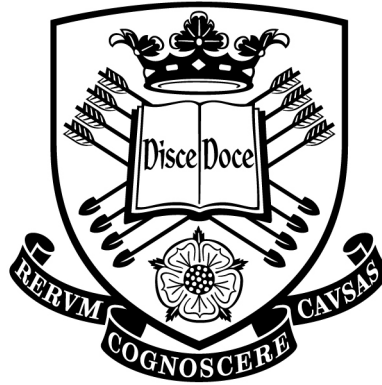# A machine learning approach to Structural Health Monitoring with a view towards wind turbines

A Thesis submitted to the University of Sheffield

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

by

Nikolaos Dervilis

Department of Mechanical Engineering

University of Sheffield

2013

ABSTRACT

The work of this thesis is centred around Structural Health Monitoring (SHM) and is divided into three main parts.

The thesis starts by exploring different architectures of auto-association. These are evaluated in order to demonstrate the ability of nonlinear auto-association of neural networks with one nonlinear hidden layer as it is of great interest in terms of reduced computational complexity. It is shown that linear PCA lacks performance for novelty detection. The novel key study which is revealed amplifies that single hidden layer auto-associators are not performing in a similar fashion to PCA.

The second part of this study concerns formulating pattern recognition algorithms for SHM purposes which could be used in the wind energy sector as SHM regarding this research field is still in an embryonic level compared to civil and aerospace engineering. The purpose of this part is to investigate the effectiveness and performance of such methods in structural damage detection. Experimental measurements such as high frequency responses functions (FRFs) were extracted from a 9m WT blade throughout a full-scale continuous fatigue test. A preliminary analysis of a model regression of virtual SCADA data from an offshore wind farm is also proposed using Gaussian processes and neural network regression techniques.

The third part of this work introduces robust multivariate statistical methods into SHM by inclusively revealing how the influence of environmental and operational variation affects features that are sensitive to damage. The algorithms that are described are the Minimum Covariance Determinant Estimator (MCD) and the

Minimum Volume Enclosing Ellipsoid (MVEE). These robust outlier methods are inclusive and in turn there is no need to pre-determine an undamaged condition data set, offering an important advantage over other multivariate methodologies. Two real life experimental applications to the Z24 bridge and to an aircraft wing are analysed. Furthermore, with the usage of the robust measures, the data variable correlation reveals linear or nonlinear connections.

Acknowledgements

First and foremost sincere gratitude must go to my supervisor Professor Keith Worden, without whom this work would not have been possible. Keith was always available to guide, help, support and add ideas throughout this thesis. I have to especially thank him for keeping up with my mental thoughts and my unlimited repulsion for coriander :). I have also to thank Keith for being a very good friend.

Second of all I would like to express my sincere thanks to my second supervisor Dr. Robert Barthorpe for his continuing help, guidance and friendship. Also, I have to thank Rob for providing me with data from the aircraft wing.

I am also grateful to Professor Wieslaw Staszewski who co-supervised my early PhD years. His contribution was important for the continuation of my PhD.

I would like to express my gratitude to Dr. Elizabeth Cross for providing me with data from the benchmark study on the Z24 highway bridge. Especially, I would like to thank Lizzy for her support, help and friendship.

A lot of thanks must also go to Dr. Charles Farrar and Professor Gyuhae Park (now in Chonnam National University, Korea) from Los Alamos National Laboratory for their support and guidance. I must especially thank Chuck and LANL engineering institute group for all the help and hospitality during my stay in Los Alamos. A large part of this work has focused on a wind turbine blade fatigue test data coming from Los Alamos National Laboratory and a lot of thanks must also go to M. Choi and S.G. Taylor.

A lot of thanks must also go to Dr. A.C.W. Creech for providing me with data from the CFD model of the wind farm and Dr. A.E. Maguire for his help on Lillgrund Wind farm.

I would like to thank everyone in the Dynamics Research Family. It is a privilege to belong and work with such friendly and bright people. Special thanks have to go to Sofia, Pete, Nevena, Claudio, Charles, Vangelis and Tara for all their help and companionship. Also, I would like to thank Les Morton provided the necessary help during the experimental work in the Lab.

My special thanks have to go to a close friend and colleague that we started together this journey. This person is Ifigeneia. With Ifi we shared a lot of time laughing, travelling and working and she has to put up with me working in opposite offices and keep talking about all the non research topics :).

Lastly, but by no means least, I would like to thank my family and friends for their unconditional love throughout my studies. I have to especially thank my unlimited crazy parents and brother for not losing their smile and craziness :). Special thanks to my grand parents (Dimitrakis & Panagoula). They know why and I know why.

As Euclid said: "There is no royal road to geometry". This journey just started and I will try my best to travel this difficult road and learn the "geometry".

P.S. I have to thank Keith for introducing Father Ted and League of Gentlemen.

NOMENCLATURE

Acronyms

AANN: Auto-associative neural network

AIC: An Information theoretic Criterion

CFRP: Carbon fibre-reinforced composites

CM: Condition Monitoring

EM: Expectation-Maximisation

EWEA: European Wind Energy Association

FPE: Final Prediction Error

FRF: Frequency response function

GFRP: Glass fibre-reinforced composites

GP: Gaussian process

LANL: Los Alamos National Laboratory

MCD: Minimum Covariance Determinant Estimator

MDOF: Multi-degree of freedom

MEMS: Microelectromechanical systems

MLP: Multi-layer perceptron

MSD: Mahalanobis Squared-Distance

MVEE: Minimum Volume Enclosing Ellipsoid

MVIE: Maximum Volume Inscribed Ellipsoid

NDE: Non destructive evaluation

NLPCA: Nonlinear Principal Component Analysis

NREL: National Renewable Energy Laboratory

NWTC: National Wind Technology Center

PCA: Principal Component Analysis

PMC: Polymer matrix composites

PPCA: Probabilistic Principal Component Analysis

RBF: Radial Basis Functions

SCADA: Supervisory Control and Data Acquisition

SDOF: Single-degree of freedom

SHM: Structural Health Monitoring

SVD: Singular value decomposition

UREX: Universal Resonant EXcitation

WT: Wind turbine

TABLE OF CONTENTS

Chapter 1

SCOPE OF THE THESIS

This chapter introduces the three main issues that are addressed throughout this work such as the auto-associative neural networks paradox, Structural Health Monitoring (SHM) of wind turbines and the critical influence of multiple outliers due to changing environmental and operational conditions. Firstly, the motivation for this research is presented and is followed by an outline of the problems with which the current work is dealing. Finally, a brief description of the layout of this thesis is summarised.

1.1 Motivation

A major novel investigation regarding nonlinear methods for novelty detection is addressed in this work. This investigation is centred around an auto-association 'paradox'. In several damage detection approaches it is never possible or it is very difficult to obtain true measurements for all possible damage classes especially in complex structures such as composite systems. Furthermore, data that is collected during a damaged state of structure is very rare. The premise of novelty detection techniques is to seek the answer to a simple question; given a newly presented measurement from the structure, does one believe it to have come from the structure in its undamaged state? The advantage of novelty detection is clear; any abnormality

defines a new situation characterised by a truly new event for the structure.

It will be demonstrated that linear tools lack performance for multimodal classification problems and novelty detection. On the other hand, the classic five layer neural network is a very powerful nonlinear tool but its complexity of finding the right number of nodes and the computational power of optimising the weights and biases is a practical drawback. This is the reason that three layer nonlinear auto-associators will be exploited in terms of their fundamental difference with Principal Component Analysis (PCA) as they could offer a significant practical speed advantage.

Over the last few years, there has been a dramatic increasing trend in the structural engineering community of monitoring structures for SHM purposes via implementing sensor networks. One of the sectors that is gaining a lot of financial and structural attention is Wind Turbines (WTs). In recent years the dramatic increase of large wind turbine installations has raised a chain of critical issues regarding the ability of such sustainable systems to compete with the traditional fossil fuel-based industry supported by new technological add-ons like CO_2 capture technologies and new generation nuclear power plants.

Reliability of WTs is one of the key cards for the successful implementation of renewable power plants in the energy arena. Failure of the WTs could cause massive financial losses especially for structures that are operating in offshore sites. Structural Health Monitoring (SHM) of WTs is essential in order to ensure not only structural safety but also avoidance of over-design of components that could lead to economic and structural inefficiency.

Within the remit of this project, the aim of the author's work is to investigate a group of pattern recognition techniques for the monitoring of wind turbine blades by using vibration data. Also, the idea of using Supervisory Control and Data Acquisition (SCADA) measurements for SHM has received very little attention from both the wind energy and SHM communities. A preliminary analysis of a simple methodology for detecting abnormal behaviour of WT systems using a data-driven approach based on artificial SCADA data extracted from a CFD model is illustrated.

And last but not least, one of the milestones of this research is to gain a greater understanding of the effects of changing operational and environmental conditions on the measured dynamic response of a structure through robust multivariate statistics.

The key novel element of this work is the introduction of robust multivariate statistical

methods to the SHM field through use of the Minimum Covariance Determinant Estimator (MCD) and the Minimum Volume Enclosing Ellipsoid (MVEE). Robust outlier statistics are investigated, focussed mainly on a high level estimation of the "masking effect" of *inclusive* outliers in order to examine the normal condition set under the suspicion that it may already include multiple abnormalities. This is very important in order to develop ways to account for the effects of these varying external factors and develop a reliable model.

The work covered in this thesis is funded by the EU Marie Curie scheme through the Initial Training Network SYSWIND dealing with structural and aerodynamic aspects of new generation wind turbines. The basic aim of University of Sheffield team is the introduction of advanced signal processing tools that could be used for SHM and CM purposes.

In the following, a short overview of the issues addressed in this work are described and an outline of the structure of this thesis is made.

1.2 A brief introduction

In each chapter there is a detailed introduction about the specific subject and the basic theory that is involved. Here there is a small presentation of the general subjects behind the thesis.

1.2.1 Auto-associative neural networks story

Auto-associative neural networks (AANN) are based on Nonlinear Principal Component Analysis (NLPCA) and in some respect are similar to the well-known linear method of Principal Component Analysis (PCA). NLPCA can be used to detect and remove correlations among variables and like PCA it can be applied as a dimensionality reduction technique, visualization of variable correlations and generally is a strong algorithm for exploring data characteristics. The key advantage of the AANN, is that compared to PCA which identifies only linear correlations among the problem variables, it can reveal both linear and nonlinear correlations between the variables without any drawback regarding the nature of the nonlinearities present or the data distribution. Analytic details about the architecture of the method

can be found in Chapter 3. Briefly, AANN operates by training a feedforward multilayer perceptron where the inputs are reproduced at the output layer. The network consists of an internal "bottleneck" layer (containing less nodes than input or output layers) and two additional hidden layers that force the AANN to learn a compressed representation of the input data.

The story goes back to 1982 and one of the first attempts connecting PCA with feature extraction and neural networks was by Oja [1] where he showed that PCA could be implemented by a linear neural network. In 1986, Hinton [2] commented of the ability of nonlinear ANNs to produce internal representations in their hidden nodes, and in 1986, Rumelhart et al.[3] demonstrated the ability of nonlinear auto-associators to solve the encoder problem, a problem that could not be solved by PCA due to the singularity of the principal components. In 1987 Ballard [4] used the idea of multiple single-hidden layer networks to solve complex encoding problems.

A serious establishment of multilayer neural networks for NLPCA is demonstrated in the work of Bourlard & Kamp [5] in 1988 and Baldi & Hornik [6] in 1989. In 1991 Kramer [7] introduced a complete work regarding the methodology of auto-associators. One of the recent advances in AANN was by Matthias Scholz et al. [8] where they proposed an algorithm that extends PCA into NLPCA through a hierarchical type of learning (h-NLPCA).

In terms of novelty detection for SHM purposes the idea of AANNs can be found in the work of Worden [9], Sohn [10] and Zhou [11].

The novel problem that this thesis addresses is a 'paradox'. The paradox has its routes in the conclusions derived by two classic papers of Bourlard & Kamp [5] and Cottrell & Munro [12]. The paradox is that even with nonlinear nodes present in the hidden layer, autoassociators trained with backpropagation are equivalent to linear methods such as PCA when the AANN consists of only a single hidden layer.

There is a limited research regarding this 'paradox' in the neural network community. This problem is exploited by a paper of Japkowicz et al.[13]. In the current study there is an investigation of the paradox in terms of novelty detection.

An analysis is performed in order to demonstrate that the ability of nonlinear auto-association of MLPs consisted of a single nonlinear hidden layer and with linear and nonlinear activation functions in the output layer is *not* equivalent to PCA. This investigation is of interest in terms of computational power and complexity compared

to the classic five layer AANN.

1.2.2 The Wind Energy interest explosion

As stated in the latest reports of the European Wind Energy Association (EWEA) [14, 15], total investments in wind turbines installations in EU region were worth between 12.8bn and 17.2bn Euros in 2012 with a total capacity of 11.895 MW. Wind energy accounted for around 26.5% of total 2012 power capacity installations.

Of the 11895 MW installed in 2012, 10729 MW was onshore and 1166 MW offshore wind farms [14, 15]. The total investment amounts demonstrate impressive numbers. Onshore wind farms had the lead with 9.4bn-12.5bn Euros and offshore wind farms between 3.4bn-4.7bn Euros. Germany was the leader market in 2012 in terms of total installations with 2415 MW with 80 MW of which were offshore farms. The UK took the second place with 1897 MW with 854 MW of which (more than 45%) were offshore, followed by Italy with 1273 MW, Spain (1122 MW), Romania (923 MW), Poland (880 MW), Sweden (845 MW) and France (757 MW).
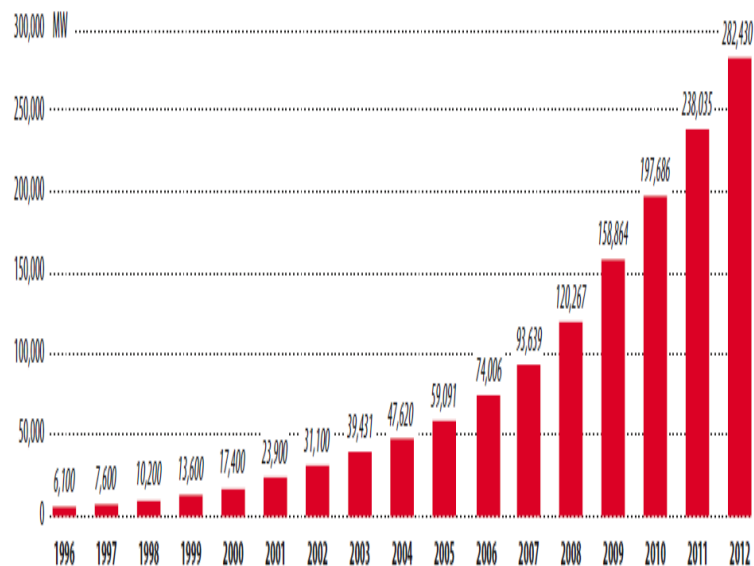
Figure 1.1: The numbers represent the global wind power installations between years 1996-2012 (cumulative) [16].
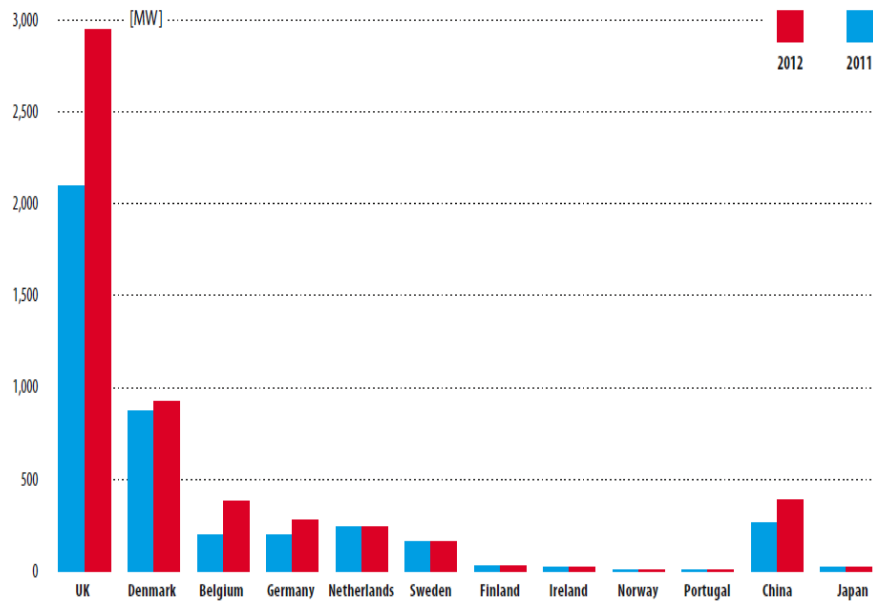
Figure 1.2: The numbers represent the global offshore wind power installation (cumulative) [16].
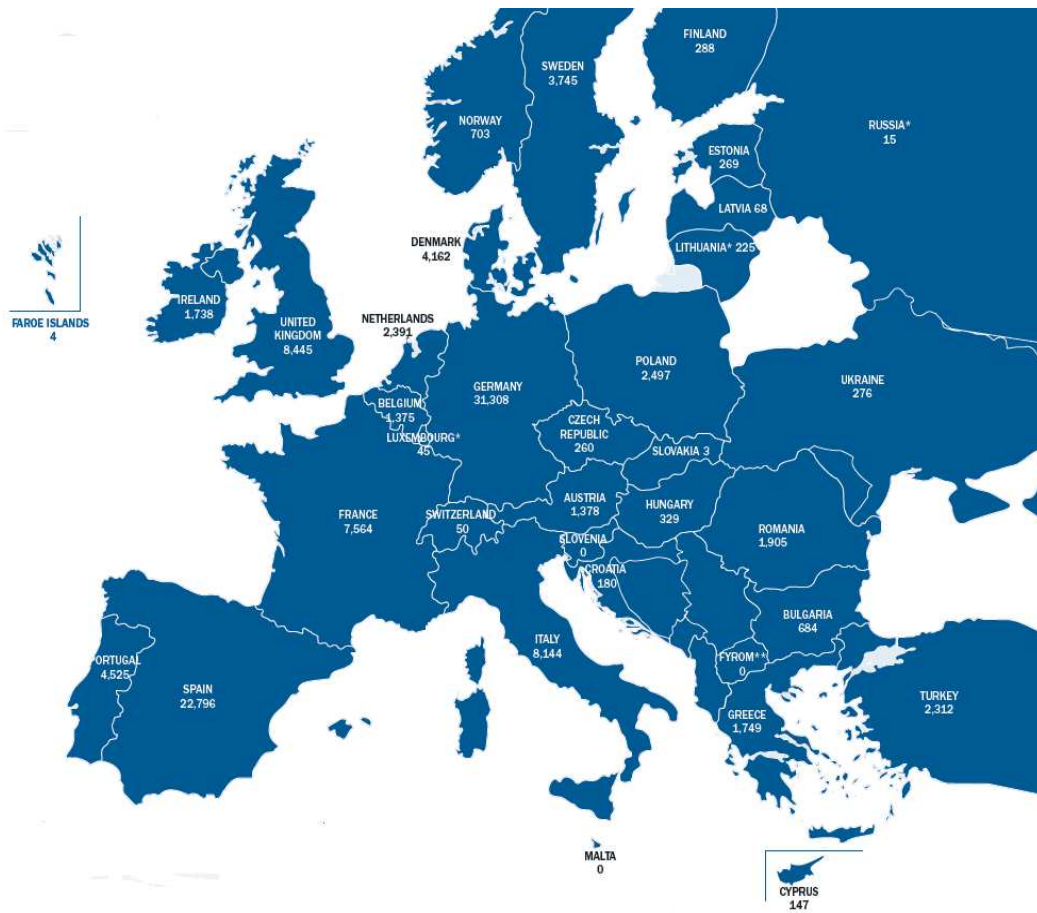
Figure 1.3: The numbers represent the wind power installed in Europe (cumulative) [14].

1.2.3 Why SHM is important for wind turbines

Almost any component of a wind turbine system is subject to possible damage from a failure of tower to a failure of the blades. In Fig.1.4 there is a schematic of a wind turbine generator.
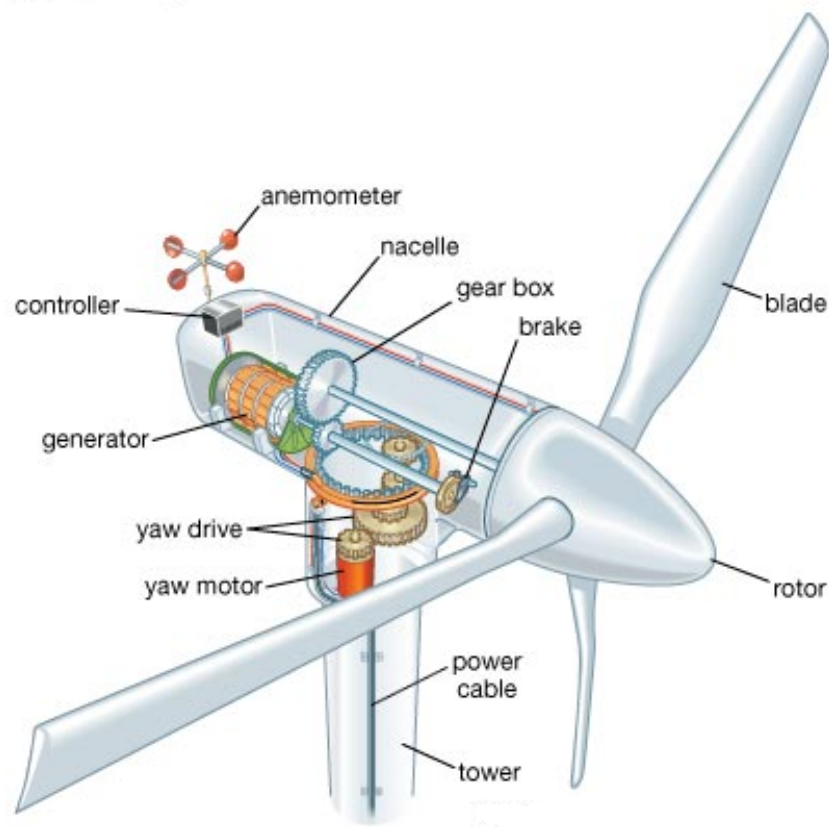
Figure 1.4: Schematic of a wind turbine generator [17].

The number of sources of information regarding reliability of WT technology lifetimes is very limited due to the highly completive market. One of the few databases and studies that evaluated operational information is that from the German 250 MW Wind Test Program [18, 19]. This specific program collected data on operational experience of WTs for a period of up to 15 years.

Up to the end of 2006, operators sent in around 63000 reports regarding maintenance or repair of WT systems and around 14000 annual reports regarding the operating costs of farms [18, 19]. The results over 15 years is shown in Fig.1.5, including failure of both structural and electrical components. In Fig.1.6 the average failure rate and the average downtime per component of the wind turbines can be seen from the same report. At this point, it can be assumed that due to the increased size of offshore

wind turbines and the trend towards direct drive permanent magnet generators (no gearbox components) the cost and the downtime of replacing a damaged blade or generator will be significantly higher.
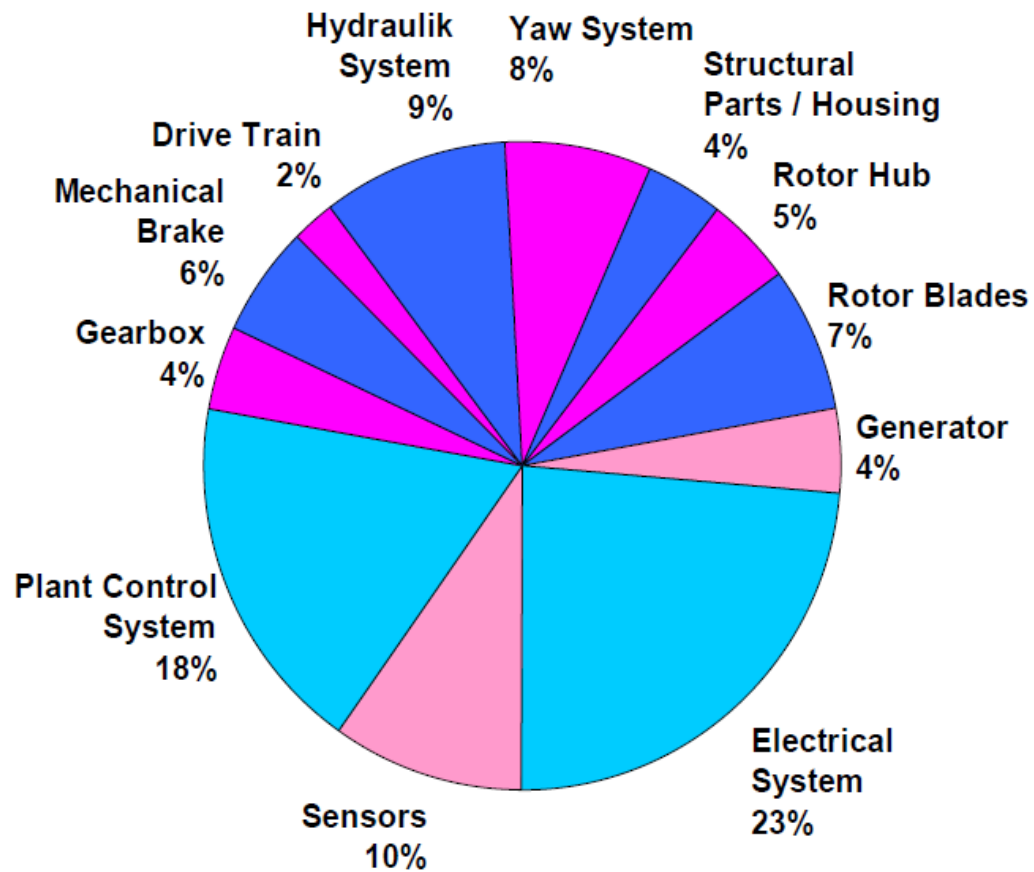
Figure 1.5: Failure of different components percentage for wind turbines at a wind farm in Germany over 15 years [18].
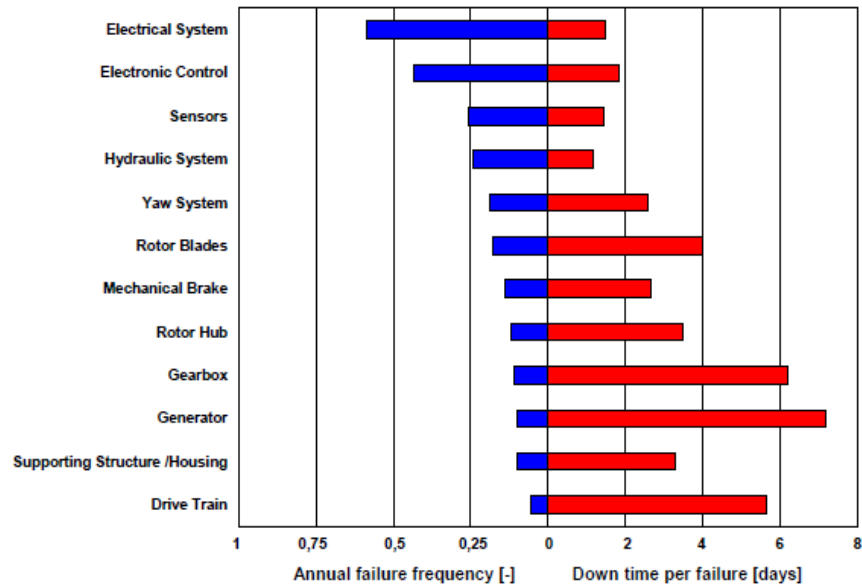
Figure 1.6: Failure frequency and downtimes of components [18].

SHM is very important step for the further development of WT farms. SHM can alert the operators to catastrophic failures and secondary effects (blade damage can cause critical failure to the whole wind turbine system-tower collapse). Online, continuous and global SHM can reduce maintenance and replacement cost by monitoring WTs at offshore and remote sides. Last but not least SHM can complement the research for further structural development of WTs (improve designs for the next generation of WTs).

Offshore wind farms are going to be the pioneers in future regarding the renewable energy sources; however, because they operate in remote areas away from land and expanding into deeper waters, SHM is an essential part of the success of these structures in the competitive market. With rotor diameters exceeding, nowadays, 150 meters (Fig.1.7) damage detection of the blades is of critical interest. An example of offshore wind turbine blade inspection and repair is shown in Fig.1.8.
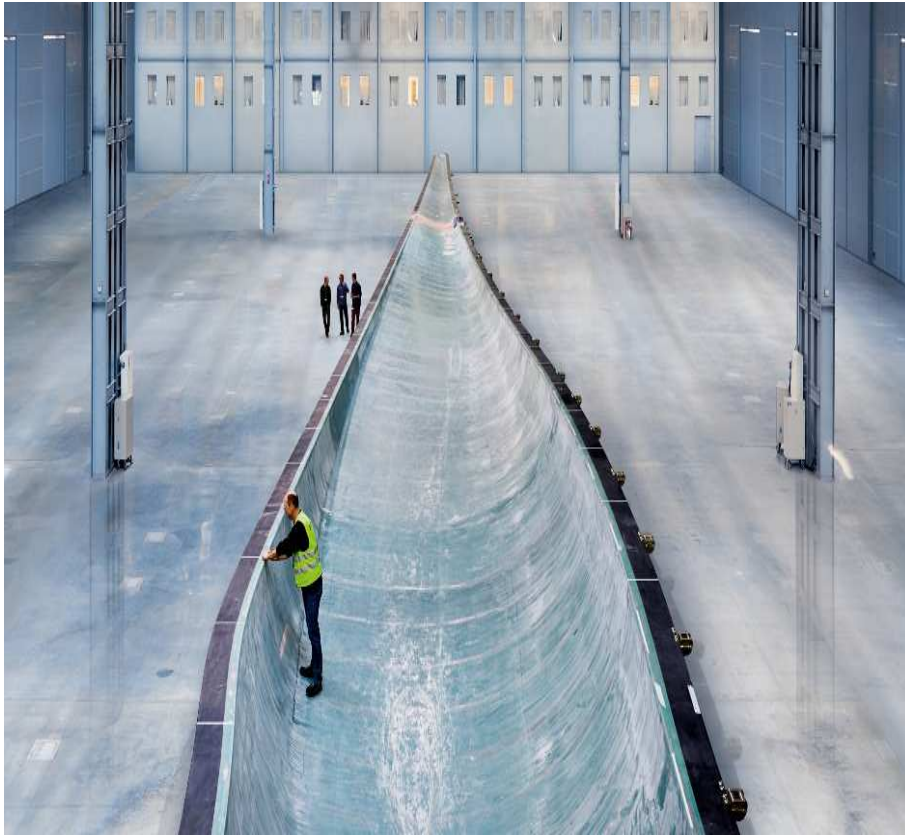
Figure 1.7: 75m blade length of a 6MW offshore Siemens wind power plant at Østerild, Denmark [20].

Following these thoughts, a view of sensitive and robust damage detection methodologies of signal processing are investigated in this thesis. One of the most important and expensive components of new generation wind turbines is the large blades that can go up to 150m in diameter and reach rated powers of over 5 MW. Automatic mechanisms for damage detection of the blades are still in the embryonic stage. A brief history of the adopted methodologies for SHM of wind turbine blades is given in Chapter 4.

The current study is based on a combination of novelty detection algorithms and high frequency dynamic responses of a wind turbine blade under continuous fatigue loading. Damage in a turbine blade is mainly caused by fatigue resulting in a type of cracking or delamination of the composite body of the blade [22]. As a further step, a preliminary analysis of using SCADA based observations from a CFD model for monitoring offshore wind farms is introduced.

Figure 1.8: Repairs of wind turbines at Windpark Prettin, GE Energy, Saxony-Anhalt [21].

1.2.4   Hide-and-seek: Inclusive outliers

One of the first critical steps in signal processing and pattern recognition techniques for SHM is understanding the data and in order to obtain this logical analysis, the detection of outlier observations is important. Outliers are often regarded as "garbage" points that they have to be removed, but a lot of times they carry critical information. Furthermore, outliers could lead to the adoption of a misleading model which can result in biased parameter estimation and wrong results. Consequently, it is important to identify and detect them before proceeding to any analysis [23, 24].

Detecting outliers in multivariate data often proves to be more difficult than in univariate data because of the additional "space" within which a potential outlier may hide. Mahalanobis distances provide the classic tool for outliers in normally distributed data. But this standard test is extremely sensitive to the presence of outliers. If there were groups of outliers already present in the training data, they would have a critical influence on the sample mean and covariance in such a way that they would subsequently indicate small distances on new observations or outlying data and thus cause the outliers to become invisible.

This is the reason that robust multivariate statistical methods are essential, especially in SHM when pattern recognition and system identification techniques are implemented.

An analytic description of the methods as well as a definition of what is an outlier can be found in Chapter 4 and 5, but before that a short literature background is presented here.

Robust estimators of the location (mean) and the shape (covariance matrix) of the data distribution can be found in early and later work of Rousseeuw et al.[25–28].

Hadi [29, 30] identifies the problem of inclusive outliers and the need for replacement of the classic arithmetic mean and covariance with updated, more efficient, versions of these multidimensional parameters.

A comparison report of multivariate outlier detection methods for the purpose of clinical laboratory safety data is presented by Penny et al.[31].

Riani et al.[32] proposed the forward search in order to calculate robust Mahalanobis distances to detect outliers in multivariate normal distributions.

Pison et al.[33] proposed small sample correction factors for the usage of the least trimmed squares estimator and the minimum covariance determinant estimators. In the same spirit Hardin et al.[34] used the minimum covariance determinant estimator by adding an F approximation to give outlier rejection points.

Atkinson et al.[35] used the series of robust estimators that is provided by the straightforward methods to explore the architecture of the Mahalanobis distances.

The key novel element of this chapter is the introduction of robust multivariate statistical methods to the Structural Health Monitoring (SHM) field through use of

the Minimum Covariance Determinant Estimator (MCD) and the Minimum Volume Enclosing Ellipsoid (MVEE).

In this study the robust outlier techniques were tested in the context of SHM and their importance in establishing a normal condition that is clear of outliers. Through the two real life experimental applications of the Z24 bridge and of an aircraft wing, the critical importance of the different uses of robust multivariate statistics are demonstrated.

1.3 Brief outline of thesis

- Chapter 2 gives a general introduction to the research field of SHM.

- Chapter 3 introduces a comparison of different architectures of auto-association which are evaluated in order to investigate the paradox as described earlier (nonlinear auto-association consisted of one nonlinear hidden layer).

- Chapter 4 demonstrates how one can combine vibration-based measurements and novelty detection techniques to reveal early damage presence in wind turbine.

- Chapter 5 introduces the concept of robust outlier analysis theory from the field of multivariate statistics and how it can be used in the context of feature selection in SHM.

- Chapter 6 builds on the findings of Chapter 4 in order to apply robust outlier tools by revealing environmental and operational trends from selected features.

- Chapter 7 concludes the thesis and future work is discussed.

<div align="right">

Chapter 2

</div>

SHM AND MACHINE LEARNING

The term *structural health monitoring* (SHM) generally refers to any kind of damage detection procedure for civil, aerospace or mechanical engineering structure. The beginnings as an area of research goes back as far as the time when visual inspection was used for fault inspection. As part of solid engineering interest it probably started around the decade of 1970s [36]. In 1969, Lifshitz and Rotem [37] demonstrated one of the first works in damage detection through vibration measurements.

Nowadays, SHM research is growing with dramatic interest, and apart from civil or aerospace structures, the game is targeting the energy sector infrastructure. This chapter aims to provide a general introduction to the field of SHM and advantages of a robust SHM system.

2.1 What is Damage

As stated by Farrar and Worden [38], damage can be defined as changes that are introduced into a system, either intentionally or unintentionally, that will affect the current or future performance of the system. This system could be a structure or a biological organism.

SHM refers to structural and mechanical systems and as a result damage can be defined as intentional or unintentional changes to the material and/or geometry architecture of the structure [38]. This changes could be in a micro-scale level (material matrix anomalies), something that can be considered present in the majority of systems as a natural formation, or in the macro-scale level such as cracks due to fatigue or corrosion or an impact.

In previous decades, the natural damage due to micro-scale abnormalities was difficult to account for in the long term performance of the structure. Nowadays, with the technological explosion in terms of materials development, scientists and engineers can account for these low scale anomalies and in turn the structure can perform as it was designed to do.

2.2 The importance of SHM

In the question of why to bother with SHM, the two quick answers are human life protection and financial motivation. Human safety is a massive motivation if someone thinks about aeroplane crashes, house or bridge collapses due to earthquake or impacts where many lives have been lost.

The second and probably dominant motivation (not surprisingly) that drives forward the research in this field is the needs of private or public industries. A significant number of structures undergo routine inspections and maintenance in order to ensure structural stability of the system. Detection of damage at an early stage could have a big economic impact.

The costs of these routine inspections could be significantly reduced if these inspections are shown to be unnecessary when a structure continues to be healthy and this could automatically be indicated by implementing an SHM system. SHM could offer robust and online monitoring and necessary maintenance or repairs could be addressed based on this technology.

Imagine the downtime cost of an offshore wind turbine or an offshore oil platform when a structure may undergo routine maintenance or emergency component replacement, which, in turn, would be an economic and environmental disaster.

Furthermore, nowadays, companies both in energy (an example is nuclear power

plants) and aerospace industry are keen on extending the initial life time of the these structures. Of course, with ageing comes "life fatigue" and economic issues are arising regarding the stability of these structures. SHM could offer a vital tool in inspecting continuously the systems for potential failures.

Last but not least, is the defence industries. The military market is keen on developing SHM technology in order to detect damage and predict operational lifetime of the structure during combat missions.

SHM is the technology that will potentially allow the time-based inspection and maintenance to move into condition-based maintenance approaches. The basic philosophy behind the condition-based maintenance is that a holistic and robust sensor network will monitor the system and via smart measurement processing will arise an alert to the operator in case of system abnormalities.

The critical steps for a holistic SHM investigation are well described in Rytter's hierarchy [39] and in Worden and Dulieu-Barton [40] with a number of small suggestions and additions to this hierarchy but without changing the nature of Rytters description. These levels can be summarised as follows:

- Level One: Existence of damage to the system (Detection).

- Level Two: Identification of where damage has appeared in the system (Localisation).

- Level Three: Which is the specific kind of damage (Type).

- Level Four: Investigation of damage severity (Quantification).

- Level Five: Prediction of the remaining useful life in the system (Prognosis).

2.2.1 Analogous areas of research with SHM

Condition Monitoring (CM) is a similar area of research with SHM, but is mainly referring to damage detection in rotating machinery [40]. CM has demonstrated success and is considered as a mature technology compared to SHM. A number of factors can be considered as the key elements for this more established approach and the basic ones are that rotating machinery is giving specific dynamic responses for specific fault classes, and as result failure detection and identification is an easier

target relative to SHM [41]. This is aided also by the fact that machineries operate in a controlled environment and their size is relatively small compared to the size of the structures SHM is targeting (an example is bridges or skyscrapers).

Another related group of methods for damage detection purposes is non destructive evaluation (NDE). NDE successfully made the transition to industry and practical engineering applications [40, 42]. In contrast with SHM that operates continuously and online, NDE is commonly carried out off-line.

Common techniques that are used for NDE are acoustic emissions, X-rays and microscopy. NDE tools are often applied to a small site of a structure where the potential damage is located.

2.3   Sensors of SHM instrumentation

For a detailed and comprehensive review of instrumentation in SHM, readers are referred to [38, 43–45]. As an introduction to SHM, a brief overview of the sensing and data acquisition methodologies is given in this section.

Instrumentation is consisted of sensors and data acquisition hardware which transforms the dynamic response of the system into a voltage signal that is analogous to the desired response [38]. The most common measurements used for SHM are of the dynamic response of a structure. Dynamic input and response quantities deliver information about the mass, stiffness and damping of a structure, quantities that are sensitive to the formation of damage.

Dynamic measurements like acceleration, force or strains are not the only data that should be acquired. As stated by Farrar and Worden [38], other quantities like damage-sensitive physical measurements (electromagnetic field or chemicals) or environmental and operational observations should be also evaluated.

The current commercially available sensors used in SHM include: microelectromechanical systems (MEMS), displacement transducers, fibre optic strain sensors and piezoelectric actuators/sensors. The latter ones have been proved to be reliable and stable sensors for SHM. However, these dynamic response data are focused on local measurements. Global sensing technologies can be considered using laser Doppler velocimeters, digital image correlation, tomography and acoustic emission sensors

[38].

The two main drawbacks when considering the dynamic response of a structure are that often in reality there is no knowledge of the excitation source (modal analysis) and that most of the sensor networks are wired systems.

Regarding the first issue, structures in practice are excited from induced operational conditions (traffic on a bridge or wind on a wind turbine blades) which, in practice, cannot be accurately measured. In such case an assumption has to be made regarding the excitation source (white noise for example) [46, 47]. A common practice is also the usage of external artificial excitation that can be calculated such as an impact vibration hammer or a shaker.

Wireless systems are a solution for wired systems in the case that structures are in remote areas [48–52]. The main disadvantage of wireless sensors for SHM are power supply of the sensors and data telemetry with energy harvesting being a field that presents an interest increase for powering such sensors [53, 54].

Guided wave technologies represent another field for damage detection tools. These wave methods use high frequency (compared to vibration based measurements) for damage assessment. There is an exhaustive literature about guided waves and the reader is referred to [55–59].

2.4   Pattern Recondition and assessment of measurements

A catholic argument is that no sensor exists that can directly measure any type of novelty. For this reason feature extraction is used to derive useful metrics from the raw data that can further be post-processed through advanced signal processing tools. The basic aim of using such features is to lower the dimensionality of the data as the raw measurements like time histories typically have high dimensions, making the assessment of the raw data impossible in practice. In the machine learning community this drawback is referred as the "curse of dimensionality".

Once a particular feature is obtained, a decision algorithm has to be introduced in order to reveal the condition of the structure. The classification of damage is a pattern recognition problem and is treated here as one of the machine learning

family. The classification of a selected feature as abnormal or not is typified by two different approaches: supervised learning or unsupervised learning (novelty detection in this context). In terms of the SHM field, supervised learning means any procedure of classification of a feature which is trained with measurements representing and labelled by all conditions of interest. At the first level this is translated simply into separation between the damaged and undamaged condition of the structure. At higher levels, via supervised learning, identification of different types of damage or localisation of damage can be obtained. In several damage detection approaches it is never possible or it is very difficult to obtain true measurements for all possible damage classes especially in high-value or complex structures such as composite systems. Furthermore, data that is collected during a damaged state of structure is very rare.

The premise of novelty detection techniques is to seek the answer to a simple question; given a newly presented measurement from the structure, does one believe it to have come from the structure in its undamaged state? Through the possession of data assured to be from the normal, undamaged condition of the structure, one can generate a statistical representation of the training data. After this training procedure, any generated data from the system can be tested and compared to the undamaged model; any suspicious deviation from normality can be said to be indicative of damage. The advantage of novelty detection is clear; any abnormality defines a new situation characterised by a truly new event for the structure.

Feature extraction is analytically described in Chapter 3 and is generally focused on dimension reduction techniques via linear, nonlinear and probabilistic principal component analysis.

Other approaches regarding feature extraction belong to *system identification* [60–64]. The basic concept of such methods is to seek to fit measurements to mathematical models or functions and through the obtained form these models can reveal useful features (like ARX, ARMA or NARX models). It is a signal processing technique that creates a relationship between an input and output signals to reproduce the equation of dynamic motion of a system. Commonly for novelty detection, a way that system identification can be applied is by using the residual error of a predictive model as a feature.

2.5   Conclusions

The general premise of SHM was presented in this introductory chapter including the basic elements of this research in order to be applied in practice. SHM is a field of research with increased interest due each major advantage of operating continuously and globally but is not a market technology yet. This is the reason that major challenges are yet to be discovered and solved. The "holy grail" of these challenges is the development of a robust and online SHM system that is capable of detecting early critical fault types during the structure operation independently of changing environmental and operational conditions.

This thesis undertakes a serious attempt to apply SHM technology in a solid and fast manner by investigating how to simplify complex machine learning approaches into a more simple form and determine a robust way of how to uncover the often encountered problem of the influence of external operational factors.

Auto-associative neural networks: a misunderstanding and novelty detection

The main body of this chapter is concerned with the investigation of a paradox. Auto-associative neural networks consisting of five layers have been used as an advanced method for novelty detection in the past. In this study, an analysis is performed in order to demonstrate the ability of nonlinear auto-associators with three layers for multimodal classification problems and novelty detection. The investigation of auto-association with only three-layer neural networks is of interest in terms of computational power and complexity. Complexity is directly connected with the architecture of a network (number of neurons, avoidance of overfitting, generalisation). In simple terms fewer layers means it is faster and simpler to optimise the neural network.

In this chapter an analysis of dimensionality reduction methods is described from the simplest technique to more complicated ones as this is essential for the understanding of later chapters. A catholic argument is that no sensor exists that can directly measure any type of novelty. For this reason feature extraction is used to derive useful metrics from the raw data that can further be post-processed through advanced signal

processing tools. The basic aim of using such features is to lower the dimensionality of the data as the raw measurements, like time histories, typically have high dimensions, making the assessment of the raw data impossible in practice. In machine learning community this drawback is referred as "curse of dimensionality".

3.1 Linear and Nonlinear data mapping

One of the simplest methods for dimensionality reduction is the selection of a subset of the data. As seen in previous works [9, 65–68], this could be proved a very successful method offering sensitive features in detecting novelty. Of course, this kind of feature selection is driven by certain criteria. Generally, any selection of a feature should be followed by a criterion which checks if this specific selection is better than another. Second, a systematic procedure should be applied that is searching through the whole range of the data. This procedure could be simple such as visual inspection or more complicated like a genetic algorithm or sequential search techniques. The search procedure could be more intensive by searching all possible subsets of features in order to be more objective. Realistically, this kind of exhaustive feature selection is impractical as it requires major computational resource.

