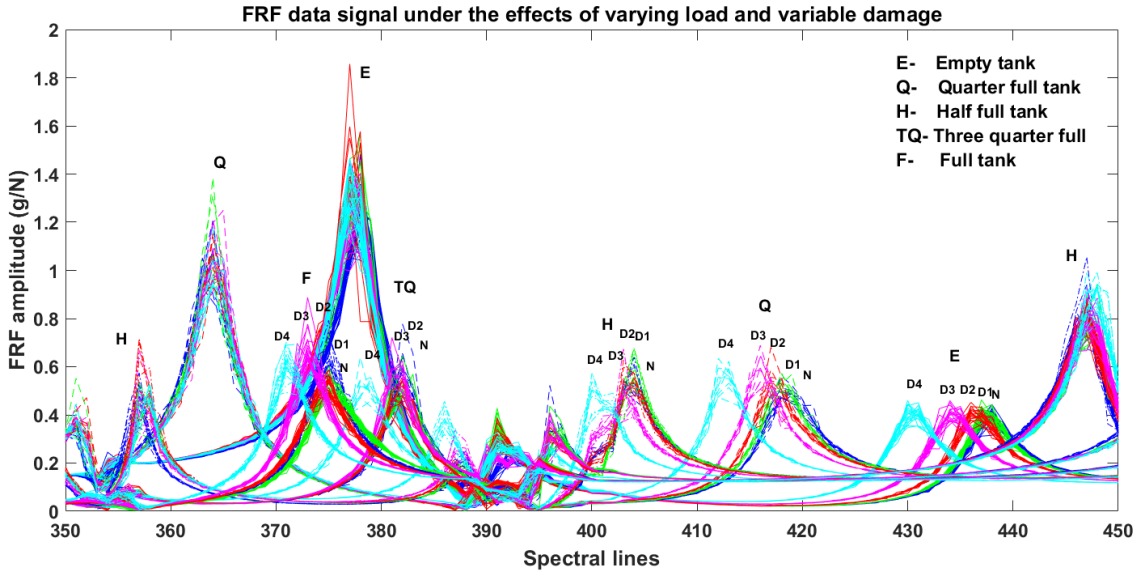


Figure 5.1: The original data feature with high spectral dimensions that inclusive of all operational loadings and damage data.

Figure 5.1 shows the original vibration data acquired in FRF series, which is obviously difficult to recognise without distinctive colour labels. The letter E, Q, H, TQ and F denote the operational loading conditions. The letter N, D1, D2, D3 and D4 represent undamaged and damage severities groups in ascending order with D1 as the lowest damage. The plot in Figure 5.1 indicates that the FRF changes due to the changing of loading variation (from E and Q and up to F load. E and Q load located between spectral lines 410 to 440 show higher variability among the data set associated to undamaged and D1, D2, D3 and D4. Whereas the FRF changes present within the spectral lines between 370 and 410 show lower variability in the ‘damage’ data sets for H, TQ and F load. This changing data characteristic would be sensitive to PCA (in terms of data variation) and would be useful to determine if this characteristic is better highlighted in the principal components projection.

- Reducing the number of spectral dimensions

As mentioned previously, the key priority is to reduce the number of spectral dimensions in the selected feature from the original high of 101 dimensions to a linear combination of the variables for better interpretation and visualization of the data variables. In PCA, the dimensionalities are reduced by transforming the original spectral dimensions into a small set of principal components of new variables without losing the the important information in the original spectral variables.

From the machine learning perspective, smaller number of components of variables are more desirable in analysing the relationship among the variables. Like in the current case study, by combining the original spectral variables in the direction where the data most spread out (maximal variance), it avoids the problem such as the singularity problems.

The key point to note is how many numbers of principal components should be retained in order to avoid a considerable loss of data information. It is desirable to have the total principal components that accounts for more than 80 percent of the total variation in the data set [33]. The projection of the PCs which having most of the data set variation is reasonably a truthful presentation of the total variables in the data set [33](Mujica et al. 2011).

5.3.1 Preparing the data prior to the PCA application

The key procedure prior to applying PCA is to prepare the data to ensure its mean equals to zero. In the current work where it involves multivariate loading and damage, it is essential to ensure the reference set or the baseline set which comprises of all undamaged data sets from all loading conditions has zero mean. All data set comprises of various loading conditions corresponding to a structural condition are subtracted by the mean of the baseline data set (undamaged state consists of all loading conditions). This is an effective step in order to get a more useful data visualization and interpretation of the multivariate data set.

PCA is very sensitive to variance and scaling of the original variables. Due to high variability in the multivariate data set, it is a good step to scale the data to a unit variance by standardizing the data. In the context of this study, standardisation refers to dividing the mean-centred data by the standard deviation of the undamaged data set. Standardised data set gives variance equals to one and the covariance lies between -1 and 1. The value of zero reflects no linear relationship between the two variables, -1 if the relationship between the two variables are perfectly inverse and +1 if the variables have perfect linear relationship [69]. In establishing the association between data sets of the undamaged and other damaged conditions, each test data set (gathered from a particular damage condition) is subtracted by the mean of the undamaged condition and standardised using the standard deviation of the undamaged data.

From Figure 5.2, it shows that the effects of standardizing the FRF test data set (damage data set from all loading conditions) by a unit variance of the undamaged baseline set. To perform standardising, each vector of test data is subtracted by the mean of baseline set and divided by the standard deviation of the baseline set. The amplitude of the test data after being standardized causes larger amplitude and larger separation between the baseline and test data.

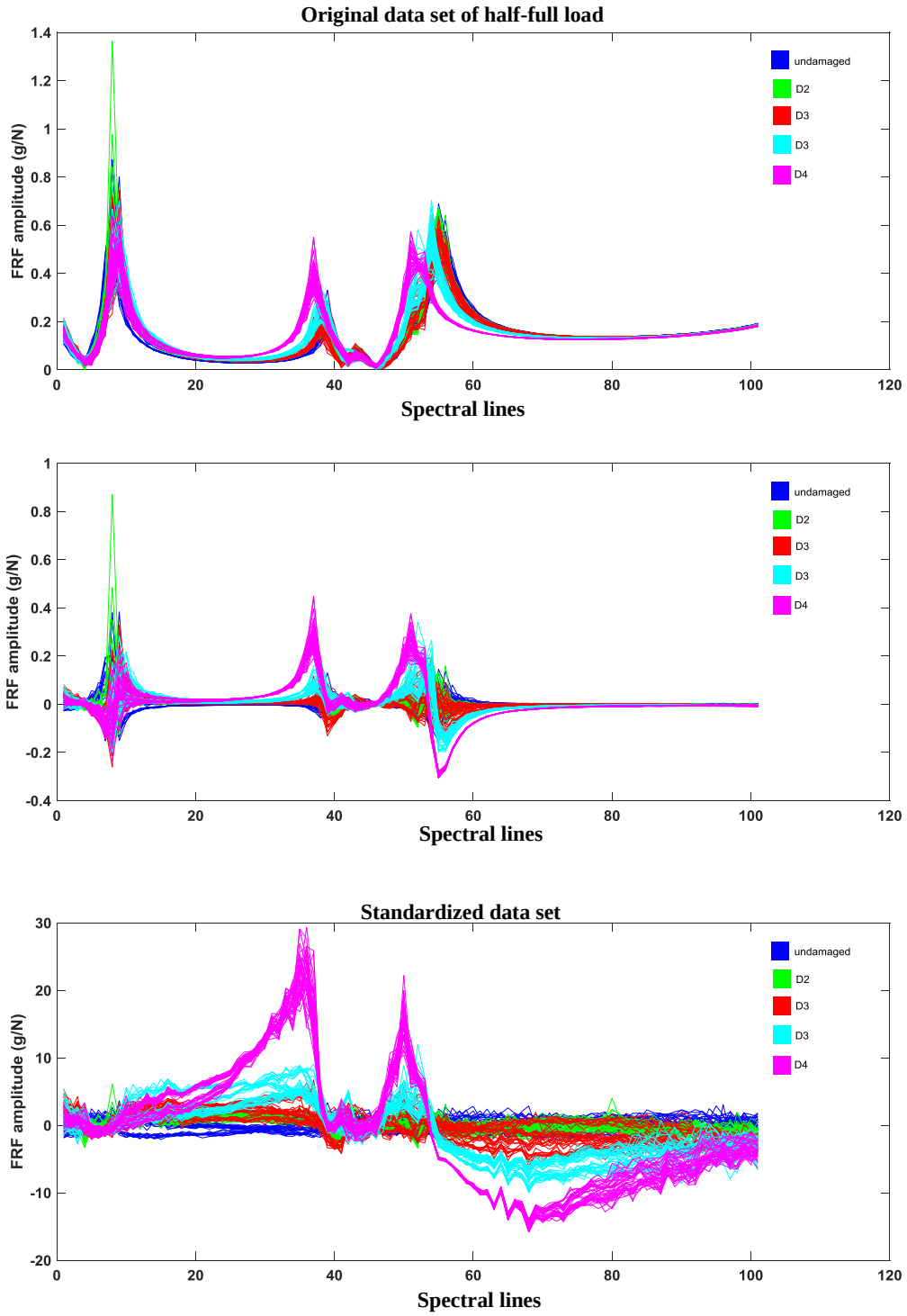


Figure 5.2: Comparisons of the effects of mean-removed data set (middle plot) and standardizing (bottom plot) with the original data set (top plot) measured in half-full loading condition.

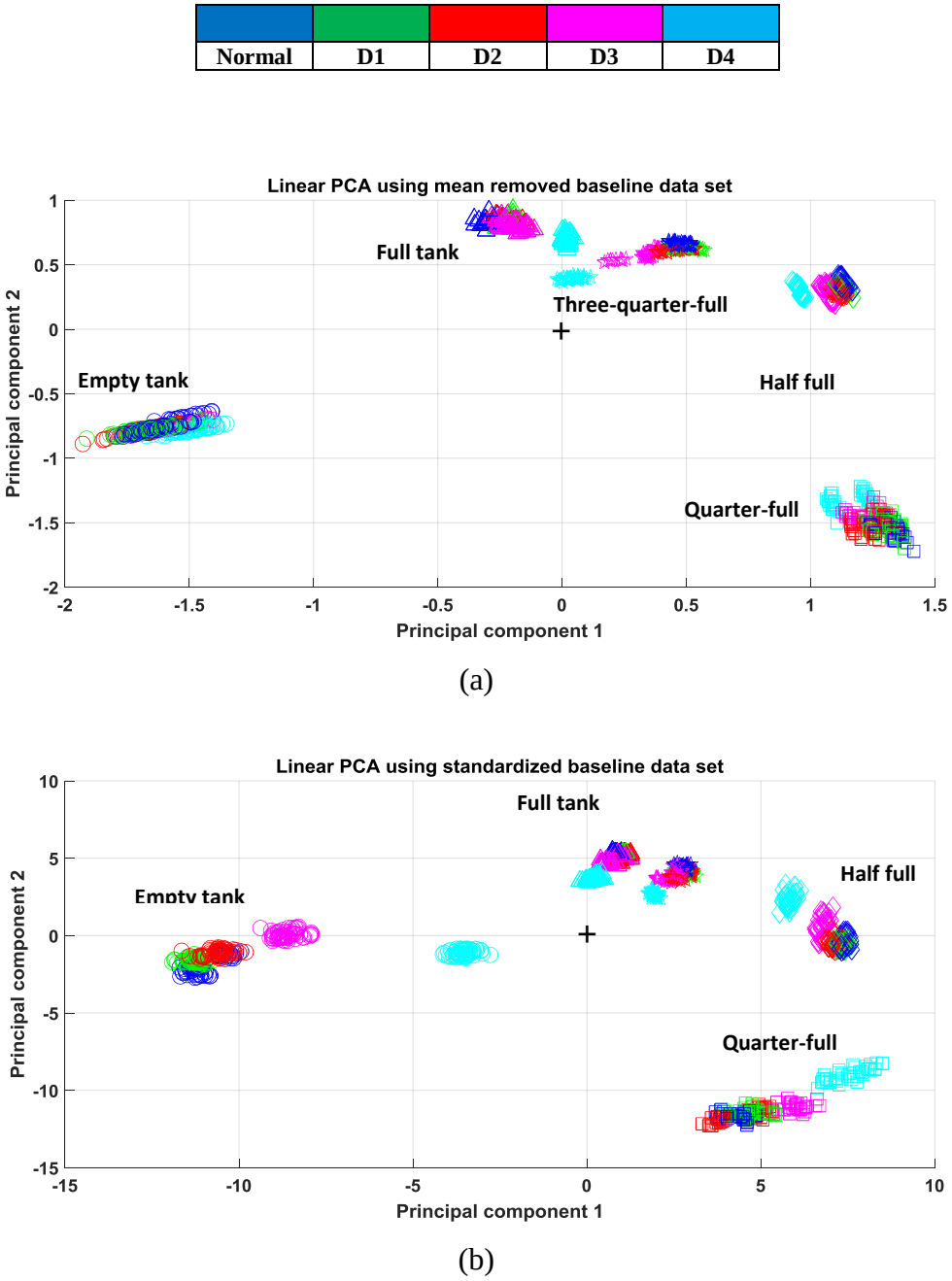


Figure 5.3: (a): Mean of baseline data set is subtracted from each test data samples. (b): Mean-removed data is standardised by the standard deviation of the baseline data set. Plus sign indicates the centre (mean) of the data set.

The transformation using the first two principal components (Figure 5.3, (a) and (b)) have shown that the loading class are clearly separable from each loading class. The principal components projections show that data correspond to smaller damage are very close to the data group of undamaged condition. These data sets seemed overlap especially in higher load class (half-full, three-quarter- full and full load).

The standardised data set shows a positive effect in making a larger separation between the undamaged and test data set in empty and quarter full load class. The results of the linear PCA shown in the top figure of Figure 5.3 display the transformation after the mean of the baseline data is removed from the data and making the overall mean of the data projected at the centre (0,0). Mean-removed of data set is the pre-requisite to a PCA implementation so that the variables are comparable.

One of the foremost observations from the linear principal components (PC) projections is that the loading groups are distinctively separated. However, the damage severities groups especially among the lower damage groups overlap considerably. In contrast, using standardized data set prior to eigenvalues decomposition, it improves the separation of damage groups especially those from smaller damage groups in empty and quarter-full load. Note that, in the previous original FRF plot, the variability among the data set from undamaged and damage conditions in the empty and quarter full load are higher and potentially more separable compared to the higher loading class (half-full, three-quarter full and full load).

The ultimate point is that, a standard PCA is shown as capable to separate the multivariate loading class effectively but the separation between the undamaged and damage severities groups in each loading class is unsatisfactory. Interestingly, the loading data sets are indicated to be on an ascending order pattern and this basically indicates that PCA is suitable to provide data interpretation and prediction regarding the system's operational loading conditions. Through PCA, a more feasible interpretation and prediction on the system's changing load is looking to be more achievable. However, the damage severities are observed to be overlapped especially for the higher loading class. Because of this factor, a kernel PCA is opted for use. It is also in the key interest of any SHM landscape

is to establish if a structure has damage or not. Another desired performance of a SHM is to be able to predict and monitor structural integrity through its data feature. It will be highlighted on these two capabilities in the next section by kernel PCA which are applied on the wing box data set.

5.4 Kernel PCA

Kernel PCA (KPCA) is nonlinear form of PCA first introduced by [52] and since then has been used in many applications including for dealing with the effects from the environmental and operational variations [70]. KPCA is vey effective in exploiting the complicated structure of the high dimensional data [26], [36], [40], [52], [55].

It is performed by using what is known as a ‘kernel trick’ by computing the dot products via the kernel Gaussian representation instead of having to transform the input data nonlinearly into the high dimensional space. In the ‘kernel trick’, the kernel function is solved in the data input space and avoid the complex computation of the transformation in the high dimensional feature space.

In separating the load data set, is is demonstrated that linear PCA can effectively be used to separate various loading groups and track the loading trajectory path. However, for a more complicated variable structure such as for the D1 and D2 test data (which are associated to the smaller damage groups), linear PCA could not separate the feature distinctively.

In this context, kernel PCA is introduced and implemented. In a previous study for facial recognition, kernel PCA has shown its distinctive performance [55]. The fact that Kernel PCA is the nonlinear form of PCA, it can reveal more complicated features and underlying patterns in the data set.

The key objective in this work is to improve the separation between baseline set (undamaged condition) and damage severities under all loading conditions. The basic

structure of the algorithm of the kernel PCA is inspired by the work in [55] and is adapted to the current multivariate-variate data problem. In the original application of kernel PCA, it was used to determine some micro expression from amazement to calmness of the human face, in contrary to linear PCA where it was shown to be limited to the basic orientation of the human face. The algorithm of the adapted kernel PCA to the wing box under operational loading and damage variables is attached on the Appendix A in this thesis.

The modifications are introduced in the algorithm specifically in the procedure for finding the the optimal Gaussian parameter σ, sigma. It involves constructing a distance matrix, also known as similarity matrix in a square form based on the Euclidean distance. Using a colour bar plot to represent the overall distance between data points from the same class or inter-class, it makes the evaluation of the optimal standard deviation more convenient and simpler.

In the following section, the procedures for implementing kernel PCA as s continuity from previous Chapter 3 are presented. A strategy for searching of the parameters related to the distance between data points and the standard deviation for the kernel Gaussian is demonstrated.

5.4.1 Constructing the kernel matrix

The 'trick' of kernel PCA is obtained by solving the kernel matrix *K* from the training data set $\{x_i\}$ using equation $\kappa(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2$, hence it avoids from solving the eigenvalues in the high dimensional feature space. Gaussian function illustrates a distance of between vector x and vector y and the standard deviation σ of the training data matrix.

The construction of kernel begins with computation of the distance matrix using Euclidean distance. A distance matrix places an important element in a kernel Gaussian function as it computes the distance of each data points that represent the data in the feature space as well as the standard deviation that governs the kernel function. The distance calculation is

based on Euclidean distance. Due to its importance and become complicated in solving multivariate problems, it is worth for brief definition and an illustration to describe of its practical use in the multivariate case problems.

An Euclidean distance between two data vectors X, which x = (x_1, x_2, …,x_n) and Y = (y_1, y_2, …,y_n) given by the Pythagorean formula:

$$d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (5.1)$$

Or it may also be represented by

$$\|X - Y\| = \sqrt{(X - Y) \cdot (X - Y)} \quad (5.2)$$

That is also equivalent to

$$\|X - Y\| = \sqrt{\|X\|^2 + \|Y\|^2 - 2X \cdot Y} \quad (5.3)$$

To illustrate the applications of this distance metric in the constructing the kernel, a simple case consists of three row vectors is demonstrated before it is used in more complicated case such in the current work.

Based on Euclidean distance, as shown by Figure 5.4, the distance is computed between the first row vector and the second row vector and from the first row vector to the third row vector until all the rows are completed. The values obtained are then stored in the first column for the E class. Once the operation using the Q class is completed, it stores the values in the second column and repeats the same distance calculation. The last column stored for the H class with the similar operation (Figure 5.5).

E = 1-by-40
Q = 1-by-40
H = 1-by-40

Figure 5.4: A simple 3 row vectors of data matrix before calculating for its distance matrix.

Dis E to E	Dis Q to E	Dis H to E
Dis E to Q	Dis Q to Q	Dis H to Q
Dis E to H	Dis Q to H	Dis H to H

Figure 5.5: The distance matrix of size 3-by-3 depicts the distance values between row vector E, Q and H.

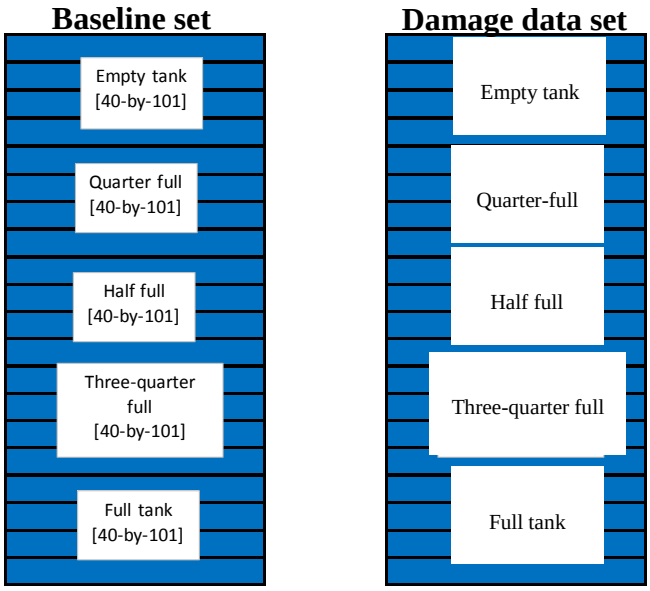


Figure 5.6: shows loading configuration for data set is grouped and arranged before Euclidean distance is computed between baseline set (left matrix) and test data set (right matrix).

Figure 5.6 shows each matrix is encompassed in a matrix of 40-by-101 (40 data observations with 101 spectral dimensions) and arranged in ascending order form empty load to full tank and follow the distance calculation.

To simplify the selection process of the parameter sigma, a distance matrix with a colour bar scale is used. It provides an alternative way to represent the inter relation of various loadings in terms of distance between various loading groups. It is also known as a similarity matrix because it compares and calculate the distance of the pair matrix for similarity. The calculation of the sigma requires details consideration from each loading set. By means of distance matrix equipped with its colour distance scale, the selection is performed based on the each block (group) of loading class. Each block represents the group of distance points in each loading class (40-by-40). It also ease the interpretation and gives summarized information in choosing the appropriate sigma values.

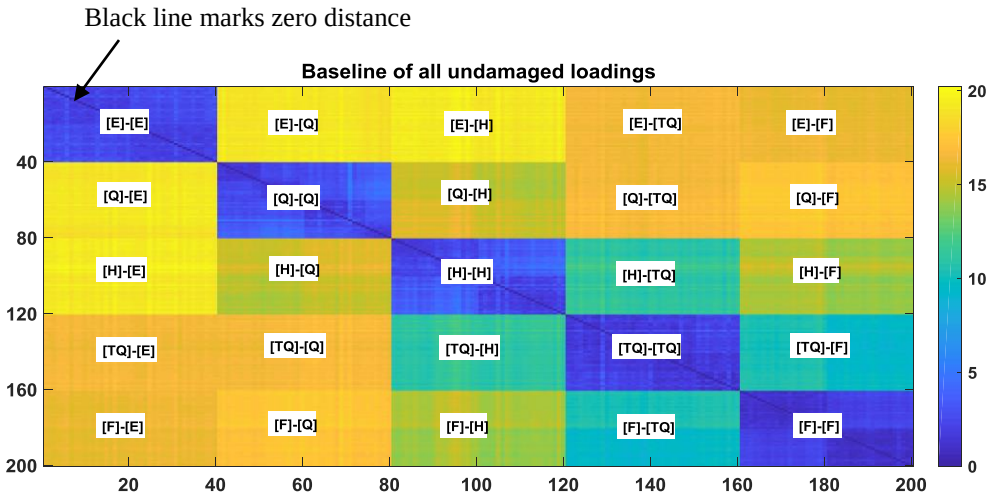


Figure 5.7: A distance matrix of the baseline set $\{L_i\} = \{E_i, Q_i, H_i, TQ_i, F_i\}$ encompasses all loadings from the undamaged condition.

The distance matrix using Euclidean distance comprises of all loading data set and arranged in ascending form, begins from empty load and up to full tank. The colour scale used in the matrix indicates the degree of distance between the pair of loading class. Pure blue regime indicates the smallest distance and pure yellow colour shows the highest

distance of data in the distance matrix. A dark line runs across the diagonal represent zero distance when the row data elements are subtracted by the same row elements in the same load class.

As expected, the smallest distance lies on the diagonal matrices of the same loading set. The highest distance occurs between the quarter full and empty load, $[Q_i] – [E_i]$ and half full and empty load, $[H_i] – [E_i]$ where i = 1, 2, …, 40 is the observations in each row of the load matrix (Figure 5.7). The black line runs across the diagonal indicates zero distance for similar data when i from one load subtracts by the same load. This distance matrix shows strong agreement with the results of the FRF plots in terms of distance of loading parameters where the data signals from empty, quarter and half-full are most far apart compared to three-quarter and full load (Figure 5.1).

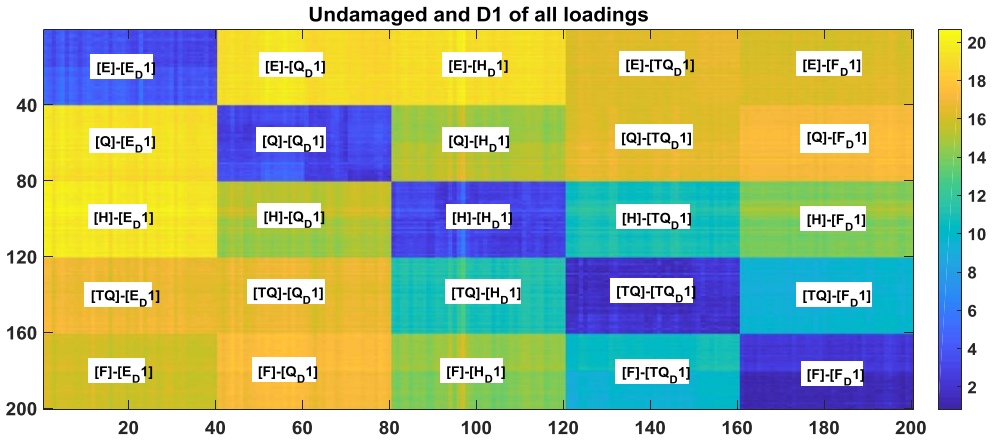


Figure 5.8: A distance matrix for pair wise distance of undamaged and D1 type test set.

The test set comprises of 4 levels of damage severities and each damage level is compared to the undamaged group separately. Figure 5.8 depicts the distance matrix between the undamaged set and D1 test data set. In comparison between the undamaged and test data set, a distance matrix is produced by subtracting the first row vector of baseline data set by every row vector in the test data matrix and the distance values are stored as the the first column in the distance matrix. The process is repeated for the next row in the baseline

set with each rows in the test data until completed. The distance matrix in colour scheme provides some summarized information of load and damage effects in term of data distance.

In the following discussion, it would be demonstrated how the parameter sigma is determined specifically under multivariate load and damage. Intuitively, the parameter sigma can be selected based on the rank of the blocks or bands of the particular test data matrix in ascending distance order. In one square band consists of 40 rows and 40 columns corresponding to the total number of measurements in one class for both load and damage class.

5.4.2 The effects of varying the standard deviation of the kernel

As stated previously, two important parameters that ensures a kernel Gaussian is effective in solving the problems are $\|x-y\|^2$ which is the Euclidean distance between the training and test data as well the σ^2, the inverse variance in the distance matrix. Both parameters are related by

$$\kappa(x,y) = \exp(-\|x-y\|^2 / 2\sigma^2 \tag{5.4}$$

After modelling and calculating the distance matrix of every load matrices in pairwise manner, it becomes more convenient to seek a suitable parameter sigma that may be used to increase the separation and variation between the data points from different damage severities groups. Intuitively, to capture the smallest data distance of two different classes, the standard deviation (σ) should be larger than the average of the minimum distance between the undamaged and D1 test data [55]. Noting that the smallest distance matrix is determined from the distance matrix of undamaged and D1 data set. Zero values is not included as they represent the distance of its own data points.

In a previous work related to face recognition [55], the parameter sigma (σ) is selected to be smaller than the inter-class distance (between training and test data set) and larger than

the inner class distance (between data points of the same class). However, uniquely for the current work, it deals with multivariate loading class and multivariate damage class. A similarity matrix plot is proposed to ease the selection of the sigma.

Emphasising that the current objective is to detect damage under the multivariate loading, thus the selection of the kernel parameters (distance and sigma) should able to improve the separation among the smaller damage groups and undamaged group. The smallest distance between undamaged and D1 of all loading classes are computed from the distance matrix and its average of the smallest distance for all the damage classes are determined.

In this study, the undamaged and test data represented in the kernel matrix is tested with some different values of parameter σ based on the values in distance matrix.

$$\sigma = \text{mean}(D_i^{NN}) \tag{5.5}$$

Where D_i^{NN} represents the smallest distance with elimination of zero distance between each data point in a row.

The first step, parameter sigma (σ) is set to be the minimum in each column in distance matrix and the average is then calculated. This strategy can highlight and increase the separation of data variables with lower variance (damage severities groups in three-quarter full and full load) as illustrated in Figure 5.12. On the other hand, using larger sigma values will result in similar results to the linear PCA. The parameter sigma selected will be used in the Gausian function as the inverse variance (Equation 5.4). This enables the separation of damage severities group to improve significantly. It capable in improving the separation between the baseline set and test data set of lower damage condition. This is very encouraging considering that loading effects can be easily detected by using linear PCA but not the case of separating damage features.

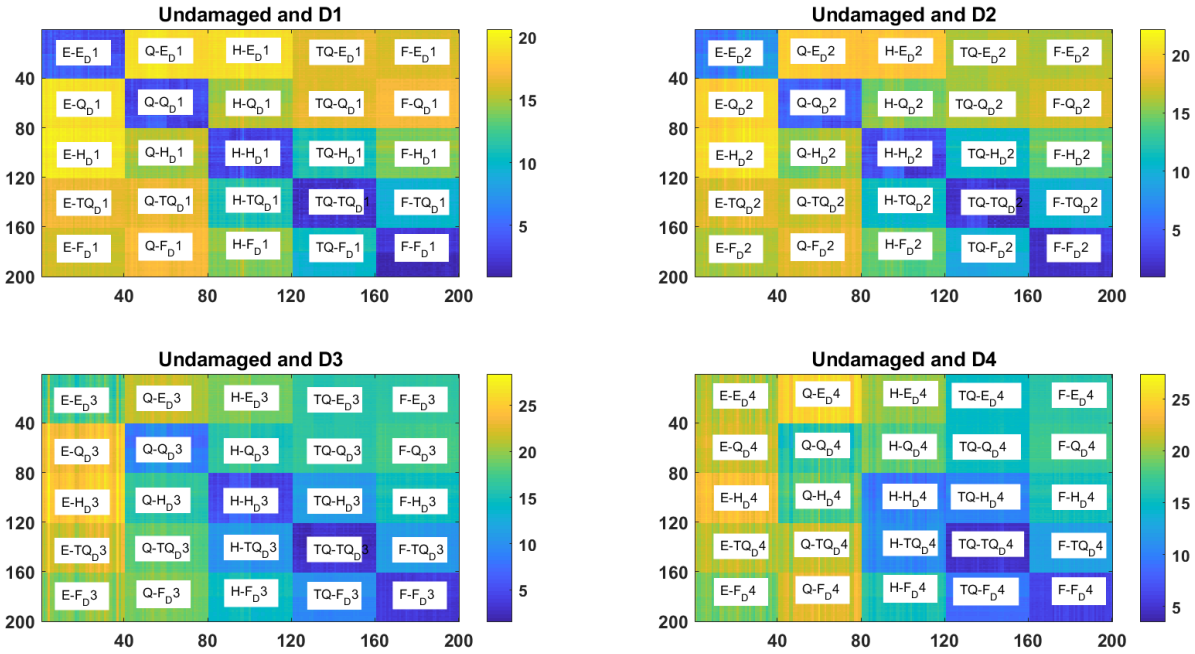


Figure 5.9: Euclidean distance between pair wise matrix of undamaged and test data corresponding to D1, D2, D3 and D4.

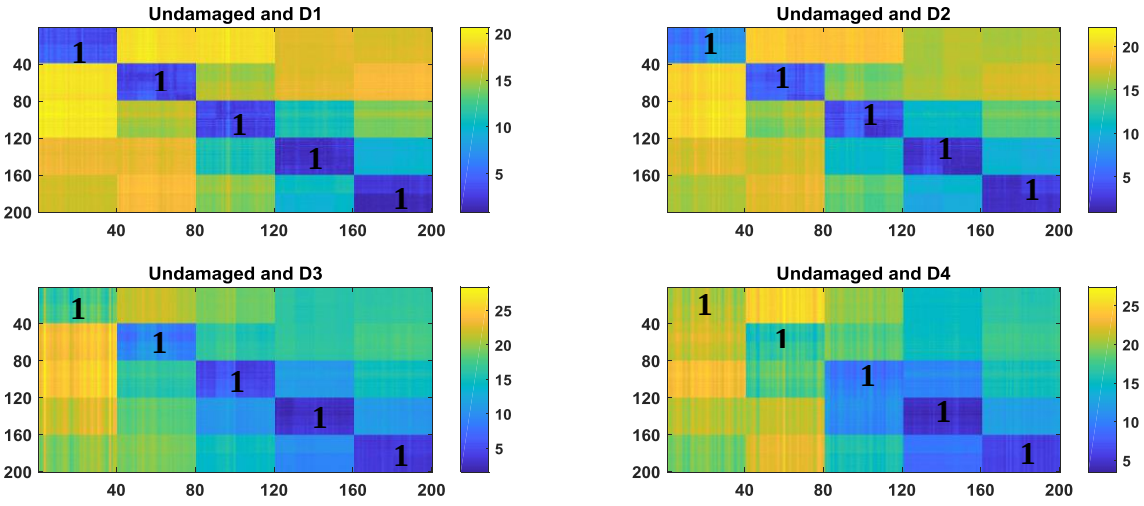


Figure 5.10: Distance matrix for undamaged and test data sets. Number one (1) indicates the smallest distance between undamaged and D1 test data of various loadings where the value of parameter sigma is calculated from.

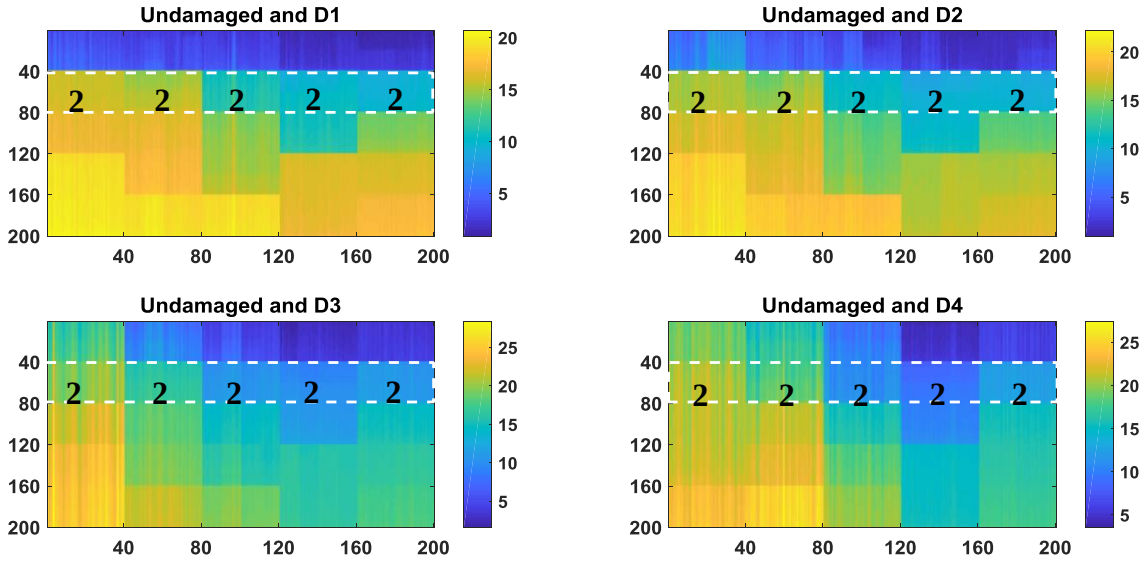


Figure 5.11: Sorting the distance matrix in ascending order.

Figure 5.9 describes a strategy of using the distance matrix in terms of colour scales representation to indicate the distance between undamaged and test data sets. It serves as an insight and guidelines for the selection of sigma parameter (σ) by focusing on the colour regime that denotes the minimum distance.

Sorting the distance vectors in ascending order allows easier selection of sigma parameter corresponding to the lowest distance based on the block of group data in one class (Figure 5.10). Selecting the parameter sigma with respect to the second lowest distance of distance matrix (Figure 5.11) creates different projection that allow separation of classes of larger distance (between loading classes) compared to the sigma of the first block. The sigma values shown to have a profound effect on the damage detection and data separability.

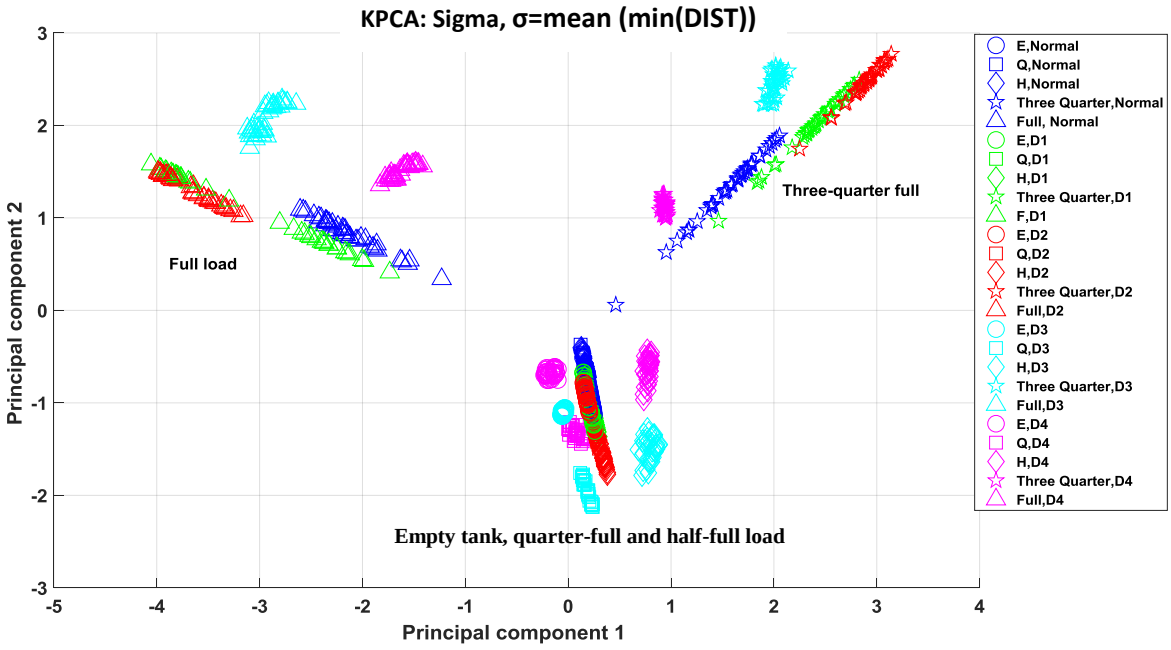


Figure 5.12: The result of using an average of the minimum distance calculated from the distance matrix.

Figure 5.13 displays the zooming in the three-quarter full class that clearly illustrates the pattern of the data variables corresponding to respective damage severities groups.

Figure 5.14 displays a zoomed view on high-density data groups using the average of the smallest distance of sigma value (σ) and the result shows no data overlapping including for groups with smaller dimensions which are group UD, D1 and D2.

The result in using the smallest parameter sigma (marked as param on figures) in the kernel function indicate that the damage separation among the damage severities groups could be achieved. From Figure 5.15, the smaller variability in three quarter full and full load causes the feature to be more separable as compared to the empty, quarter-full and half-full loading class. Subsequently, an important justification can be made based on this observation with respect to the sigma value selection.

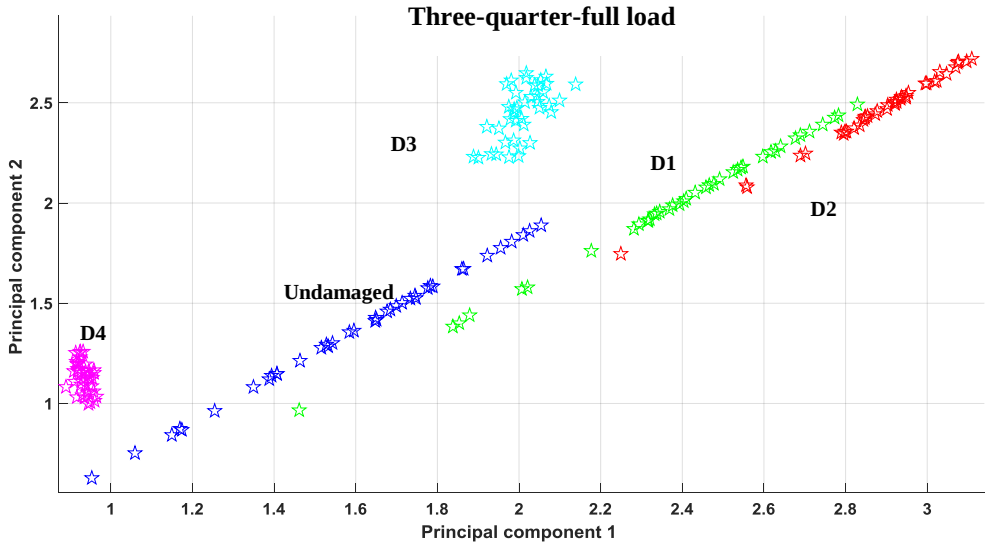


Figure 5.13: A zoomed view inside three-quarter-full load by the same parameter.

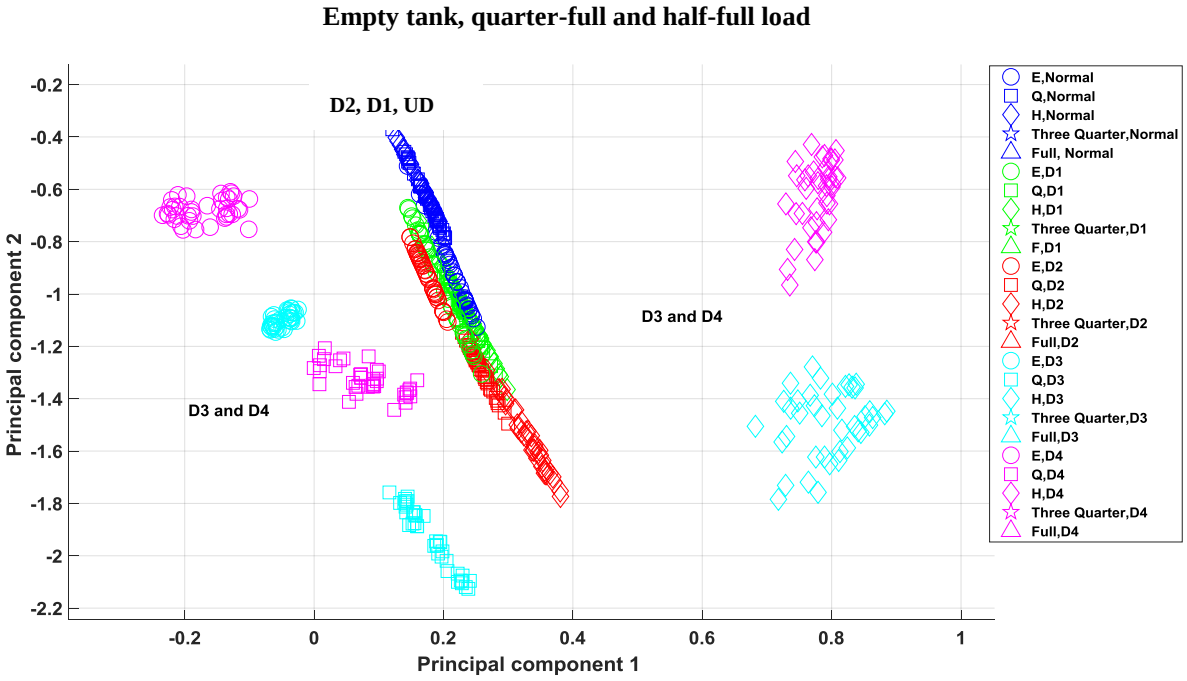


Figure 5.14: A zoomed view on highly dense data set using the same sigma parameter.

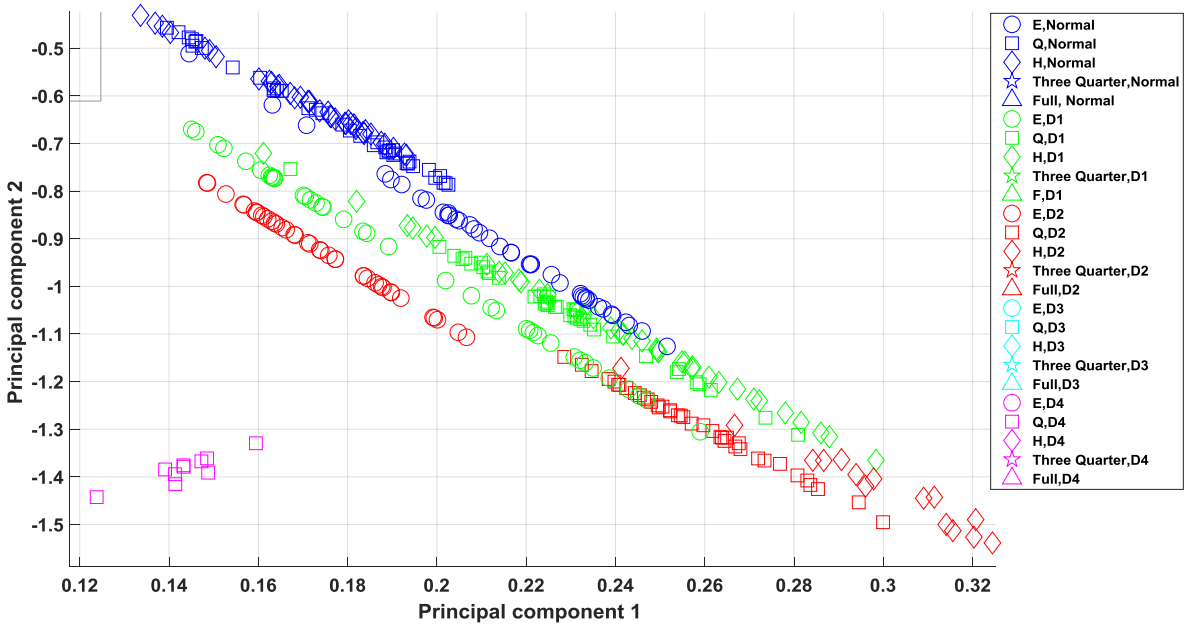


Figure 5.15: A further zoom into the high-density data groups of the Figure 5.15.

Zooming into Figure 5.14, reveals that despite the loading effects are not clearly separated on the first level, the projection manages to highlight the separation among the damage severities groups after zooming in the region (second level of feature resolution, Figure 5.15). On this level feature space, Figure 5.15 reveals the zoom-in of the variables of three-quarter full load showing distinct separation of damage variables. Next, the parameter is changed using the second block of distance block to identify any changes from the increasing the sigma. In this case, sigma (σ) from the block 2 in the distance matrix can be used (Figure 5.11). The result shows a different perspective where the projection manages to encompass all loading conditions on the feature space at the same level with good indication of damage severities (Figure 5.16).

Some underlying remarks can be derived with respect to the sigma selection. Some factors to consider in choosing parameter sigma:

- the parameter sigma (σ) can be considered from the second lowest distance block in the distance matrix based as shown in Figure 5.10 highlights the overall loading conditions. Block number 2 indicates the sigma value to be determined. This sigma can highlight complete features of both loading and damage variables on one

- projection with the capability to track the structural health conditions based on particular loading conditions (Figure 5.16).
- It reveals that, as the damage severity increases, the variables tend to move towards a focal point away from the baseline variables resembling an ellipse pattern (Figure 5.16). For normal condition, the variables from various loading classes form a large ellipse shape ascending in load in a clockwise rotation.
- Using the average of the smallest distance as in block 1, data trajectory of all loading conditions is not possible to be visualised and monitoring the changes of structural conditions are restricted. However, it is useful in novelty detection test described later in Section 5.6 with the aim of detecting the smallest damage.

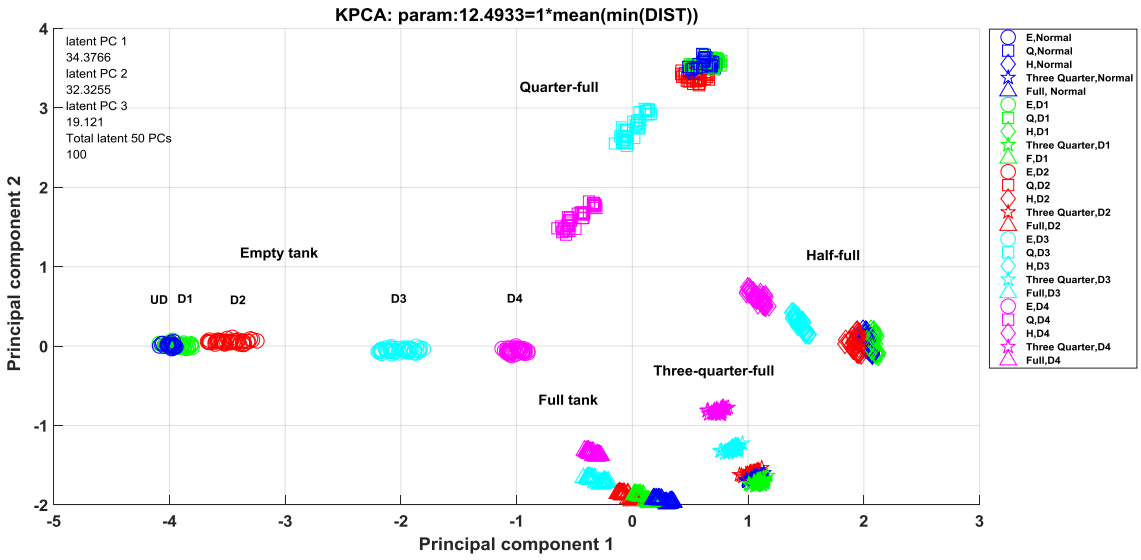


Figure 5.16: Result from using the average distance values from the second distance block as shown in Figure 5.11.

The results so far highlight the advantage in using kernel Gaussian PCA in separating the damage variables from the undamaged variables. The important parameter in the analysis is the parameter sigma (σ) which refers to as the average of the smallest distance computed from the distance matrix. The selection is dependent on the sequence of block inside the distance matrix where the smallest distance located in the first block (40-by-40 matrix)

inside the 200-by-200 distance matrix is considered. The distance measures the Euclidean distance between undamaged (baseline set) and a particular test data is generally considered.

As it is shown, the smallest distance between the two sets causes a separation between three-quarter and full load and their damage severities associated with the loading conditions at the first level. For separation between empty, quarter-full and half full, it requires a more detailed or deeper level of focus to see the separation of these classes including the associated damage severities. This characterised as hierarchical visualisation which will be covered in Chapter 7 using decision trees.

In relation to that, KPCA can generalises the different level of separation by choosing a slight higher sigma. This has a strong potential application in monitoring of structural conditions and damage progressions as demonstrated in Figure 5.16. For the purpose of monitoring, the second lowest block (row 41 to row 80 of the 200-by-200) in distance matrix has shown to provide complete overall data trajectory patterns of baseline and damage features. In the following chapter, the capability of kernel PCA to separate between undamaged and damage variables from loading variabilities is further investigated and tested via various principal component models.

5.5 PC models for discriminating loading effects

It is desirable to determine if the effects of the operational loading can be isolated from damage and to ascertain the extent of the damage severities. To achieve these goals, PCA models are built from using a reference data set, which is reconfigured according to four possible cases, named as PC model A, B, C and D. Initially, standard PCA is applied on the reference set using eigenvalues decompositions. From the PCA computation, a loading matrix (transformation matrix) is applied to the test data set and projected into the first and second principal component projection (sometimes known as score 1 and score 2). The type of training set is described under each section of the PC models. Feature projection of the test data is desirable to be variant from the reference or baseline feature set.

The separation and orientation of the test feature with respect to the feature model is the desired goal as it potentially provides an indication for presence of damage.

5.5.1 PC Model A

PC of Model A is established by incorporating an undamaged condition comprising all operational loading conditions as its baseline data set. The model is used to provide an undamaged feature in comparison to some test variables produced from different structural conditions. A separation from the baseline data path could highlight some indication for a possibility of damage in the structure. On another note, a comparison with the baseline feature in the principal component projection space heavily depends on the quality of data signal used and it should be cleansed from any unwanted outliers and excessive noise before applied with the principal components of the baseline model.

The model A is viewed as the most likely case that can be available in real scenario given that the undamaged sets can easily be obtained and recorded before damage occurs. This model is motivated by the fact that, in general, data sets from various damage severities are not always available. In this context, a test data can assume a structural health condition with unknown damage severities. The test data is applied with the principal components of the baseline and compared in terms of separation between the features of the baseline and test variables.

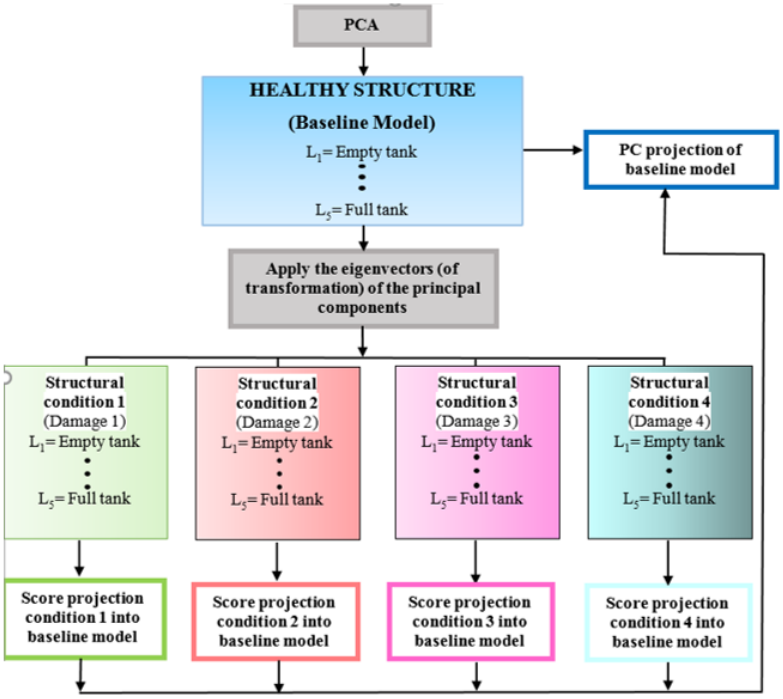


Figure 5.17: Model A: PCA is performed on a healthy structure comprises of complete operational loads

The principal components of model A are computed based on the eigenvalues decomposition of the model covariance matrix that encompasses of overall loading conditions $L_1, L_2, \ldots, L_5$ corresponding to empty load, quarter full and up to full load of undamaged structural condition (Figure 5.17). The first object to compute is the eigenvectors and then ranked them according to the highest eigenvalues in descending order and apply the principal components on the test set from various damage seventies. Any deviation from the baseline data path highlight a potential damage.

By establishing this model, the covariance, which is comprised of the baseline data set, is dominated by variance from the loading effects and no influence from the damage variables. The key question is, can the PCA manage to provide some indication of damage presence in the test test data set. The projection results from the linear PCA show the data

set is distinctly separated in terms of loading severities, however the damage severities seems to overlap between undamaged and lower damage groups (Figure 5.18).

The kernel PCA significantly outperforms the linear PCA in highlighting and separating different damage severities. For loading class identification, linear PCA is sufficient to perform the task. For damage class identification especially for the variables that lie on smaller dimensional length for instance D1 and D2 variables, nonlinear PCA has the capability to perform.

Noting that, the main purpose of establishing PC model A is to monitor the structural health conditions by analysing the separation between the baseline variables (undamaged variables) and the damage variables. Kernel PCA is shown to fulfil this expectation as highlighted in Figure 5.19.

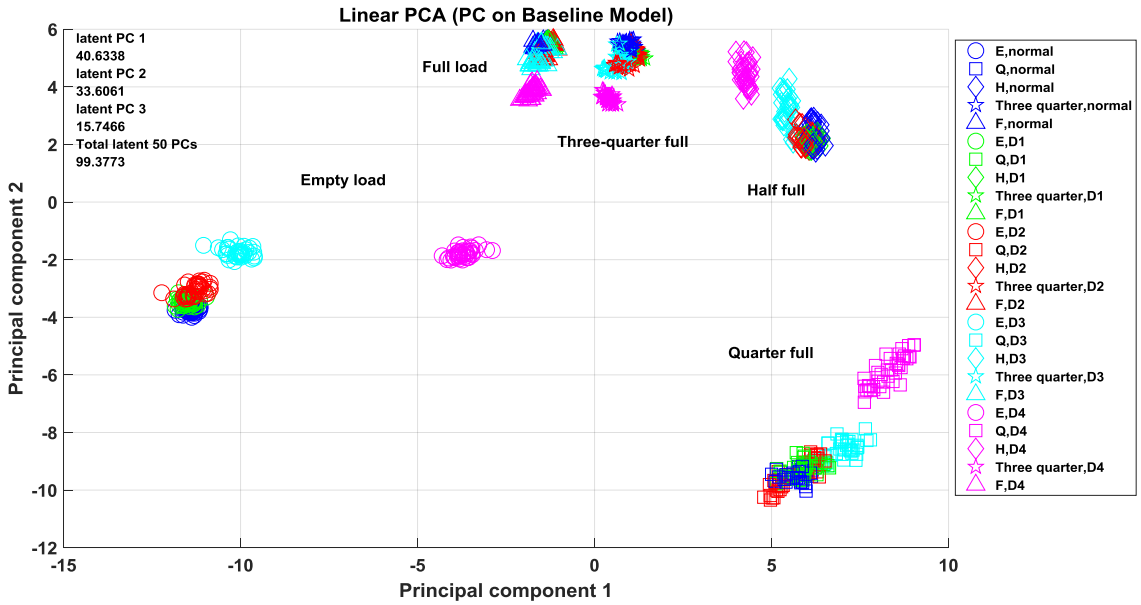


Figure 5.18: The linear PCA using eigenvalues decomposition of a covariance matrix of whole observation sets from undamaged loading conditions.

