THE UNIVERSITY
of EDINBURGH

DATA-DRIVEN FRAMEWORKS FOR ROBUST AND INTERPRETABLE DAMAGE DETECTION IN WIND TURBINE BLADES

By

Artur Movsessian

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Ph.D.)

In Institute for Infrastructure and Environment

THE UNIVERSITY OF EDINBURGH

2022

Abstract

Wind energy has rapidly become one of the pillars of the energy transition generating more than 20% of the world's renewable electricity in 2019. By the end of 2020, 743 GW of wind capacity had been installed globally. However, to limit global average temperature increases to 1.5°C, approx. 180 GW (including both onshore and offshore wind) must be installed annually until 2030.

To ramp up wind energy's contribution to the energy transition, the cost competitiveness of this technology must be addressed. The levelised cost of electricity (LCOE) from wind is heavily influenced by operational costs. Approx. 30% of the LCOE from onshore wind installations corresponds to operation and maintenance (O&M) costs. This share is higher for offshore wind installations due to the harsher marine environment, skilled labour and specialised vessels required, among other reasons.

Maintenance, particularly unscheduled interventions, contribute to downtime. Blades represent a critical component for the reliable operation of wind farms with a failure rate of 0.7 occurrences per year and per turbine in the case of offshore wind. Economic and safety consequences are anticipated when such failures are not timely identified. Blade failures have been identified as the main cause of accidents in wind turbines between 2012 and 2020. Therefore, the research presented in this dissertation is motivated by the frequency and severity of failures in wind turbine blades (WTBs).

Vibration-based structural health monitoring (VSHM) systems can contribute to reducing the number of failures and accidents, enhancing profitability, reliability and safety of wind farms. With the rapid development of sensing technologies, VSHM systems enable online and continuous remote integrity monitoring. VSHM systems, which are based on the measurement of vibration signals on structures, have been

successfully implemented in offshore oil platforms, aerospace structures and bridges where the damage location is generally unknown and/or accessibility issues arise. Wind turbines share these challenges, yet VSHM systems have not been widely implemented.

One of the prevailing challenges for damage detection in WTBs using VSHM systems is the influence of environmental and operational variabilities (EOVs) on the structure and, ultimately, on the sensor readings. EOVs such as temperature, wind conditions and moisture and humidity levels can either camouflage damage or falsely indicate damage. Despite significant efforts to mitigate the influence of EOVs on damage detection and the various approaches adopted by the scientific community to face this challenge, there is no consensus on one single approach that can be generalised and identified as the most accurate and reliable methodology for damage detection. This dissertation introduces a robust, interpretable, data-driven and semi-supervised damage detection framework based on machine learning techniques to address the influence of EOVs.

Contrary to conventional damage detection frameworks, where the damage sensitive features (DSFs) are normalised, in this dissertation's proposed framework, the normalisation of the EOVs is aimed at the novelty index. The effects of EOVs remain perceptible in the novelty indices and are later mitigated through a supervised regression model that is trained to learn the relationships between the DSFs extracted from vibration measurements from an in-operation wind turbine blade, EOVs and novelty indices.

The damage detection framework proposed results from the combination of two approaches: one approach addresses the robustness of the damage detection process and the second approach addresses the issue of interpretability of data-driven novelty detection techniques. Different combinations of techniques including artificial neural

networks (ANN), gradient boosted decision trees (XGBoost DT) and Shapley Additive Explanations (SHAP) were tested on data extracted from an operating Vestas V27 wind turbine blade and compared to the widely adopted Mahalanobis distance (MD)-based approach.

The experiments and analyses conducted throughout this dissertation have shown that the ANN-based approach provides a mitigated novelty index that leads to high damage detection accuracy and the interpretable ML techniques, i.e., the XGBoost DT - SHAP sequence, facilitate the understanding of the effect of EOVs on the monitored structure without mitigating this effect. The XGBoost - SHAP sequence proposed results in a trade-off between accuracy and interpretability, however, this approach still provides higher accuracy relative to the MD-based approach. Ultimately, the choice of approach for damage detection will be application-specific and requires domain knowledge and expertise. The operator should decide whether the trade-off between accuracy and interpretability is acceptable for the monitored structure.