A reduction in the dimensionality by mapping the data from high-dimensional spaces to lower-dimensional spaces is accompanied by loss of some information. Therefore, the goal in dimensionality reduction should be to preserve as much relevant information as possible. In this chapter a series of techniques (that are later used) combine the input variables and lead to lower-dimensional representation. All of these methods follow an unsupervised learning procedure. These unsupervised learning techniques for dimensionality reduction can rely either on linear or nonlinear transformations. The discussion begins with the most common linear methods such as Principal Component Analysis (PCA) and continues with more complex techniques such as Nonlinear Principal Component Analysis (NLPCA).

3.2 Principal Component Analysis (PCA)

Principal Component Analysis takes a multivariate data set and maps it on to a new set of variables called "principal components", which are linear combinations of the

old variables. The first principal component will account for the highest amount of the variance in the data set and the second principal component will account for the second highest variance in the data set independent of the first, and so on. The importance arises from the fact that in terms of mean-squared-error or reconstruction it is the optimal linear tool for compressing data of high dimension into data of lower dimension. The unknown parameters of the transformation can be computed directly from the raw data set and once all parameters are derived, compression and decompression are small operations based on matrix algebra [1, 69–71]. One has,

$$[X] = [K][Y] \tag{3.1}$$

Where $[Y]$ represents the original input data with size $p \times n$, with p number of variables and n the number of data sets, $[X]$ is the scores matrix of reduced dimension $q \times n$ where $q < p$ and $[K]$ is called the loading matrix. The columns of $[K]$ are the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of $[Y]$. The original data reconstruction is performed by the inverse of equation (3.1):

$$[\hat{Y}] = [K]^T[X] \tag{3.2}$$

The information loss of the mapping procedure is calculated in the reconstruction error matrix:

$$[E] = [Y] - [\hat{Y}] \tag{3.3}$$

For further information on PCA, readers are referred to any text book on multivariate analysis (examples being references [69, 70]).

The main two disadvantages of the classical PCA algorithm is that there is not a generative model as there is no density model and as a result no principled interpretation of the error surface. Furthermore, when the data is of very high dimensionality with a relatively small number of observations there are accuracy and data scarcity problems, making the usage of the classic covariance matrix approach weak and significantly low in speed of processing. The second significant drawback is that PCA is limited by its nature of being a linear technique. It may be therefore unable to capture complex nonlinear correlations. Regarding the first problem a solution can be applied such as Probabilistic Principal Component Analysis (PPCA) and regarding the second one, a series of methods will be described such as Auto-Associative Neural Networks (AANNs).

3.3 Probabilistic Principal Component Analysis (PPCA)

The usage of the Expectation-Maximisation (EM) algorithm (appendix B) for computing PPCA offers a series of advantages discussed in [69–74]; the main ones are: the introduction of a likelihood density model, the algorithm does not need to compute the sample covariance matrix, it can deal with missing data, it is a relatively simple and fast calculation of eigenvectors and eigenvalues when dealing with a large number of samples and high dimensions and, through some additions, can be extended to mixture PPCA models with Bayesian inference algorithms that can calculate a complete Gaussian probabilistic model and produce true likelihoods [69–74]. For the sake of completeness only a brief description is given of the implementation of PPCA as it is much more fully described in the following works [69–74]. Classical PCA can be transformed into a density model by using a latent variable approach, similar to factor analysis, in which the data $\{x\}$ is calculated from a linear combination of variables $\{z\}$ [69–74] via the form,

$$\{x\} = [W]\{z\} + \{\mu\} + \{\epsilon\} \tag{3.4}$$

Where $\{z\}$ has a zero mean, unit covariance, Gaussian distribution $N(0, I)$, $\{\mu\}$ is a constant (whose maximum likelihood estimator is the data mean), and $\{\epsilon\}$ is an independent noise parameter. As the main objective is the reduction of dimensions, the latent variable dimension q is chosen to be smaller than the dimension p of the data. In the PPCA procedure there is a symmetric component in the data plus an independent error term for each variable with common variance and, by assuming a noise model with isotropic variance, one can write the covariance matrix of the data as $[\psi] = \sigma^2[I]$. The probability model of PPCA can be written as a combination of the conditional distribution:

$$p\left(\{x\}|\{z\}\right) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp(-\frac{\{x\} - [W]\{z\} + \{\mu\}^2}{2\sigma^2}) \tag{3.5}$$

And latent variable distribution

$$p(\{z\}) = \frac{1}{(2\pi)^{q/2}} \exp(-\frac{\{z\}^T\{z\}}{2}) \tag{3.6}$$

By integrating out the latent variable z, one can obtain the marginal distribution of the observed data which is also Gaussian, $x \sim N(\mu, C)$, where $[C] = [W][W]^T + \sigma^2[I]$. In turn this model represents the data in the lower dimension space. To fit this model to the actual data, one can use the log-likelihood as an error index,

$$\mathcal{L} = \sum_{n=1}^{N} \log p\left(\{x_n\}\right) = -\frac{N}{2}\{d\log\left(2\pi\right) + \log\left|[C]\right| + tr\left([C]^{-1}[S]\right)\} \qquad (3.7)$$

where

$$[S] = \frac{1}{N}\sum_{n=1}^{N}\left(\{x_n\} - \{\mu\}\right)\left(\{x_n\} - \{\mu\}\right)^T \qquad (3.8)$$

is the sample covariance matrix of the observed data, provided that $\{\mu\}$ is set by maximum likelihood estimation, which in this case is the sample mean. Estimation of $[W]$ and σ^2 can be computed by the iterative maximisation of $\mathcal{L}$ using an EM algorithm [69–74]. This is the approach which is used in this study by adopting the sensible principal component analysis method as introduced by Roweis [74] and is based on software by J.J.Verbeek [72].

3.4 Neural networks

Neural networks are a well established class of algorithm and for deeper inside information, readers are referred to [69–71]. Briefly, the group of multi-layer neural networks consist of a series of connected elements called nodes (or neurons in biological terms), organised together in layers. Signals pass from the input layer nodes, progress forward via the network hidden layers and finally reach the output layer. Each node with index i is connected to each node with index j in its proceeding layer through a connection weight w_{ij}, and similarly to nodes of the following layer. The mechanism of signal processing via each node is as follows: a weighted sum is computed at node i of all signals x_j from the proceeding layer, giving the excitation z_i of the nodes; this then passed through a nonlinear activation function f to emerge as the output of the node x_i to the next layer:

$$x_i = f\left(z_i\right) = f\left(\sum_{j} w_{ij}x_j\right) \qquad (3.9)$$

Different choices for the function are available such as the hyperbolic tangent function or logistic function which are mainly used in this study. Throughout the network architecture, a bias node is used and it is important because of its connection to all other nodes in the hidden and output layer in order to perform constant offsets in the excitation z_i of each node. The first step of employing a neural network is to set suitable values for the connection weights w_{ij}. This is usually called the training/learning phase. At each training step, a set of inputs is passed through the network giving trial outputs which are then compared to the actual set of output. Regarding if the comparison error is small enough or not, the network may continue the training procedure via a back-propagation algorithm, where the error is passed backwards via the network and the weights are re-adjusted until a desired error is achieved. The simplest method of adjusting the weights is by a gradient descent algorithm.

$$J(t) = \frac{1}{2} \sum_{i=1}^{n^{(l)}} (y_i(t) - \widehat{y_i}(t))^2 \tag{3.10}$$

$$\Delta w_i = -\eta \frac{\partial J}{\partial w_i} = -\eta \nabla_i J \tag{3.11}$$

$$w_{ij}^{(m)} = w_{ij}^{(m)}(t-1) + \eta \delta_i^{(m)}(t) x_j^{(m-1)}(t). \tag{3.12}$$

Where $\delta_i^{(m)}$ is the error in the output of the $i^t h$ node in layer m, t is the index for the iteration procedure, $J(t)$ is a measure of the network error and $n^{(l)}$ is the number of output layers. Equation (3.11) represents a standard steepest-descent algorithm where an adjustment of the network parameters is performed. The error is not known beforehand but is constructed form the previously known errors (3.13). When the update is performed there is often an introduction of an additional momentum term which allows previous updates to continue (3.12):

$$\delta_i^{(l)} = y_i - \widehat{y_i} \tag{3.13}$$

$$\Delta w_{ij}^{(m)} = \eta \delta_i^{(m)}(t) x_j^{(m-1)}(t) + \alpha \Delta w_{ij}^{(m)}(t-1) \tag{3.14}$$

The coefficients η and α determine the speed of learning in the gradient descent algorithm and are called respectively the learning and momentum rates. There are many different learning algorithms beyond the classic gradient descent algorithm for training a neural network. The ones used here are the Levenberg-Marquardt method, which is very fast and suitable for regression analysis and scaled conjugate gradients whose memory requirements are relatively small, as it is much faster than standard

gradient descent methods. For a detailed description of the algorithms the reader can refer to [69, 70, 75]. Here brief descriptions of the Levenberg-Marquardt and scaled conjugate gradient methods are given.

3.4.1 Levenberg-Marquardt and Scaled Conjugate Gradient methods

The Levenberg-Marquardt algorithm was specifically designed in order to minimise the sum-of-squared-errors. Consider the general equation of the sum-of-squares errors as [70]:

$$E = \frac{1}{2} \sum_n (\varepsilon^n)^2 = \frac{1}{2} \|\{\varepsilon\}\|^2 \tag{3.15}$$

where ε^n is the error for the n_{th} variable, and $\{\varepsilon\} = \{\varepsilon^1 ... \varepsilon^n\}$. As the training evolves, if the difference $\{w'\} - \{w\}$ (where $\{w\}$ is the old weight vector and $\{w'\}$ is the new weight vector) is very small then by utilising Taylor series and keeping only the first order terms one can write:

$$\{\varepsilon(\{w'\})\} = \{\varepsilon\}(\{w\}) + \frac{\partial \varepsilon^n}{\partial w_i}(\{w'\} - \{w\}) \tag{3.16}$$

And as a result, the error function can be re-written as:

$$E = \frac{1}{2} \left\| \{\varepsilon(\{w\})\} + \frac{\partial \varepsilon^n}{\partial w_i}(\{w'\} - \{w\}) \right\|^2 \tag{3.17}$$

The next step is to minimise this error function with respect to the $\{w'\}$ and one gets:

$$\{w'\} = \{w\} - \left(\left(\frac{\partial \varepsilon^n}{\partial w_i}\right)^T \frac{\partial \varepsilon^n}{\partial w_i} \right)^{-1} \left(\frac{\partial \varepsilon^n}{\partial w_i}\right)^T \{\varepsilon(\{w\})\} \tag{3.18}$$

The update formula given in equation (3.18) can be applied in an iterative fashion in order to minimise the error. The problem that arises is that step size in equation (3.18) could be large and as a result there is no linear approximation. This critical step is addressed in Levenberg-Marquardt by minimising the error function by keeping the step size relatively small in order to ensure a linear approximation. Thus an extra term is added to equation (3.17) giving:

$$\widetilde{E} = \frac{1}{2} \left\| \{\varepsilon(\{w\})\} + \frac{\partial \varepsilon^n}{\partial w_i}(\{w'\} - \{w\}) \right\| + \lambda \|\{w'\} - \{w\}\|^2 \tag{3.19}$$

where λ controls the step size (is a regulariser). If again one minimises the error function with respect to the $\{w'\}$, one obtains:

$$\{w'\} = \{w\} - \left(\left(\frac{\partial\varepsilon^n}{\partial w_i}\right)^T\frac{\partial\varepsilon^n}{\partial w_i} + \lambda\mathbf{I}\right)^{-1}\left(\frac{\partial\varepsilon^n}{\partial w_i}\right)^T\{\varepsilon(\{w\})\} \qquad (3.20)$$

If values of λ are relatively large the updated equation (3.20) leads to the classical gradient descent method (if very small it leads to the Newton algorithm).

The Levenberg-Marquardt algorithm is a technique where the error function approximation operates within a region around the search point. The size of this region is directly controlled by the value of λ.

The conjugate gradient approach can be considered as a type of gradient descent method, in which the parameters η (determines the step length) and α (fraction of the previous step that is included to the next step) in equation (3.14) are automatically determined at each run. The conjugate gradient algorithm is a line search method. In line search methods a search direction is determined in weight space and as a second step the minimum of the error function is evaluated along that direction. These types of algorithms are much more powerful than the classical gradient descent method.

The general conjugate gradient is however followed by a series of critical drawbacks. In every line search a considerable number of error functions have to be computed. These intensive calculations evolve computationally challenging tasks. The line search procedure is driven by a number of parameters that determine the termination for each line search. As a result, the performance of the method is very sensitive to these values.

Moller [75] introduced the Scaled Conjugate Gradient algorithm which overcomes the line search method of the conventional conjugate gradient. This algorithm consists of too many complex steps to be explained in detail here, but the basic idea, is a combination of the model-trust region approach (used in the Levenberg-Marquardt algorithm), with the conjugate gradient approach. Model-trust region methods are techniques that are effective only around a small region of the search point. Technical details regarding the general conjugate gradient algorithm can be found in [69, 70, 75].

3.4.2 Auto-Associative Neural Networks (AANNs)

Multi-layer neural networks can be used to perform nonlinear dimensionality reduction and overcome some of the drawbacks of linear PCA [1, 5–7, 70, 76, 77]. Similarly to linear PCA, Nonlinear PCA (NLPCA) proceeds by adopting arbitrary nonlinear functions to seek a mapping generalising the equations (3.2),(3.3):

$$\{X\} = G(\{Y\}) \tag{3.21}$$

G is a nonlinear vector function, possibly consisting of a different number of individual nonlinear functions. The original data reconstruction is then performed by using another nonlinear function H:

$$\{\hat{Y}\} = H(\{X\}). \tag{3.22}$$

The nonlinear functions H and G can be learned from data and encoded as neural network structures. Nonlinear Principal Component Analysis (NLPCA) is based on the architecture of a five-layer neural network in the form of Fig.3.1, including the input, mapping, bottleneck, demapping and output layers. As the targets used to train the neural network are simply the same as the inputs, the network is trying to map the inputs onto themselves.

This kind of neural network architecture is called an Auto-Associative Neural Network (AANN) [5, 7, 70]. It is of course, an unsupervised learning tool, since no independent classes of target data are provided. A restriction of the mentioned topology is that the bottleneck layer must have less neurons than the input and output layers. The bottleneck structure forces the neural network to learn important features of the presented patterns; the activations of the bottleneck layer correspond to a compressed representation of the input. This kind of network can be viewed as two consecutive mappings M_1 and M_2. The first mapping M_1 projects the original data onto a lower dimensional sub-space defined by the activations of the bottleneck layer. Due to the first hidden layer of nonlinear transfer functions, this mapping is arbitrary and there is no linear restriction. The de-mapping M_2, which is the second half of the AANN is an arbitrary mapping from the lower dimensional sub-space back to the original space. It has to be noted that the neurons of the mapping and demapping layers must have nonlinear transfer functions in order to be able to encode the arbitrary functions G and H. Nonlinear transfer functions are not essential in the bottleneck
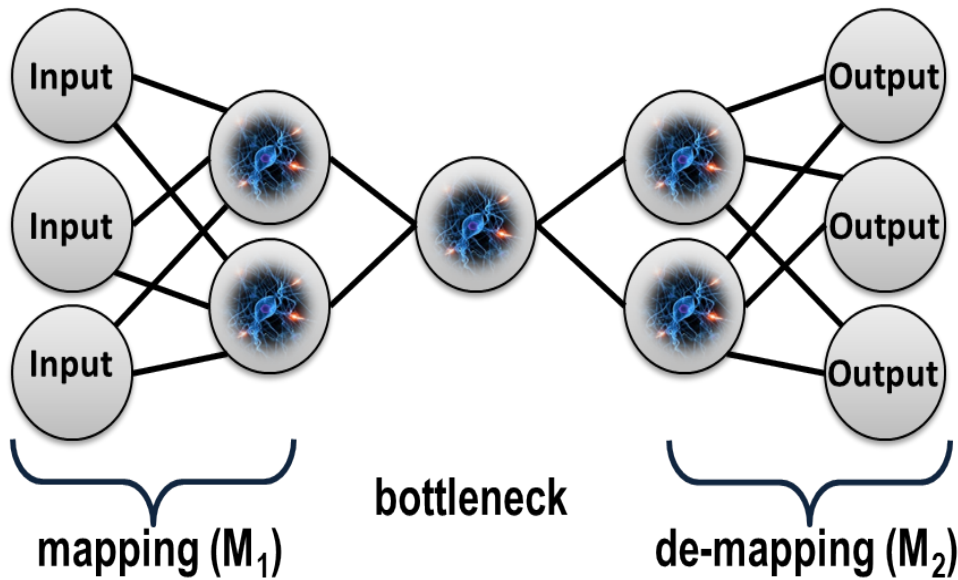
Figure 3.1: Auto-Associative Neural Network (AANN) architecture.

layer, since this central layer represents the output layer of G functions mapping modelling. However, if a bounded response [5, 7, 70] in the feature space is needed, nonlinear transfer functions can be utilised in all network nodes (something that is used in this study).

3.5 Radial Basis Functions (RBF) Networks

A brief description is given of RBF networks; for a more detailed analysis the reader is referred to the following works [69–71]. The main difference compared to the MLP concept is that instead of units that compute a nonlinear transfer function of the scalar product between the input data vector and a weight vector, in radial basis networks the activation of the hidden neurons is given by a nonlinear function of the distance between the input vector and a weight vector as in Fig.3.2. One can write a robust representation of the radial basis function network mapping as [71]:

$$y_k(t) = \sum_{j=1}^{M} w_{kj} f_{kj}(t) + w_{k0} \tag{3.23}$$

Where the f_{kj} are the basis function, w_{kj} are the output layer weights, w_{k0} is the bias weight, index M is a set of density functions labelled by the index j and t is the

distance (Euclidean) $\{x\} - \{c\}$ between the input vector $\{x\}$ and a centre vector $\{c\}$. Most of the time it is more convenient to absorb the bias weights into a summation by adding an extra basis function with a constant activation equal to one. In the current work, the basis function used is Gaussian although a number of different basis functions can be introduced eg. multiquadratic, inverse multiquadratic, cubic approximation functions and thin-plate-spline. The form of the Gaussian is:

$$f\left(t\right) = e^{-\left(\frac{t^2}{2r_i^2}\right)} \tag{3.24}$$

If it is assumed that the centres and radii r_i are fixed, the weights can be computed by a pseudo-inverse of the matrix formed from training data, alternatively, backprop-agation as in MLP training can be implemented via iterative methods. In general the most challenging task with RBF networks is to identify the centres and radii for the Gaussian distributions, as this is a nonlinear optimisation problem. The most efficient way of setting the centres is to fit a Gaussian mixture density model to the data through the expectation-maximisation algorithm (EM) and this is the approach used for this study. It has to be mentioned that in the current study the RBF network does not use logistic or softmax outputs at the output layer as the speed advantage of RBF networks compared to MLP networks will be lost [69–71]; here, linear output neurons are used. The expectation-maximisation (EM) algorithm is a well-defined algorithm and is not described in this study but the reader can be referred to the earlier references and Appendix B. As mentioned above, the main advantage of RBF networks compared to MLP networks is that they do not usually need a full and challenging nonlinear optimisation of all parameters. The total training time in all cases considered in this study was between 10-20 seconds.
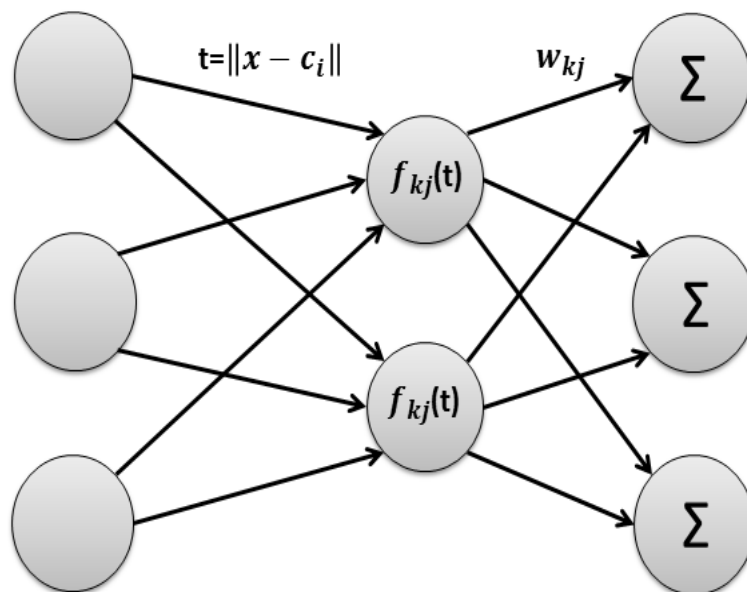
Figure 3.2: Radial Basis Functions (RBF) Networks architecture.

3.6 Data pre-processing and overfitting of a neural network

In this work, the data are transformed in terms of "standardisation" before applying the PCA or RBF networks and are "rescaled" before applying to the MLP. Standardising a vector in this case means subtracting the mean and dividing by the standard deviation. Standardising the input is of critical importance not only before applying PCA, but mainly because of the usage of a method such as an RBF network. In RBF networks the input variables are combined, as mentioned above, through a distance computation such as Euclidean distance between the input vector $\{x\}$ and a centre vector $\{c\}$. The contribution of each variable will be significantly depend on its variability compared to other variables of the input space. Regarding the MLP, rescaling the input is important for a series of practical reasons. The main reason is to avoid saturation of sigmoid transfer functions and as a result the usage of small initial random values. The data were scaled between [-1,1] or [0,1] before entering the input to the auto-associative neural network. It has to be noted that the same statistics and rescaling factors that were applied to the training data must be used also for the test or validation sets.

The dimension of the data remains a major challenge as the novelty detection

technique suffers from the "curse of dimensionality". Even if technologically the power was available for such performance when the number of observations is much smaller than the dimensions, then overfitting would be the dominant problem as the neural network could focus only on local regions of the training data, resulting in a very poor tuning of network [78]. An early-stopping criterion was implemented in order to avoid overfitting problems and achieve a better generalisation [9, 69–71]. The problem of overfitting is essentially the problem of rote learning the training data rather than learning the actual underlying function of interest. In simple terms it occurs when there are too many parameters in the network compared to the training observations or patterns. For the purposes here, a percentage of the input data was used for validation purposes and testing. The mean-square-error on the validation set is examined during the training procedure. In the case of overfitting the validation error normally begins to rise compared to the training set error. In the early-stopping technique, the right choice of the validation data is vital. The validation set should be typical of all different points in the training input. For this reason, a random sampling was used for choosing each data set. The training set was corrupted several times with different Gaussian noise vectors of different r.m.s. values, in order to tune correctly the unknown parameters (weights, bias). It is a good idea to train the network starting from multiple random initial conditions and by testing it robust network performance can be achieved. In real life, the obtained data most of the times introduce limitations on the actual number of neurons in the hidden layers. This is the second reason that the normal data used for the tuning of the networks is usually corrupted here several times with Gaussian noise.

Explicit criteria taking into account the compromise between the accuracy and the dimension of the layers can be applied [10, 69, 70]. Two classic criteria are Akaikes Final Prediction Error (FPE) and An Information theoretic Criterion (AIC), and their implementation is a trivial task.

$$FPE = \frac{a\left(1 + \frac{M_{total}}{M}\right)}{1 - \frac{M_{total}}{M}} \tag{3.25}$$

$$AIC = \ln(a) + 2\frac{M_{total}}{M} \tag{3.26}$$

$M_{total} = (n + d + 1)(L_2 + L_3) + n + d$ is the total number of weights, d is the dimension of the bottleneck layer, L_2 and L_3 the dimensions of the mapping and demapping layers, $M = pn$ is the total number of observations and $a = \frac{MSEsum}{2M}$ where $MSEsum$ is the summation of the mean squared errors. By trying different

numbers of dimensions in the hidden layer, one can find by minimising these criteria the optimal network without overfitting drawbacks. In this current study several different topologies were tested for each sensor and by applying the early-stopping criterion, regularisation and the later criteria, the best AANN architecture was found.

3.7 The Paradox

In Kramer's [7] work there is an argument which indicates a requirement of three hidden layers (so five layers in total) in order to accurately perform a nonlinear auto-association as the elimination of the mapping and demapping layers introduces severe technical drawbacks. As it is stated, a network with only one hidden layer transforms it to a bottleneck layer between inputs and targets. And if the transfer functions of the bottleneck layer are linear, this leads to linear PCA (as shown by Sanger [76]). This argument is carried a step forward by stating that even if the bottleneck nodes are nonlinear functions, the mapping capability of G and H would still produce linear combinations between the inputs and outputs, just compressed by the nonlinear nodes. It has been concluded that the performance of an autoassociative neural network with only one internal layer of nonlinear nodes performs similarly to a linear PCA.

Cottrell & Munro [12] claimed a similar argument with Kramer that for auto-association with a single hidden layer with linear output units, the optimal weight values can be derived by standard linear algebra, and therefore that the usage of nonlinear transfer functions at hidden layers may be pointless.

On the other hand, nonlinear auto-associators were widely used for their ability to solve problems that cannot be solved by SVD because of the singularity of the PCA characteristic.

A close look at a particular condition was assumed in [5, 7, 76] and leads to the investigation of a paradox. For the nonlinear transfer function $F(x)$ it is observed that if the values of input x are small enough then the nonlinear processing function $F(x)$ can be approximated by the linear part $F(x) \rightarrow a_0 + a_1 x$ arising from its power series expansion. This approximation automatically leads to a result where the hidden unit activations prior to transformation must be in the linear range of the $F(x)$ function. This in turn leads to the suggestion that when the network

inputs do not fall in the linear range of the transfer function they do not inevitably generate PCA. Furthermore, while SVD-PCA represents a unimodal reconstruction error surface by calculating a global solution to the problem, the nonlinear transfer functions can "comprehend" local valleys to the problem.

Japkowicz et al.[13] described and compared auto-associative neural networks with different topologies where interesting results can be found.

In this study a comparison of classic auto-associative neural networks, single hidden layer auto-associators and Principal Component Analysis (PCA) is performed in the context of novelty detection. It will be demonstrated that in the context of novelty detection, single layer auto-associators are not performing in a similar fashion with PCA. Although they can not fully reconstruct highly non-linear surfaces, their ability to catch some of the non-linearities of the variables is noticeable.

3.8 Representation of error surface

In order to give a live image of the the paradox described previously (supporting the results derived from [13]) and how important it is in the case on novelty detection, a two dimensional simulated problem with 200 sample points (2×200) is created which is a challenging nonlinear correlation between the two dimensions as shown in Fig.3.3. The data points were corrupted with Gaussian noise of r.m.s value 0.01. The pseudocode equations describing the simulated data are:

$$
\begin{aligned}
x_i &= (\text{alog(i/(samplepoints)))sin}(0.1 + 2\text{pi(i/(samplepoints)))} \\
y_i &= (\text{alog(i/(samplepoints)))cos}(0.1 + 2\text{pi(i/(samplepoints)))}
\end{aligned}
\tag{3.27}
$$

where a is just a constant scaling factor and x, y are the two dimensions. The reconstruction error was calculated based on the equation:

$$
res(x, y) = (residualerror(x))^2 + (residualerror(y))^2
\tag{3.28}
$$

The plots in Figs.3.4-3.7 describe the reconstruction error surfaces. The four different types of MLP are presented, including the five layer MLP, the three layer MLP with
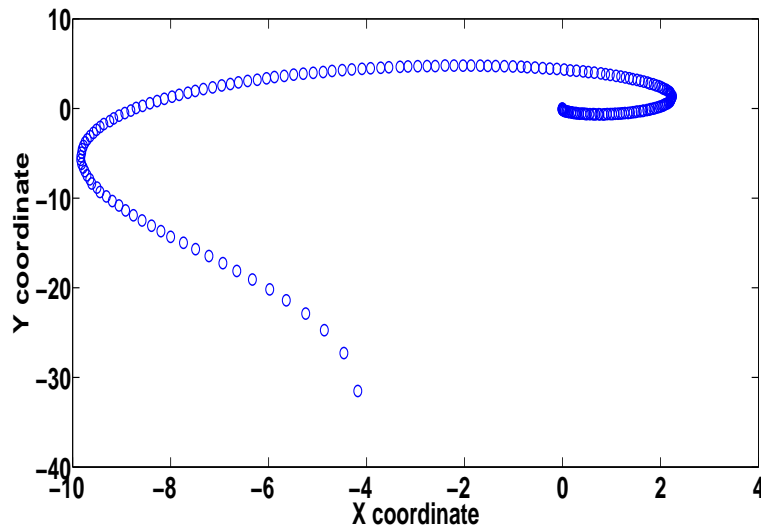
Figure 3.3: 2-D data shape.

nonlinear activations in all nodes, the three layer MLP with linear activations in the output and PCA in Figs.3.4-3.7 respectively. The $Z - axis$ is the res value at every point and $x, y - axis$ represent the replicated grid vectors of 200 points in order to produce a full 3-D grid. As the feature is a two-dimensional set, the dimension compression was to move to a lower space of one component. It can be noted that PCA reconstruction error reaches values near 400, while the two three-layer MLPs are 1 and 2 respectively. The five layer auto-associator performs much better with values around 0.2. Before visualising the original data against the actual outcome of the different types of the minimisation techniques a noticeable conclusion could be helpful. Except for the different values of the reconstruction error (res), the actual geometrical shape of the surfaces is interesting. The classic five-layer MLP exhibits a much more complicated surface compared to the other forms of MLPs. But the impressive point is the difference between the MLP with nonlinear activations in all nodes and the three layer MLP with linear activations in the output and PCA as it demonstrates a more interesting surface including extra peaks.

The plots in Figs.3.8-3.11 are particularly critical in understanding the nature of computing differences between the nonlinear auto-associators and PCA as they describe the original data against the actual outcome of the data reconstruction techniques after dimensionality reduction. It can be seen that despite the complex nonlinear body the five-layer MLP captures the data trend - something that is not surprising. In turn, the reconstruction is almost perfect. But this is something that was expected.

The main focus of this chapter is centred on the performance of the three-layer MLP against PCA in order to check the paradox described previously. The surprising observation is that none of the three-layer MLPs performs the same as PCA. In accordance to classic studies [7] it is confirmed that the five-layer MLP outperforms the single hidden layer MLPs in terms of auto-association. The three layer MLP with nonlinear transfer functions in all nodes (including the output layer) catches a nonlinear trend compared to PCA. This is a very encouraging result as it demonstrates an ability that can be proved a qualitative advantage in terms of novelty detection.

Both PCA and the three-layer MLP with linear activations in the output reproduce linearly the data but with a critical difference between them. The PCA reconstructed data is out of the original shape boundaries and this result explains why the reconstruction error reaches values around 400. On the other hand, the MLP with linear activations in the output reaches much lower values of the reconstruction error. This observation confirms that although in both of them there is a linear data reconstruction, their performance is not identical.

It is clear that PCA and MLPs with three layers do not perform the same and the paradox investigated in this study exists. Even these slight differences in their performance could be proved an extra key factor in terms of novelty detection. Indeed, PCA and MLP with linear activation functions in the output layer reconstruct the data linearly. But it has to be mentioned that this data distribution was challenging in terms of nonlinearity and usually features extracted from vibration responses do not demonstrate so strong nonlinear distribution shapes. In the next section this paradox and the distinct difference described here is more noticeable where novelty detection using these algorithms will be compared in structural damage terms.
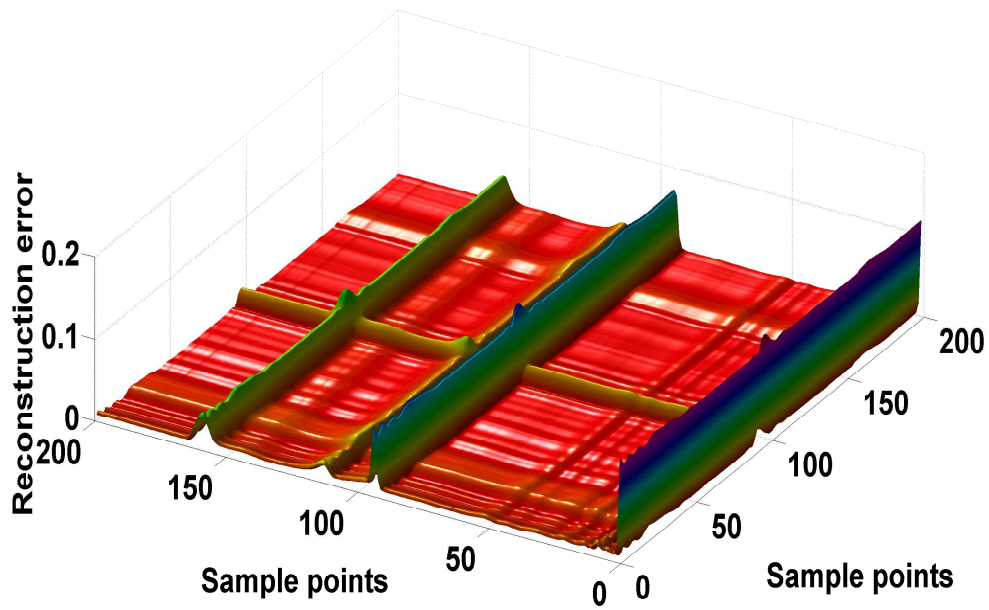
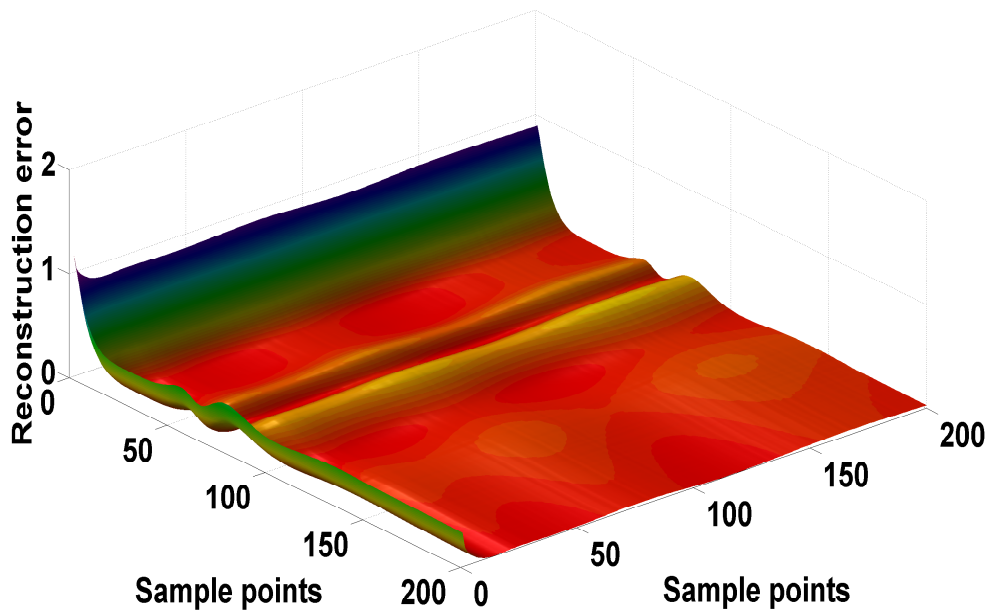Figure 3.4: Reconstruction error surface for five-layer AANN.

Figure 3.5: Reconstruction error surface for three-layer AANN with nonlinear activations in all nodes.
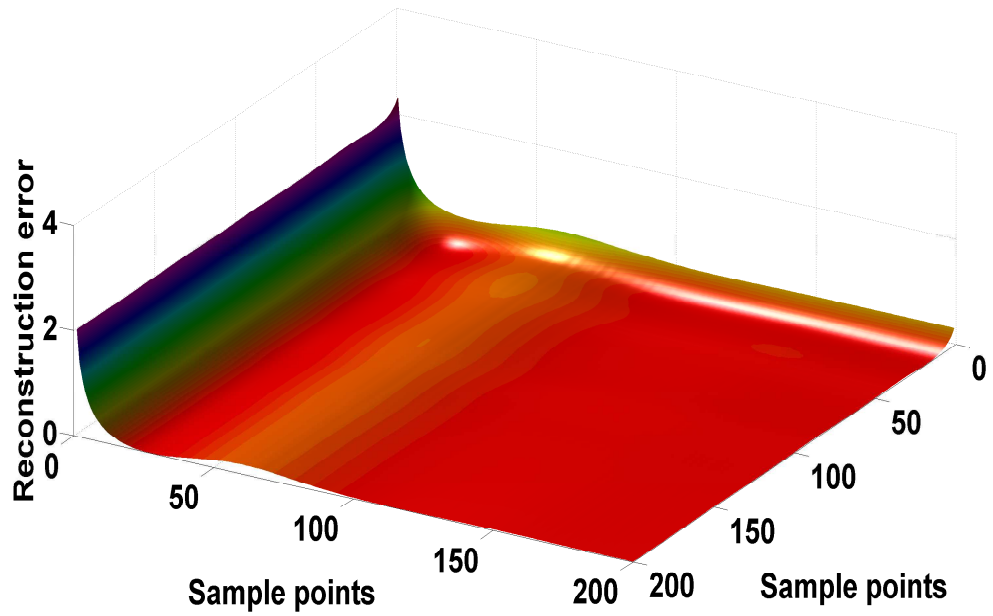
Figure 3.6: Reconstruction error surface for three-layer AANN with linear activations in output layer.
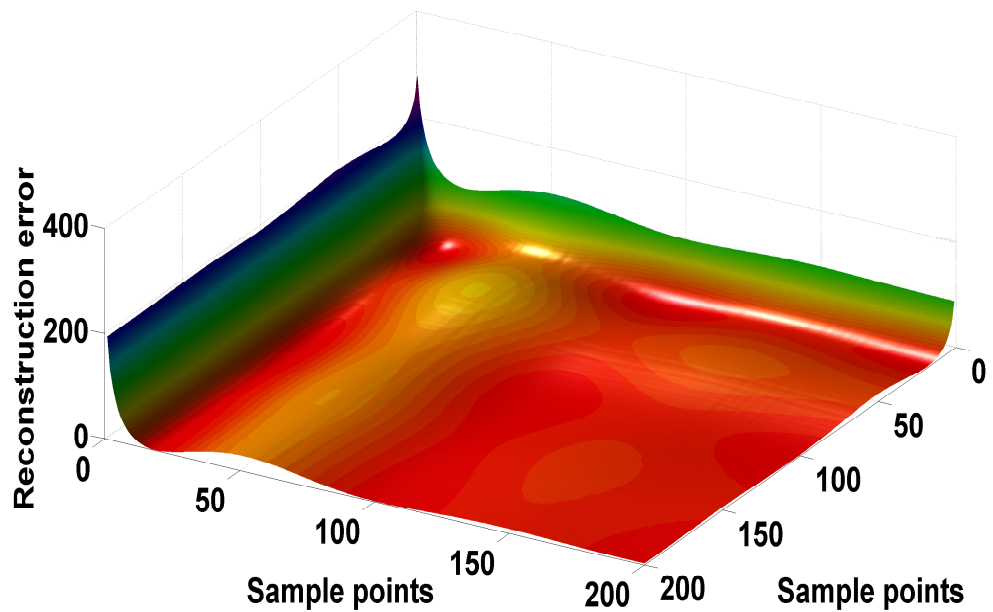

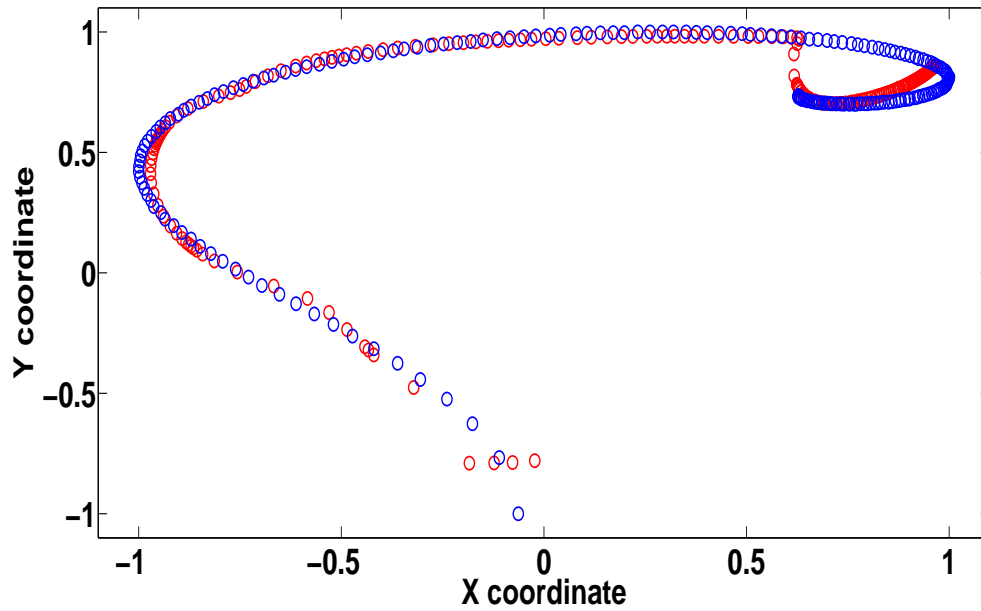Figure 3.7: Reconstruction error surface for PCA.
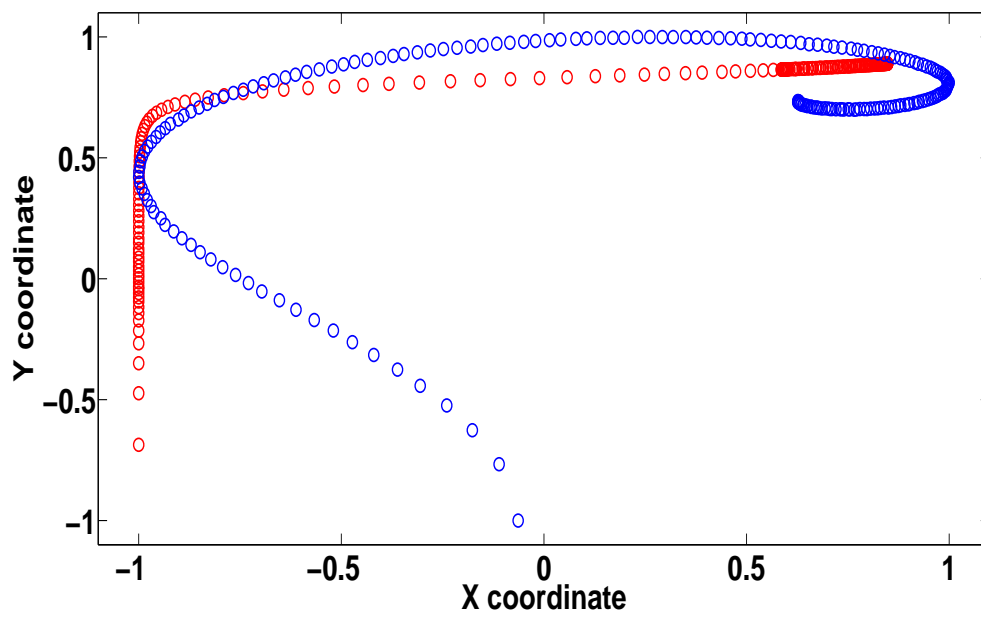
Figure 3.8: Reconstructed data for five-layer AANN.

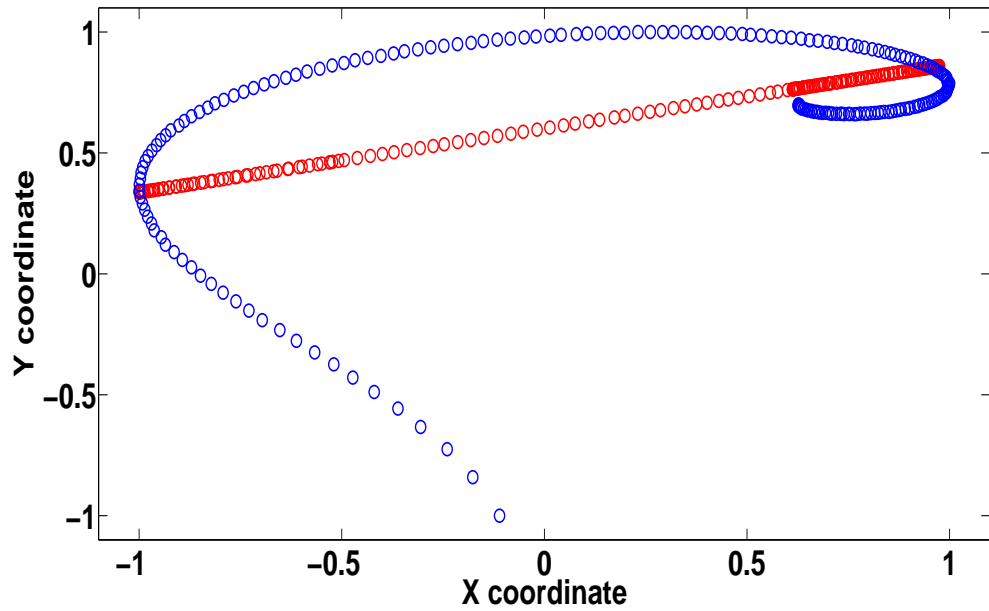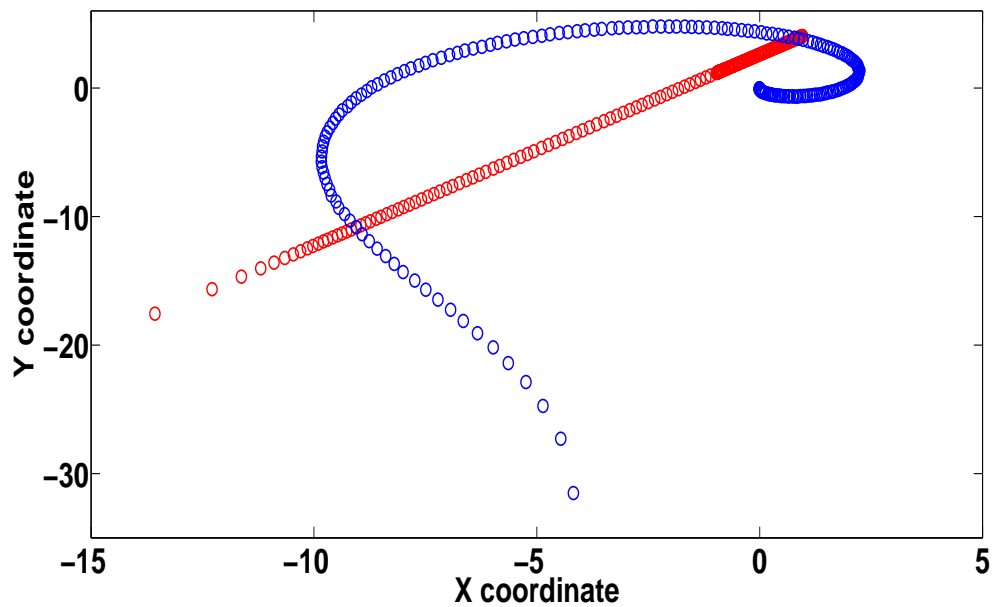Figure 3.9: Reconstructed data for three-layer AANN with nonlinear activations in all nodes.