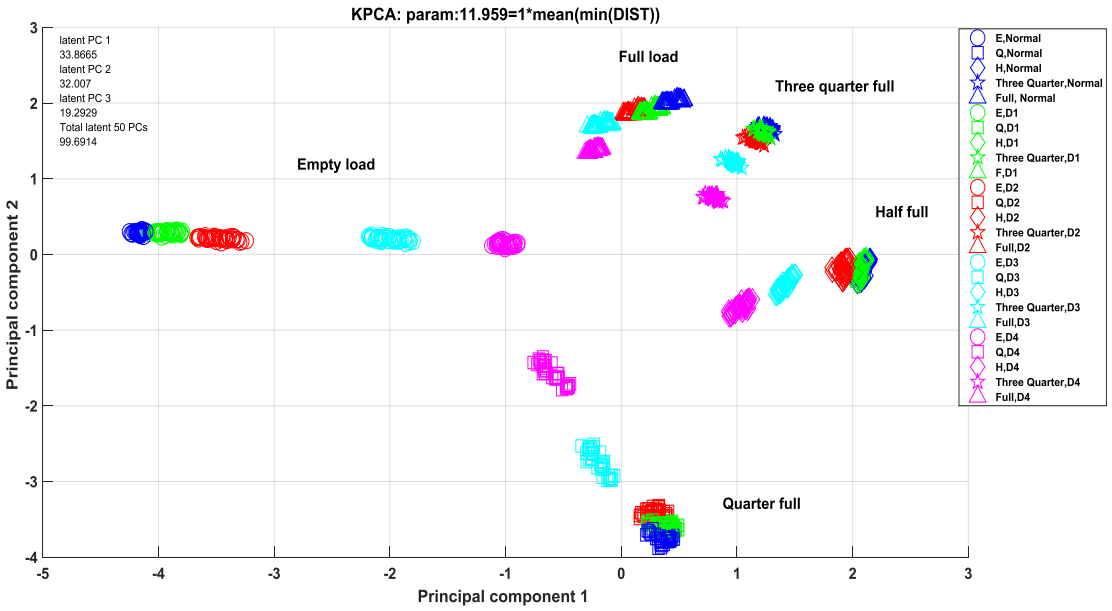


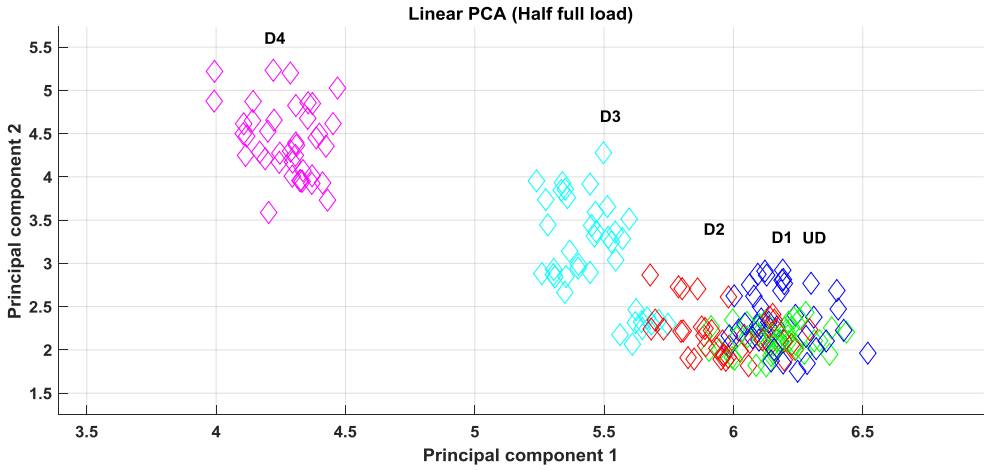
Figure 5.19: Kernel PCA applied on the similar baseline model to the linear PCA.

Deeper analysis on the results from the linear PCA (Figure 5.18) and kernel PCA projections (Figure 5.19) highlight that kernel PCA is clearly provide a model basis for tracking the data pattern to monitor the changes in structure's conditions. Whereas in the standard PCA, it is almost impossible to find a trajectory pattern for monitoring damage conditions for the whole loading system in form of close loop.

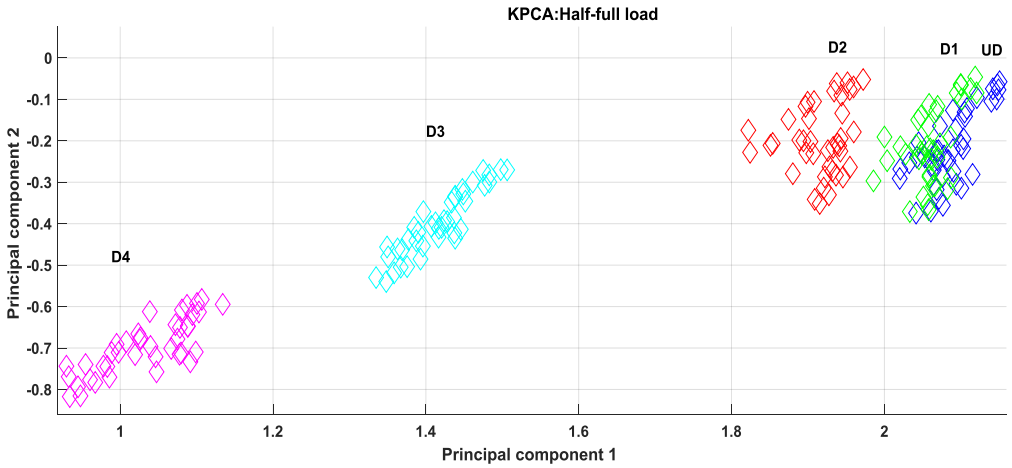
Focusing on half-full load loading class (Figure5.20 (a)) in order to examine the data separation, the standard PCA fails to distinguish between undamaged and the small damage classes (D1 and D2). In contrary to linear PCA, kernel PCA (Figure5.20 (b)) shows a trend of damage classes in correlated with damage severity levels.

The nonlinear PCA manages to separate all damage features in all loading class as shown in Figure 5.20-Figure 5.21. In many cases, linear PCA is unable to distinguish and separate the undamaged data from the lower damage class of D1, D2 and D3.Only in empty load, the separation of damage data from the undamaged data is distinct.

Contrary to the linear PCA, kernel PCA significantly improves the visualization for loading clusters as well as all level of damage classes inside one projection space. Computing the distance matrix and the parameter sigma at each data set of different structural condition, it gives a leading advantage by having a localised data model at each structural state condition by using the Gaussian function. A zoomed view at each loading state (Figure 5.20-Figure 5.21) also reveals a unique data pattern in each damage set as they moving away from the undamaged data.

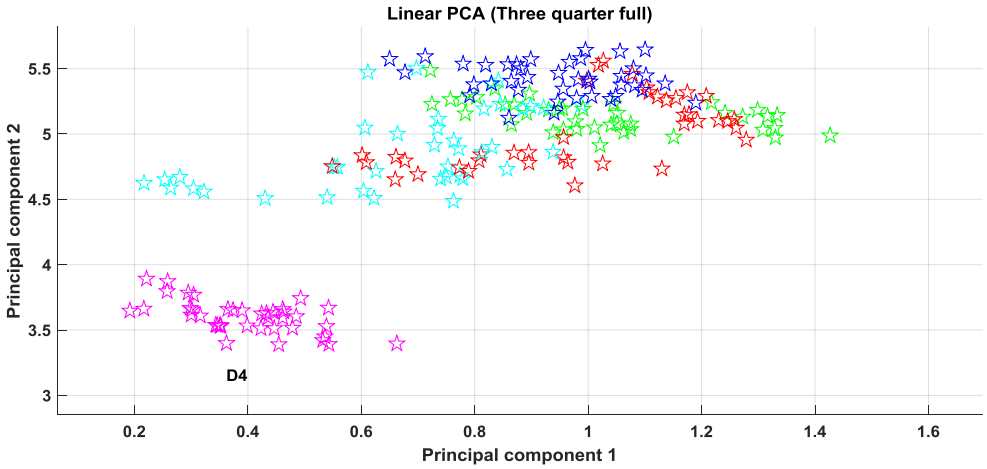


(a)

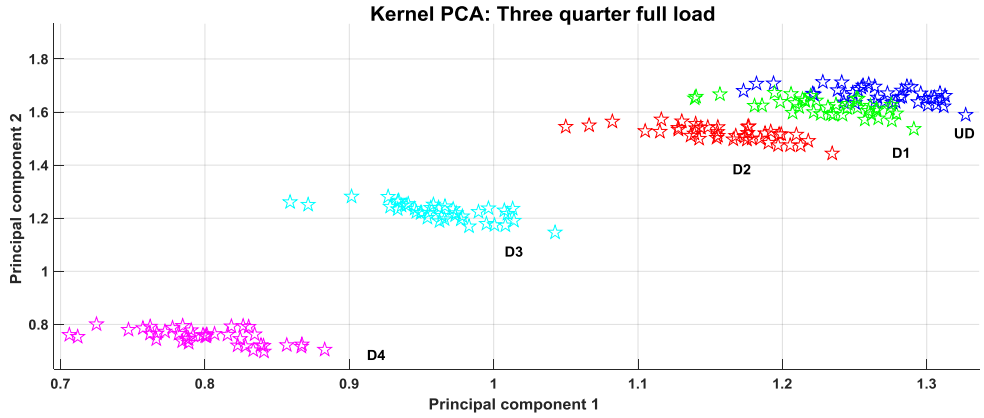


(b)

Figure 5.20: (a) A zoomed view based on Figure 5.18 to visualise the projection of linear PCA with respect to damage class separation. (b) Kernel PCA projection based on the same loading class from the overall projection of loading classes shown in Figure 5.19.



(a)



(b)

Figure 5.21: (a) The result of standard PCA after zooming three-quarter full load in Figure 5.21. (b) Kernel PCA transformation result on the zoomed view of three-quarter full load obtained from Figure 5.19.

Highlighting on three-quarter full load in Figure 5.21(b), the results from kernel Gaussian PCA, display a great performance in terms of damage classifications and data pattern recognition in the interest of tracking damage level in that load. Kernel Gaussian PCA successfully establishes a tracking pattern for the loading system. Linear PCA is found to be struggle separating the damage classes corresponding to their damage level 5.21(a).

In PC model A, the use of undamaged condition as a baseline set proves to give an advantage to a SHM model. Damage class separation between the undamaged condition

and small damage conditions is achieved through sigma taking the second lowest block in data distance.

It is shown that for the case of damage monitoring of a loading system, the sigma values in the kernel function $K(x, y) = \exp\left(\frac{-\|x-y\|}{2\sigma^2}\right)$ plays an important role in making the damage variables separable. Ideally, sigma should be chosen to be the smallest data distance in order to capture the separation between the undamaged and the smallest damage. However, in the case of an overall loading system, tracking the structural damage states is desirable and by using the second lowest sigma (from the second lowest distance block in the distance matrix (illustrated in Figure 5.11). In summary, PC model A forms a promising damage monitoring under operational loading system in the use of kernel Gaussian PCA by adapting the distance function and the inverse variance to each damage conditions (test data sets) that make kernel technique is localised with respect to each each structure's condition.

5.5.2 PC Model B

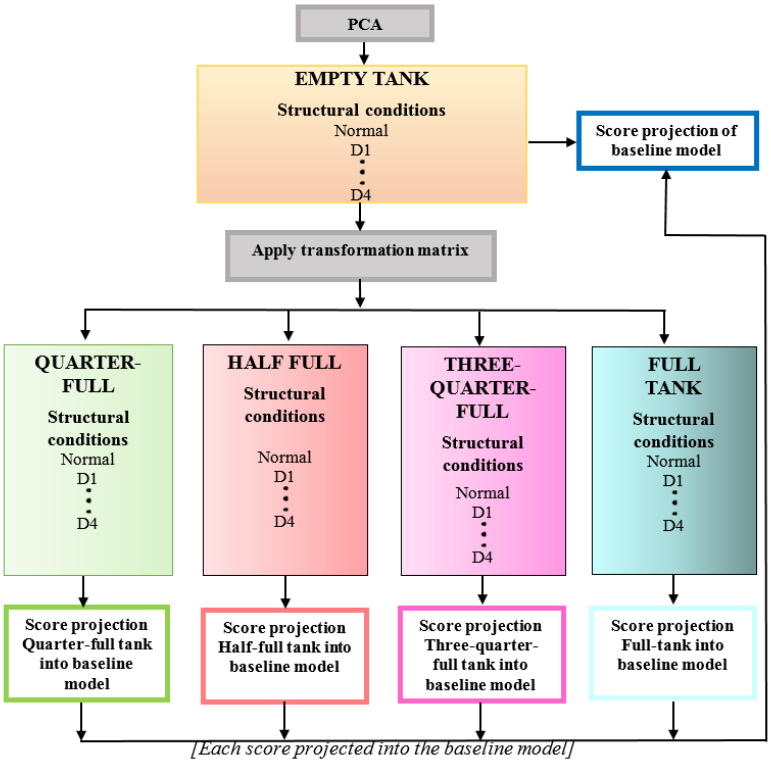


Figure 5.22: PC model based on the empty tank associated with various structural conditions as a reference set.

In model B, the study attempts to focus on the relationship in terms of data distance between the damage and undamaged variables at each loading state. It aims to provide a PCA model that is more sensitive to damage variation rather than the loading variation. It has been illustrated in the previous PC of model A, linear PCA unsatisfactorily distinguish the damage variables from the undamaged variables. This model tries to ascertain if linear and nonlinear PCA can discriminate the loading effects and focus more on the damage severities variations in the feature space.

To achieve this, an empty tank load comprising all structural conditions is incorporated as the baseline and the eigenvalues decomposition is performed (Figure 5.22) and applied any loading conditions having about similar damage type. One main concern, is that if this model has realistic values on real application of SHM. To illustrate a practical case for

aerospace maintenance, for instance some aircraft components are known to suffer from a typical type of local damage (crack) in the wing box centre section at the shear fittings area [71]. One of the strategy to implement PC model B that is to collect the data under these damage severities for data monitoring SHM. Under one loading condition, data from several damage states can be stored as a reference data set for comparison to a new test data set independence to various loading.

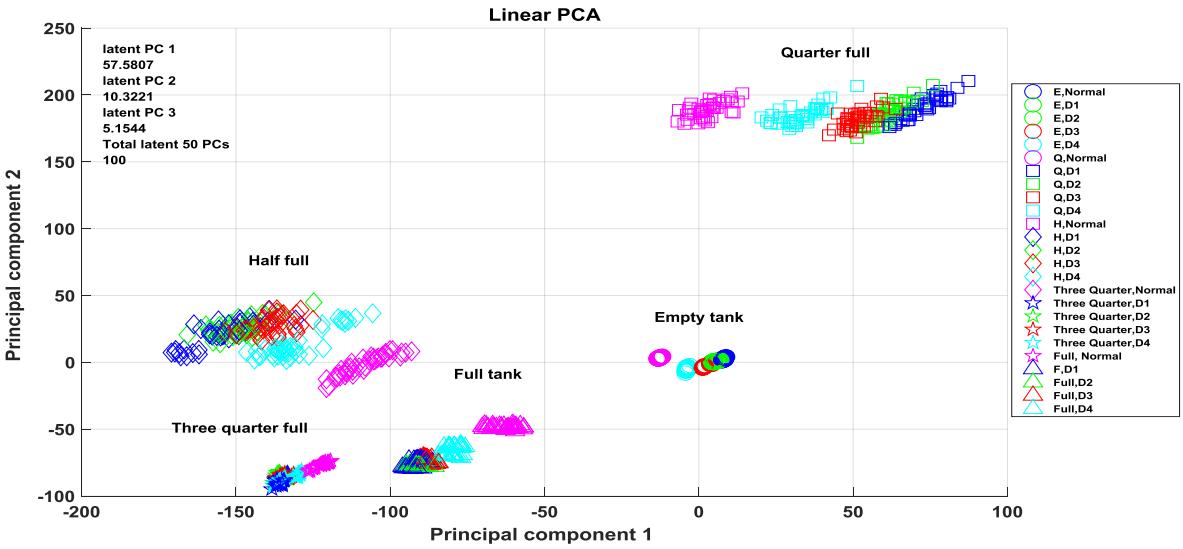


Figure 5.23: The transformation of data set by the linear PCA using damage sensitive principal components.

What is apparent based on Figure 5.23 is that the separation of damage severities are more distinct for empty and quarter full load than the half-full load and higher. There are high data overlaps within the undamaged and smaller damage severities (D1, D2 and D3). Empty tank seems to be the only loading class that has recognisable data separation.

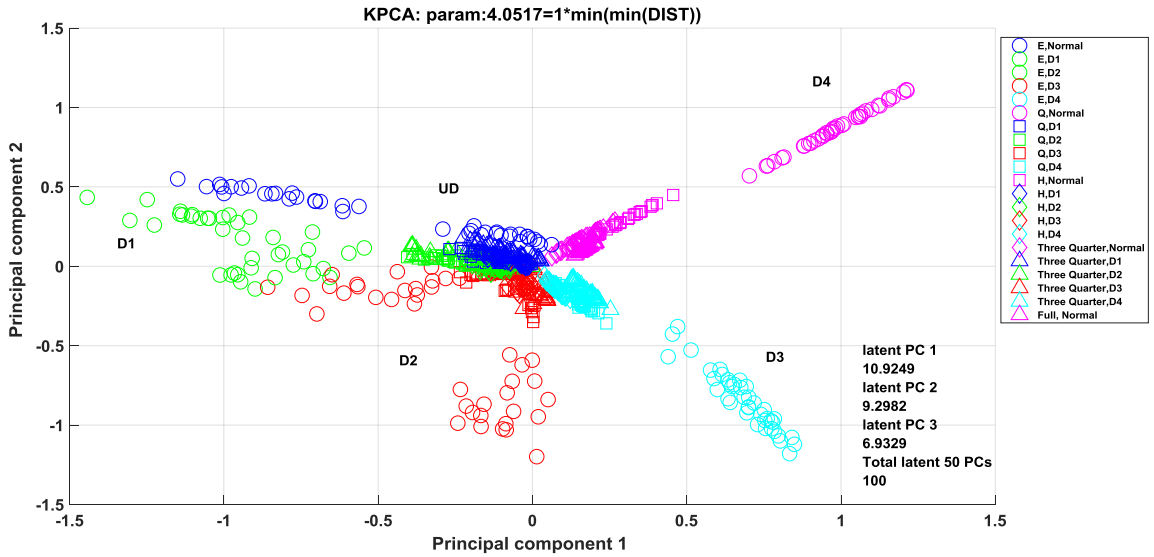


Figure 5.24: The result of kernel PCA using PC model B using damage severities as reference.

Results from Figure 5.24 reveals a very interesting finding in regards to damage severities and overshadow the loading variables. The variables are ordered in correlation with damage level, in a counter-clock wise pattern. Each damage severities group transformed in its own region with smallest data variation concentrated on the centre of the space and sets with larger data variation lie away from the centre point. Empty loading class has the largest data variance and transformed most far from the centre space. The loading groups from half-full load and and higher are concentrated nearer to the focal point (intersection point), a point where the smallest data variability can collapse.

Figure 5.25 highlights the region near the focal point where the higher loading class are more densely concentrated. It is very clear that the each damage severities are transformed corresponding to their data groups regardless of their loading class. This model effectively discriminate loading conditions from the damage features.

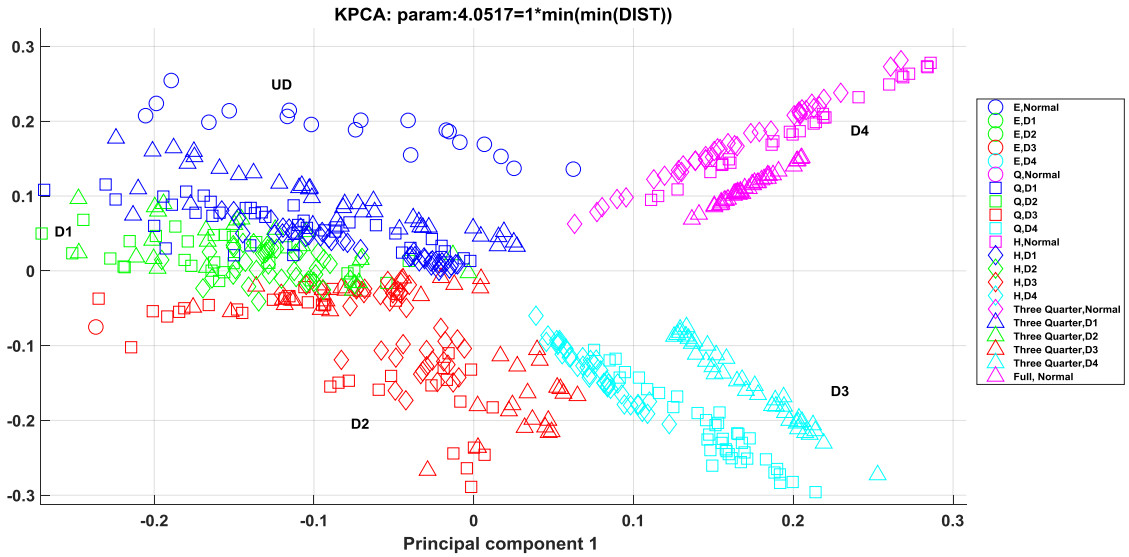


Figure 5.25: A zooming in PC model B based on damage severities of empty loading class.

For this PC model, for sigma parameter selection, the smallest distance block (block 1) from the distance matrix as described in Figure 5.10 is chosen. The smallest distance of the sigma allows the smallest distance between the data points in each damage class to be detected and used in the kernel function for calculating the principal components. Note that, the amount of variance in the first, second and third principal component in the KPCA plot is very small (Figure 5.29) compared to the kernel principal component model A is more than three times larger (Figure 5.21)

Again the main advantage drawn from kernel PCA is that each kernel is generated at each loading set consists of various damage severities. Therefore, the localised kernel functions acting on each individual loading set and principal components were computed based on the loading set's kernel. In the application aspect of this model, any new data can be applied on the principal components of the training model and identify the severities of the damage variables from the partitioning of the damage severities.

Some downsides from this model are that the training data established as the reference set has to include all possible damage types and severities. The author sees this model has a potential place in application for structure where a particular damage is typical known to

occur. In many engineering field including in aerospace, common type of damage is known. It can discriminate the dominant effects from another parameter effectively and gives enlightenment on more important parameter such as damage in the context of multivariate data analysis and damage identification in SHM.

5.5.3 PC Model C

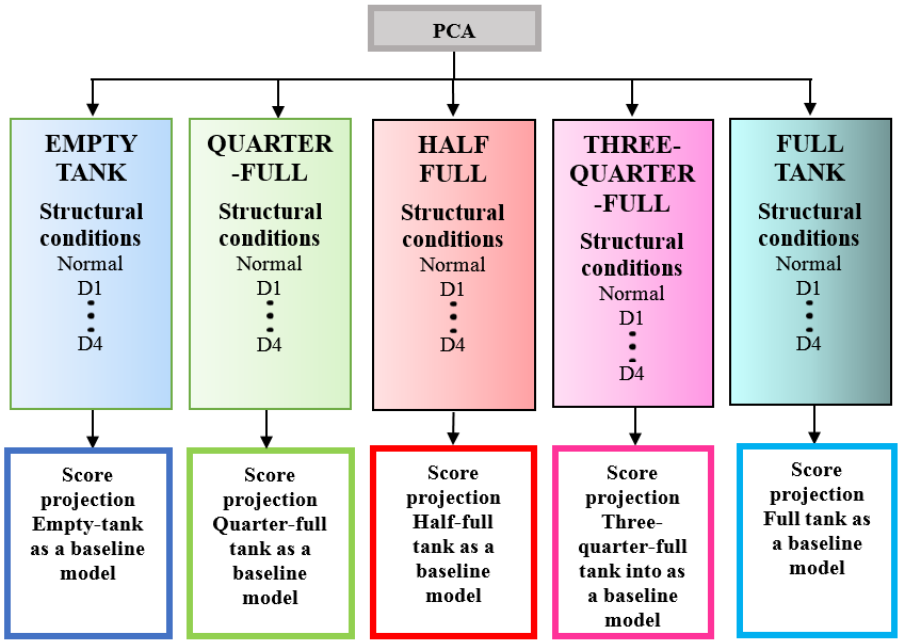
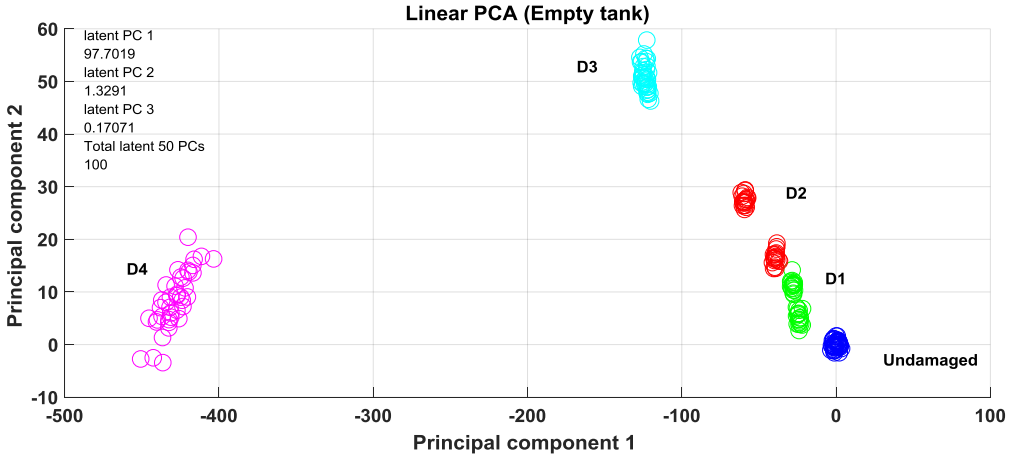


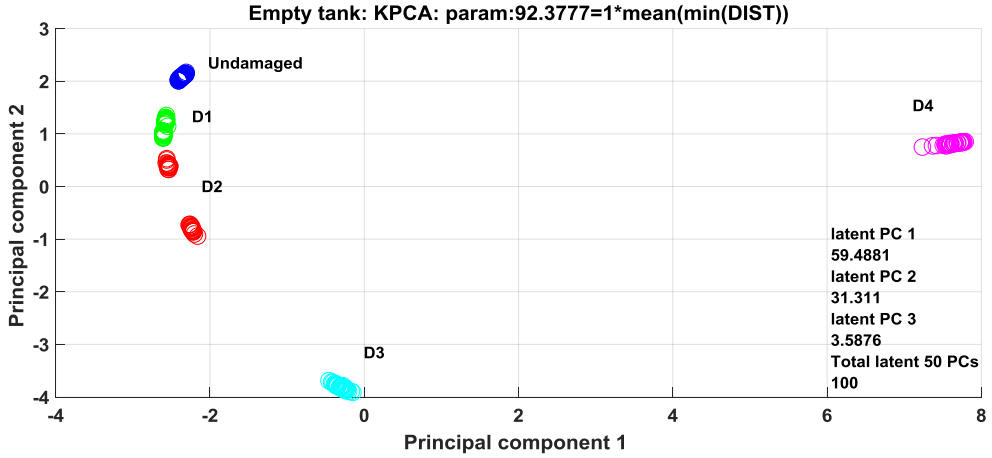
Figure 5.26: Individual baseline model of each load tank in focus of damage monitoring

The motivation behind this model is to provide a reference mapping for each loading and its structural conditions, so that it can be used as a comparison for other loadings and their structural conditons. To realize this, principal components are computed form each loading set including its damage severities occur during that loading (Figure 5.26). The main advantage of this model is the effects from the loading parameters are discriminated naturally and leaving only theeffects of structural conditions (damage-sensitive feature) in the projection space. This model should be able to capture better data separation and

damage identification given that it has only one parameter influences the eigenvalues decomposition.



(a)

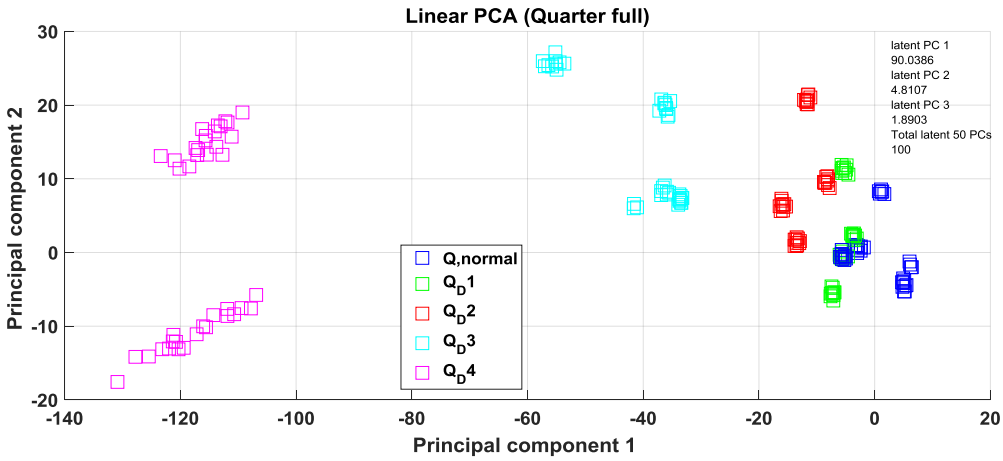


(b)

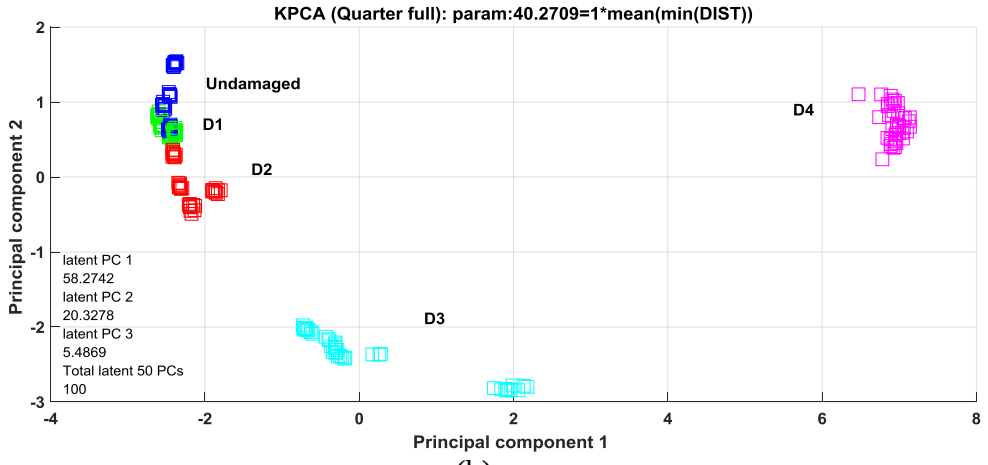
Figure 5.27: The comparison between (a) linear PCA (b) and kernel PCA for empty load.

The kernel PCA projection in Figure 5.27 shows the result by selecting the sigma where it considers the second lowest distance block (row 41-80) in distance matrix as the same procedure shown in Figure 5.11. It produces distinct track of data pattern, establishing a curvature path that can be used to monitor the structural conditions. It is observed that, for linear PCA in Figure 5.27-Figure 5.28, only straight data path is established and tracking the path of damage severities is not very convenient. This is one of the key feature of kernel

PCA that allows principal components projection on nonlinear projection trait [34], [52]. This characteristic brings strong benefit in monitoring structural damage conditions that indicates to the changes of structural health condition by looking from the data position on the track path.



(a)



(b)

Figure 5.28: Using the particular quarter-full loading data set (a) Linear PCA (b) kernel PCA

In the current model (Figure 5.28), the main interest is acquire the track of the data pattern so that the damage severities can be monitored. On other hand, choosing the smallest distance block from the distance matrix causes the focus on separation among the undamaged, D1 and D2 data variables which have about same data variance size (Figure

5.29). D3 and D4 data are not affected by the sigma because it needs larger sigma to highlight the separation of the larger damage variables.

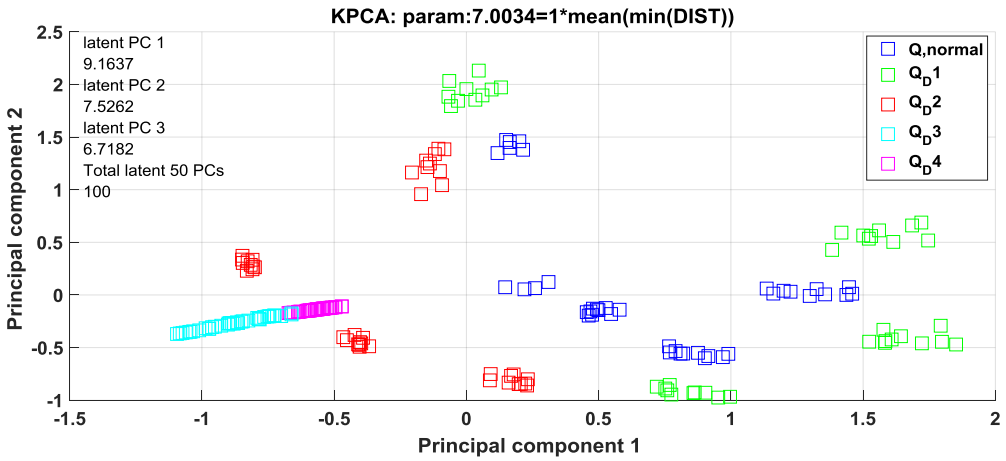
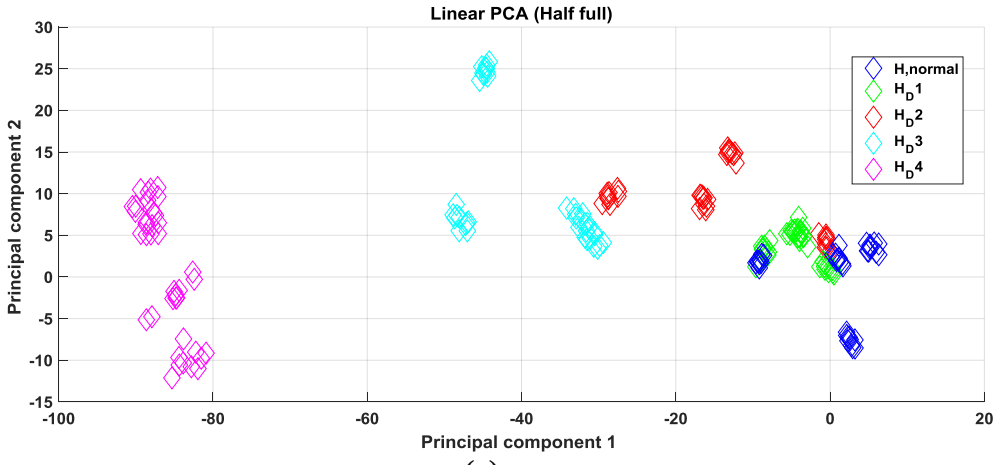
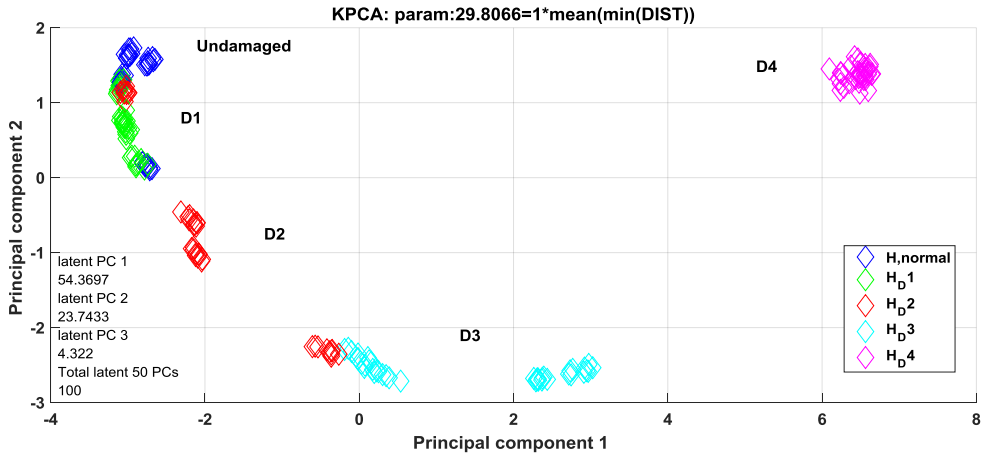


Figure 5.29: Using the sigma from the smallest distance block.

In another aspect, it successfully widen the separation between the undamaged and the smaller damage variables (D1 and D2) which is not possible by linear PCA.



(a)

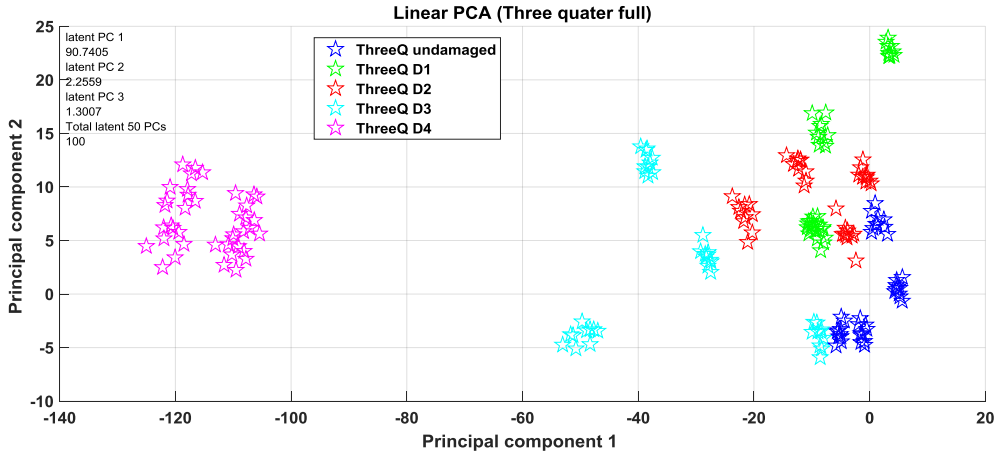


(b)

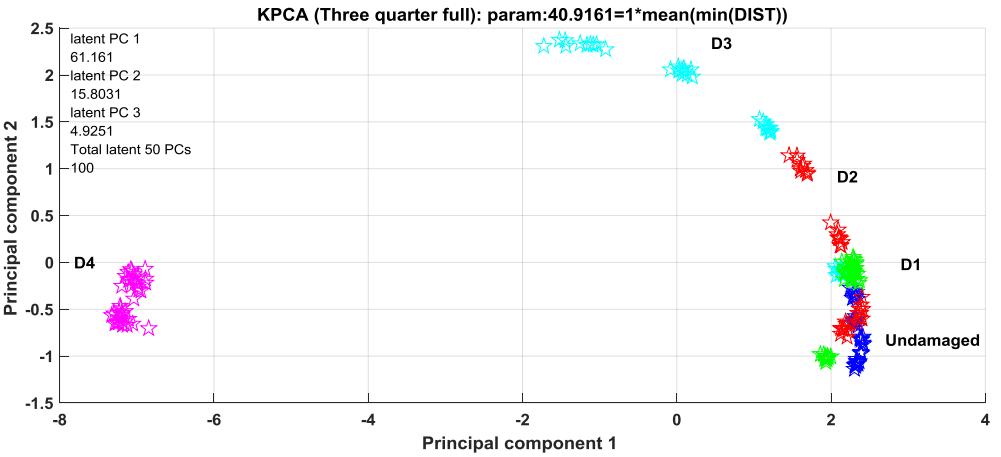
Figure 5.30: Half-full load (a) linear PCA and (b) kernel PCA

Figure 5.30 also shows the similar behaviour when using the sigma values from the second block of the lowest distance from distance matrix. It is identified that, if a higher sigma is adapted, a clear pattern of data track is not achievable.

So far, it is identified a smaller sigma enables a better capture of data sets that has smaller variance as in this case it refers to undamaged, D1 and D2 data sets. The block of distance of the second lowest distance or the middle distance block are identified can give good data tracking along the curvature. Noting that, a few misclassified data points from undamaged and D2 data points into D1 class (Figure 5.30) originated from the actual FRF data.



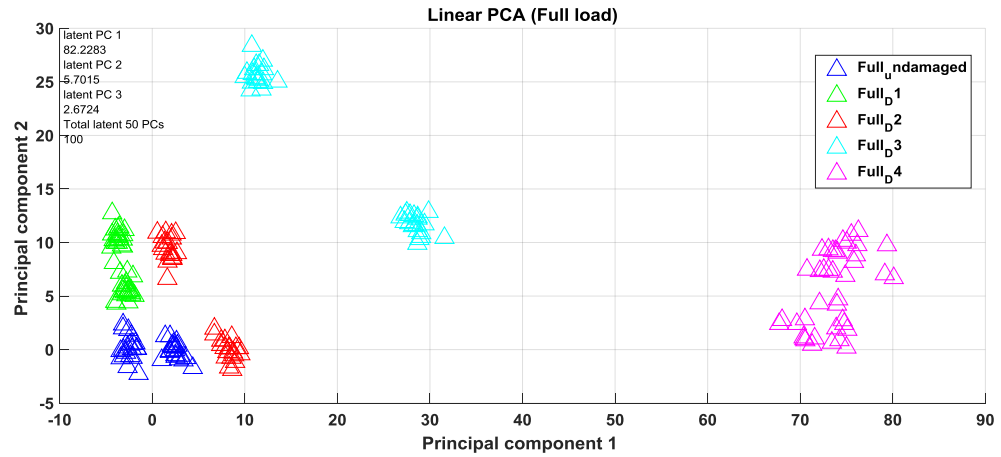
(a)



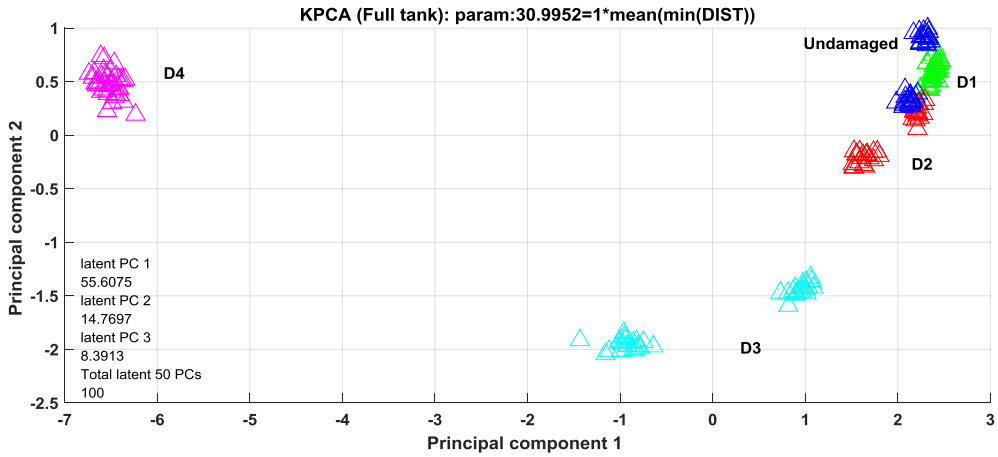
(b)

Figure 5.31: Three quarter-full load: (a) standard PCA (b)Kernel PCA

Figure 5.31 shows three-quarter full load data track from using a sigma from the middle block (row 81-120) of the distance matrix. Noting that some overlapped green and red label data points within undamaged class are preserved from the original FRF features. Linear PCA (Figure 5.31(a)) shows greater data overlap with no possibility of damage tracking.



(a)



(b)

Figure 5.32: (a) Transformation by linear PCA (b) kernel PCA

The full tank load also exhibits the similar pattern for data track with its sigma values calculated uniquely for its data sets at this loading condition with four different structural conditions (Figure 5.32). Again, sigma is computed from the mean of the middle distance block in the distance matrix in ascending order.

The PC model C highlight its potential use when some damages occurred in a particular loading and the aim is to identify damage through visualization using PCA. Using kernel PCA, it is shown that a damage monitoring system by establishing its data track, the damage severities can be traced along the curvature. In term of practicality, this model has one

drawback which is gathering the damage data set as its training set. However, it has been explained in PC model B where uses various damage data as a training set, it is useful for an application that certain damage are known to occur especially in aerospace maintenance engineering[14]. Next, another PC model is explored where in this case it assumes all loading and damage data as its training set.

5.5.4 PC Model D

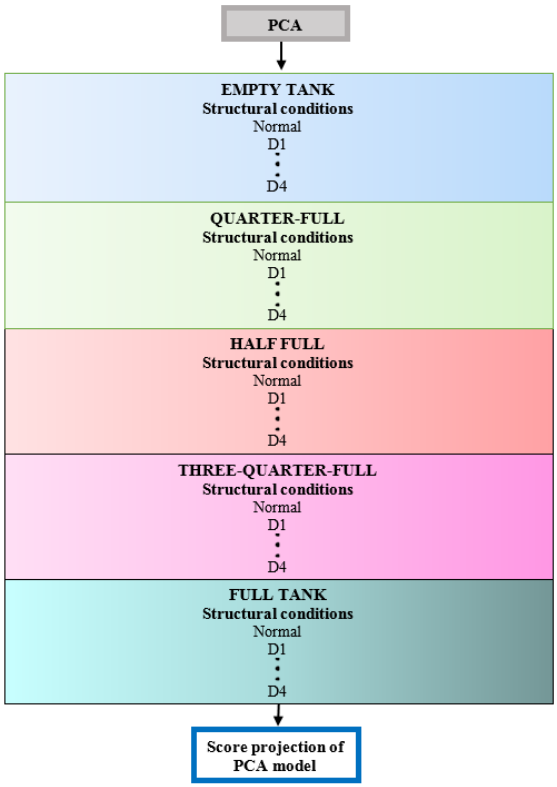


Figure 5.33: A baseline model built from complete operational loading associated with normal and various damage severities.

This model (Figure 5.33) establishes all data sets including the loading and damage data as a training set. It has been shown that how the effects of loading or damage on model PC A and model B respectively in computing the principal components.

The training data set of model D is 1000-by-101 is then standardized on the matrix before performing the eigenvalues decomposition. As mentioned previously, a distance matrix is established to analyse the block that can be used as a candidate for the parameter sigma computation. In the linear PCA projection space, we observed that much of the overlapping occur within undamaged and the lower damage level. Moving into this direction, the block with smallest distance is extracted to enable a better capture of data in the lower damage class.

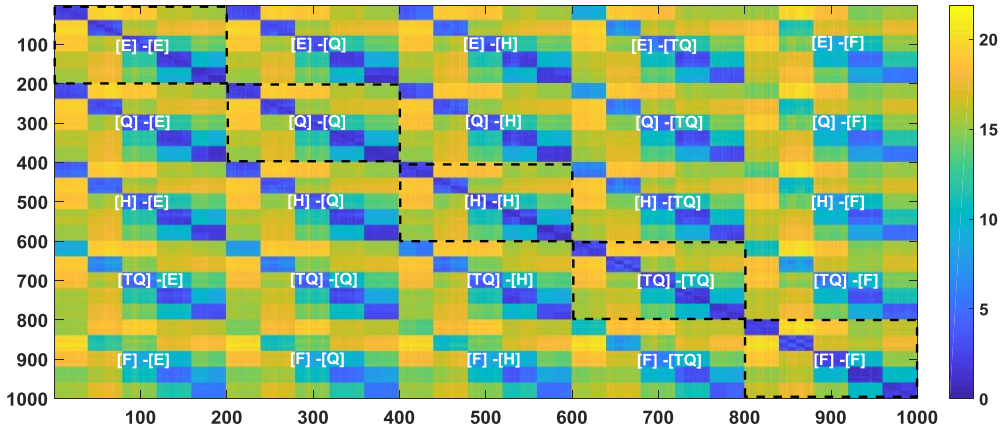


Figure 5.34: The distance matrix of combined loading classes associated with various structural conditions.

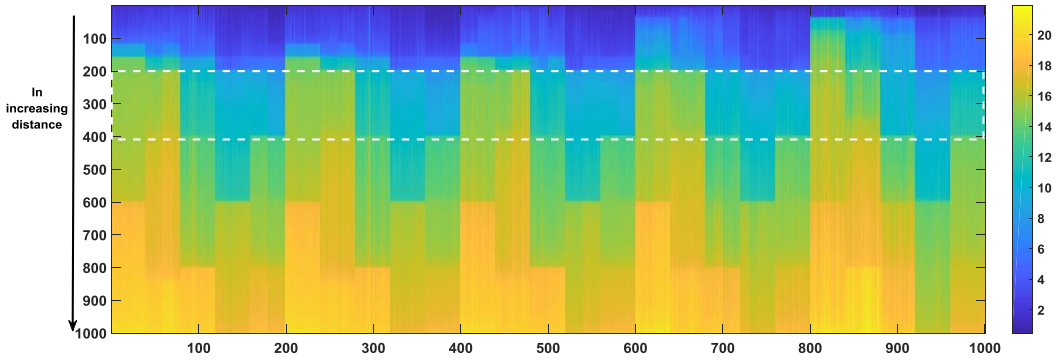


Figure 5.35: The sorted distance matrix in increasing values computed for combined data sets of PC model D

The distance matrix in Figure 5.34 shows the overall pairwise distance of the combined all data sets. It shows significant degree of variability of between undamaged and undamaged/ damaged data sets and also between two loading data sets. By sorting the distance in ascending order, clearer order of variability is obtained and sigma value can be chosen more conveniently (Figure 5.35). The sorted distance matrix in increasing values with the dash box indicates the block where the min distance is considered for the standard deviation as it shown to provide most appropriate data separation between undamaged and damage class.

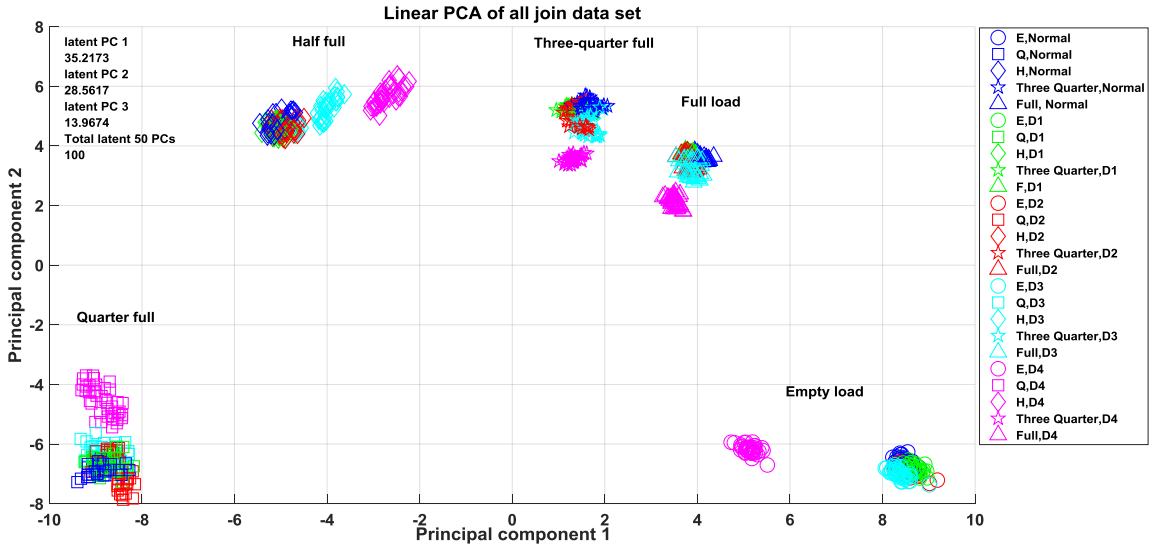


Figure 5.36: Result of the linear PCA using combined data sets

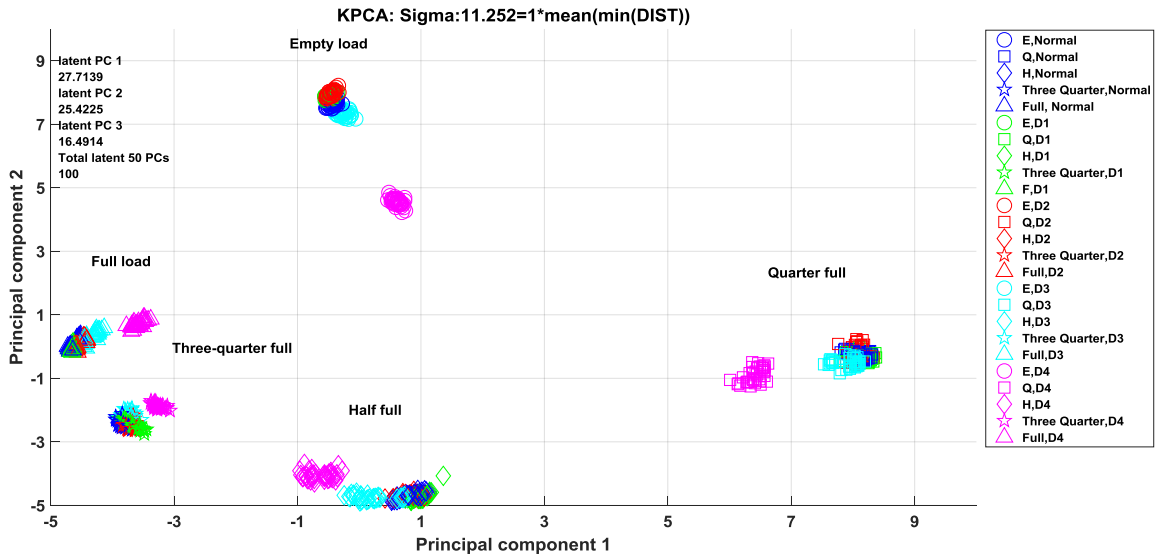


Figure 5.37: Results of using the kernel PCA

The selection of the sigma is selected from the average of minimum distance in block of the second lowest distance indicated by the dash box (Figure 5.35). As explain previously, that the second block with smallest distance allows the sigma to capture the variation between damage variables and between loading variables. Choosing the sigma from the first block with the smallest distance has shown to emphasis on the distance among damage data (inner-class distance) in local load but it neglects the inter-class variability.

The results from the first and second principal components projections of both the linear and kernel PCA (Figure 5.36 and Figure 5.37) show there is no significant improvement on the separation among the smaller damage and undamaged variables in the use of kernel PCA. The reason is because there is no adjustment on parameter sigma to adapt the variability of each data class. There is only one sigma parameter used globally for whole data sets. Certainly, this general model of variability cannot provide an appropriate and good variability model for each loading class. Variability model infers to the establishment of kernel parameters which consider the sigma that represents the inverse variance distance and distance between two vectors of data sets.

In order to improve data variability and reduce data overlapping, the large matrix can be partitioned in two separate groups. The first group comprises empty and quarter full load. The lower loading classes which are found from the original FRF to have higher variability

whereas the second group of half, three quarter and full tank load data set have smaller variance. By partitioning these two data groups which are based on variability allow the parameter sigma to be adjusted locally to their data variability. The following plots demonstrate the results of such partitioning technique.

In this partitioning case where it includes data set from empty and quarter-full load with larger variance, it is beneficial to use larger sigma to capture the separation in both loading data set that comprises undamaged and other four damage classes.

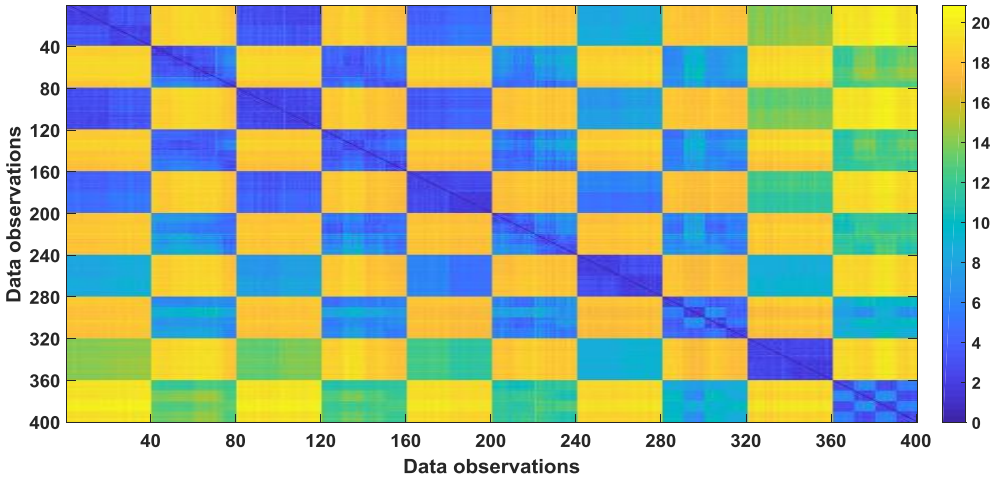


Figure 5.38: Distance matrix is used to visualize data variability and separation.

Figure 5.38 shows the separation between the data points in each variables in distance matrix form. The variables in the distance matric include undamaged and all damage classes in empty and quarter-full load. Noted that, the highest separation occur between undamaged empty and D4 quarter-full located at the last block (40-by-40) of the distance matrix. Then to choose the sigma parameter, it requires the smallest distance among the data variables. Hence, by ordered the distance matrix in ascending form as done previously, gives more convenience in which block of rows to select.

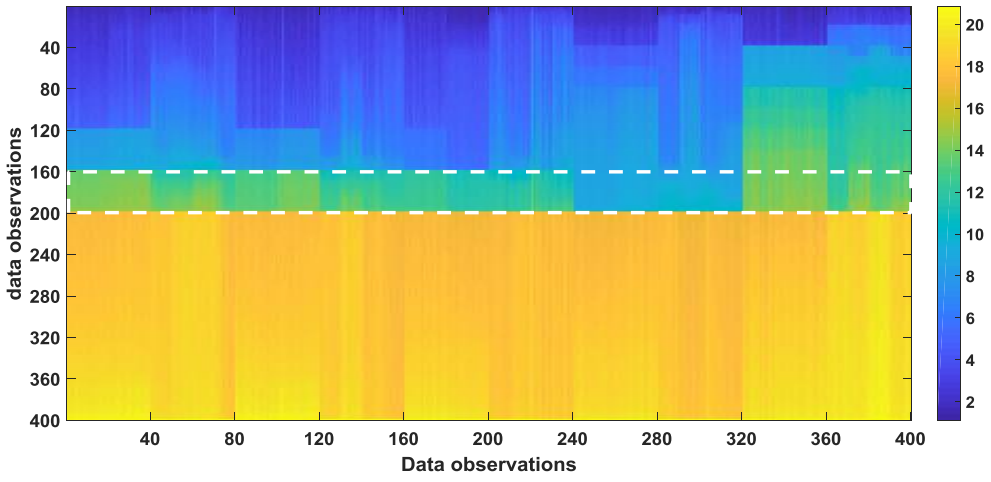


Figure 5.39: Distance matrix provides convenient way in considering the parameter sigma.

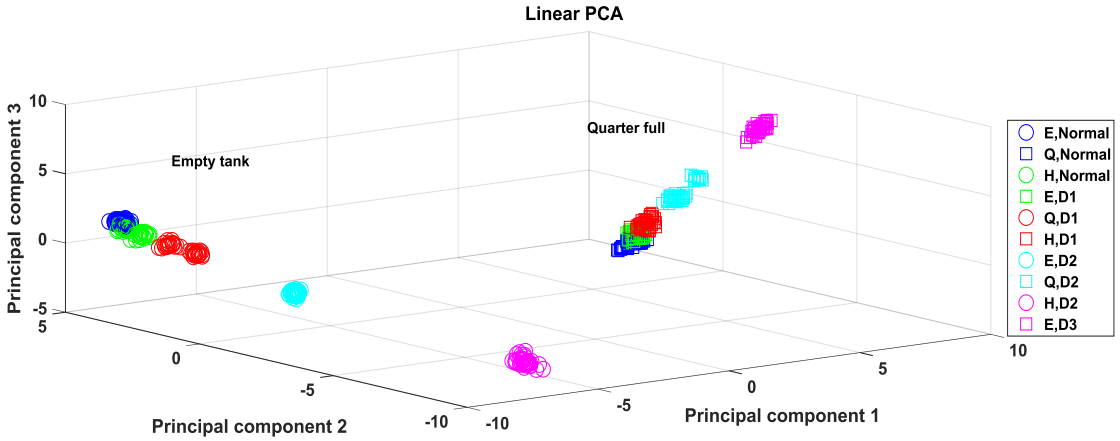


Figure 5.40: Linear PCA by portioning of empty and quarter full from other load set

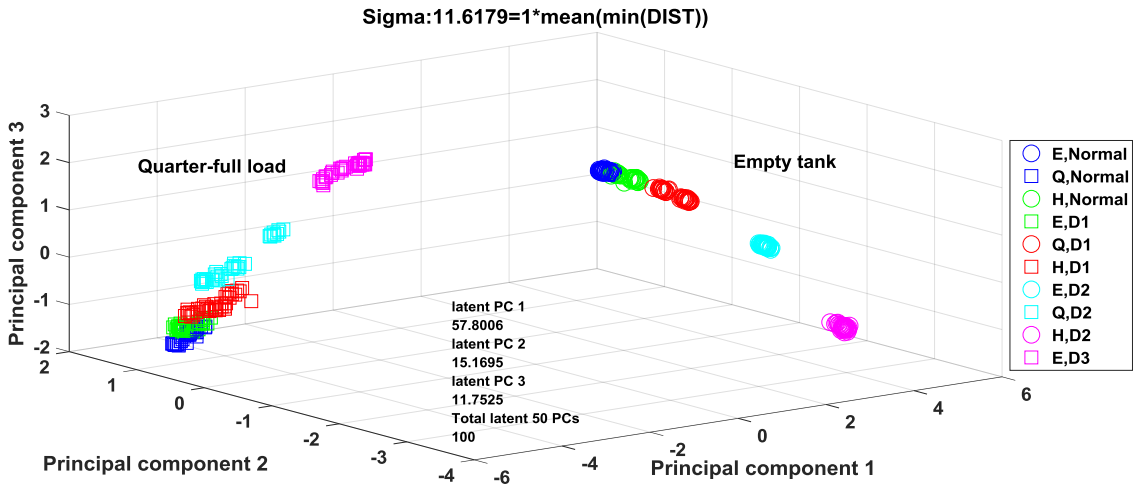


Figure 5.41: The result of partitioning empty and quarter-full load from other loading class

In case of combining empty load and quarter-full load, a higher sigma from block (row 161-200) in the corresponding ascending distance matrix reveals complete the data trend to monitor the change of structural condition (Figure 5.41). The attention is on tracking the damage in both loading sets rather than the inter-distance of loading class. In this 3-D view of this plot (Figure 5.41), it describes the structural changes nonlinear geometry of the data path. Compared to the linear PCA, the monitoring can only be taken in a straight geometry line. Noting also the high variance in the the first three principal components is stretching and enlarge the data separation.