Keywords: Structural Health Monitoring; Damage detection; Wind turbine blades; Machine learning; Novelty detection; Environmental and operational variabilities; Artificial Neural Networks; Gradient Boosted Decision Trees; Shapley Additive Explanations

Declaration

I hereby declare that the thesis entitled:

"Data-Driven Frameworks for Robust and Interpretable Damage Detection in Wind Turbine Blades"

and the work presented in the thesis and any included publications, with the exception of sources of help, which have been acknowledged, were originated entirely by myself. I confirm that the content of this thesis has previously not been submitted for any other degree or diploma.

Artur Movsessian

Date

Dedication

To Dr. Alla Silberberg and Vrezh Movsessian

Мои дорогие и любимые родители! От всей души я хочу вас поблагодарить за ту помощь и поддержку, которую вы мне оказывали при написании моей диссертации. Я всегда чувствую ваше надёжное родительское плечо. Для меня очень ценными являются ваши советы, предостережения и рекомендации. Но моя признательность вам ограничивается не только сроком написания моей научной работы. Своим достижениями и плодотворной научной деятельностью я обязан вам ещё с раннего детства, когда вы мне помогали постигать предметы школьной программы, добиваться необходимого результата. Без этого титанического труда и усидчивости не было бы и сегодняшнего результата - моей успешной научно - исследовательской работы. Вы мне привили искренний интерес к учёбе и научным познаниям. Своим успехом я обязан вам на всех этапах подготовки и выполнения диссертации. Я благодарен вам за ваше внимание, терпение, принципиальноость, доброту, родительскую мудрость и любовь ко мне. Низкий вам за это поклон!

"И вечный бой! Покой нам только снится..."

Александр Блок

Contents

List of Figures

List of Tables

Acknowledgments

Throughout my Ph.D. studies and writing the dissertation I have received a great deal of support and assistance.

First and foremost, I am extremely grateful to my supervisors, Dr. David García Cava and Dr. Dmitri Tcherniak for their invaluable advice, continuous support, and patience to all my concerns and crazy ideas during my PhD. Their immense knowledge and plentiful experience have encouraged me throughout my academic research. You have also created an environment for me, which kept me motivated and dedicated. I feel extremely fortunate that I had you as my supervisors.

Additionally, I would like to thank Dr. Kim Branner and the members of the Large Scale Facility at DTU Risø Campus, who have included me in their blade fatigue testing and granted me an opportunity to collaborate with their research. It was an exciting practical experience for me and a big contribution to my dissertation.

I would also like to use this chance to thank all the people who collaborated with me throughout my studies. Dr. Bilal Ali Qadri and Dr. Martin Ulriksen from the University of Aalborg; Panagiotis Martakis, Dr. Yves Reuland and Dr. Sai Pai from ETH Zürich; and Said Quka from the University of Bologna. I feel very fortunate to had been able to work with you all and thank you for enriching my Ph.D. experience.

Dr. Maria Mercedes Movsessian, my best friend, my wife. I would like to thank you for all the late nights you have dedicated to listening to my thoughts and concerns during this Ph.D. You have supported me in all possible ways, from academic advice to my everyday life. Thank you for keeping me motivated and on track during the studies, which took place in very unusual times. I owe you a lot.

List of Abbreviations

AANN	auto-associative neural network
AIC	Akaike's information criterion
ANN	Artificial neural networks
AR	Autoregressive (model)
ARMAX	autoregressive moving average model with exogenousinputs
ARX	Auto-Regressive with eXternal model input
BIC	Bayesian information criterion
COMAC	Coordinate modal assurance criteria
CWFI	Caithness Windfarm Information Forum
CWT	Continuous wavelet transform
DI	Damage index
DSF	Damage sensitive feature
DT	Decision trees
EOV	Environmental and operational variabilities
FEM	Finite element model
FOWT	floating offshore wind turbine
FRF	Frequency response function
GMM	Gaussian Mixture model
GPR	Gaussian process regression
GRU	Gated Recurrent Unit
GW	Gigawatts
GWEC	Global Wind Energy Council
GWIPCC	Intergovernmental Panel for Climate Change
IEA	International Energy Agency
IPCC	Intergovernmental Panel for Climate Change

LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory
MAC	Modal assurance criteria
MD	Mahalanobis Distance
ML	Machine learning
MW	Megawatt
NLPCA	Nonlinear Principal Component Analysis
NN	Neural network
O&M	Intergovernmental Panel for Climate Change
PC	Principal component
PCA	Principal component analysis
RBF	Radial Basis Function
ROC	Receiver operating characteristic
SCADA	Supervisory control and data acquisition
SHAP	Shapley Additive exPlanations
SHM	Structural Health Monitoring
SSA	Singular spectrum analysis
SVM	Support vector machines
VSHM	Vibration-based Structural Health Monitoring
WTB	Wind turbine blade
XGBoostDT	eXtreme gradient boosting decision trees

Publications

Journal papers

Movsessian, A., García Cava, D., & Tcherniak, D. (2021). An artificial neural network methodology for damage detection: Demonstration on an operating wind turbine blade. *Mechanical Systems and Signal Processing*, 159, 107766. https://doi.org/10.1016/j.ymssp.2021.107766

Movsessian, A., García Cava, D., & Tcherniak, D. (2021). Interpretable machine learning in damage detection using Shapley Additive Explanations. *ASME J. Risk Uncertainty Part B.* http://dx.doi.org/10.1115/1.4053304

Martakis, P., **Movsessian, A.**, Reuland, Y., Pai, S. G., Quqa, S., García Cava, D., Tcherniak ,D., & Chatzi, E. (2021). A semi-supervised interpretable machine learning framework for Sensor Fault detection. Smart Structures and Systems, An International Journal. http://doi.org/10.12989/sss.2022.29.1.251

Quqa, S., Martakis, P., **Movsessian, A.**, Pai, S., Reuland, Y., & Chatzi, E. (2021). Two-step approach for fatigue crack detection in steel bridges using convolutional neural networks. Journal of Civil Structural Health Monitoring, 1-14. https://doi.org/10.1007/s13349-021-00537-1

Book chapters

García Cava, D., Avendaño-Valencia, L. D., **Movsessian, A.**, Roberts, C., and Tcherniak, D. (2022).On Explicit and Implicit Procedures to Mitigate Environmental and Operational Variabilities in Data-Driven Structural Health Monitoring, pages 309–330. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-81716-9_15

Peer-reviewed conference papers

Movsessian, A., García Cava, D., & Tcherniak, D. (2020). Adaptive feature selection for enhancing blade damage diagnosis on an operational wind turbine. *In Proceedings of the 13th International Conference on Damage Assessment of Structures* (pp. 594-605). Springer, Singapore. https://doi.org/10.1007/978-981-13-8331-1_44

Movsessian, A., Qadri, B. A., Tcherniak, D., García Cava, D., & Ulriksen, M. D. (2020). Mitigation of environmental variabilities in damage detection: a comparative study of two semi-supervised approaches. In EURODYN 2020: *XI international conference on structural dynamics.* https://doi.org/10.47964/1120.9103.19618

Movsessian, A., García Cava, D., Tcherniak, D., & Janeliukstis, R. (2020, November). A methodology on interpretable novelty detection. In *International Conference on Structural Dynamics* (EURODYN 2020) (pp. 922-935). https://doi.org/10.47964/1120.9073.19621

Quqa, S., Malatesta, M., Martakis, P., & **Movsessian, A.** (2020). Investigation

on Damage Sensitive Features for Optimal Sensor Networks based on Real-Scale Recordings. In EURODYN 2020 *Proceedings of the XI International Conference on Structural Dynamics Streamed from Athens, Greece 23-26 November 2020* (Vol. 1, pp. 936-947). European Association for Structural Dynamics (EASD). http://doi.org/10.47964/1120.9074.19628

Chapter 1

Introduction

This chapter provides background information related to the development of the wind energy sector in the last decade as well as an expected outlook to 2030. Failure statistics are provided to examine the most prevailing causes of damage in wind turbines. Blades have been identified as the main cause of accidents between 2012 and 2020 and a critical component with high failure rates, particularly in offshore wind farms. The research presented in this dissertation is motivated by the frequency and severity of failures in wind turbine blades (WTBs). Common types of damages in WTBs are described in this chapter and vibration-based structural health monitoring (VSHM) is introduced as a potential solution to reduce the number of failures and accidents. Further, this chapter describes conceptual learning frameworks for VSHM along with tools for decision making such as control charts. These are fundamental componenst of online monitoring systems. The chapter ends introducing one of the prevailing challenges for damage detection in WTBs, namely the influence of environmental and operational variabilities (EOVs).