Figure 3.10: Reconstructed data for three-layer AANN with linear activations in output layer.

Figure 3.11: Reconstructed data with one component for PCA.

3.9   Novelty Index and Alarm Threshold

The objective of this section is to demonstrate the technique of novelty detection
in the context of auto-associative neural networks (AANNs) and PCA in order
to further address the paradox described previously. Radial basis function (RBF)
networks which are a robust and simple statistical technique will also be used here
in the context of auto-association (and later will be used also as a density method
for novelty detection). When a trained AANN is fed with an input data set coming
from an unprecedented state of the structure a novelty index n described in the form
of Euclidean distance will increase:

$$n\left(\{y\}\right) = \{y\} - \{\hat{y}\} \tag{3.29}$$

Where $\{y\}$ and $\{\hat{y}\}$ are each initial output and network or PCA output vectors
respectively. If the neural network learning was successful then $n\left(y\right) \approx 0$ for all the
training data set. Later on in testing, $n\left(\{y\}\right)$ will significantly depart from zero
indicating the presence of novelty $n\left(\{y\}\right) \neq 0$. It is common in SHM and condition
monitoring to introduce a threshold in order to visualise clearly the presence of
abnormal readings. In the case of a novelty index the warning level is the threshold
value after which a reading value can be considered as an abnormal quantity to
involve further investigation. The warning levels can be defined as $\bar{v} + a\sigma$ where
$\bar{v}$ and σ are, respectively the mean and standard deviation of all the values of the
novelty index over the training data. In statistical terms a controls the percentage
of false positives. For example if the distribution is purely Gaussian then a value of
1.96 will give estimates within warning levels of 95% probability. In this paper a is
set equal to 3 giving a 99.7% confidence limit.

3.10   Multi-DOF simulated structure

The simulated system that is used in this section in order to demonstrate the objective
of this study, which is the paradox and novelty detection, is a ten degree-of-freedom,
nonlinear lumped parameter system as shown in Fig.3.12. The specific values of the
undamaged system were $m = 1$, $c = 20$, $k = 10^4$ and $k_{nl} = 10^9$ where m is the mass,
c is the damping coefficient, k and k_{nl} are characterising the linear and nonlinear
springs respectively. The fault in this system was simulated by reducing the stiffness

between masses 1 and 2 by 10% of the original value, in order to make it easier for PCA to compete with the neural networks. In order to implement the methods and generate an appropriate number of samples, the unfaulted and faulted data were copied 1000 and 200 times respectively. Each copy was subsequently corrupted with different Gaussian noise vectors of r.m.s. value 0.05%. The feature that was used for the detection process was the transmisibility function between masses 1 and 2 around 5 points of the first resonance frequency Fig.3.12. In Fig.3.13 the blue line represents the normal condition and the red line the faulted one. This was calculated by simulating the response of the system to a harmonic excitation. To make the PCA task even easier, four out of five principal components were used.
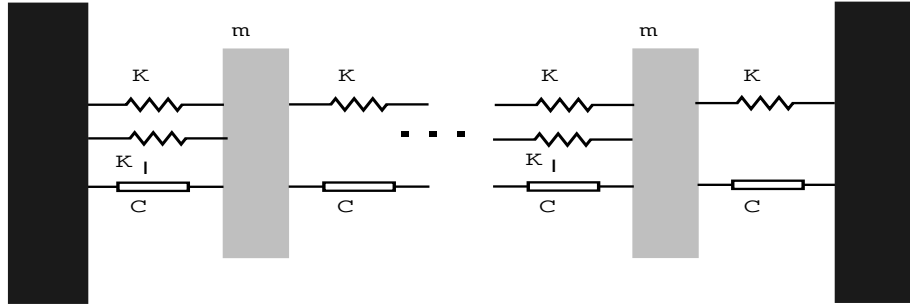

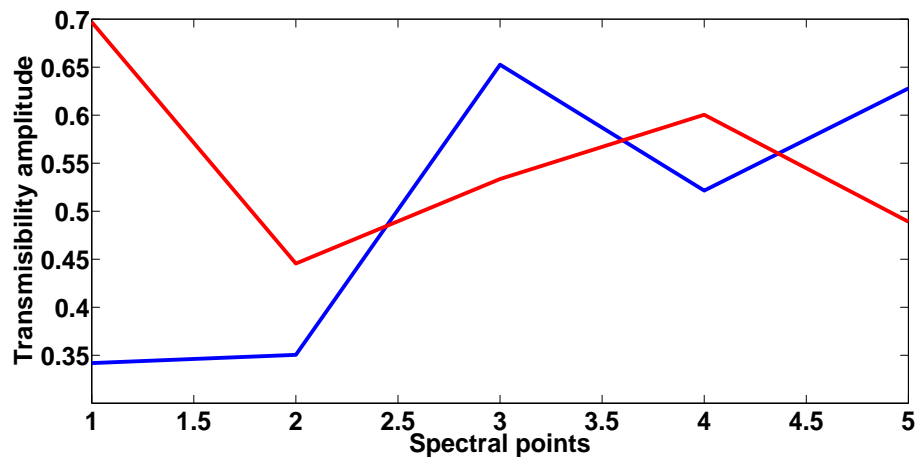Figure 3.12: Ten degrees-of-freedom lumped parameter system.

Figure 3.13: Transmisibility feature around 5 points of the first resonant frequency.

The existence of the paradox, and that PCA and MLPs with three total layers do not perform the same in terms of novelty detection is validated from the damage detection results derived in this section. In Figs.3.14-3.17 the novelty detection results for the five-dimensional feature can be seen where the damage is introduced after

observation 1000. To start, the first noticeable observation is that the three-layer auto-associator with nonlinear transfer function in all nodes (Fig.3.15) outperforms both PCA (Fig.3.17) and the MLP with linear nodes at the output layer (Fig.3.16). Also, the MLP with linear nodes at the output layer (Fig.3.16) outperforms PCA (Fig.3.17). Of course, the results regarding the single layer auto-associator with linear nodes in the output layer are not great but still the difference with PCA is considerable. The outliers detected after sample point 1000 are much more visible in MLPs compared to PCA.

As expected the five-layer MLP outperformed all of them, but the result that was not expected is the remarkable performance of the three-layer MLP in Fig.3.15 as it is in on a par with the classic five-layer MLP in Fig.3.14. Both single layer auto-associators with linear and nonlinear transfer functions in the output layer perform distinctively better than PCA, where no outliers are observed in the damaged case.
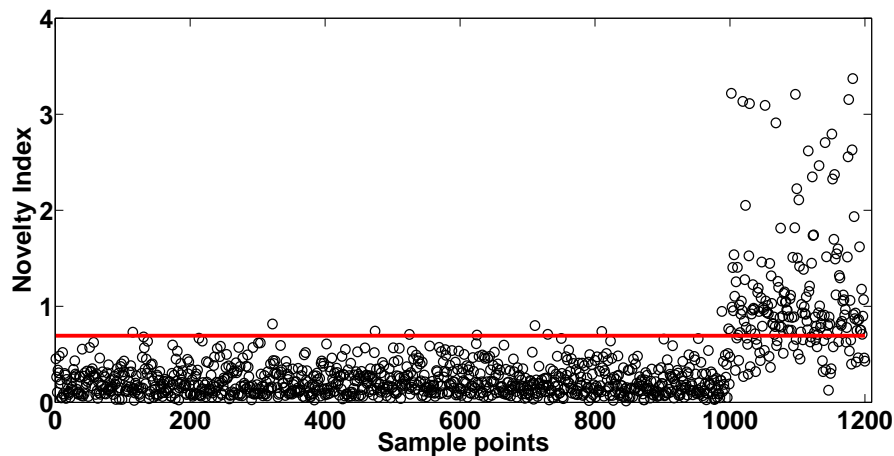

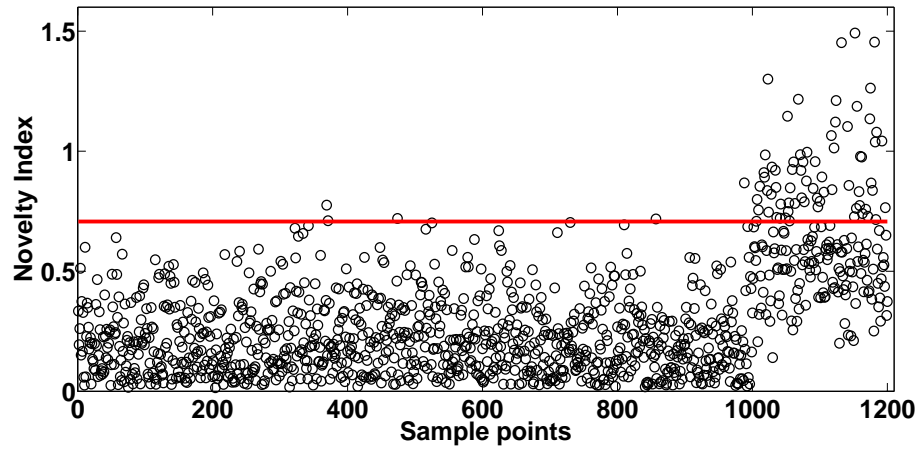Figure 3.14: Five-layer AANN novelty index

Figure 3.15: Three-layer AANN novelty index with nonlinear activations in all nodes.
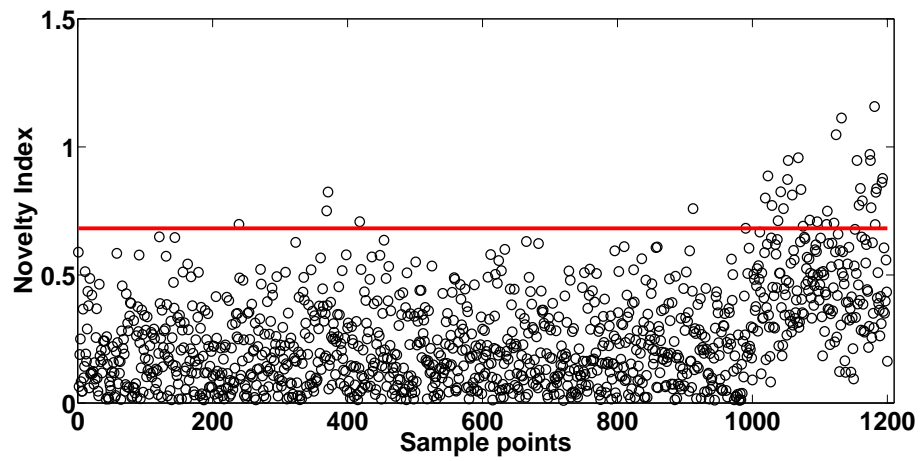

Figure 3.16: Three-layer AANN novelty index with linear activations in output layer.
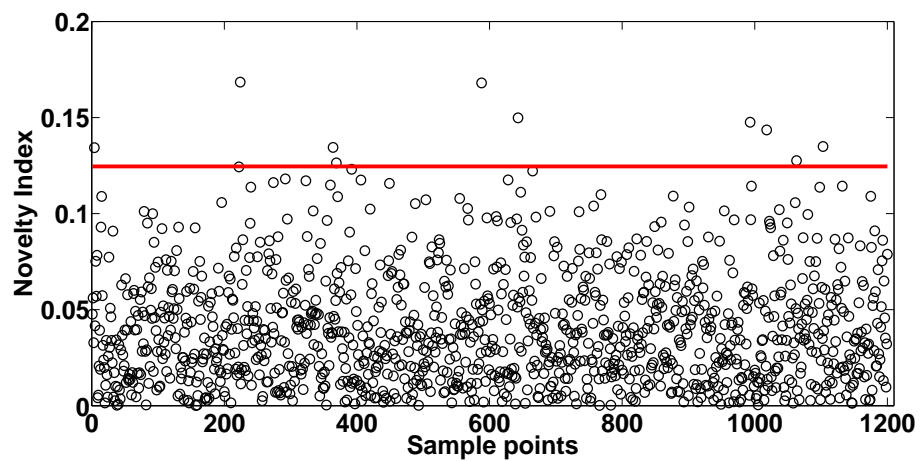

Figure 3.17: PCA novelty index.

3.10.1 RBF auto-association for novelty detection

Another novel approach to auto-association by using radial basis function networks (RBF) is demonstrated here. This means that as above and for fair comparison with MLPs and PCA only four nodes were kept in the middle layer. The main advantage of RBF networks compared to MLP networks is that they do not usually need a full and challenging nonlinear optimisation of all parameters. It is important to highlight that as RBF networks give a functional representation as a sum of local processing neurons; their sensitivity regarding the training vector is reduced. On the other hand, the MLPs compute a global approximation to a pattern mapping and this gives them the advantage to generalise to spaces distant from the training data, something that RBF architectures are unable to offer. RBF networks are significantly faster and computationally more efficient, giving them the advantage of fastest online inspection for simple novelty tasks.

As can be seen from the results presented in Fig.3.18 the RBF auto-associator performance is noteworthy as in comparison with the three-layer MLPs it is much more efficient and equal to the five-layer auto-associator (if not better). The damage is strongly detectable after observation 1000 with the majority of points exceeding the threshold.

This significant result can be explained by carefully observing Fig.3.19 where the RBF centres are projected onto the trained data distribution. The centres, despite their limited number (for the sake of auto-association), are following the input shape by covering the distribution boundaries. It could be said that the RBF auto-associator reacts as a reduced dimension density-based method. In general the most challenging task with RBF networks is to identify the centres and radii for the Gaussian distributions, as this is a nonlinear optimisation problem. In this current example, the distribution of the training data is clearly Gaussian as can be observed in Fig.3.19 making RBF networks perform even better as novelty detectors.
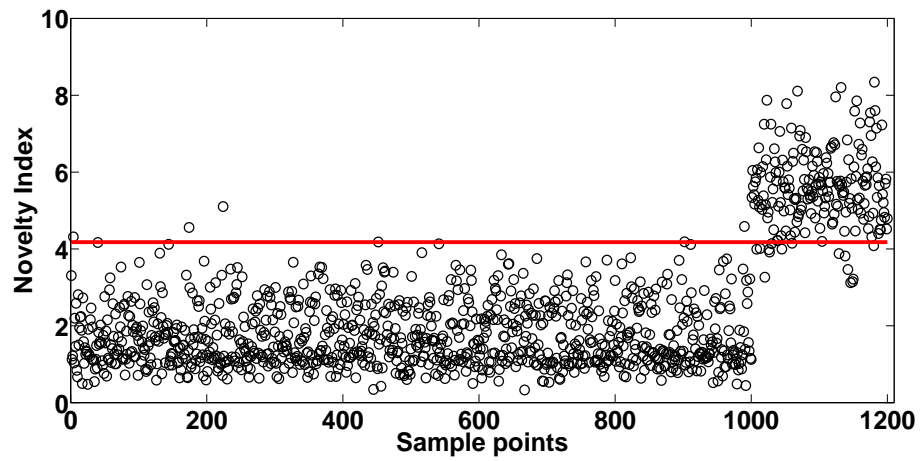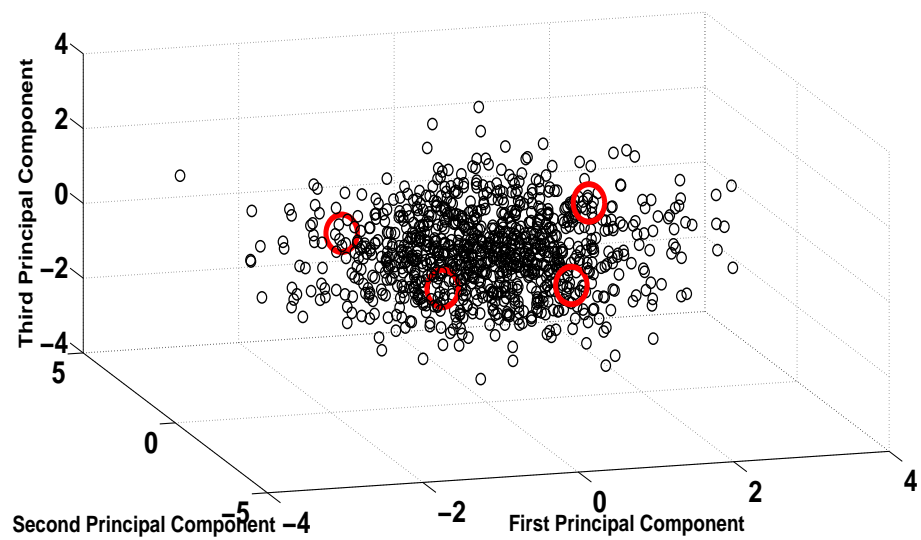
Figure 3.18: RBF-AANN novelty index.

Figure 3.19: RBF centres projection to normal data distribution.

3.11 Experimental validation

For the purposes of experimentally validating the paradox and the conclusions derived previously, the structure to be tested was a carbon fibre plate with a stiffening element. The geometry of the plate was 60cm x 14.8cm x 4mm and the panels lay-up sequence was (0/90/45/0/90), Fig.3.20. The carbon plate was suspended using soft springs to approximate free-free boundary conditions. It was decided to use Frequency Response Function (FRF) data in order to monitor the specimen. The sensor used was a single-axis piezoelectric accelerometer. The accelerometer was fixed with wax. The plate was excited using an impact hammer. The FRFs were measured using an LMS-DIFA SCADA III acquisition system controlled by LMS software. FRFs were measured in the range of 0-10240 Hz and processed using 4096 spectral lines, giving a frequency resolution of 2.5 Hz. It was anticipated that using hand impact excitation would introduce additional variability into the test relative to other methods of excitation e.g. via a shaker. However, given that the aim of the study is to diagnose damage in the presence of variability this was not deemed problematic.
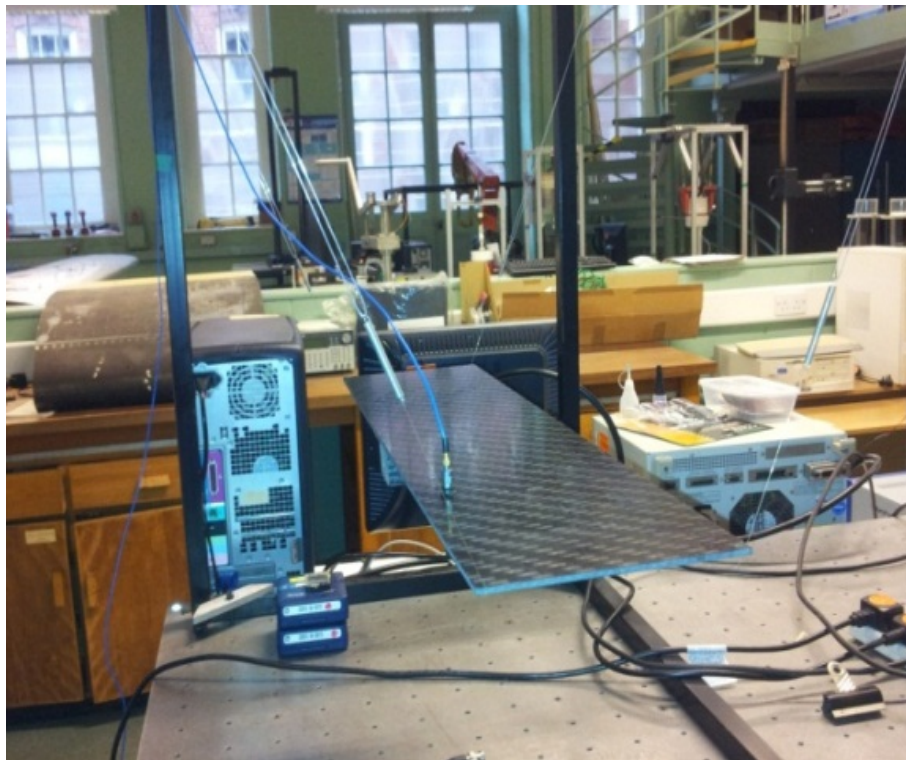

Figure 3.20: Carbon fiber reinforced polymer (CFRP) plate.

FRFs from a specific position were obtained with 5 averages for each of them as

shown in Fig.3.21. The position was chosen in order to avoid losing basic mode shapes (like torsional modes). Next, 180 measurements were repeated successively. The second series of tests involved introducing damage into the structure, and gathering data from the structure in its subsequent "damaged" conditions. An impact rig was used to apply three different levels of impact (15, 30 and 40 Joules) to the centre point of the plate. After each impact the specimen was removed from the impact rig and again placed in an approximation of a free-free condition. The same procedure of excitation with an impact hammer was applied in order to extract 60 repeated measurements of each faulty condition. These were used for the purposes of testing the novelty detector.
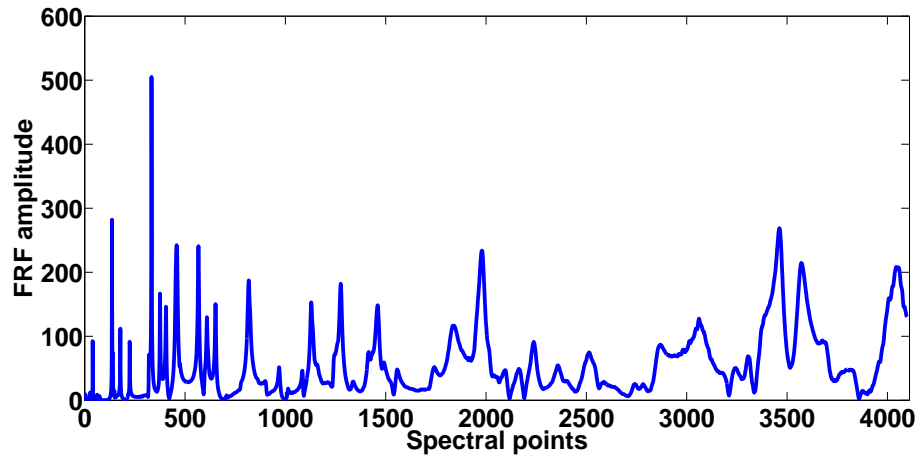

Figure 3.21: FRF spectrum.

One can introduce subsets of the data by introducing features which are sensitive to damage. The initial stage is to set up which features of the FRF spectrum will be used to individually detect damage in the specimen. The normal model condition is created by the data extracted for the undamaged situation. Later, subsequent data are compared to this undamaged model and any crucial differences are assumed to indicate damage presence. In order to implement possible features, 7 points of the resonance frequency between spectral points 464-470 were selected as shown in Fig.3.21. Three out of seven components were used for the dimensional compression, in order to make the novelty detection techniques performance challenging in terms of validating the paradox that was described. Of the 180 observations of undamaged condition, the first 120 would be used to establish the training and validation set for the auto-association analysis (avoiding overfitting), see Table.3.1.

Observations	Condition	
1-120	undamaged	train and validation
121-180	undamaged	test
181-240	impact 15 joules	test
241-300	impact 30 joules	test
301-360	impact 40 joules	test

Table 3.1: Description of observations sequence.

The existence of the paradox and that PCA and MLPs with three layers with nonlinear nodes in middle the layer and linear or nonlinear nodes in the output layer do not perform the same in terms of novelty detection is validated from the damage detection results which can be seen in Figs.3.22-3.25.
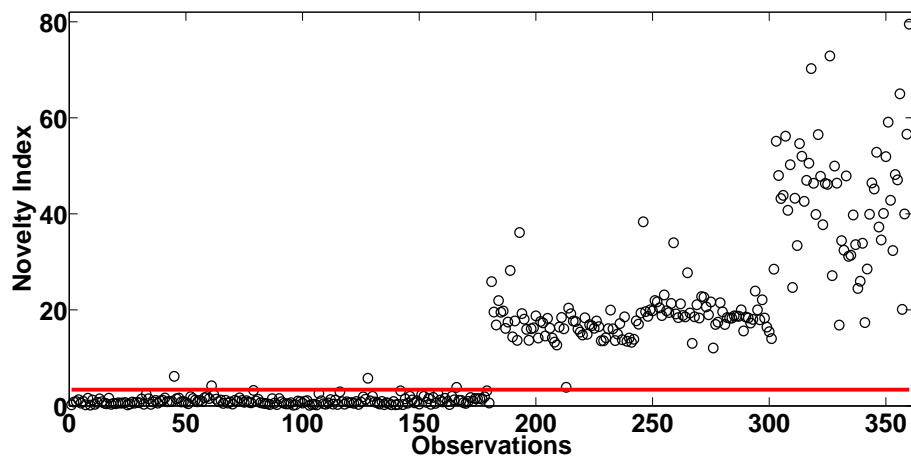

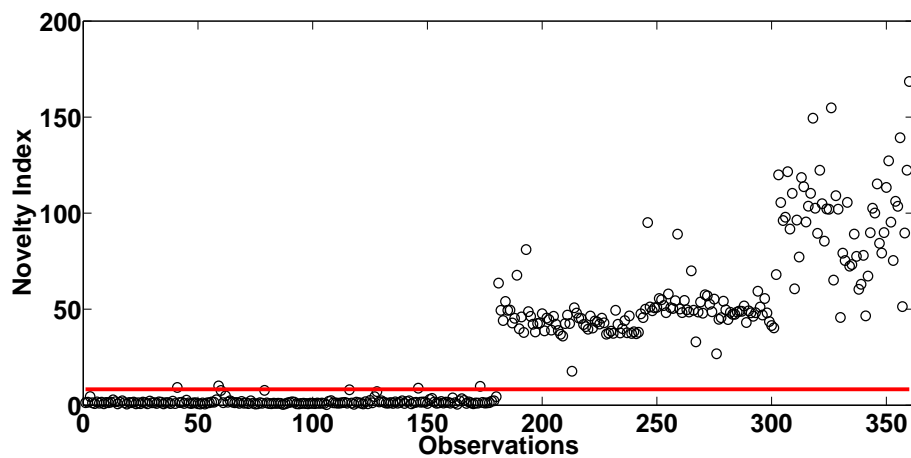Figure 3.22: Five-layer AANN novelty index

Figure 3.23: Three-layer AANN novelty index with nonlinear activations in in all nodes.
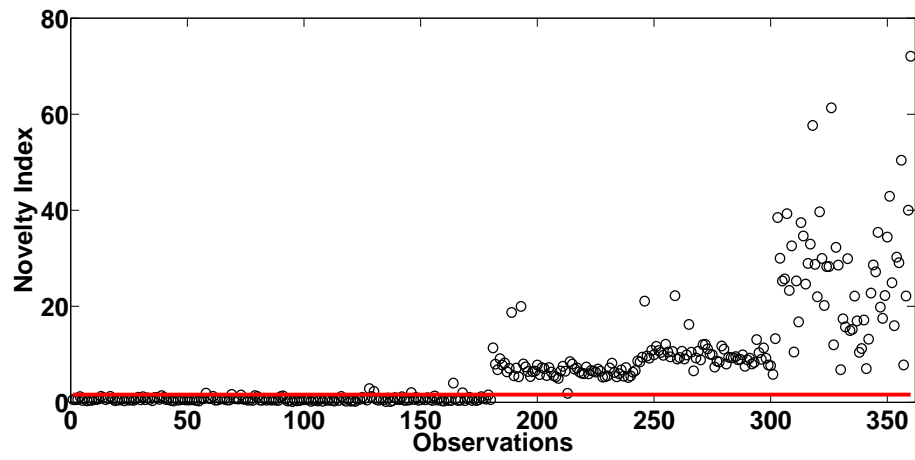
Figure 3.24: Three-layer AANN novelty index with linear activations in output layer.
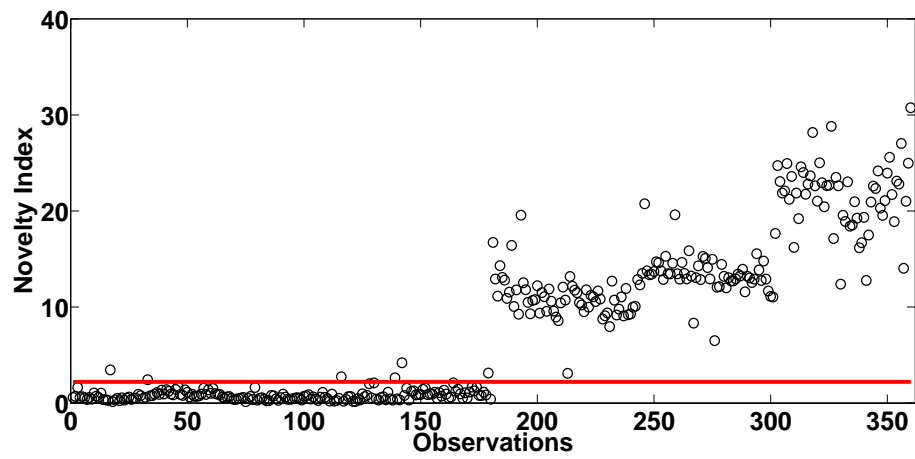

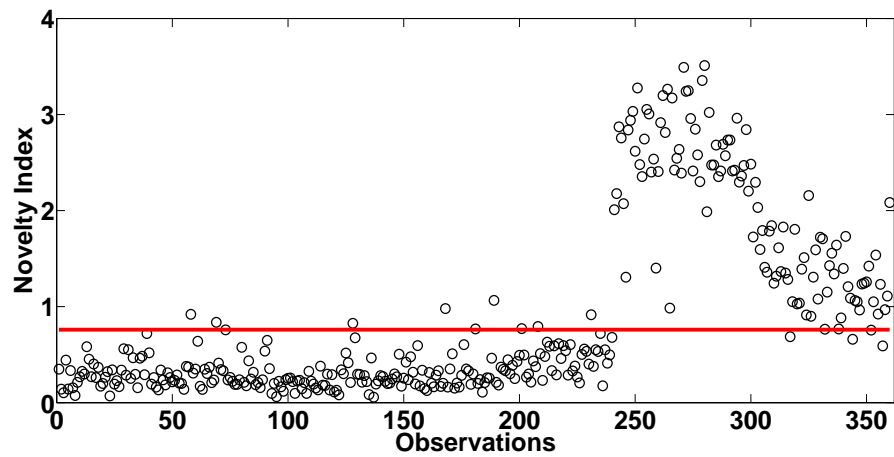Figure 3.25: RBF-AANN novelty index.

Figure 3.26: PCA novelty index.

The first obvious comment derived from the results is that MLPs with nonlinear nodes and the RBF network outperform PCA. Specifically, both the three layer MLPs and RBF network demonstrate remarkable results as they not only perform much better than PCA but also, highlight similar results with the five-layer auto-associator. PCA (Fig.3.26) is not able to detect the damage introduced by the 15 Joule impact between observations 181-240 (Table.3.1). Also, PCA is unable to distinguish the damage between the 30 and 40 Joules impact.

On the other hand, the MLPs and RBF auto-associator (Figs.3.23-3.24 and Fig.3.25) detect the 15 Joule impact and also indicate a monotonic difference between the impacts.

These results are very important as they experimentally validate the conclusions derived in previous sections and also that the paradox indeed exists. From the experimental results it can be said that for features derived from vibration responses were the correlation between the variables does not demonstrate very strong nonlinearities as in Section 3.8, the three-layer MLP and RBF networks perform as well as the five-layer MLP.

At this point, it has to be made clear that correlation between the feature variables affects the efficiency of auto-association regarding their nonlinearity and not regarding the Gaussian or non-Gaussian distribution between the variables. This is to be said as the auto-associators do not suffer from the specific distribution shape (non-Gaussian distribution).

3.12 Discussion and conclusion

In this chapter the properties of different architectures of auto-association are evaluated in order to investigate the paradox as it was described. Although, this study is very technical and seems theoretical, the conclusions that were derived involve critical practical aspects of efficient novelty detection. An analysis is performed in order to demonstrate the ability of nonlinear auto-association of MLPs consisted of one nonlinear hidden layer and with linear and nonlinear activation functions in the output layer. Also, RBF networks were implemented in terms of auto-association and novelty detection.

Auto-association with only three-layer neural networks is of great interest in terms of

reduced network complexity. The classic five-layer MLP is a very powerful tool but its complexity of finding the right number of nodes and the computational power of optimising the weights and biases is a practical drawback. This is the reason that three-layer nonlinear auto-associators offer a practical speed advantage.

It is established that linear PCA lacks performance for multimodal classification problems and novelty detection. It is a dominant conclusion throughout this chapter that single hidden layer auto-associators are not performing in a similar fashion to PCA. The only architecture that allows similar results with PCA is an MLP where all nodes are linear as can be seen in Fig.3.27 as the novelty detection results are identical with the PCA results in Fig.3.26.
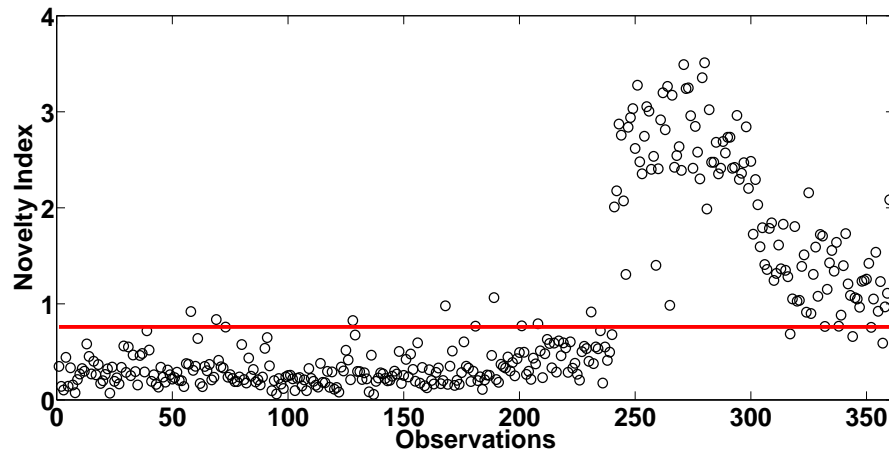

Figure 3.27: Three layers AANN novelty index with linear activations in all nodes (identical to linear PCA).

Nonlinear auto-association of MLPs consisted of one nonlinear hidden layer can not fully reconstruct highly non-linear surfaces. But, especially, in structural vibration responses where the distribution shapes are not so demanding in nonlinear terms their ability to catch some of the nonlinearity of the variables is noticeable.

This technical ability is so important as they offer significant advantage in detecting damage compared to linear tools. A lot of the methods described in this chapter will be implemented in even more complex and demanding tasks in Chapter Four and Chapter Six.

Chapter 4

WIND TURBINE DAMAGE DETECTION

In the current study, machine learning algorithms based on Artificial Neural Networks (ANNs), including an Auto-Associative Neural Network (AANN) based on a standard five-layer MLP form and a different approach (as it was analytically described in Chapter Three) to auto-association with Radial Basis Functions (RBFs) networks are used, which are optimised for fast and efficient runs.

The purpose of this chapter is to introduce such pattern recognition methods into the wind energy field and attempts to address the effectiveness of such methods by combining vibration response data with novelty detection techniques.

With the increased interest in implementation of wind turbine power plants in remote areas, structural health monitoring (SHM) will be one of the key cards in the efficient establishment of wind turbines in the energy arena. Detection of blade damage at an early stage is a critical problem as blade failure can lead to a catastrophic outcome for the entire wind turbine system. Failure of the WTs could cause massive financial losses especially for structures that are operating in offshore sites. Structural Health Monitoring (SHM) of WTs is essential in order to ensure not only structural safety, but also to avoid component over-design that could lead to economic and structural inefficiency.

Experimental measurements from vibration analysis were extracted from a 9m CX-

100 blade by researchers at Los Alamos National Laboratory (LANL) throughout a full-scale fatigue test conducted at the National Renewable Energy Laboratorys (NREL) and National Wind Technology Center (NWTC). The blade was harmonically excited at its first natural frequency using a Universal Resonant EXcitation (UREX) system.

4.1   The story so far

One of the most important and expensive components of new generation wind turbines is the large blades that can go up to 90 metres in length; such turbines can reach rated powers of over 5 MW. Automatic mechanisms for damage detection of the blades are still at an embryonic stage. Wind turbine blades are susceptible to multiple modes of failure. In addition, the continuous nature of their operation under variable loading and environmental conditions makes the employment of a damage detection system challenging.

Several researchers have tried in recent years to apply damage detection technologies and these studies were mainly in a laboratory environment [22, 79–85]. Briefly, both passive and active sensing technologies have been applied in the context of WT blade SHM [81]. In passive sensing techniques there is no external/artificial excitation as in active sensing techniques. Most of the SHM techniques and sensor systems that are discussed in the literature and available to industry have been considered for application to WT blades. However, there is not much progress on robust, successful and on-line application of these techniques in the SHM of in -service WT. The methods [22, 79–85] that have been applied to WT SHM [81] include vibration monitoring based methods (accelerometers, piezo or microelectromechanical systems (MEMSs), strain (strain gauge or fibre optic cables), ultrasonic waves which are popular with composite structures (piezoelectric transducer), smart paint (piezoelectric or fluorescent particles), acoustic emissions (usually barrel sensors), impedance techniques (piezoelectric transducer), laser vibrometry (scanning laser Doppler), impedance tomography (carbon nanotube), thermography (infrared cameras), laser ultrasound (laser devices), nanosensors (electronic nano-particles) and buckling health monitoring (piezoelectric transducer) [22, 79–85]. All of these methods are marked by disadvantages and advantages regarding their applications to WT blades.

The blades are complicated large/long structures with many different/hybrid materi-

als [86] combined and they are rotating continuously and simultaneously changing orientation. In this study, methods such as Probabilistic Principal Component Analysis (PPCA) are adopted in order to transform the raw measurements in a lower dimensional representation. Once a particular feature is obtained, a decision algorithm has to be introduced in order to reveal the condition of the structure.

The main focus of this study is the discussion of a group of machine learning techniques for the monitoring of turbine blades by using vibration data and specifically high frequency response function measurements (FRFs) [46]. The algorithms were optimised to an extent that could offer robust, fast and accurate online monitoring.

4.2 Brief description of Wind turbine blade materials and structure

New generation wind turbines are usually hybrid material structures made of polymer matrix composites materials (PMC) combined with monolithic and sandwich composite [22, 80, 81, 83, 84, 87]. The dominant composite material being used in wind turbine blades by industry is glass fibre-reinforced composites (GFRP) and carbon fibre-reinforced composites (CFRP).

Generally, carbon fibres are much stronger as well as stiffer, compared to glass fibres. This extra stiffness allows the surrounding resin of the blade to tolerate fatigue better than glass fibres. The dominant drawback up to now with carbon composites is their high cost. Carbon laminates are used practically only for very large blades (up to 90m radius) and nowadays, in a high percentage for main spar components [87].

Briefly, the main reason that large blades need carbon fibre reinforcements is because it is significantly difficult to achieve sufficient stiffness without increasing the weight and carbon composites are a lot lighter compared to glass fibre composite. Extra weight means extra cost, lower natural frequencies (may overlap with the tower passing frequency) and increase of loading due to edgewise bending as the main cause of this is the gravitational load. Sandwich materials can be explained briefly as a type of composite laminate where two of more thin, stiff, strong composite laminates are separated by a thick, lightweight core material (they can be found as part of the main spar cap of the blade)[22, 80, 81, 83, 84, 87].

The blade mainly consists of four parts as can be seen in Figs.4.1 and 4.2, [87]. The blade shell primary function is to deliver the optimal aerodynamic shape, by structurally contributing in strengthening and stiffening the spar, mainly with torsional load and edgewise bending. Composite sandwich laminates can be found across the leading and trailing edges in order to increase resistance due to edgewise load. Commonly when sandwich materials are used, they are followed by relatively thin composite laminates where the shell parts are adhesively attached to the main spar.

When high torsion is applied to blades they are undergoing shear loading and for this reason a percentage of fibres is orientated diagonally. In addition, the shells have fibres which run along the whole body. This basically helps the blade shell to withstand edgewise bending and to assist the spar cap in flapwise moments. The main spar is covering the part from the root of the blade to a position very close to the tip.

The basic function of the spar is to carry the bladewise bending loading. The top part of the spar which is connected to the shell as shown in Fig.4.2, called the spar cap or flange is to carry the highly demanding flapwise bending moment and usually is manufactured by thick monolithic composite layers which, for the new generation large blades is a combination of carbon/glass composites. To make the spar caps able to deal with this bending moment, unidirectional high axial fibre angles are used. The other part of the main spar called the internal web or stiffeners as shown in Fig.4.2, plays a significant role in the structure of the blade as they carry the flapwise shear forces and they frequently consist of composite sandwich material with polymeric or balsa wood core and biaxial thin composite face layers with $\pm 45^o$ degrees of orientation [22, 80, 81, 83, 84, 87].
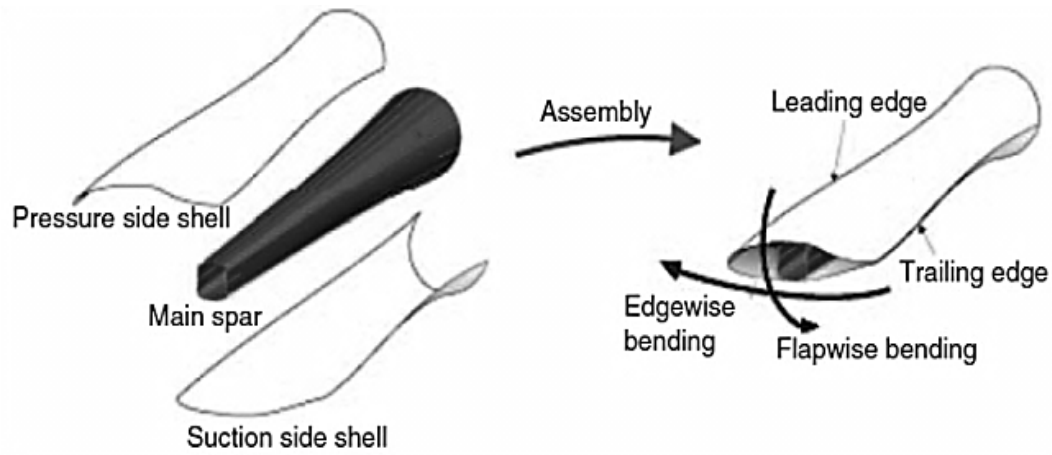
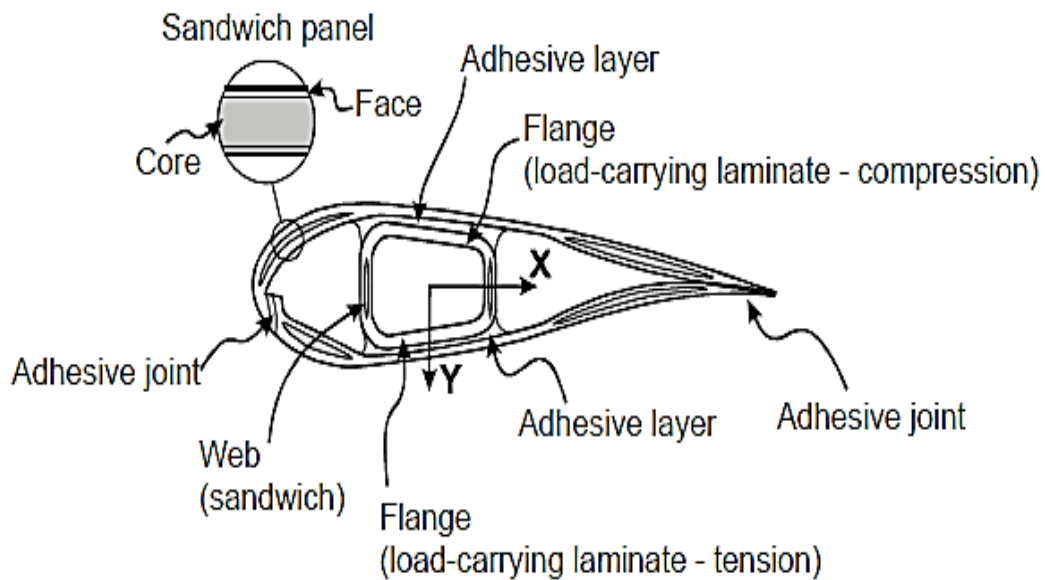Figure 4.1: Wind turbine blade assembly [87].

Figure 4.2: Wind turbine blade cross section [87]].

4.3   Motivation and Experimental overview of the testing blade

Blade loading in reality is characterised by a complicated combination of different loads [85]. Flapwise and edgewise bending are generated from pressure loading on the blade. The edgewise bending loading is mainly formed from the gravitational load and as the blades rotate the applied direction changes as a result. These two loads play a vital role in the structural efficiency and design of large wind turbine blades. It has to be mentioned that for wind turbine blades that are operating in offshore sites, the maximum values of flapwise bending load can occur when the blades are subject to extreme values of gust wind and they are in a standstill position in order to avoid failure. Other categories of blade loading are torsional loading, normal loading (inertia forces due to natural rotation of the blades) and moderately negligible loading due to some pitch acceleration or deceleration through the wind turbine operation. Flapwise bending was adopted in this experiment as it is one of the main reasons for failure of the blade body and especially the main supporting spar.