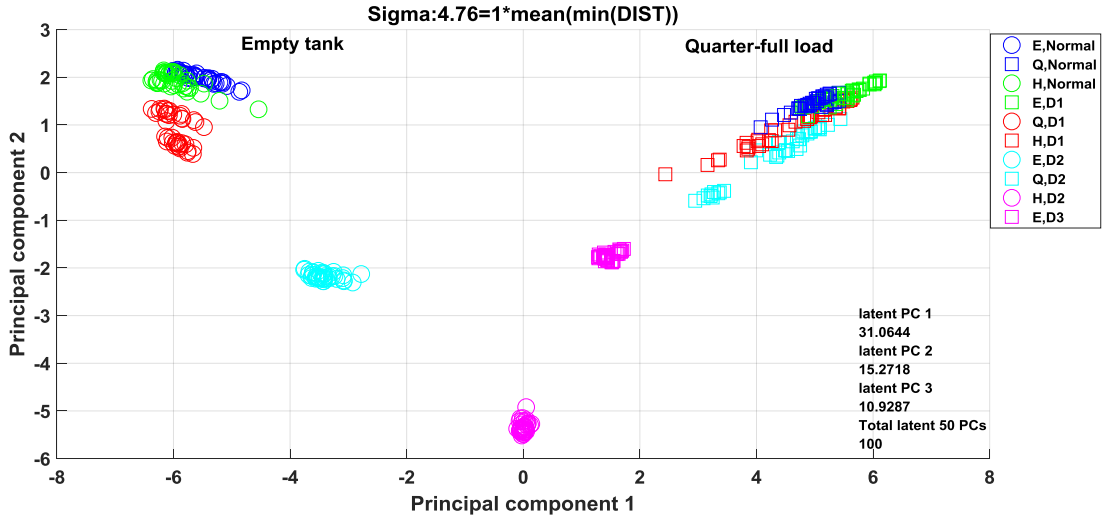


Figure 5.42: Kernel PCA using sigma from the second lowest block in distance matrix

Selecting smaller parameter sigma from the block with the second lowest distance (row 41 to 80), the feature projection shown to emphasis on the data variability among its damage variables and show less distinct data path of the two loading sets (Figure 5.42). For monitoring purposes, it is more favourable to consider a higher sigma parameter that includes high varieties of data distance preferably in the middle distance range to get reasonable data separation. Using dissimilarity or distance matrix, it can easily evaluated by looking at the colour distribution at each block rows.

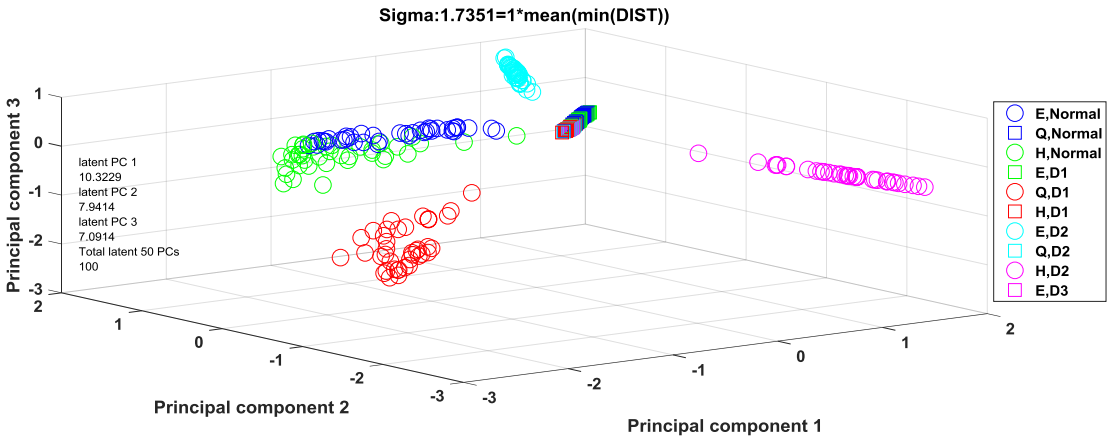


Figure 5.43: Kernel PCA using the smallest sigma from the the first block in the distance matrix

Using the smallest sigma from the row block 1-40 sigma in distance (Figure 5.43), the smaller variance allows better separation of damage variables within empty load class variables but performs poorly on quarter-full load.

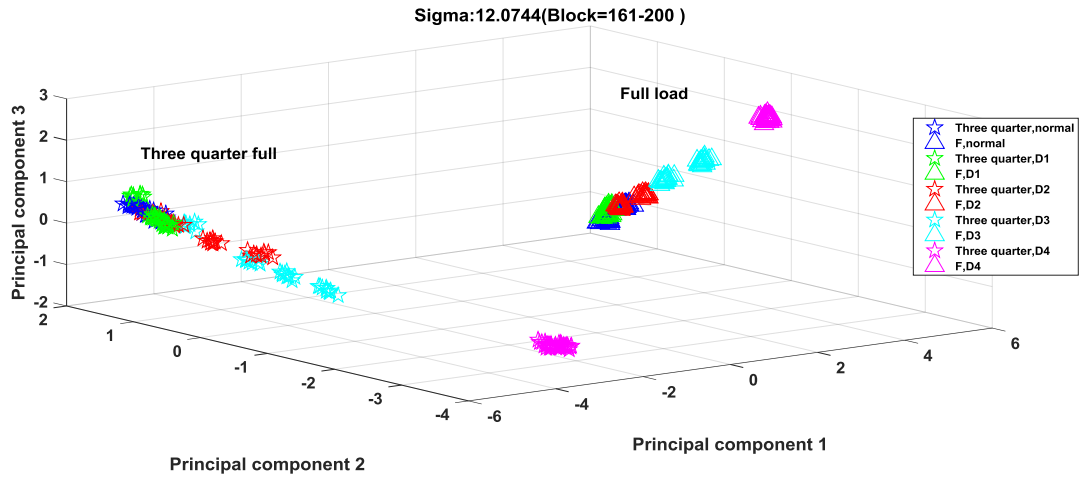


Figure 5.44: Kernel PCA on combined three-quarter full and full load data sets

Figure 5.44 shows the results using the sigma parameter acquired from the block in the distance matrix in Figure 5.39 which parameter lies in the intermediate range of data variability. The result is not encouraging due to some data overlapped especially in the three-quarter-full load.

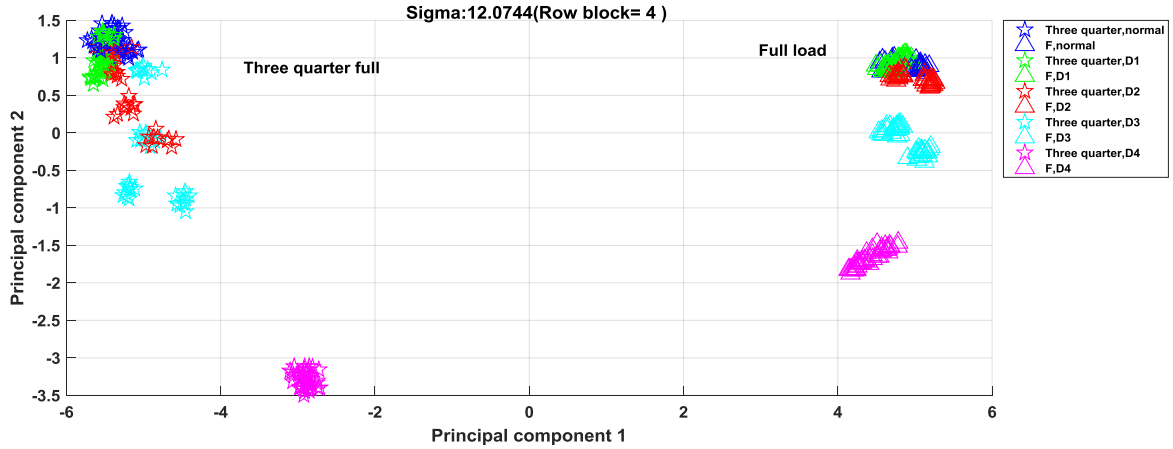


Figure 5.45: The same figure simplified into 2D plot of the above plot.

It is shown that from Figure 5.45, combining the loading classes having approximate equal size of variance despite partitioning them based on the approximately equal size in variability does not improve significantly the data separation between damage severity

classes. The result of kernel PCA (Figure 5.45) displays lower separation of smaller damage and undamaged variables in both three-quarter full and full load. As mentioned earlier, the generalized sigma used on both combined load (empty and quarter full, three-quarter full and full load) is not able to model the data variability locally or adapt to the variability of at each loading class. Having a global sigma parameter for the whole matrix for the kernel Gaussian computation seemed does not improve data separation between the undamaged and the smallest damage variables.

On a positive note, the kernel Gaussian PCA has its inherent pattern of data trajectory in the use of sigma value selected from the average of the second smallest distance in the distance matrix. Whereas, a smaller sigma parameter can potentially focus on the smaller variability in the loading set. The selection of the sigma is important to ensure it either separating the damage variables or establishing a data trajectory pattern for SHM purposes.

5.6 Novelty detection

In the previous section, features from PC models are projected into the first and second principal components projections and highlighted using 2-D and 3-D plots. The feature visualisations provide the first step in novelty detection before it proceeds to any machine learning technique. In the next stage, the features obtained from the visualization plots are applied with the T-squared statistics to measure the distance of each damage test data and the undamaged data set.

In Features Processing Techniques (Chapter 3), T-squared statistics are described in the context of PCA. In the SHM perspective and damage detection architecture, a higher T-square index is desirable because it indicates large separation between the damage condition and undamaged condition and potentially increases damage detectability.

In the following section, the reduced dimensional data from PCA are analysed in terms of its variability regarding the separation between damage severities and undamaged condition. Each of the observation in the score matrix (principal components projections)

is checked against the undamaged condition by computing MSD. In the context of SHM, damage detection is recognised if the data observation detected above the threshold. With supervised learning and labels exist for all data set, the true or false of damage detection can simply be identified.

5.6.1 Novelty detection of PC Model A

The features performance in the reduced dimensional space are measured by considering how far is the undamaged data separated from the damaged condition using T-squared statistics. The variance (eigenvalues) in the first fifty principal components are extracted to give a good representation of the most important information from the structural characteristic.

In linear PCA, combining all operational loading conditions as its baseline data set causes true damage detection only for empty and quarter full load with systematic increase in T-squared index corresponding to the increase in damage severities. However, as the load increases beyond half-full, the performance of the novelty detection becomes unsatisfactory but improving on the full tank load (Figure 5.46).

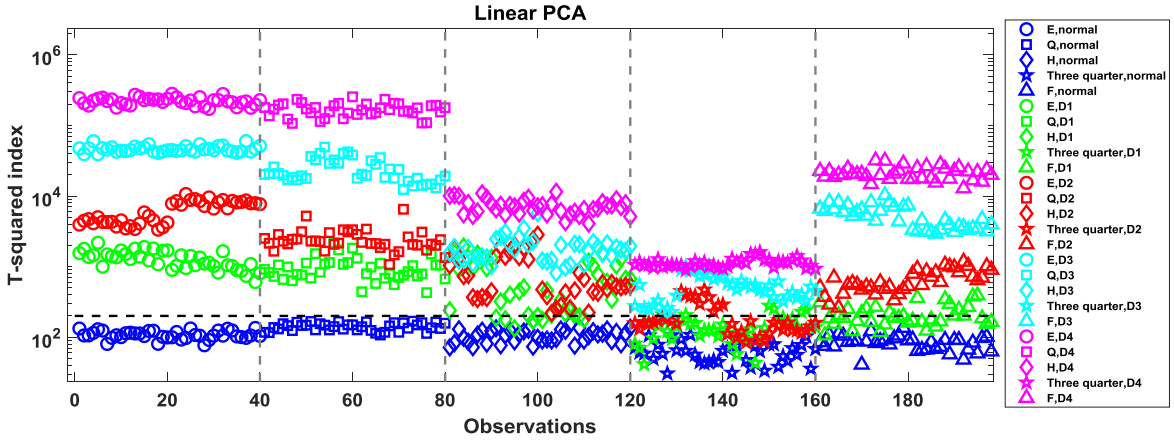


Figure 5.46: Novelty test result based on baseline model of undamaged data set that spans all loading conditions tested against all test data sets

In higher operational load, especially for three quarter full load (Figure 5.46), it fails to detect the presence of damage in a satisfactory manner. Novelty detection for empty and quarter-full load show good novelty detection. In half full and full load, the detection performs better than the three quarter full with only data points from the smallest damage, D1 measured below the threshold line and this can potentially cause a false damage detection.

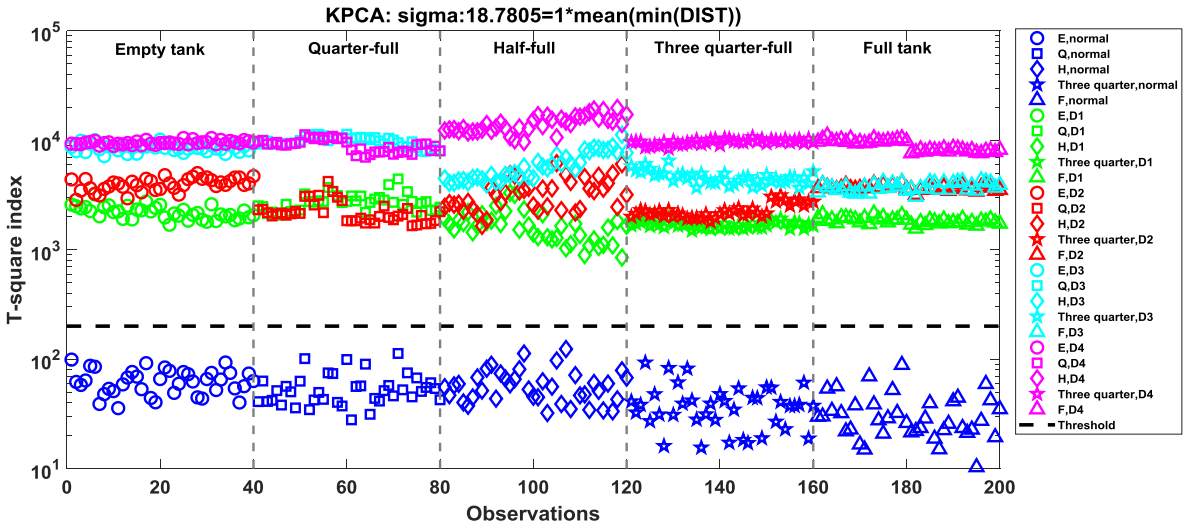


Figure 5.47: MSD distance on the principal components projections using 50 principal components.

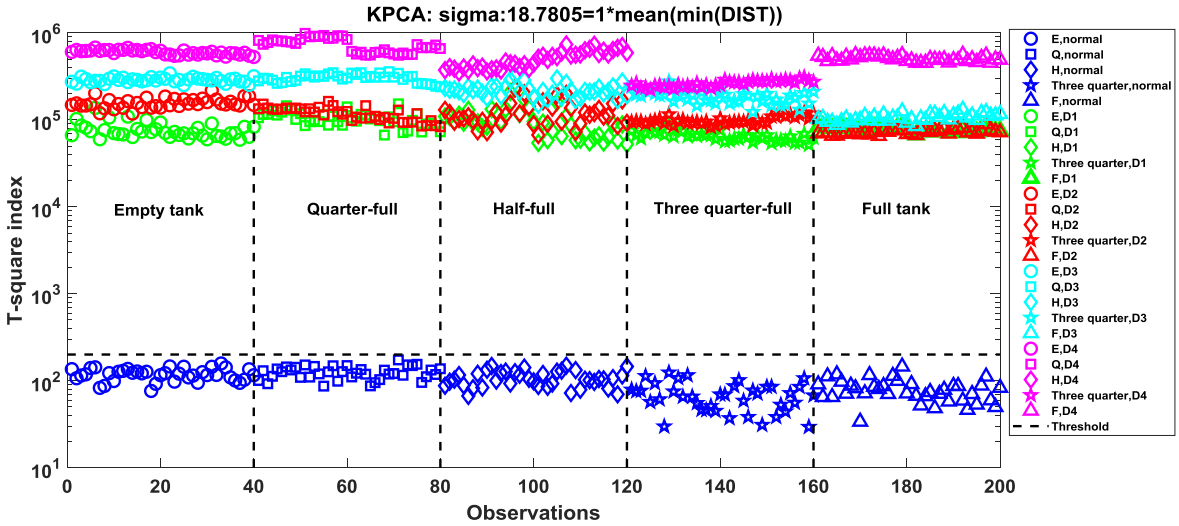


Figure 5.48: Kernel PCA using the overall baseline of undamaged data set with 100 principal components

In kernel PCA based on kernel PC model A (Figure 5.48), the separation between the test data and the undamaged baseline data is large and profound that the undamaged are separated remarkably from all damage variables in all loading classes. In fact, increasing number of PCs in the the novelty test can improve damage detection performance. All damage variables are effectively identified as damage compared to the standard PCA result. However, the correlations of damage severities are not consistent with its T-square index. This can be due to the large variabilities in the distance matrix arise from the combined loading sets and the sigma that is calculated from the distance matrix that is difficult to adjust to the high variability from the large variance.

It had been investigated and as expected that varying the sigma values alter the variability trend of damage severities and their T-squared index from the reference set. In this case, the sigma is adapted to each structural health condition (in configuration as Figure 5.17) to obtain appropriate damage variabilities. Despite some data is not correlated with the damage severities level (Figure 5.48), generally it produces novelty detection result in the context of damage detection under the influence of operational loading variations.

The results of the novelty detection should not be identical to the visualisation plot of the same model described in Figure 5.19 because the sigma values are adjusted differently for these two cases, used either for novelty detection or for health monitoring of the structure. As stated earlier, for visualisation model, the aim is to establish a data tracking technique to monitor the structural health conditions. In novelty detection, the object is detect damage presence by measuring the data from the undamaged set and conclude as damage if the data points are positioned above the threshold.

The results also demonstrate that if more eigenvectors (more PCs) are considered in the novelty detection (outlier analysis), potentially it can produce better damage detection. The distance index between test data and the baseline increased due to more of the principal components selected in the eigenvalues decomposition, which is hundred PCs (Figure 5.48). Noting that, the visualisation plot compared to Figure 5.19 considers only two main leading eigenvectors in the feature projection compared to more of principal components are selected for the T-squared analysis. This is another reason for the incompatible results between T-squared plot and KPCA visualisation plot.

It is important to note that, despite the kernel PCA produces profound novelty detection, the performance relies mainly on the parameter sigma which are computed locally at each structural condition in the case of PC model A. The values can be tuned so that the distance of the test data from the baseline set are distinct with good damage consistency.

5.6.2 Novelty detection of PC Model B

In this model, a reference set encompasses all of the damaged data sets of an empty tank load is applied. If only one loading condition is used as a reference set in the MSD function, it means that the mean and covariance matrix is based only on one loading condition and it may not be so robust in terms of loading variability because in reality, the loading can easily change during the operation. Hence, the test data is compared with the overall loading conditions before calculating the T-squared index using the MSD function.

The T-squared results using the transformed data from the linear PCA evidently show relatively inaccurate damage detection where many test data from various damage severities detected as undamaged (Figure 5.49) and this can cause false negative damage detection. On contrary, kernel PCA (Figure 5.50) effectively produces accurate novelty detection for all test data sets. The results are very encouraging as all damage severities including the smallest damage are positively detected as damage and lies above the similar threshold line (Figure 5.50) without not much tuning of the sigma. The sigma is just simply calculated from the average of the minimum distance of each distance matrix corresponding to the each loading conditions.

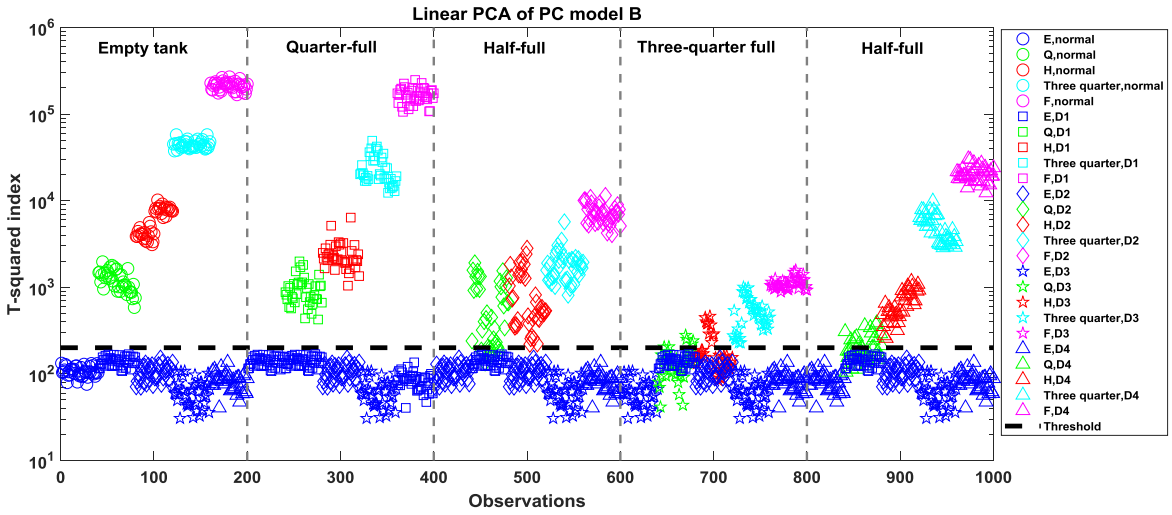


Figure 5.49: Novelty test based on linear PCA data variables

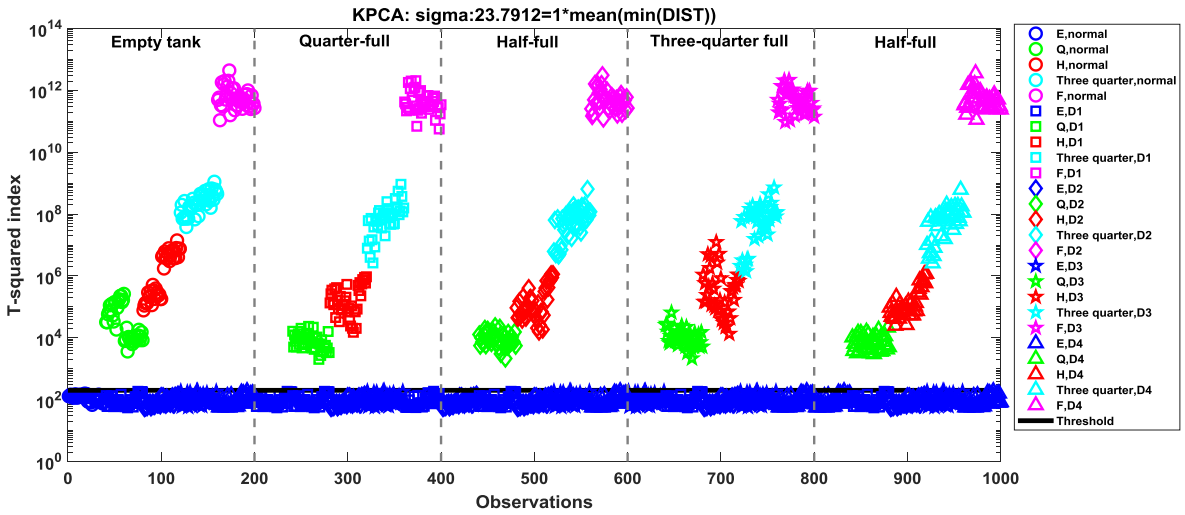


Figure 5.50: Novelty test based on kernel PCA using PC model B

PC model B via its kernel algorithm successfully isolate the loading variability from the damage variables. Via kernel PCA, the damage variables are measured indicating the distance index in an ascending order corresponding to the damage severities (Figure 5.53). Revisit the construction of covariance of this model, it organizes an empty load with its corresponding damage states, the eigenvalues decomposition is computed based on only the empty load consists of all damage severities in the loading class. The T-squared plot demonstrates an excellent result with the damage severities increase corresponding to the

higher distance index in all loading classes. This PC model stresses the advantage and feasibility of computing the sigma parameter for small variability data set in each loading class. The small variability is produced by each of damage data set and because of that, the sigma is simply taken as the smallest distance excluding zero. In comparison to the PC model A (Figure 5.17), each structural condition class has huge variability due to the combined of various loading conditions. Small variability in each class enables the sigma to be better generalised as variable predictor as it switches from one loading to another loading class.

5.6.3 Novelty detection of PC Model C

As discussed previously, model C has no effects from operational loading variations in regards to its principal components as the eigenvalues decomposition is performed on each covariance of each loading condition individually. Each loading class contains its own damage variables under each loading condition. For kernel PCA, a distance matrix with a sigma calculated the smallest value based on the distance matrix are established from each loading class and plotted in Figure 5.52.

There is a clear trend of different variability in each loading class consisting all damage data set in T-squared of the linear PCA where there is switch from high variability in empty and quarter-full load to lower variance in half-full, three-quarter full and full tank load (Figure 5.51). The novelty detection is satisfactory; however note that, there is few data points falsely detected as undamaged in the half-full load.

The trend of switching of variability throughout the loading class becomes more systematic in kernel PCA. Noting that, the variability in quarter-full is increased and it seems the smallest sigma in the kernel function on each loading class has normalised the distance matrix (the denominator of the distance matrix which is $2 \times \sigma^2$). When computing the T-squared in MSD function, the effects from the kernel Gaussian produced a more stabilised variance throughout the loading class. This is one of the key finding from the current work which involves multivariate data (Figure 5.52).

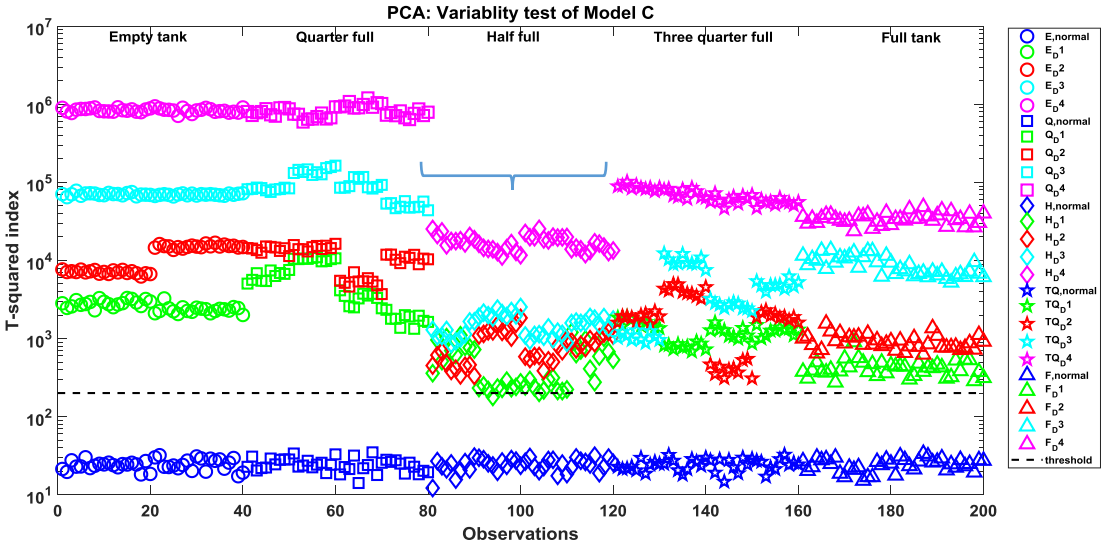


Figure 5.51: A switching trend of damage variability through the loading class.

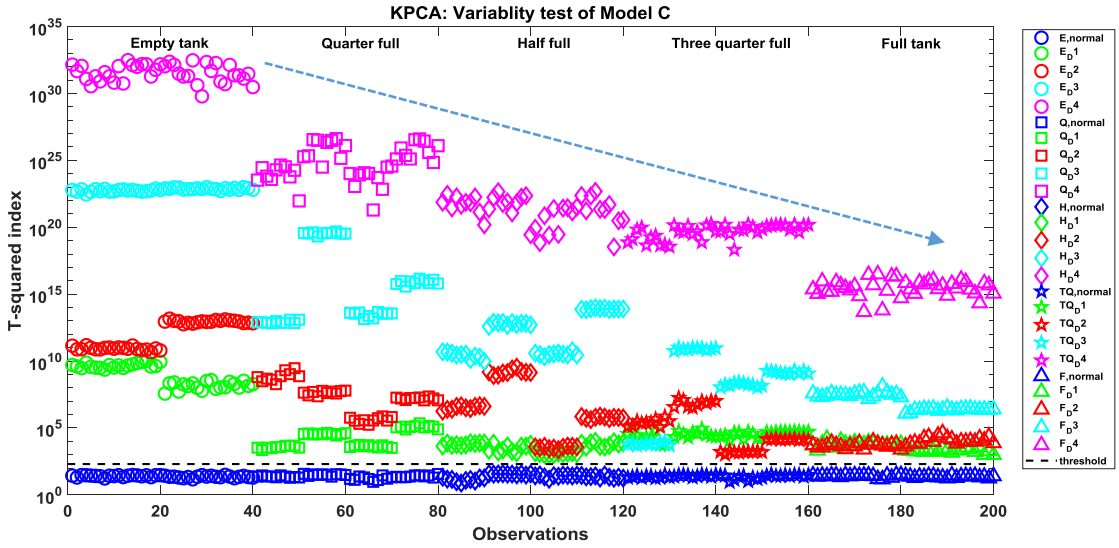


Figure 5.52: Novelty test of kernel PCA using PC model C of which each loading class is join subsequently.

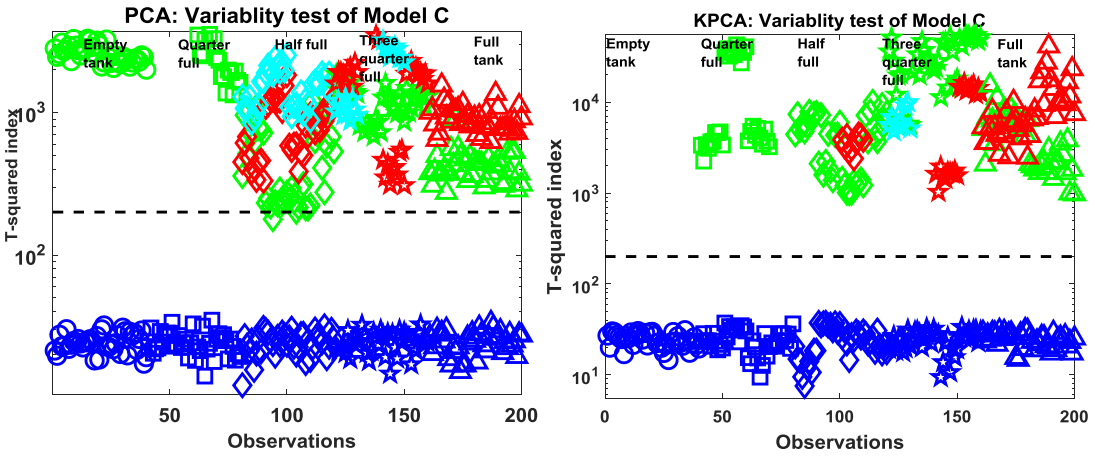


Figure 5.53: A zoomed view of the previous figure reveals novelty detection for linear PCA (right) and kernel PCA (KPCA) application (left).

As anticipated, the kernel PCA of model C performs good novelty detection (Figure 5.52 and Figure 5.53) using one sigma parameter and a distance matrix corresponding to each loading condition. Kernel PCA manages to detect all damage class including the smallest damage class and the data points of damage class are projected well above the threshold line (Figure 5.53, right plot).

However, notice that there is some data 'jump' from D3 and D2 data variables on smaller damage group on the T-squared plot observed in the three-quarter full load. This 'jump' is inherited from the original FRF data signal. The data behaviour can also be observed on the linear PC model A in Figure 5.21 (a) and Figure 5.45. The cause of such behaviour can be due to the noise and variations during the implementation of the experimental procedure (loading of water, shaker stringer replacement and different tightening of bolt on the structure).

By selecting the smallest distance in the distance matrix, the selected sigma gives good performance in separating the lowest damage data set from the undamaged data set. This is the main benefit in kernel PCA where the parameter sigma can be adjusted according to the aim of analysing the multivariate data, it is either for tracking the structural damage or for detecting the minimum damage.

5.6.4 Novelty detection of PC Model D

In previous models, it is highlighted how one of the effects from loading or damage could be isolated by some arrangements of data set. In current Model D, the data matrix combines all undamaged and damage data sets into one matrix in which the principal components are computed. It causes both loading and damage effects on the variability from the combined data class in the matrix.

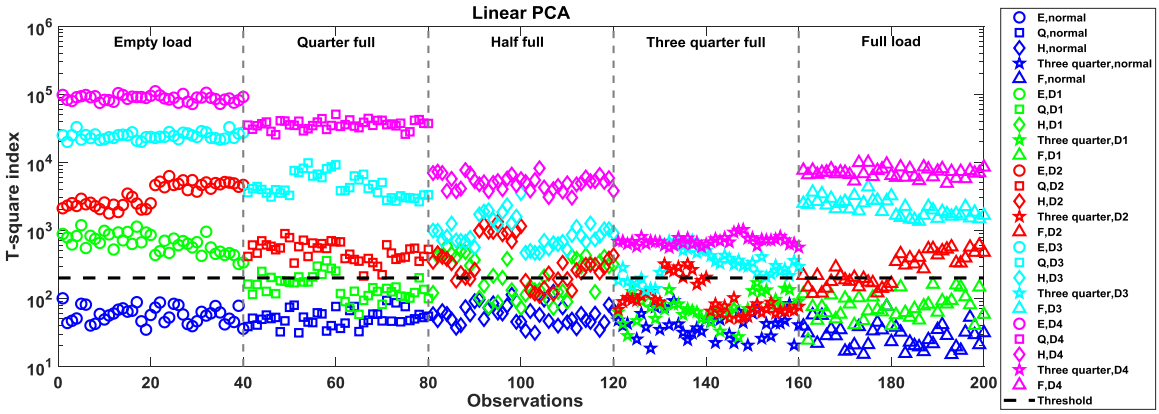


Figure 5.54: The variability results for joining all data sets in which the eigenvalues decomposition is performed.

From the linear PCA projection in Figure 5.54 using the combination of whole data sets, it is observed that there is a clear switch in variability across the loading class. The MSD plot shows a trend of a larger variance in empty and quarter-full load data set and reduced variance at half-full and full load. The three-quarter full class seemed to exhibit the smallest damage variability in its loading class. The damage detection is not encouraging especially for the three-quarter full load.

In kernel PCA, the same procedure is performed that is a distance matrix is established for the large matrix (1000-by-1000). The sigma parameter is selected from the smallest distance on the distance matrix (excluding zero). MSD is computed based on the data matrix which follows the configuration in Figure 5.6. Of course, the key advantage from

the kernel PCA application is that the damage detection is giving nearly 100% accurate (Figure 5.55). Notice that, the effect of sigma squared normalising the distance matrix is insignificant compared to the individual loading class (model C). Nevertheless, it is successful in the novelty detection test compared to the linear PCA. This result signifies how a local model of distance matrix and its sigma is more effective in increasing the data separation than the global distance matrix and its sigma parameter.

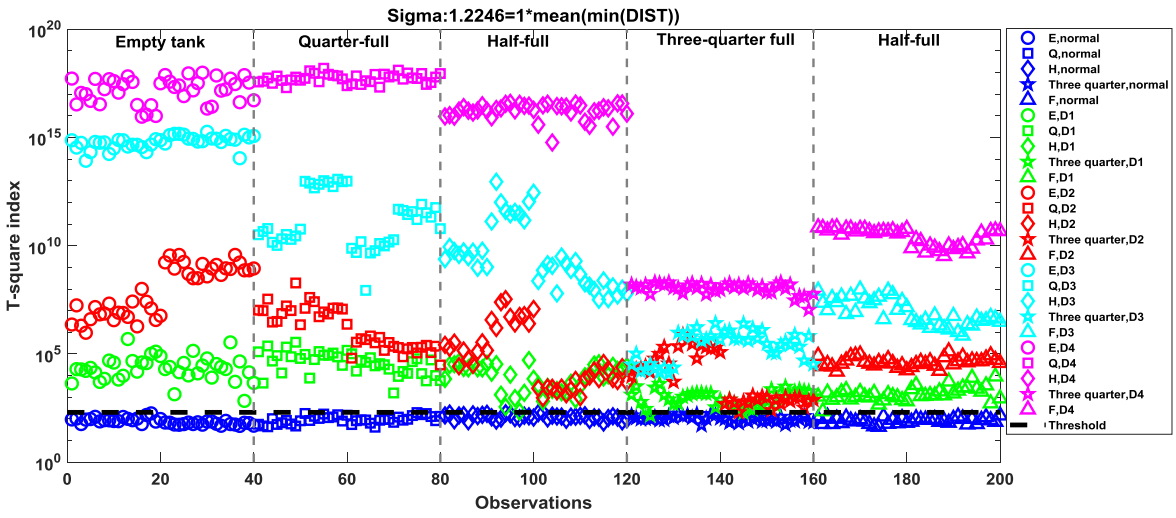


Figure 5.55: The combined of all data sets matrix applied with the kernel PCA and compared with reference set of all undamaged loading data set.

Obviously, as highlighted in the previous model, three-quarter full load is shown to have the smallest data variability compared to other loading classes. Due to the advantage of the kernel Gaussian function, the distance is still enlarged and separated from the undamaged data set. As a result, the observation data are all escaped above the threshold to give accurate novelty detection. Figure 5.56 highlights the data points fall close to the threshold.

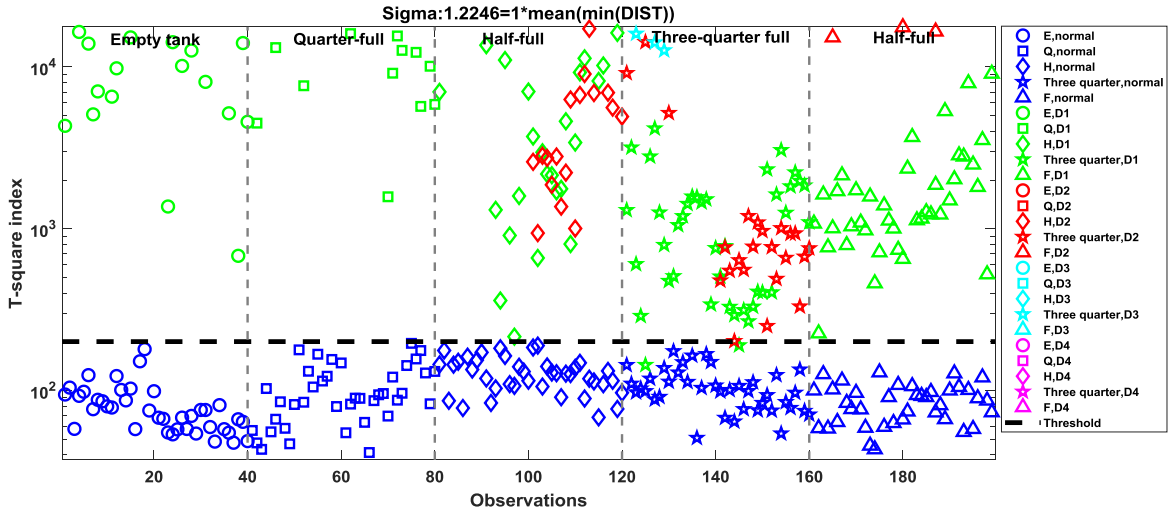


Figure 5.56: The zoomed view from the previous figure focusing on the observations data close to the threshold line.

A small sigma reveals better variability or separation distance for damaged and undamaged data sets under empty, quarter and half-full loading due to a higher variance in the class. The distinct and separated data sets from these classes inherited from the FRF data which are also observed to exhibit FRF peaks that are distinguished and separated from each classes. These characteristics heavily influence the novelty detection using the T-squared and the projection in the feature space as illustrated by the kernel PCA and the linear PCA. Kernel PCA enables larger separation between data set of the test data set that associated to various structural conditions.

As described previously, the parameter sigma must be set to the average of the minimum distance of multivariate data class to capture the smallest distance of neighbouring data points from test data sets but should be it must be larger than the distance between loadings class. This strategy applied in order to determine the sigma value by calculating the minimum distance for the mean of each data point distance as illustrated by equation (8).

In respect to the usage of kernel PCA, separating the loading data based on significant different of variability can improve damage detectability. Data set from higher variability such as the empty and quarter full load requires larger parameter sigma in order to capture the neighbourhood distance. On the other hand, data groups from lower variance, for instance the three-quarter full and full load give better performance with much smaller sigma as been highlighted in Figure 5.55.

5.6.5 Novelty detection on partitioning data class

Having knowledge about the data variability characteristics and the advantage having a local kernel Gaussian model on each loading class, it is not a bad idea to partition the loading classes between the high variance and much lower variance data class. This is performed to explore if it is still possible to gain the benefits of kernel Gaussian algorithm and produce better novelty detection compared to the previous model, which is a combination of all data classes.

The results of the first partitioned group from higher variance group, which is the empty and quarter full load, reveal better novelty detection in both linear and kernel PCA with a sign of systematic ascending order of damage severities (Figure 5.57 and Figure 5.58). Kernel PCA is shown to perform extremely well with 100% accurate damage detection compared to the linear PCA which has false positive damage detection on its quarter-full load.

In kernel PCA, using the same procedure (building a distance matrix and find the smallest distance from the sigma), the algorithm effectively separates between each damaged/undamaged class as shown in Figure 5.58.

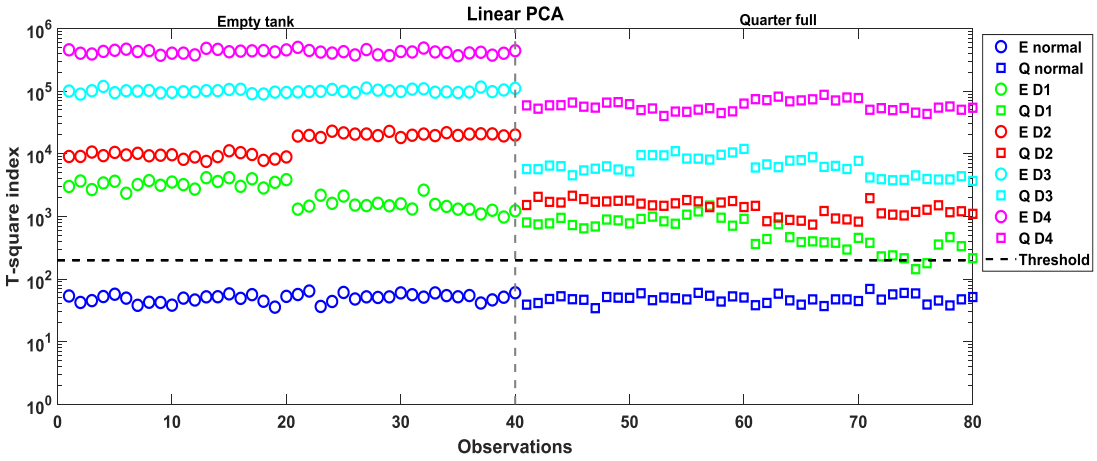


Figure 5.57: Novelty test using linear PCA on combined data from empty and quarter full load.

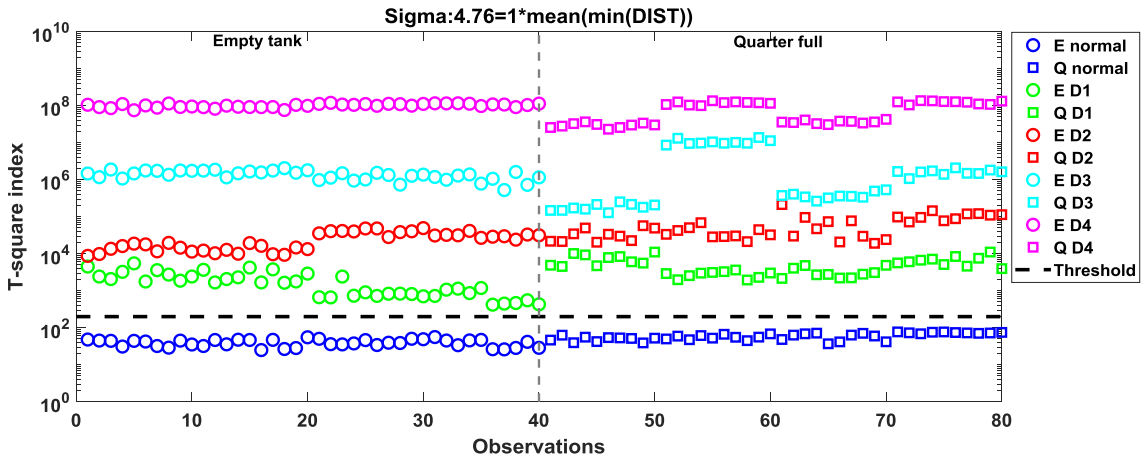


Figure 5.58: Novelty test using kernel PCA on partitioned empty and quarter full load

In another portioned group, comprised of half-full, three-quarter full and full load, the kernel PCA processed data manage to produce generally accurate novelty detection (Figure 5.59). However, the performance is shown to be better for smaller partitioned group comprising three-quarter full and full load (Figure 5.60). Another smaller partition,

which comprises of the three-quarter full and full load is performed with the kernel Gaussian algorithm. It reveals more accurate damage detection with indication of similar variability in damage variables (Figure 5.60). The result indicates that the smaller partitioned class with about similar data variability can improve novelty detection.

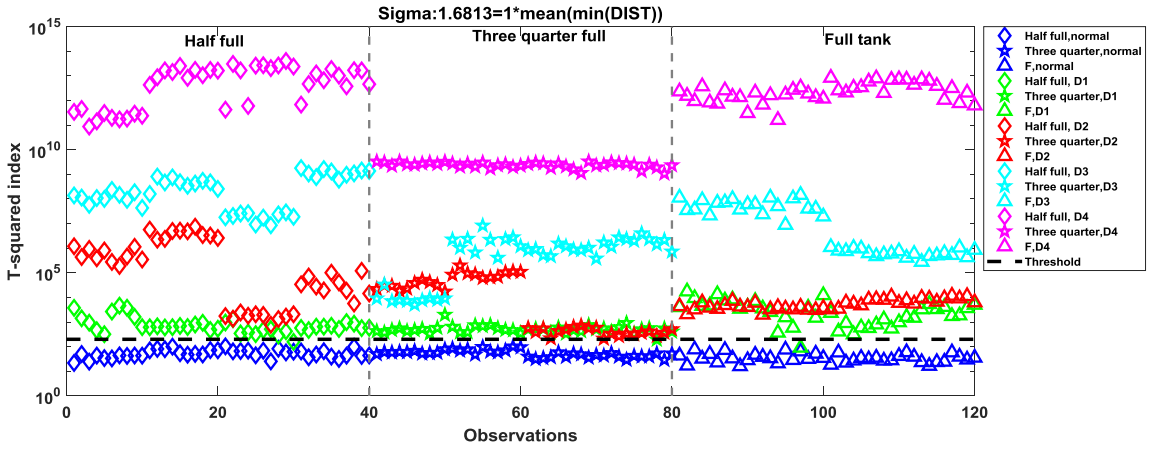


Figure 5.59: T-squared distance using kernel PCA on joining data sets of lower dimensional length (half full, three-quarter and full load)

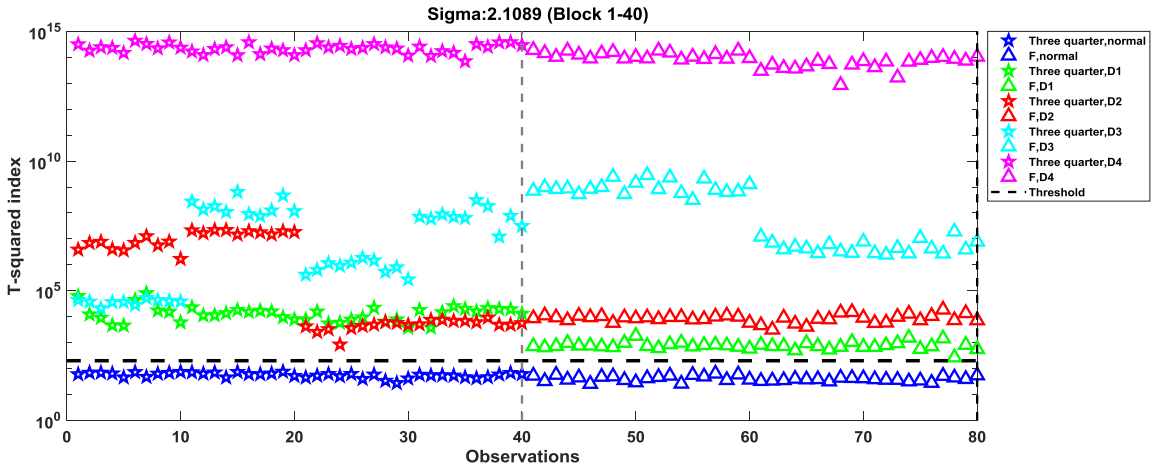


Figure 5.60: T-squared index using kernel PCA based on data partitioning by including three quarter full and full load from data set

5.7 Discussions and Summary

This chapter presents various ways of data representation for the covariance matrix before performing eigenvalues decomposition. Standard and kernel PCA are applied on the PC models showing more promising results for the generalised nonlinear PCA. Two main findings are derived from using both techniques; visualisation of structural health changes and novelty damage detection using Mahalanobis squared distance (MSD). Some important findings related to these works are discussed and summarised in the following section.

5.7.1 PCA as the inheritance to the FRF peaks

In this chapter, nonlinear PCA based on Gaussian kernel are demonstrated and compared with linear PCA. PCA is applied on the vibration data acquired from the wing box experiment under varying effects of liquid loading and damage severities. PCA is used on the data set to reduce the high dimensions and to obtain only the most significant feature with all the unwanted noise and data redundancies are omitted from the feature. The PCA data are shown to inherit a very important data behaviour from the operational loading conditions. The PCA data (all the principal components projections) inherit the peaks of the FRF magnitude which represent the structure's natural frequencies.

It is observed that there is a changing trend of data variability during the loading transition as the loading switched from one loading condition to another loading condition. One underlying finding observed from the PC projections is that the features' variability are intrinsic to the characteristic of the FRF peaks produced from the experiment. Given the data signal is highly coherence (less noise), the FRF peaks of the spectral variables which correspond to largest variance, the effects would be indicated on the PC projections. If the data signals are close or overlaid to one another, the resulted features from the PCA associated to the overlaid signals have the tendency of overlapping. Here, the effects of proper feature selection play an important role in avoiding this scenario.

5.7.2 Kernel PCA as a damage severities separator

In this work, kernel PCA is applied on the data set to ascertain if it can improve the separation between the overlapping data features that are present in the standard PCA. Kernel PCA is shown that it can separate the features associated to damage severities group. It is indicated that kernel PCA can transform the features into a curved orientation in its projection space whereas a standard PCA is limited to only a straight line projection. In the context of SHM, this emerge a useful benefit in monitoring structural integrity.

5.7.3 The fundamentals process in modelling kernel PCA

Incorporating kernel Gaussian PCA into the data set, an essential step is to create a distance matrix which represents the distance between each data observation in data sets. Each data matrix is created for each loading condition (for baseline model of PC model A) or for each structural conditions (damage model of PC model B). Each distance matrix is also created for each loading conditions (PC model C) and also for a combined loading and test data sets (PC model D).

From the created distance matrix respective to each model, the parameter sigma, which represents the average of the smallest distance, computed on the block in the distance matrix represents the data distance between undamaged and the smallest damage class. The block matrix that highlights the location in the distance matrix where the desirable distance were calculated based on the indication of the colour scheme indicated the distance matrix. Each cell in the distance matrix displays a similar colour scale for the same data class and varies as it moves to different cell along the distance matrix. Each block has the same size of each damage class. To ease further on the computation of sigma, each rows in the distance matrix is rearranged in ascending form.

With the aim to make larger separations between the data classes, the strategy is to choose a sigma that is large enough to capture the neighbouring data points and much smaller than the inter-class distance [55]. The key finding revealed that, for the purpose of monitoring the overall structural health conditions throughout all loading conditions, the sigma should

not be calculated from the first block of the smallest distance in the distance matrix but at least on the second block or higher in order to capture a better generalized model for larger data separation. For novelty damage detection, the sigma is selected from the first block of the distance matrix in order to choose the smallest distance. Both parameters of sigma and distance matrix are then applied with the Gaussian function to increase separation of each data points.

5.7.4 The advantage of establishing different PC models

Another key finding reveals that the main advantage for using and adopting different PCA models in isolating the varying loading conditions from damage detection. In visualization technique, the main concern is to visualise the projection and to discover how the structural health conditions are changing with respect to the first and second principal components projections. The first and second principal components projections correspond to the largest data variance of the variable components of the data set with respect to the dynamic property (natural frequencies) contained in the feature.

PC model A

Model A represents a more practical and realistic case where a training set is established from an undamaged condition consists of all loading conditions. The principal components projections using kernel PCA algorithm for this model demonstrates that based on a baseline set comprises all undamaged conditions, damage severities can be tracked along its path of data pattern representing the systematic loading, which each path represents different structural conditions. Each track which indicates one specific structural condition of all operational loading conditions has specific trait for different damage severities. Within a loading condition, a distinct data pattern effective show correlated increase in data separation as the damage severities increase.

PC model B

For PC model B, a reference data set is established based on an empty tank load consists of a particular structural condition that in reality it might contain with one potential damage severity. In this model, damage severities are the highlights in the principal components projection. The results effectively highlight the classifications of different damage severities in a systematic order of damage severities. As mentioned earlier, model B has a potential practical use in some fields in engineering such as the aerospace where a typical damage type that is common to occur in some aircraft structure can be used as a reference model. The damage propagation present in the structure can be monitored and send alarm if the novelty detection is far above the threshold.

PC model C

Model C reveals a direct approach where a particular loading condition is analysed individually by performing linear and nonlinear Gaussian PCA on each data set. Obviously, using this model, loading effects are discriminated from the PC features. Novelty detection for kernel PCA illustrates an improved detection for the smallest damage severity group compared to the standard PCA.

PC model D

In model D, the multivariate data from undamaged and damaged condition are combined into one single matrix. Here, the results have shown that kernel PCA is not effective in separating the damage severities group. This study reaffirms of the need to implement kernel PCA locally at every structural conditions (for PC model A) or loading conditions (model B and C). Finally, the finding reveals how the partitioning of data features associated to higher variability from those with smaller variability can produce reliable novelty damage detection.

5.7.5 Summary

By establishing distinct PC models, it is demonstrated how loading effects on the feature could be discriminated from damage effects. The PC models are established on numbers of possible scenarios which for instance, the ‘data’ from structure under normal condition is available. The second situation is when the ‘data’ form the structure under various damage conditions are found. This operational loading conditions are known except the damage conditions, a specific damage type is anticipated to occur under unspecified loading conditions or both loading and damage are unknown. These PC models have shown to produce different visualisations and the benefits that mainly driven by its field of application and the availability of data information.

The study has shown that the changes in loading and structural conditions could be better visualized and tracked through its data trajectory and pattern recognition using kernel Gaussian PCA. Using kernel PCA, it naturally increases the separation between the baseline and test data, hence offer great potential in SHM for system under similar conditions.

Another main highlight from the findings in this chapter is that kernel PCA has superior performance compared to the linear PCA with respect to the orientation and separation of the feature. Kernel PCA algorithm via Gaussian function is found to expand and widen the feature space and as a result, the features are more spread out compared with linear PCA.

Kernel PCA with Gaussian function has shown its superior performance by means of creating a kernel Gaussian model locally within each structural conditions or loading conditions. The main challenge of the kernel Gaussian PCA lies on finding the sigma parameter (inverse variance) as to optimise the data separation under the multivariate case.

	Sigma: Visualisations
Baseline model (under all loadings)	Baseline: 11.959 D1: 11.75 D2: 11.96 D3: 12.82 D4: 13.67
Damage severities model	9.81
Loading class model	E: 92.377 Q: 40.271 H:29.807 TQ:40.916 F:30.995
Combined load and damage model	11.252

	Novelty detection
Baseline model (under all loadings)	Baseline: 1.24 D1: 2.84 D2: 3.17 D3: 7.64 D4: 11.16
Damage severities model	5.51
Loading class model	E: 14.04 Q:7 H:7.48 TQ:10.14 F:9.85
Combined load and damage model	1.225

Table 5.1: Comparisons of parameter sigma in each data set concerning the use in visualisation and novelty detection.

Table 5.1 describes the values of the parameter sigma (σ) for different PC models and structural conditions. This parameter is incorporated as the inverse variance ($1/\sigma^2$) in the Gaussian function and also known as the precision [34]. It determines the quality of the data projected on the feature space and therefore plays an important role to ensure the separation and pattern of the data trajectory.

In this study, it has shown that the parameter values can be adapted according to the type of analysis to be performed. For visualisation purposes, the parameters assume slightly higher sigma compared to using its application in novelty detection. One remarkable finding is that, each data set from a loading class associated with a particular structural condition has distinct data characteristic or trait with respect to variance and that can be visualised better using kernel PCA. The parameters are adapted to each loading class or structural conditions locally and this is the superiority and flexibility of the kernel PCA.

To conclude this chapter, it outlines some underlying principles with regards to PCA.

- Standard/ linear PCA can be an effective as the first tool for dimensional reduction and ‘data variability discriminator’. The latter term means under a situation where two variabilities are present in the data, linear PCA can be applied to discriminate or overshadow one of the variabilities. As a result, only the effect from one variable parameter is present in the data set. Visualizations and novelty detection test for data

- sets driven/ influenced by only one variable parameter (as for this study, the effect form the variable damage is in the main interest).
- Kernel technique for generalized non-linear PCA can be used to examine in depth the changing trend in data variabilities using both the visualizations of feature space in high dimensions. The trend of data variability as subjected to changes in system properties (as in this case is mass) can be displayed in more informative and widespread manner. In this case, for example, the variability in damage classes (Figure 5.52) under the changes in loading demonstrates a decreasing trend in data variability when the system's mass is increasing.