1.1 Wind energy

According to the Intergovernmental Panel for Climate Change (IPCC), renewable energy needs to supply 70-85% of the global electricity by 2050 to limit global warming to 1.5 °C [80]. Wind energy has rapidly become one of the pillars of the energy transition generating approx. 21% of the world's renewable electricity in 2019 [131]. This energy sector is currently the largest clean energy industry after hydropower with 743 GW installed by the end of 2020 [68]. Despite this progress, an average of 180 GW need to be installed annually to limit global average temperature increases to no more than 1.5 °C as can be seen in Figure 1.1.

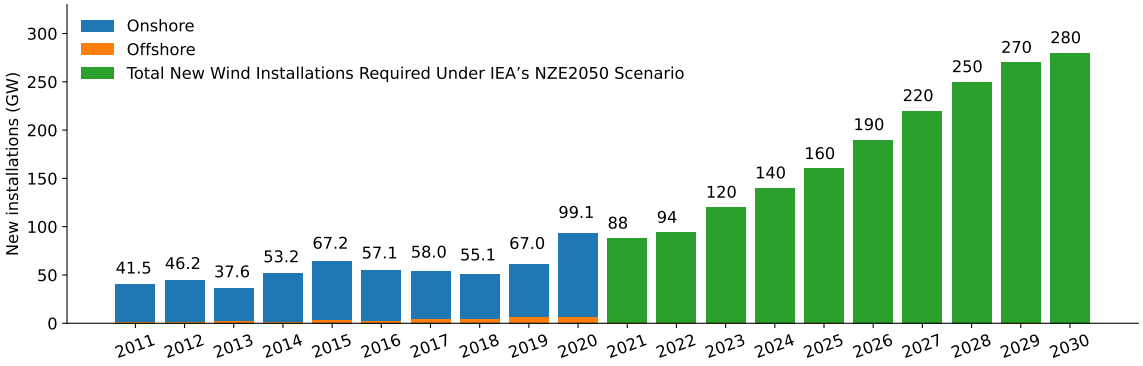


Figure 1.1: Historic development of annual new wind energy installations and outlook. Source: Global Wind Energy Council (GWEC) [68] based on GWEC Market Intelligence and IEA World Energy Outlook [79].

A key aspect of both the energy transition and the success of the wind energy sector is cost competitiveness, which is heavily influenced by operational costs. Operation and maintenance (O&M) costs represent up to 30% of the levelised cost of electricity for onshore wind [81]. However, these costs are expected to be higher for offshore wind turbines due to the harsher marine environment, the need for skilled labour and specialised vessels and availability and accessibility issues.

Wind turbines are designed to operate for a minimum of 20 years [73]. Throughout this lifetime, wind turbines are continuously exposed to fluctuating loads and harsh environmental conditions, especially in locations with highly turbulent wind. Generally, scheduled maintenance requiring between 24 and 50 man-hours per turbine is performed a few times per year on the wind turbines [117]. On the contrary, unscheduled maintenance, i.e., failure repairs, is several orders of magnitude more costly to conduct, requiring significantly more time per turbine and increasing downtime.

Some components and subsystems are particularly critical. With information from different databases and over 18000 wind turbines, Dao et al. [36] compiled the failure rates of different components and subsystems in both onshore and offshore wind turbines. Figure 1.2 shows that electrical, control system, generator, pitch and blades and hub experience high failure rates in both types of installations, however, the average failure rates are higher offshore than onshore.