The main motivation of the experiment that is described was to implement a variety of different sensor techniques [88, 89] and find capable features that, via advanced machine learning tools, could detect damage at early stages before visual abnormalities were presented.

The full-scale fatigue test was performed between the dates 11/08/2011 and 9/11/2011, see Fig.4.3. The test ran continuously for approximately 8.5 million cycles until a visible crack was observed on 9/11/2011, see Fig.4.6. The crack nature was a through-thickness crack that appeared on the surface of the blade in the root area near the leading edge as shown in Figs.4.7-4.8.

The blade is made of a fibre-glass body (shell) and a carbon-fibre spar cap with balsa wood frame (a small percentage of the spar cap of the root section is made of glass fibre with some carbon-fibre layers in the thick skin) as shown in Figs.4.2 and Fig.4.5.
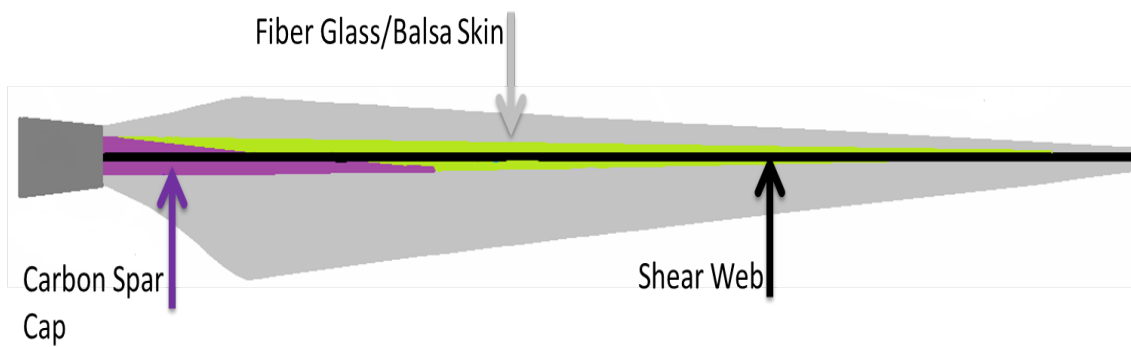
Figure 4.3: Wind turbine blade experiment.

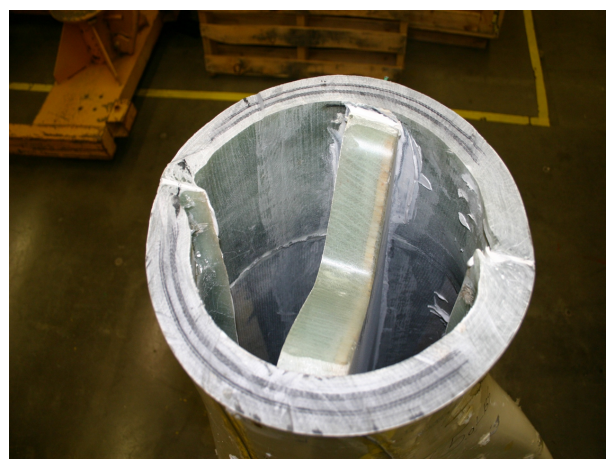Figure 4.4: CX-100 blade main material regions.

Figure 4.5: Carbon fibre spar cap with balsa wood frame.

The specimen was excited at its first natural frequency near 1.8 Hz. The UREX (Universal Resonant EXciter) was implemented in order to apply the test loading. The UREX is a system that was developed at NREL with a specific design in order to test bend-twist coupled blades. The UREX consists of a pair of hydraulic actuators which are mounted onto the blade through a ballast saddle. The actuators are located on each side of the wind turbine blade (red square boxes) as shown in Fig.4.3. These specific saddle positions offer important advantages compared to single-actuator concepts. The main advantage of the UREX is that the rotational inertia is minimized rather than mounting the actuator and added mass above the center of the blade, which could lead to introduction of torsional loading and as a result influencing the bend-twist response of the structure.

Two saddle positions were implemented at 1.6 m and 6.75 m from the root and on 13/09/2011, 10/10/2011 and 18/10/2011 extra mass was added on the first saddle at 1.6 m, leading to an increase of mass from 582.4 kg initially to 642.3, 702.16, 762.04 kg respectively. This extra mass was introduced due to the fact that after 2 million cycles no visual damage was observed on the blade body. The structural health monitoring sensor system consisted of several different systems implementing both active sensing and passive sensing [88, 89].

In this study, active sensing measurements were used for the novelty detection methods. Using this active sensing system (LASER sensing system) two different sensor arrays were implemented called the INNER and OUTER sensor arrays; they consisted of 6 and 7 sensors respectively and an actuator was used in each of them (a1), see Figs.4.9-4.10. The excitation frequency bandwidth was between 5 kHz and 40 kHz with a sampling rate of 96 kHz giving a resolution of 7200 spectral points for the FRFs that were measured and used for this study. The Frequency Response Function (FRF) in simplified terms is the measure of any system's output spectrum in response to an input signal spectrum (appendix A). The observations are between 11/08/2011 and 9/11/2011 and correspond to 565 in number for the INNER sensors and 534 for the OUTER sensors (crack observed at observation 560 for INNER and 529 for OUTER array).
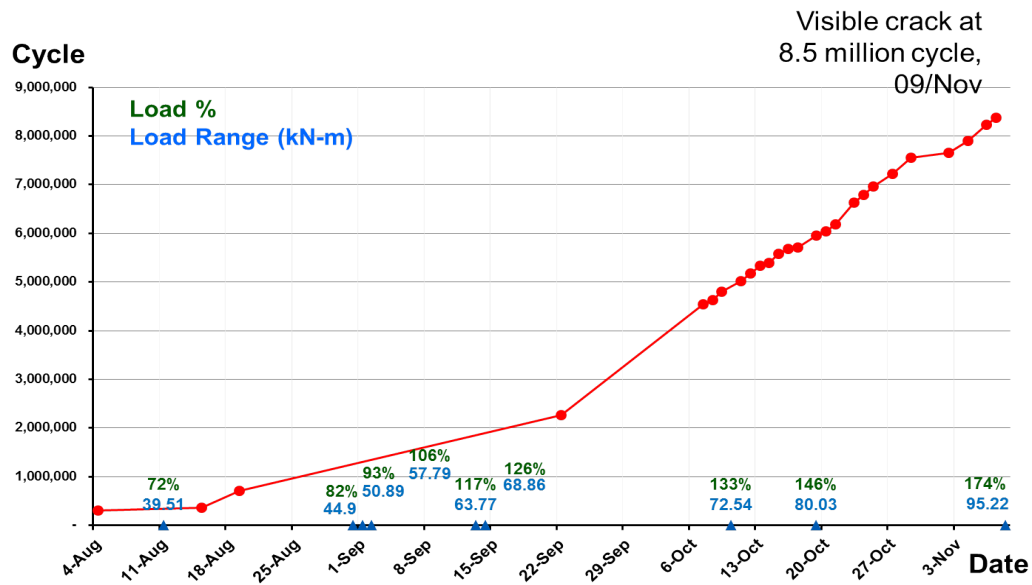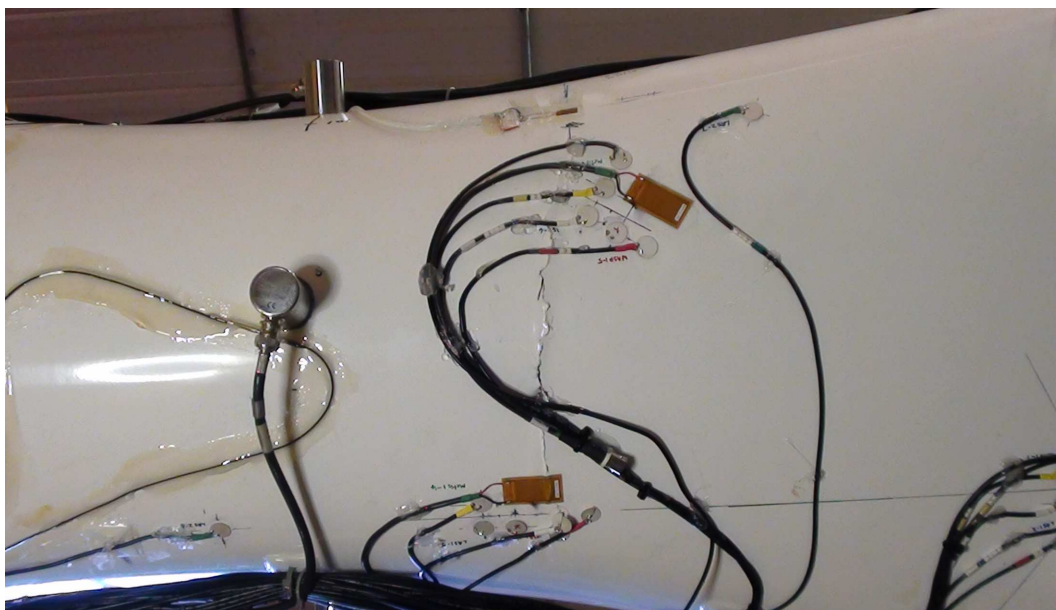
Figure 4.6: Fatigue test history.

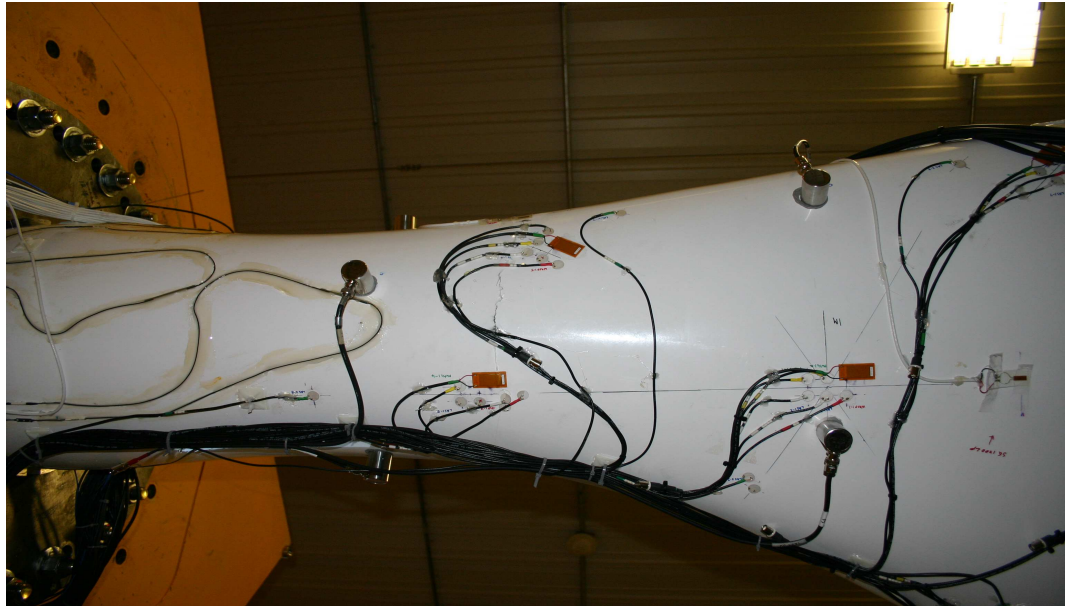Figure 4.7: Visible crack was observed on 9/11/2011.

Figure 4.8: Crack that appeared on the surface of the blade in the root
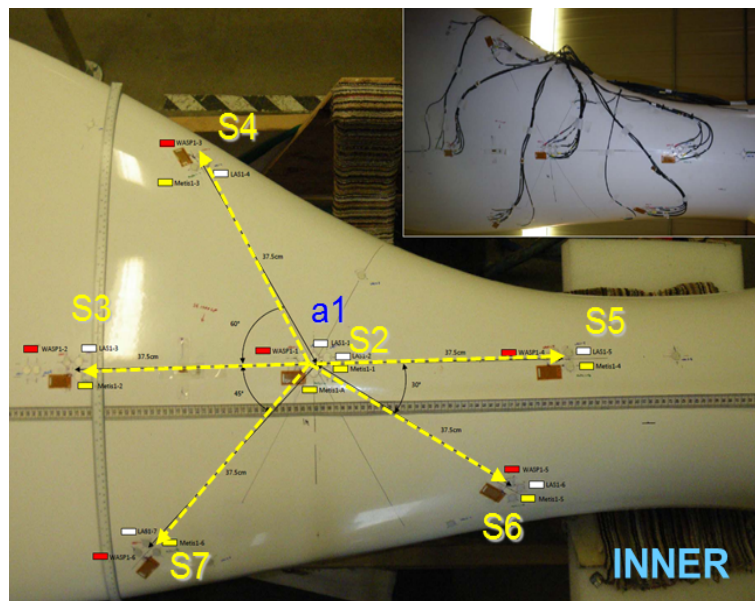area near the leading edge.

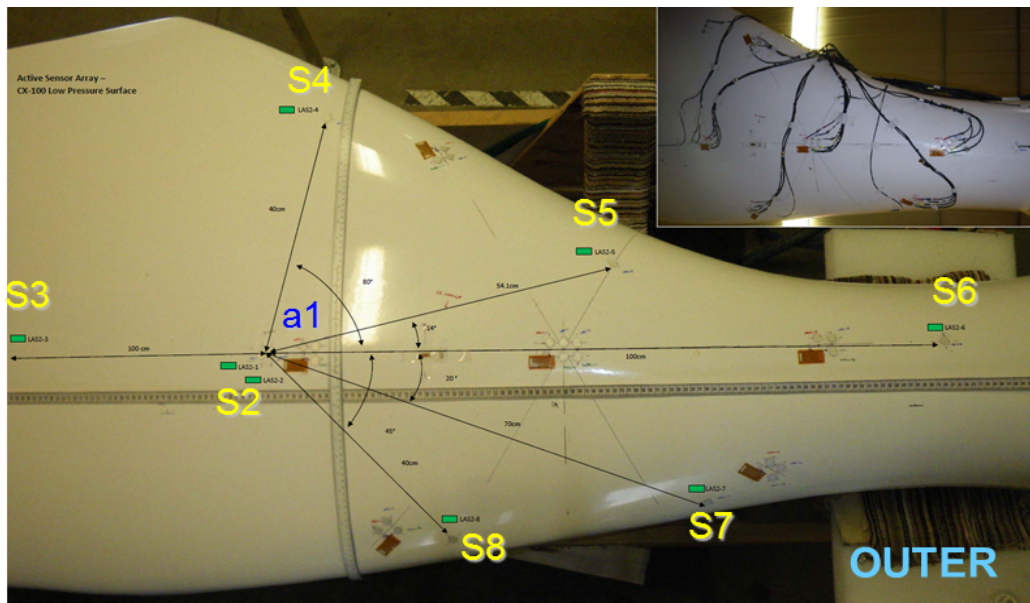

Figure 4.9: INNER sensor array.

Figure 4.10: OUTER sensor array.

4.4 Novelty detection results

As demonstrated in the previous Chapter Three, one can introduce subsets of the data by introducing features which are sensitive to damage. In this section an alternative approach is presented by reducing the dimension of the FRF data using PPCA. It was decided to retain the first 5 principal components and thus give a five dimensional feature. It has to be mentioned that the reason that classical PCA is not used is because the FRF dimension is several thousand spectral points making the usage of the classic covariance matrix approach weak in accuracy (if not impossible) and significantly low in speed processing. PPCA was implemented in two parts; a low frequency band up to point 100 which contains all the basic resonance frequencies and a high frequency one between points 101-7200 which includes the high frequency response components (corresponding to resonance frequencies of more complex modal operations). In turn, the results are presented in two groups.

After selecting the training, validation and testing data for each of the 13 sensors labelled INNER and OUTER, a trained five-layer MLP AANN was used via unsupervised learning to model the normal condition. The first 300 observations (**black** colour points in novelty detection results) were selected for training after being corrupted by noise and reproduced several times to enlarge the normal condition set. 80% of the data was used for training and the rest 20% of data for validation in order to avoid over-fitting.

A novelty distance was calculated by feeding the two different network algorithms with testing data (blue points in novelty detection results). In Table.4.1 and Table.4.2 there is a summary of the novelty detection results and in Figs.4.11-4.23 is shown the analytic results of two sensor arrays in the INNER and OUTER families by comparing the low and high frequencies families for the same sensor.

Inner Sensor	Observation	Date
1	326	17 Oct
2	305	14 Oct
3	311	14 Oct
4	318	16 Oct
5	317	15 Oct
6	306	14 Oct

Table 4.1: Summary of the novelty detection results of Inner sensors for the high frequency responses.

Outer Sensor	Observation	Date
1	331	18 Oct
2	335	19 Oct
3	331	18 Oct
4	328	17 Oct
5	330	17 Oct
6	313	16 Oct
7	329	18 Oct

Table 4.2: Summary of the novelty detection results of Outer sensors for the high frequency responses.
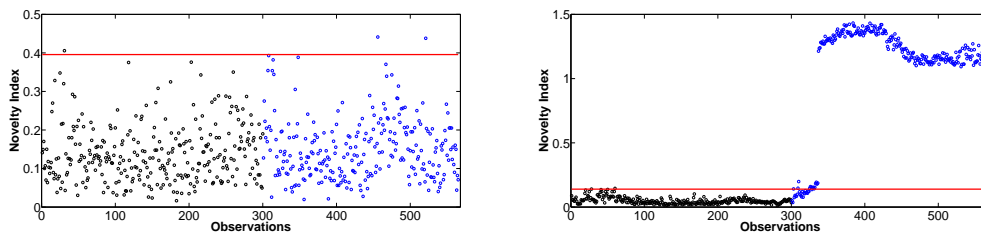

Figure 4.11: Low frequencies range novelty index and high frequencies range novelty index for Inner sensor 1.

Figure 4.12: Low frequencies range novelty index (left) and high frequencies range novelty index for Inner sensor 2 (right).

Figure 4.13: Low frequencies range novelty index (left) and high frequencies range novelty index for Inner sensor 3 (right).

Figure 4.14: Low frequencies range novelty index (left) and high frequencies range novelty index for Inner sensor 4 (right).

Figure 4.15: Low frequencies range novelty index (left) and high frequencies range novelty index for Inner sensor 5 (right).
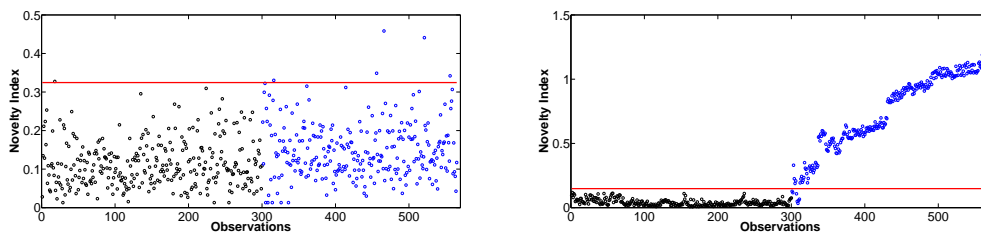

Figure 4.16: Low frequencies range novelty index (left) and high frequencies range novelty index for Inner sensor 6 (right).

Figure 4.17: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 1 (right).
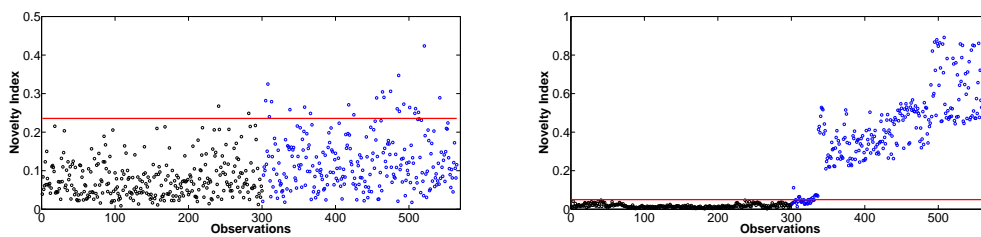

Figure 4.18: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 2 (right).

Figure 4.19: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 3 (right).
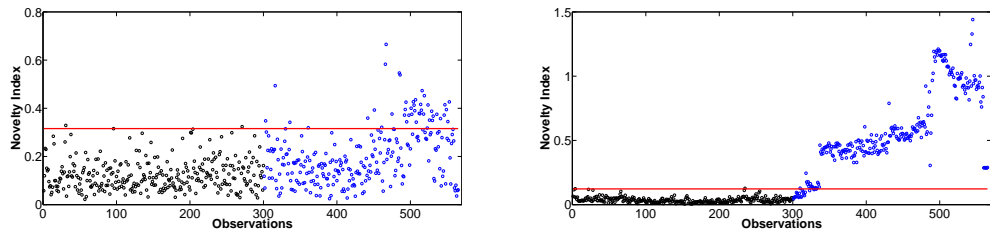

Figure 4.20: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 4 (right).
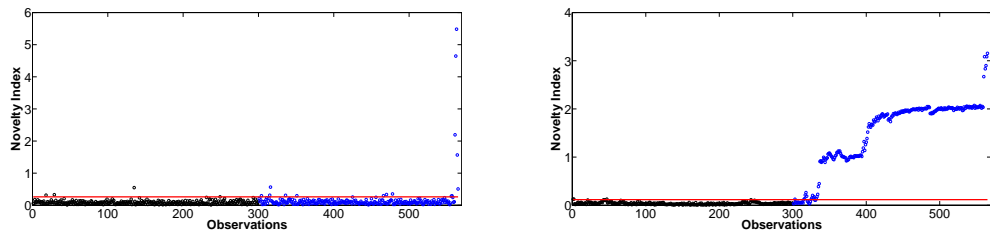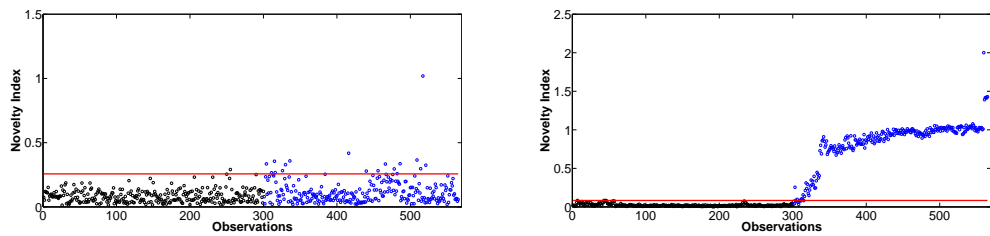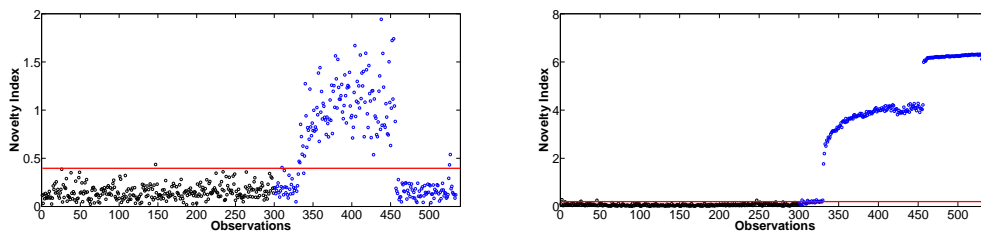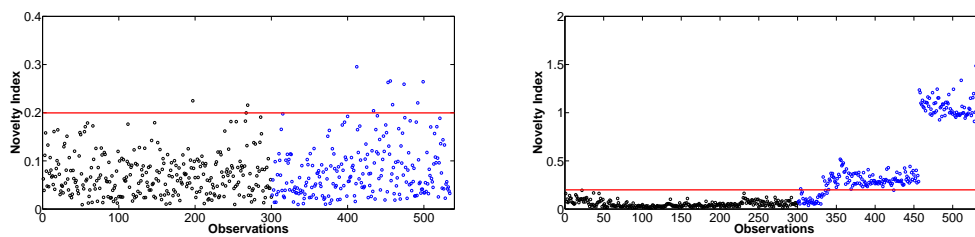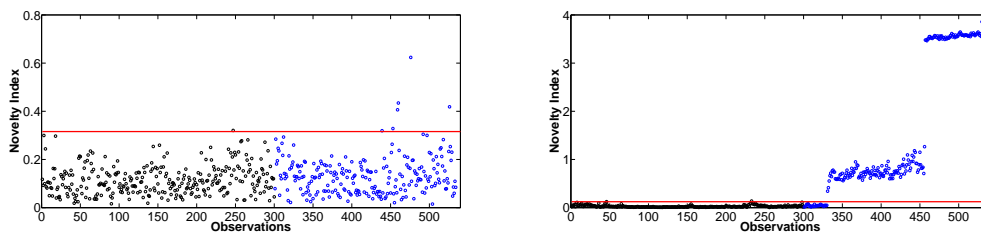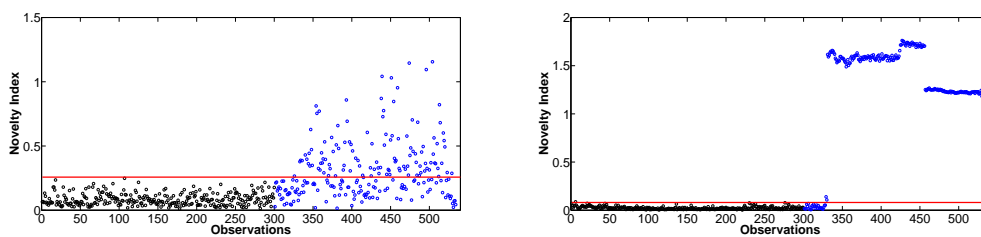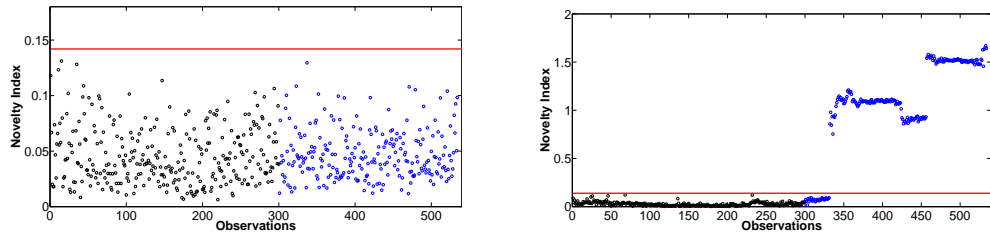
Figure 4.21: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 5 (right).
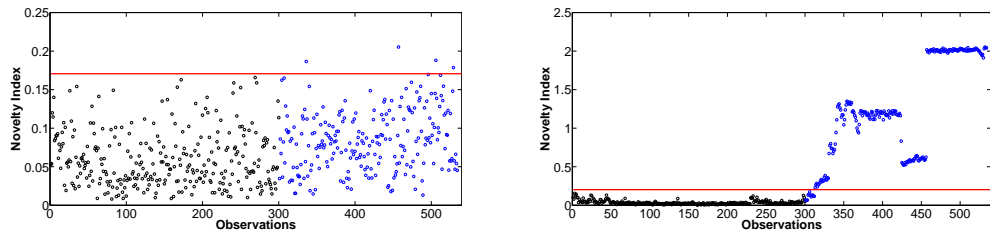

Figure 4.22: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 6 (right).
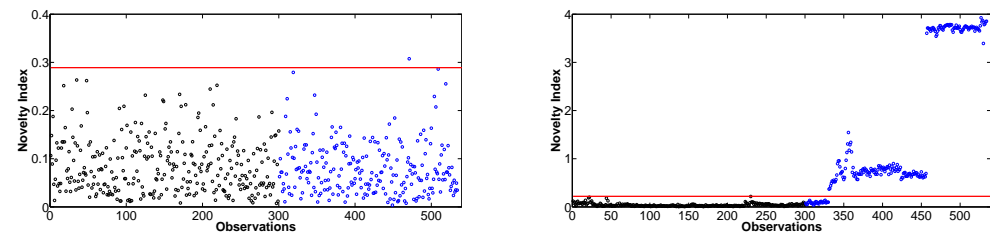

Figure 4.23: Low frequencies range novelty index (left) and high frequencies range novelty index for Outer sensor 7 (right).

The first obvious comment is that low range of frequency responses is not a useful feature regarding damage detection as in most of sensors no strong indication of outliers is detected. Simply, there is no detectable change in low frequency responses except the results in Outer sensors 1 and 4.

On the other hand, high frequency response functions, a technique that was applied in the current study, was proved to be a significant asset in detecting potential damage. The advantage of using high frequency vibrations to monitor the blade for damage is that the structure parameters such as stiffness, mass, or energy dissipation of the whole blade system will have a critical connection with high frequency components of the response. Also, the wavelength of the modes is smaller, so more sensitive to smaller damage.

The results are strongly indicating that the measured FRFs by utilising the piezoelectric active-sensors allow the methodology to be sensitive to small structural changes (as it happened). Novelty detection applied to each sensor's FRF measurements indicate early novelty, about 20-25 days before the visible crack was observed.

The results point out that the initial damage was introduced internally and started from the main carbon spar. In the experiment, the excitation of the fatigue test was the first bending moment with a natural frequency at 1.8 Hz. This specific load is carried from the main spar and in high proportion at the blade root. If a failure occurs in the spar components then the blade's shell is not able to carry the high loading and usually damage will appear inevitably to the shell's body. The shell mainly is for aerodynamic reasons and for carrying the edgewise loading but also its design is playing a small structural role by helping in stiffening and strengthening the spar. After performing an autopsy on the blade's shell the former assumption was validated, as can be seen in Fig.4.24.

A spanwise surface crack was formed at the junction of the carbon spar cap and the trailing edge of the shell. By the 6 million cycles level, the crack had progressed via the thickness of the skin and at 8.5 million cycles level a visible crack on the surface was noted and then the test was stopped after 5 more runs.

It can be assumed with confidence that once the crack had fully penetrated the skin the damage could dramatically increase progressively. In reality, this critical blade failure could lead to a catastrophic outcome for the entire wind turbine system.

Regarding, the Outer sensor 1 results, one can see a low frequency responses it can
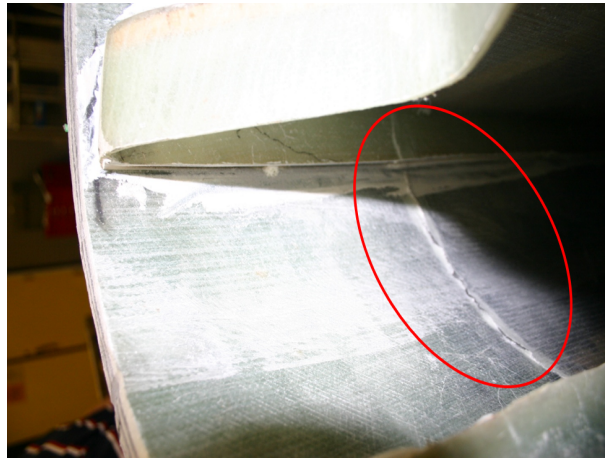
Figure 4.24: Initial damage was introduced internally and started from the main carbon spar.

be seen a novelty detection after observation 335 where discordance index values are following a cyclic path up to observation 450 and then a sudden reduction of the index values before the structure comes back to normal condition. A possible answer to this phenomenon could be that the crack or delamination introduced in the carbon spar can open or close during the fatigue loading in such composite structures and thus mask the possible damage in the low frequency responses.

In some sensor results, one can notice a monotonic jump as the days pass, indicating a noticeable novelty change about 10-15 days before a visible crack was observed. This noticeable observation is aided by the ability of the FRF method because as a tool it can determine global damage detection capabilities.

In addition, this monotonic damage index increase is directly related to the relative propagation path of the damage location (from the internal spar to the shell body). This feature could be utilised to classify the location of the damage. This is a very strong conclusion as similar results were obtained for all different sensors. These combined high frequency vibration analyse with MLPs AANNs offers the the potential comparative advantage of limiting the number of sensors that are required to cover the entire blade body.

4.5 Auto-association using Radial Basis Functions

As analytically described in a previous Chapter Three, the RBF network can be used in terms of auto-association. The main reason that RBF networks are presented in this section as an alternative to MLPs is that RBF networks compared to MLPs do not usually need a full and challenging nonlinear optimisation of all the parameters. Furthermore, RBF networks are significantly faster and computationally more efficient, giving them the advantage of fastest online inspection for simple damage detection tasks for wind turbine blades.

A few results will be presented in order to demonstrate the efficiency of an RBF in comparison with a five-layer MLP that was used previously for the high frequency responses. In Figs.4.25-4.28 the comparative results are displayed.

As can be seen from the results presented, the RBF auto-associator performance is noteworthy as in comparison with the five-layer auto-associator it offers a high performance in terms of damage detection. The results validate the conclusions of Chapter Three that auto-association with RBF networks is possible for simple novelty tasks as they perform (even with limited centres) as a nonlinear density model.
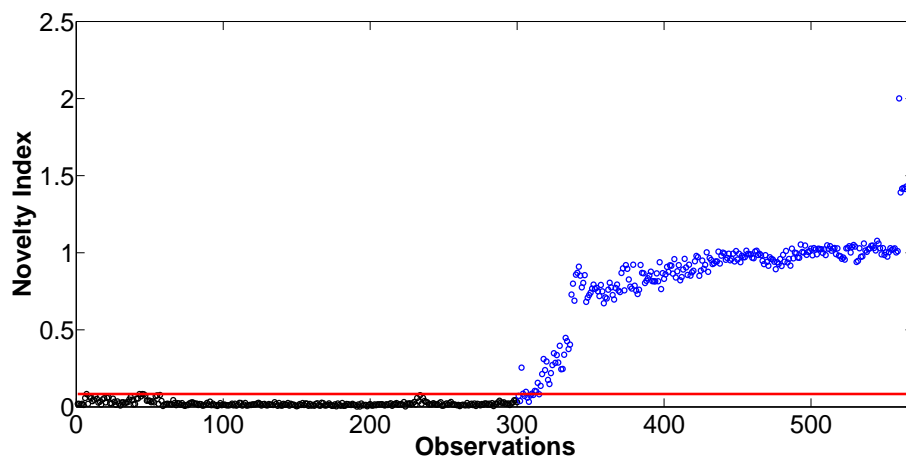

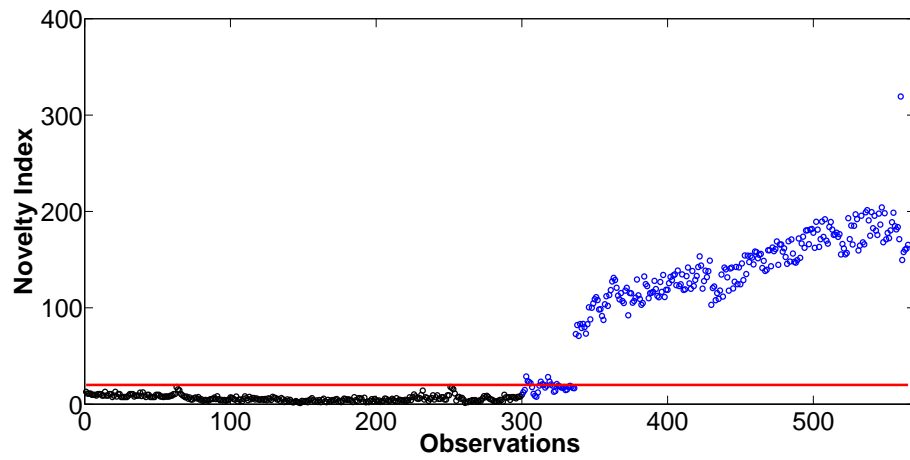Figure 4.25: RBF novelty index for Inner sensor 6.

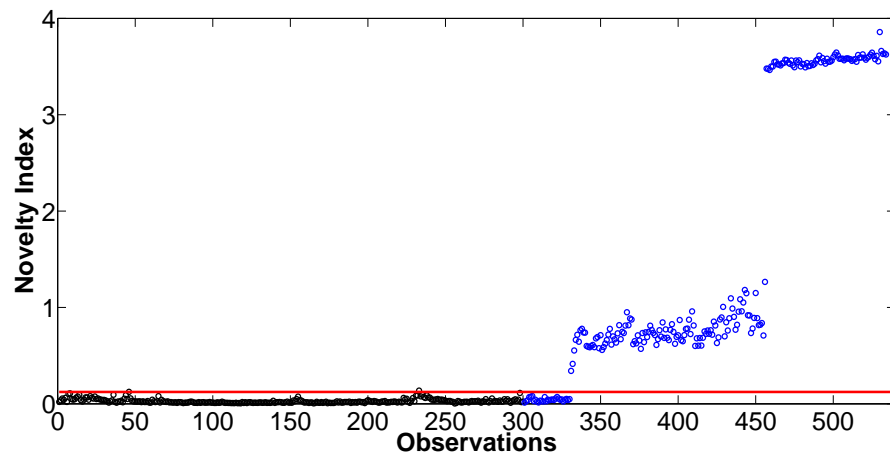Figure 4.26: Five layer MLP novelty index for Inner sensor 6.

Figure 4.27: RBF novelty index for Outer sensor 3.
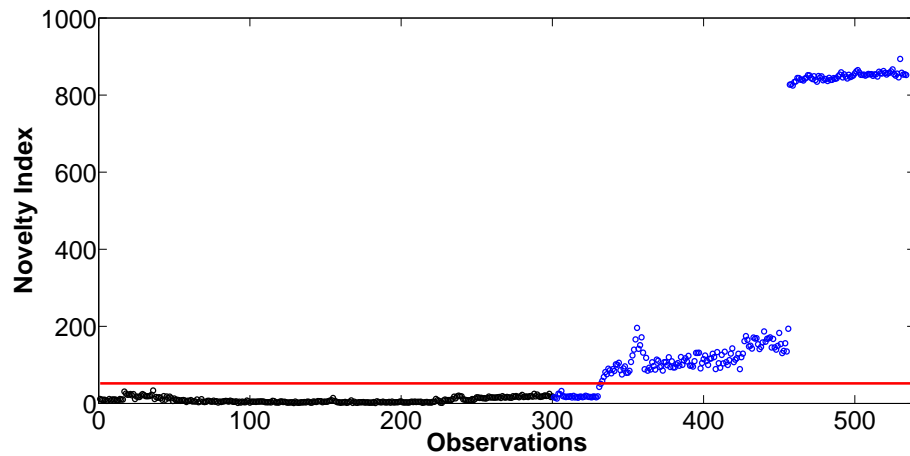
Figure 4.28: Five layer MLP novelty index for Outer sensor 3.

4.6 An exploration of virtual SCADA data of a simulated offshore Wind Farm for SHM

The fully operational Lillgrund wind farm is situated in the sea area between Denmark and Sweden, consisting of 48 WTs of rated power of 2.3 MW turbines [90]. The analysis here of a machine learning approach is based on virtual SCADA data generated by a validated CFD model of Lillgrund [91–93]. The idea of using SCADA measurements for SHM and CM has received very little attention from both the wind energy and SHM communities. In order to maintain a qualitative profit with large offshore wind farms, a major challenge is to keep operational and maintenance costs to the lowest level by insuring reliable and robust SHM systems, as these kind of costs are significantly higher compared to those of onshore wind farms. For this reason, data mining and machine learning are promising approaches for modelling wind energy aspects such as power prediction or wind load forecasting.

This section proposes a simple methodology for detecting abnormal behaviour of WT systems using a data-driven approach illustrated by CFD modelling of Lillgrund Wind farm based on artificial SCADA data extraction. The analysis is based on neural network and Gaussian process regression and is used to predict the measurement of each WT from the measurements of other WTs in the farm. Regression model error is used as an index of abnormal response.

4.7 Farm Description and CFD modelling

Each turbine is labelled as a combination of a letter (rows A-D) and number (rows 1-8). For simplicity, the turbines have also been numbered from 1 to 48 as shown in Fig.4.29. The separation between the turbines in the row is $3.3 \times D$ where D is the diameter of the turbine and the rows are separated by $4.3 \times D$, Fig.4.29. The wind turbines are Siemens SWT-2.3-93, characterised by a rotor diameter of 92.6m and the hub height is 65m giving a rated power of 2.3 MW, Fig.4.30. The maximum rated power is reached when wind speeds take values of 12m/s (rated wind speed).

The data used in this section was provided from Creech et al. [91–93]. Briefly, the CFD model of Lillgrund [91, 92] used the dynamic actuator volume approach for the turbine [91], which provides time-varying turbine diagnostics. Fluidity, the

open-source CFD solver from Imperial College, was used to resolve the flow [94]. Aerodynamic performance of the blades was based upon NACA aerofoil data, which closely matched the specifications of the Siemens blades [95]. Large eddy simulation (LES) modelled the unsteady turbulent flow, along with the Synthetic Eddy Method (SEM) [96] to inject atmospheric turbulence into the simulation. The unsteady modelling of turbulence allowed the turbines to experience (and react to) varying flow conditions, and so exhibit time-varying performance and blade loading.

The computational domain was characterised by a cuboid volume of $8.1 \times 8.1 \times 0.6$km, with the wind farm in the centre in such a way that the first turbine was at least 2km from the inflow boundary, to allow the simulated turbulence to develop fully before encountering the farm as seen in Fig.4.31. The inlet conditions were a steady flow with a logarithmic vertical profile, upon which the SEM-generated turbulent eddies were imposed.

The simulations were run for the first 2000s without turbines, to allow the atmospheric turbulence to fully develop, at which point the turbines were then activated. A further 600s of simulation time was then used to allow the turbines to spin-up. This was checked by monitoring the turbine diagnostics, which had a sampling time of 0.5s.
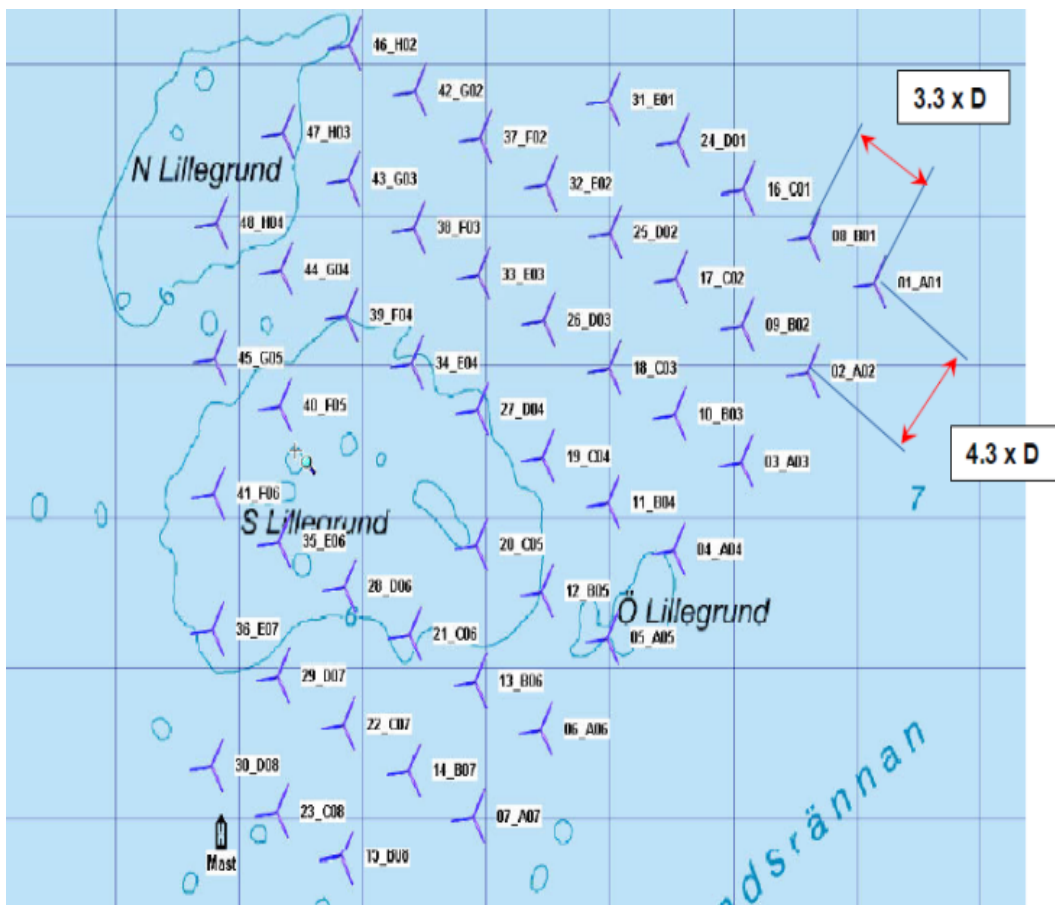
Figure 4.29: The picture shows the location of the 48 turbines in Lillgrund wind farm [90].

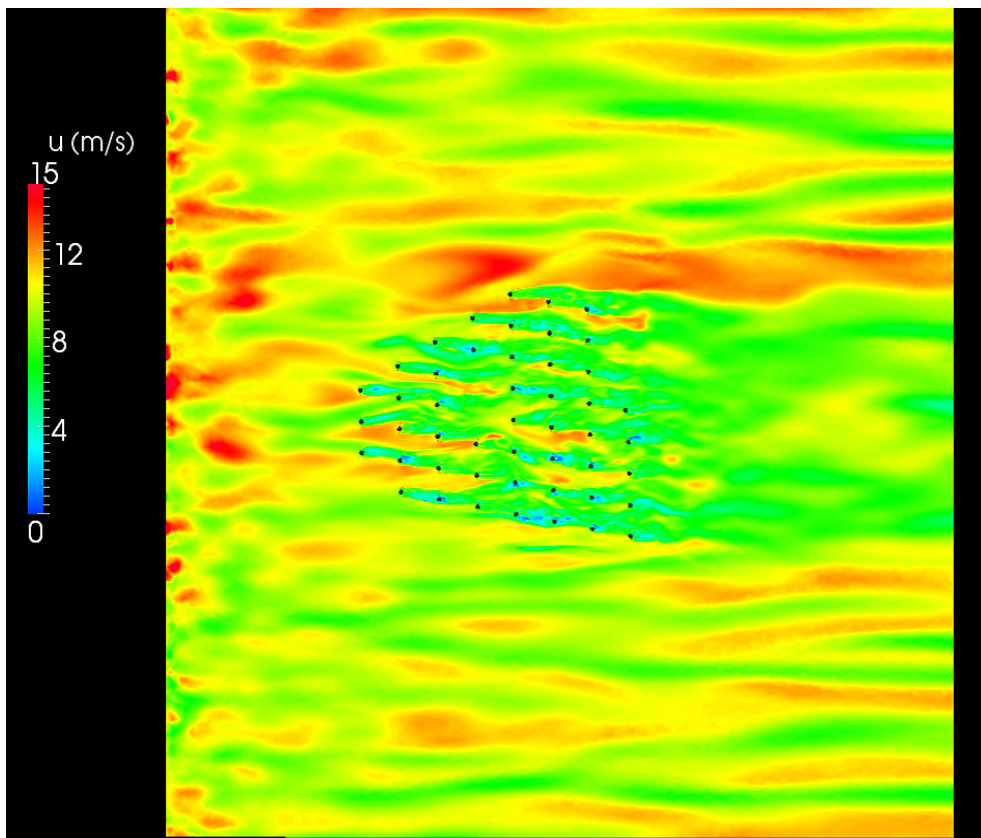Figure 4.30: The picture shows a Siemens SWT-2.3-93 in Lillgrund [90].