It would be of great interest, to consider other machine learning techniques such as data clustering using the feature obtained via the kernel Gaussian PCA for this data set. A probabilistic model framework that can be used as a prediction tool for various damage severities class will be discussed in Chapter 6. For this purpose, the graphical Gaussian Mixture model via clustering method will be explored in the following chapter to evaluate the prediction of the data class from various damage severities under varying operational loading conditions. In the next chapter, a technique where the basis functions can be adjusted and tuned during training will be discussed. It is called an Artificial Neural Network (ANN). It will be demonstrated how this technique greatly offer different perspective of damage detection and structure's health monitoring than the current's visualisation and novelty detection techniques.

Chapter 6

CLUSTERING AND NON-LINEAR REGRESSION FOR AN SHM SYSTEM UNDER OPERATIONAL LOADING

6.1 Overview

In Chapter 5, the study has presented four PC models for data visualisations based on different data set arrangements in the covariance matrix before performing PCA. It is also demonstrated that by using kernel PCA as a generalised nonlinear PCA, a more efficient way for extracting hidden and nonlinear features via a nonlinear transformation can be achieved. The results derived from standard PCA highlight dimensional reduction and feature extraction on a basis of linear transformation whereas kernel PCA presents a holistic approach in solving the problems related to data overlapping by improving the data class separation. This is addressed by adapting a basis function (kernel Gaussian) to the training set (baseline set) and to each test set (associated with damaged conditions). Using the data sets obtained from the PCA in the previous chapter, this chapter will demonstrate two different approaches by using clustering and nonlinear regression analysis.

The main concern is to obtain a probabilistic classifier model based on a parametric model approach by using GMM. This chapter is mainly divided into two parts. In the first part, probabilistic models are established based on Gaussian mixture model (GMM) with respect to building clusters in the context of latent and observed variables using the data sets. In the context of visualisation, GMM offers an interesting solution by classifying the variables (obtained from the kernel PCA) in the maximum likelihood estimation (MLE) approach. To implement GMM, the efficient Expectation Maximisation (EM) algorithm is used. The goal is to classify various damage classes by incorporating clusters on each group of the data points that are likely to assume different Gaussian distribution models. The data sets are obtained from the feature derived from the kernel PCA as described in the previous chapter.

In the second part, a nonlinear regression method as performed by neural network (ANN) will be presented. The input data is acquired from the feature corresponding to the standard PCA results. ANN provides a different perspective in adapting its basis functions in parametric forms during training, improving the performance and testing the network with a 'new' data set. The goal of performing ANN is to predict data inputs of various damage severities and loading classes given the respective target vectors. The key advantage of using ANN for this work lies in the use of nonlinear functions, parameter optimisations and training algorithms in improving damage severity identification based on the maximum likelihood and Bayesian framework. By specifying the neural network with a sequence of target vectors, the network can learn the relationship between the variables (obtained from the PCA) and the specified target vectors.

With respect to GMM, the concept of posterior probabilities, which refers to the responsibilities of each distribution model in producing each data point in the data set, is introduced.

With the benefit of the availability of the data labels, which is commonly referred to as a supervised learning, the results of the GMM clustering can be simply determined if they are accurate. In a practical scenario, complete data labels will not always be available. However, in the context of SHM, the minimum requirement is that one should have a

baseline data of undamaged condition as a reference set. Therefore, any damage present can be identified by observing distinct separation from the baseline feature.The study will present a different approach of using initial values for the GMM parameters in the beginning of the EM iteration process. These initial values related to the information of parameters (mean and covariance) is associated with the class of baseline data set, which are usually available. Noted that, EM is found to be sensitive to the choice of initial values, thus different sets of initial values would display different clustering results [72].

6.2 Background of Gaussians mixture model (GMM)

The motivation for choosing Gaussian mixture model for the current study is to establish a predictive model for each data cluster based on Gaussian distribution models. The basic idea is that generally data points are assumed to adopt a normal distribution and each of the density models is associated to different damage severities. This chrematistic depends on the mean and covariance as well the mixing proportions in the probability framework. In the study, the data is justified to follow the Gaussian distribution as it is generated using Random signal in addition to the theory of Central Limit, which states that enough random samples from any distribution will lead towards a normal distribution.

GMM is a superposition of linear combinations of the basic Gaussian distributions as shown in Figure 6.1. This combination is essential when it comes to model real data sets to improve characterisation of the data set. This is due to the fact that with only a simple Gaussian density model, it is unable to capture the characterization of the data set [34]. Such superposition formulates probabilistic models are known as Gaussian Mixture Model in which parameters are defined by the means and covariances as well as the mixing proportions of the Gaussian components.

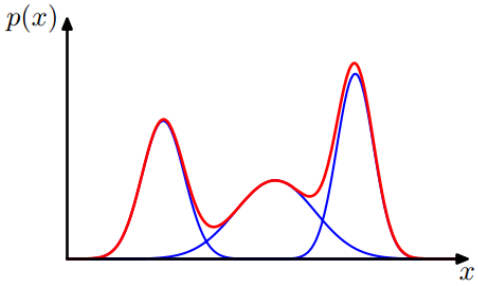


Figure 6.1: Three Gaussian distributions in one dimension (blue lines) shown with a Gaussian mixture distribution as the sum of individual distributions (red line) [34].

The fitting of the Gaussian mixture distribution is implemented using maximum likelihood and the iterative Expectation-Maximisation (EM) algorithm to estimate the parameters (means and covariances) of the Gaussian distributions. In this work, EM is utilised in the GMM framework to estimate the maximum likelihood solutions. Latent variables are introduced as the unobserved variables in the framework so that the joint distributions are more tractable over the expanded space of observed and latent variables.

GMM can be formulated by choosing enough number of Gaussian components *C*, means μ and covariance Σ to describe the Gaussian distributions for the data set. It uses an Expectation and Maximisation algorithm (EM) to fit the parameters of the model, consisting of the means, covariances and mixing proportions of the Gaussians components.

GMM is an extended version of k-means clustering. In k-means clustering, each cluster is specified by its mean in feature space and each data point is assigned to the cluster with the nearest means using Euclidean distance. In the case where data groups may overlap (Figure 6.2) or if described by non-circular shape, it will be difficult for k-means to assign the data points to their right cluster, as k-means clustering does not consider the size and orientation of the cluster. This shortcoming of k-means clustering is avoided with the GMM.

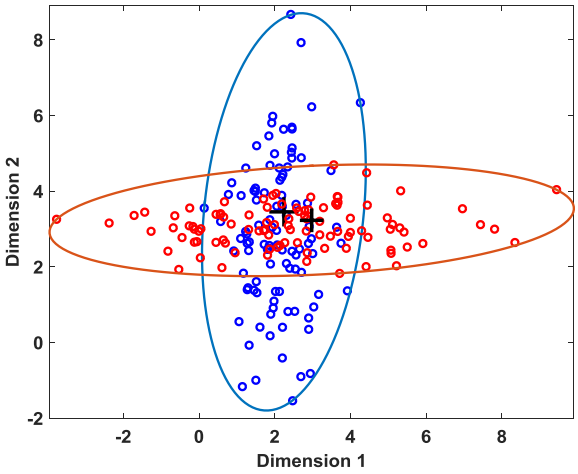


Figure 6.2: An example showing the case of data sets overlap randomly generated using two different Gaussian distribution models

In GMM, the clusters are modelled using a combination of Gaussian components, specified not only by their means but also by the covariance matrix that describes the ellipsoidal shape of the data density. Gaussian mixture model fits a Gaussian distribution to the data points by maximising the likelihood solution using the EM algorithm. GMM is a popular technique for many classification problems such as an unsupervised learning, in which the basic properties such as the number of cluster components, their means and covariance matrix can be unknown. With respect to the non-probabilistic k-means algorithm, GMM has another advantage because it can deal with overlapping data classes that assumes different orientations and shapes. This makes GMM as a more flexible clustering technique compared to the linear k-means clustering. With the implementation of maximum likelihood and the EM algorithm, this allows a soft clustering method feasible on the data sets. Soft clustering relates that the data points can be assigned to more than one component (cluster) of the Gaussians.

Figure 6.3 shows contour ellipses representing different covariance matrices with different orientations to describe the size and orientation of each cluster. In the case of variables dependency and correlation between the variables X and Y as observed in the current data set, a full covariance type whose ellipse has nonzero values at its off-diagonal is used. Diagonal and identity matrix which has zero values on the covariance elements (off-diagonal) shows non-correlation. A full covariance structure is used to represent the

correlations between the spectral variables. In the case of the other two diagonal covariance matrix, the covariance matrices assume zero elements for their off-diagonal terms to describe zero relationship between variables.

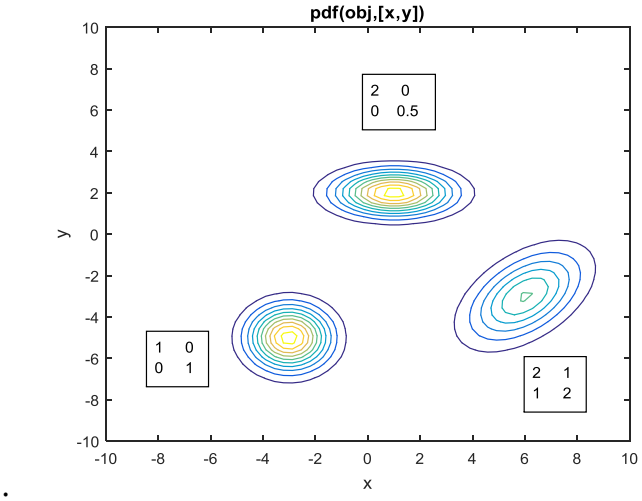


Figure 6.3: The comparisons of different covariance types to show the correlation between variable X and Y

In the next section, the maximum likelihood and Expectation and Maximisation (EM) are described briefly in the context of the parameters of the Gaussian components, consisting of the means, covariances and the mixing proportions. The probabilistic model generated by GMM assigns each data point to the relevant components of the mixture using Bayes' theorem. Bayes' theorem introduces conditional probabilities in which the posterior and prior probabilities have an important role in this GMM. In chapter 3, a detailed description on the mathematical backgrounds on the joint distribution of the Gaussians mixture have been described. In the context of posterior probability, each component (cluster) in the mixture distribution has a computed responsibility value in which it denotes a degree of responsibility the cluster has for generating each data point.

6.3 EM for GMM

Before proceeding with the implementation of an elegant and powerful method using the iterative EM algorithm in the maximum likelihood framework, it is worth emphasizing that there is a significant problem related to the GMM with the maximum likelihood framework. The problem arises due to the presence of singularities. This is a common problem that occurs whenever the Gaussian components collapse onto a particular data point. This implies that when one of the components of the mixture model has its mean exactly equal to one of the data points, this would result the covariance term in the denominator term to go to infinity and the maximum log likelihood function will also go to infinity [34]. Next, an elegant and powerful EM algorithm is motivated here by giving a general treatment for GMM to obtain the maximum likelihood solutions with the latent variables.

Let the process begin by reinstating the equation (3.34) from Chapter 3:

$$\ln p(\mathrm{X}|\pi,\mu,\Sigma)=\sum_{n=1}^{N}\ln\left\{\sum_{c=1}^{C}\pi_c\mathcal{N}(\mathrm{x}_n|\mu_c,\Sigma_c)\right\} \tag{6.1}$$

where the Gaussian distribution is based on the multivariate form,

$$\mathcal{N}(\mathrm{x}|\mu,\Sigma)=\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\mathrm{x}-\mu)\Sigma^{-1}(\mathrm{x}-\mu)^{\mathrm{T}}\right\} \tag{6.2}$$

where μ is the *D*-dimensional mean vector, Σ is the *D* X *D* covariance matrix and $|\Sigma|$ denotes the determinant of Σ.

Setting the derivatives of equation (6.1) with respect to the means μ_c of the Gaussians components to zero, the maximum of the log likelihood function is obtained:

$$0 = -\sum_{n=1}^{N} \overbrace{\frac{\pi_c \mathcal{N}(\mathrm{x}_n | \mu_c, \Sigma_c)}{\sum_j \pi_j \mathcal{N}(\mathrm{x}_n | \mu_j, \Sigma_j)}}^{r(z_{nc})} \sum_c (\mathrm{x}_n - \mu_c) \tag{6.3}$$

Note that the posterior probabilities or the responsibilities appear on the right hand side of the derivatives of the log likelihood function in equation (6.1).

Multiplying by $\sum_c^{-1}$ and rearrange the equation (6.3) gives

$$\mu_c = \frac{1}{N_c} \sum_{n=1}^{N} r(z_{nc}) \mathrm{x}_n \tag{6.4}$$

The mean for the c^{th} Gaussian component μ_c is computed by taking a weighted mean of all data points in the data set whose weighting factor for data point x_n is given by the posterior probability $r(z_{nc})$ in which the component c is responsible for generating the data point x_n. Defining N_c as total number of data points effectively assigned to cluster c. The sum of this soft membership or fractional weight assigned to cluster c is described as

$$N_c = \sum_{n=1}^{N} r(z_{nc}) \tag{6.5}$$

The covariance matrix in the frame of maximum likelihood solution is given in terms of the weighted responsibility for the component c that generates the data point can be stated as

$$\Sigma_c = \frac{1}{N_c} \sum_{n=1}^{N} r(\mathrm{z}_{nc})(\mathrm{x}_n - \mu_c)^T (\mathrm{x}_n - \mu_c) \tag{6.6}$$

It is the equivalent form with the corresponding result for a single Gaussian with each data point weighted by the corresponding posterior probability $r(z_{nc})$ and multiplied by the inverse N_c.

π_c is the mixing coefficient for the c^{th} component given by the average responsibility which the component takes for explaining the data points. In other words, it is the total responsibility allocated to cluster *c* normalised by the total number of data *N*, given as

$$\pi_c = \frac{N_c}{N} \tag{6.7}$$

EM implements GMM iteratively in two steps. The first step is the E-step that choose the parameter means μ and covariance $\sum$ and mixing coefficients as fixed values. The current values for the parameters are used to evaluate the posterior probabilities or responsibilities given in equation (3.34). Then for each data point x_n and each cluster *c*, the responsibility $r(z_{nc})$ or the posterior probabilities that measures the corresponding probability which the component *c* is responsible for generating x_n is computed. The latent variable *z* is used in corresponding with the observed observations x as a joint distribution. The probability of x, $p(x)$ assigned to component *c* with a weighted Gaussian π_c is normalised by the total values of *c* which gives a similar form of equation (3.33). Here the responsibility of data point x belongs component *c* is reinstated as:

$$r(z_{nc}) = \frac{\pi_c \mathcal{N}(x|\mu_c, \Sigma_c)}{\sum_{j=1}^{C} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)} \tag{6.8}$$

Practically, $r(z_{nc})$ is a number of data points by number of clusters that sums to one over the index *c*. If x is very likely to be under the Gaussian component *c*, it will get high responsibility value $r(z_{nc})$. The denominator just makes the sum of $r(z_{nc})$ equal to one.

The second step in EM is the maximisation step. It starts with the probability assignment $r(z_{nc})$ and update the clusters' parameters μ_c, $\sum_c$ and π_c. The parameters are weighted by $r(z_{nc})$ so that if x_n is a strong member of cluster *c*, this weight will be nearly one, but if x_n is not very well explained by cluster *c* then it will not influence the average very much.

For each cluster *c*, the parameters are updated using the estimate weighted by the probabilities $r(z_{nc})$. By updating the new means, covariances and mixing coefficients

followed by maximisation, it guarantees to increase the log likelihood solution and make it converge. The EM assumes many iterations to reach a convergence and therefore requires significantly higher computation compared to the k-means algorithm.

In this work a confidence ellipsoid with an obtained score that satisfies a 99% probability threshold for each component is used as this set is typically used in many applications [72]. These scores determine the length of major and minor axes of the ellipsoids.

In the next section, GMM is applied onto the data feature acquired from the kernel PCA baseline model A (eigenvalues decomposition from baseline model of undamaged condition from all loading sets) and kernel PCA model D (matrix of all combined data set of undamaged and damaged conditions from all loading sets) under some different initialisations.

6.3.1 EM steps for GMM

The goal of implementing GMM is to construct a classifier by maximising the likelihood function with respect to the parameters (the means, covariances of the components and the mixing coefficients) [34].

1. Initialise the means μ_c, covariances $\sum_c$ and mixing coefficients π_c and evaluate the initial value of the log likelihood.

2. In E step, evaluate the responsibilities using the current parameter values as illustrated in equation (6.8).

3. In M step, update the parameters μ_c, $\sum_c$ and π_c using the current responsibilities using

$$\mu_c^{new} = \frac{1}{N_c}\sum_{n=1}^{N} r(z_{nc})\mathrm{x}_n \tag{6.9}$$

$$\sum_c^{new} = \frac{1}{N_c}\sum_{n=1}^{N} r(z_{nc})(\mathrm{x}_n - \mu_c^{new})^T(\mathrm{x}_n - \mu_c^{new}) \tag{6.10}$$

$$\pi_c^{new} = \frac{N_c}{N} \tag{6.11}$$

where

$$N_c = \sum_{n=1}^{N} r(z_{nc}) \tag{6.12}$$

4. Evaluate the log likelihood and check for the convergence

$$\ln p(\mathrm{X}|\mu, \Sigma, \pi) = \sum_{n=1}^{N} \ln \left\{ \sum_{c=1}^{C} \pi_c \mathcal{N}(\mathrm{x}_n|\mu_c, \Sigma_c) \right\} \tag{6.13}$$

If the convergence is not achieved, continue to step 2 and repeat the steps.

6.4 The application of GMM as a damage predictive model

Having informally described the treatment of EM in the context of GMM, the algorithm will then be applied on the data set obtained from the kernel PCA. Each data point will be assigned to a Gaussian mixture model according to the maximum likelihood framework. The implementation of GMM in this work is performed using the Statistics and Machine Learning toolbox of User's Guide (R2017b) [72]. The rationale of taking the kernel PCA as the data input for the GMM is because the data class is more discrete compared to those obtained from the linear PCA. Consequently, this can help to avoid the problem due to

singularity during the EM iteration process. To implement GMM, some criteria are introduced in terms of the parameter initialisations. The initialisations for mean, covariance and mixing proportions of which values are fixed during the beginning of the E step takes one of the following forms:

- The initial values of the parameters for the means, covariance and number of components correspond to the parameter values of the undamaged condition components.
- Each data point in the data set is associated and connected with the categorical class label its data points.
- Naturally perform the clustering without using any information about the number of components and their parameters (means and covariances), only by using k-means algorithm.

6.4.1 Classifier model using kernel PC model A data input

The first step is to construct a GMM classifier based on the undamaged baseline model obtained from kernel PC model A (Figure 6.4). This GMM classifier takes the initial values from the baseline model. The advantage of such a model is due to its convenience in that it only needs to establish the parameters from the undamaged state compared to a more costly damaged state. Practically, new test data obtained using the same kernel PCA model will be used with the same initial values of the parameters belonging to the undamaged condition.

EM algorithm will then update the parameter values, evaluate the maximum log likelihood of the posterior probability and repeat the steps until the criteria for convergence is met. The goal of this model is to obtain a probability model based on the mixture of Gaussian distribution model for each data group associated with a damage severity through all loading conditions. By fitting the ellipsoidal shapes on the data clusters based on the maximum log likelihood function, the results are then compared with the original data groups to check its performance. This study aims to validate if GMM based on the maximum likelihood function can correctly classify the damage severity groups and form

the clusters that form ellipsoidal shapes on the data points based on the means and covariances of the undamaged state.

The results from Figure 6.5 show the fitting of ellipsoids on each data test using the GMM-EM algorithm that takes the number of Gaussian components as five, which represents the five loading conditions. The GMM algorithm fits the ellipsoidal on each Gaussian component produced from different structural health conditions. The highlight of the results is that it shows a potential application as a predictive model with the condition that the data should be pre-processed by PCA prior to GMM so that the features are present in discrete data groups as shown in Figure 6.4. The advantage of this model is that it only requires the parameters from the baseline data set. Note that if a higher number of components are used, the complexity of the likelihood problem also increases.

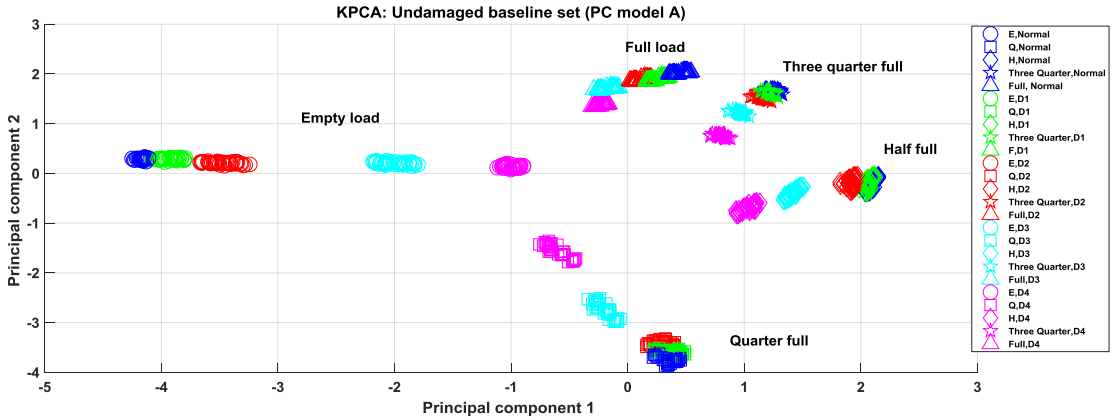


Figure 6.4: The original data from kernel PCA utilised in the GMM.

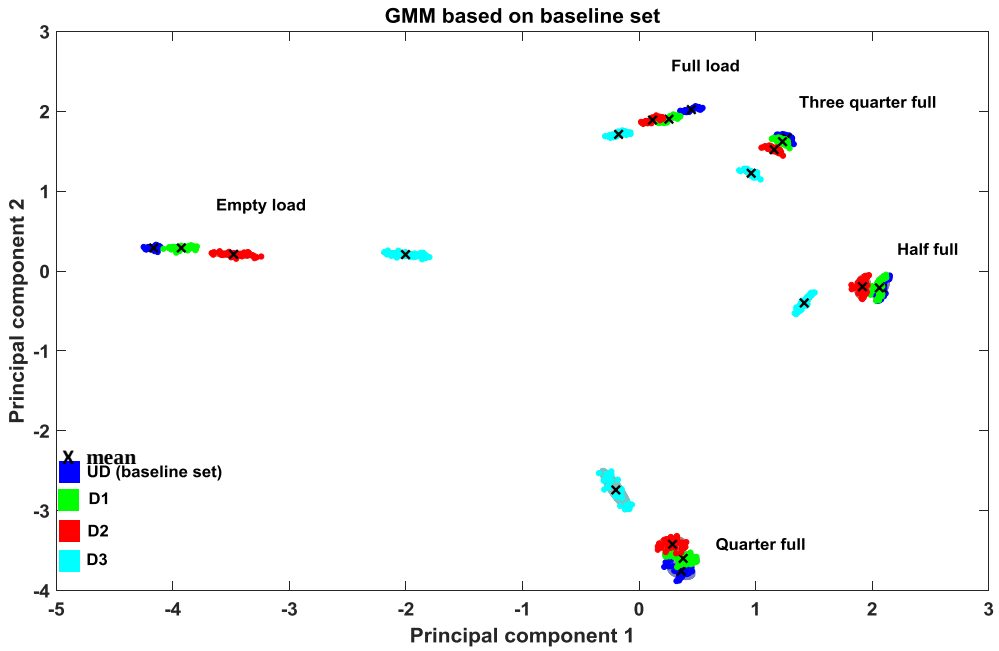


Figure 6.5: Using only the initial means and covariance of the baseline data group (undamaged condition).

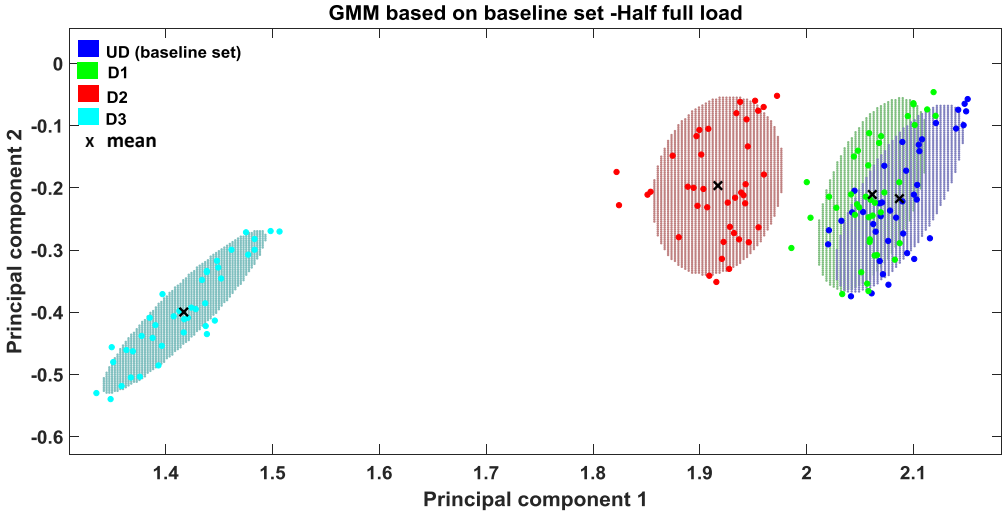


Figure 6.6: Zooming in on the half-full load to highlight the cluster characteristics with respect to changes in damage severity.

In Figure 6.5, the GMM algorithm assigns the new data points to their corresponding Gaussian components arbitrary well when the data set corresponding to a specific damage condition is applied to the algorithm. For subsequent new data sets as indicated by green, red, cyan colours, their initial values for the means and covariances are all corresponding to the baseline parameters. A zoomed view on the half-full load class results in Figure 6.6 shows a consistent cluster translation, moving away from the undamaged (UD) state as the damage level grows. It reveals very accurate prediction of test data that uses only the baseline set as the initial conditions. The result is encouraging, considering that all the means and covariances of the initial values of the test data are based only on the baseline model of the undamaged condition.

However, the main setback of this model is that the number of Gaussians should be specified and each new data set is applied corresponding to a specific damage condition. This situation should not give many obstacles since the number of components can be easily determined from the PCA analysis. As has been demonstrated earlier in the last chapter, the loading effects are effectively isolated from the damage effects by using PCA and with GMMs, the clusters associated with various damage states are grouped together denoting a specific operational loading condition (Figure 6.5).

Practically, this model can be used to establish a predictive model by observing the separation of the cluster from the clusters belonging to the baseline set. It implies that as the clusters of damage states move away from the baseline clusters, the damage is more severe and require greater attention from the maintenance personnel.

In the next figure, the overall baseline and test data are applied all together at once, which means the algorithm computes all 25 classes based only on their labels (Figure 6.7).

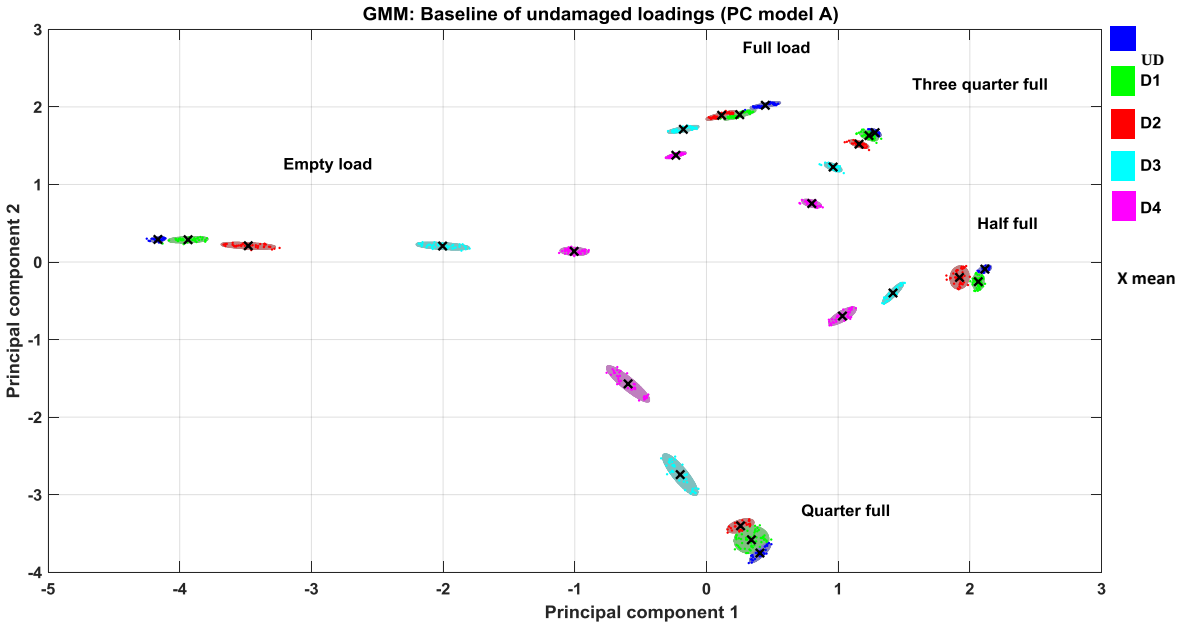


Figure 6.7: Fitting GMM by connecting each observation to its label for all 1000 observation data points.

The results of GMM classification of all data points by linking each data point to its respective labels or names are generally excellent (Figure 6.7). All the ellipsoids revealed good consistency in correlation to damage severities that describe the increase in damage as the clusters demonstrate higher separation from the baseline cluster. No other parameters are required except the total number of components which needs to be specified. A close view onto the damage clusters at the quarter-full load (Figure 6.9) reveals the orientation and shape of the damage clusters in more detail, which is in agreement with the actual feature as in Figure 6.8.

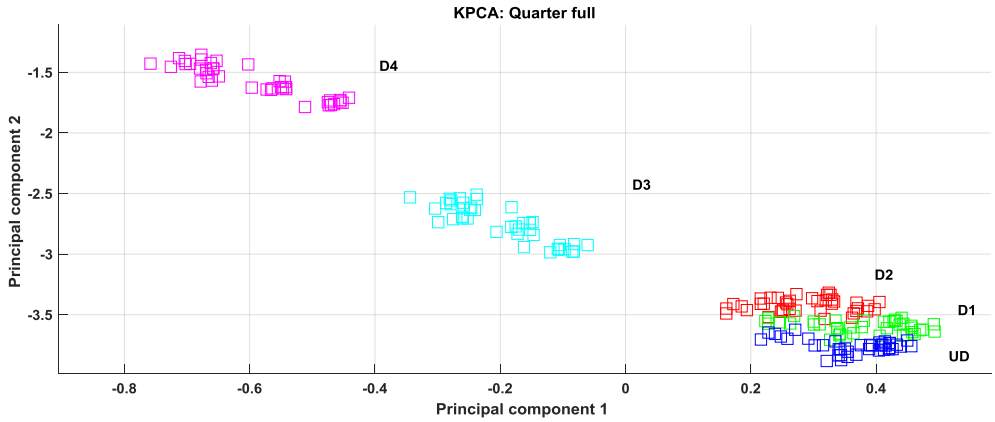


Figure 6.8: The original data points associated with the corresponding damage severity.

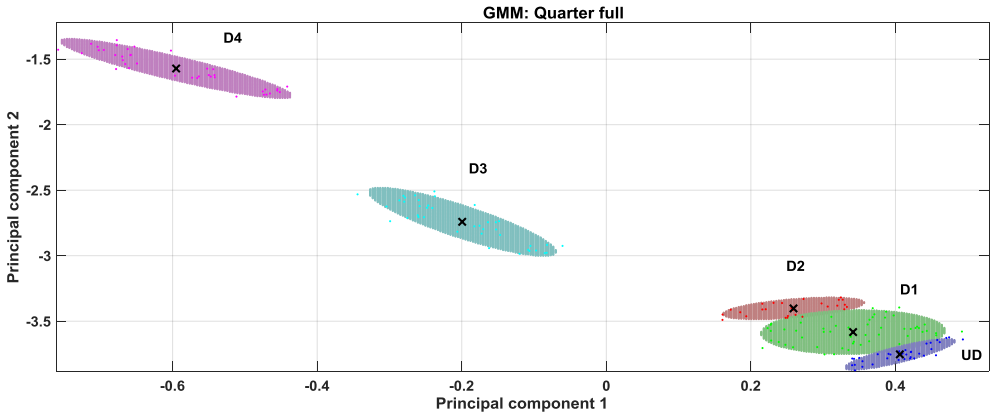


Figure 6.9: A zoom in on a quarter full load that describes the Gaussian density model.

However, it is observed that, the EM-GMM algorithm includes some top end data points from undamaged (UD) class (as illustrated in Figure 6.8 by a dash-oval) into D1 cluster (Figure 6.9). This is expected as the small portion of data points of UD class lay close to the D1 groups and predicted to be generated by the Gaussian component belonging to D1 class. As a result, the small set of data points is included into the D1 ellipsoid as illustrated in Figure 6.9.

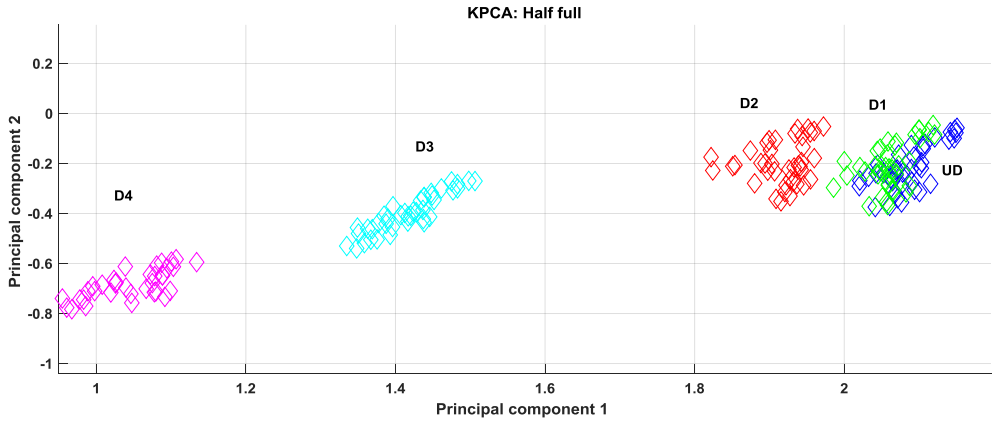


Figure 6.10: The original kPCA feature showing the position of each damage severity groups.

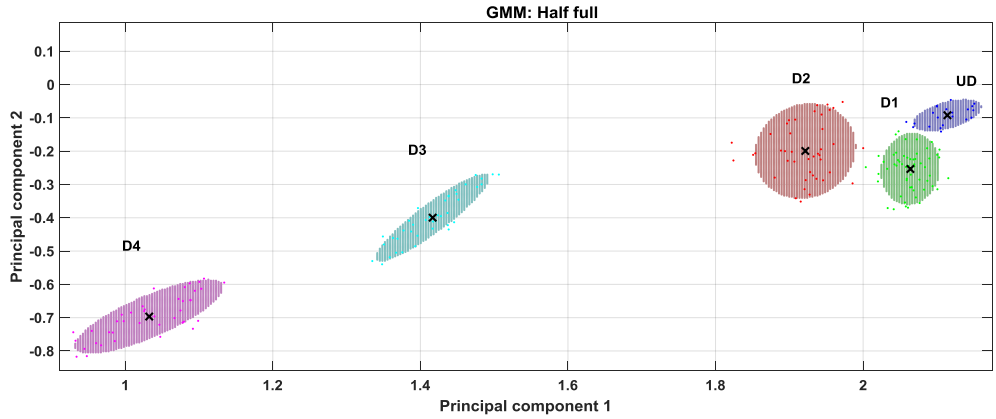


Figure 6.11: A zoom in on half-full load to show clearer results of the classification.

In the case of the half-full load, it is observed that the UD and D1 clusters (Figure 6.11) have a different data density model by the way the ellipsoids are made with respect to their original feature in Figure 6.10 after the GMM implementation. Rationally, the data points (belonging to UD and D1 class) that seemed to be likely generated by the two distinct Gaussian distributions are clustered independently as two components as illustrated by Figure 6.11. Nevertheless, the rest of the clusters are correlated with damage severities.

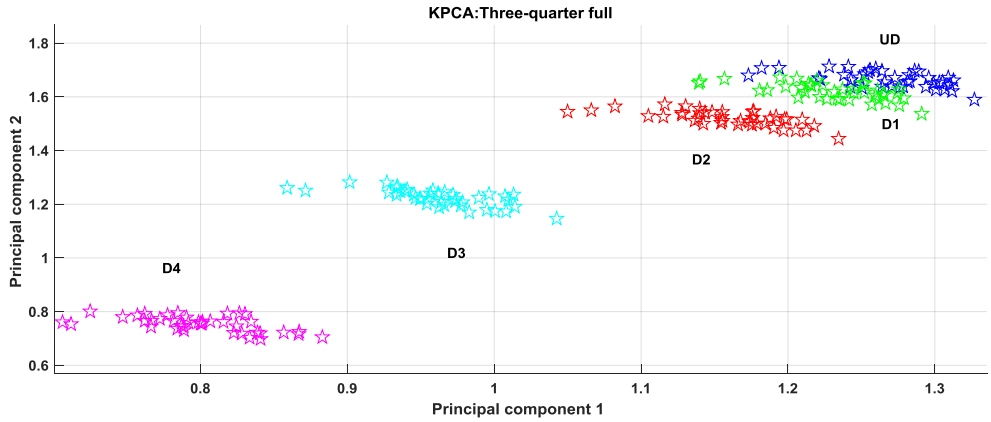


Figure 6.12: The original feature showing all the data groups corresponding to various classes of damage severity.

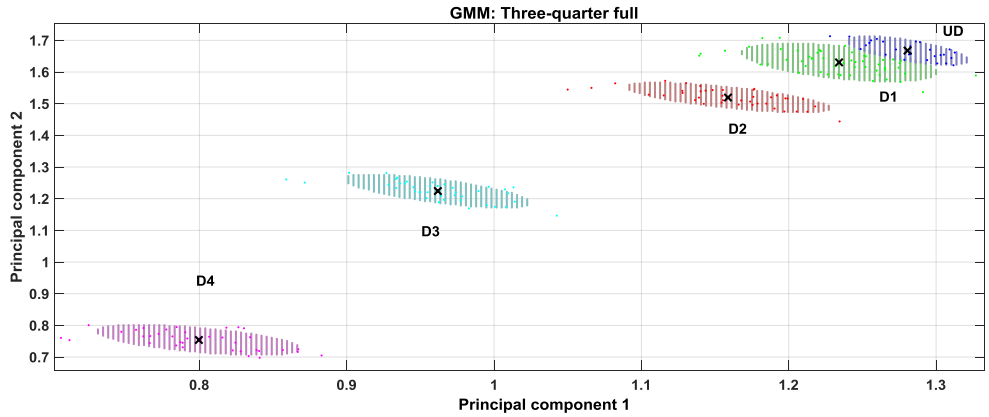


Figure 6.13: The result of the GMM classification for varying damage severities in three-quarter full load.

In the three-quarter full load class (Figure 6.13), the clustering of the data points belonging to various damage severities are showing some encouraging results. All the data clusters are highly correlated with the original feature (Figure 6.12) with respect to the damage level. With regard to the data group UD laying close to D1 class, the corresponding data points are assigned proportionally to their respective classes (Figure 6.13) by comparing with the original feature. It can be stated that from these observations (Figure 6.12 and Figure 6.13), the orientation of data groups consisting of densely distributed data points play a significant role in classifying them as one Gaussian component.

The results from these two approaches (initialisation by the baseline parameters and connecting data points to corresponding labels) have shown positive outcomes as the data classifiers model based on the EM algorithm and the mixture of Gaussian distributions framework. One main concern found in the analysis is that one needs to specify the number of Gaussian components (or clusters) to be suitably created by the algorithm. On the other hand, the positive remark is that the number of components can be easily determined from the PCA results. Another setback realised is that given the data groups sit very close together, the orientation of clusters or to an extent the data points, can be wrongly assigned as clusters corresponding to a different class.

Generally, based on these findings, GMM shows relatively promising results for damage cluster classifications given the true labels in the categorical array and also based on the means and covariances values of the baseline data set. The result has also shown that the GMM-EM algorithm performs very well with respect to correlation with the level of damage severity.

In the next GMM-EM test, the data set derived from the kernel PC model D is implemented using a different approach in selecting the initial GMM parameters.

6.4.2 Utilising kernel PC model D in GMM prediction analysis

In the following analysis, the GMM-EM algorithm is applied to classify the various damage severities under one specific loading condition. This is in comparison to the earlier model in which data are classified under various loading classes. In this model, it concerns the accuracy and orientation of the data clusters corresponding to the damage severity.

The data set from model D of the empty load and quarter full load is applied. In the first test, only data from the empty load is applied. It is useful to test the data groups of each loading set individually as given more components in the algorithm, the computation becomes more complicated and its iteration takes much longer. In this test, the means and covariance parameters are explicitly specified for each component parameter.

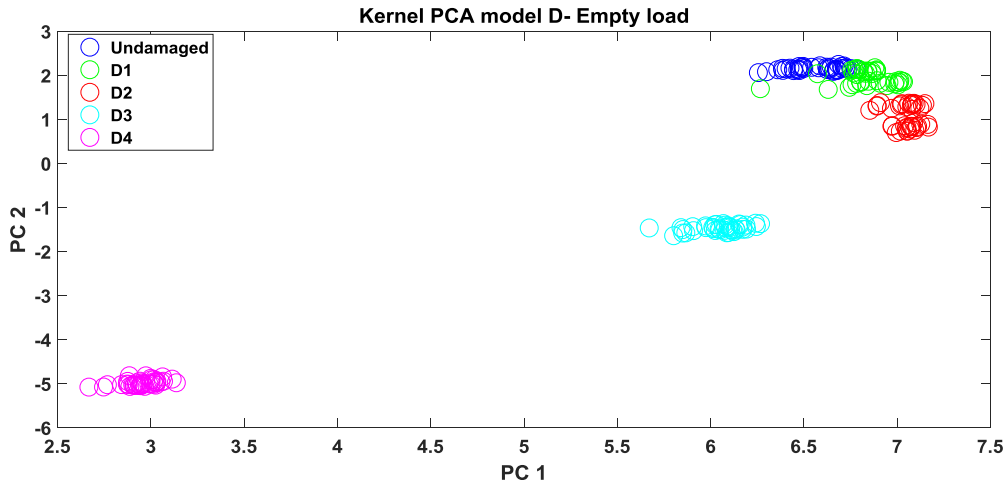


Figure 6.14: The original data set with damage specifications.

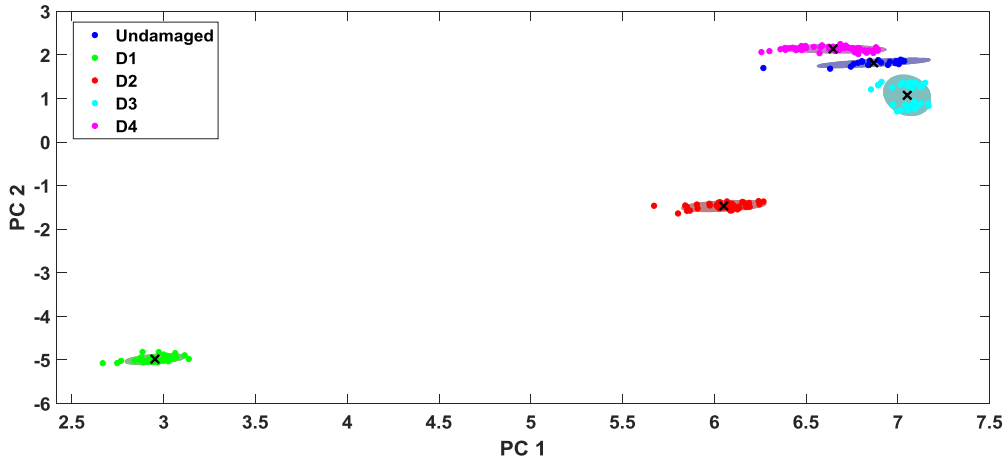


Figure 6.15: The first approach for finding the initial values with k-means algorithm.

The result obtained in Figure 6.15 shows a promising classification model using k-means algorithm initialisation method. Using k-means algorithm, the initial choice for the means is performed arbitrarily and the solution in the log likelihood is to converge when the change in the function falls below a certain threshold. The threshold is calculated based on a chi-square distribution of which its value exceeds 99% of the samples. The classifier successfully classifies the data groups into each cluster in comparison to the original data groups in Figure 6.14.

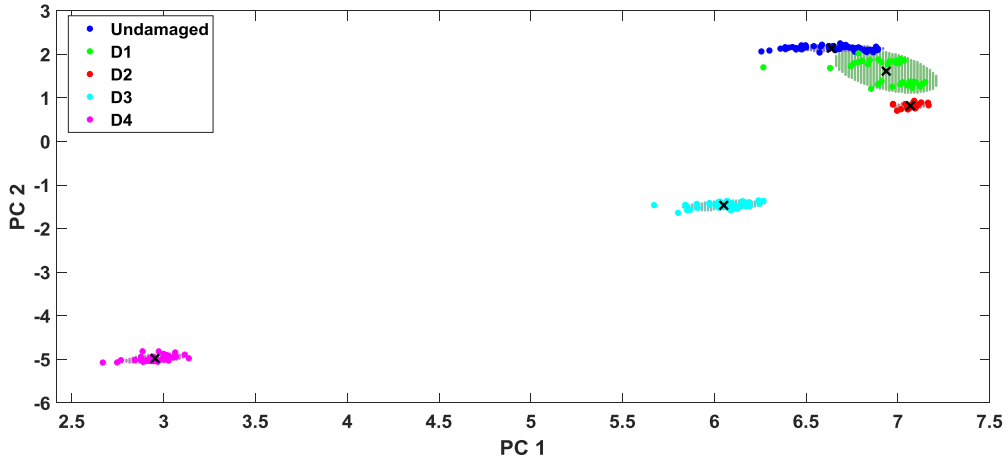


Figure 6.16: The second approach using the initial mean and covariance of the first cluster.

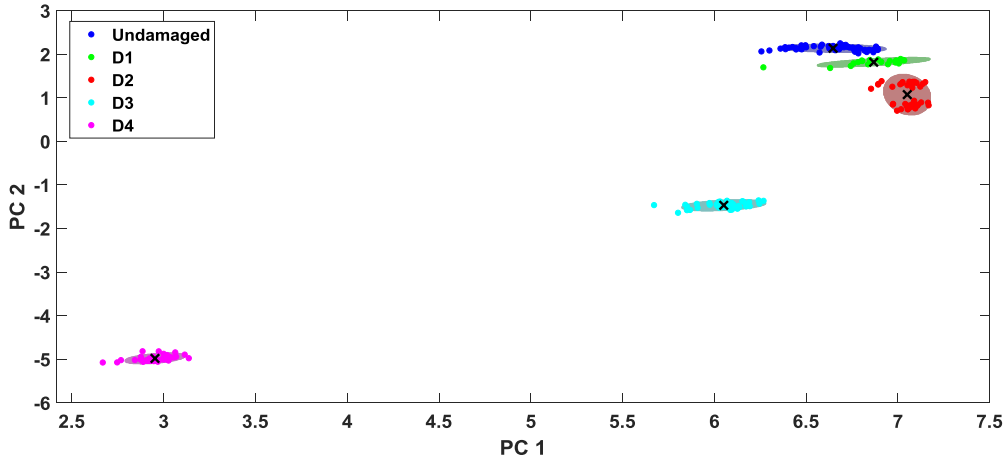


Figure 6.17: The third approach by connecting each observations to its label.

Overall, the results in using these three approaches are generally promising with the initialisation method by k-means clustering (Figure 6.15) and connecting data points with their labels (Figure 6.17) which are comparatively accurate. Again, the main concern are the data groups that are located close together. Here, UD, D1 and D2 clustering are slightly different in comparison to the original feature in Figure 6.14 when the initial parameters are initially fixed by the undamaged parameters (Figure 6.16).

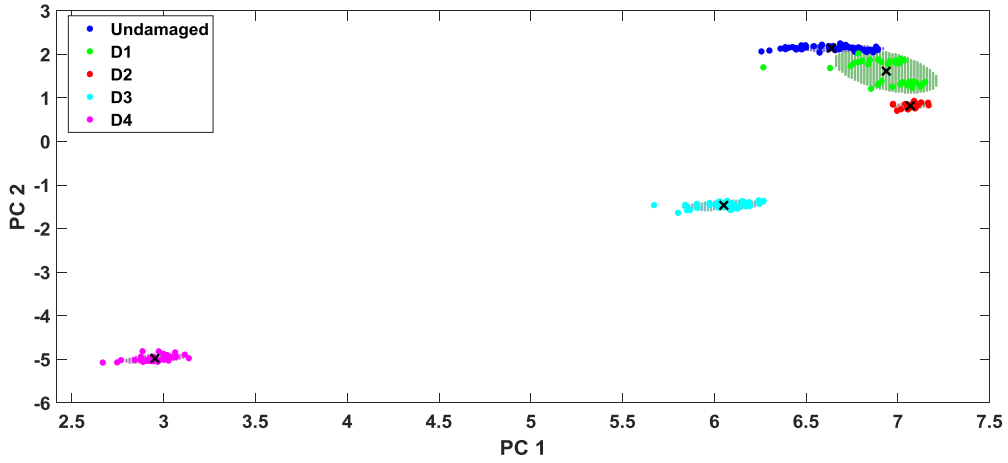


Figure 6.18: The second approach using the initial mean and covariance of the first cluster.

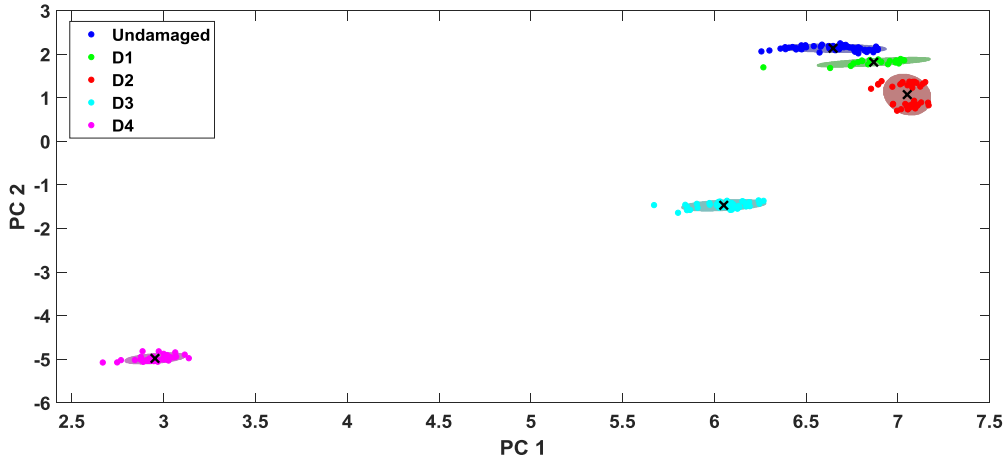


Figure 6.19: The third approach by connecting each observations to its label.

As overall, the results in using these three approaches generally are promising with the initialisation method by k-means clustering (Figure 6.15) and connecting data points with their labels (Figure 6.17) which are comparatively accurate. Again, the main concern is that the data groups that located close together. Here, UD, D1 and D2 clustering are slightly different in comparison to the original feature in Figure 6.14 when the initial parameters are initially fixed by the undamaged parameters (Figure 6.16).

A zoomed view on result of the approach using initialised baseline parameters and the data points labels (Figure 6.18), it is indicated that classification of D1 data has wrongly recognised the upper portion of D2 data group as part of D1 data group by the second approach of using baseline's initial parameters. In respect to k-means clustering, the methods successfully classify these data groups. It is understandably, that high data separation in D2 data class allow the upper half of the data groups to be assigned to either D1 class or correctly included into D2 data class.

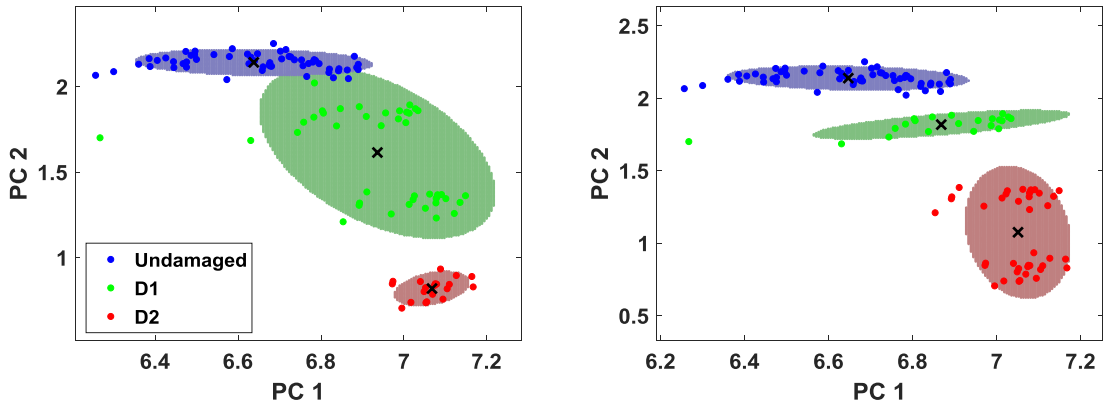


Figure 6.20: The comparison between the second approach (baseline parameters) (left) and third approach corresponding to data points label (right plot).

It can be identified from the earlier PCA result that in fact the actual number of components (clusters) is ten. However, for the purpose of verifying the robustness of this technique and to test the algorithm on the current problem, the actual number of components are implicitly unspecified. The algorithm is then used to generate clusters based on the number of components specified from the range between 1 to 12 (Figure 6.19). During the 13th iteration, when computing the covariance for the 8th components, an error shows up in the process, indicating the error is due to an ill-conditioned covariance created. This is not a real surprise as it is recognised that, making the algorithm to generate more than 8 components, there is high possibility for singularity to occur as the data points are overlap for UD and D1 data class as illustrated in the Figure 6.20.

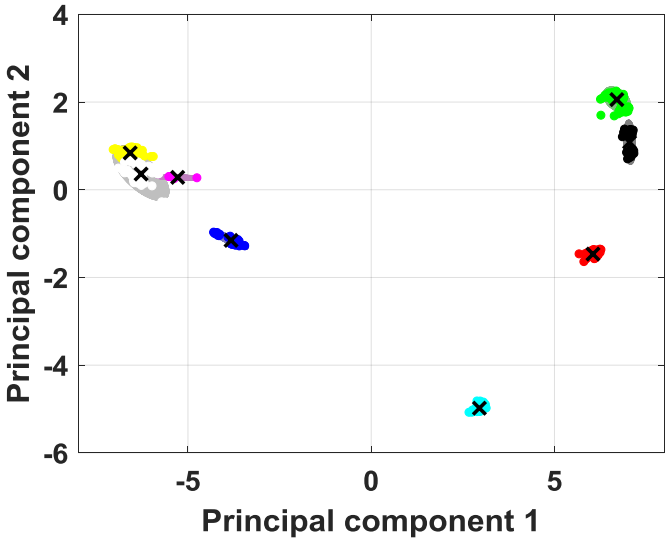


Figure 6.21: GMM naturally assigns eight clusters for overall data groups.

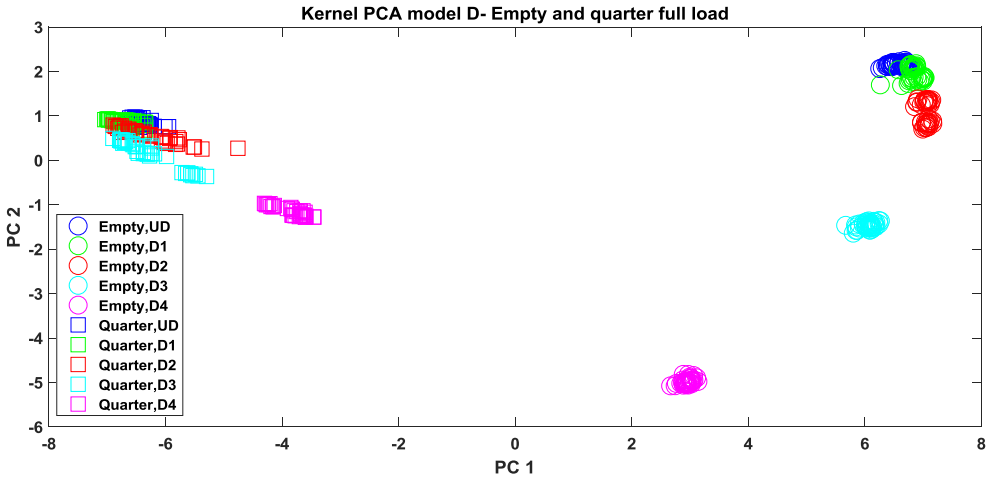


Figure 6.22: The original data groups corresponding to respective damage class.

The result in Figure 6.19 corresponds to the original feature (Figure 6.20) describing the clustering of the data groups by iteratively specifying the number of components from 1 to 10. During the iteration, it is found that the algorithm only managed to perform the iterations up to 8 components before reaching the convergence of the maximum log likelihood solution (Equation 6.3). The singularity issue is raised when the components are

overlap as illustrated in the actual feature plot of Figure 6.20. Nevertheless, it is acceptable to say that, the result in Figure 6.19 indicates that the data distribution model generated in this experiment can potentially be modelled based on GMM with some restrictions to the overlap data class. Nevertheless, the mixture of Gaussians can be used to provide some general indications to the data behaviours in respect to changing of loading conditions and damage severities based on the PCA data features as the input to the GMM-EM algorithm.

As a brief conclusion, the maximum log likelihood framework has the prospect to be used in SHM in respect to the effects of loading variations with some limitations due to singularity. Potentially, it can be used as a generative model given under the situation that there is a lack of real data set. It can be used as a predictive model of a parametric model with the goal of monitoring the structural health. It is important to note that, the role of PCA in providing data inputs for the GMM-EM algorithm. The quality of the features obtained from the PCA can affect the performance of the GMM classification. It is highlighted that, the main issue raised in using GMM is with overlap data groups where the singularity problem can occur.

Previously, with GMM, a classifier model based on a probability model is established. Data points of different damage class are assigned to each clusters based on mixtures of Gaussian distributions and maximum likelihood framework. In the following work, a non-linear regression analysis based on the powerful Artificial Neural Network (ANN) will be introduced. It will present the input-target mapping model which is a non-linear regression model by using the data inputs from the PCA.

In that section, ANN will be addressed with respect to training algorithms and a regularization approach as to generate accurate fitting results. It shall begin by introducing neural network, its descriptions of underlying parameters and highlights the optimization technique in order to avoid overfitting of the data set. The final finding will conclude its results for comparing the maximum likelihood with the Bayesian approach concerning of determining the optimal network parameters.

6.5 Introduction to neural network

Neural network finds its origins back in the1940s as scientists and engineers attempted to find mathematical representations of information processing in the way of biological systems [34]. At present, with the advancement of computer technology, the field of ANN is evolving quickly and has became a powerful tool for many real applications. It has good records of accomplishment in solving large-scale problems especially where linear models have limitations [34].