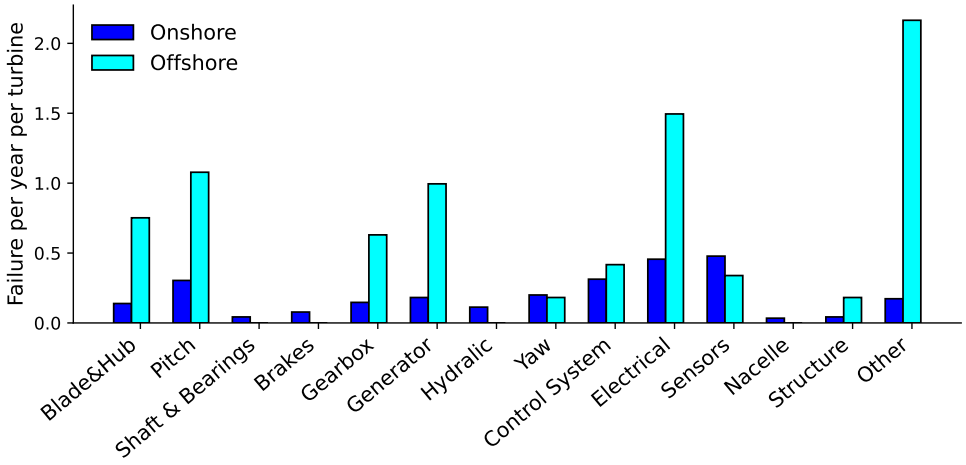


Figure 1.2: Comparison of weighted failure rates between onshore and offshore databases [36].
Note: "Other" refers to equipment in the turbine that could not be associated with any other specific category, e.g., small components that might not necessarily be associated with the function of the turbine such as cabinets, lighting, etc.

If these failures are not timely identified, economic and safety consequences are to be expected. Particularly blade failures can cause significant damage to a wind turbine

as shown in Figure 1.3.

Figure 1.3: Blade Failure. Source: [74]

The Caithness Windfarm Information Forum (CWFI) reported around 238 accidents related to blade failure between 2012 and 2020. As can be seen in Figure 1.4, blade failure has been the main cause of accidents in wind turbines during this period. This suggests that this component is critical in terms of reliability, productivity, cost competitiveness and, most importantly, safety.

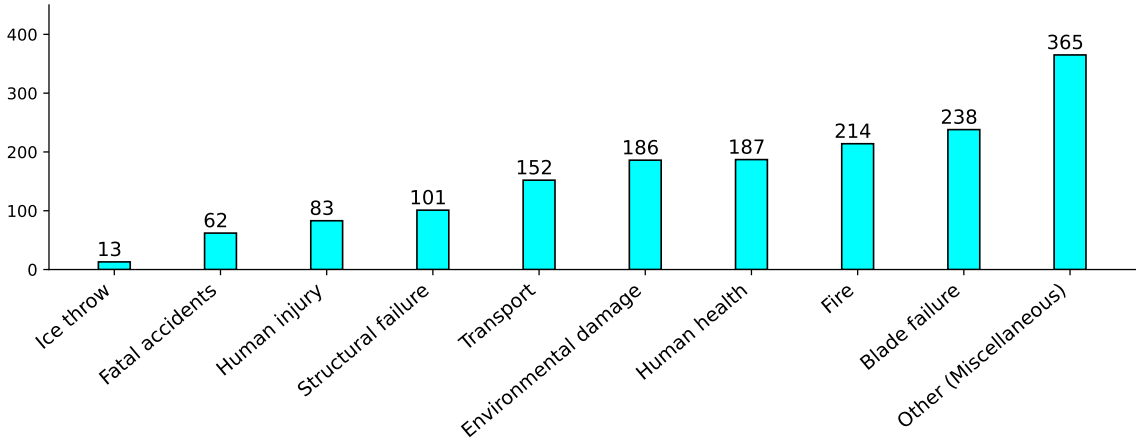


Figure 1.4: Wind turbine accidents registered between 2012 and 2020. Source: CWFI [18]

VSHM systems can contribute towards enhancing the profitability, reliability and safety of wind farms. Furthermore, these systems can reduce O&M costs and provide an assessment of the structural integrity of the turbines and their blades.