Figure 4.31: Snapshot of streamwise velocity in a horizontal section at hub height. The grey dots indicate the turbine locations [92].

4.8 Neural Network Regression Process

Given suitable training data from the power output, the neural network model should be able to accurately predict blade loading response, not only of the WT that was used as a reference, but potentially for each WT if the structure continues to operate in a similar way as in the time period where the observations for training were recorded. In turn, the regression model error can be used as an indicator of abnormal structural response. The central purpose of the simulations here is to demonstrate how the models that were built can predict the blade loading of different wind turbines, based only on the training data of one reference turbine.

Power output and blade loading were used as input and target to the regression model in order to build a relationship between the two of them. As a reference, the C07 wind turbine was chosen randomly, and as test subject, the E01 wind turbine. The first 100 and last 300 points of data were removed because they represent the simulation beginning and ending respectively. A change was introduced artificially by reducing the blade loading by 20, 40, 60% gradually after observation 1400 for every 200 points in order to simulate a progressive variation. Only the first 900 points of the power output of C07 were used as the training set in order to predict the blade loading of E01. The warning levels for regression model error can be defined as $\bar{v} + a\sigma$ where $\bar{v}$ and σ are, respectively the mean and standard deviation of all values of residual index over the training data. In statistical terms a controls the percentage of false positives. For example if the distribution is purely normal then a value of 1.96 will give estimates within warning levels of 95% probability. In this study a is set equal to 3 giving a 99.7% confidence limit.

In Figs.4.32 and 4.34 the model prediction of the blade loading for E01 based on the C07 WT for the uncorrupted and corrupted case are shown. It can be seen that the model in Fig.4.34 is catching the general trend of the blade loading although it is based on the training data (power output response) of another wind turbine.

The results in Figs.4.33 and 4.35 show the residual error between the estimations of the model and the actual values for the uncorrupted and corrupted system. It can be seen that the residual error is within the threshold values for the unfaulted system. In Fig.4.35 the residual error is exceeding the lower bound error when the system is corrupted, afterwards giving a monotonic novelty index for the different levels of fault.

Figure 4.32: Model prediction of the blade loading for E01 based on C07 Wind Turbine for the undamaged case.

Figure 4.33: Residual error of the model prediction for E01 with no introduced change.

Figure 4.34: Model prediction of the blade loading for E01 based on C07 Wind Turbine for the damaged case.
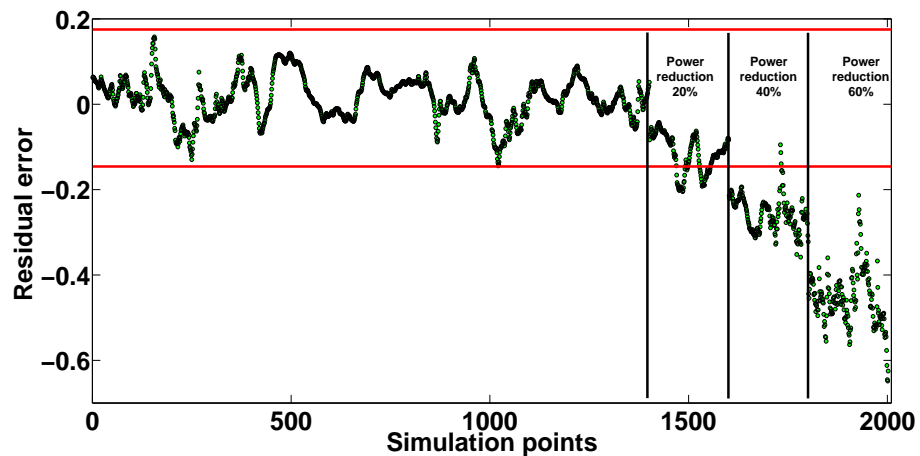
Figure 4.35: Residual error of the model prediction for E01 with introduced change.

4.9 Gaussian Process Regression

What becomes clear through this thesis, is that for simple tasks such as nonlinear regression, neural networks although they present a very powerful tool, sometimes can make it difficult and demanding to achieve the right tuning. The hard questions that have to be asked while MLPs are implemented are: which is the right architecture? How many nodes? What transfer functions? What momentum or learning rate? How many times they should run for different initial conditions?

In the area of monitoring a wind turbine via a regression analysis and in the exact

same philosophy as the one described earlier, another powerful technique can be adopted which is much simpler and faster. This technique is the Gaussian process for regression.

Gaussian process is a research area of increasing interest not only for regression but also for classification purposes. For more details readers are referred to Appendix C and [97]. The use of Gaussian processes (GPs) is a stochastic nonparametric Bayesian approach to regression and classification problems. These Gaussian processes are computationally very efficient and the nonlinear learning is relatively easy. Gaussian process regression takes into account all possible functions that fit to the training data vector and gives a predictive distribution of a single prediction for a given input vector. As a result, a mean prediction and confidence intervals on this prediction can be calculated from this predictive distribution.

The initial and basic steps in order to apply Gaussian process regression is to obtain a mean and covariance function. These functions are specified separately, and consist of a specification of a functional form and a set of parameters called hyperparameters. Here, a zero-mean function and a squared-exponential covariance function are applied (see Appendix C or [97]). When the mean and covariance functions are defined then the inference method specifies the calculation of the exact model and in simple terms describes how to compute hyperparameters by determining the minimisation of the negative log marginal likelihood. The software used for the implementation of GP regression was provided by [97].

In Figs.4.36 and 4.38 the model prediction of the loading for H02 based on the B05 WT blade power for the uncorrupted and corrupted case are shown. It can be seen that the model in Fig.4.36 is catching the general trend of the blade loading output although it is based on the training data (power output) of another wind turbine as it happened when MLPs were implemented instead of GPs.

The results in Figs.4.37 and 4.39 show the residual error between the estimations of the model and the actual values for the uncorrupted and corrupted system. It can be seen that the residual error is within the threshold values for the unfaulted system. In Fig.4.39 the residual error is exceeding the lower bound error when the system is corrupted, after giving a monotonic novelty index for the different levels of fault.
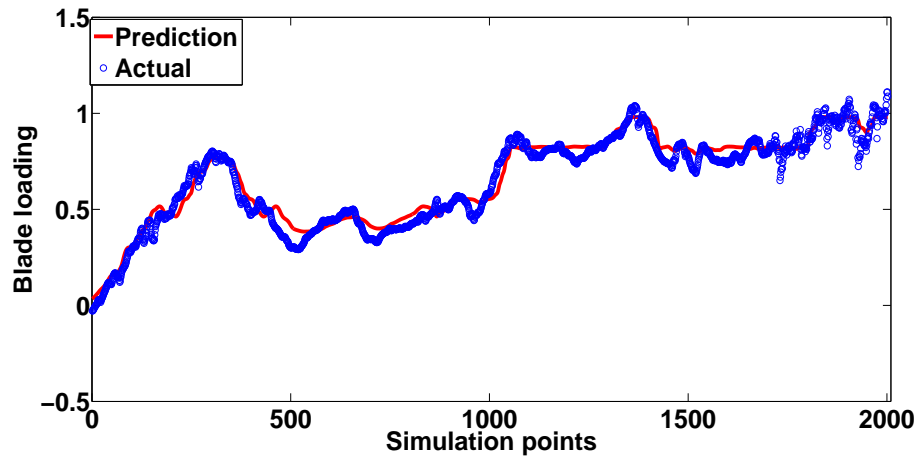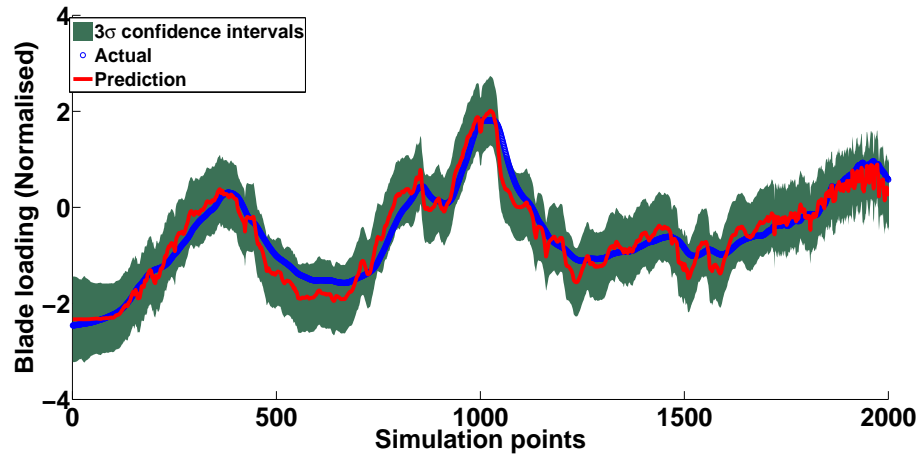
Figure 4.36: Model prediction of the blade loading for H02 based on B05 Wind Turbine for the undamaged case.
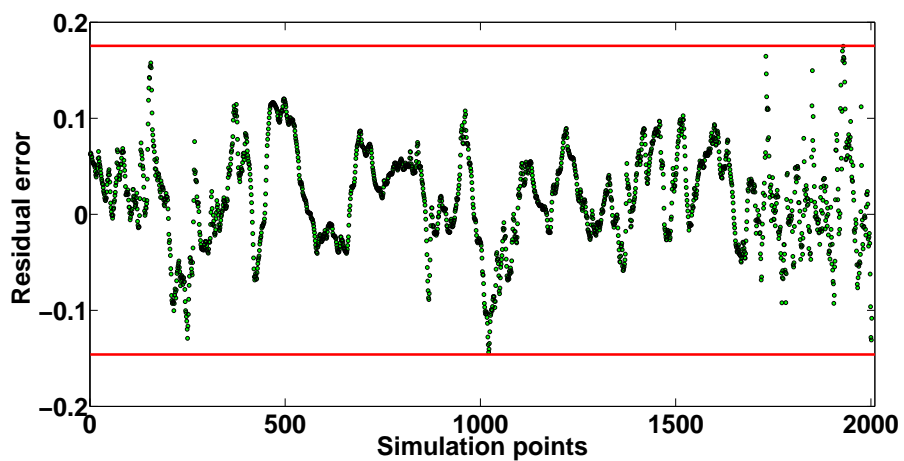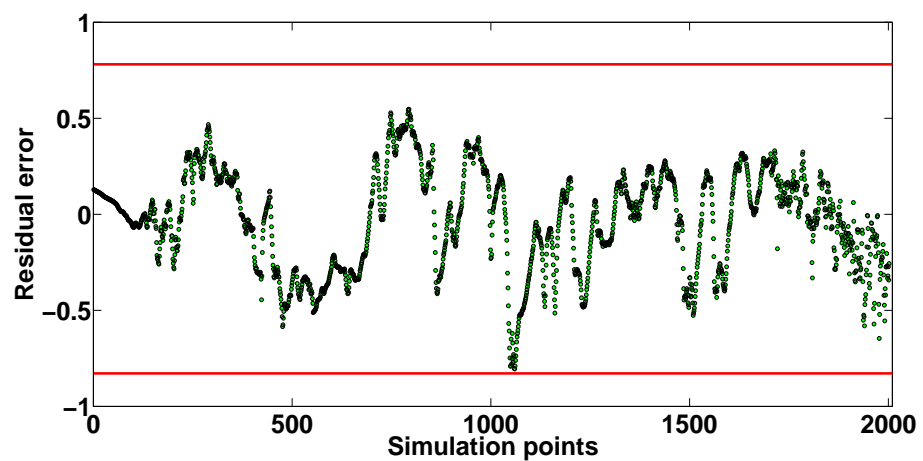

Figure 4.37: Residual error of the model prediction for H02 with no introduced change.
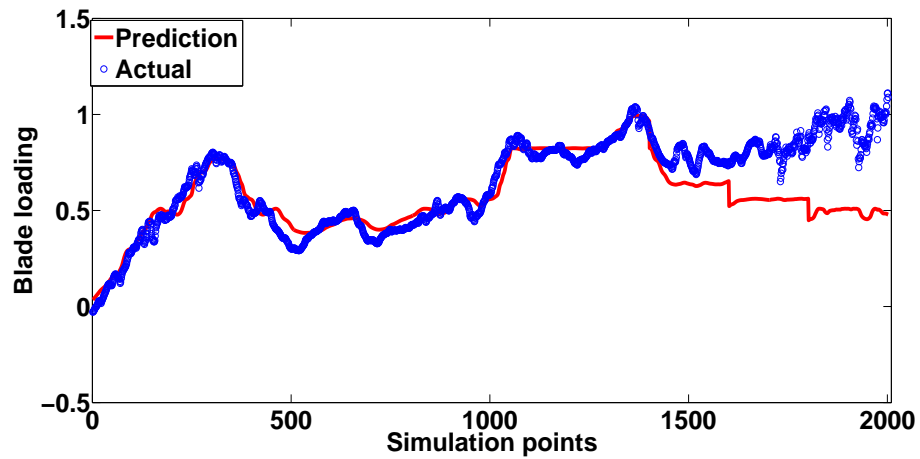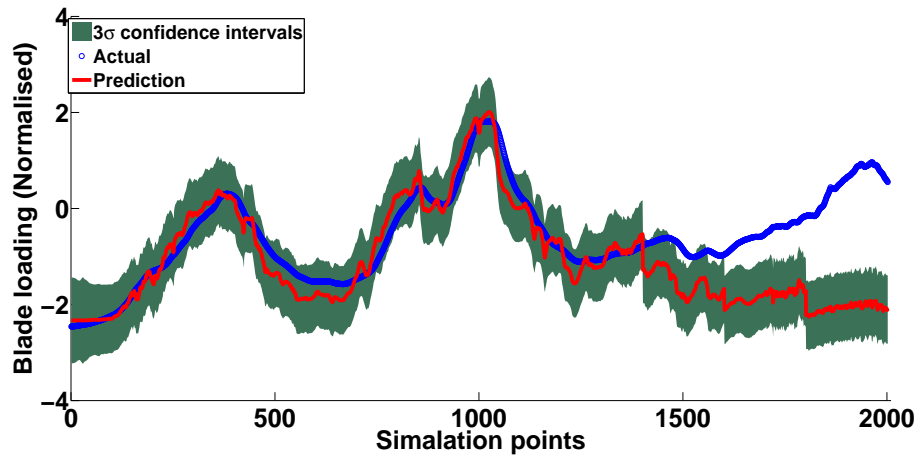
Figure 4.38: Model prediction of the blade loading for H02 based on B05 Wind Turbine for the damaged case.
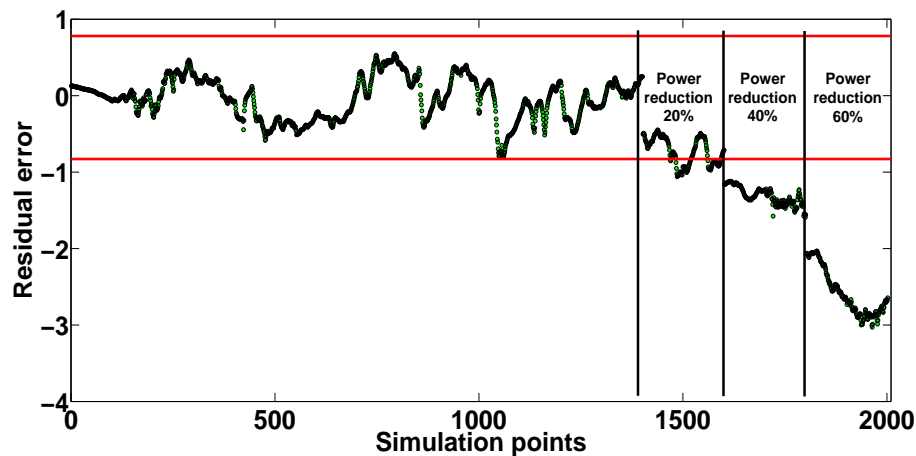
Figure 4.39: Residual error of the model prediction for H02 with introduced change.

4.10 Discussion

The central target of this work is to provide machine learning approaches to structural health monitoring of wind turbine blades by introducing advanced dimension reduction and novelty detection techniques.

It can be concluded that high frequency FRFs are a robust SHM feature for detecting or even locating damage in wind turbine blades. Furthermore, the results point out that they have the potential of reducing the actual number of sensors required to cover the entire body of the blade.

Also in this section, a preliminary study of a model regression analysis between power output and blade loading that can be used for early damage detection of a wind turbine was demonstrated. The results have shown that an automatic interpretation of SCADA data is feasible. This data driven view combined with machine learning approaches could provide important information to the offshore wind farm industry regarding the qualitative importance of different collections of SCADA data volumes regarding the monitoring of the turbine. In the current analysis only power output response and blade loading were used but the proposed approach can be extended by combining many different measurements of SCADA data by developing a population-based architecture for an online automated damage detection system for the wind farm in total.

Chapter 5

ROBUST OUTLIER DETECTION IN THE CONTEXT OF SHM

The key novel element of this chapter is the introduction of robust multivariate statistical methods into the Structural Health Monitoring (SHM) field through the use of the Minimum Covariance Determinant Estimator (MCD) and the Minimum Volume Enclosing Ellipsoid (MVEE). Robust outlier statistics are investigated, focussed mainly on a high level estimation of the "masking effect" of inclusive outliers, not only for determining the presence or absence of novelty - something that is of fundamental interest - but also to examine the normal condition set under the suspicion that it may already include multiple abnormalities. By identifying and detecting variability at an early stage, the prospects of achieving good generalisation and establishing a correct normal condition classifier may be increased. It is critical to highlight that there is no *a priori* division between the damaged or undamaged condition data when the algorithms are implemented, offering a significant advantage over other methodologies. In summary, this chapter explains the technical details of the robust distances and introduces a new scheme for feature selection and damage detection in SHM by exploiting robust multivariate outlier statistics.

5.1 Definition

In SHM and Condition Monitoring, a relatively large number of variables are extracted. One of the first simple steps is to observe points in the data that appear as outliers. Outliers are often regarded as a sensor error or noise but most of the time these abnormal observations might carry important information. Detecting possible outliers is a vital step when working with machine learning algorithms (classification or regression) as they otherwise may lead to a wrong model or parameter estimates and thus give wrong results.

There are many definitions regarding what is an outlier. Hawkins [98] and Barnett and Lewis [99] give two general definitions. Barnett and Lewis indicate that "an outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs". Hawkins defines an outlier "as an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism". But which are the main mechanisms that produce outliers? According to Hawkins [98] the are two basic mechanisms that are responsible for the contamination of data with outliers. The first mechanism assumes that the data is coming from heavy tailed distributions (like the $t-distribution$). The second mechanism assumes that data is coming from different kinds of distributions. One of these may be the general distribution that generates the "good" measurements and the other one may be the "infecting" distribution, that gives the contaminated measurements. If the contamination distribution has tails that are heavier that those of the good, then there will be a possibility for the observations coming from the infecting distribution to appear as outliers.

The categorisation of outlier detection techniques can be divided between univariate methods and multivariate methods. Most of the initial outlier detection tools were based on univariate methods and they were relying on the assumption of an underlying known distribution of the observation family. A lot of discordancy detection methods assume that the distribution of the data parameters and the nature of the outliers are *a priori* known [99]. Of course, these simplified assumptions are not realistic with real applications. In multi-variable observations when each of the variables is regarded as independent; outliers can not always be detected. The discordancy test for outliers must take into account the relationship between the different variables that are present.

A further step in classifying outlier detection methods is between parametric (or statistical) methods and non-parametric. In parametric methods the assumption of a known underlying distribution of the observations is considered [25, 98–100] or either statistical estimates of unknown distribution parameters are considered [29, 101]. The main mechanism behind these methodologies is indicating as outliers these points that deviate from the assumed model parameters.

Distance-based methods are a major part of the non-parametric outlier detection techniques and are mainly based on extracting local distance measures [23, 102–105]. Clustering and spatial outlier methods are also present in the literature [106–108].

5.2 Motivation and novelty detection

Outlier detection methods have been utilised for many different applications and cover a broad range of fields of research such as econometrics, computer sciences, medical and biological sciences, meteorology or even political science [25, 27, 31, 98, 99, 109–113].

Before presenting the robust solution for detecting groups of outliers, it has to be mentioned that a significant parameter when someone is dealing with outliers is the *breakdown point*. The breakdown point identifies the ability of an estimation technique to find the unbiased estimates for parameters with data containing contaminated observations. It is generally defined as the smallest percentage of contamination (e.g. observations that do not follow the general pattern of the population) that can force the estimator to obtain biased values distant from the true estimations [25, 98–100]. The maximum value that can be considered for a breakdown point is 50% as, it becomes clear that when more than 50% of the data points are infected and labelled as outliers then obviously it is impossible to distinguish which of the observations is entitled the characteristic of the "good" or "bad point".

The idea of novelty detection by utilising outlier analysis for structural damage detection is not new, as it was introduced and developed in [9, 65–68]; however, effective use of the technique still faces some issues. Detecting outliers in multivariate data often proves to be more difficult than in univariate data because of the additional "space" within which a potential outlier may hide. If one is confident that Gaussian statistics apply the shape and size of a multivariate dataset may be characterised

by a mean vector and a covariance matrix. If this is established, a quantitative discordancy test may be applied to help evaluate whether a new observation is likely to have come from a particular condition. The classic discordancy measure used in the previous studies is the Mahalanobis Squared-Distance (MSD), which is given by the following equation,

$$D_i^2 = (\{x_i\} - \{\mu_x\})^T [\Sigma]^{-1} (\{x_i\} - \{\mu_x\}) \tag{5.1}$$

where $\{x_i\}$ is the potential outlier, $\{\mu_x\}$ is the mean of the sample observations and $[\Sigma]$ is the sample covariance matrix. The mean and covariance matrix could be inclusive or exclusive measures. That is to say, that the statistics may or may not have been computed from data where outliers are already present. Generally, in many different fields the test set (outlier) is not known *a priori* and an inclusive approach is a necessity.

The MSD tells one how far away a specific measurement is from the centre of the training data cloud, relative to the size of the cloud. The main disadvantage of the classical MSD is that suffers from a multiple outlier "masking effect". If there were groups of outliers already present in the training data, they would have a critical influence on the sample mean and covariance in such a way that they would subsequently indicate small distances on new observations or outlying data and thus cause the outliers to become invisible. The arithmetic mean and unbiased covariance matrix of MSD are statistics that suffer heavily from multiple outliers present in the data. Specifically, when outliers from a cluster cloud that lie inside the data are present then they will directly move the arithmetic mean towards them and even expand the classical tolerance ellipsoid in their direction [25]. The MSD arguably performs best when a training set of normal condition samples alone is established and then a testing data set is compared using this training mean and covariance matrix as *exclusive* measures. However, in the context of SHM or condition monitoring this situation presents a series of drawbacks regarding the use of multivariate statistics. In many practical applications the normal condition may be contaminated for several different external factors and thus give a misleading representation of the normal condition and mask faults due to these external factors.

The motivation for new approaches is clear, as all the outliers discussed here are considered inclusive, this means that there is no *a priori* knowledge assumed of the damaged or undamaged condition. The need of revealing groups of outliers when

someone is not 100% confident about the normal condition is demanding in SHM.
Of course, it will be demonstrated that the robust distances could be used as a
reliable damage detection tools without the need of a pre-defined normal condition.
It is very important to mention that in multivariate data, one can lose the ability of
viewing directly the full data set. Visualisation of data is limited to two or three
dimensions and a common procedure is (as seen in Chapter Three) the introduction
of dimension reduction techniques (like PCA). But the multi-variable nature of the
data and the complex covariance between the variables may result a masking effect of
some outliers or even indicate some points as abnormal [25, 27, 114, 115]. And thus,
using plots from multivariate analyses (like linear PCA) outliers or classes of different
patterns may not be visible as estimation of covariance matrices or correlation of the
data structure may be biased by the outliers presence.

It will be shown in this study that an inclusive MSD computation is not an efficient
method to reveal novelty changes when groups of outliers are present in the normal
condition set. To overcome this difficulty, robust estimators of the mean and
covariance matrix that can resist inclusive outliers are needed and for this reason
the Minimum Volume Enclosing Ellipsoid (MVEE) and the Minimum Covariance
Determinant Estimator (MCD) are used.

In this chapter, technical details of the robust methods are given, followed by a
demonstration example (a simulation application). Later a more extensive view on
challenging real life experimental applications of the Z24 bridge where environmental
trends are revealed and a more complex experiment on an aircraft wing, where
boundary conditions are changing, are presented.

5.3   The Minimum Volume Enclosing Ellipsoid (MVEE)

The description here of the algorithm that is used closely follows the paper [116]
where the idea behind the minimum volume enclosing ellipsoid (MVEE) is motivated
by the problem of finding an approximation to the projected area of an object of
interest in an image plane [116]. There, the proposed method attempts to find an
ellipse with minimum volume that encloses all the pixels of a coloured box [116].

The method and algorithm of [116] were used here and via a re-arrangement of

the MVEE method can be applied to a new novelty detection method by revealing inclusive outliers. A much more detailed analytical description of the MVEE can be found in [116]. If the MVEE can be identified, then outliers can be detected as they are essentially points on or near the boundaries of the minimum volume ellipsoid. Outlier detection can be used later for feature selection or as a damage index.

Khachiyan and Todd [117, 118] established a linear-time reduction of the MVEE problem to the problem of computing a Maximum Volume Inscribed Ellipsoid (MVIE) in a polytope described by a finite number of inequalities. Therefore, the MVEE problem can also be solved using the algorithms developed for the MVIE problem [116–120].

Consider a set of m points in an n-dimensional space: $S = \{\{x_1\}...\{x_m\}\} \in R^n$. Denote the minimum volume enclosing ellipsoid of the set S by *MVEE(S)*. The ellipsoid should have positive volume and in centre form is given by [116]:

$$E = \{\{x\} \in R^n | (\{x\} - \{c\})^T [A](\{x\} - \{c\}) \leq 1\} \tag{5.2}$$

where $\{c\} \in R^n$ is the centre of the ellipse E and $[A] \in S_{++}^n$ (which is the set of $n \times n$ positive definite matrices), describes the axes. The points $\{x_i\}$ of the multivariate set S should each satisfy the constraint:

$$(\{x_i\} - \{c\})^T [A](\{x_i\} - \{c\}) \leq 1 \tag{5.3}$$

The volume of E which will be minimised is given by [116],

$$vol\,(E) = \frac{u_0}{\sqrt{\det([A])}} = u_0 \det\left([A]^{-1}\right)^{-1/2} \tag{5.4}$$

Where u_0 is the volume of the unit hypersphere in dimension n. In summary, the problem of determining the ellipsoid of least volume containing the points of S is equivalent to finding a vector $\{c\} \in R^n$ and an $n \times n$ positive definite symmetric matrix $[A]$ which minimises $det([A]^{-1})$ subject to the constraint (5.3). The natural formulation of the problem is:

$$\textit{By varying } [A], \{c\}, \textit{ Minimise } det([A]^{-1}) \text{ subject to the constraints:}$$
$$(\{x_i\} - \{c\})^T [A](\{x_i\} - \{c\}) \leq 1 \textit{ when } i = 1, ..., m \tag{5.5}$$

There are several different methods available in order to obtain a solution of the problem; the one presented here is the dual formulation method based on Khachiyan's algorithm [116, 117, 119, 120]. It is essential to briefly explain Khachiyan's algorithm [117, 118], its transform of the problem into a convex optimisation problem and its solution via a dual formulation by applying a conditional gradient method.

A lifting of the space S to R^{n+1} is given by $S' = \{\pm q_1, ..., \pm q_m\}$ where $\{q_i\}^T = [\{x_i\}^T, 1]$ for every $i = 1, ..., m$. In this case every point $\{x_i\}$ is lifted to the hyperplane $H = \{(\{x\}, \{x_{n+1}\}) \in R^{n+1} \mid \{x_{n+1}\} = 1\}$. By lifting into the higher dimensional space (in this case R^{n+1}). S' is made centrally symmetric i.e. $S' = -S'$ and $MVEE(S')$ is therefore centered on the origin. The minimum volume enclosing ellipsoid of the original problem is recovered as the intersection of H with the $MVEE$ containing the lifted points $\{q_i\}$, $MVEE(S) = MVEE(S') \cap H$ and the optimisation problem becomes [116, 118]:

$$By\ varying\ [M] \in S_{++}^{n+1},\ Minimise\ log(det([M]^{-1}))$$
$$subject\ to\ the\ constraint: \qquad (5.6)$$
$$\{q_i\}^T[M]\{q_i\} \le 1 \ \forall\, i = 1, ..., m.$$

Now that the minimum-volume ellipsoid of the initial problem is re-arranged as the intersection of the hyperplane H and the MVEE centered at the origin containing the lifted points, a concave optimisation problem is formed and is solved by utilising a conditional gradient descent method. Khachiyan [117, 118] observed that this is a line-search problem with a closed-form solution. For more details of the applied algorithm and method, readers can look to [116–118].

Briefly, if $[P]$ is a matrix of m points in n dimensional space ($n \times m$) where the columns are vectors $\{p_i\}$, $[P] = [p_i, ...p_m]$, then $[Q]$ has column vectors $\{q_i\}$ with $[Q] = [q_1, \ldots q_m] = \begin{bmatrix} [P] \\ [1]^T \end{bmatrix} \in R^{(n+1)\times m}$. The Lagrangian dual problem is given by

[116, 118]:

$$Maximise \ \{z\} \rightarrow log(det(V(\{z\})))$$

$$subject \ to \ the \ constraint:$$

$$[1]^T z = d + [1] \ for \ \{z\} \geq 0 \tag{5.7}$$

$$where \ V(\{z\}) = [Q]diag(\{z\})[Q]^T \ and \ \{z\} \in R^m$$

$$is \ the \ decision \ variable.$$

The change of variable $z = (d + 1)u$ results in the following dual problem:

$$Maximise \ \{u\} \rightarrow log(det(V(\{u\}))) \ subject \ to \ the \ constraint:$$
$$[1]^T \{u\} = 1 \ for \ \{u\} \geq 0 \tag{5.8}$$

where $V(\{u\}) = [Q]diag(\{u\})[Q]^T$ and $\{u\} \in R^{(m \times m)}$. This problem is a concave optimisation problem and is solved by utilising a conditional gradient descent method. The ascent direction $\triangle u$ and the step size α have to be defined. The gradient of the objective function of equations (5.8) is $g(\{u\}) = [g_1(\{u\}), ..., g_m(\{u\})]^T$ and as a result:

$$g_i(\{u\}) = \frac{\partial logdetV(\{u\})}{\partial u_i} = \{q_i\}^T V(\{u\})^{-1}\{q_i\} \tag{5.9}$$

First $\{u\}$ is initialised as $\{u\} \leftarrow \frac{1}{n}\{1\}$ and then a repeating iteration procedure is followed. The descent direction is given by $g_i(\{u\}) = \{q_i\}^T V(\{u\})^{-1}\{q_i\}$ and if $j = \text{argmax} g_i(\{u\})$ then $\triangle\{u\} = \{e_j\} - \{u\}$ is the ascent direction. A line search seeks to solve the problem: $\text{maxlogdet} V(\{\bar{u}\} + a(\{e_j\} - \{\bar{u}\}))$ where $\{\bar{u}\}$ is the iterate value for every $a \in [0, 1]$. The final closed form solution is given by,

$$a \leftarrow \frac{g_i(\{u\}) - (n + 1)}{(n + 1)(g_i(\{u\}) - 1)} \tag{5.10}$$

and the update is given every time by $u \rightarrow u + a\triangle u$ until a stopping criterion is satisfied. In practical terms the essential part is to compute, via the previous solution the parameters of the covering ellipse. Consider the lifting points $\{q_i\}$, then the problem becomes [116, 118]:

$$Minimise \ [M] \leftarrow log(det([M]^{-1})) \ subject \ to \ the \ constraint:$$
$$\{q_i\}^T [M]\{q_i\} \leq 1 \ when \ i = 1, ..., m \ and \ [M] > 0 \tag{5.11}$$

The Lagrangian of the previous problem is given by [116]:

$$L\left([M], \{z\}\right) = -logdet[M] + \sum_{i=1}^{m} \{z_i\}(\{q_i\}^T[M]\{q_i\} - 1) \tag{5.12}$$

By differentiation (8) one obtains:

$$\frac{\partial L}{\partial M} = -[M]^{-1} + \sum_{i=1}^{m} \{z_i\}\{q_i\}\{q_i\}^T = -[M]^{-1} + [Q][Z][Q]^T = 0 \tag{5.13}$$

where $[Z] = diag(\{z\})$ and $[Q] = [q_1, ..., q_m]$. When a positive define matrix $[M]^* \in R^{(n+1)\times(n+1)}$ is optimal for the dual problem (5.11) and the Lagrangian multipliers $\{z\}^* \in R^m$ then one has that: $V(\{z\}^*) = [Q][Z][Q]^T = ([M]^*)^{-1} = (d+1)V(\{u\}^*)$ and given that $\{q_i\}^T = [\{x_i\}^T, 1]$ the equation of the ellipsoid is given by [116]:

$$MVEE\left(S\right) = \left\{x \in R^n | \{q\}[M]^*\{q\}^T \leq 1\right\}$$
$$= \left\{x \in R^n | \frac{1}{d+1}\{q\}^T V(\{u\}^*)^{-1}\{q\} \leq 1\right\} \tag{5.14}$$

In turn, after obtaining the dual problem solution from (5.8), the equation of the ellipse can be found [116].

$$V\left(\{u\}\right) = [Q]diag\left(\{u\}\right)[Q]^T = \begin{bmatrix} [P]diag(\{u\})[P]^T & [P]\{u\} \\ ([P]\{u\})^T & [1]^T\{u\} \end{bmatrix} \tag{5.15}$$

which can be re-written as [116]:

$$V\left(\{u\}\right) = \begin{bmatrix} I & [P]\{u\} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} [A]^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ ([P]\{u\})^T & 1 \end{bmatrix} \tag{5.16}$$

where $[A]^{-1} = [P]diag(\{u\})[P]^T - [P]\{u\}([P]\{u\})^T$ and the inverse is given by the equation [116]:

$$V(\{u\})^{-1} = \begin{bmatrix} I & 0 \\ -([P]\{u\})^T & 1 \end{bmatrix} \begin{bmatrix} [A] & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & -[P]\{u\} \\ 0 & 1 \end{bmatrix} \tag{5.17}$$

and if $[P]\{u\} = \{c\}$ then one obtains:

$$\{q\}^T V(\{u\})^{-1}\{q\} = (\{x\} - \{c\})^T[A](\{x\} - \{c\}) \tag{5.18}$$

Finally the ellipsoid of the dual solution is given by [116, 118]

$$MVEE\,(S) = \left\{x \in R^n | (\{x\} - \{c\}^*)^T [A]^* (\{x\} - \{c\}^*) \le 1\right\} \tag{5.19}$$

$$[A]^* = \frac{1}{d}([P]diag\,(\{u\}^*)\,[P]^T - [P]\{u\}^*([P]\{u\}^*)^T)^{-1} \tag{5.20}$$

The discordancy test (MVE index) is similar to equation (5.1) and calculates the squared distance from the centre of the ellipse to each data vector. In order to illustrate how the algorithm works, a data matrix of dimensions 3×1000 was constructed with each individual element a randomly generated vector from a normal distribution with zero mean and unit standard deviation and the minimum volume enclosing ellipsoid was calculated. A tolerance parameter can be used to allow a proportion of points to escape the ellipse. This is determined in much the same way as the *soft margin* used in training a support vector machine [116]. The tolerance of the algorithm was set to 0.1 in Fig.5.1 and 0.001 in Fig.5.2, in order to demonstrate the difference on covering all the extreme false positives values. It can be noticed that when the error tolerance is 0.1 there are some points that escape from the ellipse but when the error tolerance is increased the ellipse covers much more points (except two).
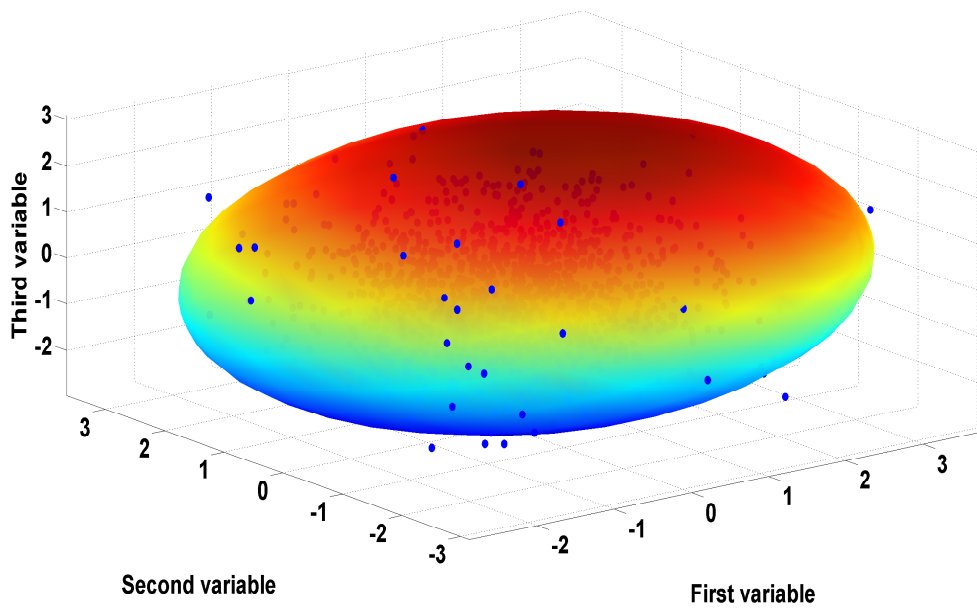
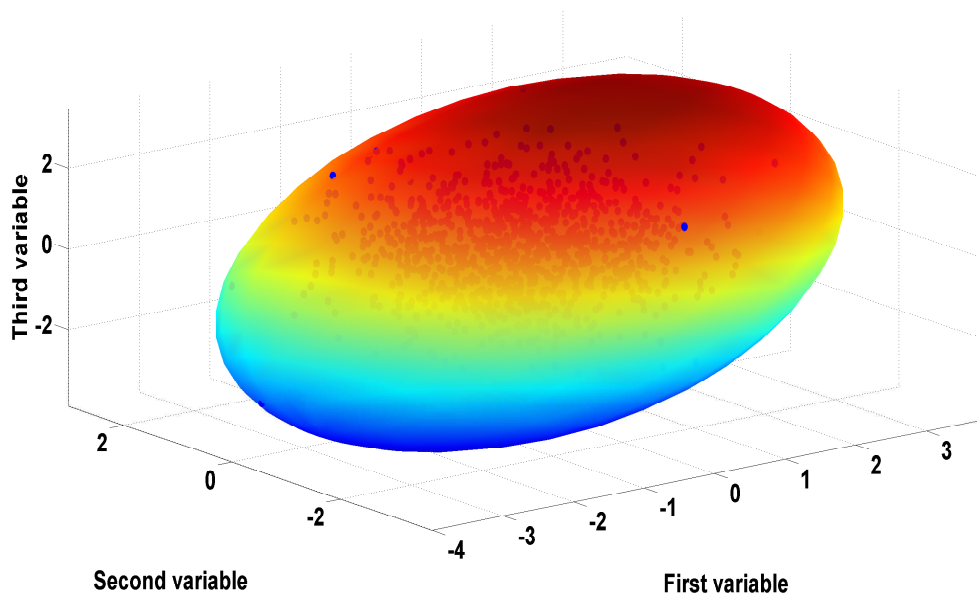Figure 5.1: Error 0.1 in the solution of MVEE with respect to the optimal value (tolerance).

Figure 5.2: Error 0.001 in the solution of MVEE with respect to the optimal value (tolerance).

5.4 The Minimum Covariance Determinant Estimator (MCD)

The application of robust computation of location and covariance estimation of multivariate data is of significant interest in the investigation for inclusive outliers. The second method that is introduced here is the minimum covariance determinant estimator (MCD). The computation of the MCD estimator is not a trivial procedure and requires an extensive calculation. In the current study, the FAST-MCD algorithm is implemented [25–27, 109, 121, 122]. The algorithm is given in detail in the references [25–27, 109, 121, 122], and the code was provided via a statistical Matlab library called LIBRA [121]. A brief description of the algorithm is provided by explaining the basic steps of the FAST-MCD technique.

A multivariate data matrix $[X] = (\{x_1\}, ..., \{x_m\})^T$ is assumed of m points in n dimensional space $(n \times m)$ where $\{x_i\} = (x_{i1}, ..., x_{in})^T$ is the observation. Robust estimates of the centre $\{\mu\}$ and the scatter matrix $[\sigma]$ of X can be calculated by the MCD estimator. The MCD tool looks for the $h(> \frac{m}{2})$ observations out of m whose classical covariance matrix has the lowest possible determinant. The raw MCD estimate of location is then computed from averaging these h points and the raw MCD estimation of scatter is the covariance matrix multiplied by a consistency factor. Based on these raw MCD estimates a reweighting step can be added in order to increase the sampling efficiency mentioned before. The advantage is that MCD estimates can resist up to $(m - h)$ outliers and in turn, the number h (or equally $a = \frac{h}{m}$) controls the robustness of the estimator. The highest resistance compared to contamination is achieved by calculating $h = \frac{(n+m+1)}{2}$. It is proposed that when a large proportion of contamination is assumed then $h = an$ with $a = 0.5$. Detecting outliers can be challenging when m/n is small because some data points can become coplanar. It is recommended [121] that when $\frac{m}{n} > 5$, a should be 0.5. Generally, the MCD estimates of location and scatter are affine equivariant which means that they are invariant under affine transformation behaviour. This is crucial as the underlying model is then immune to different variable scales and data rotations. Rousseeuw and Van Driessen (1999) [26] developed the FAST-MCD algorithm based on a Concentration step (C-step).

A C-step selects the h observations with the smallest distances and the scatter matrix with the lower determinant [26] and the main details are given now. Assume

$[X] = (\{x_1\}, ..., \{x_m\})$ and let $h - subset \ H_1 \rightarrow (1, ..., n)$ which is $|H_1| = h$. Then $\{\widehat{\mu_1}\} = \frac{1}{h} \sum\limits_{i \in H_1} \{x_i\}$ and $[\widehat{\Sigma_1}] = \frac{1}{h} \sum\limits_{i \in H_1} (\{x_i\} - \{\widehat{\mu_1}\})(\{x_i\} - \{\widehat{\mu_1}\})^T$ and when $[\widehat{\Sigma_1}] \neq 0$ relative distances can be defined as:

$$\{d_1(i)\}^2 = (\{x_i\} - \{\widehat{\mu_1}\})^T [\widehat{\Sigma_1}]^{-1} (\{x_i\} - \{\widehat{\mu_1}\}) \ for \ i = 1, ...m \qquad (5.21)$$

The procedure continues by selecting an appropriate $H_2 \rightarrow \{\{d_1(i)\}; i \in H_2\} = \{(d_1)_{1:m}, ..., (d_1)_{h:m}\}$ where $(d_1)_{1:m} \leq (d_1)_{2:m} \leq ... \leq (d_1)_{m:m}$ are the ordered distances and the then $\{\widehat{\mu_2}\}$ and $[\widehat{\Sigma_2}]$ are calculated based on H_2. In turn, one should have $\det([\widehat{\Sigma_2}]) \leq \det([\widehat{\Sigma_1}])$. If $\det([\widehat{\Sigma_1}]) > 0$, the C-step leads to $[\widehat{\Sigma_2}]$ with $\det([\widehat{\Sigma_2}]) \leq \det([\widehat{\Sigma_1}])$. When $\det\left([\widehat{\Sigma_1}]\right) = 0$ then the minimum value is obtained. This C-step repeated condition is followed in the algorithm until a stopping criterion is fulfilled. The stopping criterion is when $\det([\widehat{\Sigma}_{new}] = 0)$ or when $\det([\widehat{\Sigma}_{new}]) = \det([\widehat{\Sigma}_{old}])$. The calculation chain of determinants is converged in finite steps, as one has a finite number of h-subsets. However, the final calculation of $\det([\widehat{\Sigma}_{new}])$ may not converge to the global minimum of the MCD objective function. This is the reason why an approximation of the MCD solution is obtained by introducing a large number of random initial conditions for H_1 and via the C-step the lowest determinant is kept. In practical terms a resampling technique is followed.

5.5 Threshold calculation

Setting an appropriate threshold in the absence of any damage-state data, as is the case in this study, is a non-trivial task. In many studies presented in the published literature, the assumption made is that the multivariate data are normally distributed, with the MSD subsequently approximated by a chi-squared distribution in p-dimensional space. For the purposes of this study, another method for setting the threshold was followed; a Monte Carlo simulation based on extreme value statistics was used. The procedure that was conducted in order to calculate the threshold is as follows:

- A $p \times n$ *(number of dimensions number of observations)* matrix is constructed with each individual element a randomly generated number from a normal distribution with zero mean and unit standard deviation.