The main advantage of neural network is that it allows the nonlinear basis functions to be adaptive and adjustable in the form of parametric model during training. Training here refers to the process of learning from the given sets of input features by the network neurons (nodes) by adjusting the weights. By doing so, the neurons can perform the pattern recognition task correctly. The most successful neural network in pattern recognition landscape is the feed-forward neural network that is also known as a multilayer perceptron (MLP) [34], [73]. The benefit of the MLP lies on the continuous nonlinearities activation functions through its multiple layers where the input data is transformed into the output. The network comprises mathematical representations has greater practical values compared to the multiple perceptron models with discontinuous nonlinearities [34].

In this work, the attention will be restricted to the regression neural network using feed forward fitting function. The aim of any neural network is to minimise the sum-of-squares error function based on the selected network parameters. The error function is a quadratic function with a linear derivative in respect to the parameter w's which can be written as

$$E(\mathrm{w}) = \frac{\mathbf{1}}{\mathbf{2}} \sum_{n=\mathbf{1}}^{N} \left\| \mathrm{y}(x_n, \mathrm{w}) - \mathrm{t}_n \right\|^{\mathbf{2}} \tag{6.14}$$

where $\{x_n\}$ is input vectors and $\{t_n\}$ is target vectors and n=1,…,N.

The work in this study centred on regression analysis where the prediction of a continuous variable y from an input vector x. Its aim is to create a mapping from input variables of vectors $\{x_n\}$ to output values $y(x_n, w)$ by approximating the outputs to the given target vectors $\{t_n\}$ where n=1,…,N as closely as possible. As stated previously, the pre-processed data expressed in terms of the PCA feature variables have particular patterns are used as the neural network inputs which their corresponding categorical conditions are known. This application is known as a supervised learning and the purpose is to train the neural network. The recognition performance of the trained network is then used to evaluate the new test data which is not included in the training database.

In this study, the exercise is implemented using Matlab Neural Network toolbox for its comprehensive and convenience toolbox. The highlights of this work includes a Bayesian treatment used to determine the network parameters in comparison to the maximum likelihood approach. Both approaches are finally compared with respect of minimising the error function.

6.6 ANN architecture

A feed-forward neural network with multilayer perceptron is one of the most effective and successful pattern recognition technique for model regression and classification when dealing with large data sets and high data dimensionalities [34]. In most neural network applications, the main interest will be to obtain the best generalisation performance of the trained network on a given test set which is not included in a training database. Good generalisation performance ensures the network with new data input produces optimum balance between over-fitting and under-fitting.

It has been identified that a two-layer MLP with sigmoid non-linearities can approximate most functions with arbitrary accuracy [34], [57]. Based on this justification, the problem of defining the network architecture reduces to one of choosing the number of hidden nodes. The schematic diagram of the two-layer MLP network with the transformation from

the inputs to the outputs are equivalent to those shown in Figure 3.2 and Figure 3.3 of Chapter 3.

The central element in neural network is the network training based on the data inputs which is performed by a specified training algorithm through its nonlinear activation functions located in the hidden layer (s). Each of data inputs are transformed using a differentiable and nonlinear activation functions in a forward propagation into the next output layer before transformed to give the network outputs.

The parameters weights and biases are adjusted with respect to minimizing the errors of the mean square function performed by the training function. An important consideration to note that if the number of hidden nodes are too high or the layer size is too large, there is a tendency for the neural network to over fit the data and it can give poor data generalization when testing the network with new input vectors. However, it can allow the network to train and solve for problems that are more complex if it provided with more hidden layers. The recommended practise is to begin with a simple one hidden layer and observe if the fitting error is small [34], [74]. Most of the practical and medium sized problems perform well with one hidden layer when given appropriate parameters [34].

6.7 Define the architecture

The first step is to define the type of multilayer network to use in the study and select the nonlinear transformation function in the hidden layer suitable for the current study. There are numerous terminologies used regarding the number of layers in the neural network. In this study, the terminology recommended by [34], [57] such as in Figure 6.21 is adopted in which the layers consisting the weights regarded as the layers of the network. Therefore, the network is considered as a 2-layer neural network.

By designing the neural network architecture, better understanding of the transformation process of the input variables into the output variables via the continuous and differentiable nonlinear activation functions in terms of parameter W can be attained. Figure 7.21

describes the neural network architecture used in the current study. The terminologies used in the architecture are defined as follow:

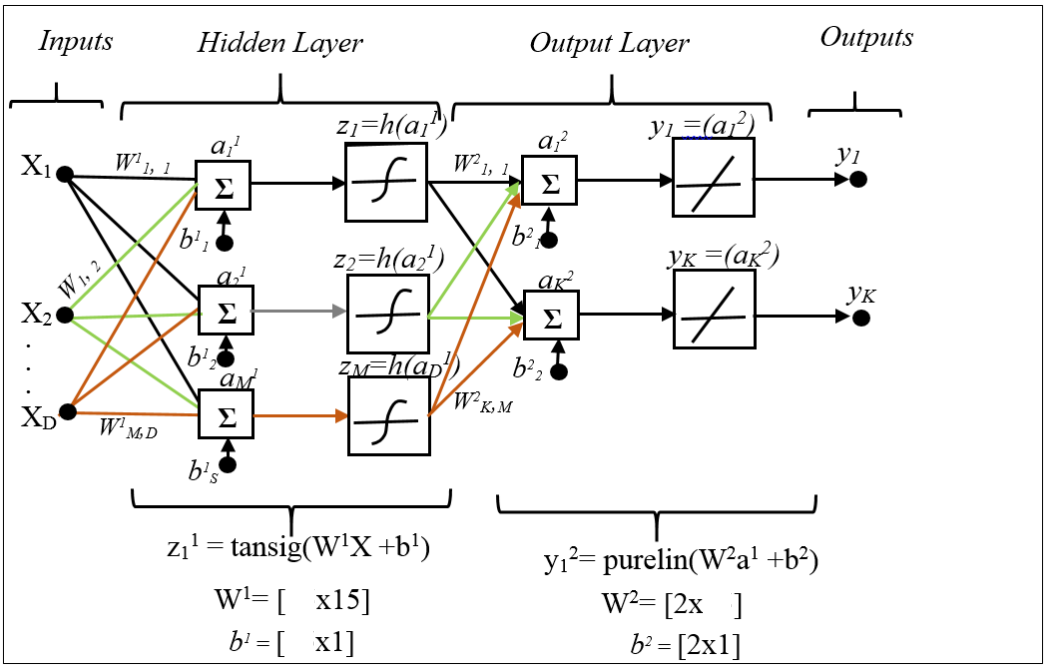


Figure 6.23: Network diagram for the two layer neural network corresponding to the current network architecture used for this study.

Defining the terminologies used in Figure 6.21, $X_{1,\ldots,D}$ is the inputs variables (principal components), D is the total number of input variables, $W^1_{1,\ldots,}$ $W^1_{M,\ D}$ are the parameter weights of the hidden layer, M is total number of hidden nodes, and K is the total number of output variables. The superscript (1) and (2) indicates the corresponding parameter in the hidden layer (first layer) and the output layer (second layer) of the network respectively. Bias b^1 and b^2 are the biases for the first and second layers. $a_1{}^1$ is the input unit activations and $a_1{}^2$ is the output unit activations. Each of the input unit activations a^1 units is transformed using a differentiable nonlinear activation function h to give hidden units z. Often the nonlinear activation function is chosen to be sigmoidal functions such as 'tanh'. In the second layer, the output unit activations a^2 are transformed using an appropriate activation function to give the network outputs y_K. In case of regression problem, a linear function which is the identity $y_K = a_K$ is used. The number of weights in the first layer is determined by the number of hidden nodes and number of input variables used in the

network as in this network it is W^1 = [23 x 15]. In the same concept, the number of weights for the output layer is given by W^2 = [2 x 23], where the value X_D = 15 is the number of input variables D, z_M = 23 is number of hidden nodes and y_K = 2 is the number of output variables.

The network learns the nonlinearity of the training data by using a powerful yet a simple training method which is a backpropagation technique. It uses a gradient descent which is applied to a MSE function. Generally, the backpropagation method refers to the modification of weights that begins between the hidden layer and output layers so that the fitting error can be decreased based on the MSE function. The errors are propagated backwards trough the network to evaluate the derivative function. For full details of the technique can be found in [74].

Despite the clear advantage of ANN in solving complex practical problems in pattern recognition, it is important to acknowledge some of the challenges and limitation concerning neural network. The main challenge is that it requires an estimate of the parameter hidden nodes in establishing the network so that the number of parameter weights can be determined. In the following section, it will describe some of the general best practice in ANN implementation. This posed as a general guideline and a summary of steps in implementing ANN for the current study.

As mentioned earlier, in general the problem of the 2-layer feedforward MLP neural network is mainly due to determine the optimal number of hidden nodes for the network. The numbers of these hidden nodes contributes to the overall network performance. This problem can affect the generalisation performance of the neural network. Other selection of network properties depends on the nature of the problems either it is regression or classification. This includes type of activation function, training algorithms used in the network training and regularisation to be used. Essentially, the free selection of the hidden nodes control the weights and biases in the network and they can be changed and adjusted during training to produce better predictive performance.

If too many of hidden neurons included, it potentially results in overfittings or poor generalization of the training data. On other hand, if very few of them is used, the network

might missed some features of the training data and learn them inadequately; this may result in large error in training data set and result to poor fittings of the input variables onto the target vectors. In general, the number of nodes in the hidden layer should be more than the number of variable inputs and outputs to avoid loss of information loss.

The selection of hidden nodes (neurons) should ideally balance the performance of the network in respect to the prediction error and overfitting of the features as well to control the total number of weights in the network. The optimum selection of number in neurons would provide good generalization of target prediction using new test data by the trained network.

6.8 General practice in ANN

The main challenge encountered in processing the data using neural network at the first time is to determine the appropriate use of number of hidden nodes, M. Because M is a free parameter, it can be adjusted to give the best generalisation performance. The important of M is significant because it determines the weights and biases for the network in case there is limited number of data inputs and the number of outputs in the network. This section provides

6.8.1 Number of training examples

Information theory recommends that the number of weights, W should be of the same order as the number of training inputs used in the network, P [57]. The free parameters of network weights W in relation to the network topology I-M-O multilayer perceptron is given as

$$W = (I + 1)\, M + (M + 1)\, O \tag{6.15}$$

where I is the number of inputs, M is the number of hidden nodes and O is the number of output units. Each of the hidden and output units has a bias weight associated with the units. Tarassenko 1998 summarised the recommendations for estimating the parameter W as [57]

A lower bound for training patterns P = W
A realistic upper bound for P = 10 W

Applying the guidelines as given in equation (7.15) and the realistic upper bound to the data set together with the parameters I = 15, P = 700, O = 2, gives

W = 18 M + 2
where P = W for lower bound and P = 10 W for upper bound
Using equation (7.15) for lower bound: 700 = 18 M + 2
and for the upper bound: 70 = 18 M + 2
the estimates give $39 \geq M \geq 4$

A practical approach is to perform iterations using the training inputs corresponding the fixed neural network architecture while changing the range values of M and plot against the MSE. The details of the approach is discussed in Section 7.8.6. In this current work, a more detailed study on the selection of the optimal number of hidden nodes is carried out in order to achieve the best generalized performance of the network.

A practical approach on choosing the optimum number of the hidden nodes can be based on the performance of validation set result corresponding to its smallest MSE as illustrated in Section 7.8.6. Based on the practical approach, M = 23 is found to be the appropriate value with respect to good generalisation performance of new input data set.

6.8.2 Selecting the number of PCs for network input

It should be noted that, neural network is a data driven model and the quality of the model depends on the quality of the data set used in network training. Therefore, with respect to

making the PCA results as the input to the neural network, the selection of the total number of PCs to be selected is an important factor.

The input data acquired from the standard PCA features is given as: Nx1000 matrix where N is the number of principal components (PCs) for 1000 observation points. The choice of N should ideally represent most of the variance calculated from the combination of principal components. This is essential so that most of the information related to the structure dynamic behaviour under the effects of the loading changes is retained in the selected number of PCs.

In the case of current work, it is observed that when N= 15, the total variance meets the 97% of total variance. Obviously, a higher N (N>15) gives higher data variability. A comparison of N=15, 20, 25 and 30 principal components initially are used as the inputs for the neural network and the results are compared to identify which N gives the most accurate fitting of the target values. It is shown, the values of Ns give about the same performance. Thus, for a better computation of the network, N= 15 is chosen.

Target data takes the matrix in the form of 2-by-1000 matrix. The target for 1000 data observations has two dimensions in which the first dimension represents the load and the second dimension takes on the damage severity class. The best practise is to normalise both the inputs and the target vectors to ensure the network output always fall into a normalized range. With the previous PCA implementation, the data used for the neural network inputs has been standardised. The advantage of standardising the data set is discussed in Section 5.3.1. Using the feedforward function in the neural network toolbox, it is automatically normalized the inputs and targets data so that the outputs fall into the range between -1 and 1 [75].

6.8.3 The configuration of data for training, validation and testing set

Before training of the network begins, the general practise is to divide the data set into three subsets. The first subsets is the training data sets where most of the data is applied for training the network. The observations consist of 1000 data points that are divided into:

70% for training (700 samples) - this set is used only for network training of the network

15% for validation (150 samples) - this validation set generalizes the performance and halt the training when the error in validation set stops improving.

15% for testing (150 samples) - this is a complete independent test and no effect of the training or network generalization. This data set is used later for calculating the network performance.

It is important to have the majority of data used for training to ensure most of data pattern and characteristics are learnt by the basis functions of the network. In this case, 70% of total data set is used for training of the network [75].

The way the data is divided can be randomly or in predetermined way. The arrangement of the data set of the network input randomly selected from the overall data set, corresponding to the configuration set described above.

6.8.4 Specifying nonlinear activation function

The transformation of previous outputs and the weights into the next network layer should be performed via a continuous nonlinear activation function. The nonlinear function selected in this work, make use of tan sigmoid (hyperbolic tangent, tanh) for hidden layer(s) and a linear function for output layer. The choice of sigmoid non-linearity for the hidden nodes that are continuous and differentiable functions allow the parameter w the error function E (Figure 6.14) to be differentiated with respect to each weight in the network.

The choice of sigmoid nonlinearities for the hidden nodes together with the use of a squared error criterion at the output of the network provide the basis of the learning algorithm for the MLP [34], [57].

The choice of these activation functions depend on the nature of the data set and the distribution of the target variables [34]. The choice of tan sigmoid is suitable for most pattern recognition problems. In the relation to the current study, the radial base function

(RBF) is also applied as to investigate whether its use as the nonlinear transformation function can improve the generalisation performance. It was indicated that as in the context of current data, the radial base function (Gaussian function) performs relatively in similar performance as the hyper tangent (tansig) transfer function.

In this study, neural network is performed using Matlab neural network toolbox where a two layer-feed forward network with tan sigmoid transfer function is used in its hidden layer and a linear function for the output layer of the activation functions.

6.8.5 Specifying a function to train the network

The process of training the network involves the tuning values of the weights and biases of the network in order to optimize the performance of the network training. For training of the network specifically for function-fitting problem, the Levenberg-Marquardt (LM) back-propagation is used to calculate the gradient. Back propagation describes the training of a multi-layer perceptron using gradient descent that applies the sum-of-squares function. This training function uses the Jacobian for the calculation and applies mean squared errors to calculate its performance. The benefits and detailed explanation of this algorithm in terms of neural network application can be found in [76].

LM drives the fastest training in this feedforward network with better performance especially for fitting function (nonlinear regression) compared to other available training functions [77]. The training of the network using LM is determined by the gradient performance which it stops training when the gradient is less than 1e-5 or when the performance of the validation set fails to improve after 6 iterations (whichever comes first). This method is recognized as the early stopping used for neural network regularization.

6.8.6 Specifying the number of hidden layers and neurons

In most cases of using input-output fitting network, the usage of one hidden layer would be appropriate and with sufficient number of neurons in the hidden layer, the network can

generally arbitrary map the input data very well. Generally, more hidden layers require more computational resources and cause high complexities in the network. For best practise, it is a good step to start with one hidden layer and see if it gives the desire performance [34].

In addition to that, it has been proven that the two-layer network perceptron can fit high dimensional data arbitrary well [34][57]. Based on this premise, a prediction model of two-layer network is used with the goal to establish a relationship between the input data from PCA features and the corresponding load and damage as the network target vectors.

One practical approach in finding the optimum number of hidden nodes, M is to plot a MSE against M (Figure 6.22). Using the fixed neural network architecture and the training data set, the training using the 2-layer feed forward neural network is performed with its properties (activation functions, training algorithm) equivalent to those described earlier in this general practice steps. Here the range of M = 2 up to M= 30 are used in the training in which every M has 10 random starts (iterations) as illustrated in Figure 6.22.

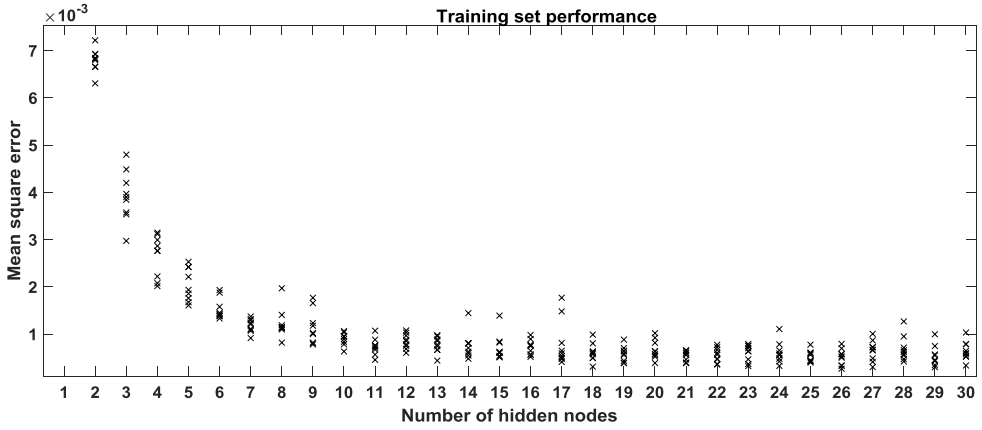


Figure 6.24: Performance of the training data set using 10 different initialization values for hidden nodes from 1 to 30.

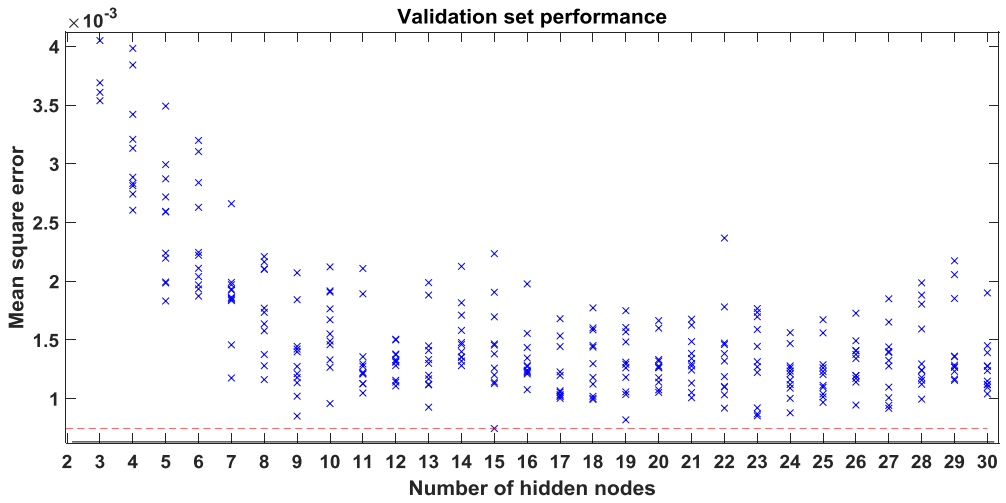


Figure 6.25: The minimum error of the validation set as indicated by the red line (M=15).

In order to determine the optimum number of hidden nodes M with respect to the best generalisation performance, one has to examine the mean square error of the validation set (Figure 6.23). In practice, M can be selected by choosing the node that has the smallest validation error. The validation set is required in respect to generalisation performance because the training stops after the error in the validation set begins to rise. A simple task which to verify the selection of M is to train the network several times using the same value of M to ensure that the repeated MSE values are in the close and acceptable range. It can be confirmed again that if M is the right choice by using the test set on the network. If the error increases or there is large difference in error between the test and validation set after using the M hidden nodes, it indicates an overfitting problem. M should provide a stable performance of the network with different initializations during the iteration. Based on Figure 6.23, the smallest error occurs at M=15. Initially, the network is tested several times using M=2 up to M= 30 and it shows after several iterations, M=23 gives better stability and good generalization in respect to validation and test set.

After it is assured that the selection of M gives the best generalisation performance based on the network using the parameters that gives some of the lowest MSE, the network is saved for testing. A similar test of the same network architecture is then performed using M=23 hidden nodes in which the performance results from the training, validation and test

sets are described in 6.24. The Bayesian framework is also compared with LM training algorithm.

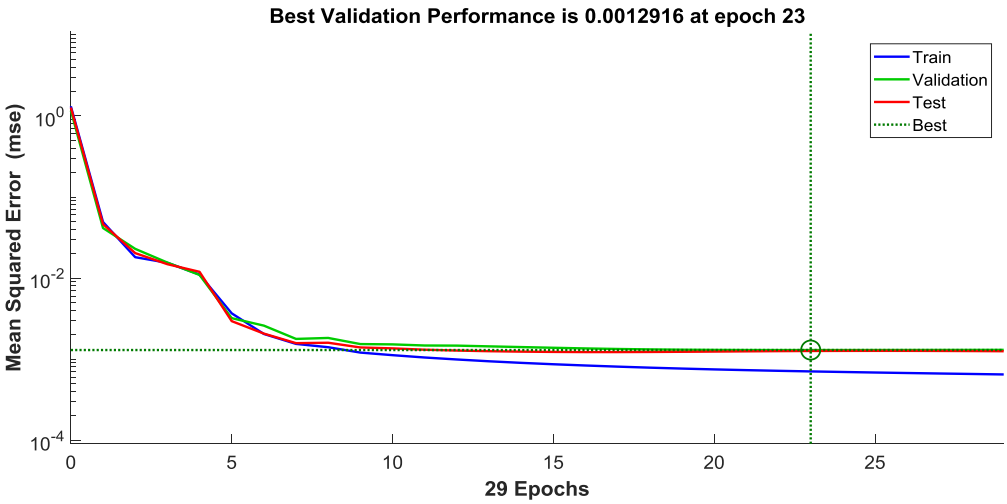


Figure 6.26: The performance analysis based on the training, validation and test set showing a vertical green dot line where the network best performance is attained.

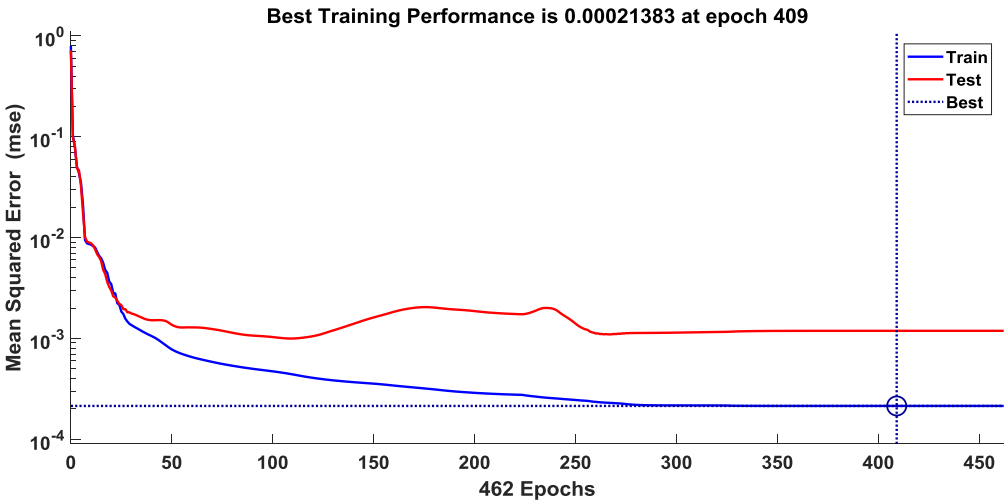


Figure 6.27: In comparison to a training using Bayesian framework. Notice of the much longer iterations used.

Figure 6.24 shows the performance of training that occurs at the iteration 23 using M hidden nodes. The validation and test set showing the similar characteristics in the mean square error which indicates that the test data is not overfit.

From this training of the network, the final result of the mean square error is small with the validation and test errors have similar characteristics. At the iteration 23, no significant of overfitting is detected. Finally, using a similar network architecture and parameters, prediction of the input variables to the target vectors linear regression between the networks outputs and the corresponding targets using the tool provided in training window in Matlab.

Using Bayesian approach based on the same network architecture shows improved prediction performance (Figure 6.25). It significantly takes longer iterations to achieve the convex solution however it does not require validation data set as it uses all data set. The mapping results of the input variables using the PCA features to the corresponding load and damage severities target vectors using the defined neural network parameters are presented in Section 6.10. Before that, general principles of maximum likelihood and Bayesian frameworks in the regression analysis perspective are discussed in the following section.

6.9 Comparisons of maximum likelihood and Bayesian neural network

So far, the exercise performed in this study involves the use of maximum likelihood estimation (MLE) approach to find the model parameters. In addition to that, Bayesian approach is described in ANN as a comparison to the training MLE algorithm. Note that, both training algorithms are used as to regularize the optimization in finding the best generalised network and to avoid from over fitting the data.

The main reason for applying of these methods of training algorithm is to estimate the network parameters which are the weights and biases. It is essential to understand the main difference between these two estimators before applying in the network. Here, it will briefly describe briefly the principles of these two methods used for parameter optimization in the scope of the current work.

Maximum likelihood estimation (MLE) approach is commonly found in many machine learning algorithms including as it is previously used for clustering technique of Gaussian mixture model (GMM) in the first part of this chapter. For a simple illustration to relate directly to the current work, the explanation will be centred on a regression analysis. Previously, maximum log likelihood function is used to fit data points to Gaussian models based on the means, covariances and the mixing proportions. MLE is applied again in neural network in the context of fitting the inputs variables on the corresponding target vectors with respect to obtain the minimum mean-squared-error MSE. Note that, minimum MSE is equivalent to maximising loglikehood function.

In terms of probabilistic interpretation, the neural outputs can be predicted using conditional distribution. For a regression problem, consider the target vector t that can take any real values, assume a Gaussian distribution with an x-dependent mean and becomes the predicted output of the neural network specified by

$$p(t|\mathrm{x},\mathrm{w}) = \mathcal{N}(t|\mathrm{y}(\mathrm{x},\mathrm{w}), \beta^{-1}) \tag{6.16}$$

Where β represents the precision or also known as inverse variance of the Gaussian noise. This is a conditional distribution that shows to be sufficient to take the output unit activation function to be the identity (linear function x = y) because such network can approximate any continuous function from x to y. Given a data set of N independent, identically distributed observations $\mathrm{X} = \{\mathrm{x}_1, \ldots, \mathrm{x}_N\}$ with the corresponding target values $t = \{t_1, \ldots, t_N\}$, the corresponding likelihood function can be established as

$$p(\mathrm{t}|\mathrm{X},\mathrm{w}) = \prod_{n=1}^{N} p(t_n|\mathrm{x}_n,\mathrm{w}), \beta) \tag{6.10}$$

The error function can be produced by taking the negative logarithm as

$$\frac{\beta}{2}\sum_{n=1}^{N}\{y(\mathrm{x}_n,\mathrm{w}) - t_n\}^2 - \frac{N}{2}\ln\beta + \frac{N}{2}\ln(2\pi) \tag{6.11}$$

where the parameters w and β are learnt using the error function. The selections of obtaining the w and β here are based on minimizing an error function rather than

maximizing the log likelihood as usually done in the neural network applications [34]. For analytical reasons and simplicity, the computation is done in the logarithm of the log likelihood than the likelihood itself. This is because the logarithm is monotonically increasing (strictly increasing) and maximizing the log likelihood is also maximizing the likelihood. Maximising the likelihood is minimizing the sum-of-squares function as described by

$$E(\mathrm{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(\mathrm{x}_n, \mathrm{w}) - t_n\}^2 \tag{6.12}$$

The value $E(\mathrm{w})$ is denoted by w_{ML} which corresponds to the maximum likelihood. Once the iterative for the optimization is done, the value of β can be determined using

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}\{y(\mathrm{x}_n, \mathrm{w}) - t_n\}^2 \tag{6.13}$$

For the case of multiple targets as it is usually be, the conditional distribution of the target values is described as

$$p(\mathrm{t}|\mathrm{x},\mathrm{w}) = \mathcal{N}(t|y(\mathrm{x},\mathrm{w}), \beta^{-1}\mathrm{I}) \tag{6.14}$$

The noise precision where K is the number of target variables, can be written as

$$\frac{1}{\beta_{ML}} = \frac{1}{NK}\sum_{n=1}^{N}\|y(\mathrm{x}_n, \mathrm{w}_{ML}) - t_n\|^2 \tag{6.15}$$

Next, Bayesian framework for optimisation of neural network parameters will be addressed. The main difference of MLE and Bayesian learning approach is that MLE views the true parameters that it seeks as fixed where as in Bayesian approach, it considers the parameters to be random variables and training of data set allows the distribution corresponded to the variable to be converted into a posterior probability density [78].

In multi-layer network, the log of the posterior distribution will be non-convex, corresponding to multiple local minima in the error function. As a result, no exact solution

can be found. To encounter this, two assumptions are made: the posterior approximated by a Gaussian and the covariance of this Gaussian is small. With these assumptions, the network function is approximately linear with respect to the parameters over the the space of the posterior probability which is non-zero. Having established the models, it provides an evidence frame work to provide point estimates for the hyper parameters and to explore models with different number of hidden nodes [34], [79], [80].

Consider predicting a single continuous target variable t from the inputs vector x and suppose that the conditional distribution P (t | x) is Gaussian with an x-dependant mean given by the output of a neural network y(x,w) and with precision β (inverse variance), written as

$$p(\mathrm{t}\,|\,\mathrm{x},\mathrm{w},\beta) = \mathcal{N}(t\,|\,y(\mathrm{x},\mathrm{w}),\beta^{-1}\mathrm{I}) \tag{6.16}$$

A prior distribution over the weights w is selected that is Gaussian in the form of

$$p(\mathrm{w}\,|\,\alpha) = \mathcal{N}(\mathrm{w}\,|\,0,\alpha^{-1}\mathrm{I}) \tag{6.17}$$

For an i.i.d (identical and independently distributed) data set of N observations $x_1, \ldots, x_N$ with a corresponding set of targets values $\mathbf{t}=\{t_1, \ldots, t_N\}$, the likelihood function is stated as

$$p(\mathbf{t}\,|\,\mathrm{w},\beta) = \prod_{n=1}^{N}\mathcal{N}(t_n\,|\,y(\mathrm{x}_n,\mathrm{w}),\beta^{-1}) \tag{6.18}$$

then the resulting posterior distribution is written as

$$p(\mathrm{w}\,|\,\mathbf{t},\alpha,\beta) \propto p(\mathrm{w}\,|\,\alpha)\,p(\mathbf{t}\,|\,\mathrm{w},\beta) \tag{6.19}$$

As a result, the nonlinear dependence of y(x,w) on w will be non-Gaussian. A Gaussian approximation is found to to the posterior distribution using the Laplace approximation. A local maximum is searched first of the posterior using iterative optimization. Maximizing the logarithm of the posterior which can be written as

$$\ln p(\mathrm{w}|\mathbf{t}) = -\frac{\alpha}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} - \frac{\beta}{2}\sum_{n=1}^{N}\{y(\mathrm{x}_n,\mathrm{w}) - t_n\}^2 + \mathrm{const} \tag{6.20}$$

The above equation corresponds to a regularized sum-of-squares error function. Assume for a moment that α and β are fixed, a maximum posterior can be found using a standard nonlinear optimization algorithms such as conjugate gradients and error backpropagation to calculate the derivatives. For details of this Bayesian method, readers are encouraged to refer to [34], [74] for further explanations on this method. Results of the neural network

After describing the general procedures and framework with respect to neural network implementation, now the network will be used to test the data sets from the standard PCA. The results from the standard PCA of a combined data set (PC model D) are used as inputs with a selected number of principal components.

Optimal number of nodes (neurons) in the hidden layer, total number of principal components as the inputs, the activation functions and the training algorithm all been selected so that the mapping performance of the neural network with the lower mean square error is achieved. It is essential to recognise that, there are other regularization methods applicable in neural network that can be used for parameters optimization to enhance the network performance[34], [74].

6.9.1 Developing a benchmark for damage level and loading

One of the novelty found in this study, is to introduce a reference target set for operational loading system that possibly has some damage on. Using features obtained from the standard PCA, the mapping onto a target benchmark divides different damage severities into different level built within each loading condition as shown in Figure 6.27. This benchmark consists of constant vectors depending on the number of loading class and damage class considered in the study. These constant vectors represent each damage class and load class respectively. The 1000-by-2 matrix of the target network with the first

column represents the load vector and the second column of the damage vector class respectively.

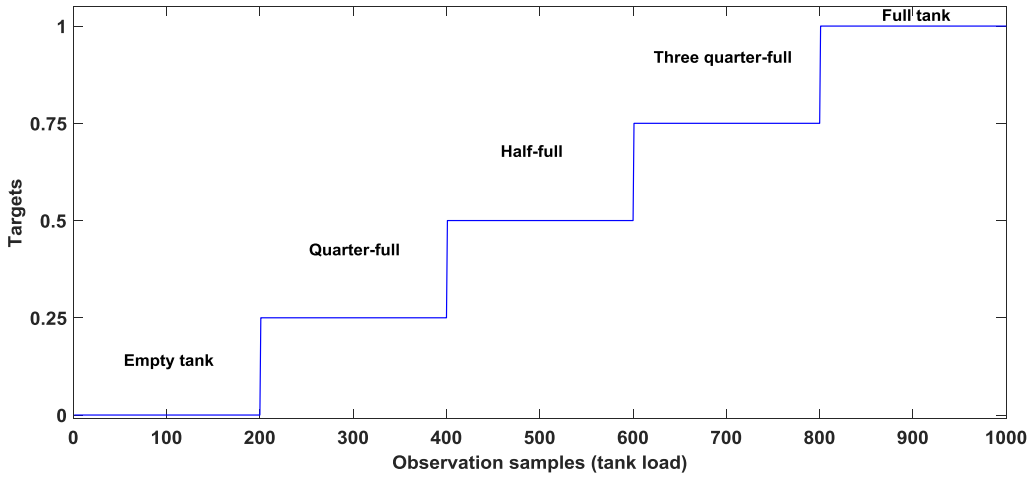


Figure 6.28: The target of loading vectors assume undamaged data set from the operational loading class to produce the output.

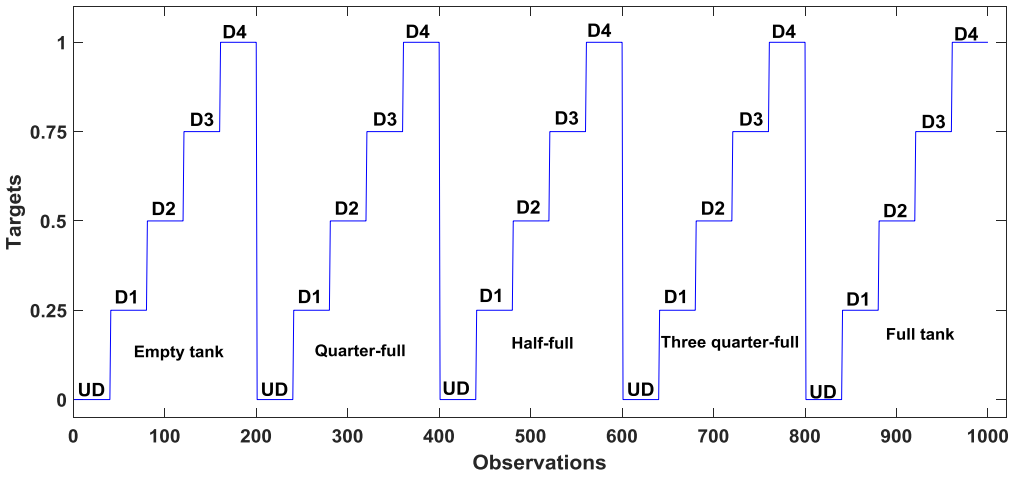


Figure 6.29: The Operational-load-effects-SHM model of its target vectors are constants values divided by the number of classes of loading and damage conditions.

Figure 6.26 displays the total number of loading conditions with each loading class includes its total number of observations. The damage severities shown in Figure 6.27

describes how the damage level is divided according to its severities. This benchmark represents target values in order of damage severities and loading class. In fact, this benchmark encompasses part of the novelty in the study where it aims to establish a relationship between damage severities and loading class through this regression analysis. A matrix comprises 1000-by-2 matrix is established that marks 1000 observation samples in each rows. In the first column, loading class L_i given a value {0, 0.25, 0.5, 0.75, 1} represents empty tank, quarter-full load, half-full load, three-quarter-full load and full load respectively. In the second column of the target matrix, damage severities D_i is incorporated to be as {0, 0.25, 0.5, 0.75, 1} placed within each loading class which takes the value of L_i. Similar to previous chapters, D_i represents Undamaged, D1, D2, D3 and D4.

It is in the interest of the study to investigate if neural network is capable in mapping the input variables obtained from the standard PCA (of PC model D) which the matrix of each loading and damage class are join into one matrix before eigen decomposition is performed on the single matrix. In the context of a supervised learning problem, the inputs data have to be arranged in correlation to the target values so that the input vectors match the order of damage severities in each loading class.

6.9.2 Fitting of the input-target variables using the network

The plots presented here are produced using the neural network with the parameters as described in previously in this chapter. In this section, the mapping of input PCA features to the target benchmark using ANN is presented. The performance results of the defined neural network model include a comparison between LM training function (maximum likelihood framework) and the Bayesian framework training algorithm. The results of testing the neural network model shown here are based on the combination of the training, validation and test sets so that there will be complete data observations in the analysis.

It is obvious that, the network predicts the data pattern corresponding to undamaged condition of all loading conditions very well (Figure 6.28 and Figure 6.29). In this scope,

it is observed that Bayesian Regularization (BR) outperforms the Levenberg-Marquadt (LM) training network.

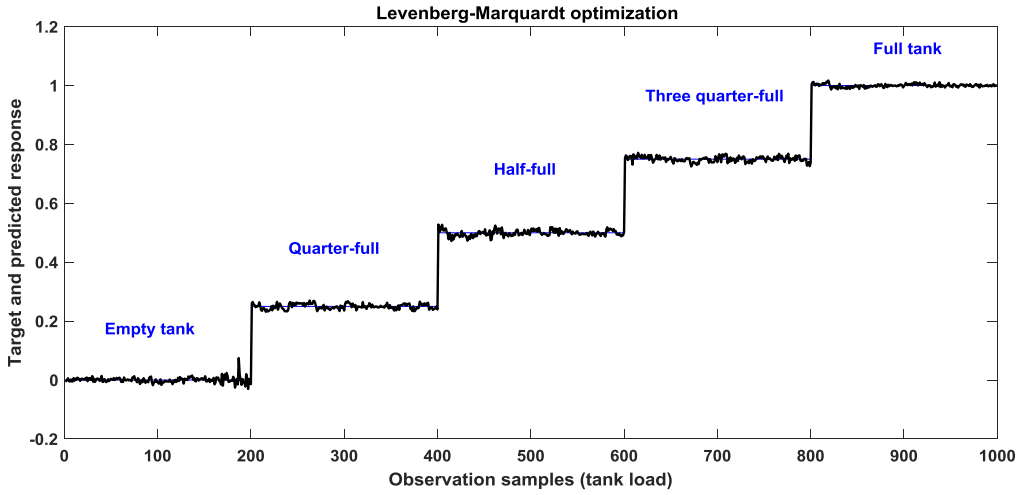


Figure 6.30: Training the data set with LM algorithm and test the network with overall undamaged data set.

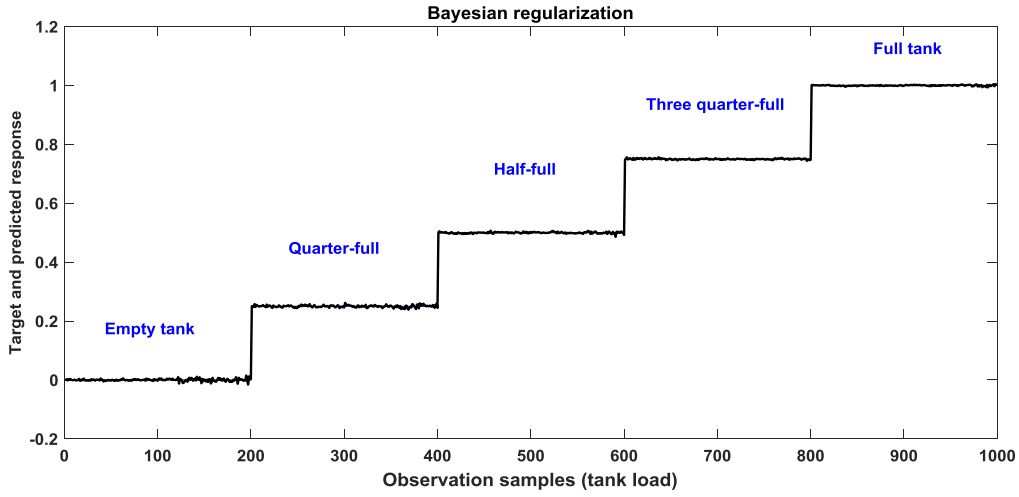
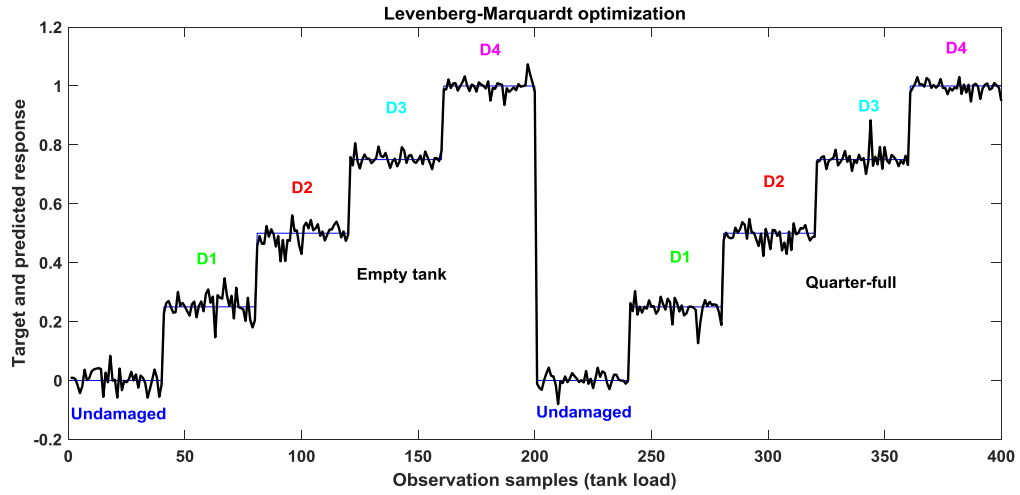


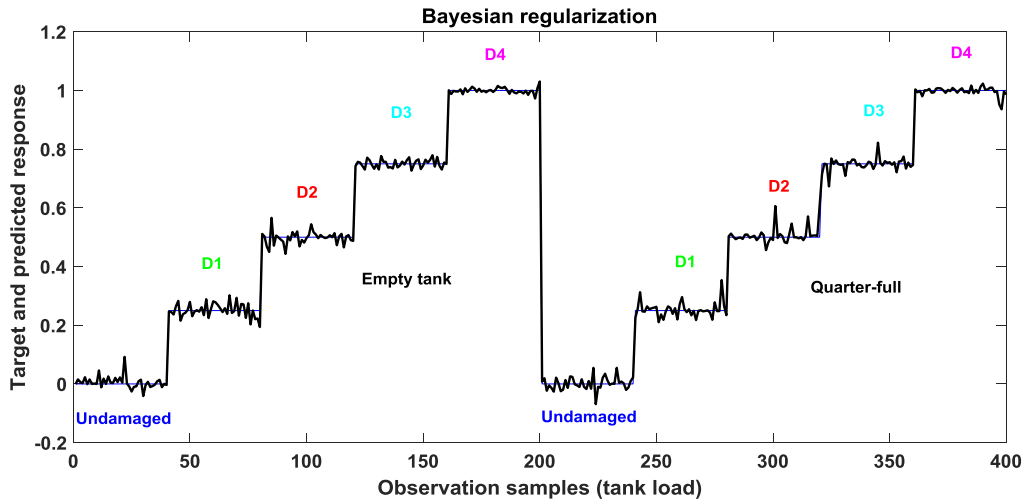
Figure 6.31: By comparison with the above plot, Bayesian regularization approach is described.

The prediction results raised from various loading conditions show excellent separation for the undamaged condition. This is expected as the data inputs initially obtained from

the standard PCA result have previously shown good separation for all loading class as indicated in the PCA visualisation result (Figure 6.36). The higher interest will now focus on the identification of damage severities especially for the smaller damage severities groups that previously shown to be overlap (Figure 6.36).



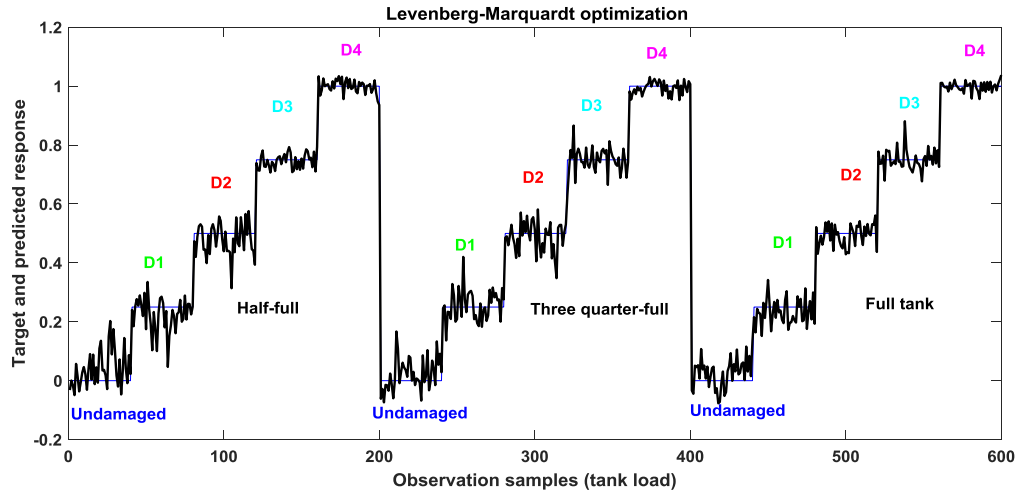
(a)



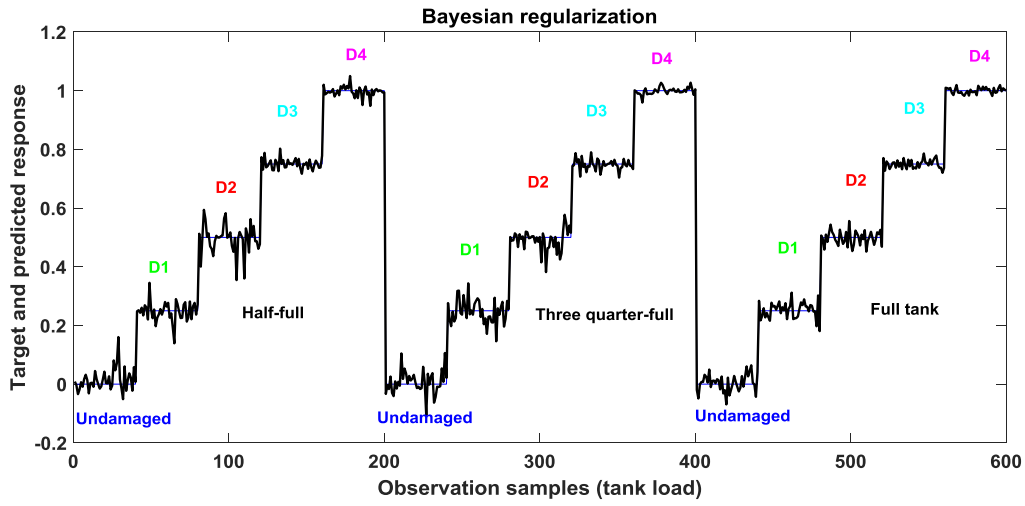
(b)

Figure 6.32: The plots show the all damage severities within each loading class Empty tank and Quarter-full tank using (a) LM training function and (b) Bayesian Regularisation approach.

The mapping of the PCA data to the correct target benchmark (shown in Figure 6.30) corresponding to various damage severities and loading conditions is excellent. BR has shown to perform better than LM using the overall 1000 data observation. As expected, the data associated with the highest damage D4 display the smallest error in mapping the data input to the target vectors.



(a)



(b)

Figure 6.33: The continuous prediction results from the previous plot showing half-full, three-quarter full and full tank in comparison of training functions to LM and Bayesian approach.

It is interesting to note that, all damage severities are capable to differentiate themselves and mapped into respective damage targets distinctly. It is also identified that BR outperforms LM approach in predicting the mapping input features from the PCA to the target benchmark as illustrated in Figure 6.31.

As overall, the trained ANN architype successfully perform the mapping given the PCA features as the network input and the novelty benchmark as their target.

Method	MSE
LM	8.75E-04
BR	3.60E-04
LM/BR	2.4297

Table 6.6.1: The results in mean of sum-squared-error between LM and Bayesian regularization (BR)

In this test for the trained network, the overall ANN results reveal that using BR, the MSE is 2.4 times smaller compared to that MSE produced by LM training function (Table 6.1). This justify that the MSE can be reduced if using BR approach. However, it is worth to mention that, the duration of training for the network takes much longer compared to the LM approach. Another positive point to consider is that, by using BR the needs for optimum network size and the concern of overfitting the data are not exist. The non-evident framework also generalises better in a smaller data set as it requires no validation set; all data set is used in the training [77].

The logical argument is whether a neural network is reliable given a new test data set that it has not 'seen' before in the training. To implement this by using LM evident based approach, 50% of the 1000 data observations is used to create the network and another 50% is used to fully test the network. In this way, the trained network is tested with data set which is not previously in the training set. The result has shown to be comparable those from the 70% training set + 15% testing set + 15% validation sets which are combined to test the network as shown in the next figures of Figure 6.32. It is indicated that the MSE has increased slightly when the test data is new data set and the number of

training data is reduced. However, the network has shown to be able to map the PCA features into the corresponding target vectors reasonably well (Figure 6.32).

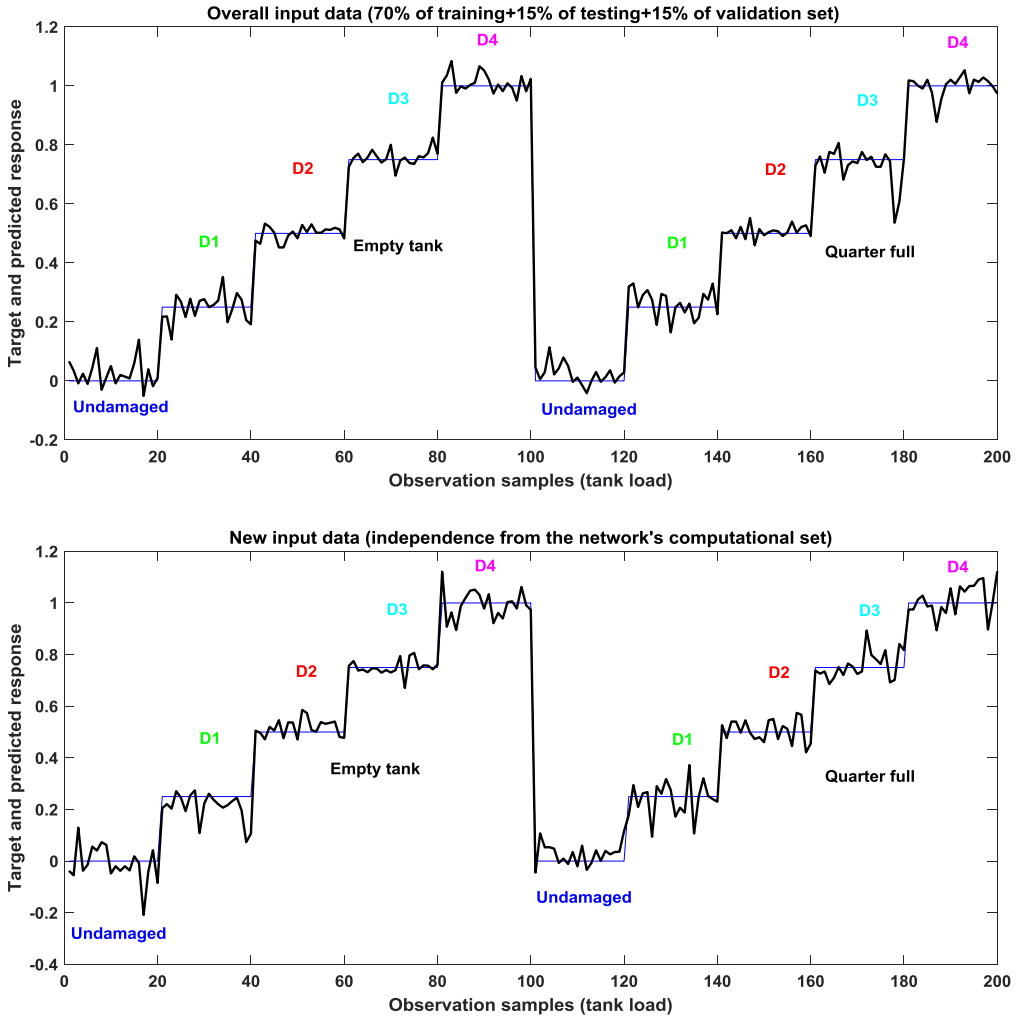


Figure 6.34: Using 50% of the data set for training, validation and testing of the network. Another 50% of the unused data set are reserved for testing the network.

The overall ANN results are promising when network is linked to the PCA results. The integration between the linear PCA (as the network inputs) and ANN (to produce nonlinear mapping) enables excellent damage identification under various loading conditions. It is worth highlighting here that, referring to Figure 5.36 in Chapter 5 where similar data sets

are used in the neural network. It reflects how mapping of the input patterns from the standard PCA to a defined target matrix through a two-layer (single hidden layer) feed-forward neural network can solve the problem related data interpretation and identification of damage severities due high overlapping of data classes.

6.10 Conclusion

The findings first highlight the applications of using GMM before it addresses a nonlinear regression analysis based on neural network for separating and identifying damage severities under operational loading conditions. Using mixtures of Gaussian distribution, the study explores how the data distribution raised from different loading conditions can be identified based on the data density characteristics. The results show how damage clustering can be established via GMM function based on the mean, covariances and mixing coefficients of the components. Based on maximum likelihood framework, EM algorithm is used to estimate the parameter means and covariances for each components and then optimised the weights for each data set.

In the section related to GMM, the study has demonstrated how different initializations selected for the mean and covariance either by using baseline set or by k-means algorithm can produce different outcomes in clustering the data groups corresponding to different damage severity. GMM can also potentially used to generate a probabilistic predictive model in monitoring of damage severity for a structure under the operational loadings changes. It can predict relatively accurate data points to be assigned to the most likely distribution model belonged to different damage severity when using data sets obtained from the kernel PCA.

The establishment of the ellipsoids during clustering the data feature can help a user to understand the behaviour of the data set and it potentially can become a generative model for SHM. This implied that if the undamaged data set is not available, damage monitoring can also be done by generating the data point equivalent to the undamaged model. If a new

data set is mapped into a cluster that far from the undamaged condition cluster, it should alert the user about the high chance of the presence of damage.

Using MLP neural network via nonlinear regression analysis, it has been illustrated how damage severities under the effects of operational loading conditions can be effectively mapped into their corresponding target values. The target benchmark consists of constant vectors that represent the number of loading class and the number of damage severities class provides a distinctive damage identification model for varying loading system . Utilising feed forward network functions and backpropogation error in evaluating the gradient of the mean-square error function, the network effectively mapped the input from the linear PCA data into their respective target vectors.

The study also highlights a comparison of training algorithm LM with Bayesian framework that shows trained network using Bayesian approach performs twice better if using LM training function. However Bayesian approach takes much longer time for its iteration and it requires more computational resources to reach a minimum solution for the MSE. In this case, it is suitable only for smaller number of observations of around 1000 data samples. For instance, when using with the Jabiru wing data samples that consist of more than 2000 data samples, the iteration takes significantly longer time and produce not far better result than using LM algorithm.

Using the simplest network consisting one hidden layer and an output layer with the logistic sigmoid function (‘tanh’) for the hidden layer and a linear identity function for the output layer, the optimised network is successfully map the data input from the PCA data onto the target vectors effectively representing the number of loading class and the damage severity class. A conclusive remark about the ANN, using the data inputs produced form the standard PCA, the prediction of the PCA features mapped into the target vectors corresponding to the number of damage and loading groups is encouraging and has a considerable place of opportunity in SHM for damage identification for structure under varying load.

Chapter 7

A DAMAGE DETECTION IN A FULL-SCALE AIRCRAFT WING

7.1 Overview

In the previous chapter, a dimensional reduction technique via standard PCA is implemented in the aim of reducing the high number of spectral dimensions and subsequently extract the underlying feature. Kernel PCA is also used as a nonlinear extension of the standard PCA to extract some hidden features related to nonlinearities and to improve separation between the data class.

This chapter extends the work of the previous two chapters of Chapter 5 and Chapter 6 by applying the previous machine learning algorithms on data set produced by a full-scale aircraft wing VBDD test as opposed to the wing box structure. The majority of the work in this chapter mirrors the previous work in that it investigates the effects of variable loading on damage detection.

In the previous chapters, the essential and well-known machine learning algorithms such as the kernel PCA, GMM and ANN regression analysis were implemented on the wing box data set, which may be considered as a laboratory structure. One may question the simplicity of the structure and it may be the case that it does not possess the same level of complexity as a real operating structure.

In this chapter, the aim remains the same as the previous chapters but now, it moves a step closer to real application by performing on a full-scale aircraft wing. This will be achieved through an experimental test of the wing in which its fuel tank undergoing incremental loading changes. The full-scale wing was obtained from a Jabiru aircraft, manufactured by Jabiru Ltd from Australia. The general view of the wing and the aircraft is shown in Figure 7.1. Due to some safety issues, the aircraft was grounded and the wing was detached from the main fuselage and purchased by the Dynamics Research Group for research at the University of Sheffield. The novelty of this work lies in the attempt to perform damage detection on a real aircraft wing in the presence of varying fuel tank loading in the interest of SHM.

Having produced some encouraging results from the kernel PCA on the wing box data, the work now shifts into a full-scale wing in an effort to establish data trajectory for monitoring the structural conditions. It was shown previously in Chapter 5 that kernel PCA, under the effects of loadings, could provide a way of monitoring structural health by tracking the data set's trajectory in the feature space. Now, nonlinear regression analysis using ANN is to be applied on the Jabiru wing (JW) data set after the promising result obtained for the wing-box structure. In the last section of this chapter, a decision tree classification model using Random Forest algorithm is presented as an alternative to the earlier GMM technique.

In the early part of this chapter, FRF characteristics of the JW of an undamaged condition will be examined under varying fuel tank loading. It will be analysed again when some damage severities classes are introduced on the wing part under the similar effects of varying fuel tank load.

Figure 7.1: The exact model of which the aircraft wing used in the experimental study. The Jabiru J-170.

7.2 Description of the Jabiru aircraft wing

The Jabiru aircraft is an Australian ultralight and light-sport aircraft consists of 2-seater capacity with the maximum gross weight of 600 kg (Figure 7.1). The wing is a semi-cantilever, stressed skin type with a main spar. It is a moulded structure with a series of ribs, bonded to the fibreglass skin, fuels tanks and to the spar through the moulding process. The wing is a strut-braced high-wing of which the strut supports each wing and connects to the fuselage. Each wing weights about 27 kg without fuel load.

- Jabiru aircraft specification

Model: J170-UL
Aircraft maximum weight: 600 kg
Wing span (from one wingtip to the other wingtip): 9.66m
Fuel capacity: 135 litres
Wing area: 9.56 m^2
Wing loading: 62.8 kg/m^2

Figure 7.2: The overall wing structure fitted to the metal rig at the wing root

Figure 7.3: Overall experimental set-up with its data acquisition system for the Jabiru wing.

The starboard wing was bolted to a steel support at the wing root and has a free support provided by a modified aluminium strut inserted through the wing bracket located below the wing of which the strut and the bracket could be visualised clearer in Figure 7.6. The

strut was a substituted part as the original strut was removed when the wing was detached from the aircraft. The newly replaced strut was suspended from a special-built aluminium frame using nylon lines that hold both ends of the rod to the high durable springs attached to the frame (Figure 7.3).

7.3 Experimental overview

To implement the VBDD test, a random excitation signal was applied on the structure using an electrodynamic shaker. The sampling frequency used in this test was 4096 Hz with frequency resolution of 0.25 Hz. In the results, the calculated frequency was 2048 Hz which was the Nyquist frequency. From this point, the data acquisition will be referred to in terms of spectral lines. The 0-2048 Hz frequency range with 0.25 Hz frequency resolution will result in 8192 spectral lines. A higher frequency range of 0-4096 Hz was used so that it be would potentially manage to detect a localised damage at higher frequencies.

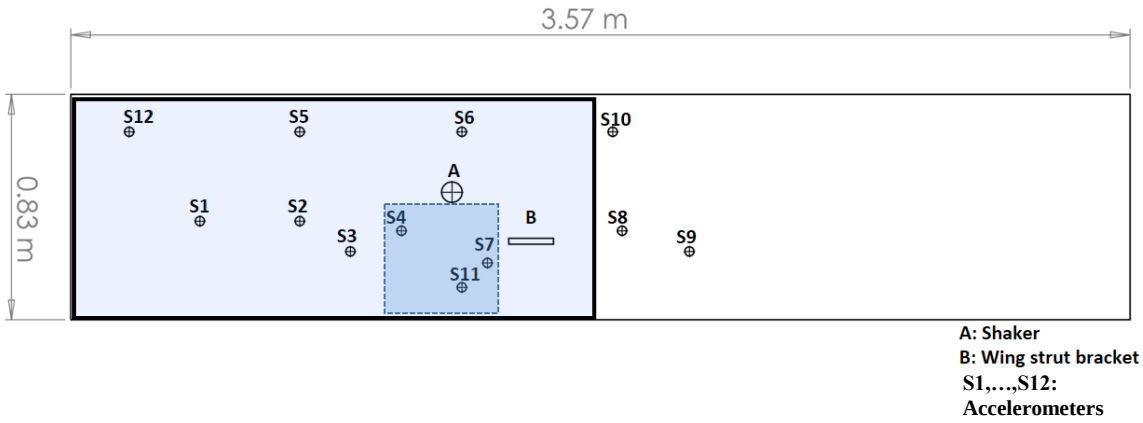


Figure 7.4: The overall placement of accelerometers fitted below the wing.

The vibration response was acquired and processed DIFA SCADAS III of 16-channel and high speed data acquisition system, controlled by the LMS software running on a Dell desktop PC. The wing was excited using Ling electrodynamic shaker, powered by an amplifier used to generate a random excitation. The shaker unit was attached below the fuel

tank of the wing (Figure 7.4). Approximately half of the wing area marked with light blue shown in Figure 7.4 indicates the approximated area of the fuel tank inside the wing. The base measurements used in the test were the FRFs acquired using a force sensor. Similar window processing used for the wing box was applied again on the Jabiru wing. The signal was applied using the most common one, the Hanning window so that better frequency definitions (resolutions) could be generated. The measurements were performed with 8-averages to produce an appropriate smoothing effect on the frequency spectrum [61] [1].

12 PCB piezoelectric type accelerometers were used to measure the response and they were all mounted below the wing using a cyanoacrylate (Loctite brand) of a super glue type adhesive. This was performed in line with the recommendation in the [81] as the amount (thickness) can contribute in achieving good frequency response. A relatively large number of accelerometers were considered because it was attempt to include a vast area of the wing to determine the best signal response. The signals obtained were evaluated and those with good frequency definition and sensitive to the properties changes in the structure were selected. The prior criteria considered when positioning the accelerometers was that they should not be mounted directly onto the wing's ribs to avoid the low amplitude of the natural frequencies due to their stiffness. Initial observation on FRF signals produced by sensors mounted on this location generally result in smaller FRF amplitude.

The natural frequencies captured from all the accelerometers were examined and found that those accelerometers in the region indicated by the dashed lines as shown in Figure 8.4 were relatively high in natural frequencies. The enclosed area as indicated by dashed lines was in the vicinity of the shaker and the wing bracket for the strut (Figure 8.4). It can be observed that, the response were judged by the definition of the natural frequencies, meaning that how distinct the peaks were with minimum noise. At this point, the natural frequencies becomes the selection criteria on the basis of the underlying assumption that the damage and loading will alter the system's dynamic characteristic that is governed by the stiffness, mass and dissipation energy of the system [1]. The response data will be discussed in more details in Section 7.4

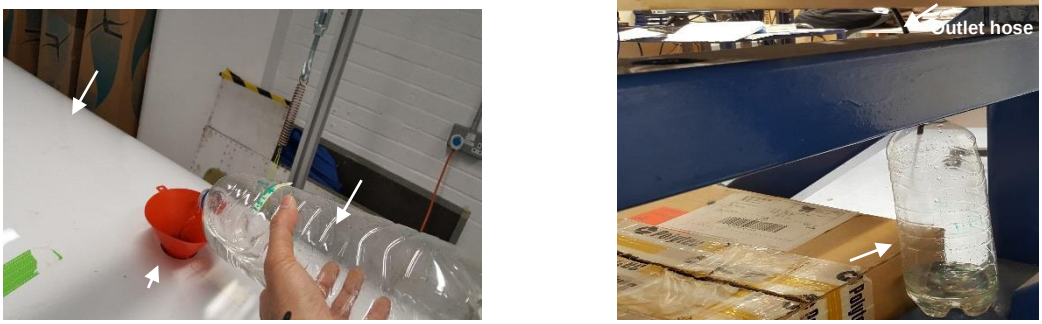


Figure 7.5: (left) Showing the process of water loading into the fuel tank and extracting the water from the fuel tank (right)

7.3.1 Loading definition

The loading of water into the wing's fuel tank strictly followed a systematic program that is sequential. Beginning from the empty tank, the loading was done in incremental and sequential manner until reaching specific loading states at 1.5L, 3L with an increment of 1.5L then loaded with a smaller increment of 0.3L to achieve 3.3L, 3.6L, 3.9L, 4.2L, 4.5L consecutively and finally achieving 6L with higher 1.5L of water increment as illustrated in Figure 7.5 (left). Upon reaching each loading state, the vibration test was conducted and the vibration measurements in the FRF domain was recorded.

After reaching 6L fuel tank load capacity limit, the load was then extracted from the fuel tank reduced to 4.5L, 3.9L, 3.6L, 3.3L, 3L, 1.5L and empty tank in sequential and incremental order as in the previous loading process as illustrated in Figure 7.5 (right). This process was conducted mechanically using an electrical pump that extracted the water via a tube connected to the fuel tank outlet at the wing root. A valve was used to stop the water flow once it reached the determined amount. The process of loading and unloading of fuel tank load was performed in the specified order in two cycles with one cycle comprising loading and unloading the fuel tank and repeated for the next cycle with the same configuration. The motivation of performing the loading and offloading in such way is to simulate a more realistic change of operational fuel loading under small and large

variations. Previously in wing-box experiment, the loading of water was into two special containers attached to the top of the structure, currently in Jabiru wing experiment, the loading was introduced directly into the fuel tank in smaller loading variation. The purpose is to investigate the capability of the previously applied machine learning algorithms with the more realistic scenario.

7.3.2 Damage types definition

In the experimental test, a real saw-cut damage is initiated into the wing bracket, which forms as part of the wing internal ribs. The bracket holds the strut onto the wing. The strut acts as as secondary support by holding the wing at its the middle length and attach it to the ground part of the fuselage

The saw-cut damage was introduced into the wing bracket to which the wing support strut was attached in progressive manner (as shown in Figure 7.6). Due to a removal of the strut from the wing, as to provide extra support to the wing, a replaced strut was fitted into a newly made hole on the bracket (Figure 7.6).

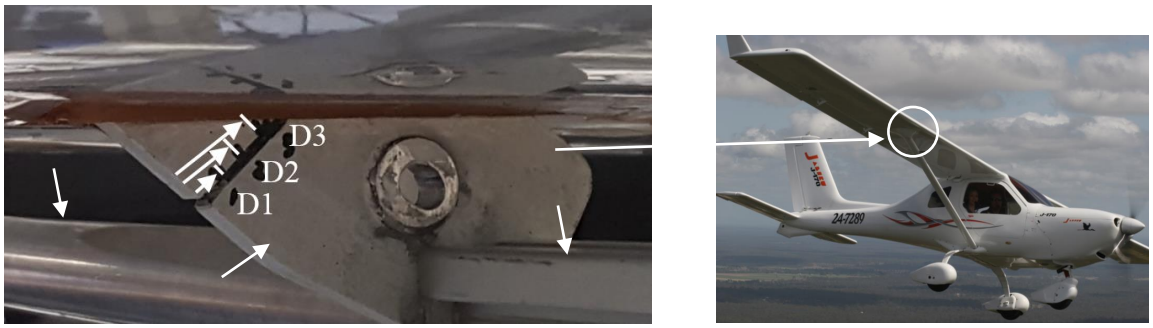


Figure 7.6: The saw-cut damage consists of three different lengths indicated by length of arrows introduced on the strut bracket underneath the wing.

Three saw-cut damages introduced into the bracket corresponded to the cut measured approximately 8 mm, 12 mm and 16 mm represented by D1, D2 and D3 respectively as shown in Figure 7.6. Real damage was decided on the wing rather than artificial damage as

to introduce the real damage effects into the structure and avoid the effects of fixing conditions if to use other than real damage [38].

7.3.3 Data acquisition

In this experiment, a total of 1120 observation samples were measured in terms of FRF domain, inclusive of the real and imaginary parts for the undamaged condition set and grouped as one matrix and labelled as a baseline set (reference set). After completing the data acquisitions for the undamaged condition and the signals were examined if they had provided desired response in terms of high frequency amplitude, frequency definition and sharp transition due to small loading variations.

Since the experiments were not repeatable after damage was introduced, any issues regarding to accelerometers placements and mountings, adjustments of the shaker and other issues pertaining to experimental configurations were solved before damage was introduced. The samples associated with each damaged conditions were recorded after introducing damage and completing the loading procedure in similar procedure as the undamaged condition. Upon completing the FRF measurements for the all three damaged conditions, they were compared with the baseline set under similar loading conditions. In total, 1140 samples were recorded inclusive of three sets of damage conditions in which each set comprised 380 samples and associated with D1, D2 and D3 structural condition respectively. The data was grouped corresponding to the structural conditions of the wing and its fuel tank loadings as described in the following Table 7.7:

Wing condition	Tank load	Wing condition	Tank load
N	E x 160 samples	D1	Ex 60 samples
N	1.5Lx 120 samples	D1	1.5Lx 40 samples
N	3Lx 120 samples	D1	3Lx 40 samples
N	3.3Lx 120 samples	D1	3.3Lx 40 samples
N	3.6Lx 120 samples	D1	3.6Lx 40 samples
N	3.9Lx 120 samples	D1	3.9Lx 40 samples
N	4.2Lx 120 samples	D1	4.2Lx 40 samples
N	4.5Lx 120 samples	D1	4.5Lx 40 samples
N	6Lx 120 samples	D1	6Lx 40 samples
Total	1120 samples		380 samples
Wing condition	Tank load	Wing condition	Tank load
D2	Ex 60 samples	D3	Ex 60 samples
D2	1.5Lx 40 samples	D3	1.5Lx 40 samples
D2	3Lx 40 samples	D3	3Lx 40 samples
D2	3.3Lx 40 samples	D3	3.3Lx 40 samples
D2	3.6Lx 40 samples	D3	3.6Lx 40 samples
D2	3.9Lx 40 samples	D3	3.9Lx 40 samples
D2	4.2Lx 40 samples	D3	4.2Lx 40 samples
D2	4.5Lx 40 samples	D3	4.5Lx 40 samples
D2	6Lx 40 samples	D3	6Lx 40 samples
Total	380 samples		380 samples

Table 7.7.1: Data samples descriptions acquired and recorded in the test.

7.4 Dynamic Response Data

The dynamic response of the aircraft wing under variable loadings will be discussed before compared with the structure when implicated with damage. The vibration response of the Jabiru wing in FRF was analysed with the main attention on the frequency peaks that represent the resonance frequencies of the Jabiru wing. In total, 12 accelerometers were used to measure the vibration response of the wing which were distributed widely across the wing. All signal from all the accelerometers were examined and the response corresponded to sharp transitions and high frequency definition were selected. It is revealed that the accelerometers located under the fuel tank along the trailing edge result in better frequency definitions. The accelerometers numbered 4, 7 and 11 (shown in Figure 8.4) were concluded for further analysis in data processing and machine learning application.

The key finding from the FRF analysis is that it reveals of the global and local modes of natural frequencies throughout the frequency range of the wing. The global modes are associated with the vibration of the whole structure. The global natural frequencies modes can be identified as frequencies that are clearly observed at all the sensors. Their presence was detected in the lower region mode. On the other hand, local modes relates to vibration of individual of the individual panels of aircraft wing structure and can be found at much higher frequency range, as in this structure, the local modes are detected above 500Hz. Figure 7.7 shows the global frequencies mode that are distinctively detected in the lower region with better frequency definition compared to those local frequencies mode in the higher region.

Some previous studies described the advantage of using local modes when detecting local damage in the structure [1], [82]. In fact, it is revealed that damage is typically a local phenomenon and it is usually detected in the higher frequency range of the vibration. In the interest of detecting damage, the local modes located in the higher frequency range will be considered for feature selection.

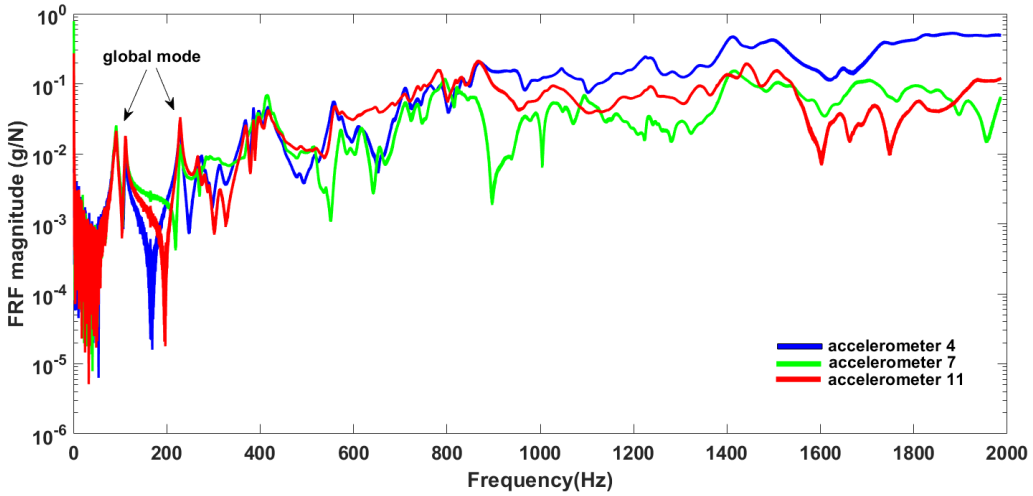


Figure 7.7: FRF plots measured from an undamaged JW using the selected accelerometers in the presence of global frequency modes.

The vibration test also reveals some interesting findings of the anti-resonance, where the amplitude at certain frequency drops to a minimum value. There are some studies related to using anti-resonance to provide useful dynamic properties of a test structure [83]. The phenomenon of anti-resonance is caused by destructive interference that can happened due to an interaction between the external driving force and with another oscillator [84]. In the scope of the current study, the focus is on a feature selection process and damage detection under the effects of operational loading variabilities by the use of machine learning algorithms.

7.5 Feature selection

An important step prior to the machine learning stage is to prepare the data set by cleansing and organizing them so that a more convenient machine learning implementation can be achieved. During cleansing process, key attention was given whether to accept or reject the data signal from the selected feature. For instance, if a noisy data signal measured from an undamaged condition or a data influenced by change of fixing conditions in the experiment is included in the feature selection, it can possibly be identified as an outlier and results in false positive damage detection. It is highlighted in many SHM literatures that performance of a machine learning is highly influenced by the quality of the selected feature [1], [6], [85].

The initial steps should be to plot the FRF plot acquired form accelerometer 11 as the most appropriate data signal in respect to frequency high definition and good transitions between loading class. This includes only main loading class (E, 1.5L, 3L, 4.5L and 6L) before examining all the peaks. Figure 7.8 displays a frequency range between 500Hz and 900 Hz that shows good sensitivity to the variation of water loading. The peaks are in better frequency definition and have larger amplitudes at the frequency range. Note that, at this point, only data set from undamaged conditions are considered. This approach is relatively practical as loading effects shown to be more distinguished on the shifting of the FRF peaks than the damage effects. The global frequency modes located in the lower

frequency range of 100 Hz to 300 Hz aslo show to be loading-sensitive but later found to show little sensitivity to damage.

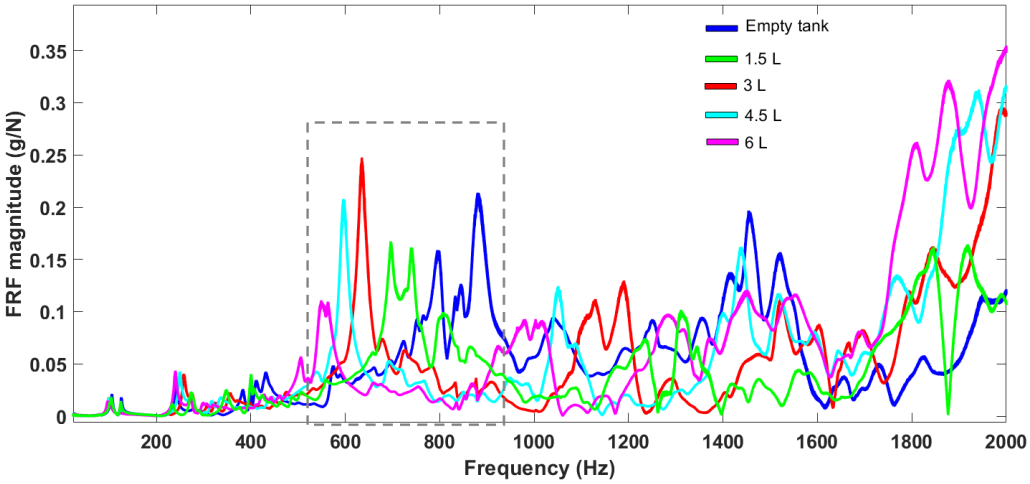


Figure 7.8: Overall FRF plot comprises 20 observations for each loading conditions spans from 20 Hz to 2000 Hz (the noise at the start and end the of FRF has been eliminated).

When the feature from undamaged condition is compared with the smallest damage (D1 data set) under the same loading variables, some indications of the peaks shifting due to damage can be detected (Figure 7.9). It is obvious that the shifting due to damage is significantly smaller compared to the shifting caused by loading variations. This characteristic is expected as the loading variations introduced are significantly larger compared to the damage and this is a typical phenomenon in SHM.

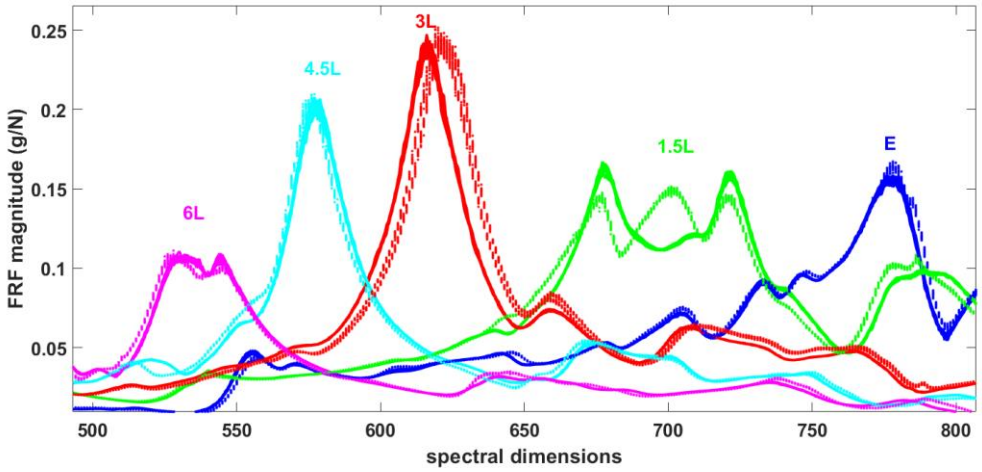


Figure 7.9: The selected feature: Comparison of undamaged (solid lines) and D1 saw-cut (dash lines) respective to the main loading groups.

At this stage of data pre-processing, the data signal is labelled and organised according to their class. Here, a feature is selected based on the sensitivity to load changes and damage. In the previous wing box experiment, it was indicated that some features located in the higher frequency range shown to be sensitive to load variability would also showed some sensitivities to damage.

So far, a potential feature is considered using a simple method by plotting the original data signal on one plot and comparing the feature from undamaged condition with the damaged state. In reality, however, this will not be a typical case. The main challenge is that when the labels of the data class are not always available. One way to overcome this problem, is to implement PCA that is well-known dimensional reduction technique for unsupervised data problem. Here it is utilised to transform the original data to combinations of data variables so that the effects of variability from the loading and damage can be better visualised and identified. At this stage, it is essential to examine the result of the PCA transformation to ensure all data points collapse into the right class.

Here, PCA was also utilized to verify if a data point associated with a class falls into its corresponding class. If a small number of observations in which the data point collapse

into another data class, it has to be discriminated from the observations before introducing test data. If the outliers are remained in the data set, it can result in false interpretation of the loading class as illustrated in Figure 7.10. The cause of these outliers can be due to many reasons as for this case; it is mainly because of the nature of the large experimental work such as the variability from the instrumentation (adjusting the wing stabilizer and the height of the the stringer) and water loading variances during the filling and extracting of water process.

It was found out that, there was a possibility, there few of the data points corresponding to the smaller inter-class variance loading class (3L, 3.3L, 3.6L, 3.9L, 4.2L and 4.5L) falsely projected into its next neighbouring loading class. Inter-class variance here infers to the variance of different data class and not the inner variance of within the class itself. This behaviour was not observed in the loading case of higher inter-class variance (E, 1.5L and 6L) class. In Figure 7.10, a small group of data variables associated with 3L class (circled in red line) was observed to fall into its neighbouring class of 3.3L loading class. The same problem raised for 3.3 loading class (in red circle) where it collapsed into 3.6L loading class. The small data group that deviated from the majority of the data points associated with the same loading class can be considered as outliers and was discriminated from the loading class. This phenomenon was also observed in the plot of their original data signals so PCA helped to verify this problem. This is an essential step before applying test data set from various damage severities classes in comparison to the undamaged baseline data set, so that it helps to avoid a false damage or loading interpretation.

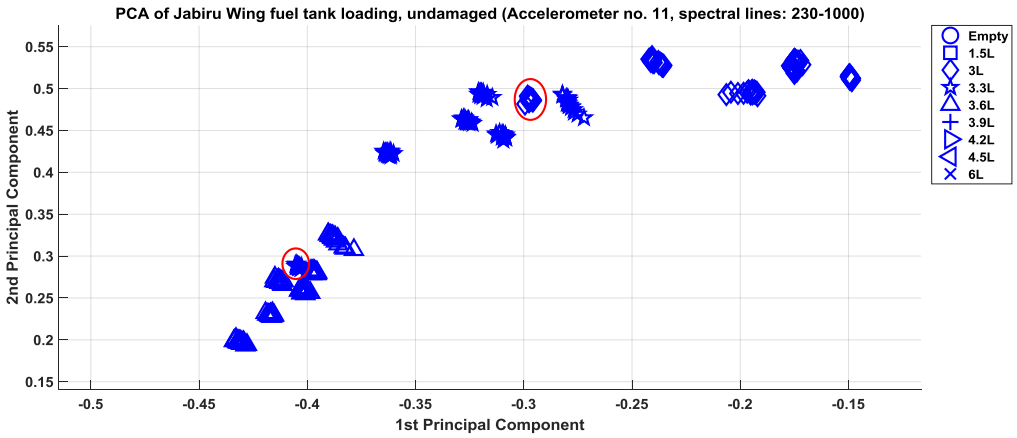


Figure 7.10: Initial PCA application reveals of some outliers from 3L class and 3.3L class that included into a wrong class belonged to the adjacent loading class.

It is essential to examine all of the frequency range including the higher frequency spectrum so that the selected feature is the most appropriate one. Having feature that is sensitive to loading and damage, it can establish useful visualisation results and can give better damage detection used in outlier analysis at later stage. Using a standard PCA, it was performed on each block comprised 1000 spectral dimensions as shown in Figure 7.11. Note that, the first block begins at spectral line 230 and the last block at spectral 7000 to exclude the high noise signals. A complete data characteristic from E to 6L of loading process can be obtained and included on a single projection when consider a full trend of the FRF signals as a result in changing of load. In the following section, it will be described the use of PCA on each block of spectral dimensions and to observe different characteristics in respect to loading variations.

Note that, at this stage, data sets from damage conditions are not included because the main concern at this point is to obtain visualisation of the complete trajectory corresponding to the baseline data variables. Data sets corresponding to damage conditions will be applied on the baseline set after selecting the appropriate feature.

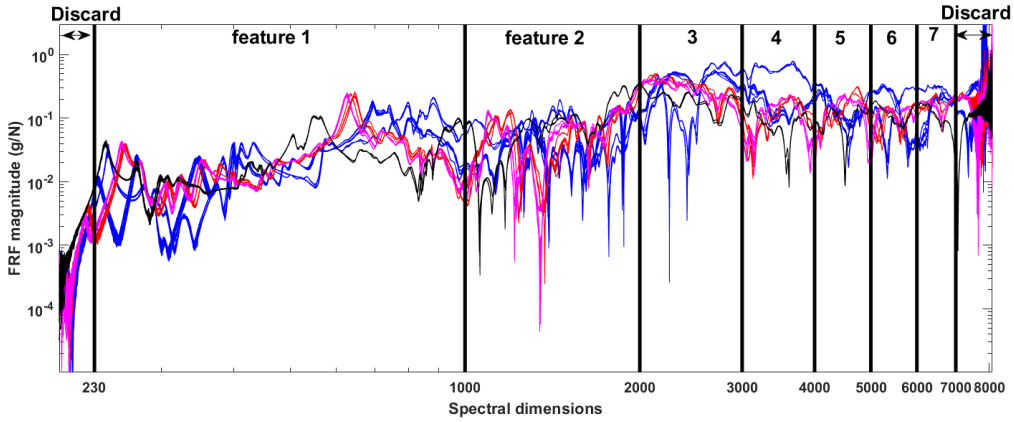


Figure 7.11: An illustration that shows how feature selection is divided and analysed in the JW data set.

The main indicator of the selected feature should be that the data class of different loadings display distinctive separations for all nine loading classes and assume a trajectory pattern for the purpose of monitoring the structural condition. In the previous visualisation results of wing-box experiment as highlighted by Figure 5.18 and Figure 5.19, there is an advantage if the projection can describe the trajectory or track pattern in respect to SHM. Adding the data set from the damage conditions under similar loading conditions, as the wing-box result highlighted, the test data path (damage conditions) will assume the similar track pattern. The expected challenge is that the possibility of data overlap between the baseline and test set.

Practically, distinguishing the effects of operational loadings would not always be the main issue. Furthermore, loading parameter is typically a controlled parameter compared to the damage parameter, which is under normal circumstances, is unknown. In the following results, using standard PCA, it will be shown the process of extracting desired FRF features before it is considered as the main feature for the Jabiru wing.

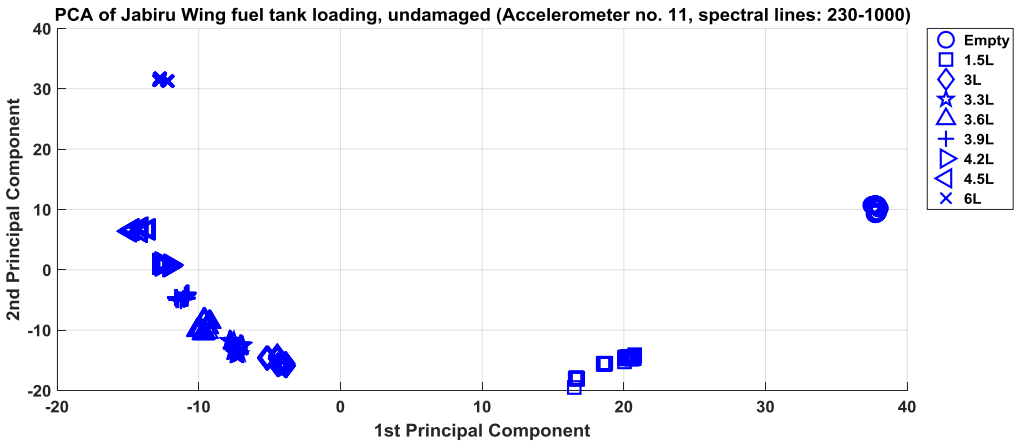


Figure 7.12: The potential feature candidate due to the distinctive data separation including for class with smaller incremental load.

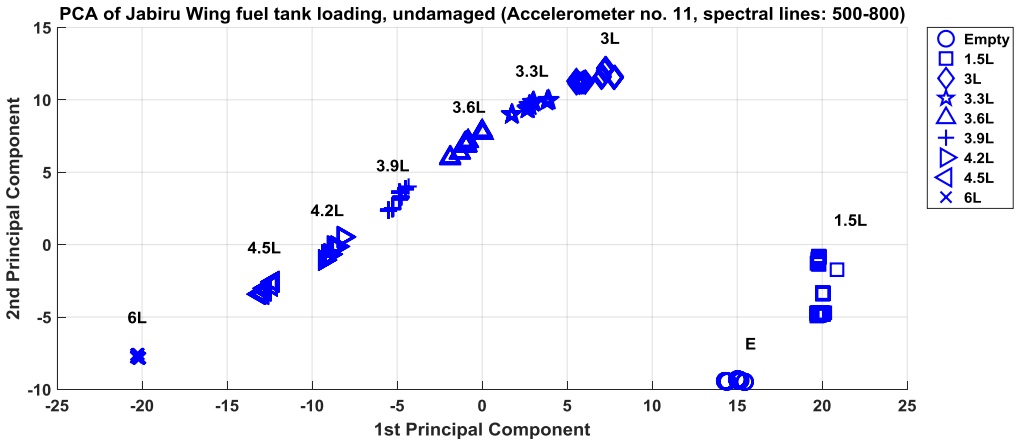


Figure 7.13: Distinct data separation but capture incomplete trace of data pattern.

Figure 7.12 displays systematic order of loading class with distinctive separation of class including the smaller inter-class variance. It exhibits a full tracking pattern that can be useful in SHM when test data is projected along the baseline set. In comparison to Figure 7.12, the projection shown in Figure 7.13 indicates the effects of using a shorter range of spectral dimensions of 500-800 spectral lines in the PCA. The effects that can be seen from using shorter dimensional range is that there relatively incomplete trajectory (more information loss due to less number of dimensions) with more of straight trajectory pattern.

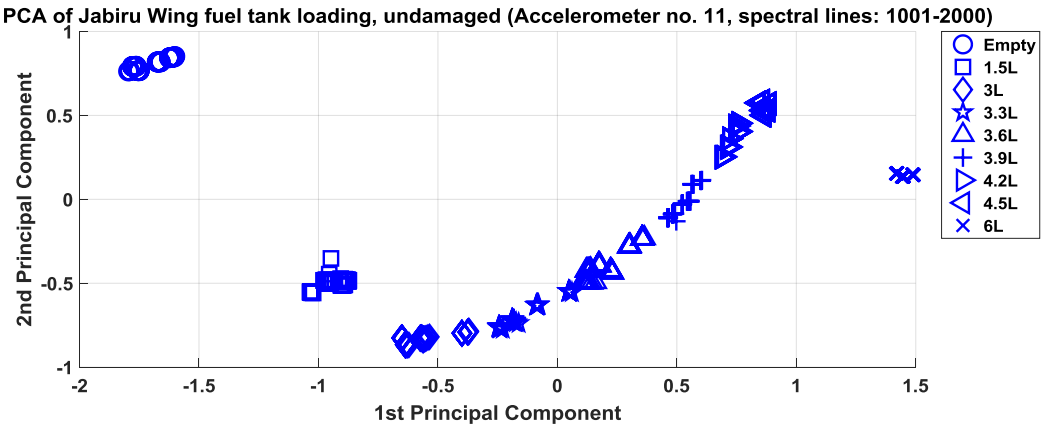


Figure 7.14: PCA on higher spectral dimensions using baseline data sets.

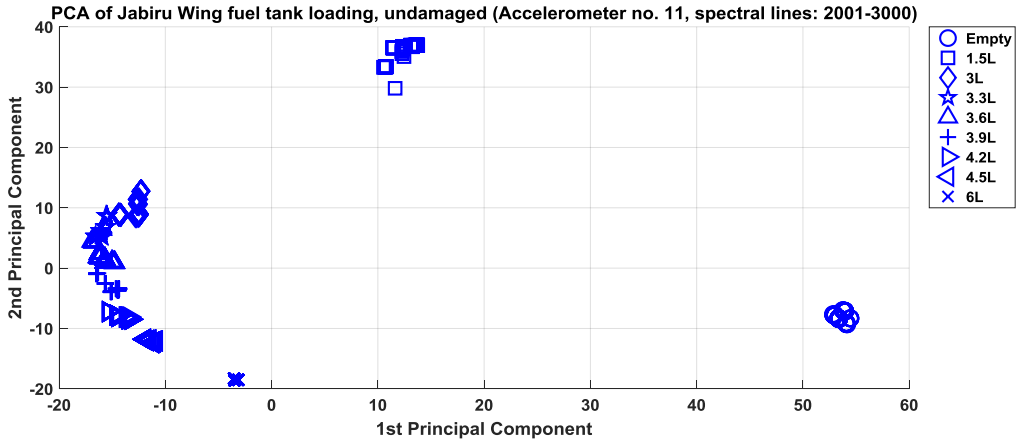


Figure 7.15: PCA on the next block consists of spectral lines 2001 – 3000.

Moving into higher spectral dimensional range, PCA is applied on the selected range of spectral lines as shown in Figure 7.14 and Figure 7.15, it is indicated that the separation between data classes with smaller incremental load is less distinctive with some data overlaps especially between 3.3L and 3.6L loading class compared to the previous feature with 230-1000 spectral lines. Presence of overlapping within these small incremental loading classes make this feature is not favourable for use.

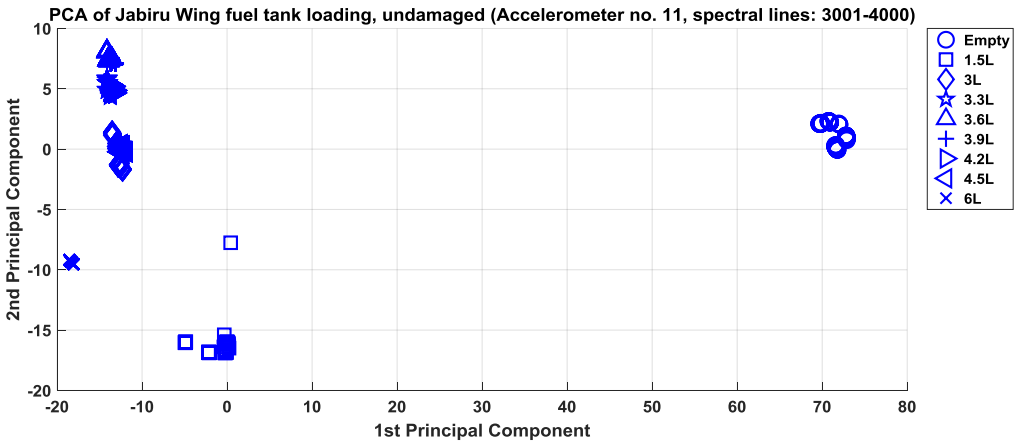


Figure 7.16: Using PCA for feature extraction performed for the 3001- 4000 spectral lines of frequency range 750 Hz – 1000 Hz.

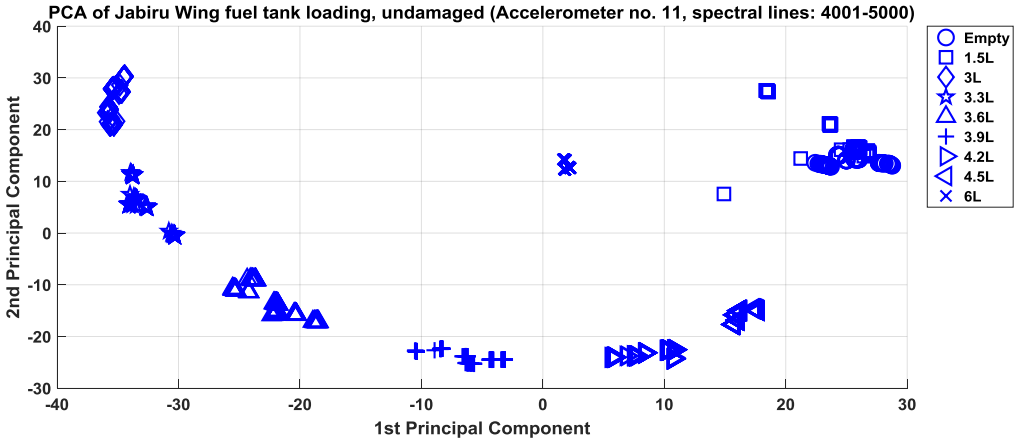


Figure 7.17: PCA extraction for the frequency range of 1000 Hz to 1250 Hz.

The results from Figure 7.16 show very high of data overlapping around the loading classes of small loading increments between spectral lines 3001 and 4000. The next feature extraction using higher frequency range 1000Hz to 2000 Hz indicates encouraging variables separation particularly for smaller inter-class variance. However, there is highly overlapped data variables between E and 1.5 loading class. In the interest of establishing a damage-sensitive feature, this spectral range seems appropriate but the results show less sensitive to large variability make it unfavourable for selection. From the results of feature extraction, it can be described that, the feature comprises 230th to 1000th spectral lines is

the most appropriate to use as the feature for the JW. This feature set will be usedn and will be applied with the test data consists of damage 1 (D1), damage 2 (D2) and damage 3 (D3) with comparison to the baseline data set.

In this section, it has been demonstrated how a standard PCA can be utilized in the process of selecting a feature that exhibits different data characteristic in terms of the variables separation and data trajectory. In the next section, the focus will be to examine the test data sets and applying them with the coefficients (loading matrix) of the baseline PCA (of undamaged condition) after standardising the test set over the baseline set.

7.5.1 Q-statistic for inspecting outliers

Before implementing the PCA dataset with any machine learning algorithms, it is essential to verify whether the data set has a good-fit based on the full dimensional data sets. Q-statistics are the residual or the difference between the score projection in full dimensional space and the projection in reduced dimensional space. In the interest of determining how well each data set in the PCA model conforms to the original dimensional space space.

In Chapter 2 of Section 3.3.2, definitions Q-statistic was explained. To implement Q-statistic, the difference between the T-squared values in the full space and the reduced space was computed. The residuals from test data set and baseline set are compared and computed using a Mahalanobis Square Distance (MSD) function to seek any data abnormalities lie in the data space.

Q-statistic also indicates the extent of each spectral variable contributing to the sum of squared error due to the lack of fit between the full dimensions and lower dimensions. For this case study, Q-statistic can explain if the reduction of dimensions contribute to high error. If the values of Q-statistic is high which indicates large error between the PC models and the original model, a straightforward solution will be to consider higher number of principal components to reduce the residual error.

The first step is would be to perform PCA based on the baseline set (undamaged condition) and apply the coeeficients on the test data. Prior to that, the baseline set is standardised and the test set is standardised using the mean and standard deviation of the baseline set. This model is parallel to the PC model A described previously in Chapter 6 under Section 6.5.1. In dealing with larger data sets consist of many variables, the technique based on Q-statistic can shed some quality of data acquisitions and some outliers related to variability due to the high noise which are not caused by damage can be removed.

Some outliers are detected when dealing with the D2 and D3 data set. There is few data points in D2 and D3 that observed to deviated from the most of data points corresponding the same data group as displayed by Figure 7.18. One way to solve it is to remove these outliers from the test set detected in the residuals using a confidence interval. Based on 99.7% confidence interval with an equivalent constant of the value of 3, the data points are evaluated if they fall within the mean of the Q-statistic (Figure 7.19). The following results illustrate the effects of outliers are being removed from the damage data set.

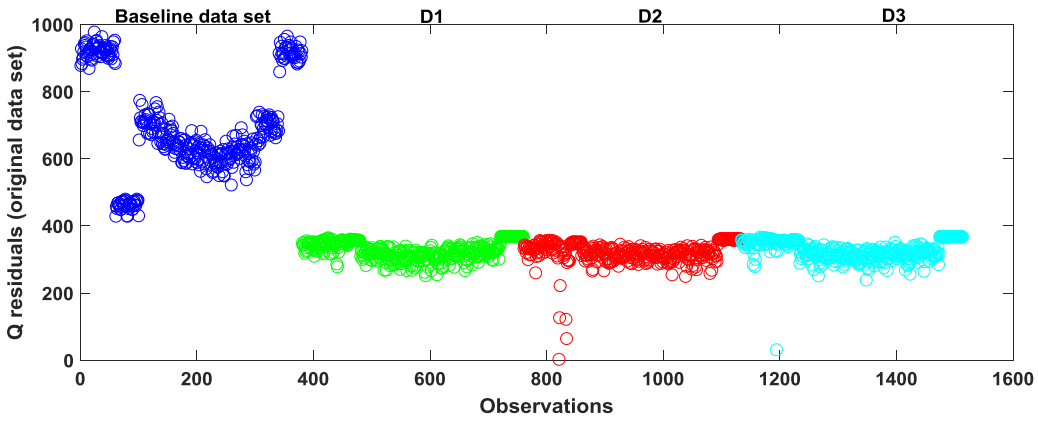


Figure 7.18: The results using Q residuals test for test data set.

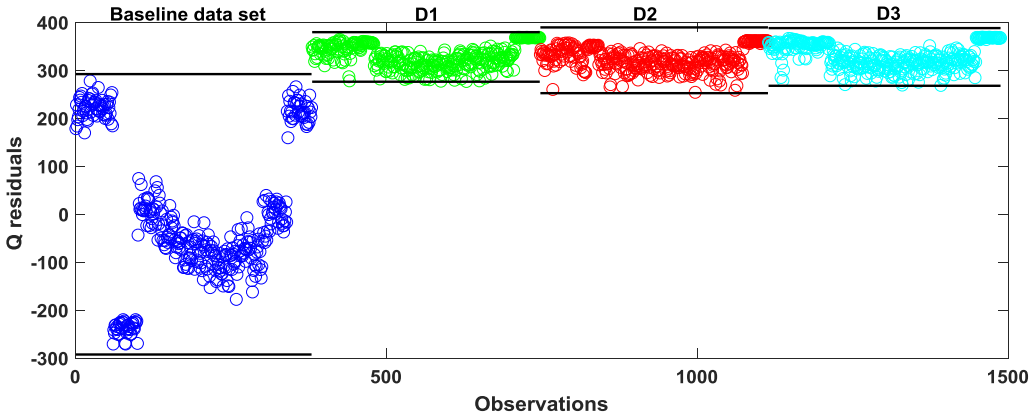


Figure 7.19: The result of Q residuals after removing the outliers from D2 and D3 test set.

The Q residuals obtained (Figure 7.18) show the raw data lack-of-fit between the full space observations and the dimensional reduced space. It indicates some extreme outliers within D1 and D2 data set that cause inaccurate comparison with the baseline data set. As MSD is very sensitive to outliers, any extreme values that deviate from the rest of the data set would be exposed in the plot (Figure 7.18) and the associated data samples are excluded in the data set before implementing further machine learning task. By removing outliers in the data set, the damage detection results provide more accurate and reliable damage detection (Figure 7.19).

7.6 Data visualization and feature extraction

The goal in this section is to demonstrate the data transformation of the vibration data set from the JW, originally in high dimensional space to a reduced two-dimensional space for the purpose of visualization and pattern recognition. This is achieved by using a standard PCA and the kernel Gaussian PCA. The key advantage of introducing kernel PCA as a nonlinear form of PCA lies on the capability of computing the principal components in high dimensional feature space related to input space by using nonlinear mapping function by means of kernel function computed in input space [52]. By obtaining principal

components in high dimensional space, it potentially provides better data characteristics and deeper features that normally lives in a possibly high dimensional space including the more complete data trajectory pattern from the Jabiru wing data. Previously, in the wing-box data set, kernel PCA had been utilised very well to expose the hidden variables by improving the data separation.

In this chapter, based on the baseline model that follows the procedure explained in Section 5.5.1 of Chapter 5, standard PCA is initially performed before implementing the kernel PCA. The PCs projections associated with test data sets are then projected onto the same plot as the baseline projection.

To implement kernel PCA on the JW data set, the procedure follows the similar steps as finding the inverse variance (parameter sigma) based on the mean of the selected smallest distance and the distance matrix applied as described in Chapter 5 of Section 5.4 for the baseline set. Basically, the technique involves dot products $k(\mathrm{x},\mathrm{y}) = (\phi(\mathrm{x}) \cdot \phi(\mathrm{y}))$ and instead of having to solve the nonlinear mapping ϕ into high dimensional space, the 'kernel trick' representations is used to solve the nonlinear transformations. The kernel used in this work is given by $k(\mathrm{x},\mathrm{y}) = \exp\left(-\|\mathrm{x}-\mathrm{y}\|^2 / \mathbf{2}\sigma^2\right)$ similar to the one used in the wing-box data set and follow the mathematical principals in Section 3.2.2 of Chapter 2.

The first step is to display the PCs projections of baseline data set using the selected feature set as described previously. In the following results, comparisons of standard PCA and kernel PCA are illustrated.

Based on the result from Figure 7.20 using standard PCA, it is indicated that variables from each loading conditions are clearly separated including the loading classes associated with the smaller loading increments. For the kernel PCA transformation as shown in Figure 7.21, the separation of the loading class has shown to improve and build better data trajectory.

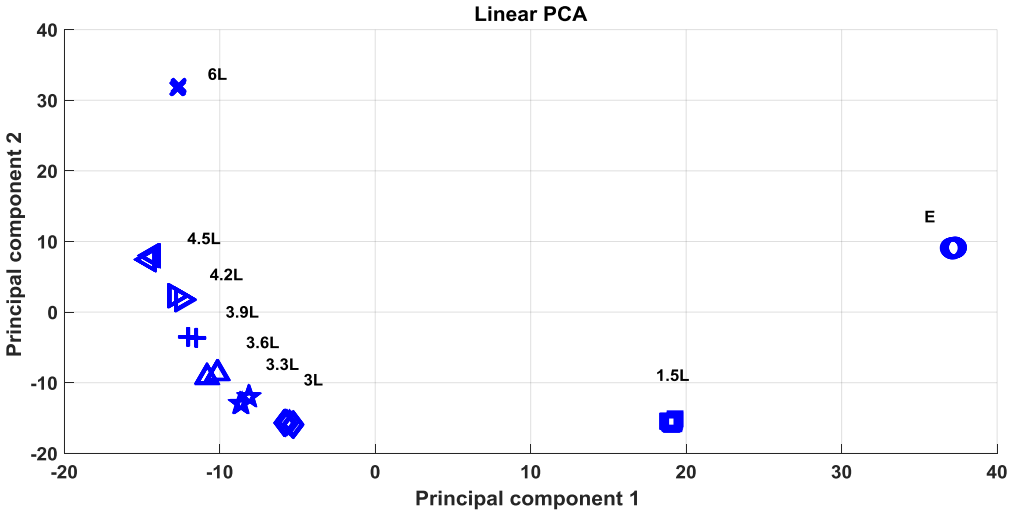


Figure 7.20: 2-D visualization using the undamaged data set (baseline set).

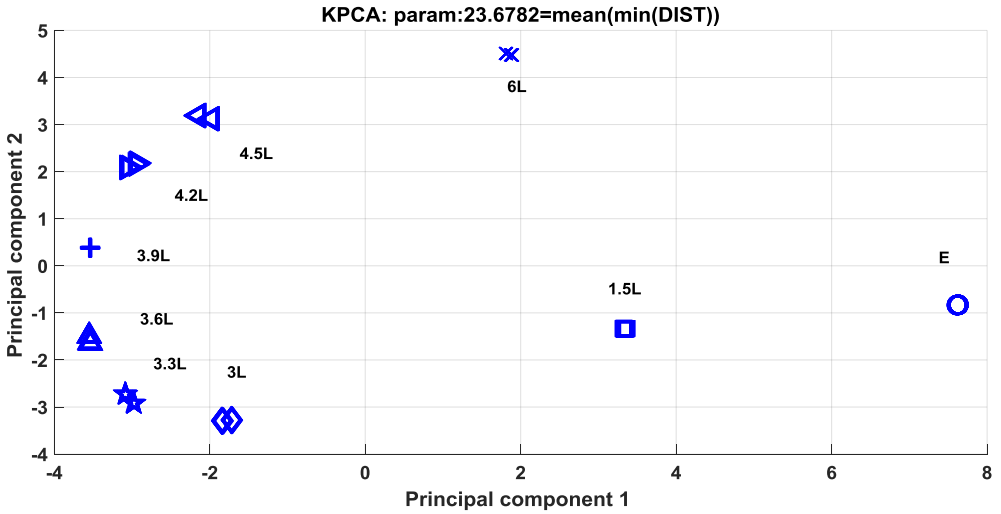


Figure 7.21: The result of performing kernel PCA on the similar baseline set.

7.6.1 PCA for saw-cut damage

At this stage, PCA is applied on the test data set from the JW damage conditions and projected on the plot consists of PCs projections corresponding to undamaged condition. The goal is to visualise the PCs projections and to verify if the new projections associated

with damage conditions can provide an indication for a damage possibility in the wing. This is achieved by comparing the new PCs projections to the existing PC projections (from undamaged condition) with respect to their separation.

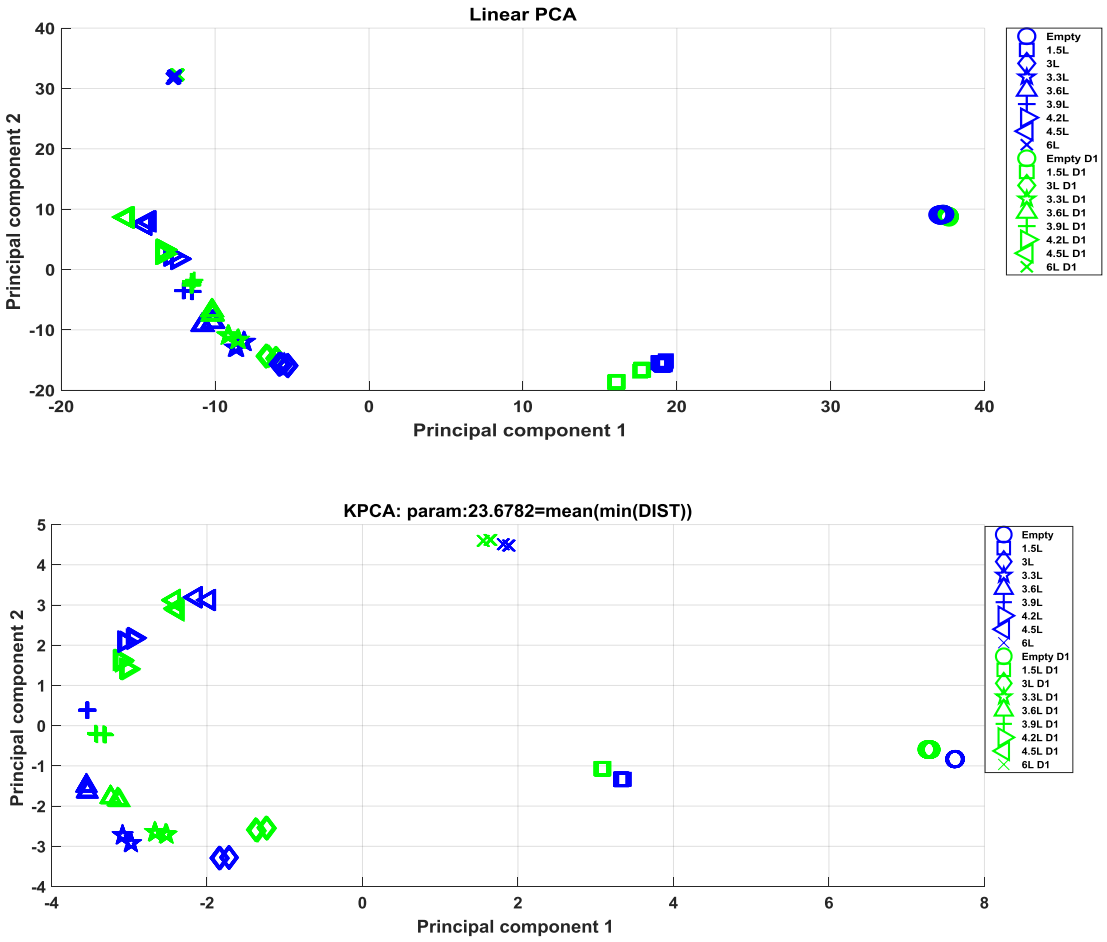


Figure 7.22: A comparison of the data transformation between the linear PCA (top figure) and the Gaussian kernel PCA (bottom figure) using undamaged condition data set as the baseline with respect to D1 data set.

Referring to Figure 7.22, kernel PCA has achieved a greater degree of data separation between undamaged and D1 data set. The data class from the smallest saw-cut (D1) shown not overlap with the PC projections of the baseline set. It has also established better data trajectory in form of nonlinear curve compared to the linear PCA which is restricted to a straight trajectory. Adding PC projections from other test data from higher damage

severity, D2 and D3 test set, the results in the standard PCA show significant data overlapping between the baseline and all test data PC variables.

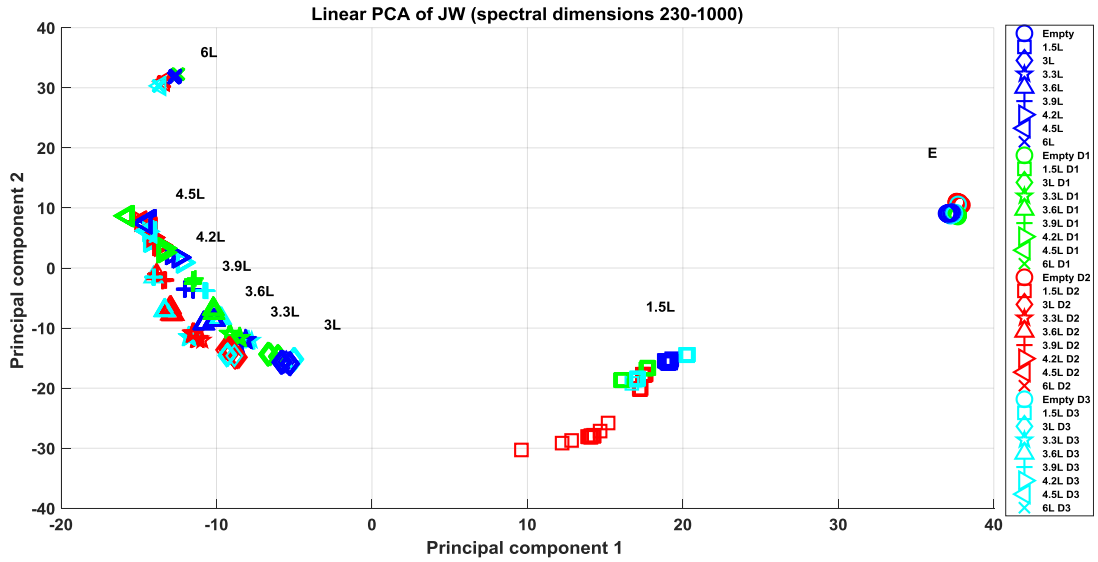


Figure 7.23: Linear PCA transformation in the comparison of undamaged and all damaged conditions based on the undamaged data set as the baseline.

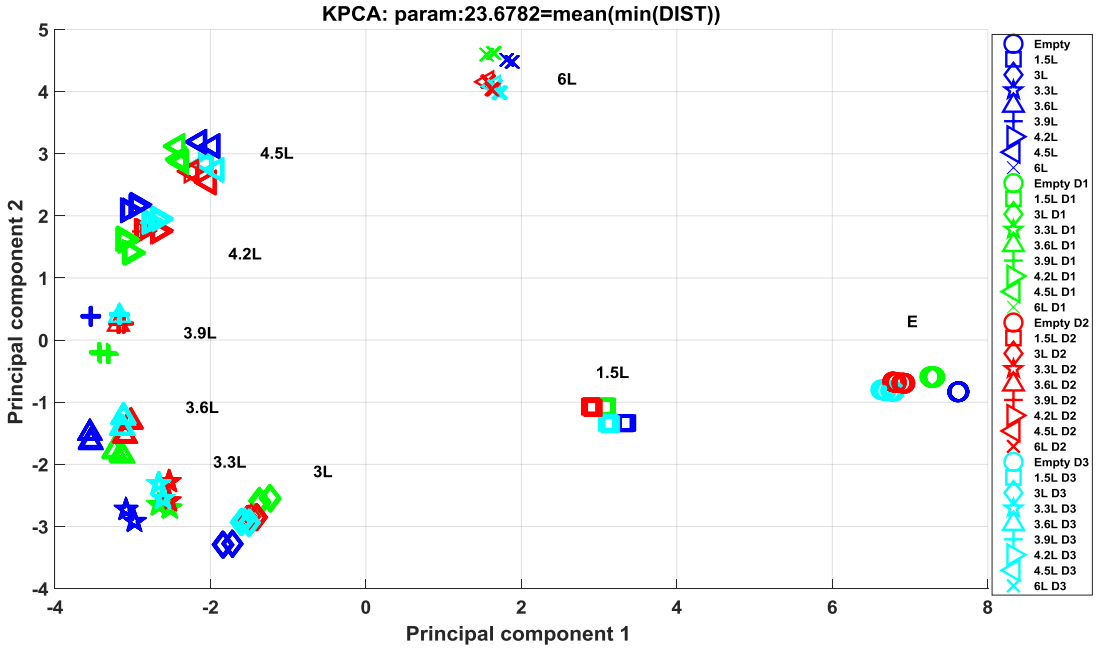


Figure 7.24: The result obtained from nonlinear transformation via kernel

In contrary, kernel PCA has set a greater degree of data separation among the baseline set and the all test sets. Noting the damage severities are not correlated with the degree of separation except for the empty fuel load. In the author's perception this can be due to the degree of damage introduced in the wing bracket which is only 4 mm increase from D1 to D2 and D3 respectively.