- The discordancy value as described in previous sections is evaluated for all

matrix values, where the mean, centres of ellipse, robust and classic covariance and minimum volume ellipse variance matrix are inclusive. The largest (i.e. extreme), value recorded for each trial matrix is stored.

- The process is repeated for a large number of trials (10000) in order to generate an array of "extreme" distance calculations. Next, all the values are ordered in terms of magnitude. The critical values (alpha value, α) can take different values such as 5% or 1% for a test of discordancy. In this paper α is set equal to 1% giving a 99% confidence limit.

5.6 Simulated structure

The simulated system that is used in this part in order to demonstrate the different novelty detection techniques is a three degree-of-freedom lumped-parameter system where the first mass is also connected to ground through a nonlinear spring with cubic stiffness. The specific values of the undamaged system parameters were $m = 1$, $c = 10^{-4}$, $k = 10^4$ and $k_{nl} = 10^9$ where m is the mass, c is the damping coefficient, k and k_{nl} are characterising the linear and nonlinear springs respectively, see Fig.5.3. The fault in this system was simulated by reducing the linear stiffness between masses 1 and 2 by 0.5, 1, 2 and 3% of the original value. In order to implement the outlier techniques and generate an appropriate number of samples, the unfaulted and faulted data are copied several times and each copy was subsequently corrupted with different Gaussian noise vectors of r.m.s. value of 0.05. The feature that was used for the detection process was the transmissibility function (see appendix A) between masses 1 and 2, see Fig.5.4. This was calculated by simulating the response of the system to a random excitation. For the purpose of this demonstration only the magnitude of the transmissibility was used. The transmissibility function was sampled at 50 points in the frequency domain to give the feature used in this study. The testing features were two different matrices of dimensions 2200×50 where the first 2000 observations include the unfaulted system and the next 200 observations include the faulted system (200 data points for each faulted approximation). Regarding the MVEE method, the tolerance of the algorithm was set to 0.001 and regarding the MCD technique the value of α was 0.5 and the number of subsets h was set to $h = am$ which resulted in a value of 1100.

In Figs.5.5-5.8 can be seen the results of the analysis. The MVEE method seems

to be more efficient in unmasking the inclusive outliers when stiffness reduction of 0.5,1,2% occurs compared to the MCD method. The classic MSD remains immune to inclusive outliers. When the stiffness reduction increases to 3%, both robust methods reveal the introduced outliers by giving a monotonic increase of novelty measure. Even when the stiffness is reduced by 3%, the classic MSD cannot detect any inclusive outliers. In all the cases, it is noticeable that the classic Mahalanobis distance is unable to find multiple outliers. The MCD and MVEE methods could be very useful tools in feature selection, novelty detection (due to change of environmental or boundary conditions) and damage detection when *no a priori* separation of health condition is known within the training data and thus a true normal condition could not be established.
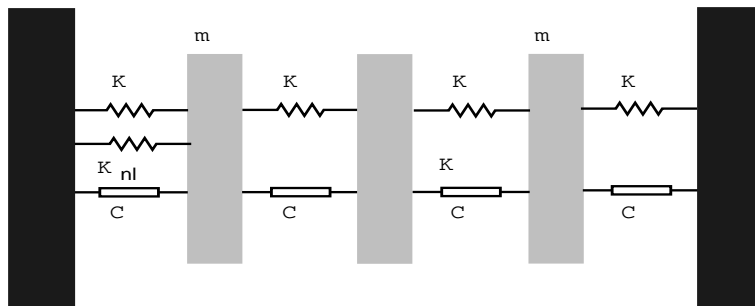
Figure 5.3: The three-degree-of-freedom simulated system.
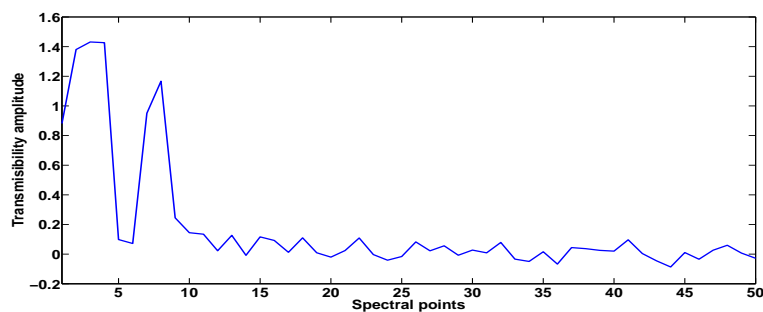
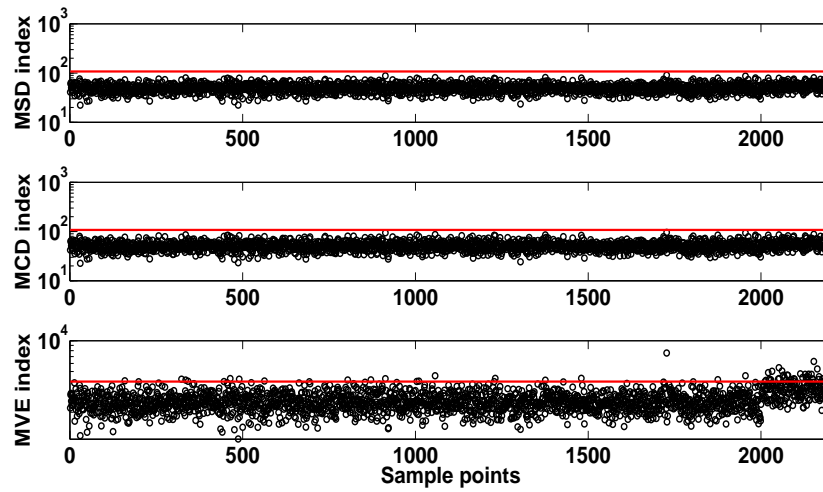Figure 5.4: Transmissibility feature.

Figure 5.5: Novelty detection for stiffness reduction 0.5%.
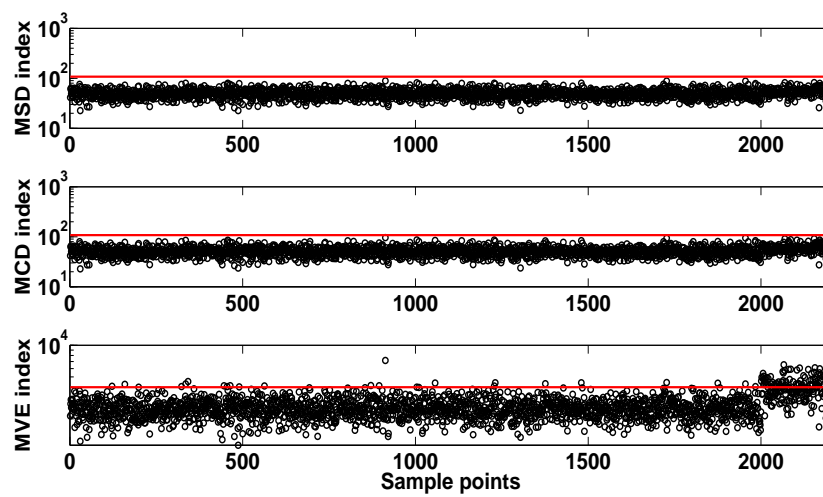
Figure 5.6: Novelty detection for stiffness reduction 1%.
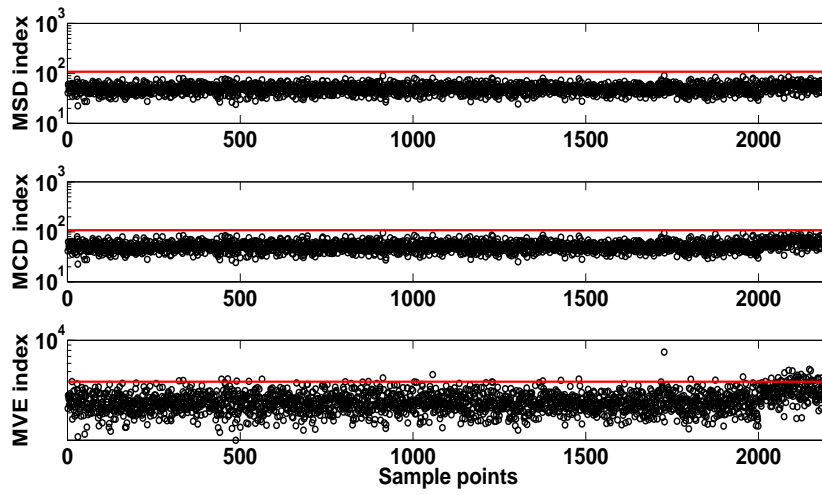
Figure 5.7: Novelty detection for stiffness reduction 2%.
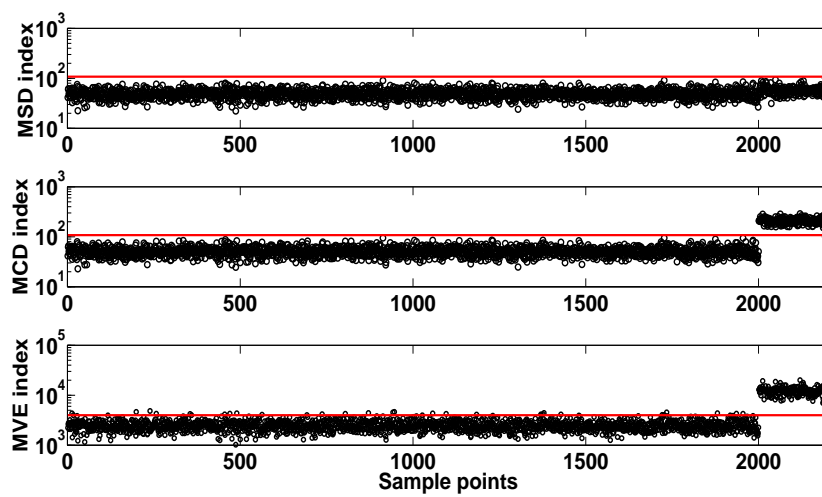
Figure 5.8: Novelty detection for stiffness reduction 3%.

5.7 Conclusion

The central target of this study is to provide alternative machine learning approaches to SHM with the use of robust multivariate statistical novelty detection. Damage detection and identification is a procedure that is hierarchical in nature. At its most sophisticated, diagnosis of the damage could include localisation, classification and severity assessment and even go so far as to estimate the time-to-failure of the structure.

In this chapter a detailed technical description of the Minimum Volume Enclosing Ellipsoid (MVE) and the Minimum Covariance Determinant Estimator (MCD) is given. These two robust methods are investigated and compared to the classical MSD index in the context of SHM. In the simulation example it can be clearly seen that both of the novel methods (in the SHM context) outperform the classical approach in multivariate discordancy tests regarding the inclusive outliers. The masking effect described in this section is so strong that MSD calculation can not detect any of the inclusive outliers. In the next chapter these two robust methods are investigated in depth through challenging experimental applications and validated.

Chapter 6

APPLICATIONS OF ROBUST OUTLIERS METHODS WITH CHANGING ENVIRONMENTAL AND OPERATIONAL CONDITIONS

In Chapter Four, the technical details of the Minimum Covariance Determinant Estimator (MCD) and the Minimum Volume Enclosing Ellipsoid (MVEE) were described. A simulated example was tested in order to validate the robust methods in the context of SHM. In this Chapter, real and more complex experimental examples are analysed. It will be shown that robust multivariate statistical analysis could prove an important step in revealing changing environmental and operational conditions.

6.1   Environmental changes of the Z24 Bridge

The Z24 Bridge was a concrete highway structure in Switzerland connecting Koppigen and Utzenstorf and in the late 1990s; before its demolishment procedure, it was used for condition monitoring purposes under the "SIMCES" project [123, 124].

Z24 is a pre-stressed bridge consisting of three spans and two lanes with an overall length of 60m (Fig.6.1). During a whole year of monitoring of the bridge, a series of sensor systems captured modal parameter measurements, as well as a family of environmental measurements such as air temperature, soil temperature, humidity, wind speed etc. The critical point in this benchmark project was the introduction of different types of real progressive damage scenarios towards the end of the monitoring year, (Table.6.1).

Figure 6.1: View of Z24 Bridge.

Sequence	Damage scenarios
1	Settlement of foundation
2	Tilt of foundation
3	Spalling of concrete at soffit
4	Landslide
5	Failure of concrete hinges
6	Failure of anchor heads
7	Number of post-tensioning tendon failures

Table 6.1: Progressive damage scenarios.

For the purposes of this study, the four natural frequencies that were extracted over a period of year, including the period of structural failure of the bridge are used. Fig.6.2 shows the four natural frequencies with values between 0-12 Hz. The beginning of the introduced failure occurs at observation 2496. It has to be mentioned that values of failed measurements have been removed. In the first instance it can be noted that there are some visible fluctuations between observations 1250-1460 but there are no dramatic visible fluctuations after the introduction of damage, making the frequencies sequence nonstationary. These fluctuations are highly related to periods of very cold temperatures under zero degrees Celsius and there is a direct connection
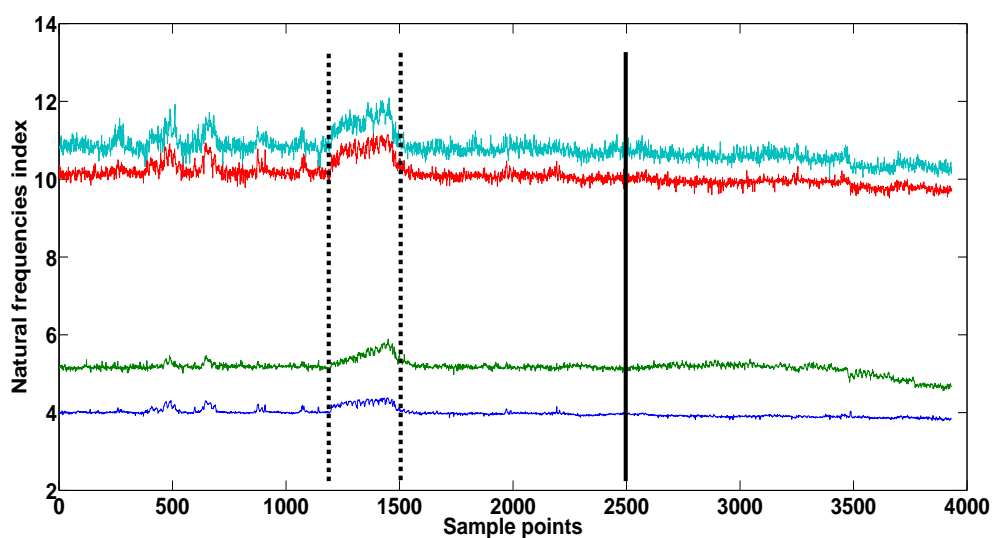
Figure 6.2: Time history of frequencies.

with increased stiffness based on the freezing of the asphalt layer of the bridge deck, see Fig.6.3. These large temperature fluctuations are suitable candidates in order to check the sensitivity of the robust outlier methods and the classic MSD index. The results of applying the MCD and MVEE robust methods on the data matrix (with dimensions 4 (natural frequencies)$\times$ 3932 (observations)) are shown in Fig.6.4. The first step is to observe that Mahalanobis squared-distance is insensitive to the multiple outliers that occur during the freezing period but also fails to uncover the damage presence. On the other hand, both the MCD and MVEE squared distances not only reveal the large fluctuations due to temperature but also reveal the multiple fault presence. The damage presence in the MCD index is visible after observation 3500. The MVEE index seems to be more sensitive, in this case showing outliers just after observation 2496 when the damage is introduced. It is essential to note that due to large temperature variations after observation 2496, the damage is not clearly detectable as the fluctuations cover the novelty. Once the outliers are revealed, a decision for a suitable feature has to be made and a more complicated algorithm has to be applied in order to remove external influences like environmental factors and create a suitable normal condition training set. This extra step is investigated in the next session by combining the robust distances with NLPCA and PCA algorithms.
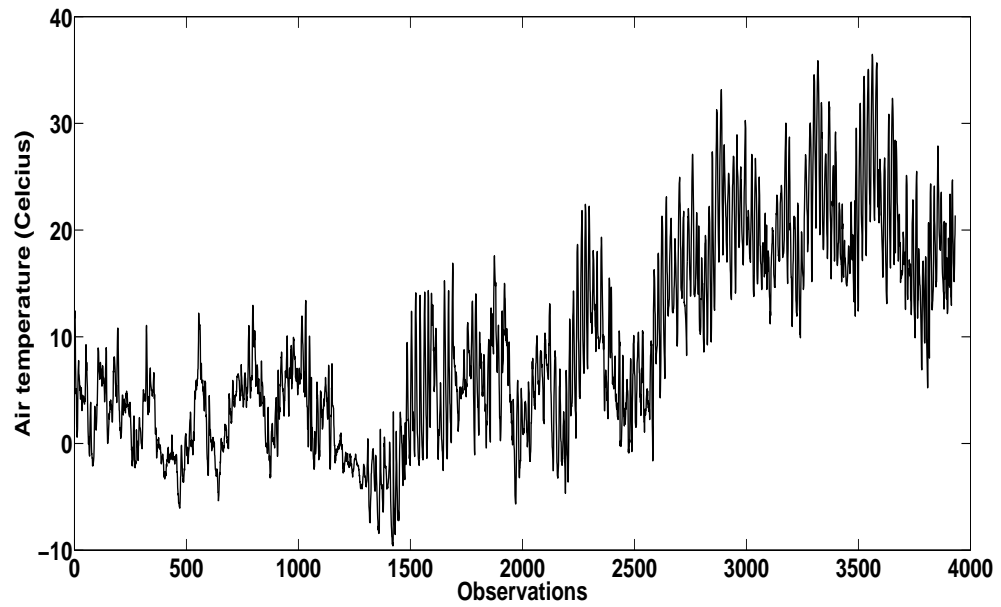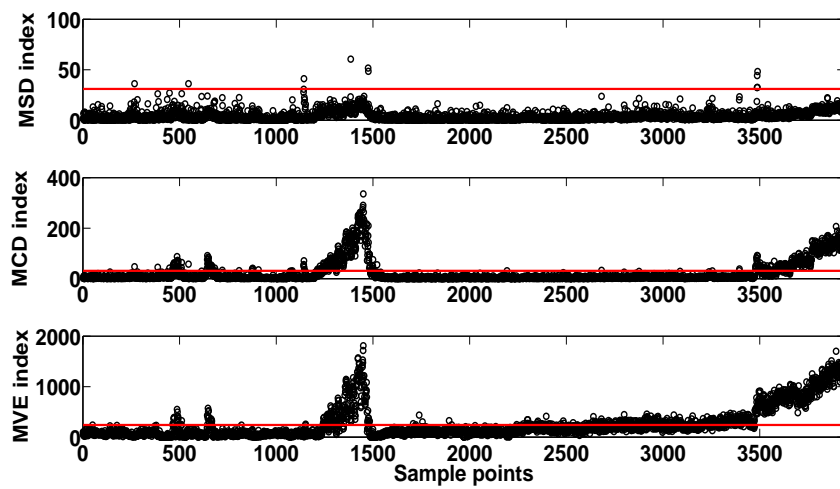
Figure 6.3: Time history of air temperature.

Figure 6.4: Novelty detection of Z24 Bridge.

6.2 Correlation between the natural frequencies

The robust multivariate statistics can be used via an alternative re-arrangement for a further and in depth exploration of data patterns, in order to reveal the variation (correlation between the variables) of the undamaged condition of the structure against the different operational conditions. Afterwards, advanced algorithms can be applied in order to establish a normal condition candidate which is able to cope with false alarms due to external influences on normal data like temperature variations in the Z24 bridge. The current analysis is an extension of the analysis performed in [124], regarding the linear or nonlinear relationship between modal frequencies of the Z24 Bridge. A further step will be applied by comparing PCA and NLPCA (AANN) algorithms regarding their ability to remove the environmental trends after revealing the frequencies correlation and establishing a normal condition.

In the first instance, the univariate MCD robust distance was calculated for each of the four natural frequencies in order to reveal each frequency's internal fluctuations, as shown in Fig.6.5. This analysis is critical, as important conclusions can be derived. The obvious observation could be that the second natural frequency leads the fluctuations as it reveals the dramatic index increase between the observations 1250-1460 where the cold temperatures are operating. Also, the index value of the damaged condition increases after observation 3500. The other three frequencies present a common pattern where again the MCD index increases during the cold period, but after observation 3500 the damage is not clearly noticeable. Based only on that someone could assume that the first, third and fourth frequencies do not offer vital information.

However, important conclusions can be found by looking at the presented patterns of Fig.6.5 more carefully. Between observations 430-520 and 590-690 two distinctive peaks can be seen that correspond also to temperatures below zero. An analogous index increase appears again when cold temperatures occur. To make it more simple and quantitative, the actual MCD index value with respect with the temperature can be checked. In turn, for the first, third and fourth frequencies when the temperature reaches values of -6^oC at point 643 the robust distance takes values of around 29 and when the temperature decreases to -10^oC at point 1420, it takes a value around 38. On the other hand, it is noticeable that for the second natural frequency there is a dramatic difference in the robust distance value between the two peaks before 800
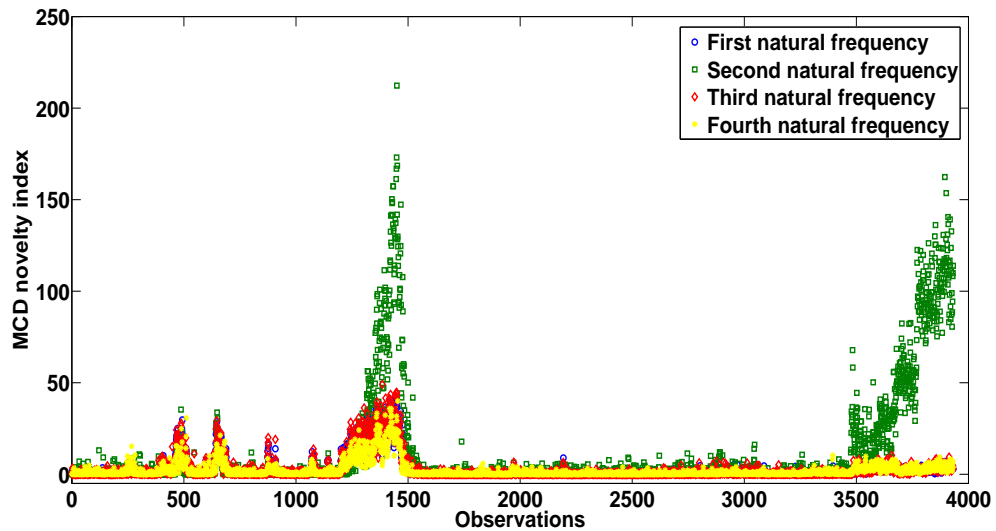
Figure 6.5: MCD univariate robust distances of four natural frequencies.

and the area of points 1250-1460 when again temperature is at values below zero. So, when the temperature reaches values of -6^oC at point 643 the robust distance takes values of around 33 and when the temperature decrease in -10^oC at point 1420 takes a value around 117. It is clear that a further analysis is needed.

It becomes desirable to reveal the correlation between the different frequencies in order to understand the characteristics of the structure behaviour. For this reason the MVE method was used in order to construct minimum covering ellipsoids around the frequency clouds and check geometrically how the distributions correlate with each other. The results can be see in Figs.6.6-6.11. In Table.6.2 the description of ellipsoid colours is given. The results reveal vital information. When the second frequency is present the nonlinearity is dominant as the covering ellipse of the damaged condition is shifted in respect to the normal condition and the centres pattern reveal a highly nonlinear connection as seen in Figs.6.6,6.9,6.10. This is something that was expected as the second natural frequency is driving the nonlinear behaviour. The interesting point is where the second frequency is not present. The distribution clouds between frequencies seem to follow a critical shifting in Fig.6.7,6.8 and 6.11 as the centre of the green ellipsoid is not inside the blue one. Also, in centre connections a weak nonlinearity can be observed between the first and third natural frequency in Fig.6.7.

Observation	Condition	Colour
1-1250	Undamaged	Blue
1251-1460	Cold temperature	Green
1461-2495	Undamaged	Black
2496-3932	Damaged	Red

Table 6.2: Description of minimum covering ellipsoids colours.

Figure 6.6: Minimum covering ellipsoids between first and second natural frequency.

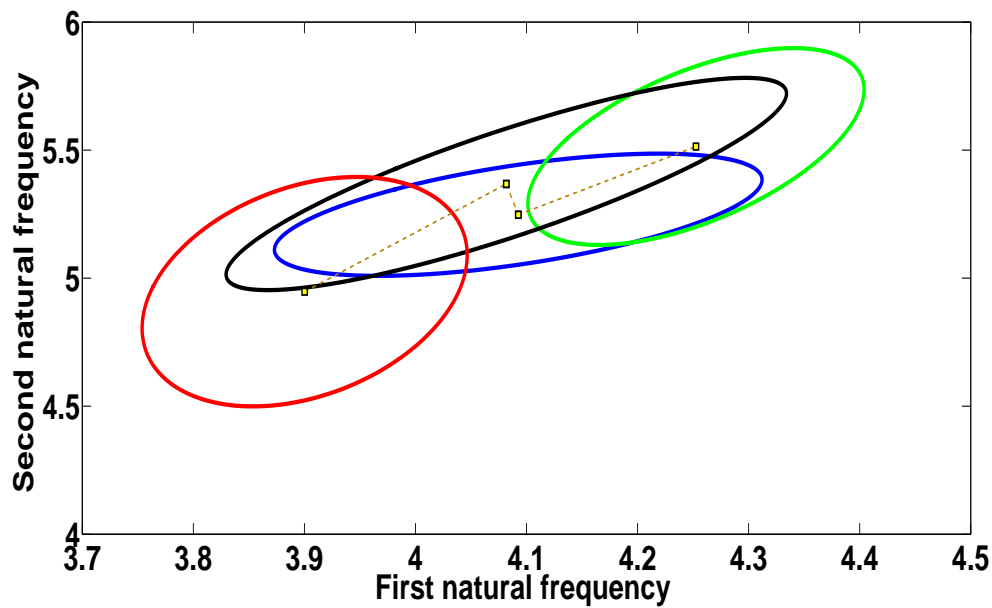Figure 6.7: Minimum covering ellipsoids between first and third natural frequency.

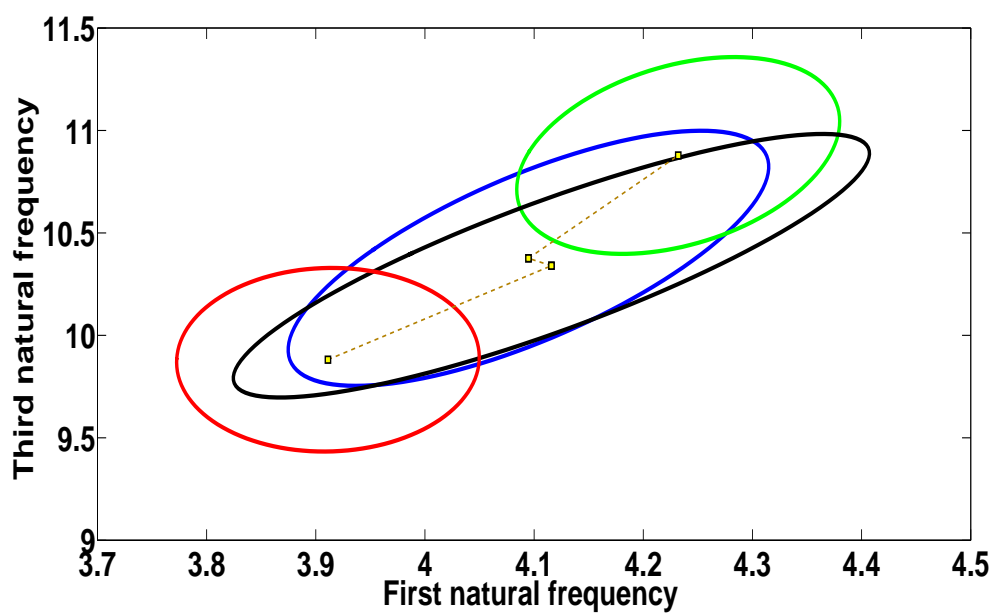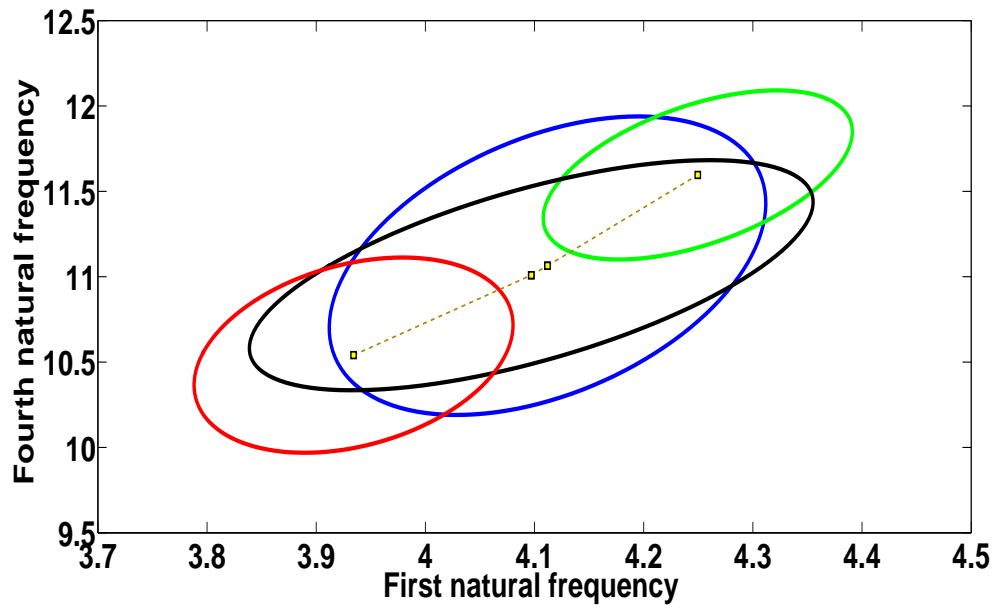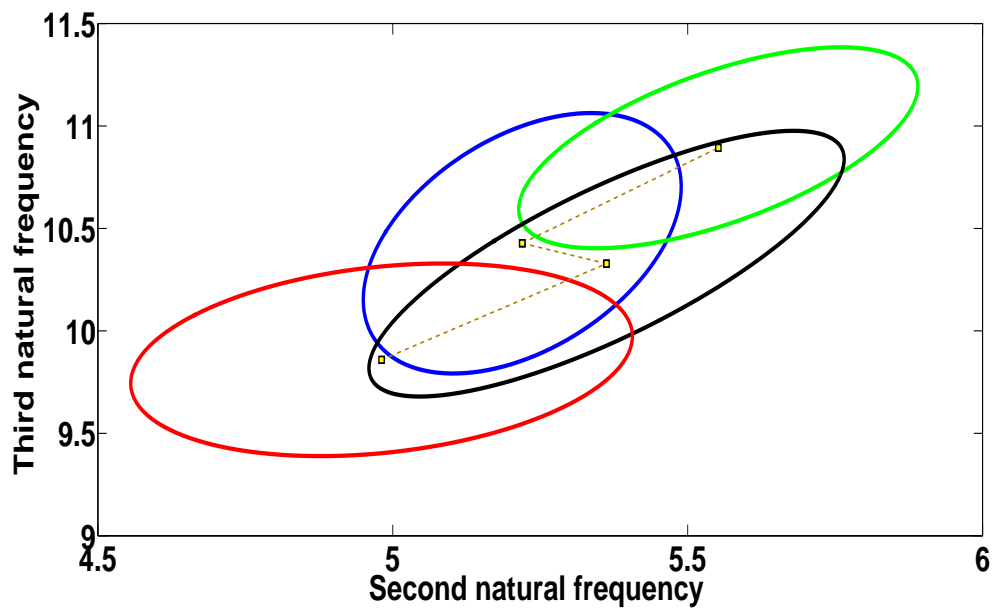Figure 6.8: Minimum covering ellipsoids between first and fourth natural frequency.

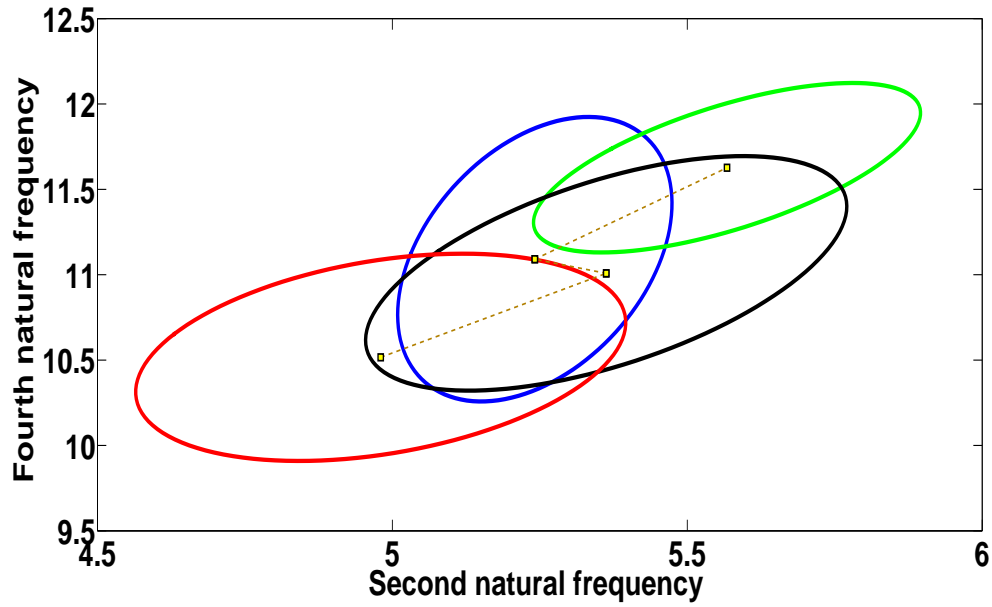Figure 6.9: Minimum covering ellipsoids between second and third natural frequency.

Figure 6.10: Minimum covering ellipsoids between second and fourth natural frequency.
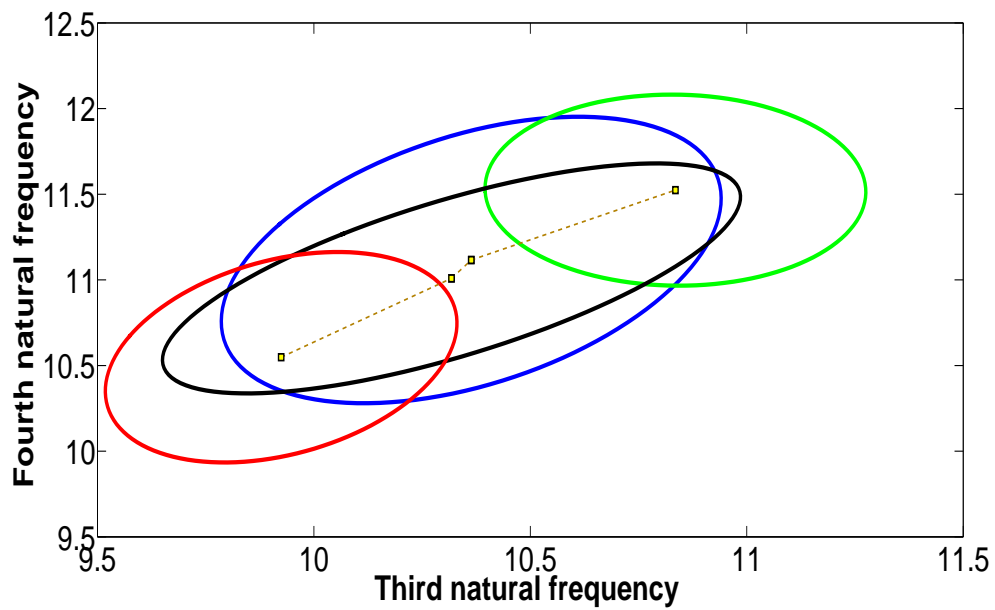

Figure 6.11: Minimum covering ellipsoids between third and fourth natural frequency.

To validate this conclusion, linear PCA and nonlinear PCA implemented via an AANN were applied by using as a training set the points 1-1000, including the two peaks observed previously. Results of the novelty detection techniques are presented in Figs.6.12-6.23 in the same fashion as the methods used in Chapter Three and Chapter Four by calculating the Euclidean distance between the algorithms output and the actual data.

The results display the test set and this is the reason that there are 2932 observations in *x-axis* with damage occurring at observation **1496**.

In both cases, one principal component for PCA and a single neuron in the bottleneck layer for the AANN were used. What can be seen from looking at PCA results is that the normal condition of the first, third and fourth frequency has a linear correlation with damage condition. And this is the reason that no novelty appears after observation 1496. On the other hand, NLPCA utilised by the auto-associator reveals the slight nonlinearity observed previously by indicating novelty when cold temperatures are present something that PCA could not reveal in full extent. Furthermore, damage is detectable after observation 1496. This indicates that AANN was able to uncover the faulty observation by capturing the pattern difference between the normal and damaged condition. A positive result can be observed in Fig.6.22 where the ability of auto-associator to generalise and learn the temperatures fluctuation in some extent can be seen. If one compares it with result in Fig.6.23 where PCA is used the difference is visible as PCA can not detect any damage. The training of AANN to understand the weak nonlinearities in the damaged condition compared to the normal condition is noticeable. This conclusion makes strong the advantages of auto-association tools despite their high complexity compared to linear methodologies.

The big drawback is that AANN is not always able to successfully generalise if in training set the nonlinearity trend is not present. This could be the future advantage of methods such as nonlinear cointegration (an approach to cointegration for SHM can be found in references [124–126]) or deep belief networks.

When, the second natural frequency is added as a new variable then things become more complicated. It is clear in Figs.6.6, 6.9 and 6.10 the different patterns between the cold temperatures in training condition and cold temperatures at points 1250-1460. The variation is so strong that PCA is able to reveal it. The results using NLPCA, are not better but again the different type of nonlinearity between the cold

temperatures in training condition and cold temperatures between points 1250-1460 can not be overcome as they are not similar patterns in training data.

Figure 6.12: AANN novelty index between first and second natural frequency.

Figure 6.13: PCA novelty index between first and second natural frequency.

Figure 6.14: AANN novelty index between first and third natural frequency.

Figure 6.15: PCA novelty between first and third natural frequency.

Figure 6.16: AANN novelty index between first and fourth natural frequency.

Figure 6.17: PCA novelty index between first and fourth natural frequency.
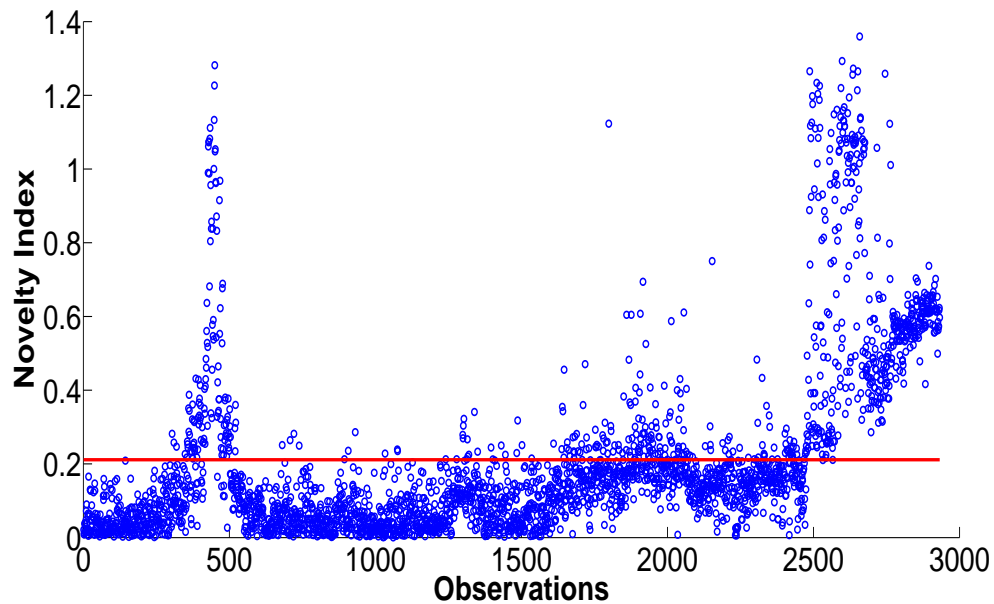
Figure 6.18: AANN novelty index between second and third natural frequency.
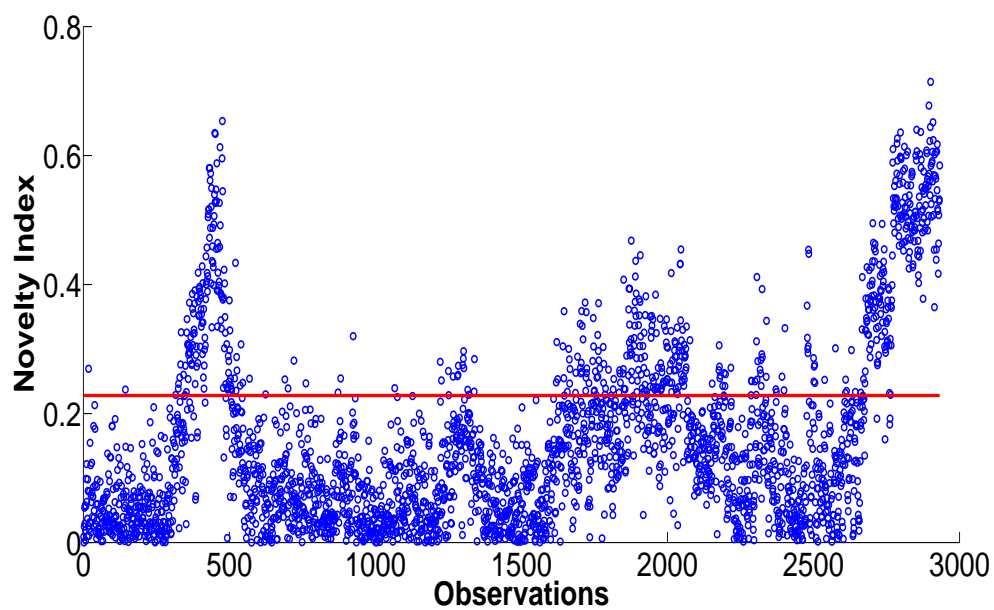
Figure 6.19: PCA novelty index between second and third natural frequency.
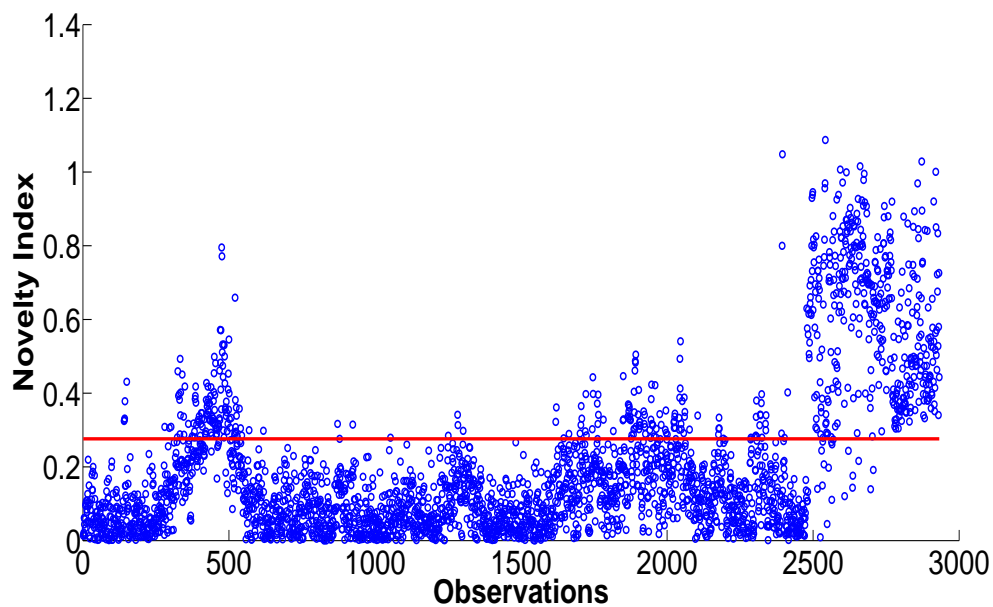
Figure 6.20: AANN novelty index between second and fourth natural frequency.
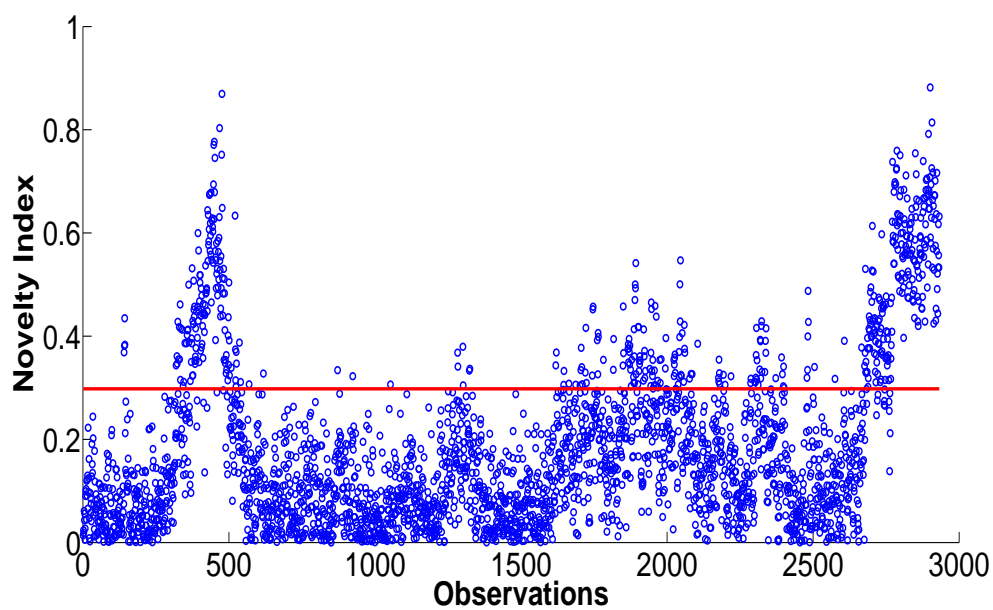
Figure 6.21: PCA novelty index between second and fourth natural frequency.
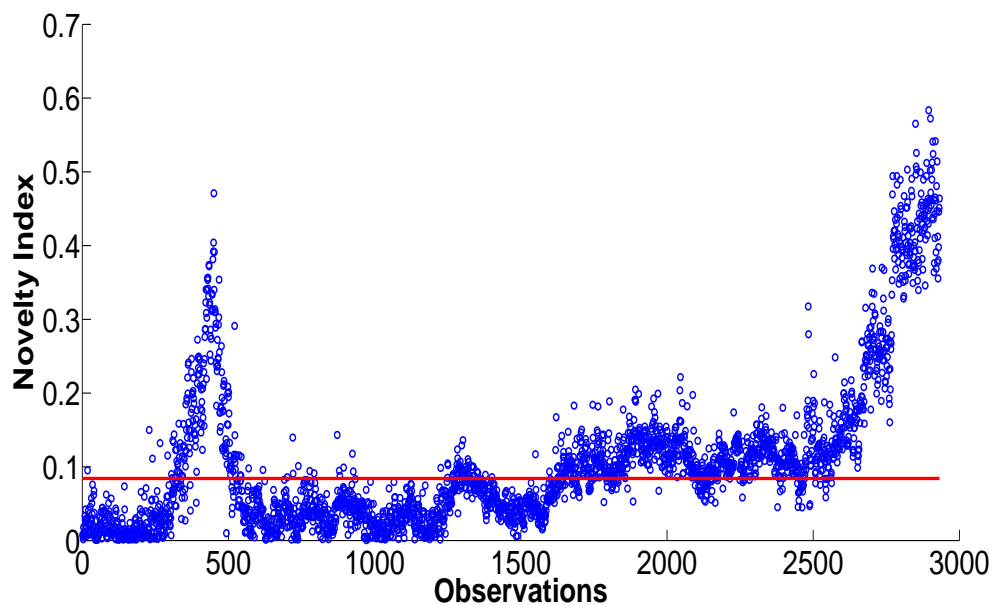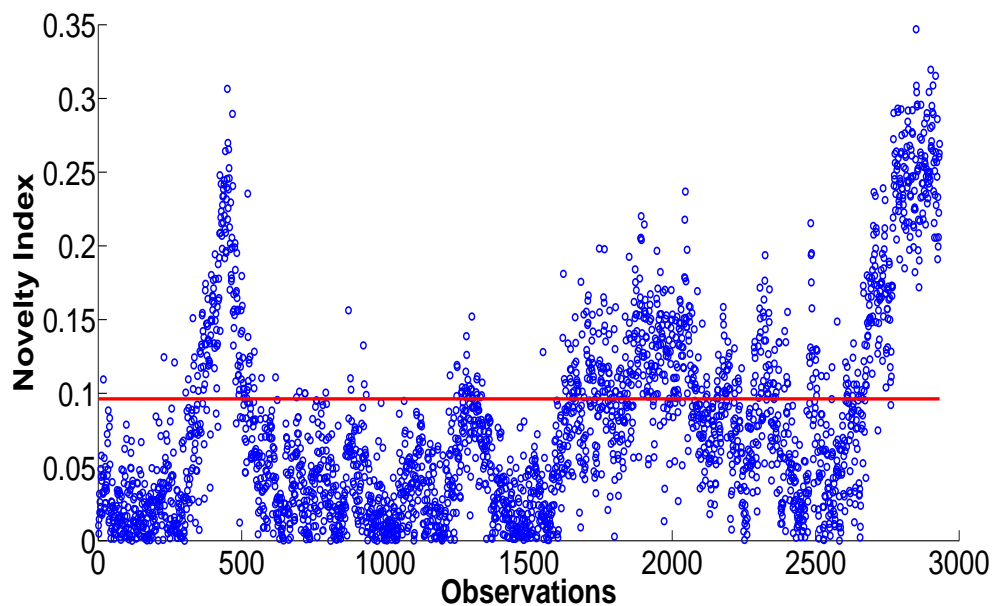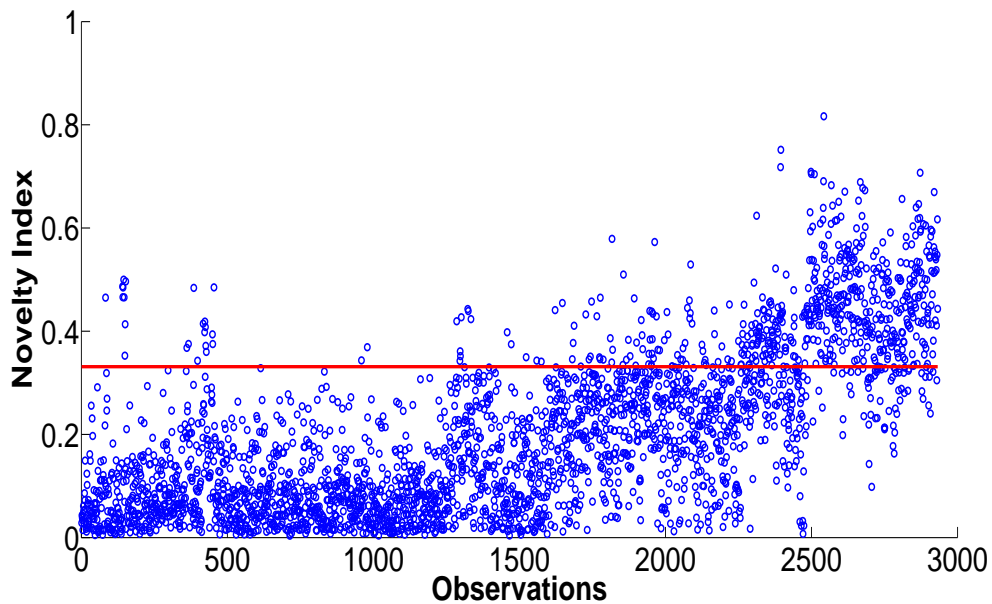
Figure 6.22: AANN novelty index between third and fourth natural frequency.