A close view around the class associated with small loading increments as illustrated in Figure 7.25, it reveals some effects of small loading variability on damage variability. From the linear PCA result, it shows that due to the effects of small loading variability and damage severities, linear PCA can project the variables associated with damage condition into variables corresponding to next loading class associated with undamaged condition given the loading variability is small (marked by red ellipse). Pertaining to the result shown in Figure 7.25, variables belonged to the class associated with D2 and D3 structural conditions and 3L loading condition are projected into subsequent 3.3L loading class belonged to undamaged condition. These variables marked by red ellipse indicate the variables having such problem as illustrated in Figure 7.25.

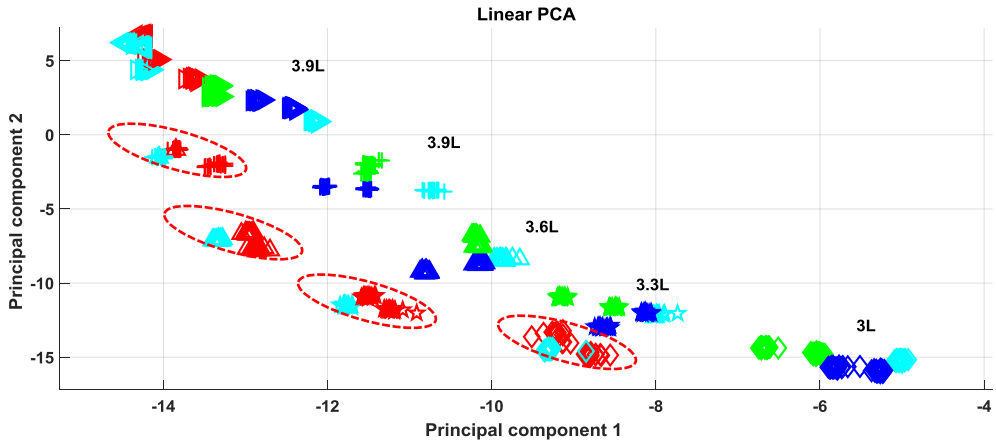


Figure 7.25: A zoom-in view around the classes of small loading increment. The red-dash ellipses mark the variables projected into wrong data classes.

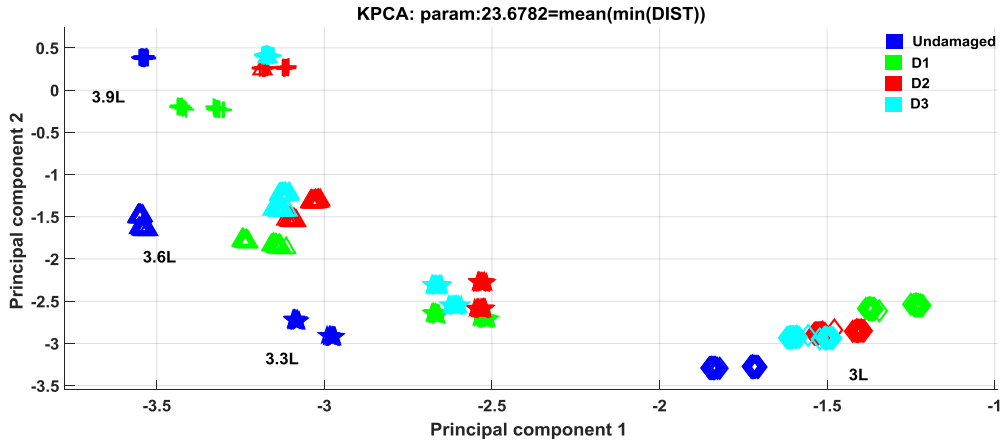


Figure 7.26: Damage severities corresponding to classes of small incremental loading using kernel PCA.

This clearly shown that standard PCA has restricted capability in discriminating damage conditions from loading conditions when it involves small variability. This is the underlying problem in SHM where both small loading variability and damage variability present similar effects on the FRF data signal and there is a limitation using linear PCA for such problem.

On the other hand, the study verifies that by utilising kernel Gaussian PCA, the problem related to the linear PCA projection particularly for variables with small loading variability and damage severities classes can be greatly reduced. As seen in Figure 7.26, all variables irrespective of different damage conditions are projected into their particular loading classes. There are several possible explanations for the superior performance of kernel PCA in respect to this JW case study:

- The kernel function created from distance matrix is computed uniquely based on each test data set in which each set is associated with one damage condition and encompassed all loading conditions.
- The inverse variance is also adapted from each test set via the distance matrix. As the distance matrix measures the data distance between baseline set and the test set, the inverse variance which is calculated based on this distance of the selected smallest distance represent the variability in the test data.

- The effects of the inverse variance in the denominator of the exponential term in the kernel $k(\mathrm{x},\mathrm{y}) = \exp\left(-\|\mathrm{x}-\mathrm{y}\|^2 / \mathbf{2}\sigma^2\right)$ provides a capability for the variables to be better identified and more separated when the inverse variance is small. Note that $\|\mathrm{x}-\mathrm{y}\|^2$ denotes the Euclidean distance calculated in the distance matrix.

Kernel Gaussian PCA has proven of its ability to enlarge the distance between variables of different test data before general treatment of PCA is applied on the adaptive kernel. By eigenvalues decomposition, $[\mathrm{V},\mathrm{E}] = \mathrm{eig}(K)$, PCs are computed from the eigenvectors V which are sorted with respect to the eigenvalues E in ascending order using the centralised kernel matrix *K*.

The zoomed view on of the loading class in Figure 7.27 below signifies the capability of the kernel representation in the interest of SHM and damage identification. The results from standard PCA and kernel PCA are compared to illustrate the positive effects on using the kernel PCA. The variables from undamaged, D1 and D3 are highly overlapped among each groups in comparison to variables projected using the kernel PCA which are clearly separated and correlated properly with damage severities.

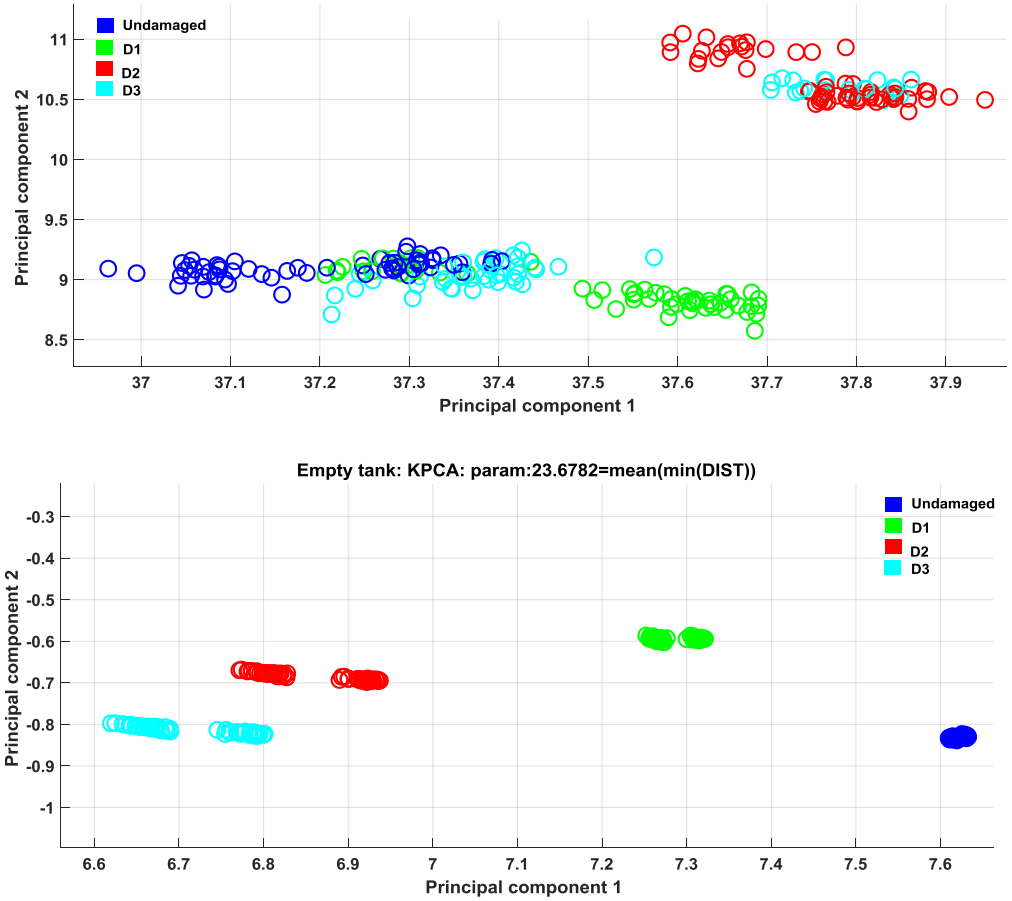


Figure 7.27: The results of standard PCA and kernel PCA focusing on empty fuel load.

7.6.2 JW novelty detection

In this analysis, MSD is used to compute the distance of the principal components (PCs) projections between baseline set and each test sets. It is worth mentioning that, in kernel PCA, it involves finding the eigenvectors of the N x N matrix (number of observations, N) rather than the D x D matrix (number of spectral dimensions, D). Therefore, the user has to determine approximate the number PCs to include the outlier analysis.

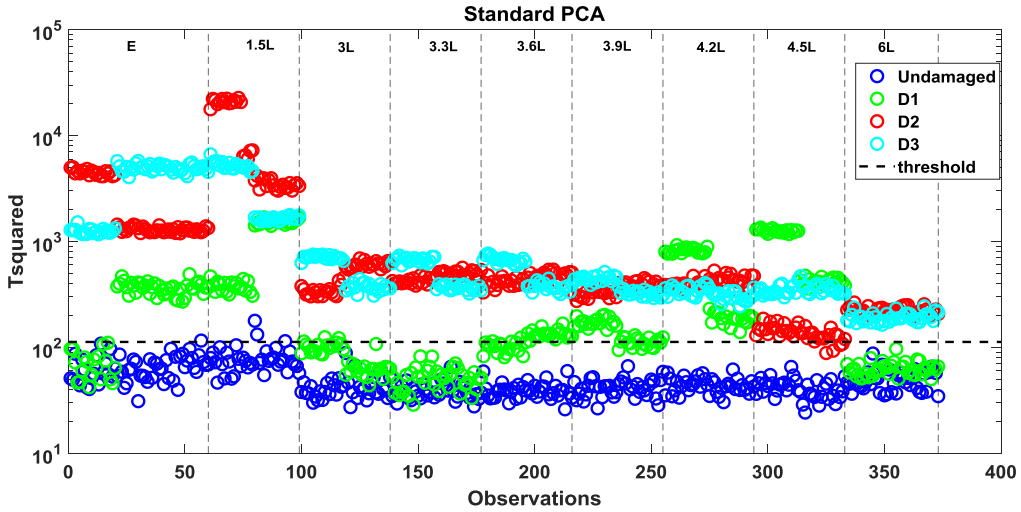


Figure 7.28: Novelty detection for the JW using standard PCA.

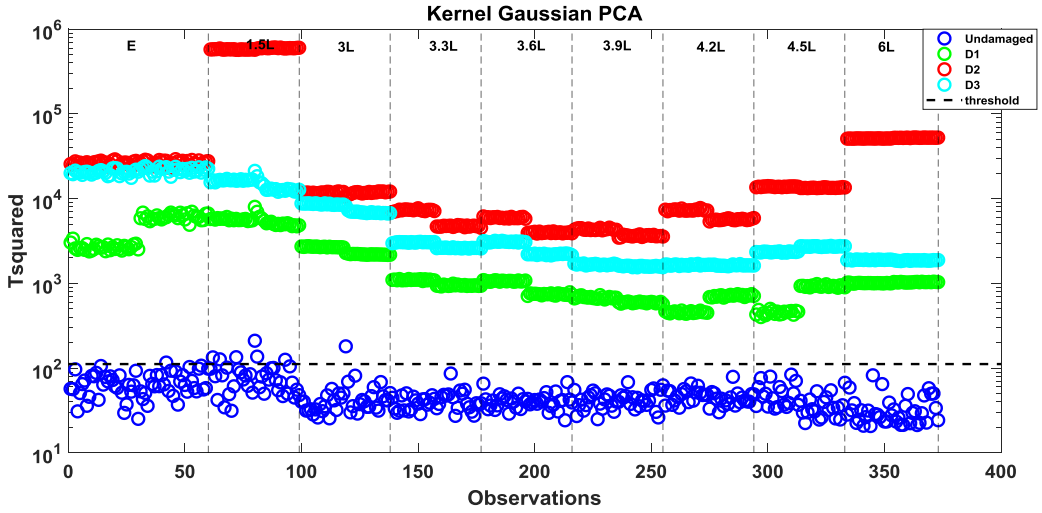


Figure 7.29: The novelty detection result using PCs variables obtained from kernel PCA.

In respect to this JW data set, it is identified that 50 principal components in the reduced dimensional space is appropriate in the outlier analysis. A higher PCs than 50 is shown to cause no much improvement on the novelty detection and if lower number of PCs makes the novelty detection to perform unsatisfactorily.

The novelty detection from the standard PCA shows the damage detection fails to detect the lowest damage D1 as damage for almost all loading class. The variables of D1 is falsely

identified as undamaged known as false negative detection. This scenario can make the damage unnoticed and can increase the possibility of higher severed damage in shorter period of time if the initial damage is overlooked and unresolved

On the other hand, the damage detection for all variables obtained from the kernel PCA clearly gives 100% accurate damage prediction for all loading class (Figure 7.29). In addition, the score variables associated with each test set clearly separated from the threshold line. The variables also show to be correlated well for D1 test set but switched between D2 and D3. It is observed of a trend in each test data through all data observations as the loading is slowly switched from one class to another due the large and small increase of loading. This effect of a slow changing between data variations as the mass loading is changing from one class to the next class gives better identification of damage severities in addition to the accurate damage detection which cannot be indicated in a standard PC. A similar trend of switching latent variables is higlighted recently using Treed Gaussian Process in [32].

7.6.3 Identifying key spectral variables

Using the features computed from the PCA, the highest variance represented by the dimensional variables in the principal components can be found via the coefficients or the loadings matrix of the PCA. It tells the contribution of the spectral variables in the two first principal components. Principal component 1 and 2 (coefficients) constitute the eigenvectors corresponding to the largest eigenvalues or highest data variability measured for all the dimensions. Using this criteria, the spectral variable correspond to the most significant FRF peaks in the PCA can be identified and should be included in the feature selection.

Figure 7.30 represents all spectral variables (230 -1000 spectral lines) included in the PCA. The blue lines denote the vectors of each 771 spectral variables showing their direction and length. They vectors indicate how each variables contribute to the principal components in the PCA plot. The red dots in the plot represent the data observations with

their their coordinates indicate their scores values for the first PC which is the horizontal axis and second PC indicated by the vertical axis [72].

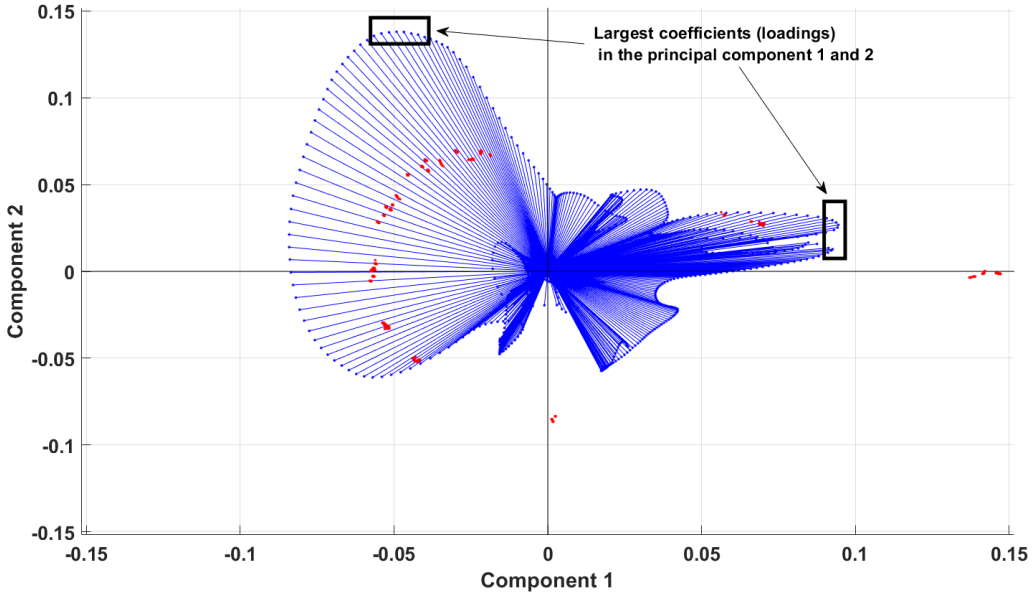


Figure 7.30: The plot of coefficient vectors on the standard PCA of the baseline data set.

The information about the spectral dimensions (spectral variables) associated with the FRF peaks that gives the highest coefficients on the PC projection can provide indication for spectral range should be considered in feature selection. The highest coefficients of the first principal component (horizontal axis) indicate the highest variance or eigenvalues corresponding to spectral variable 569 and 570 which is one of the FRF peaks as indicated in the FRF plot in Figure 7.33. The spectral number 650, which shows significant contribution on the PCA (Figure 7.31) in terms of the highest variance, is also one of the main FRF peaks in empty fuel tank load as indicated in the FRF plot in Figure 7.33. This technique allows better decision on feature selection by focusing on the variables that contribute higher coefficients (loading) values in the PCA plot. Hence, using this information, one can reduce the feature's dimensions by concentrating more on the spectral variables that have highest coefficients in the first two principal components.

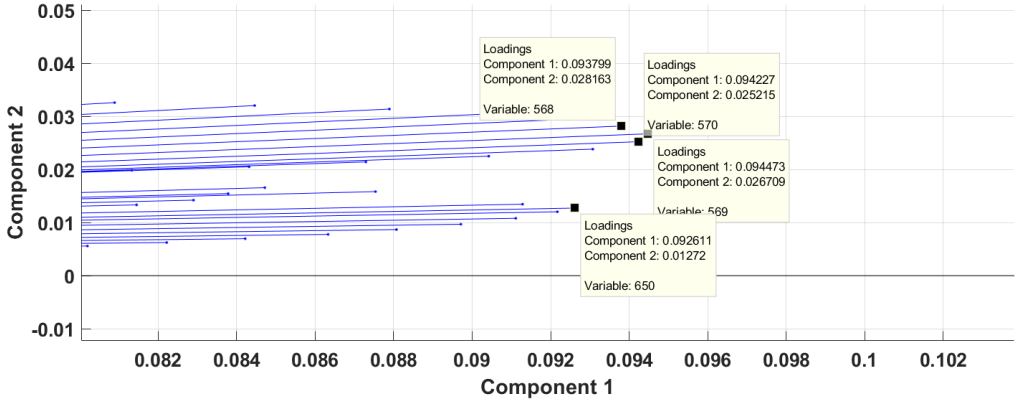


Figure 7.31: Zooming-in into the component 1 of the above plot:

The zoomed view on component 1 in Figure 7.31 highlights the spectral variable 650 and 569-570 that show the largest coefficients values on the first PC. These spectral variables, which has the highest coefficients, correspond to the FRF peaks as shown in Figure 7.33. Figure 7.32 shows the spectral variables (in blue lines) on the second PC which indicate spectral line 407 to 409 and this range of spectral dimensions should be included in the feature to ensure good data variability and structure's characteristics in term of the change of frequency peaks.

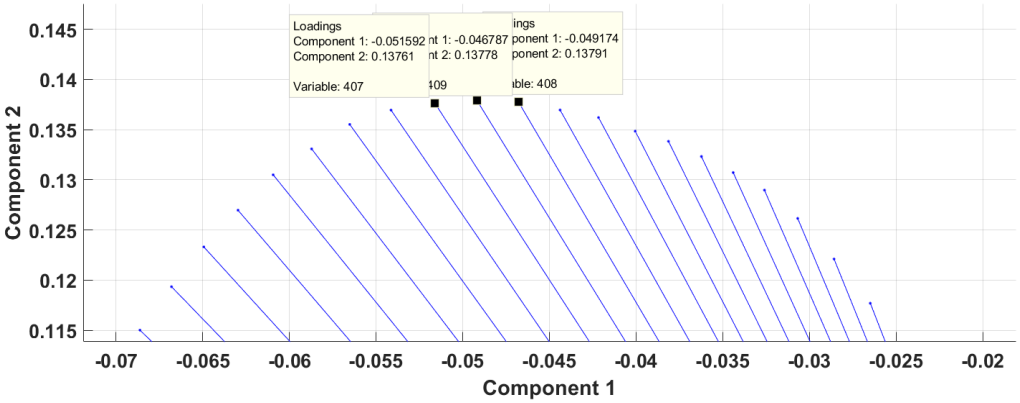


Figure 7.32: Zooming-into the component 2:

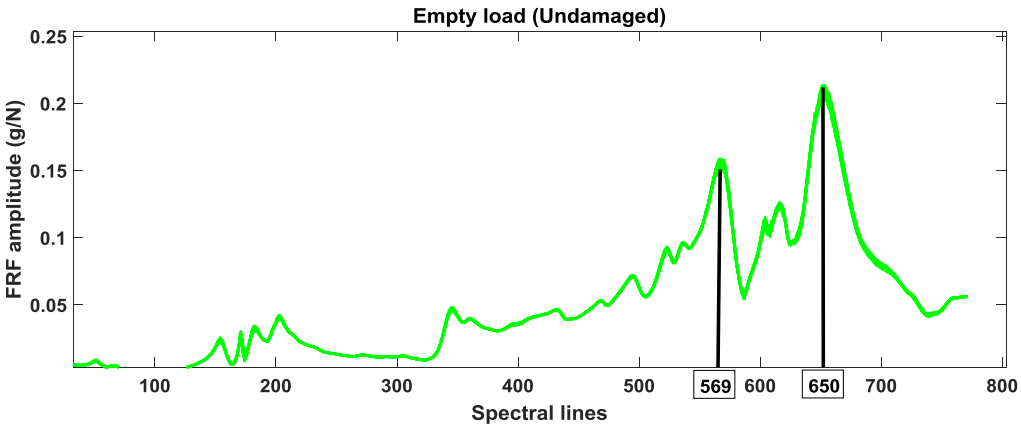


Figure 7.33: The peaks associated to the highest coefficient of the principal component 1is shown on the FRF plot.

The technique via PCA in finding the spectral variables of the FRF that contribute significantly to the PCA is discussed. Having appropriate number of feature dimensions are favourable in the machine learning perspective. In high data dimensions comprised of many loading classes, one of the challenge is will be to choose the appropriate number of spectral dimensions for its feature as shown in Figure 7.34. Plotting the coefficients of the first and second PCs, one is able to identify the spectral dimensions that contribute to the highest variance in the PCA and identify them on the FRF plot for its feature selection. Using this information, one can include these FRF peaks in the feature and possibly reduce the size of the feature by excluding the variables associated with low variance as illustrated by Figure 7.34.

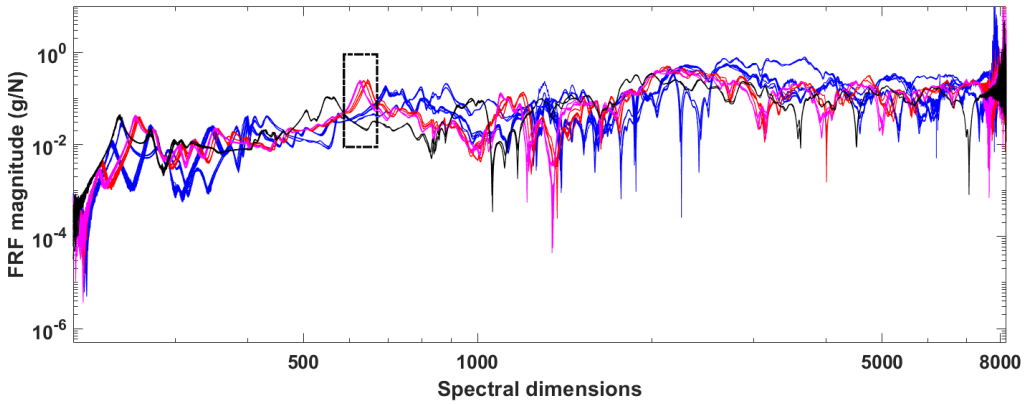


Figure 7.34: The location where the peaks with highest variance from the selected feature is chosen (from 230 to 1000 spectral dimensions).

7.7 Neural Network Regression model

In this chapter related to neural network regression which data labels are known, the aim is to map the input features obtained from the standard PCA to the target vectors. The target benchmark is divided into 2-column matrix with some arbitrary values corresponding to the numbers of damage and loading classes in similar process for the wing-box case study. The network is created based on the general procedures and guidelines explained in Chapter 6 of Section 6.8.

The network designed for the JW data set is a multilayer perceptron consists of one hidden layer and one output layer. The transfer function used in the hidden layer is a radial basis function (RBF). It has been found that the network gives good generalisation performance with 25 hidden nodes and 20 PCs as the input vectors. In this case, training algorithm Levenberg-Marquete will be used for training the network. The training is performed on the 70% of the overall dataset that is randomly selected from the overall data set. 15% of data set is used for validation and another 15% for a new testing.

For the JW case, radial base function (Gaussian function) is used as the the network transfer function. Using this transfer function, each inputs values is squashed with a bell-

shaped function centred at 0 in to the interval [0 1] with fixed variances. Here the net input for the RBF is the vector distance between the input vectors p and the weight matrix w. Several benefits have been identified using RBF (Radial basis function) [86]. Among them is the decision boundary that is hyper-ellipse compared to sigmoidal unit which uses hyperplane. With RBF activation function, it produces local response function through the exponential of distance squared $\exp\left(-\|w-p\|^2\right)$ rather than the dot product $x \cdot w$ as used in sigmoidal unit to produce linear response. RBF trains faster than sigmoidal unit function but it requires more hidden nodes (one hidden node for each input pattern) for its local representation. The network specification used for the mapping of PCA features into damage and loading target vectors for the JW is summarised as follow:

Network	Network architype	Training method	Training algorithm
MLP Feedforward	I-H-O (1 input layer, 1 hiddenlayer , 1 output layer	Backprogation	Levenberg-Marquadt

Number of hidden nodes	Number of inputs patterns	Number of output nodes	Hidden nodes transfer function	Output nodes transfer function
25	20 PCs	2	Radial basis function (RBF)	Linear

	Training set	Validation set	Testing set	Error function
Data set configuration	0.7	0.15	0.15	MSE

Table 7.7.2: Neural network specification for the JW

In designing three neural network models based on the specified parameters in Table 7.2, the study aims to investigate

- the effects of having unequal proportion size of data set associated with undamaged and damage conditions in a training data se

- reducing number of data dimensions.

7.7.1 The neural network with the higher dimensional spectrum

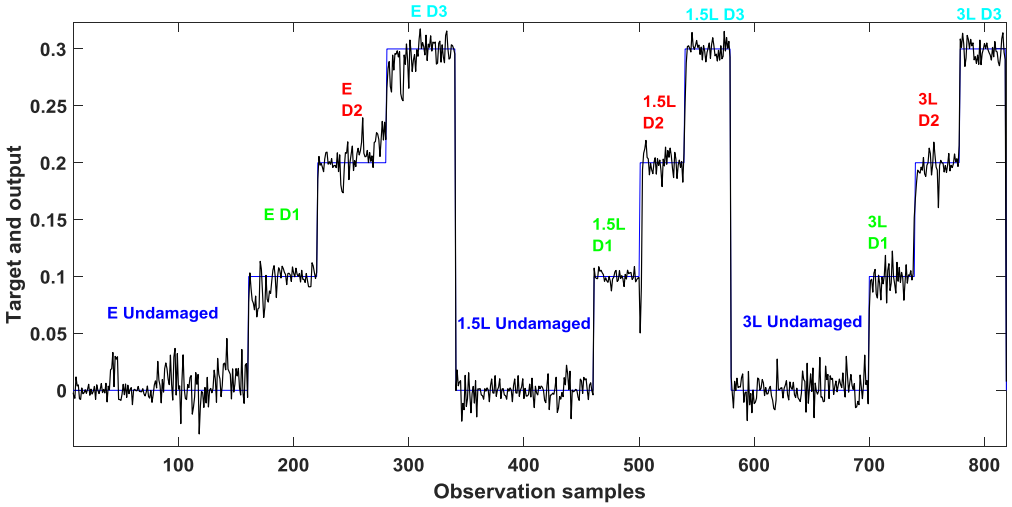
The input of overall data samples are randomly divided into subsets of training set, validation set and testing set using the 'dividerand' function based on the ratio specified in Table 7.3. Initially, the network is created based on the specified network parameters defined in Table 7.2 using the overall data sample 2253 with 771 spectral dimensions ranged between 230 and 1000. The network is trained several times until good accuracy is found. In the first network model, testing the trained network with respect to using all data samples gives the performance error as 1.898e-4. The overall configurations of the first network is summarised as

Network	Baseline set	Damage1	Damage2	Damage3	Damage sets
1	1120	380	373	380	1133

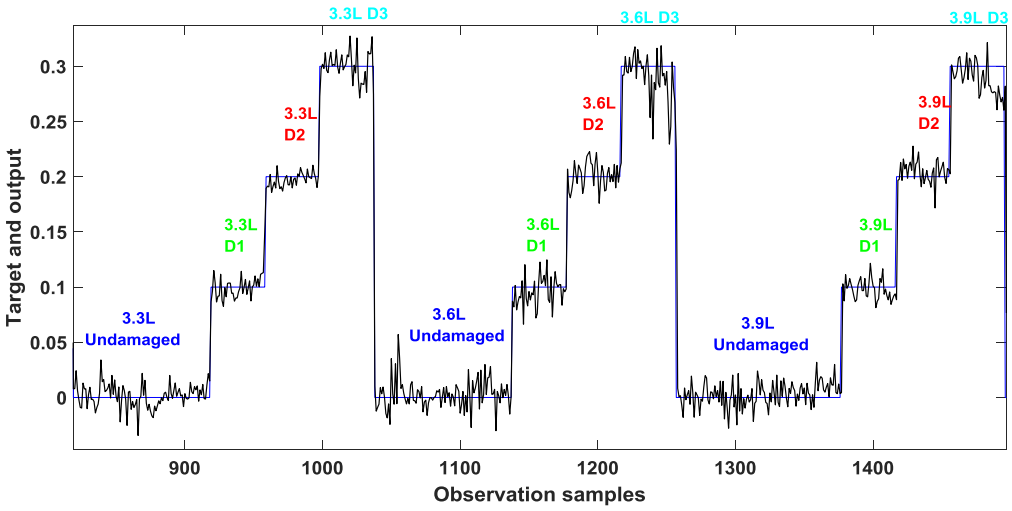
Network	Dimensions	Total observations	Training set (70%)	Testing set (15%)	MSE
1	771	2253	1577	338	1.898e-04.

Table 7.7.3: Data configuration of the network 1of the JW

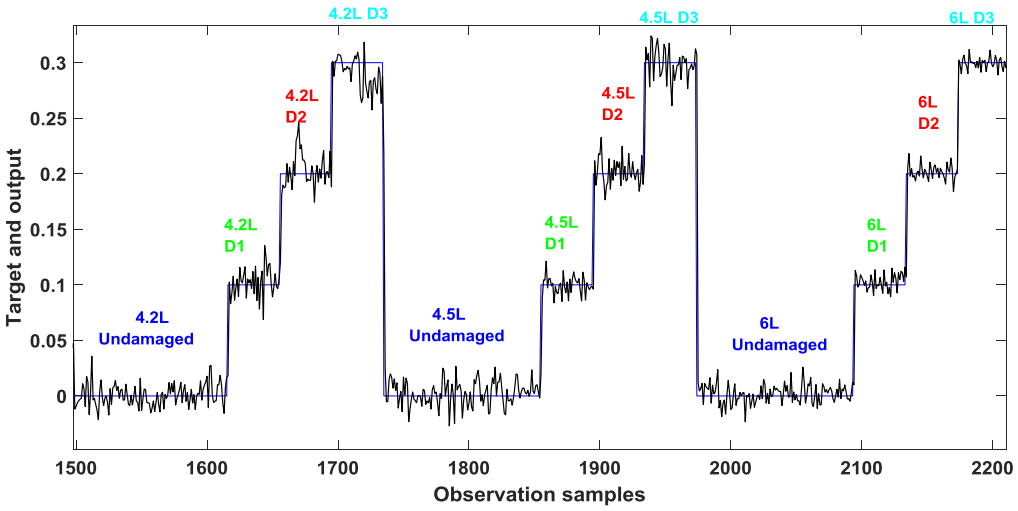
The mapping results using the first network model is presented in the following page. It is clear that the trained network effectively mapped the features from the linear PCA to their respective loading and damage classes. In each plots, there are 3 loading classes in which each loading class includes 4 structural conditions (Undamaged, D1, D2 and D3). The network has to map each of 20 feature patterns into one of those 36 classes (comprised of 9 loading class in which each loading class consists of 4 structural conditions). There are no significant change in performance in terms of the fitting error between networks outputs and the network target when the mapping is performed on loading class of either small increments or large loading increments.



(a): Fuel tank loading class E, 1.5L and 3 L inclusive of 4 structural health conditions.



(b): Fuel tank of smaller incremental loading- 3.3L, 3.6L and 3.9 L inclusive of 4 structural health conditions.



(c): Subsequent fuel tank loading class 4.2L, 4.5L and 6 L inclusive of 4 structural health conditions.

Figure 7.35: Overall mapping of input pattern from standard PCA score data to the benchmark target of 4 damage class and 9 loading classes using defined JW neural network.

7.7.2 Reducing the baseline samples and dimensions in neural network

In this section, the network generalization is evaluated by reducing the number of samples of the baseline data set as well to equalise the number of samples for the baseline and damaged data set (and referred as network 2). In the previous network (network 1), the baseline data is made up of 2.84 times more number of observations than each damage sets. Despite the high number of overall data samples that is 2253 data samples, the main concern is that whether the generalisation can be affected due to the disproportion number of data samples between undamaged and damaged set.

To verify this ambiguity, the number of samples in baseline set and in all damage sets are made to be proportionate. Using the same network's parameters and training function, the network 2 composed of reduced number of observations is trained and tested using the similar network parameters and architype. The performance of the network 2 is computed

and after testing the trained network several times to find the smallest error which is 6.588e-4.

The third network model is implemented using the same network parameters and training function but now both data samples and number of dimensions are reduced. The number of spectral dimensions is reduced from 771 to 301 which reduced to more than half of its dimensional size. The input feature used for the network 3 is shown in Figure 7.36 which shown to be almost impossible for anyone to differentiate between the undamaged and all damaged sets based on visualisation. It is also shown the loading class of small loading increments (3L to 3.9L) overlapped with their subsequent class and this visualisation brings no meaning in the context of damage monitoring without further use of an appropriate machine learning.

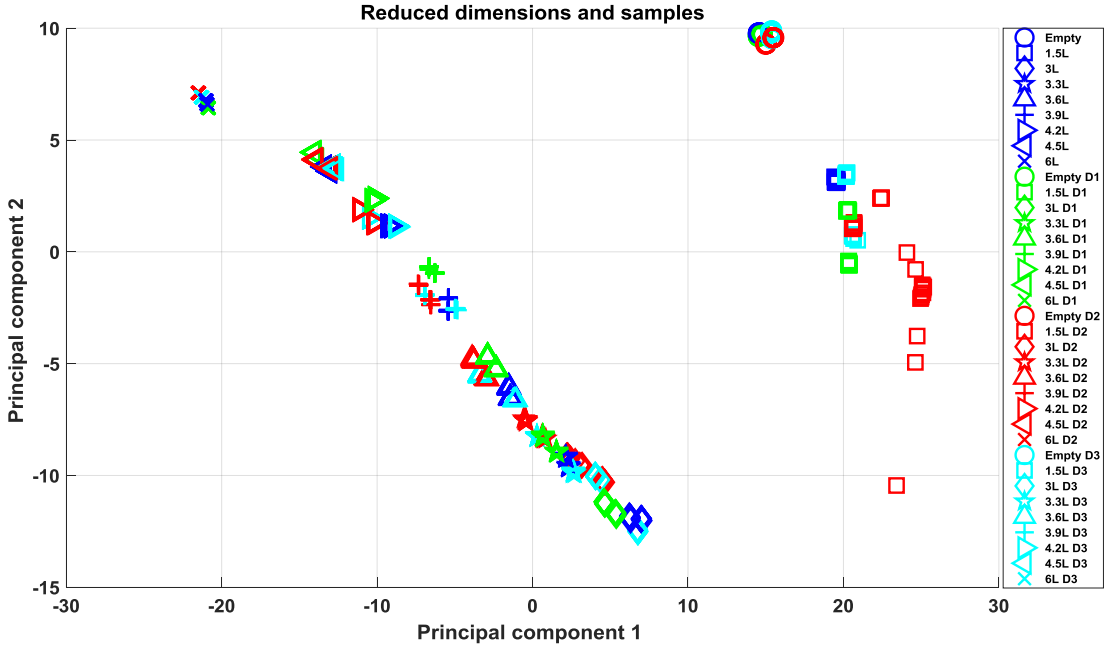


Figure 7.36: The input feature from the standard PCA with reduced number of samples and dimensional size.

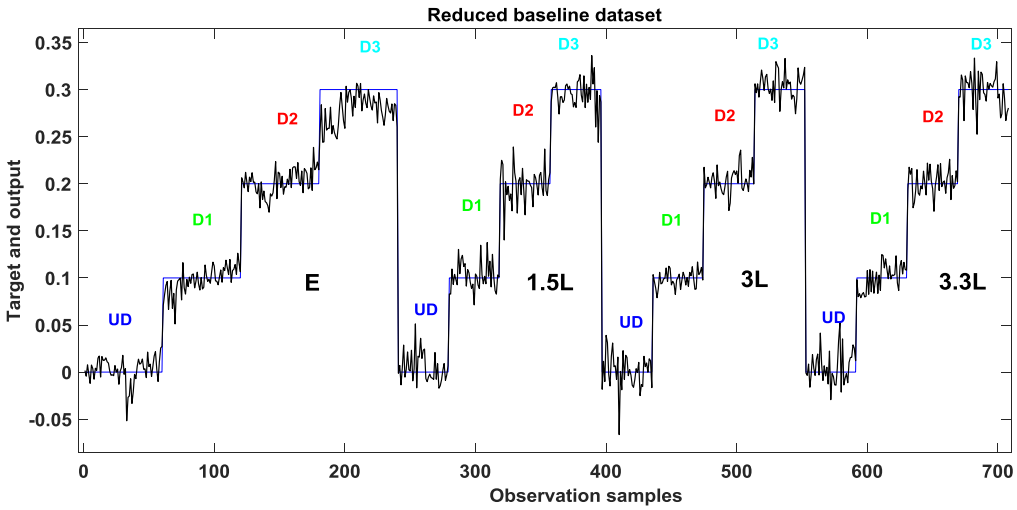
The results of the mapping from the PCA input features to the target vectors for network 2 (reduced samples number) and network 3 (reduced samples number and dimensional

size) are presented on the following page. Each tests comprising 9 loading classes (E to 6 L)in which each class consist of 4 structural conditions (Undamaged, D1, D2, D3). For clarity, the first plot displays 4 of the loading conditions and succeeding 5 loading conditions on the other plot.

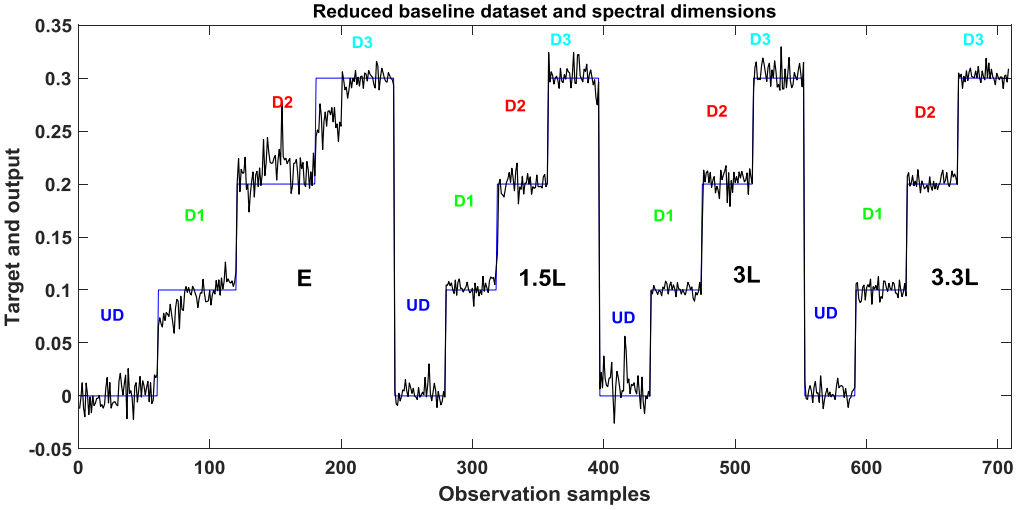
Initial examination of the damage detection and identification based on the mapping of PCA features to the arbitrary stepwise target benchmark produces distinguishable damage severities on each steps with each loading class is highly distinguished (Figure 7.37 – Figure 7.38). The result of using more spectral dimensions (Figure 7.37 (a)) produces slightly better prediction compared to the result in lower spectral dimensions (Figure 7.37 (b)).

The accuracy of mapping the features to the stepwise target for each loading classes with small incremental loading is lightly reduced compared to the classes with larger incremental loading (Figure 7.38(a) and Figure 7.38 (b)).

The calculation of the network performance in terms of MSE for the second network which has double of spectral dimensions compared to network 3 is 6.588e-4. Network 3, which has much smaller number of spectral dimensions, gives slightly lower performance at 6.988e-4.

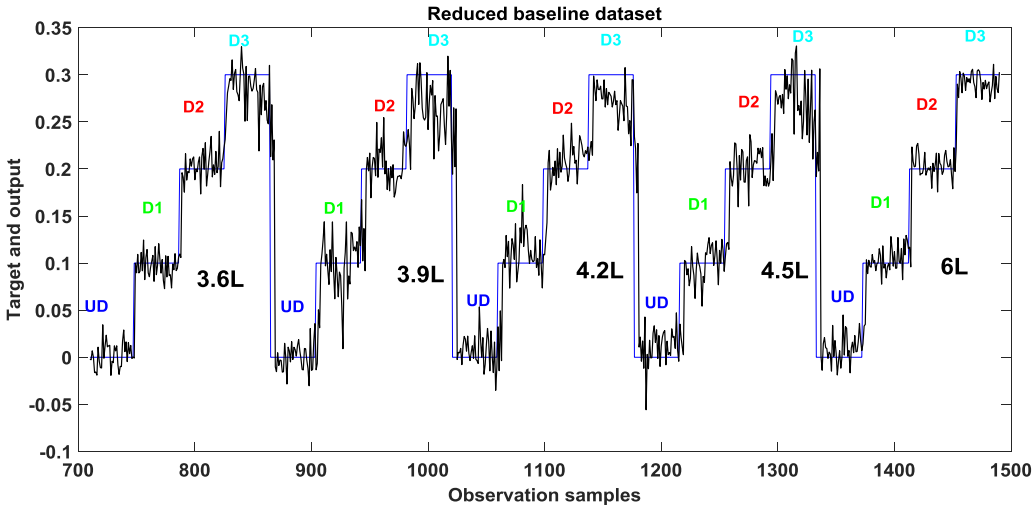


(a) Equal numbers of data observations in each structural conditions (771 spectral dimensions)

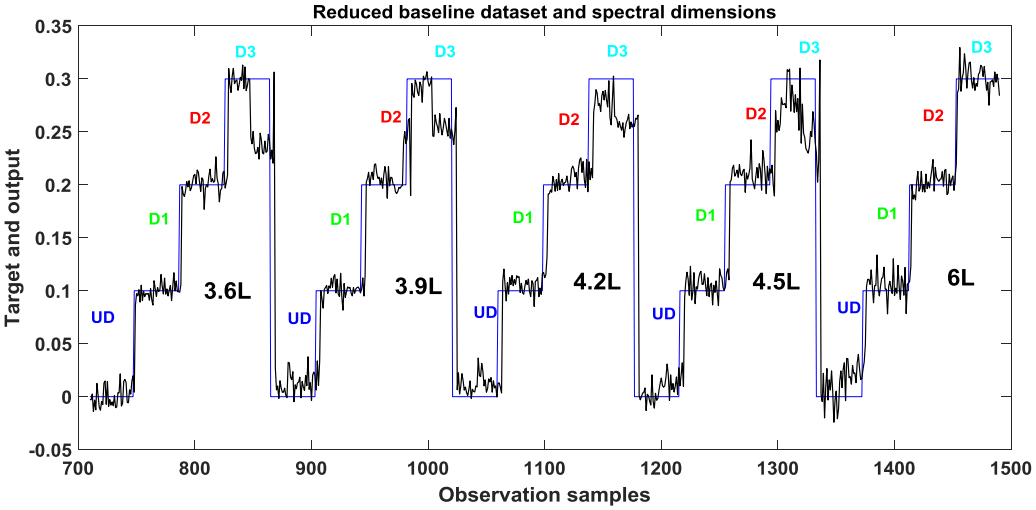


(b) Equal numbers of data observations in each structural conditions and reduced dimensions (301 spectral dimensions)

Figure 7.37: (a) A comparison of the network result when using the inputs of much higher dimensions (230-1000 spectral lines) and (b) the lower dimensions of 500-800 spectral lines.



(a) Equal numbers of samples in each structural conditions with 771 spectral dimensions in smaller incremental loading.



(b) Equal numbers of samples in each structural conditions and reduced dimensions

Figure 7.38: Comparison of the results among smaller incremental loading classes.

The pinnacle of the work has established that the neural network again remarkably distinguish the features of various loading and damage severity classes and mapped them appropriately to their respective stepwise target vectors. Having investigated the effects of data samples between the damaged and undamaged condition of unequal size, it is found

that the size of samples play significant role in improving the neural network performance. In this study, it shows that by having 50% more data samples for training the network can improve the network performance 3 times better using standard MLP feedforward backpropagation with single hidden layer network. Having higher number of spectral dimensions does not necessarily guarantee better network performance. In this case of JW, the network performance with higher dimensional inputs performs only 1.06 times better than the network having less than half of the dimensional inputs. The finding also describes that the MSE is shown to increase for neural network to fit the features for 3.6L , 3.9L, 4.2L and 4.5L loading class due to its small incremental loading. The details of the neural network in comparison their samples size distribution among the baseline set (undamaged set) and the three damage conditions are presented in Table 7.4 as follow. The table also describes a comparison between Network 2 and Network 3 with respect to the size of input dimensions in terms of MSE performance.

Network	Baseline set	Damage1	Damage2	Damage3	Damage sets
1	1120	380	373	380	1133
2	373	373	373	373	1119
3	373	373	373	373	1119

Network	Dimensions	Total observations	Training set (70%)	Testing set (15%)	MSE
1	771	2253	1577	338	1.898e-04.
2	771	1492	1044	224	6.59E-04
3	301	1492	1044	224	6.99E-04

Network	MSE ratio to network 1
1	1
2	3.471
3	3.681

Table 7.7.4: The comparisons between the training and testing sets in their numbers distribution in each network.

As a conclusion, the results summarise the improved generalisation of the network when using higher size of data samples compared to using higher number of spectral dimensions. Having more data samples for training regardless of the distribution size among the undamaged or damaged condition sets can improve significantly the network performance. As a result, the network can learn the system characteristics more details and provide good generalisation and reducing the MSE. Having significantly, higher number of spectral dimensions (more than double) does not guarantee an improved network performance as highlighted by the network 2 and network 3 in Table 7.42. The study successfully provides an effective technique in discriminating the loading effects from damage identification process using only one hidden layer standard feedforward backpropagation network.

7.8 Classification trees

In chapter 6, the study has described GMM as a parametric classification model in predicting the data features of various loading classes associated with undamaged and damaged conditions by establishing clustering onto the mixtures of Gaussian distribution density model. In this chapter, a non-parametric method in classifying multi-class variables performed in a hierarchical and powerful ensemble method of decision trees is presented. The classification tree model described here is based on the technique proposed by Breiman et al. [87]

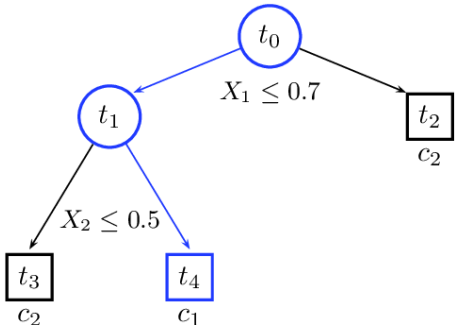


Figure 7.39: A simple decision tree for a binary classification from an input space X_1 and X_2 [88].

In a single decision tree (Figure 7.39), the tree predicts classifications based on two predictors, predictors X_1 and X_2. Beginning at top node (t_0), it examines the values of the variables or predictors to decide which branch to follow by creating splits at first level to reach t_1 and t_2. At the second level (at node t_1), the tree makes another partition to separate t_3 from t_4 and assigned each those as c_2 and c_1. When the branch has reached a leaf node, the data is classified. The tree model develops multilevel clusters in a hierarchical way where clusters at one level are linked to the clusters at the next level.

7.8.1 Prediction trees

In this section, further descriptions on how a decision tree predicts each class for the training samples is described. Prediction trees uses recursive partition as a representation of its splitting of its tree branches. Recursive partitioning can be inferred as the partitioning the data space into smaller regions. The partitioning of the data space is continued and repeated until it reaches a data set that is purer than the parent set so that it can be fitted with a simple model with less complicated feature. In each cell (or leaf), a model is just a constant estimate of a particular class. Suppose the data samples (x_i, y_i), (x_2, y_2), …, (x_n, y_n) are all the samples belong to the leaf k. The model for leaf k is described as just $\hat{y} = \frac{1}{n}\sum_i^n y_i$ which is the sample mean of the response variable in that cell (leaf) k.

In this chapter, a technique known as Bagging and the Random Forest algorithm are discussed as they are interrelated and used in the tree model. Bagging is a technique applied as a step in the Random Forest algorithm. They applied on the JW data sets of reduced dimensions using Matlab Statistics and Machine Learning toolbox under ‘TreeBagger’ function.

7.8.2 Explanation of bagging

Bagging is a technique for generating multiple versions of a predictor and then find an aggregated predictor based on these generated predictors. The aggregation averages over the versions when it predicts a numerical outcome and gives numerous votes when predicting a class. Multiple versions are produced by making bootstraps that replicates the learning sets and making these as new learning sets. The underlying of bagging technique used here is based on Bagging Predictors by Breimen L (1996) [89].

Bootstrap aggregating is the acronym for bagging. It describes the way the resampling is done. Bootstrapping means a type of resampling where large number of smaller samples are withdrawn repeatedly from the original sample set with replacement. Replacement here means the smaller samples are replaced back into the main sample set each time when the smaller sample is withdrawn from the main sample. The bootstrap aggregation of an ensemble of many tree predictors can reduce the effects of overfitting and improves generalization. It grows the decision trees in the ensemble using bootstrap samples extracted randomly from the input data set.

Bagging classification is a superior method in tree predictions category. It works by averaging the noisy and unbiased models so that a new model with low variance is created. The technique works on the random forests algorithm that grows many classification trees and gives the best prediction of the class with the most votes over all the trees in the forest [88].

7.8.3 How Random Forest algorithm works

Having described the Bagging technique which is done as a step in Random Forest algorithm, now a brief description of the algorithm is briefly described. Random Forests are a combination of tree predictors in such that each tree has its own random vectors values sampled independently with the same distribution for all trees in the forest [88]. The bagging technique is done as a step in Random Forest algorithm [88] which works on

a large collection of decorrelated decision trees. The details of the algorithm are presented by the author, Breiman in which the theorems and proof can be found in [88]. Before proceeding to the bagging method, consider the criteria for each tree:

- The number of training data samples are selected randomly from the original sample with replacement and would be the training set for growing the tree.
- The input number of variables m is chosen randomly out of M variables (m < M). The selected variables m determines the split at each node that is based on the best split of the variables m. The number of m is made unchanged throughout the forest growing of many prediction trees.
- Each tree is grown to the deepest level and it involves no pruning.

It is important to note that cross-validation and separating the testing set are all conducting internally in the algorithm. Now the process of bagging of trees is illustrated. Suppose that the training samples S that submitted to the algorithm to build the classification model- the training matrix has feature variables of f_{a1}, f_{b1} and f_{c1} of the first sample up to the N samples. L_1 up to L_N are the training class. These are the feature variables a, b and c in the training class and the aim is to create a random forest in order to classify this sample set. The sample set S is specified as

$$S = \begin{bmatrix} f_{a1} & f_{b1} & f_{c1} & L_1 \\ \vdots & & \vdots & \\ f_{aN} & f_{bN} & f_{cN} & L_N \end{bmatrix} \tag{8.1}$$

From the sample set, Random Forests create individual random subsets and establish its own decision trees for each subset. All the samples are used for the training and one third of the samples would be retained for testing the tree.

It is important to note that, the choice of subsets variables are randomly selected from the sample sets so that subsets would have many variation in the decision trees. The decisions trees are then ranked as explained next. The process can be described as follows

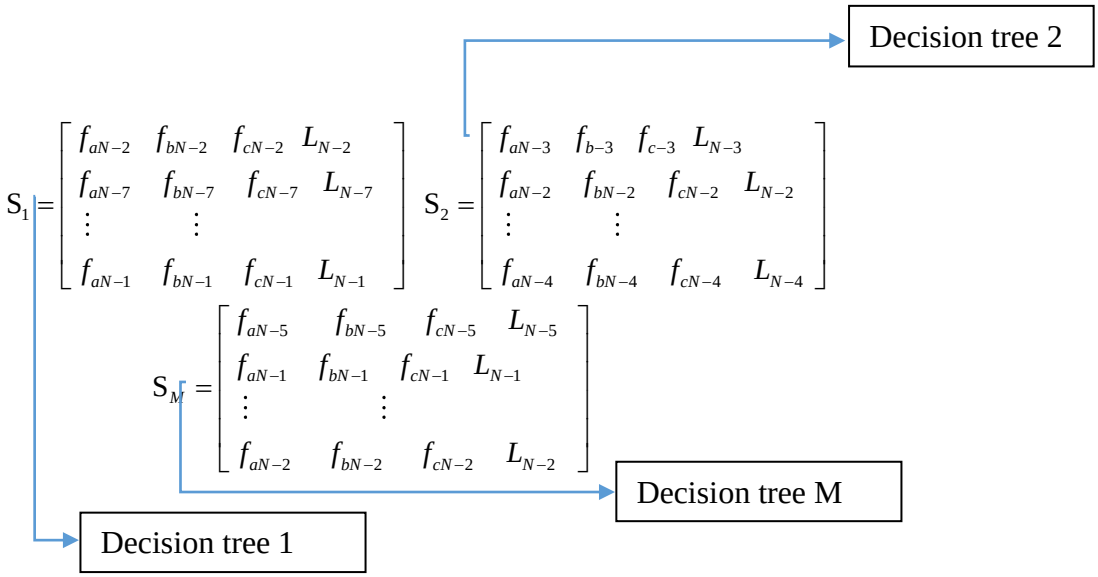


Figure 7.40: An illustration that shows how the random forests is created

Note that each subset S_1 up to S_M has different values from the sample set and would have its own decision tree. As a result, the decision trees of the subsets would have different classification variation. All the decision trees are then used to create the ranking of the classifier. The decision tree that receives the highest vote is the most popular classifier.

To describe this process more clearly, the linear PCA dataset from empty fuel tank load data set, E from the JW is provided with its corresponding categorical labels of E(N), E(D1), E(D2) and E(D3) for the training samples and implement the steps as described in the Random Forests algorithm (Figure 7.40).

The bagging of trees classification is performed based on Matlab Statistics and Machine learning toolbox utilising Random Forests algorithm [90]. The inputs comprise the principal component 1 and 2 and their corresponding labels set. Via the bagged classification technique and the Random Forest algorithm, the decision trees are selected based on the number of times the tree classifier is described by the algorithm or referred as the tree that receives the highest vote (most popular).

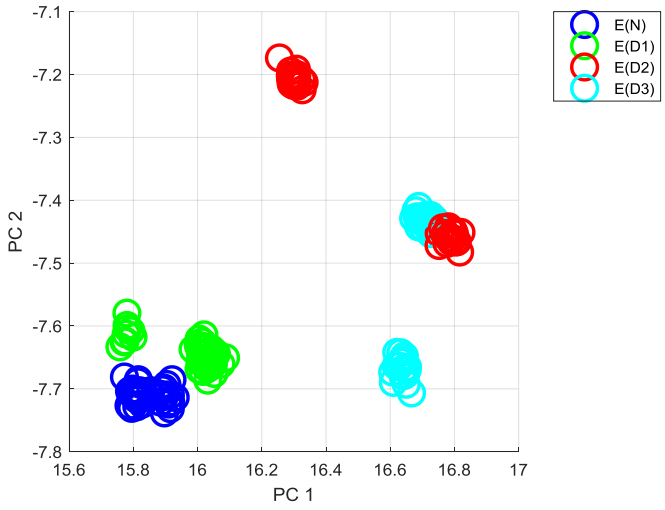
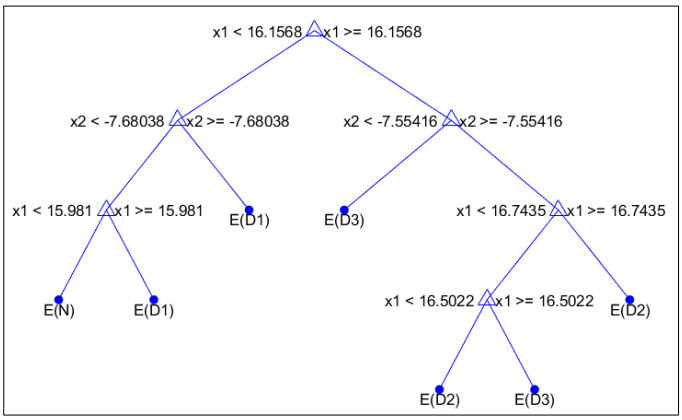


Figure 7.41: The training data samples of which the empty tank load is used as the training sample.

Figure 7.42 provides the ranking results of the bagged classification trees obtained using the bootstrap-aggregated decision trees via the random forest algorithm as described previously. The class 1 is the most popular tree classifier for the sample set and shall be used as the classifier model for the structure of the similar characteristic.