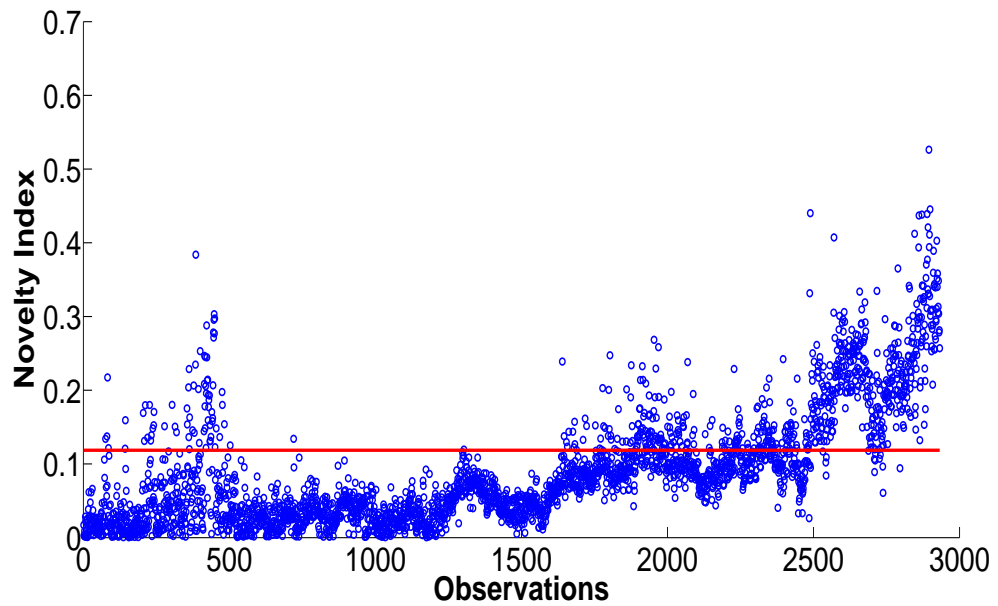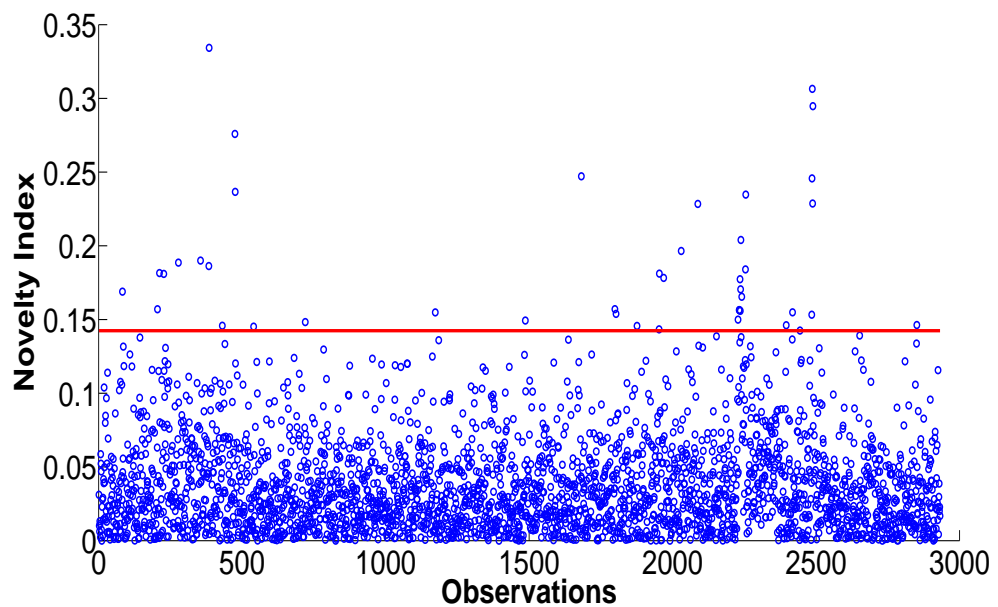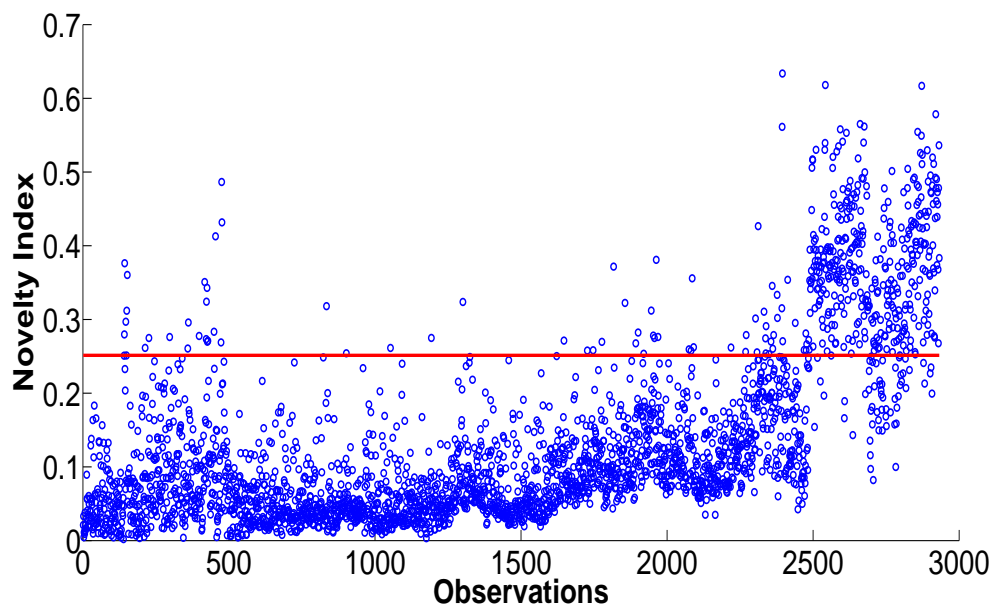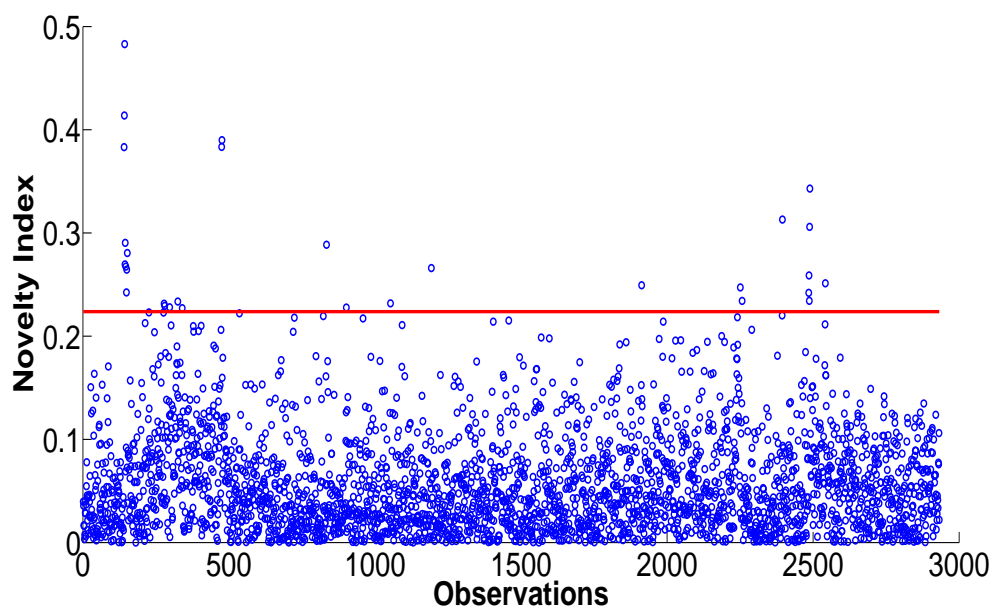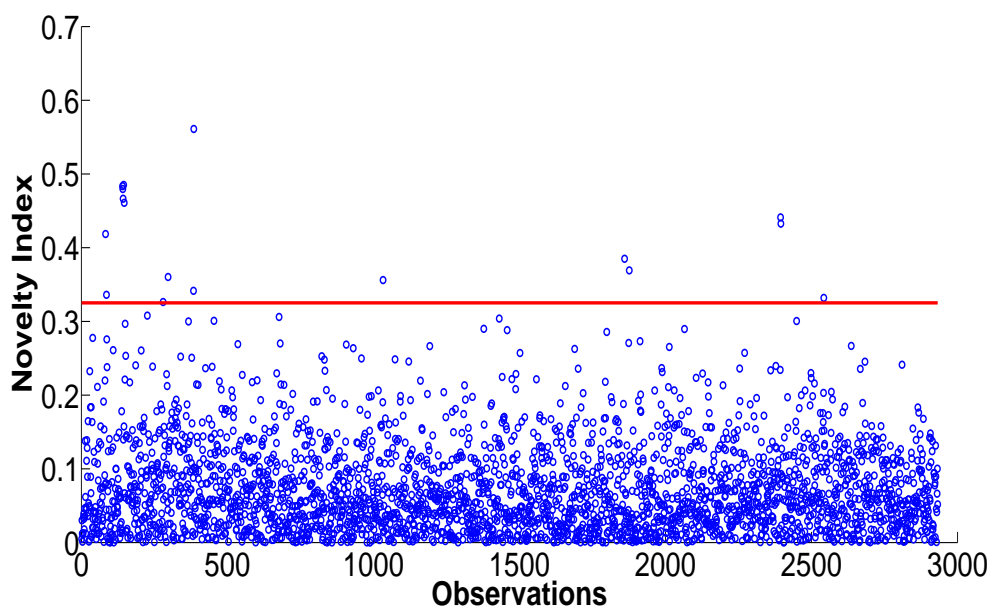Figure 6.23: PCA novelty index between third and fourth natural frequency.

As a conclusion it could be assumed that the change in temperature gradient and
the formed ice seems to introduce different types of nonlinearity along the whole
data of the second natural frequency but as well as to the first, third and fourth.
Although, AANN was able to learn in some extent the correlation pattern between
the third and fourth frequency.

6.3 Operational changes of the Piper Tomahawk aircraft wing experiment

The experimental structure presented is an aluminium aircraft wing [127], as shown
in Fig.6.24. The wing is mounted in a cantilevered fashion on a substantial, sand-
filled steel frame. Fifteen PCB piezoelectric accelerometers were mounted on the
upper (as mounted) surface of the wing using ceramic cement. The location of the
sensors, inspection panels and sub-surface stiffening elements (dotted lines) are shown
schematically in Fig.6.25. The sensors are denoted S1 to S15. Experimental data
acquisition was performed using a DIFA SCADAS III system controlled by LMS
software. All measurements were recorded within a frequency range of 0-2048 Hz
with a resolution of 0.5 Hz. The structure was excited with a band-limited white
Gaussian signal using a Gearing and Watson amplifier and shaker mounted beneath
the wing. Both the real and imaginary parts of the accelerance FRFs (see Fig.6.27)
were recorded at 15 response locations using single-axis accelerometers. Five-average
samples were recorded in all cases as this was found to offer a good compromise
between noise reduction and acquisition time. In order to introduce damage in a
repeatable and realistic manner, the inspection panels were modified on the underside
of the wing, with the wing being mounted upside-down to enable access. Five panels
were considered and all of them had the same dimensions and orientation (Fig.6.26).
The test sequence was arranged into two rounds of 5 blocks resulting in 10 blocks.
Each block contains 3 runs: a normal condition run, a damage (saw-cut) run and
a damage (panel-off) run. Each run contains 100 observations. Within each block,
only the panel of interest is removed or saw cut panel replaced, the remaining four
panels remain in place. In turn, at the end of the first run all 5 panels were removed
and the second run started by repeating the procedure sequence again, in Table.6.3.

The way the experimental procedure was performed offered the advantage of high
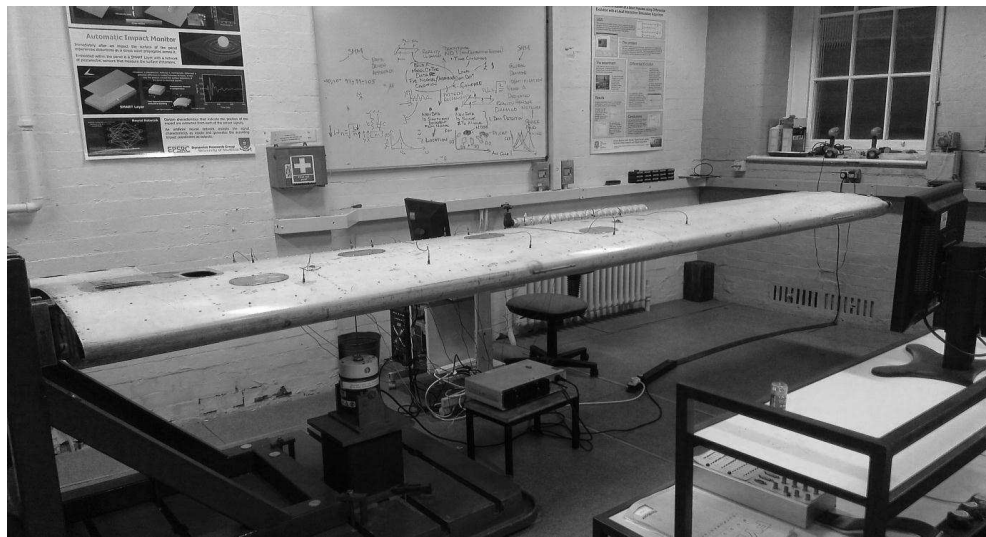
Figure 6.24: Piper Tomahawk aircraft wing.
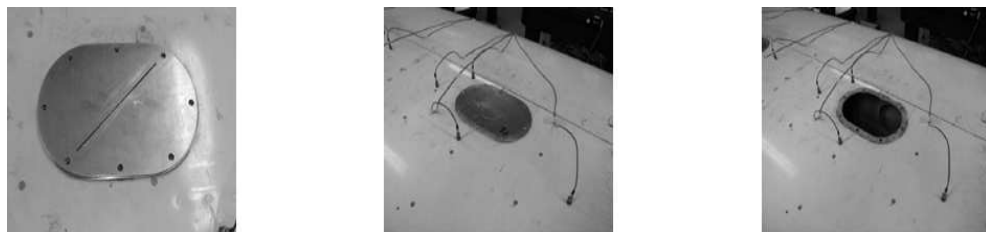
Figure 6.25: Inspection panel in normal condition (middle), removed
panel (right), saw cut (left).
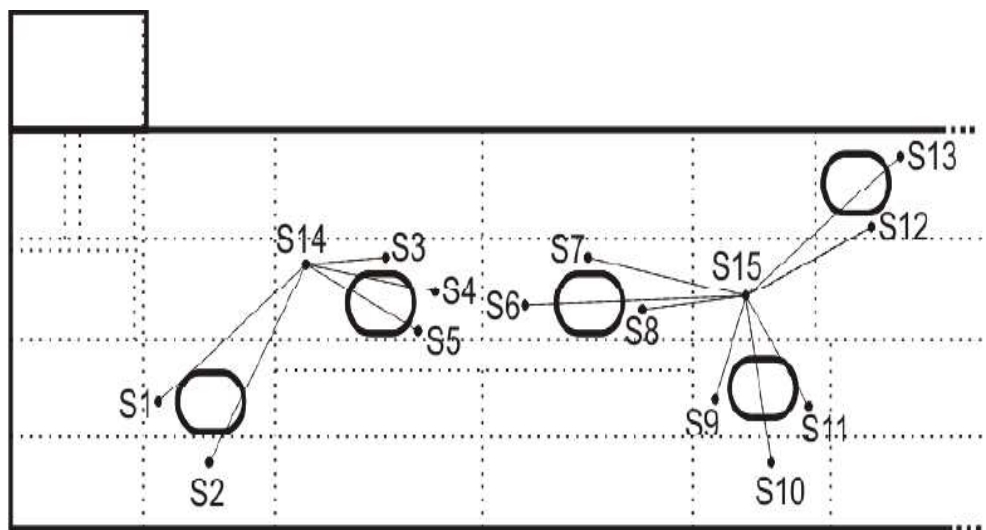
Figure 6.26: Schematic sensor placement diagram.

Round one	Normal	Panel removal	Panel saw cut	
Sensor 1	100	100	100	Observations
⋮				
Sensor 15	100	100	100	Observations
Round two				
Sensor 1	100	100	100	Observations
⋮				
Sensor 15	100	100	100	Observations

Table 6.3: Piper Tomahawk aircraft wing measurements sequence.

variability in normal condition FRF measurements due to the fact of the removal and reattachment of the panel. As the aim of this study is to uncover inclusive outliers without having to pre-set a normal training data reference the experimental sequence followed was ideal. By identifying and detecting variability at an early stage, the prospects of achieving good generalisation and establishing a correct normal condition training classifier may be increased. The inspection panels are attached using eight screws and the change of boundary conditions due to the removal and reattachment of the panel could introduce increased sensitivity to the FRF measurements.

The dimension of the data remains a major challenge as the novelty detection technique suffers from the "curse of dimensionality". As demonstrated in previous works [9, 65–68], one can introduce subsets of the data by introducing features which are sensitive to damage. In this section something analogous was followed as the resonance frequency between spectral points 780-820 was used as can be seen in Fig.6.28. It is obvious that several feature combinations can be tested, by combining the normal condition with each panel removal and each of the sensor measurements in order to check the sensitivity of damage detection regarding the position of the sensor.

The first step in applying the robust outlier analysis of the MVC and MVEE techniques is to construct a testing feature of the 2 rounds by creating a 1000×50 matrix for each of 15 sensors, Figs.6.29-6.32. For the results that are shown in Figs.6.33-6.36 the first 1000 samples include the natural, undamaged condition of the structure of the 10 blocks (10 blocks per diagram); the next 200 samples include the panel removal (100 samples for each different run) and the next 200 samples include the panel saw cut damage (100 samples for each different run). Before discussing the results, it has to be made clear that the purpose of this study is not an extended investigation of all sensor measurements in respect to panel location and as a result
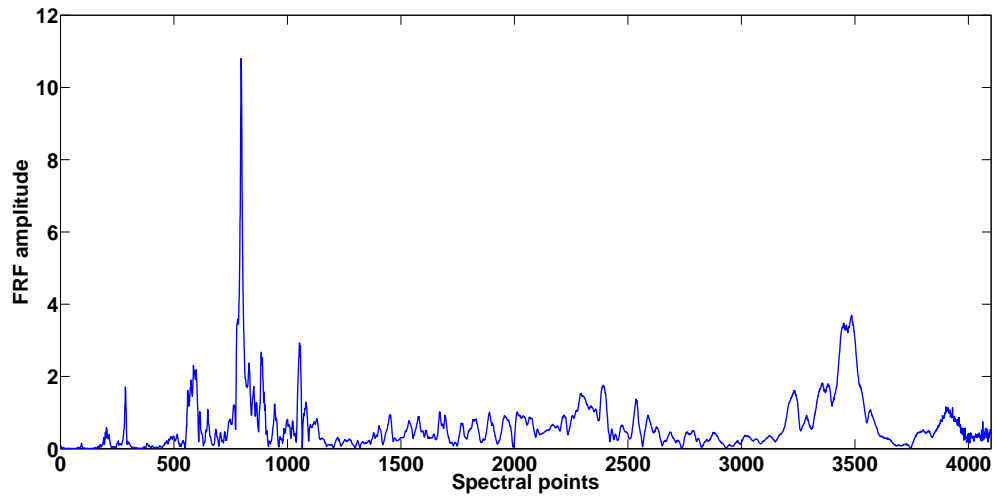
Figure 6.27: Typical FRF spectrum.
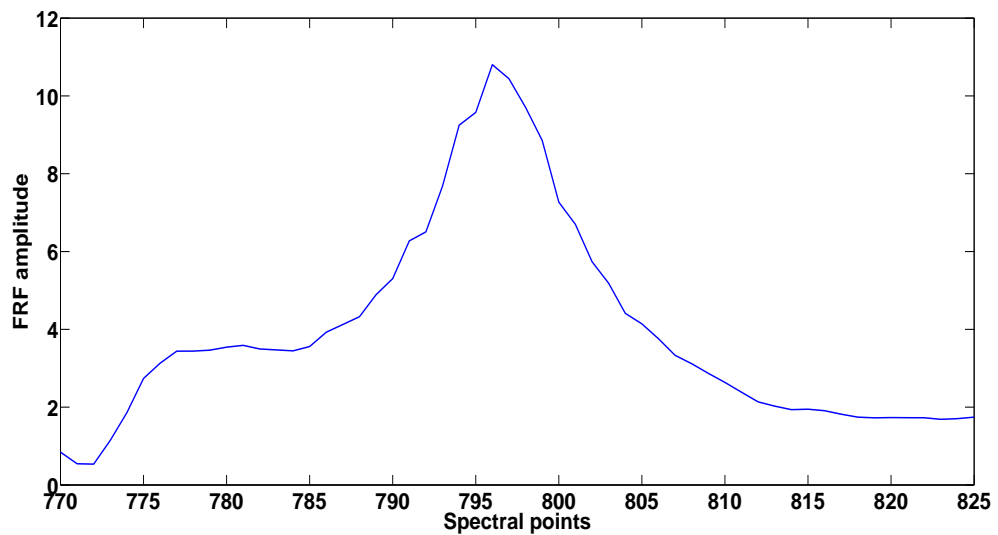
Figure 6.28: Resonant frequency selected as a feature with 50 points
around the pick.

their sensitivity in the multiple damage presence. This extended overview can be
found in reference [127]. The purpose is rather to test if the selected feature can
unveil if there is any variability due to screwing and unscrewing the panel after each
run in order to reattach the panel. This is the reason that not all combinations in
respect of sensors location, are presented.

The variation of normal condition is dominant in Figs.6.29-6.32. Once again it is
validated that the MSD discordance test is not suitable when inclusive outliers are
present. The MVE index seems to be much more sensitive than MCD index in this
experimental validation of the robust distances. The interesting conclusion derived
from the application of the minimum volume ellipsoid on this particular feature is
that almost all the normal condition (even the same sensor measurements in the two
different rounds) are above the threshold and not following a normal distribution.
It becomes clear how important are these results in order to establish a normal
condition free from external variations. If the target of doing the experiment was an
extensive application of classification or novelty detection analysis, it is clear that
due to the extended variation between the measurements the output of the applied
machine learning algorithms could be weak.
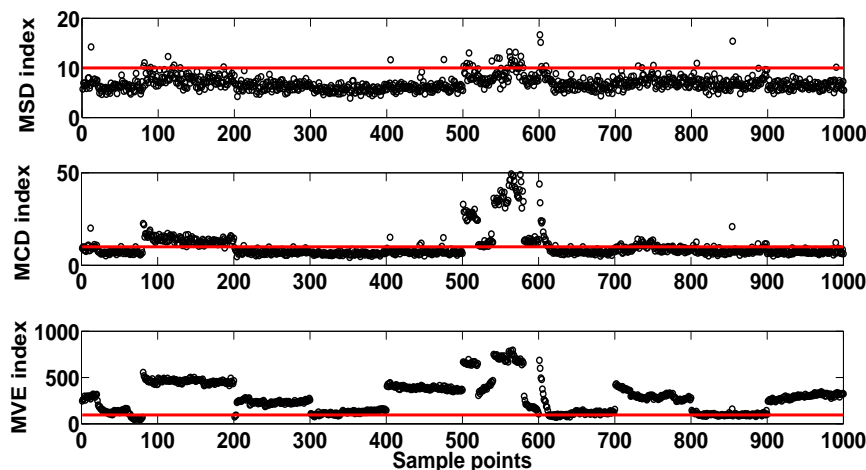
Figure 6.29: Novelty detection of normal condition of sensor 1.
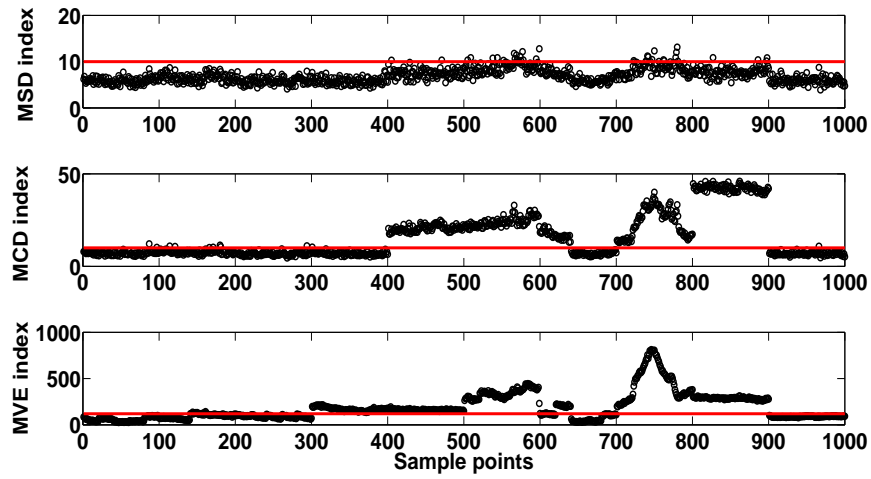
Figure 6.30: Novelty detection of normal condition of sensor 5.
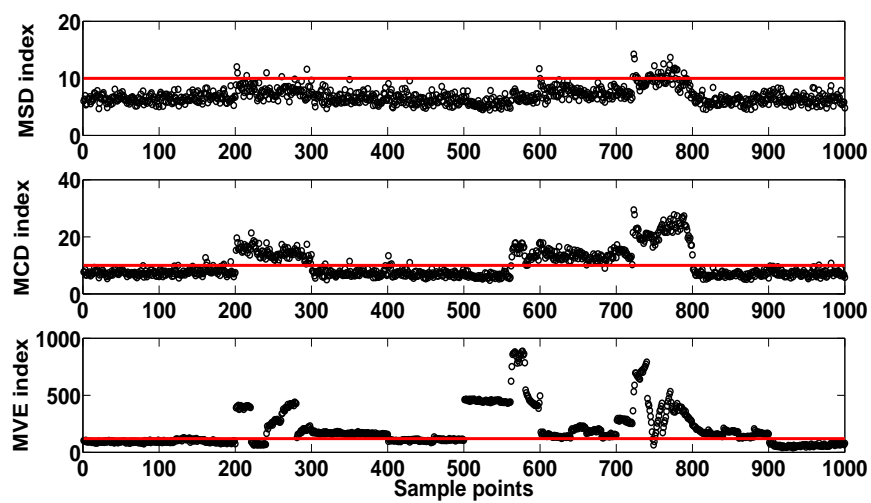
Figure 6.31: Novelty detection of normal condition of sensor 8.

The results that are shown in Figs.6.33-6.36, are an analysis including the damage
scenarios on the wing. MSD index classifies almost all the observations as normal
below the threshold. The MCD and MVE index reveal the induced variation with
MCD index in Fig.6.33 to be more sensitive compared to the MVE index giving
a strong appearance of damage after sample point 1000. On the other hand, the
MVE index is very sensitive to the fluctuations of normal condition. This makes the
combined novelty detection of both robust methods critical as they can complement
each other. The changes on normal condition are so dramatic that indicate higher
discordance values compared to the damaged scenarios. Last but not least is the
remarkable difference between the first and second round measurements. Operational
variations (like the change of boundaries conditions)can severally alter the raw signals
that are extracted.
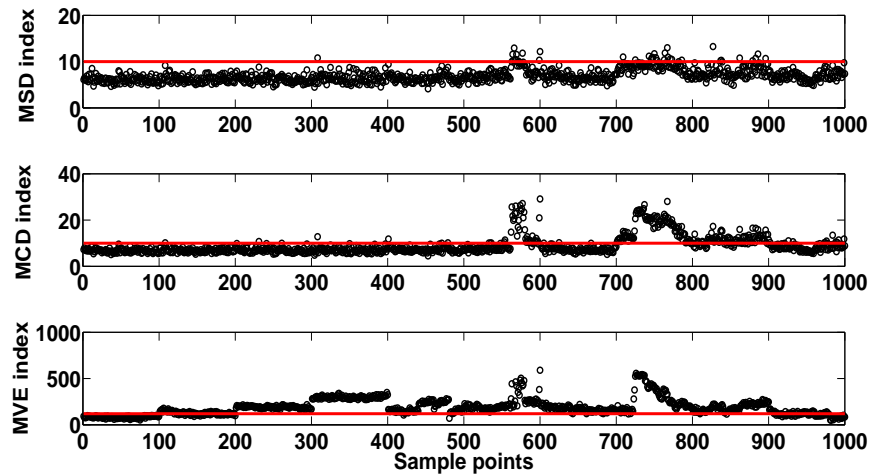
Figure 6.32: Novelty detection of normal condition of sensor 12.
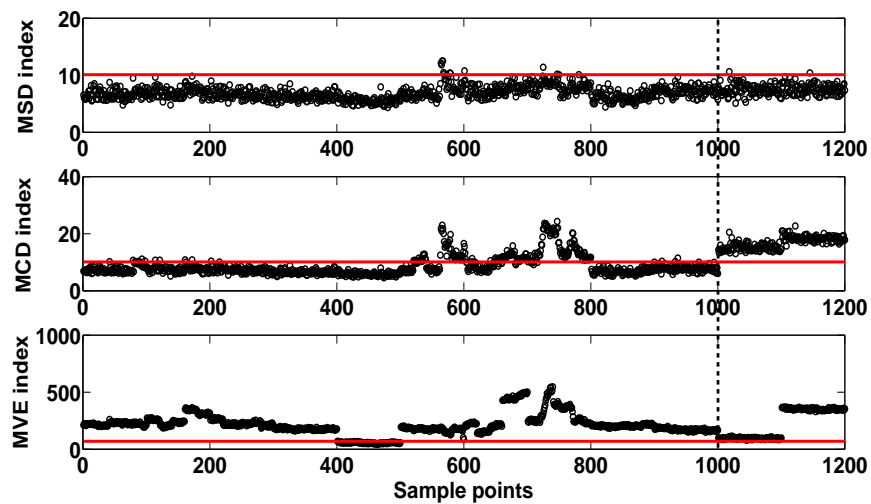
Figure 6.33: Novelty detection of sensor 14 in respect to panel 3, including
panel removal.
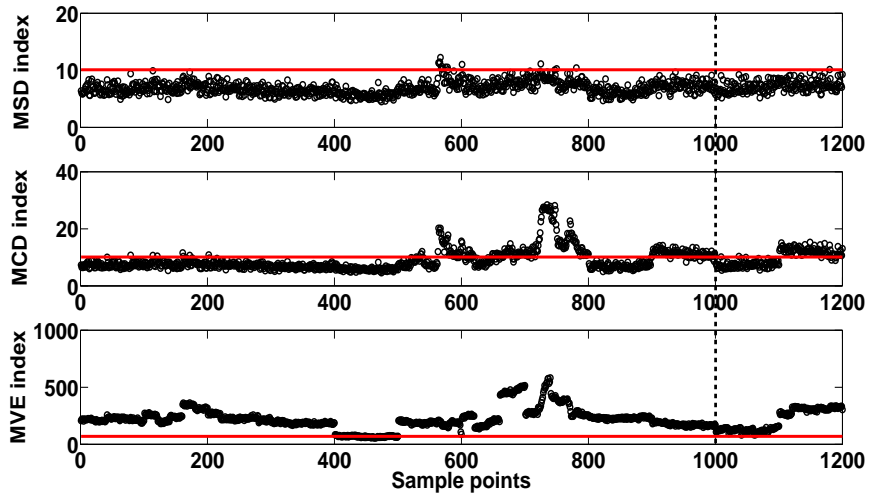
Figure 6.34: Novelty detection of sensor 14 in respect to panel 3, including
saw cut panel.
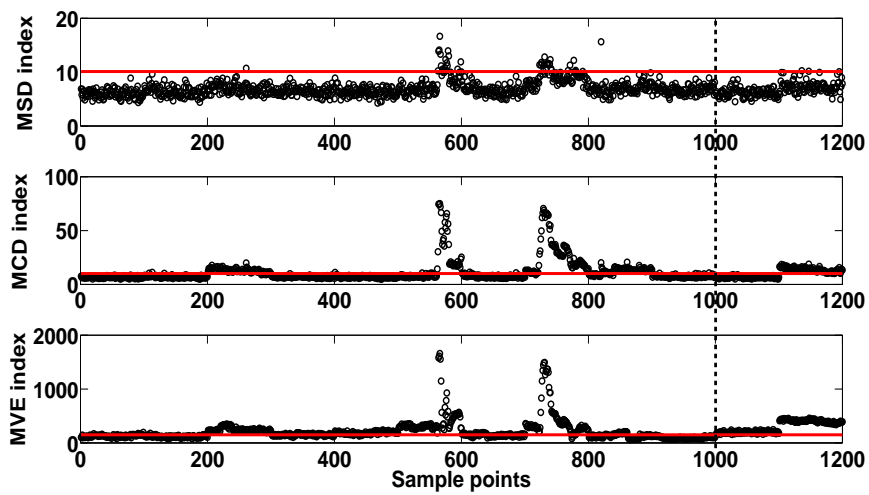
Figure 6.35: Novelty detection of sensor 15 in respect to panel 1, including
panel removal.
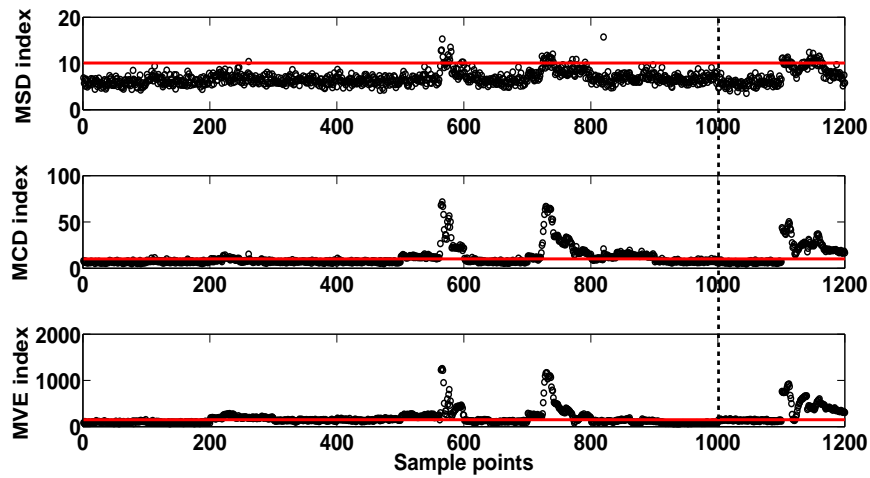
Figure 6.36: Novelty detection of sensor 15 in respect to panel 1, including saw cut panel.

6.4 Conclusion

In this chapter the robust outlier techniques were tested in the context of SHM and their importance in establishing a normal condition that is clear of outliers is shown. Through the two real life experimental applications to the Z24 bridge and to an aircraft wing, the critical importance of the different uses of robust multivariate statistics were demonstrated. With the usage of minimum covering ellipsoids a geometrical representation of "what is actually happening to the data" was presented by revealing the frequency correlations in the Z24 bridge. As a further step for validation purposes novelty detection algorithms such as linear PCA and NLPCA were tested in order to make clear the conclusions derived by geometrically looking to the data. By utilising the vibration responses of an aircraft wing the dramatic impact of changing boundary conditions on the normal condition identification was visualised. The most important observation of all is that the outlier analysis was performed by utilising inclusive parameters without setting a training set. This could prove an important tool in the context of SHM and CM data exploration.

It is of critical interest to investigate in detail factors such as variability, loading and environmental conditions, boundary conditions and feature selection that will all affect the performance of the classifier, especially when an unfaulted condition classifier has to be present for more sophisticated damage diagnosis. In this chapter, robust multivariate statistics were investigated focused mainly on a high level estimation of

the outliers which determines the presence or absence of novelty - something that is of fundamental interest.

Since different novelty detection techniques are based on disjoints sets of assumptions and different technical bases a direct comparison between them is not always fair and possible. In a lot of real life cases, the data structure and the outlier generating mechanisms dictate which method will outperform the others and reveal the internal variability. This the reason that both of the presented methods should be combined in order to ensure a viable result. A promising method regarding the latter could be proved to be least trimmed squares (LTS) regression [128].

Chapter 7

CONCLUSIONS AND FURTHER WORK

This thesis has focused on three main issues that are addressed through this work such as the auto-associative neural networks architectures, aspects of SHM for wind turbines and the introduction of robust multivariate statistics tools for the investigation of critical influences of multiple outliers due to changing environmental and operational conditions.

7.1 The truth behind AANN architectures

In Chapter Three, the structure of different architectures of auto-association are evaluated in order to investigate conclusions derived from classic literature. The observations that were derived involve critical practical aspects of efficient novelty detection of nonlinear patterns. An analysis is performed in order to demonstrate the ability of nonlinear auto-association of MLPs consisted of one nonlinear hidden layer and with linear and nonlinear nodes in the output layer. Auto-association with only one hidden layer is of great interest in terms of reduced network complexity.

It is established that linear PCA lacks performance for multimodal classification problems and novelty detection. It is a dominant conclusion throughout this chapter that single hidden layer auto-associators are not performing in a similar fashion to

PCA. Nonlinear auto-association of MLPs consisted of one nonlinear hidden layer
can not fully reconstruct highly nonlinear surfaces but in terms of novelty detection
they perform much better compared to linear PCA.

The analysis was validated via simulated and experimental data sets. The three-layer
MLPs and RBF network demonstrate remarkable results as they not only perform
much better than PCA but also, highlight similar results with the five-layer auto-
associator. Furthermore, regarding the experimental validation, PCA was not able
to detect the damage introduced by a 15 Joule impact to a composite plate and was
unable to distinguish the damage between the different levels of impact.

7.2 Machine learning algorithms for SHM purposes on WTs

In Chapter Four the findings and algorithms investigated in Chapter Three like,
machine learning tools based on Artificial Neural Networks (ANNs), including
an Auto-Associative Neural Network (AANN), auto-association with Radial Basis
Functions (RBFs) networks and a powerful technique of Gaussian processes for
regression are used.

Advanced signal processing for SHM purposes in the wind energy sector is still in a
preliminary stage compared to civil and aerospace engineering. The purpose of this
chapter is to introduce such pattern recognition methods into the wind energy field
and to investigate the effectiveness and performance of such methods.

Experimental measurements from vibration analysis were extracted from a 9m blade
throughout a full-scale continuous fatigue test. High frequency responses functions
(FRFs) combined with dimensionality reduction techniques and nonlinear novelty
detection were able to reveal a sensitive feature for detecting or even locating damage
in wind turbine blades. Furthermore, because the cost of multiple sensors remains a
fundamental financial drawback, the results point out that there is a perspective of
reducing the actual number of sensors required to cover the entire body of the blade.

The chapter concludes with a preliminary analysis of a model regression between
power output and blade loading of a simulated offshore wind farm that can be used
for early detection of performance issues. The results have shown that an automatic

interpretation of SCADA data is feasible. The proposed approach can be extended by combining many different measurements of SCADA data by developing a population-based architecture for an online automated damage detection system for the wind farm in total with a low financial cost.

7.3 The introduction of robust outliers algorithms

Chapter Five introduces robust multivariate statistical methods to Structural Health Monitoring (SHM). The algorithms that were described are the Minimum Covariance Determinant Estimator (MCD) and the Minimum Volume Enclosing Ellipsoid (MVEE). The main novel element of this work is the high level estimation of the "masking effect" of inclusive outliers in order to identify and detect variability at an early stage and establish a normal condition clear from internal anomalies. It is very important to be mentioned that these robust outlier methods are inclusive and in turn there is no need to pre-determine an undamaged condition data set, offering a vital advantage over other multivariate methodologies.

Chapter Six builds on the findings of Chapter Five through the two real life experimental applications to the Z24 Bridge and to an aircraft wing. Both of the robust methods outperform the classical approach in multivariate discordancy tests regarding the inclusive outliers. MSD calculation can not detect any of the inclusive outliers. Furthermore, with the usage of minimum covering ellipsoids, a geometrical representation of the data variables correlation was presented by revealing linear on nonlinear connections. At the end, a combination with the methods of Chapter Three (linear PCA and NLPCA) is applied as an extra step for validation purposes.

Again, the significant observation of all the analysis is that the robust outlier tools implemented by utilising inclusive parameters without setting a training set (normal condition). This could prove an important tool in the context of SHM and CM data exploration before applying pattern recognition techniques.

7.4 Further work

This is one of the few works to use different neural networks architectures for novelty detection. But there are still hard questions that have to be investigated regarding

auto-association. Is generalisation needed for novelty detection when neural networks are used or not? If the objective is to learn precisely the normal condition even when the network is left at the "mercy" of overfitting by learning the raw data, it should become much more sensitive to outlier detection in the test set. On the other hand, one of the great challenges of novelty detection methods is their ability to detect damage independently of the operational and environmental fluctuations that may alter the natural dynamic characteristics and indicate wrongly a fault signal. In this context the need for generalisation should be essential for the network to learn underlying parameters inside the normal condition. Furthermore, there is no accurate measure of evaluating the exact number of nodes in the mapping and demapping layers and in some percent the practical approach of running the network for different initial conditions and different number of nodes is computationally very intensive.

The threshold calculation is very important regarding auto-associators. In this study the assumption that the normal condition is following Gaussian statistics was assumed but this idea has the critic drawback that the auto-associators do not suffer from the specific distribution shape and can learn any kind of non-Gaussian distribution (undamaged condition might not be Gaussian).

A field that has gained limited (or no) attention in SHM field and nonlinear novelty detection is deep belief networks (DBN) which are dense architectures with several layers composed of stochastic, latent variables [129–133]. These networks are graphical models which learn to extract a deep hierarchical representation of the training data with incredible accuracy. A DBN is a directed acyclic graph consisted of stochastic variables as mentioned before. The big advantage is that one can solve the inference problem (infer the states of the unobserved variables) and the learning problem (tune and find the correlation between variables in order to make the network more likely to generate the observed data) [129–133]. These networks require massive amounts of observations in order to be implemented (although the usage of Gaussian processes could be proved beneficial) and they could be highly beneficial for data coming from a family of wind turbine blades of an offshore farm (SCADA observations), offering a holistic approach. Also, data that is derived from high sampling frequencies like acoustic emission and ultrasounds could be used for such systems.

In the wind energy sector, SHM technology has slowly developed and in reality there is no commercial application of combined advanced signal processing and complete

sensor network. One of the reasons is the limited access to data from the wind industry. Without data, the market will not accept the vital importance of structural health monitoring and the significant financial advantages.

A second drawback is the lack of integrated sensor networks to offshore wind farms, the powering of such systems as well as the data transmission. As mentioned in the introduction the cost of faulty generators and blades could be catastrophic and a future attention of cost benefit analysis between integrated sensor technologies and early damage detection is critical. Wireless systems, fibre optics, energy harvesting and compressed data sensing are promising tools.

At last, it is of critical interest to investigate in detail factors such as variability, structure of the real blade, loading and environmental conditions, boundary conditions and feature selection that will all affect the performance of the classifier. All these factors are crucial as the general trend leads wind energy to offshore investments where the environmental and loading conditions are under a high variability as they often operate in harsh environments.

Regarding the robust novelty detection techniques, since they are based on disjoint sets of assumptions and different technical bases as mentioned in Chapter Six, a direct comparison between them is not always fair and possible. In a lot of real life cases, the data structure and the outlier generating mechanisms dictate which method will outperform the others and reveal the internal variability. This is the reason that both of the presented methods should be combined in order to ensure a viable result. A promising method regarding the latter could be proved to be least trimmed squares (LTS) regression [128].

The main remaining problem with these outliers techniques is that the majority in this work have used a version of the control chart in order to visualise a threshold. Extreme value statistics were implemented in order to provide a threshold that may be less susceptible to false positive outliers. This concept requires further investigation and attention. Different techniques of calculating a robust threshold should be used. The idea of an adaptive threshold calculation is very interesting.

7.5 General Challenges in SHM

SHM is a challenging research field and this difficulty is reflected in the minor progression of the technology into the market world. There are several reasons that SHM has to overcome in order to become a part of the real world and escape from the lab environment.

Industrial operators have yet to be convinced that structural health monitoring could be an integral technology for the new generation large structures that can offer comparative advantage against competition. Here comes the obvious answer, that SHM is not present because it is not ready. This is partly true as the industry itself is not trusting the SHM community with the necessary data and equipment to prove that SHM technology is mature enough in order to operate at large scale. It seems that in civil and aerospace engineering, the SHM tools are entering the game with better prospectives. Although, the energy sector and especially wind and wave energy power plants and as well as nuclear power sector will need the addition of a reliable SHM system. There will be always operators that believe that such technologies are unnecessary and in order to monitor a bridge the visual inspection or the hearing of the sound is enough.

History shows that SHM systems can be expensive, unreliable or very complicated for an industrial manager to understand. In turn this could lead to a critical misunderstanding regarding SHM and introduce a barrier to even put sensor systems for testing on real structures under operation.

On a realistic approach, the sensors used for SHM are introducing challenges themselves and especially when large numbers of data have to be collected and processed in a continuous and on-line level. In practical terms, the implementation of sensors on a blade of an offshore wind farm under harsh environmental conditions is a considerable challenge. The challenges one faces are for example sensor failures, power sources, data collection and transferral etc.

As carefully mentioned, damage detection and identification is a procedure that is hierarchical in nature. At its most sophisticated, diagnosis of the damage could include localisation, classification and severity assessment and even go so far as to estimate the time-to-failure of the structure. And all of these are becoming more difficult if one has to consider factors such as variability, loading and environmental

conditions (Z24 bridge in Chapter Six), boundary conditions (aircraft wing in Chapter Six) and feature selection that will all affect the performance of the classifier. More particularly validation for SHM systems is very difficult as no company will damage and test several submarines or bridges in order to validate the technology. As seen through this thesis a system (signal processing methods) could fail to detect critical outliers (false negative detection) or can be very successful on detecting abnormalities that have no connection with damage alarms (false positive). As seen, detecting as faults the change of temperature would have severe financial impact and could negatively colour the reasons for adopting an SHM.

This is the greatest challenge that the field of SHM has to face and solve. The optimist says the glass is half full. The pessimist says the glass is half empty. The market says the glass is twice as big as it needs to be and the realist says the glass contains half the required amount of liquid for it to overflow. The truth lies somewhere in the middle and the research has to address the issue of validation otherwise SHM technology will remain inside lab environments and scenarios without being able to replace visual inspections of safety critical components.

<div align="right">

Appendix A

</div>

Frequency Response Functions

Almost all the analysis in this thesis is based on vibration-based SHM. In this appendix a short description behind the extraction of frequency response functions (FRFs) is given. For a more detailed and holistic description of modal analysis and vibration responses the reader can look at any dynamics engineering book (for example [46, 134, 135]).

A.1 Laplace Domain and FRF

The dynamic properties of a mechanical system can be described by its mass, stiffness and damping. They are properties that characterise inertia, dissipative and elastic forces. This is the reason that modelling a real system where all features interact, is a very complex procedure. In this section it is essential to give the basic concept of Laplace transform before preceding to to define FRFs. For simplicity a single-degree-of-freedom system is given.

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \tag{A.1}$$

where m is the mass, k is stiffness, c is the damping coefficient, $f(t)$ is the time dependent excitation force which is applied to the system and $x(t)$ is the corresponding displacement response. The Laplace method can be used in order to derive the dynamic response characteristics of a mechanical system under any type of excitation. The main advantage of Laplace transform is that it replace the differential equations with algebraic ones making easier to work with. The Laplace domain is given by the equation:

$$X(s) = \mathcal{L}[x(t)] = \int_0^{+\infty} e^{-st} x(t) dt \tag{A.2}$$

where s is Laplace variable and is a complex quantity and $x(t)$ is a function. If one applies the transform to (A.1) can obtain:

$$\mathcal{L}[m\ddot{x}(t) + c\dot{x}(t) + kx(t)]$$

$$= m[s^2 X(s) - sx(0) - \dot{x}(0)] + c[sX(s) - x(0)] + kX(s) \tag{A.3}$$

$$= (ms^2 + cs + k)X(s) - msx(0) - m\dot{x}(0) - cx(0)$$

and $\mathcal{L}[f(t)] = F(s)$ in turn one gets:

$$(ms^2 + cs + k)X(s) = F(s) + m\dot{x}(0) + (ms + c)x(0) \tag{A.4}$$

where $x(0)$ and $\dot{x}(0)$ are the initial conditions. If the initial conditions are zero one can obtain the ratio:

$$H(s) = \frac{X(s)}{F(s)}$$

$$\tag{A.5}$$

$$H(s) = \frac{1}{ms^2 + cs + k}$$

$H(s)$ is known as the system transfer function and is a complex function lying in the Laplace domain. The denominator is the characteristic equation and if one assumes two roots of the characteristic equation s_1, s_2 then the transfer function can be written as:

$$H(s) = \frac{1}{m(s - s_1)(s - s_2)} \tag{A.6}$$

The Laplace domain characterises the system is terms of poles and residues. The transfer function can be described in the frequency domain and in turn, it can give

the Frequency Response Function (FRF):

$$H(w) = \frac{1}{k - w^2 m + iwc} \tag{A.7}$$

The FRF may be presented in terms of displacement, velocity or acceleration as can be seen in the following equations:

$$a(w) = \frac{displacement\ response}{force\ excitation}$$

$$b(w) = \frac{velocity\ response}{force\ excitation} \tag{A.8}$$

$$c(w) = \frac{acceleration\ response}{force\ excitation}$$

There are a lot of advanced signal processing tools in order to estimate experimentally the FRFs of a mechanical system and details can be found for example in [38, 46, 134–140].

Expectation-Maximisation algorithm

The expectation-maximization (EM) algorithm was introduced by Dempster et al.[141] in 1977 and is a general method in order to solve maximum likelihood estimation problems. This algorithm is simple and fast as it avoids calculations and storage of derivatives and it can be used to many other probabilistic methods such as probabilistic principal component analysis, radial basis function networks, Gaussian mixture models or could be extended to incorporate in the probabilistic graphical models [142]. The general idea behind the EM algorithm can also be found in Hidden Markov models [143]. A brief description of the basic steps of the algorithm are given, following the thoughts and the description of [69–71] where analytic details can be found.

B.1 Algorithm theory

In order to start describing the EM algorithm the framework of the problem from which the EM algorithm is useful is given [144]. Let $\{x\}$ be a random data set vector with probability density function $p(\{x\}|\{\theta\})$, where $\{\theta\}$ is an unknown parameter

vector. The basic principle is based on maximizing the data likelihood function or minimising the negative log likelihood of the data vector, Error function $= -\mathcal{L} =$ $-\sum_{n=1}^{N} \log p(\{x\}^n)$. Except in situations where the purpose is estimating the arithmetic mean and variance of a Gaussian population (Gaussian mixture models), the idea of finding the maximum data likelihood has no closed-form solutions. The global minimum of $-\mathcal{L}$ is approaching $-\infty$ and most of the times several numerical routines are needed to check the variance at each iteration. Also, when a big number of local minima are present then density function estimation is approximated by poor models.

In this respect, the EM algorithm is a method designed to efficiently solve the minimisation of the negative log likelihood. EM performs iterations in a such manner that maximises successive local approximations of the likelihood function. In turn, each iteration consists of two steps: one that computes the approximation (the E-step) and the one that maximizes it (the M-step). The critical advantage of EM is that it does not rely on calculation and storage of derivatives.

The negative of log likelihood can be regarded as an error function and can be re-written as:

$$E = -\mathcal{L} = -\sum_{n=1}^{N} \log p(\{x\}^n) = -\sum_{n=1}^{N} \log \left\{ \sum_{j=1}^{N} \ln \{p(x^n \,|\, j)\} \, p(j) \right\} \qquad (B.1)$$

where $j = 1, ..., M$

As mentioned above minimising this error function is not a trivial task. In this analysis one assumes that the training data set is coming from an unknown mixture model because if the component of which each point of data $\{x\}$ sampled from was known then it would be easy to calculate the model parameters. This assumption of assuming a hypothetical complete data is the reason the algorithm can deal with missing data points. In turn, if for each data point $\{x\}^n$ there is a random variable z^n which is an integer of $1, ..., M$ then $\{y\}^n$ is the complete data $(\{x\}^n, z^n)$ and $\{w\}$ is the corresponding parameters vector of the model. The EM iterations will produce a series of $\{w\}^{(m)}$ starting from an initial vector values of $\{w\}^{(0)}$.

The likelihood of a complete data point for $\{z\} = \{j\}$ is given be the equation:

$$p(({\{x\}, \{j\}})|\{w\}) = p(\{x\}|\{j\}, \{w\})P(\{j\}|\{w\}) = p(\{x\}|\{\theta\}_j)P(\{j\}|\{w\}) \quad \text{(B.2)}$$

where $\{\theta\}_j$ is the density function parameters vector which is essentially mean and variance for each individual component of $\{j\}$. To obtain the likelihood of $\{x\}$ is straightforward as it is a discrete variable by performing a summation on (B.2),

$$p(\{x\}|\{w\}) = \sum_{j=1}^{M} P(\{j\}|\{w\})p(\{x\}|\{\theta\}_j) \quad \text{(B.3)}$$

If the current vector of parameters is given $\{w\}^{(m)}$ and based on the equations that give the prior $P(j)$, the mean $\{\mu\}_j$ and the variance $\{\sigma\}_j^2$,

$$P(j) = \frac{|I_j|}{N} \quad \text{(B.4)}$$

$$\{\mu\}_j = \frac{1}{|I_j|} \sum_{i \in I_j} \{x\}^i \quad \text{(B.5)}$$

$$\{\sigma\}_j^2 = \frac{1}{d\,|I_j|} \sum_{i \in I_j} \left\| \{x\}^i - \{\mu\}_j \right\|^2 \quad \text{(B.6)}$$

one wants to calculate the next set of parameters $\{w\}^{(m+1)}$ then although the class labels are not known the knowledge of the used probability distribution can use the expected values of the class labels in respect of the current parameters.

$$
\begin{aligned}
Q(\{w\}|\{w\}^{(m)}) &= E(\log p(\{y\}|\{w\}))p(\{z\}^n|\{x\}^n, \{w\}^{(m)}) \\
&= \sum_{j=1}^{M} \sum_{n=1}^{N} (\log(p(\{x\}^n, \{z\}^n|\{w\})))P(\{j\}|\{x\}^n, \{w\}^{(m)}) \\
&= \sum_{j=1}^{M} \sum_{n=1}^{N} (\log P(j) + \log p(\{x\}^n|\{\theta\}_j))P^{(m)}(\{j\}|\{x\}^n)
\end{aligned} \quad \text{(B.7)}
$$

where $P^{(m)}(\{j\}|\{x\}^n)$ is the expected posterior distribution.

As a result the E-step is the computation of the Q function and the M-step is the optimisation of $Q(\{w\}|\{w\}^{(m)})$ for the new parameters $\{w\}^{(m+1)}$.

EM algorithm forces the likelihood to converge monotonically and the algorithm will run until it convergences. This is something that can be proven and details can be found in [70, 71]. For Gaussian mixture models this optimisation can be expressed analytically but for more complicated structures the M-step is identified by checking that each time, the Q value increases.

Gaussian Process Regression algorithm

Rasmussen and Williams [97] define a Gaussian process (GP) as "a collection of random variables, any finite number of which have a joint Gaussian distribution". In recent years GPs are gaining a lot of attention as an alternative machine learning approach to MLPs in the area of regression (or classification) analysis as they offer fast and simple computations. Gaussian process regression is a robust tool which takes into account all possible functions that fit to the training data set and gives a predictive distribution of a single prediction for a given input vector. As result, a mean prediction and confidence intervals on this prediction can be calculated from this predictive distribution. The basic details of the algorithm are presented following the steps in [97]. The algorithm that was used in Chapter 4 is also coming from Rasmussen and Williams [97].

C.1   Algorithm theory

The initial and basic step in order to apply Gaussian process regression is to obtain a mean $m(\{x\})$ and covariance function $k(\{x\}, \{x'\})$ as GPs are completely specified

by them, $\{x\}$ represents the input vector. So for any real process $f(\{x\})$ one can define:

$$m(\{x\}) = E[f(\{x\})] \tag{C.1}$$

$$k(\{x\}, \{x'\}) = E[(f(\{x\}) - m(\{x\}))(f(\{x'\} - m(\{x'\})] \tag{C.2}$$

where E represents the expectation. Often, for practical reasons because of notation purposes (simplicity) and little knowledge about the data at the initial stage the prior mean function is set to zero. The Gaussian processes can be defined as:

$$f(\{x\}) \sim GP(m(\{x\}), k(\{x\}, \{x'\})) \tag{C.3}$$

As was assumed a zero-mean function, the covariance function can be described as:

$$\text{cov}(f(\{x\}_p), f(\{x\}_q)) = k(\{x\}_p, \{x\}_q)$$
$$= \exp\left(-\frac{1}{2}\Big|\{x\}_p - \{x\}_q\Big|^2 - \frac{1}{2}\Big|\{x\}_p - \{x\}_q\Big|^2\right) \tag{C.4}$$

This is the squared-exponential covariance function (although not the only option). It is very important to mention an advantage of the previous equation as the covariance is written as a function only of the inputs. For the squared-exponential covariance can be noted that it obtains unity values between variables that their inputs are very close and starts to decrease as the variables distance in the input space increases.

Assuming now that one has a set of training outputs $\{f\}$ and a set of test outputs $\{f\}_*$ one has the prior:

$$\begin{bmatrix} \{f\} \\ \{f\}_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \tag{C.5}$$

where the capital letters represent matrices. As can be seen, the covariance matrix must be symmetrical about the main diagonal and this is the reason that one of the few possible choices available is the squared exponential mentioned before.

As the prior has been generated by the mean and covariance functions, in order to specify the posterior distribution over the functions, one needs to limit the prior distribution in a such a way that includes only these functions that agree with actual data points. An obvious way to do that is by generating functions from the prior and select only the ones that agree with the actual points. Of course, this is not a realistic way of doing it as it would consume a lot of computational power. In a probabilistic manner this can be done easily via conditioning the joint prior on the observations and this will give (for more details see [70, 71, 97]):

$$\{f\}_*|[X]_*, [X], \{f\}$$

$$\sim N \left(\begin{array}{c} K([X_*], [X])K([X], [X])^{-1}\{f\}, K([X_*], [X_*]) \\ -K([X_*], [X])K([X], [X])^{-1}K([X], [X_*]) \end{array} \right) \tag{C.6}$$

Function values $\{f\}_*$ can be generated by sampling from the joint posterior distribution and the same time evaluating the mean and covariance matrices from (C.6).

The covariance functions used in this study are usually accompanied by some extra parameters in order to obtain a better control over the types of functions that are considered for the inference. As an example, the squared-exponential covariance function can take the form (1-dimensional):

$$k_y(k_p, k_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(k_p - k_q)^2\right) + \sigma_n^2 \delta_{pq} \tag{C.7}$$

where k_y is the covariance for the noise target set y. The length-scale l (determines how far one needs to move in input space for the function values to become uncorrelated), the variance σ_f^2 of the signal and the noise variance σ_n^2 are free parameters that can be varied. These free parameters are called hyperparameters.

The tool that has to be applied for selecting the model for choosing the optimal hyperparameters for GP regression, is the maximum marginal likelihood of the predictions $\mathrm{p}(\{y\}|[X], \{\theta\})$ with respect to the hyperparameters θ:

$$\mathrm{logp}(\{y\}|[X], \{\theta\}) = -\frac{1}{2}\{y\}^T[K]_y^{-1}\{y\} - \frac{1}{2}log|[K_y]| - \frac{n}{2}\log 2\pi \tag{C.8}$$

where $[K_y] = [K_f] + \sigma_n^2 I$ is the covariance matrix of the noise test set $\{y\}$ and $[K_f]$

is the noise noise-free covariance matrix. In order to optimise these hyperparameters through maximising the marginal log likelihood the partial derivatives give the solution:

$$\frac{\partial}{\partial \theta_j} \log p(\{y\}|[X], \{\theta\})$$

$$= \frac{1}{2}\{y\}^T [K]^{-1} \frac{\partial [K]}{\partial \theta_j} [K]^{-1}\{y\} - \frac{1}{2}tr\left([K]^{-1}\frac{\partial [K]}{\partial \theta_j}\right) \tag{C.9}$$

$$= \frac{1}{2}tr\left((\alpha\alpha^T - [K]^{-1})\frac{\partial [K]}{\partial \theta_j}\right)$$

where $\alpha = [K]^{-1}\{y\}$. Of course this solution is not a trivial procedure and for specific details readers are refereed to [97].

BIBLIOGRAPHY

[1] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[2] Geoffrey E Hinton. Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the cognitive science society*, pages 1–12. Amherst, MA, 1986.

[3] David E Rumelhart, Geoffrey E Hintont, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[4] Dana H Ballard. Modular learning in neural networks. In *AAAI*, pages 279–284, 1987.

[5] Hervé Bourlard and Yves Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4):291–294, 1988.

[6] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural networks*, 2 (1):53–58, 1989.

[7] Mark A Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243, 1991.

[8] Matthias Scholz and Ricardo Vigário. Nonlinear pca: a new hierarchical approach. In *Proceedings ESANN*, pages 439–444, 2002.

[9] Keith Worden. Structural fault detection using a novelty measure. *Journal of Sound and vibration*, 201(1):85–101, 1997.

[10] Hoon Sohn, Keith Worden, and Charles R Farrar. Novelty detection under changing environmental conditions. In *SPIE's 8th Annual International Symposium on Smart Structures and Materials*, pages 108–118. International Society for Optics and Photonics, 2001.

[11] HF Zhou, YQ Ni, and JM Ko. Structural damage alarming using auto-associative neural network technique: exploration of environment-tolerant capacity and setup of alarming threshold. *Mechanical Systems and Signal Processing*, 25(5):1508–1526, 2011.

[12] Garrison W Cottrell and Paul Munro. Principal components analysis of images via back propagation. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, volume 1001, pages 1070–1077. Cambridge, MA, 1988.

[13] Nathalie Japkowicz, Stephen Jose Hanson, and Mark A Gluck. Nonlinear autoassociation is not equivalent to pca. *Neural Computation*, 12(3):531–545, 2000.

[14] EWEA. Wind in power: 2012 european statistics. February 2013.

[15] EWEA. The european offshore wind industry - key trends and statistics 1st half 2013. July 2013.

[16] GWEC. Global wind energy: Solid growth in 2012. February 2013.

[17] Encyclopaedia Britannica. http://www.britannica.com/ebchecked/media/125134/components-of-a-wind-turbine.

[18] Berthold Hahn, Michael Durstewitz, and Kurt Rohrig. Reliability of wind turbines. In *Wind Energy*, pages 329–332. Springer, 2007.

[19] E Echavarria, B Hahn, GJW Van Bussel, and T Tomiyama. Reliability of wind turbine technology through time. *Journal of solar energy engineering*, 130(3): 31005, 2008.

[20] http://www.altitec.co.uk/rope-access-services-made-in-germany/altitec-references-worldwide.html.

[21] http://www.siemens.com/press/en/feature/2012/energy/2012-07-rotorblade.php.

[22] Bent F Sørensen, Erik Jørgensen, Christian P Debel, Find M Jensen, Henrik Myhre Jensen, Torben K Jacobsen, and Kaj M Halling. Improved design of large wind turbine blade of fibre composites based on studies of scale effects. Technical report, Danmarks Tekniske Universitet, Risø Nationallaboratoriet for Bæredygtig Energi, 2004.

[23] Graham Williams, Rohan Baxter, Hongxing He, Simon Hawkins, and Lifang Gu. A comparative study of rnn for outlier detection in data mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 709–712. IEEE, 2002.

[24] Hancong Liu, Sirish Shah, and Wei Jiang. On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9):1635–1647, 2004.

[25] Annick M Leroy and Peter J Rousseeuw. *Robust regression and outlier detection*, volume 1. 1987.

[26] Peter J Rousseeuw and Katrien Van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.

[27] Peter J Rousseeuw and Bert C Van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411): 633–639, 1990.

[28] Peter J Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications Vol. B*, pages 283–297, 1985.

[29] Ali S Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 761–771, 1992.

[30] Ali S Hadi. A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–396, 1994.

[31] Kay I Penny and Ian T Jolliffe. A comparison of multivariate outlier detection methods for clinical laboratory safety data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(3):295–307, 2001.

[32] Marco Riani, Anthony C Atkinson, and Andrea Cerioli. Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 71(2):447–466, 2009.

[33] Greet Pison, Stefan Van Aelst, and G Willems. Small sample corrections for lts and mcd. *Metrika*, 55(1-2):111–123, 2002.

[34] Johanna Hardin and David M Rocke. The distribution of robust distances. *Journal of Computational and Graphical Statistics*, 14(4), 2005.

[35] Anthony C Atkinson and Marco Riani. Distribution theory and simulations for tests of outliers in regression. *Journal of Computational and Graphical Statistics*, 15(2):460–476, 2006.

[36] Scott W Doebling, Charles R Farrar, Michael B Prime, and Daniel W Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review. Technical report, Los Alamos National Lab., NM (United States), 1996.

[37] Jacob M Lifshitz and Assa Rotem. Determination of reinforcement unbonding of composites by a vibration technique. *Journal of Composite Materials*, 3(3): 412–423, 1969.

[38] Charles R Farrar and Keith Worden. *Structural Health Monitoring: A Machine Learning Perspective*. Wiley, 2012.

[39] Anders Rytter. *Vibrational based inspection of civil engineering structures*. PhD thesis, unknown, 1993.

[40] K Worden and JM Dulieu-Barton. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring*, 3(1):85–98, 2004.

[41] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[42] Peter J Shull. *Nondestructive evaluation: theory, techniques, and applications*. CRC press, 2002.

[43] Arnaud Deraemaeker and Keith Worden. *New Trends in Vibration Based Structural Health Monitoring*. Number 520. Springer, 2010.

[44] E Peter Carden and Paul Fanning. Vibration based condition monitoring: a review. *Structural Health Monitoring*, 3(4):355–377, 2004.

[45] Hoon Sohn, Charles R Farrar, Francois M Hemez, and Jerry J Czarnecki. A review of structural health review of structural health monitoring literature

1996-2001. Technical report, Los Alamos National Laboratory, 2002.

[46] Nuno Manuel Mendes Maia and Júlio Martins Montalvão e Silva. *Theoretical and experimental modal analysis*. Research Studies Press Baldock, 1997.

[47] Carlo Rainieri and Giovanni Fabbrocino. Automated output-only dynamic identification of civil engineering structures. *Mechanical Systems and Signal Processing*, 24(3):678–695, 2010.

[48] Jerome Peter Lynch. *Decentralization of wireless monitoring and control technologies for smart civil structures*. PhD thesis, Stanford University, 2002.

[49] Jerome Peter Lynch. An overview of wireless structural health monitoring for civil structures. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):345–372, 2007.

[50] Billie F Spencer, Manuel E Ruiz-Sandoval, and Narito Kurata. Smart sensing technology: opportunities and challenges. *Structural Control and Health Monitoring*, 11(4):349–368, 2004.

[51] R Andrew Swartz and Jerome P Lynch. A multirate recursive arx algorithm for energy efficient wireless structural monitoring. In *Proc. 4th World Conf. on Structural Control and Monitoring*, 2006.

[52] Andrew T Zimmerman, Michihito Shiraishi, R Andrew Swartz, and Jerome P Lynch. Automated modal parameter estimation by parallel processing within wireless monitoring systems. *Journal of Infrastructure Systems*, 14(1):102–113, 2008.

[53] Gyuhae Park, Tajana Rosing, Michael D Todd, Charles R Farrar, and William Hodgkiss. Energy harvesting for structural health monitoring sensor networks. *Journal of Infrastructure Systems*, 14(1):64–79, 2008.

[54] David L Mascarenas, Eric B Flynn, Michael D Todd, Timothy G Overly, Kevin M Farinholt, Gyuhae Park, and Charles R Farrar. Experimental studies of using wireless energy transmission for powering embedded sensor nodes. *Journal of Sound and Vibration*, 329(12):2421–2433, 2010.

[55] Joseph L Rose. A baseline and vision of ultrasonic guided wave inspection potential. *Transactions-American Society of Mechanical Engineers Journal of Pressure Vessel Technology*, 124(3):273–282, 2002.

[56] Zhongqing Su, Lin Ye, and Ye Lu. Guided lamb waves for identification of damage in composite structures: A review. *Journal of sound and vibration*, 295(3):753–780, 2006.

[57] Ajay Raghavan and Carlos ES Cesnik. Review of guided-wave structural health monitoring. *Shock and Vibration Digest*, 39(2):91–116, 2007.

[58] Keith Worden, S Gaerth Pierce, Graeme Manson, WR Philp, Wieslaw J Staszewski, and B Culshaw. Detection of defects in composite plates using lamb waves and novelty detection. *International Journal of Systems Science*, 31(11):1397–1409, 2000.

[59] Keith Worden. Rayleigh and lamb waves-basic principles. *Strain*, 37(4):167–172, 2001.

[60] Lennart Ljung. *System identification*. Wiley Online Library, 1999.

[61] Lennart Ljung. System identification: theory for the user. *Prenice Hall Inf and System Sciencess Series, New Jersey*, 7632, 1987.

[62] Jonas Sjöberg, Håkan Hjalmarsson, and Lennart Ljung. Neural networks in system identification. 1994.

[63] Lennart Ljung and Torkel Glad. Modeling of dynamic systems. 1994.

[64] Keith Worden, Wieslaw J Staszewski, and James J Hensman. Natural computing for mechanical systems research: A tutorial overview. *Mechanical Systems and Signal Processing*, 25(1):4–111, 2011.

[65] Graeme Manson, Keith Worden, and D Allman. Experimental validation of a structural health monitoring methodology: Part ii. novelty detection on a gnat aircraft. *Journal of Sound and Vibration*, 259(2):345–363, 2003.

[66] Graeme Manson, Keith Worden, and D Allman. Experimental validation of a structural health monitoring methodology: Part iii. damage location on an aircraft wing. *Journal of Sound and Vibration*, 259(2):365–385, 2003.

[67] Keith Worden, Graeme Manson, and D Allman. Experimental validation of a structural health monitoring methodology: Part i. novelty detection on a laboratory structure. *Journal of Sound and Vibration*, 259(2):323–343, 2003.

[68] Keith Worden, Graeme Manson, and NRJ Fieller. Damage detection using

outlier analysis. *Journal of Sound and Vibration*, 229(3):647–667, 2000.

[69] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.

[70] Christopher M Bishop et al. *Neural networks for pattern recognition*. Clarendon press Oxford, 1995.

[71] Ian T Nabney. *NETLAB: algorithms for pattern recognition*. Springer, 2004.

[72] Jakob J Verbeek. http://lear.inrialpes.fr/ verbeek/software.php, 2006.

[73] Alexander Ilin and Tapani Raiko. Practical approaches to principal component analysis in the presence of missing values. *The Journal of Machine Learning Research*, 11:1957–2000, 2010.

[74] Sam Roweis. Em algorithms for pca and spca. *Advances in neural information processing systems*, pages 626–632, 1998.

[75] Martin Fodslette Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4):525–533, 1993.

[76] Terence D Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural networks*, 2(6):459–473, 1989.

[77] Peter Foldiak. Adaptive network for optimal linear feature extraction. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 401–405. IEEE, 1989.

[78] C Zang and M Imregun. Structural damage detection using artificial neural networks and measured frf data reduced via principal component projection. *Journal of Sound and Vibration*, 242(5):813–827, 2001.

[79] Douglas Adams, Jonathan White, Mark Rumsey, and Charles Farrar. Structural health monitoring of wind turbines: method and application to a hawt. *Wind Energy*, 14(4):603–623, 2011.

[80] Erik R Jørgensen, Kaj K Borum, Malcolm McGugan, CL Thomsen, Find Mølholt Jensen, CP Debel, and Bent F Sørensen. Full scale testing of wind turbine blade to failure-flapwise loading. *Report Risø*, 2004.

[81] Goutham R Kirikera, Mannur Sundaresan, Francis Nkrumah, Gangadhararao

Grandhi, Bashir Ali, Sai L Mullapudi, Vesselin Shanov, and Mark Schulz. Wind turbines. *Encyclopedia of Structural Health Monitoring.*

[82] Ole JD Kristensen, Malcolm McGugan, Peter Sendrup, Jørgen Rheinländer, Jens Rusborg, Anders M Hansen, Christian P Debel, and Bent F Sørensen. *Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades- a Preproject, Annex E: Full-scale Test of Wind Turbine Blade, Using Sensors and NDT.* Risø National Laboratory, 2002.

[83] LCT Overgaard and E Lund. Structural collapse of a wind turbine blade. part b: Progressive interlaminar failure models. *Composites Part A: Applied Science and Manufacturing*, 41(2):271–283, 2010.

[84] Lars CT Overgaard, Erik Lund, and Ole Thybo Thomsen. Structural collapse of a wind turbine blade. part a: Static test and equivalent single layered models. *Composites Part A: Applied Science and Manufacturing*, 41(2):257–270, 2010.

[85] Jonathan R White, Douglas E Adams, and Mark A Rumsey. Modal analysis of cx-100 rotor blade and micon 65/13 wind turbine. *Structural Dynamics and Renewable Energy, Volume 1*, pages 15–27, 2011.

[86] Bhagwan D Agarwal, Lawrence J Broutman, and K Chandrashekhara. *Analysis and performance of fiber composites.* Wiley, 2006.

[87] Ole Thybo Thomsen. Sandwich materials for wind turbine bladespresent and future. *Journal of Sandwich Structures and Materials*, 11(1):7–26, 2009.

[88] Kevin M Farinholt, Stuart G Taylor, Gyuhae Park, and Curtt M Ammerman. Full-scale fatigue tests of cx-100 wind turbine blades. part i: testing. In *Proc. of SPIE Vol*, volume 8343, pages 83430P–1, 2012.

[89] Stuart G Taylor, Hyomi Jeong, Jae Kyeong Jang, Gyuhae Park, Kevin M Farinholt, Michael D Todd, and Curtt M Ammerman. Full-scale fatigue tests of cx-100 wind turbine blades. part ii: analysis. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 83430Q– 83430Q. International Society for Optics and Photonics, 2012.

[90] James A Dahlberg. Assessment of the lillgrund windfarm: Power performance, wake effects. technical report. *Vattenfall AB*, September 2009.

[91] Angus CW Creech, Wolf-Gerrit Früh, and Peter Clive. Actuator volumes and

hr-adaptive methods for three-dimensional simulation of wind turbine wakes and performance. *Wind Energy*, 15(6):847–863, 2012.

[92] Angus CW Creech, Wolf-Gerrit Früh, and A Eoghan Maguire. High-resolution cfd modelling of lillgrund wind farm. *International Conference on Renewable Energies and Power Quality, Bilbao (Spain)*, 2013.

[93] Angus CW Creech, Wolf-Gerrit Früh, and A Eoghan Maguire. Full-scale simulations of a wind farm using large eddy simulation and a dynamic actuator volume model. *Journal of Fluid Mechanics*, 2013 (submitted).

[94] Matthew D Piggott, Christopher C Pain, Gerard J Gorman, PW Power, and Antony JH Goddard. h, r, and hr adaptivity with applications in numerical ocean modelling. *Ocean Modelling*, 10(1):95–113, 2005.

[95] Franck Bertagnolio, Niels N Sørensen, Jeppe Johansen, and Peter Fuglsang. *Wind turbine airfoil catalogue*. Technical report, Risø National Laboratory, 2001.

[96] Nicolas Jarrin, Sofiane Benhamadouche, Dominique Laurence, and Robert Prosser. A synthetic-eddy-method for generating inflow conditions for large-eddy simulations. *International Journal of Heat and Fluid Flow*, 27(4):585–593, 2006.

[97] Carl Edward Rasmussen and CKI Williams. Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, 38:715–719, 2006.

[98] Douglas M Hawkins. *Identification of outliers*, volume 11. Chapman and Hall London, 1980.

[99] Vic Barnett and Toby Lewis. *Outliers in statistical data*, volume 1. 1984.

[100] Laurie Davies and Ursula Gather. The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423):782–792, 1993.

[101] H Caussinus and A Ruiz. Interesting projections of multidimensional data by means of generalized principal component analyses. In *Compstat*, pages 121–126. Springer, 1990.

[102] Edwin M Knox and Raymond T Ng. Algorithms for mining distance-based outliers in large datasets. In *Proceedings of the International Conference on*

Very Large Data Bases. Citeseer, 1998.

[103] Edwin M Knorr, Raymond T Ng, and Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3-4):237–253, 2000.

[104] Graham J Williams and Zhexue Huang. Mining the knowledge mine. In *Advanced Topics in Artificial Intelligence*, pages 340–348. Springer, 1997.

[105] Stephen D Bay and Mark Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 29–38. ACM, 2003.

[106] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. Wiley-Interscience, 2009.

[107] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *ACM SIGMOD Record*, volume 29, pages 427–438. ACM, 2000.

[108] Shashi Shekhar, Chang-Tien Lu, and Pusheng Zhang. A unified approach to spatial outliers detection. *GeoInformatica*, 7(2):139–166, 2003.

[109] Donald A Jackson and Yong Chen. Robust principal component analysis and outlier detection with ecological data. *Environmetrics*, 15(2):129–139, 2004.

[110] Simon Hawkins, Hongxing He, Graham Williams, and Rohan Baxter. Outlier detection using replicator neural networks. In *Data Warehousing and Knowledge Discovery*, pages 170–180. Springer, 2002.

[111] Tom Fawcett and Foster Provost. Adaptive fraud detection. *Data mining and knowledge discovery*, 1(3):291–316, 1997.

[112] Richard Arnold Johnson and Dean W Wichern. *Applied multivariate statistical analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 2002.

[113] Edgar Acuna and Caroline Rodriguez. A meta analysis study of outlier detection methods in classification. *Technical paper*, 2004.

[114] Claudia Becker and Ursula Gather. The masking breakdown point of multivariate outlier identification rules. *Journal of the American Statistical Association*,

94(447):947–955, 1999.

[115] Randy J Pell. Multiple outlier detection for multivariate calibration using robust statistical techniques. *Chemometrics and Intelligent Laboratory Systems*, 52(1):87–104, 2000.

[116] Nima Moshtagh. Minimum volume enclosing ellipsoid. *Convex Optimization*, 2005.

[117] Leonid G Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2):307–320, 1996.

[118] Leonid G Khachiyan and Michael J Todd. On the complexity of approximating the maximal inscribed ellipsoid for a polytope. *Mathematical Programming*, 61 (1):137–159, 1993.

[119] Piyush Kumar and EA Yildirim. Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1):1–21, 2005.

[120] Peng Sun and Robert M Freund. Computation of minimum-volume covering ellipsoids. *Operations Research*, pages 690–706, 2004.

[121] Sabine Verboven and Mia Hubert. Libra: a matlab library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, 75(2):127–136, 2005.

[122] Mia Hubert, Peter J Rousseeuw, and Stefan Van Aelst. High-breakdown robust multivariate methods. *Statistical Science*, pages 92–119, 2008.

[123] Guido De Roeck. The state-of-the-art of damage detection by vibration monitoring: the simces experience. *Journal of Structural Control*, 10(2):127–134, 2003.

[124] Elizabeth J Cross. *On structural health monitoring in changing environmantl and operational conditions*. PhD thesis, May 2012.

[125] Ifigeneia Antoniadou, Elizabeth J Cross, and Keith Worden. Cointegration and the empirical mode decomposition for the analysis of diagnostic data. *Key Engineering Materials*, 569:884–891, 2013.

[126] Elizabeth J Cross, Keith Worden, and Qian Chen. Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data. *Proceedings of the Royal Society A: Mathematical, Physical and*

Engineering Science, 467(2133):2712–2732, 2011.

[127] Robert J Barthorpe. *On Model-and Data-based Approaches to Structural Health Monitoring*. PhD thesis, 2010.

[128] Peter J Rousseeuw and Katrien Van Driessen. Computing lts regression for large data sets. *Data Mining and Knowledge Discovery*, 12(1):29–45, 2006.

[129] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[130] Geoffrey E Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.

[131] MarcAurelio Ranzato, Fu Jie Huang, Y-L Boureau, and Yann Lecun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[132] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[133] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[134] William Tyrrell Thomson. *Vibration theory and applications*. Prentice-Hall Englewood Cliffs, NJ, 1965.

[135] Robert K Vierck. *Vibration analysis*. Crowell, 1979.

[136] David J Ewins. *Modal testing: theory and practice*, volume 79. Research studies press Letchworth, 1984.

[137] Peter Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *Audio and Electroacoustics, IEEE Transactions on*, 15(2):70–73, 1967.

[138] Monson H Hayes. *Statistical digital signal processing and modeling*. Wiley. com, 2009.

[139] Craig Marven, Gillian Ewers, and Texas Instruments. *A simple approach to*

digital signal processing. Wiley, 1996.

[140] Stéphane Mallat. *A wavelet tour of signal processing.* Access Online via Elsevier, 1999.

[141] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[142] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[143] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171, 1970.

[144] Alexis Roche. Em algorithm and variants: An informal tutorial. *arXiv preprint arXiv:1105.1476*, 2011.

PUBLICATIONS

Author publications to date

Journal Papers

- Dervilis, N., Choi, M., Taylor, S. G., Barthorpe, R. J., Park, G., Farrar, C. R., Worden, K. On Damage Diagnosis for a Wind Turbine Blade using Pattern Recognition. Accepted to Journal of Sound and Vibration.

- Dervilis, N., Cross, E. J., Barthorpe, R. J., Worden, K. Robust methods of inclusive outlier analysis for Structural Health Monitoring. Submitted to Journal of Sound and Vibration.

Reviewed Conference Papers

- Dervilis, N., Barthorpe, R., Staszewski, W. J., Worden, K., (2012). Structural Health Monitoring of Composite Material Typical of Wind Turbine Blades by Novelty Detection on Vibration Response. Key Engineering Materials, 518, 319-327.

- Dervilis, N., Choi, M., Antoniadou, I., Farinholt, K. M., Taylor, S. G., Barthorpe, Park, G., R. J., Worden, K., Farrar, C. R. (2012). Novelty detection applied to vibration data from a CX-100 wind turbine blade under fatigue loading. In Journal of Physics: Conference Series (Vol. 382, No. 1, p. 012047). IOP Publishing.

- Dervilis, N., Choi, M., Antoniadou, I., Farinholt, K. M., Taylor, S. G., Barthorpe, R. J., Park, G., Farrar, C. R., Worden, K., (2013). Machine Learning Applications for a Wind Turbine Blade under Continuous Fatigue Loading. Key Engineering Materials. (in press)

- Antoniadou, I., Manson, G., Taylor, S. G., Dervilis, N., Worden, K., Farrar, C. R., (2013).Damage detection of RAPTOR Telescope systems using time-frequency analysis methods. Key Engineering Materials. (in press)

- Dervilis, N., Barthorpe, R. J., Worden, K. (2013). Comparative Study of Robust Novelty Detection Techniques. Key Engineering Materials, 569, 1109-1115.

- Antoniadou, I., Dervilis, N., Barthorpe, R. J., Manson, G., Worden, K., (2013). Advanced Tools for Damage Detection in Wind Turbines. Key Engineering Materials, 569, 547-554.

- Dervilis, N., Creech, A. C. W., Barthorpe, R., Maguire, A. E., Antoniadou, I., Worden, K., (2013). An SHM view of a CFD model of Lillgrund Wind farm. Applied Mechanics and Materials. (submitted)

- Antoniadou, I., Howard, T. P., Dwyer-Joyce, R. S., Marshall, M. B., Naumann, J., Dervilis, N., Worden, K., (2013). Envelope analysis using the Teager-Kaiser energy operator for condition monitoring of a wind turbine bearing. Applied Mechanics and Materials. (submitted)

Conference Papers

- Dervilis, N., Barthorpe, R., Antoniadou, I., Staszewski, W. J., Worden, K. (2012). Damage detection in carbon composite material typical of wind turbine blades using auto-associative neural networks. In Proc. of SPIE Vol (Vol. 8348, pp. 834806-1).

- Antoniadou, I., Manson, G., Dervilis, N., Staszewski, W. J., Worden, K., (2012). On damage detection in wind turbine gearboxes using outlier analysis. In SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring (pp. 83430N-83430N). International Society for Optics and Photonics.

- Dervilis, N., Barthorpe, R., Antoniadou, I., Worden, K., (2012). Impact

Damage Detection for Composite Material Typical of Wind Turbine Blades Using Novelty Detection. $In6^th$ European Workshop on Structural Health Monitoring, EWSHM, Dresden.

- Antoniadou, I., Manson, G., Dervilis, N., Barszcz, T., Staszewski, W. J., Worden, K., (2012). Condition Monitoring of a Wind Turbine Gearbox Using the Empirical Mode Decomposition Method and Outlier Analysis. In $6^{t}h$ European Workshop on Structural Health Monitoring, EWSHM, Dresden..

- Antoniadou, I., Manson, G., Dervilis, N., Barszcz, T., Staszewski, W. J., Worden, K., (2012). Instantaneous Characteristics Estimation using the Teager-Kaiser Operator for the Condition Monitoring of a Wind Turbine Gearbox. In Proceedings of International Conference on Noise and Vibration Engineering, ISMA, Leuven, Belgium.

- Dervilis, N., Barthorpe, R., Worden, K., (2013). Auto-Association and Novelty Detection: Truths and Myths?. In Proceedings of International Workshop on Structural Health Monitoring, IWSHM, Stanford .