Class 1

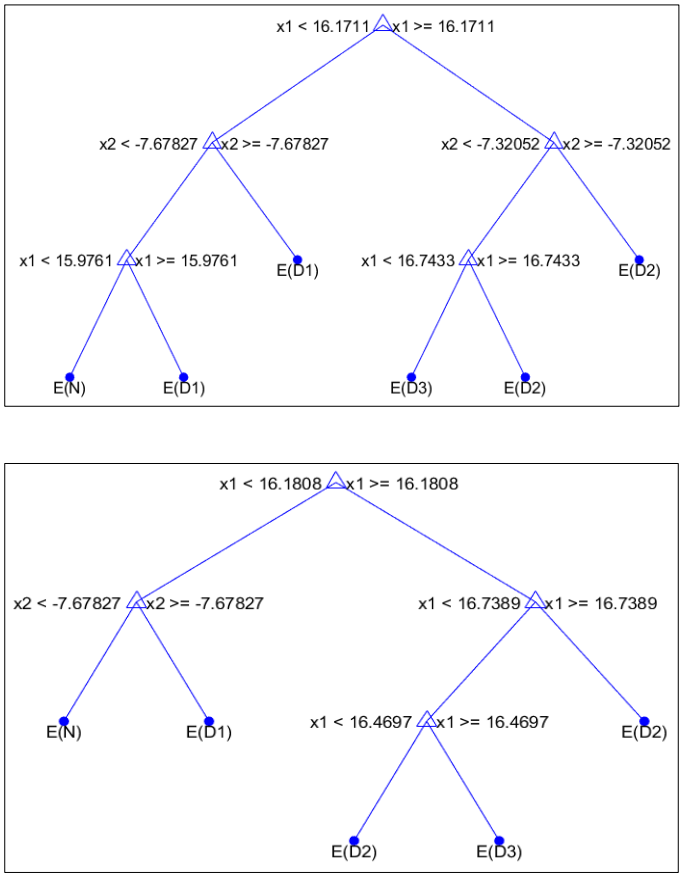


Figure 7.42: These decisions trees are classified and ranked with the highest vote for the most popular class at the input x1 and x2.

Random Forest algorithm classifies Class 1 as the most likely predicted models followed by class 2 and class 3. Noticed that in Figure 7.42, the first prediction tree, which has the highest votes, has seven leaves nodes and that is equivalent to the number damage clusters in the original PCA data space. Analysing the prediction tree of Class 1, the split of the damage groups and prediction of the damage classes are very excellent in reference to the actual PCA data in Figure 7.41.

Some interpretations can also be derived from the random forest classifier in terms of the proximities between the splits that represent the correlation of each class variables. The most popular classifier of class 1, based on the proximity of their splits, variable E(N) on the bottom left corner is more correlated to the E(D1) due to their same split. Leaves that are most far from each other and on opposite branch are very decorrelated. These are the

four variables class belonged to E(D2) and E(D3) class located on the right side of the tree branch separated by the main branch of the first classifier.

7.8.4 The JW data set on bagged classification model

The data samples obtained from the linear PCA from the JW experiment applies Random Forest algorithm (Figure 7.43) for tree classification. Spanning across the PCA data space lies a region of high data density which involves classes of small incremental loading. Before the bagged trees classification is applied, a simple partitioning of the data groups into subgroups is described as what the prediction tree is inspired and motivates the binary decision trees approach. The divisions represent the split of the nodes which is carried out automatically with the Random Forest algorithm.

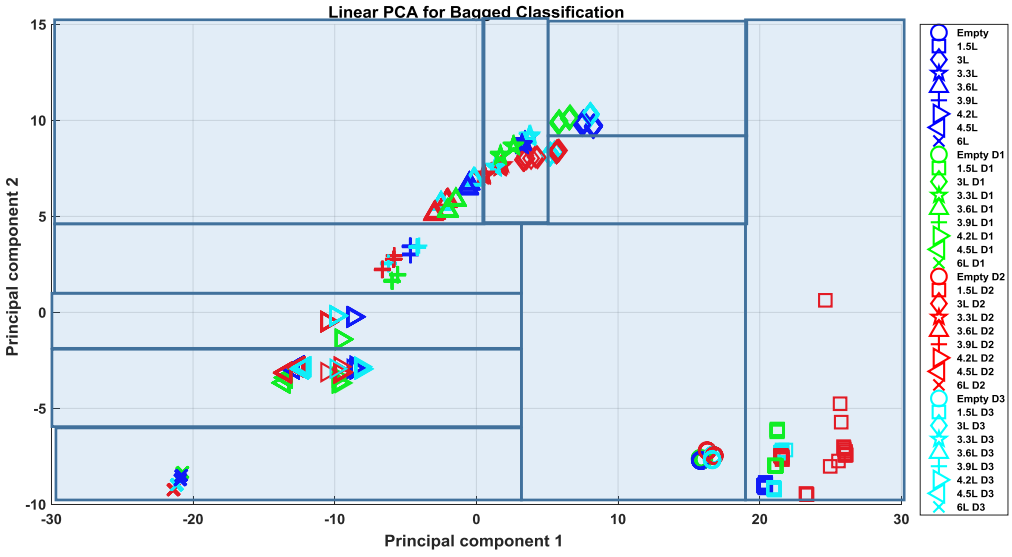


Figure 7.43: Simple partitions of the standard PCA scores of the JW

Figure 7.43 demonstrates the use of partitions to classify damage groups of various loading conditions with restriction on dividing the groups in deeper level which requires higher resolution.

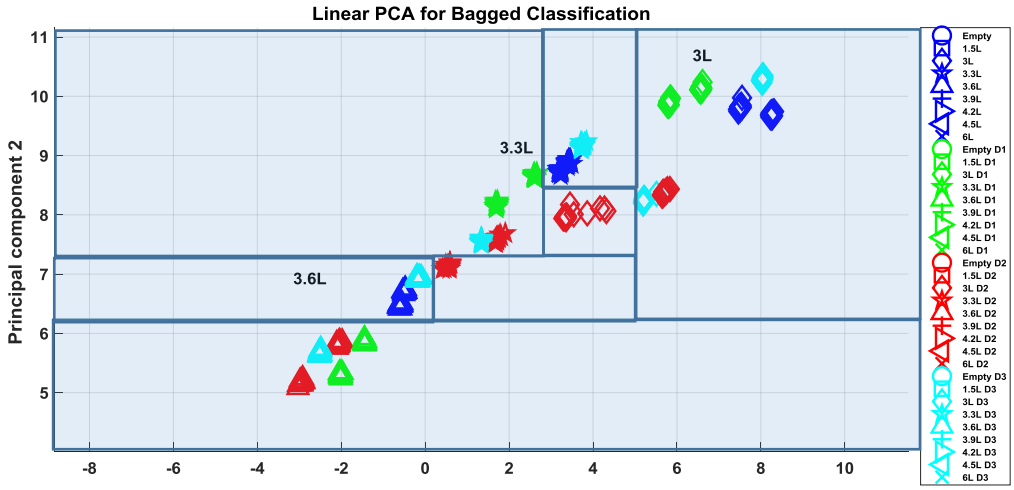


Figure 7.44: A zoomed view into the regions of concentrated data points

Figure 7.44 shows the subdivisions of damage class is implemented automatically using the Random Forest algorithm. The advantage of using bagged classification trees in the JW data set is related to that the nature of the data set which is hierarchical. This translates into meaning that the loading classes should be in the upper nodes while damage level states assume the nodes in the leaves (the deepest nodes).

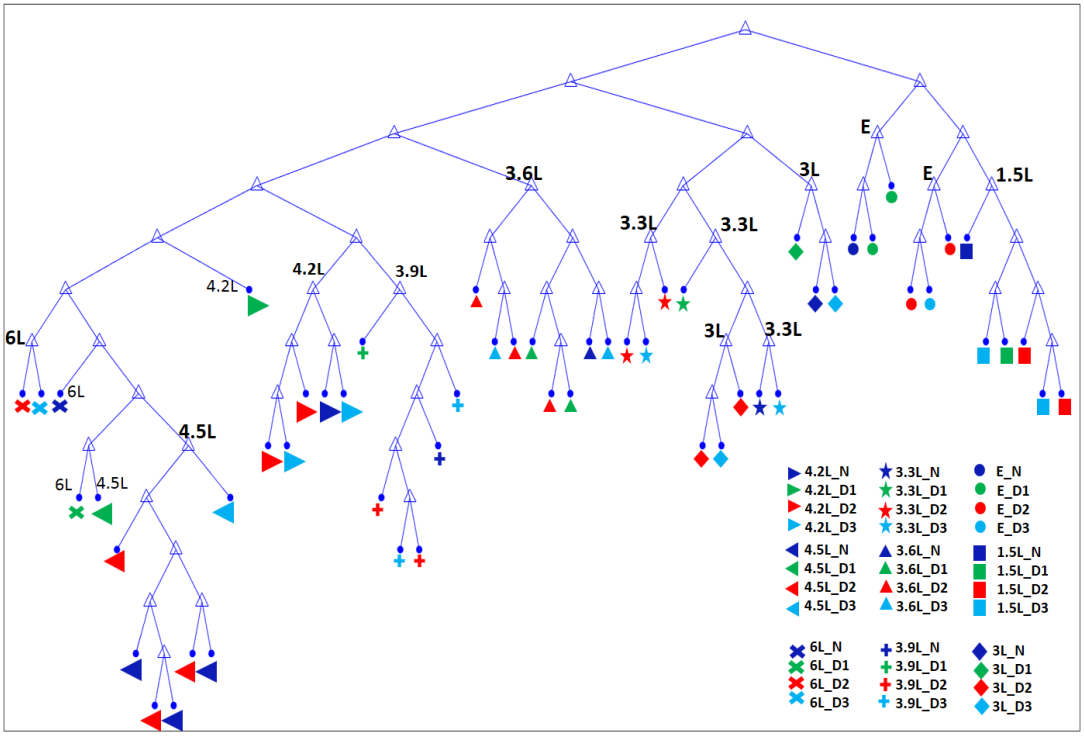


Figure 7.45: A classification tree using overall PCA data sets of the JW.

Figure 7.45 shows the complete JW data set obtained from the standard PCA where the branches make up the loading sets. The final split of the branches (no more division partitioning) reach the damage class that lie in the sub-divisions of the data space and find the data purer than the upper node.

The classification tree derived from the random forest algorithm has revealed some interesting findings that the GMM classifier does not address. This result of Figure 7.45 creates 55 splits in the tree classifier. In total, there are 4 structural conditions associated with 9 loading classes which make up to 36 classes over all data points. The main advantage of this prediction tree is that the leaf nodes represent single damage class thus eliminate the problem of overlapping data sets. The prediction tree describes the relationship of each data class as shown by the proximities of the loading and damage variables. The bagging method effectively produce many deeper trees naturally, which represents damage severities (indicated by colour shapes). The prediction of the damage severities classes are excellent which makes the classification using prediction trees very

promising. Another advantage derived from this classification tree is the branches of the tree also follow the direction and systematic pattern as displayed by the linear PCA data sets in Figure 7.36. The other branch leads to portioning of divisions that represent the other loadings. In addition to that, it is observed that, the sequence of the division is in agreement with the loading sequence in the PCA data space.

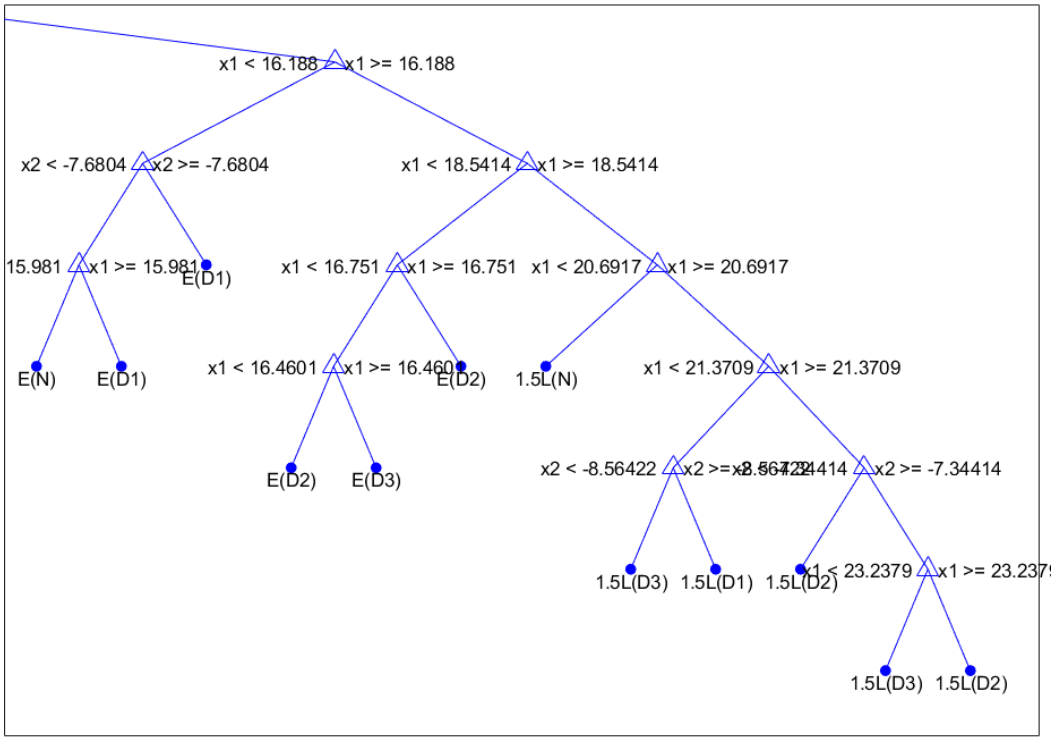


Figure 7.46: The highlighted part of the lowest branch on the right side of the tree.

The split is done by aggregating the binary values the principal component 1or principal component 2 to reach the data sets that are purer compared to their parent data sets of previous node. To explain further on prediction trees in Figure 7.46, say the process begins at terminal node (top node) by simply gives a sequence of binary questions related to the features position with respect to predictor x1 and x2. For instance, for data point, x is > x1==16.188 or x < x1==16.188, the next branch leads to the next partition and it asks another question until x falls in the corresponding cell of the right partition. The question asked depends on the previous question and the process of asking binary question

continues until the data points in the partition is pure, which now indicates the group belongs to the right damage class. The next Figure 7.47 highlights the most left and lowest branch of the same decision tree, validated by the PCA data scores of the 4.5L loading class.

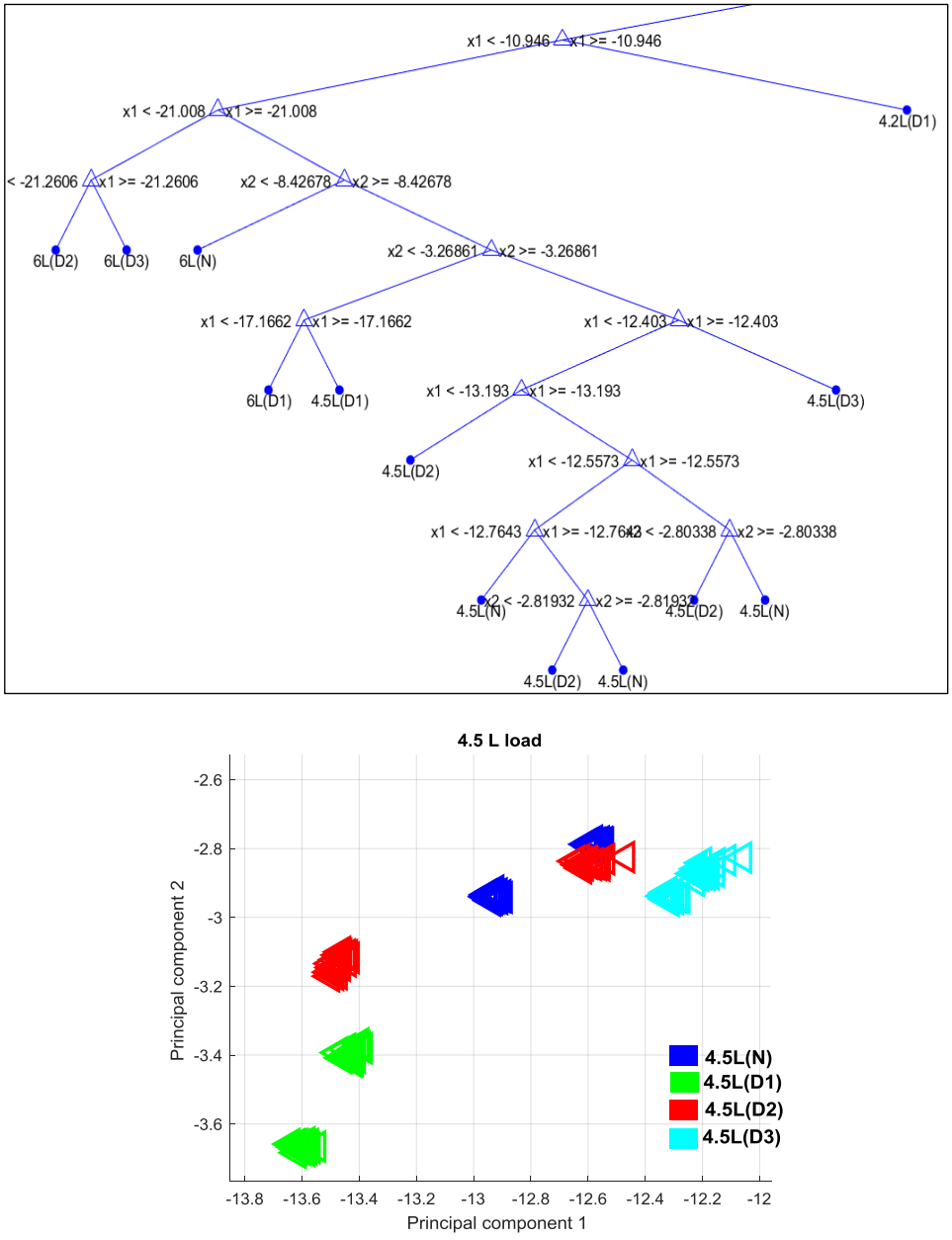


Figure 7.47: Validating the tree classifier results with the zoom-in PCA data plot into 4.5L data class.

It focuses the left side of tree branches of the classifier over all data sets and it reveals an accurate damage class prediction through their leaf nodes representations. Next plot shows how the bagged classification performance can be evaluated.

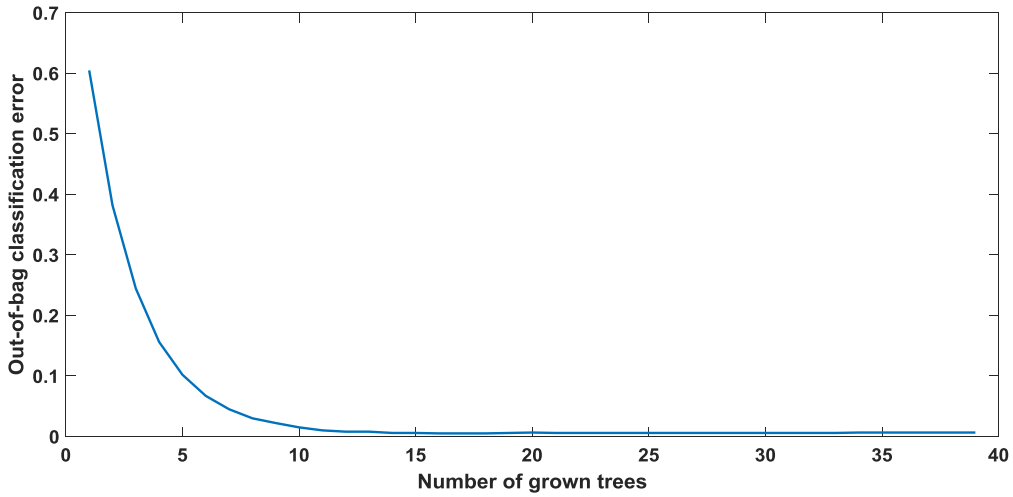


Figure 7.48: It shown that the out-of-bag error decreases with the increase number of grown trees.

Out-of-bag error estimates the data samples that are out of bag for each tree. The result in Figure 7.48 shows that accuracy of the tree prediction increases when using more trees in the forest. In JW prediction of ensemble trees, 39 of grown trees are specified in 'Bagged Tree' classification function in Matlab Statistical Machine Learning Classification. This performance validation for the bagged classification tree proves to be highly accurate. The performance corresponds to the overall JW tree ensemble in Figure 7.45. This concludes that using many trees in ensemble trees prediction, there is good prospect for the variables to be correctly classified and become unbiased in many tests with many trees.

7.9 Conclusions

In this chapter, the FRF measured from the JW aircraft was analysed and discussed. The response shows that the structure exhibits global mode and local mode of vibration. The local mode of vibration provides an advantage in detecting damage and therefore it is used as a potential feature for machine learning application.

Applying the kernel Gaussian PCA using parameterised Gaussian function locally, improves the data separation in principled manner. Locally here means that the parameter for the smallest distance is computed at each level of structural conditions of the data sets consisting of all loading conditions. In fact, this is an advantage because the training or the reference set are established from undamaged condition only comprising all loading conditions. Having new test data produced by different structural conditions, the test set can compare to the reference set if the data deviates significantly from the threshold line.

The standard PCA has shown to be robust by means of finding and extracting an appropriate feature of high dimensional data sets. Coupled with Q-statistic and T-squared, it can compose a comprehensive technique in evaluating the data variability including for the damage detection. In fact, PCA as a powerful non-supervised technique, that can be used for finding the classes of high variability (loading variables). Incorporated with the kernel Gaussian PCA, nonlinear projection of data with small variability benign the damage class can be discovered with distinction and remove overlap data.

In neural network regression analysis using JW data sets, it is shown that the MSE performance is not significantly difference from the high dimensional data compared to low dimensional data of the same feature. Having different size of data sets for baseline or damage sets do not necessarily effect the overall performance of the neural network. The more important criteria is to have more more data samples for the network training so that

the network can learn adequate data patterns before testing of new data set. In terms of dimensional size, reducing the number of data dimensions does not affect the network performance considerably compared to if reducing the number of data samples. For large data sets such as the JW, training function of Bayesian framework is not recommended to use as it assumes high computational resources and significant processing time and brings no improved results.

In the nonlinear classification using prediction tree, the real advantage observed is that algorithm based on Random Forest works exceptionally well for large data sets with no problems about the curse of dimensionalities. Despite a supervised learning mode, prediction trees has a high advantage due to its simplicity for being utilised in the SHM if incorporated with the PCA. PCA works on an unsupervised mode, has shown to be effective in discriminating the loading effects from the smaller damage effects.

By using the bootstrap aggregation in making many replicates of training sets using its samples and building many decision trees in the random forest algorithm, the most popular tree is selected. The result from this work has shown to produce a prediction tree as a damage classifier with very small classification error. Another advantage of the classification tree- the splits of the nodes can go naturally deep to find the purest data set that is the true class of the data variable with no issues of class overlapping

It is a simple binary classification but can produce remarkable classification especially for large data sets in which have different level of data variability. In fact, bagged classification trees has the natural capability to classify the data hierarchically, creating the variables of the highest variance take up the upper branch and the variables with the smallest variance occupy the deepest leaves or be at the lowest branch in the prediction tree.

Chapter 8

CONCLUSIONS AND FUTURE WORK

The main contribution from this work is centered on the visualisation techniques via KPCA projection plots and novelty detection via outlier analysis using KPCA data sets. The advantage of using KPCA can be seen from the ability to adapt the parameter (σ) locally according to the each test sets. Another secondary contribution comes from the application of ANN (for establishing the mapping between linear PCA input data and the target output vectors), GMM (modelling the KPCA data set according Gaussian distribution density model) and Bagged decision trees (as a hierarchical classification and prediction tree based model). This thesis has outlined several main methods and procedures of damage detection under the effects of operational loading conditions. These can be divided into six main categories:

- Feature extraction and dimensional reduction using a standard PCA
- Generalised nonlinear feature extraction using kernel PCA
- Damage novelty detection via statistical model of T^2-statistic
- Probabilistic classifier model using GMM
- Nonlinear mapping of PCA features to binary stepwise target benchmark via ANN

- Hierarchical model of prediction tree model using bagging and Random Forest algorithm

The current work has highlighted its attempt to discriminate the features developed by operational loading variations from those caused by damage. In this context, the above procedures were performed on the data sets derived from the aircraft wing which underwent incremental loading inside its fuel tank and was also exposed to various damage severities. The aim of this thesis is to unravel underlying patterns in this experimental data and provide insight into the effects of loading on damage features.

8.1 Discriminating the loading effects by using PCA

The first evaluation obtained from the PCA reveals that the effects of operational loading can be discriminated from damage-sensitive features. PCA was found to be capable of producing positive results in isolating loading effects from the vibration response and causing the damage features to overshadow the loading effects. The results of the PCA have indicated that the data signal processed with PCA can be transformed as a combination of new variables and projected according to a particular variable group associated with each loading class. In the case of the wing-box test, which involves large variance in loading changes, PCA can effectively isolate all features of damage severity from the influence of more dominant loading effects.

On the contrary, for the Jabiru wing (JW) vibration-based damage detection (VBDD) test, loading classes with smaller incremental loading are introduced. The problem arises when the feature caused by damage and small loading increments are transformed on the region space which does not occur in the case of higher incremental loading classes. Without utilising machine learning algorithm, the damage feature might be mistaken and interpreted as undamaged data belonging to a higher loading class (Figure 7.25). In this context, kernel PCA, ANN and ensemble of prediction tree show unparalleled and outstanding result within their own framework.

In the context of data pre-processing for the FRF response of the JW, data cleansing of the data set is thought to be necessary to remove extreme outliers. These outliers are caused by some changes of fixing conditions of the stringer's height due to the significant changes of the structural weight and the possibility of water sloshing effects in the tank during the test. This is seen as an essential step before the novelty test and other machine learning algorithms are used on the data set to prepare for quality data sets. The phrase 'garbage in, garbage out' is particularly applicable at this stage.

As an advantage of supervised learning, it effectively establishes various PC models based on different data arrangements described in the covariance matrix. The aim is to isolate the effects of operational loading variations, so that the score variables (PC projections) can manifest signs of damage severities in the feature space.

The study has proven that standard PCA exhibits some limitations on both novelty detection and visualisation particularly for small damage severity class. On a positive note, due to its capability in revealing the data sets with the highest variance and to work even under unsupervised learning, standard PCA has been used intensively in the earlier part of the work. The result suggests that PCA can effectively serve as the first data investigative tool to acquire the underlying structure of the data before attempting to use more advanced machine learning algorithms.

8.2 Kernel Gaussian PCA as a nonlinear feature extractor

In addition to the standard PCA as a tool in discriminating the loading effects from damage signs on the features, a kernel Gaussian PCA is utilised. Kernel PCA is applied as a second step tool as a nonlinear form of PCA to extract the hidden features located in high dimensional space, which is restricted in the linear PCA projections. In this context, the technique has been more accurate in separating two classes of different structural conditions compared to the linear PCA, with less parameterisations required in comparison to other nonlinear PCA

techniques such as the auto-associative neural network. The results have shown to be very encouraging with respect to data visualisation and novelty detection (T-squared analysis) in comparison to standard PCA where problems related to data overlapping are normally encountered. Having positive outcomes from the kernel PCA, the technique requires a selection of an important hyper parameter sigma (σ) that is used as an inverse variance ($1/\sigma^2$) in the nonlinear mapping transformation function based on the kernel Gaussian or Radial Based Function (RBF). The work has established a novel way of selecting the parameter using a distance matrix colour-scaling plot for the purpose of assisting the user to visualise the distance between the data class. Based on Euclidean distance matrix, the inverse variance can be computed on the basis that the sigma parameter value (σ) should be smaller than the average inter-class data distance and larger than the average of inner class data distance. Another main benefit of using kernel Gaussian PCA, besides improving data separation between undamaged condition and damaged condition classes, is the formation of full data trajectory. Having established a full data trajectory facilitates a monitoring of the structural condition changes through its nonlinear form of feature patterns produced by the nonlinear kernel function.

The superiority of the kernel Gaussian PCA lies on its flexibility of the kernel that is introduced locally at each test data set and adapted to the variability of each set. The positive indication from this is that the variability trend, as the loading switches from one loading class to the next, is more evident and viable for damage detection and prediction model (Figure 7.29).

8.3 Underlying multivariate data phenomenon due to incremental loading

Under substantial loading variations, the study has discovered that the load-sensitive features are more apparent on the lower region of the frequency spectrum while damage-sensitive features are more prominent and localised on the higher range of the FRF data. This discovery is in agreement with the basic premise of VBDD that the damage is typically a local

phenomenon and the substantial alteration of the structural stiffness, mass and energy dissipation generally affects the global dynamic properties of the system [1].

The first essential step involved in data pre-processing is to identify and select the feature that exhibits underlying data information that best describes the changes in structural health condition. With respect to JW data, more challenges were observed as a result of smaller incremental loading, which produced the features and similar effects as those caused by damage. In addition, more attention is necessary to remove any significant outliers which can be validated using Q-statistic test. Failure in handling those extreme outliers can potentially result in false damage detection. Kernel PCA has performed effectively to overcome such problems and making the damage-sensitive features more separable from those loading-sensitive features.

The major discovery with respect to data characteristics under the influence of loading and damage variations is the amount of the incremental loading is a determined factor in the ability to detect the presence of damage in the PCA processed data. For instance, a small incremental loading in the system produced a similar effect of damage in the PCA projection, hence leaving both effects of damage and loading on the same region in the feature space. In view of dealing with this issue, kernel PCA has shown to be able to mask the damage effects from the loading features effectively in the feature space (Figure 7.26).

8.4 GMM as a predictive model in SHM with loading variation influence

In this study, Gaussian mixture model (GMM) is implemented with the aim of verifying the technique to classify the data groups according to their true damage conditions in the probability framework. The results of this study have shown that the performance of algorithm is sensitive to the initial values for the GMM parameters (mean and covariance and number of components). The study has used a training data set (undamaged state under all loading

conditions) in order to establish a baseline set as a reference for further test data sets (damaged states under all loading conditions) undergoing similar operational loading conditions.

It has to be stressed here about the issue of singularity of the covariance that may occur during the iteration of the EM algorithm. This case happens when the data points are produced by a small change of damage severity between different groups or when the data points are highly correlated among different data classes. Despite this challenge, Gaussian mixture probability models can potentially be used to provide a good predictive model by manipulating the parameter values related to the GMMs. In other words, the means and the ellipsoids of the corresponding GMMs can be used to validate if the new test set can be inferred with damage. GMM establishes clustering naturally over all data group densities that is most likely to be generated by the particular Gaussian distribution models. The advantage is that the data sets that assume and fit very well to the Gaussian distribution model will result in effective GMM results.

On the other hand, GMM technique can provide good data classifications for the purpose of damage severity identification even if the data labels are not available. Using k-mean algorithm to calculate the initial conditions, the algorithm naturally predicts each data point that is likely to be generated by a specified Gaussian distribution model (Figure 6.19). GMM can be used to generate a predictive probabilistic model based on the data features acquired from the kernel PCA. It can also potentially be established as a generative model, generating a similar GMM for the baseline data set if the baseline set is not available. In view of SHM, a comparison of the baseline feature of an undamaged condition to a new test set can be carried out to determine any departed features from the baseline set which can indicate the possibility of damage present in the system.

8.5 Neural network regression for damage prediction model under loading variables

In Chapters 6 and 7, neural network regression was performed on the multivariate data set. In fact, ANN forms one of the backbone in this study with the aim of producing an effective damage detection framework. This nonlinear regression method of ANN is established byfeed-forward MLP using only a single hidden layer network. Despite using only a single hidden layer, the result has shown to be effective in predicting damage and loading levels respectively. Using only standard PCA data scores as the network input, the neural network successfully map them to their respective target vectors. These target vectors are simply arbitrary numbers determined by the number of damage severities and loading classesconsidered in the network input. Again, ANN requires data labels to be known in order to establish successful mapping of input to target values.

The main concern of using ANN is that it requires a process of selecting an appropriate number of hidden layer networks, optimum number of hidden nodes, suitable training functions for the network training, and that it may require some regularisation before one can reach the optimal solution for the network. It should be noted that it is recommended practice to start training the network with the most simple network architecture and adjust the network size if the results are not satisfactory [34], [77].

Another important finding is the comparison between maximum likelihood and Bayesian approach in calculating the network parameters and fitting the data output on the targets, which shows that the Bayesian approach can potentially perform better than the maximum likelihood approach based on the LM training function given the size of the data samples are not large. The study using the JW data sets reveals that the performance of the network is mostly affected by the number of data samples used in the training rather than the number of spectral dimensions and the size distribution among the data categories (undamaged and damaged classes) (Table 7.4).

8.6 Random forest model as hierarchical damage classifier

Finally, Random Forest classification is tested on the multivariate loading and damage data set to discover the capability of the Random Forest algorithm to predict the damage classes over all the different loading conditions. In this work, bagged classification trees performed by the Random Forest algorithm has represented an impressive method in predicting the damage variables under the effects of loading variables that require no parameterisation. The splits of the nodes can go naturally deep to find the pure data set that signifies zero data overlap between damage classes. It is a simple binary classification technique but is able to work very well with large data sets without any issues pertaining to singularity. The results from the prediction tree have indicated a good correlation with the linear PCA data sets.

The model provides a greater extent of prediction and visualisation performance of damage class through the distance between their splits. The overall structure of the prediction tree represents a more convenient way with better interpretability for large data variables that involve different damage classes associated with different loading conditions in comparison to a direct visualisation from the PCA score plot. Bagged classification trees have the capability to enhance the data classification and work better with larger data sets. Essentially, classification tree models have a natural way of classifying the data variables hierarchically, whereby the variables with the highest variance assume the upper branch and the variables with the smallest variance occupy the deepest leaves nodes in the prediction tree.

8.7 Future work

In view of the GMM approach, some restrictions are observed while performing the technique, including the need to specify the number of Gaussian components before performing the EM algorithm, in addition to the singularity problem that can exist. One approach of overcoming this limitation is by introducing a Bayesian treatment in the framework of variational

inference. Bishop et al. has explained this preferred approach in using Gaussian mixture classification model in comparison to the present maximum likelihood estimation (MLE) [34].

A generalised nonlinear PCA implemented via kernel method has shown to improve the separation between the classes of damage severity. The sigma selection for the inverse variance in the kernel Gaussian function represents the essential parameter in order to obtain a successful kernel PCA transformation model. In relation to the current work, the sigma selection is performed manually through the distance matrix. In this context, it can be more convenient and comprehensive if the selection process can be conducted in an iterative and automatic manner. This will help the user to choose the inverse variance that is validated against each test set and the baseline set and rank the value according to the T-square index. The EM algorithm is ideal for this purpose as it involves the Expectation step (calculate the Euclidean distance between data class) and Maximisation step (maximise the distance between baseline and damage class) in the MLE approach.

It will be beneficial to compare PCA with other dimensional reduction techniques such as the Fisher's linear discriminant. The technique is not far different from PCA, only in that this linear classification model takes account of the class labels to provide much better class separation. Thus, it assumes a supervised learning mode compared to PCA that can use both supervised and non-supervised learning criteria.

An improvement for this study should consider the removal of external outliers using MCD which has shown positive SHM result in real bridges [28]. The result obtained from the MSD computation using the KPCA variables show there are Type I errors that indicate a false positive error on the T-squared index plot (Figure 7.29) even after introducing Q-statistic analysis as an effort to remove outliers from data set.

On the side of data instrumentation, currently piezo-electric sensors have been utilised to capture the vibration response and measured experimental data in the FRF domain. It will be promising to use other SHM based techniques such as the Guided wave due to its high damage sensitivity and long monitoring range [1], [91][92][93]. From the aspect of practicality, it will be promising to adopt an active sensing system particularly for detecting many types of damage on metal and composite including cracks, corrosion, delamination etc. [94]. This

active sensing involves an excitation signal being sent to an actuator and other sensors picking up the structure response from the excitation signal which potentially works well on the composite surface of an aircraft wing [91], [93] .

In terms of incorporating a robust SHM damage detection technology for an aircraft structure, a more comprehensive understanding of all EOVs pertaining to an inflight aircraft is vital. The current work is concerned with the operational loading variation due to fuel tank loading.

To achieve a more robust SHM system, the next step would be to consider another variability such as temperature variation in addition to the present operational loading variabilities. In the context of practical application for operating aircrafts, the sensing system has to be extremely durable to withstand the harsh operating environment of an operating inflight aircraft. Realistically it is expected to be more challenging and costly, but essentially may bring the realisation of VBDD based SHM into practical application for operating inflight aircrafts one step closer.

Appendix A

The kernel PCA Algorithm for the SHM

Kernel PCA is the nonlinear form of PCA which better exploits the hidden and overlapped data features of high dimensional features space. In this appendix, the actual algorithm used in Matlab adapted into multivariate loading and damage data sets is presented. The purpose is to demonstrate to the reader of how the technique of selecting appropriate parameter sigma with respect to the SHM application is done. The selection is based whether it is for monitoring (visualisation) or damage detection (outlier analysis). This algorithm is inspired from the work by[55] that is used successfully to construct human face expression models. This algorithm has been adapted to the current work concerning the SHM of the wing box structure under the changing of operational mass loading. It can be used directly on Matlab software after giving the input data.

==========CONSTRUCTING KERNAL PCA FOR SHM==========

%For kernel PCA, two data sets are used to compute the Eucludean distance,

% undamaged and test set

%Require inverse variance (sigma parameter squared)

```
%--------baseline set of all joined undamaged loading conditions---------
Normal_baseline_dataset=[training_data_empty;training_data_quarter_tank;...

training_data_half_tank;training_data_threeQ_tank; training_data_full_tank];

%centralised and stadardised baseline set
mean_BaselineData = mean(Normal_baseline_dataset);
rep_meanBaselineData= repmat(mean_BaselineData,200,1);
centered_baseline=Normal_baseline_dataset- rep_meanBaselineData;
std_train=std(Normal_baseline_dataset);
std_trainData=zscore(Normal_baseline_dataset);
%establish kernel matrix X
%compute the covariance of the Gaussian kernel matrix K for the baseline
% X= (Normal_baseline_dataset-rep_meanBaselineData);
X=std_trainData;
N=size(X,1);

%based on Eucludian distance, calculate the distance matrix, D of data set
XX=sum(X.*X,2);
XX1=repmat(XX,1,N);
XX2=repmat(XX',N,1);

D=XX1+XX2-2*(X*X');
D(D<0)=0;
D=sqrt(D);
D(D==0)=inf; %excluding zero values
D(D<1e-6)=inf; %excluding very small values
D_sort=sort(D); %arrange the distance in ascending order
D_2ndminBlock=D_sort(41:80,:);%consider distance block from
%the second smallest distance

D_2nd_min=min(D_2ndminBlock);%find the minimum distance from the block
```

```
D_2nd_mean=mean(D_2nd_min);%find the average
para=D_2nd_mean %this is parameter sigma
%compute the kernel matrix k:
%squared distance and inverse variance (sigma squared)
 K=D.^2;
K0=exp(-K./(2*para.^2));

%centering the feature space for Sum(transformation(Xk))=0
N=size(X,1);
oneN=ones(N,N)/N;
K_norm=K0-oneN*K0-K0*oneN+oneN*K0*oneN;

%% eigenvalue analysis
[V,E]=eig(K_norm/N);
eigValue=diag(E);
[~,IX]=sort(eigValue,'descend');
eigVector=V(:,IX);
eigValue=eigValue(IX);
norm_eigVector=sqrt(sum(eigVector.^2));
eigVector=eigVector./repmat(norm_eigVector,size(eigVector,1),1);

% dimensionality reduction
% specify number of eigenvectors (number of Principal Components)
eigVector=eigVector(:,1:100);

%calculate the PC variance (latent)
eigValue=eigValue(1:101,1);
latent=eigValue(1:50,1);
latent_baselineA1=100*eigValue(1,1)/sum(eigValue(:,1))
latent_baselineA2=100*eigValue(2,1)/sum(eigValue(:,1))
latent_baselineA3=100*eigValue(3,1)/sum(eigValue(:,1))
```

```
%project the principal components on the high dimensional space
Y3 = K_norm*eigVector;
%plot 3D dimensional space of the KPCA projections

%------------------Damage 1 data set--------------------------------------

%construct test set 1 using all loading conditions corresponding to D1

test_D1_dataset = [testing_data_empty_D1;test_data_D1_quarter_tank;...
test_data_D1_half_tank; test_data_D1_threeQ_tank;test_data_D1_full_tank];

%standardised the test set 1 agaist the baseline set
test_D1_norm=(test_D1_dataset) - (rep_meanBaselineData);
std_testD1=test_D1_norm./std_train;
X_D1= std_testD1;
N_D1=size(X_D1,1);

%calculate Eucludean distance between baseline and test set 1
XXn=sum(X.*X,2);
XXd=sum(X_D1.*X_D1,2);

XX1=repmat(XXn,1);
XX2=repmat(XXd',N,1);

D1=XX1+XX2-2*(X*X_D1');
D1(D1<0)=0;
D1=sqrt(D1);
D1(D1==0)=inf; %exclude zero
D1(D1<1e-6)=inf;%exclude very small distance
D_min=min(D1); %calculate the smallest distance from distance matrix
D1_sort=sort(D1)
%consider the second block from the rank distance D1_sort (similiar to
```

```
%baseline set
D1_2ndminBlock=D1_sort(41:80,:); % each block has 40 rows (40
observations)D1_2nd_mean=mean(D1_2ndminBlock);
D1_2nd_mean=mean(D1_2nd_mean);
paraD1=mean(D1_2nd_mean)

%construct kernel of the test set1
D1(D1==inf)=0;
K=D1.^2; % squared Eucludiean distance
Kd1=exp(-K./(2*paraD1.^2)); %compute the kernal matrix k

%centralised the feature space of the kernal using the Gram matrix
oneN_D1=ones(N_D1,N_D1)/N_D1;
K_D1=Kd1-oneN_D1*Kd1-Kd1*oneN_D1+oneN_D1*Kd1*oneN_D1;

%Project test data D1 in high dimensional space based on the
%loading coeeficient of the baseline setPCs of baseline baseline

Y3_D1=K_D1*eigVector;

%plot 3D of the kPCA projections
%repeat the same process for following  damage test sets in same procedure
% as Damage 1 test set
```

BIBLIOGRAPHY

[1] C. R. Farrar and K. Worden, *Structural Health Monitoring: A machine learning perspective*, 2013th ed. John Wiley & Sons, Ltd, 2013.

[2] K. Worden, H. Sohn, and C. R. Farrar, "Novelty Detection in a Changing Environment: Regression and Interpolation Approaches," vol. 258, pp. 741–761, 2002.

[3] K. Worden, T. Baldacchino, J. Rowson, and E. J. Cross, "Some Recent Developments in SHM Based on Nonstationary Time Series Analysis," *Proc. IEEE*, vol. 104, 2016.

[4] H. Sohn, K. Worden, and C. R. Farrar, "Novelty Detection under Changing Environmental Conditions," *SPIE's 8th Annu. Int. Symp. Smart Struct. Mater.*, 2001.

[5] E. J. Cross, G. Manson, K. Worden, and S. G. Pierce, "Features for damage detection with insensitivity to environmental and operational variations," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 468, no. 2148, pp. 4098–4122, 2012.

[6] E. Figueiredo, G. Park, C. R. Farrar, K. Worden, and J. Figueiras, "Machine learning algorithms for damage detection under operational and environmental variability," *Struct. Heal. Monit.*, vol. 10, no. 6, pp. 559–572, 2011.

[7] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*. Wiley-Blackwell, 2007.

[8] C. R. Farrar, T. A. Duffey, S. W. Doebling, and D. A. Nix, "A Statistical Pattern Recognition Paradigm for Vibration-Based Structural Health Monitoring," in *International Workshop on Structural Health Monitoring*, 1999, pp. 764–773.

[9] L. Wiskott, N. Kr, and C. Von Der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *Intell. Biometric Tech. Fingerpr. Face Recognit.*, pp. 1–23, 1999.

[10] D. Valentin, H. Abdi, A. J. O. Toole, and G. W. Cottrell, "Connectionist Models of

Face Processing: a Survey," *Pattern Recognit.*, vol. 27, no. 9, 1994.

[11] M. K. Heming and G. W. Cotwell, "Categorization of Faces Using Unsupervised Feature Extraction," in *1990 IJCNN International Joint Conference on Neural Networks*, 1977, pp. 65–70.

[12] J. Yang, D. Zhang, S. Member, A. F. Frangi, and J. Yang, "Two-Dimensional PCA : A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.

[13] M. A. Grudin, "On internal representations in face recognition systems," *Pattern Recognit.*, vol. 33, no. September 1998, pp. 1161–1177, 2000.

[14] D. Piotrowski, D. Roach, A. Melton, J. Bohler, T. O. M. Rice, S. Neidigk, and J. Linn, "Implementation of Structural Health Monitoring (SHM) into an Airline Maintenance Programme," 2014.

[15] A. Deraemaeker and K. ; Worden, *New Trends in Vibration Based Structural Health Monitoring*. Springer Wien New York, 2010.

[16] D. Piotrowski, J. Linn, P. Swindell, D. Roach, T. Rice, P. Swindell, I. Won, and M. Freisthler, "Validation of a Structural Health Monitoring (SHM) System and Integration Into an Airline Maintenance Program (Part 1) Boeing Structural Health Monitoring – Integration into Routine Maintenance," 2014.

[17] S. W. Doebling, C. R. Farrar, M. B. Prime, and D. W. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibartion characteristics: A literature review," 1996.

[18] S. W. Doebling, C. R. Farrar, and M. B. Prime, "A Summary Review of Vibration-Based Damage Detection Identification Methods," *Identif. Methods," Shock Vib. Dig.*, vol. 30, pp. 91–105, 1998.

[19] L. E. Mujica, J. ; Rodellar, J. ; Vehí, K. ; Worden, and W. Staszewski, "Extended PCA visualisation of system damage features under environmental and operational variations," in *Proceedings of SPIE*, 2009, vol. 7286.

[20] J. Sierra-Pérez, A. Güemes, and L. E. Mujica, "Damage detection by using FBGs and strain field pattern recognition techniques," *Smart Mater. Struct.*, vol. 22, no. 2, p. 025011, 2013.

[21] A. Güemes, "SHM Technologies and Applications in Aircraft Structures," in *5th International Symposium on NDT in Aerospace*, 2013, pp. 13–15.

[22] E. P. Carden and P. Fanning, “Vibration Based Condition Monitoring : A Review,” vol. 3, no. 4, pp. 355–377, 2004.

[23] I. Lopez and N. Sarigul-Klijn, “A review of uncertainty in flight vehicle structural damage monitoring, diagnosis and control: Challenges and opportunities,” *Prog. Aerosp. Sci.*, vol. 46, no. 7, pp. 247–273, 2010.

[24] H. Sohn, “Effects of environmental and operational variability on structural health monitoring.,” *Philos. Trans. A. Math. Phys. Eng. Sci.*, vol. 365, no. 1851, pp. 539–560, 2007.

[25] M. A. Rumsey and J. A. Paquette, “Structural health monitoring of wind turbine blades,” *Proc. SPIE 6933, Smart Sens. Phenomena, Technol. Networks, Syst.*, pp. 1–15, 2008.

[26] H. J. Lim, M. K. Kim, H. Sohn, and C. Y. Park, “Impedance based damage detection under varying temperature and loading conditions,” *NDT E Int.*, vol. 44, no. 8, pp. 740–750, 2011.

[27] L. Qiu, S. Yuan, H. Mei, and F. Fang, “An improved gaussian mixture model for damage propagation monitoring of an aircraftwing spar under changing structural boundary conditions,” *Sensors (Switzerland)*, vol. 16, no. 3, pp. 1–18, 2016.

[28] N. Dervilis, K. Worden, and E. J. Cross, “On robust regression analysis as a means of exploring environmental and operational conditions for SHM data,” *J. Sound Vib.*, vol. 347, pp. 279–296, 2015.

[29] D. A. Tibaduiza, L. E. Mujica, and J. Rodellar, “Damage classification in structural health monitoring using principal component analysis and self-organizing maps,” *Struct. Control Heal. Monit.*, vol. 20, no. Struct. Control Health Monit. 2013; 20:1303–1316, pp. 1303–1316, 2013.

[30] F. Gharibnezhad, L. E. Mujica, and J. Rodellar, “Comparison of two robust PCA methods for damage detection in presence of outliers Comparison of two robust PCA methods for damage detection in presence of outliers,” *J. Phys. Conf. Ser.*, vol. 305, 2011.

[31] H. Sohn, K. Worden, and C. R. Farrar, “Statistical Damage Classification Under Changing Environmental and Operational Conditions,” *J. Intell. Mater. Syst. Struct.*, vol. 13, no. 9, pp. 561–574, 2002.

[32] K. Worden and E. J. Cross, “On switching response surface models , with applications to the structural health monitoring of bridges,” *Mech. Syst. Signal Process.*, vol. 98, pp. 139–156, 2018.

[33] I. T. Jolliffe, “Principal Component Analysis, Second Edition,” *Encycl. Stat. Behav. Sci.*, vol. 30, no. 3, p. 487, 2002.

[34] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[35] J. Sierra-pérez and J. Alvarez-montoya, “Damage detection methodology under variable load conditions based on strain fi eld pattern recognition using FBGs , nonlinear principal component analysis , and clustering techniques,” *Smart Mater. Struct.*, vol. 27, 2018.

[36] C. K. Oh and H. Sohn, “Unsupervised Support Vector Machine Based Principal Component Analysis for Structural Health Monitoring,” *Science (80-.).*, vol. 8, no. 3, pp. 91–99, 2008.

[37] F. Gharibnezhad, L. E. Mujica, and J. Rodellar, “Applying robust variant of Principal Component Analysis as a damage detector in the presence of outliers,” *Mech. Syst. Signal Process.*, vol. 50–51, pp. 467–479, Jan. 2015.

[38] K. Worden, G. Manson, and D. Allman, “Experimental Validation of a Structural Health Monitoring Methodology: Part I. Novelty Detection on a Laboratory Structure,” *J. Sound Vib.*, vol. 259, no. 2, pp. 323–343, 2003.

[39] A. D. F. Santos, M. F. M. Silva, C. S. Sales, J. C. W. A. Costa, and R. A. Corrˆ, “Applicability of linear and nonlinear principal component analysis for damage detection,” in *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015.

[40] C. K. Oh and H. Sohn, “Damage diagnosis under environmental and operational variations using unsupervised support vector machine,” *J. Sound Vib.*, vol. 325, pp. 224–239, 2009.

[41] J. Sierra-Perez, A. Güemes, L. E. Mujica, and M. Ruiz, “Damage detection in composite materials structures under variable loads conditions by using fiber Bragg gratings and principal component analysis, involving new unfolding and scaling methods,” *J. Intell. Mater. Syst. Struct.*, vol. 26, no. 11, p. 14, 2014.

[42] M. A. Torres-arredondo, G. Alfredo, L. E. Mujica, G. Alfredo, L. E. Mujica, G. Alfredo, and L. E. Mujica, “Structural Health Monitoring by Means of Strain Field Pattern

Recognition on the basis of PCA and Automatic Clustering Techniques Based on SOM," in *17th IFAC Symposium on System Identification SYSID 2015*, 2015, vol. 28, pp. 987–992.

[43] G. Manson, "Identifying damage sensitive, environment insensitive features for damage detection," *Proc. 3rd Conf. Identif. Eng. Syst. Univ. Wales, Swansea*, 2002.

[44] C. Tschöpe and M. Wolff, "Statistical Classifiers for Structural Health Monitoring," *IEEE Sens. J.*, vol. 9, no. 11, pp. 1567–1576, 2009.

[45] S. O. B. Anerjee, X. I. P. Q. Ing, S. H. B. Eard, and F. U. U. O. C. Hang, "Prediction of Progressive Damage State at the Hot Spots using Statistical Estimation," *J. Intell. Mater. Syst. Struct.*, vol. 21, 2010.

[46] L. Qiu, S. Yuan, F. Chang, Q. Bao, and H. Mei, "On-line updating Gaussian mixture model for aircraft wing spar damage evaluation under time-varying boundary condition," *Smart Mater. Struct.*, vol. 23, 2014.

[47] T. K. Geoffrey McLachlan, *The EM Algorithm and Extensions, 2nd Edition*. Wiley, 2008.

[48] D. Chakraborty and N. Kovvali, "An adaptive learning damage estimation method for structural health monitoring," *J. Intell. Mater. Syst. Struct.*, vol. 26, no. 2, pp. 125–143, 2015.

[49] L. Qiu, S. Yuan, Q. Bao, H. Mei, and Y. Ren, "Crack propagation monitoring in a full-scale aircraft fatigue test based on guided wave- Gaussian mixture model," *Smart Mater. Struct.*, vol. 25, 2016.

[50] N. Mechbal, J. S. Uribe, M. Rebillat, N. Mechbal, J. S. Uribe, M. Rebillat, and A. P. M. Classifier, "A Probabilistic Multi-class Classifier for Structural Health Monitoring," *Sci. Arts Métiers*, 2015.

[51] G. Madzarov, D. Gjorgjevikj, and I. Chorbev, "A Multi-class SVM Classifier Utilizing Binary Decision Tree Support vector machines for pattern recognition," *Informatica*, vol. 33, pp. 233–241, 2009.

[52] K.-R. Schölkopf, B. ; Smola, A. ; Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," *Neural Comput.*, no. 10, pp. 1299–1319, 1998.

[53] S. O. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Pearson, 2009.

[54] T. M. Cover, "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Electron. Comput.*, 1965.

[55] Q. Wang, “Face Recognition and Active Shape Models,” 2012.

[56] K. Worden, G. Manson, and N. R. J. Fieller, “Damage Detection Using Outlier Analysis,” *J. Sound Vib.*, vol. 229, no. 3, pp. 647–667, 2000.

[57] L. Tarassenko, *A Guide To Neural Computing Applications*, 1st ed. Butterworth-Heinemann, 1998.

[58] E. Papatheou, G. Manson, R. J. Barthorpe, and K. Worden, “The use of pseudo-faults for novelty detection in SHM,” *J. Sound Vib.*, vol. 329, no. 12, pp. 2349–2366, 2010.

[59] LMS International, *Tutorial:Impact Testing Exercise: exicite the structure with a hammer*. 2013, pp. 1–8.

[60] LMS International, *Spectral processing LMS Test.Lab*. 2013.

[61] R. Barthorpe and K. Worden, *Course module MEC 409/6409: Signal Processing and Instrumentation*. 2015.

[62] K. Worden and G. Manson, “Experimental Validation of a Structural Health Monitoring Methodology: Part II. Novelty Detection on a Gnat aircraft,” *J. Sound Vib.*, vol. 259, no. 2, pp. 323–343, 2002.

[63] K. Worden, C. R. Farrar, G. Manson, and G. Park, “The fundamental axioms of structural health monitoring,” *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 463, no. 2082, pp. 1639–1664, 2007.

[64] H. Sohn, C. R. Farrar, F. Hemez, and J. Czarnecki, “A Review of Structural Health Monitoring Literature 1996 – 2001,” 2002.

[65] L. E. Mujica, J. Rodellar, A. Fernández, and A. Güemes, “Q-statistic and T2-statistic PCA-based measures for damage assessment in structures,” *Struct. Heal. Monit.*, vol. 10, no. 5, pp. 539–553, 2011.

[66] D. A. Tibaduiza, L. E. Mujica, J. Rodellar, and A. Guemes, “Structural damage detection using principal component analysis and damage indices,” *J. Intell. Mater. Syst. Struct.*, vol. 27, no. 51, pp. 233–248, 2015.

[67] L. E. Mujica, J. Vehi, M. Ruiz, M. Verleysen, W. Staszewski, and K. Worden, “Multivariate statistics process control for dimensionality reduction in structural assessment,” *Mech. Syst. Signal Process.*, vol. 22, no. 1, pp. 155–171, 2008.

[68] Jonathon Shlens, *A Tutorial on Principal Component Analysis*. 2014.

[69] S. Sharma, *Applied Multivariate Techniques*. John Wiley & Sons, Ltd, 1996.

[70] E. Reynders, G. Wursten, and G. De Roeck, “Output-only structural health monitoring

in changing environmental conditions by means of nonlinear system identification," *Struct. Heal. Monit.*, vol. 13, no. 1, pp. 82–93, 2014.

[71] G. Gardiner, "Structural health monitoring: NDT-integrated aerostructures enter service," *CompositesWorld*, 2015.

[72] MathWorks: Statistics and Machine Learning Toolbox, *Statistics and Machine Learning toolbox: Cluster Analysis-Gaussian Mixture Model, User's Guide*. 2017.

[73] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[74] D. G. Duda, R., Hart, Peter E., & Stork, *Pattern classification*. 2nd ed., Wiley-Interscience publication, 2001.

[75] MathWorks: Neural network toolbox, *Neural network toolbox: Multilayer-neural-networks, User's Guide (R2017b)*. 2017.

[76] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 2–6, 1994.

[77] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural Network Toolbox ™ User ' s Guide MathWorks*. 2018.

[78] F. D. Foresee and M. T. Hagan, "Gauss-Newton Approximation to Bayesian Learning," *Proc. 1997 Int. Jt. Conf. Neural Networks*, 1997.

[79] D. J. C. Mackay, "A Practical Bayesian Framework for Backprop Networks," *Neural Comput. 4*, vol. 74, no. 2, pp. 1–20, 1992.

[80] D. J. C. Mackay and N. Systems, "The Evidence Framework applied to Classification Networks," *Neural Comput. 4*, pp. 1–16, 1992.

[81] Endevco, *Guide to adhesively mounting accelerometers*. 2013.

[82] G. Z. Qi, X. Guo, P. Chang, X. Qi, and W. Dong, "Local Modes of Vibrations for Civil Structural Health Monitoring," in *The 14 World Conference on Earthquake Engineering*, 2008, no. 2.

[83] F. Wahl and G. Schmidt, "On the significance of Anti-resonance Frequencies in Experimental Structures," *J. Sound Vib.*, vol. 219, no. 3, pp. 379–394, 1999.

[84] D. J. Ewins, *Modal Testing: Theory and Practice*. Wiley, 1984.

[85] C. R. Farrar and K. Worden, "An introduction to structural health monitoring.," *Philos. Trans. R. Soc.*, vol. 365, no. 1851, pp. 303–15, 2007.

[86] D. S. Touretzky, "Biological Inspiration for RBFs," pp. 1–41, 2006.

[87] L. Breiman, Jerome H. Friedman, R. A. Olshen, and Charles J. Stone, *Classification And Regression Trees*. Chapman & Hall/CRC, 1984.

[88] L. E. O. Breiman, "Random Forests," in *Machine Learning*, Kluwer Academic Publishers, 2001, pp. 5–32.

[89] L. E. O. Breiman, "Bagging Predictors," in *Machine Learning*, Kluwer Academic Publishers, 1996, pp. 123–140.

[90] E. Ayyıldız and V. Purutc, "A Short Note on Resolving Singularity Problems in Covariance Matrices," vol. 1, no. 2, pp. 113–118, 2012.

[91] J. Franklin, M. Rodrigues, N. Tremblay, and P. Belanger, "The Feasibility of Structural Health Monitoring Using the Fundamental Shear Horizontal Guided Wave in a Thin Aluminum Plate," *Materials (Basel).*, vol. 10, 2017.

[92] W. J. Staszewski, S. Mahzan, and R. Traynor, "Health monitoring of aerospace composite structures – Active and passive approach," *Compos. Sci. Technol.*, vol. 69, no. 11–12, pp. 1678–1685, 2009.

[93] W. J. Staszewski, C. Boller, and G. R. Tomlinson, *Health Monitoring of Aerospace Structures: Smart Sensor Technologies and Signal Processing*. John Wiley & Sons, Ltd, 2004.

[94] J.-B. Ihn and F.-K. Chang, "Pitch-catch Active Sensing Methods in Structural Health Monitoring for Aircraft Structures," *Struct. Heal. Monit.*, vol. 7, no. 1, pp. 5–19, 2008.