1.1.1 Wind turbine blades

Blades are one of the most expensive components of a wind turbine and arguably the most important [14]. As a consequence of the rapid expansion of wind energy in the last decades, wind turbines have been continuously growing in size and rated power since their first commercial introduction in the 1980's. Currently, the largest installed turbines have rotor diameters of 220m and rated power of 12 MW [44] but larger installations are expected in the future, particularly offshore. The increased blade weight requires extensive laboratory testing to ensure safety and reliability over the lifetime of the turbine.

WTBs are manufactured from composite materials. This mainly includes glass fiber, carbon fiber, balsa wood or foam. The proper choice of composite materials can increase the strength to weight ratio and, thereby, contribute to safety and reliability. Typically, a WTB is formed by a shell in the suction side and a shell in the pressure side. One or several vertical shear webs are fitted between the shells, assembling the shells together and enabling the transfer of shear loads. Some designs include a box beam formed by upper and lower spar caps and vertical shear webs (see Figure 1.5). This structural design provides bending stiffness and torsional rigidity to resist the flapwise bending which refers to the bending moment in the Y direction shown in Figure 1.5.

Generally, WTBs were manufactured using the wet hand lay-up technology in open molds. In this process, the two shells, the shear webs and the spar caps are bonded to the trailing and leading edges through a highly-toughness adhesive. However, this process implies high labour costs, lower quality of products and some environmental concerns [105]. More recently, the prepreg technology has been adapted from the

aircraft industry and adopted in the wind industry. This technology allows the industrial impregnation of composite fibers to form them into complex shapes. The resin infusion technology is widely used to manufacture long blades. The outside of the blade is covered by a gel coat to provide protection against ultraviolet degradation and water penetration.

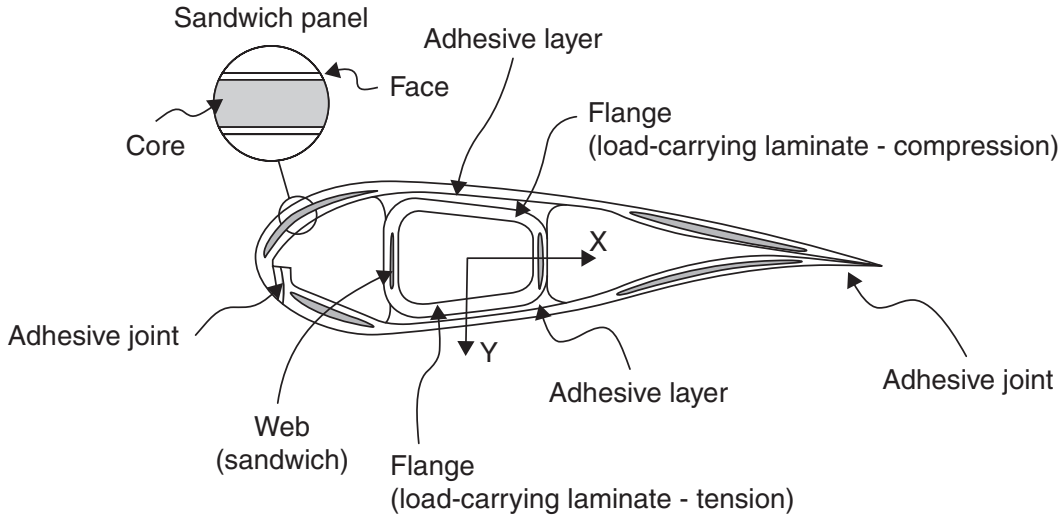


Figure 1.5: Example of a WTB design with a box beam [82]

Despite the advanced designs and technologies adopted for the manufacture of WTBs, they remain susceptible to damage as previously discussed in section 1.1.

1.1.2 Types of damages

The design of a WTB implies a trade-off between aerodynamic and structural efficiency. Therefore, WTBs are not optimised with regards to their structural strength. There are several causes for failures in blades, namely:

- *Leading Edge Erosion*: Abrasive airborne particles can impact and gradually

destroy the leading edge of the blade during operation. The erosion especially impacts blades operating at high rpm and particularly the regions closer to the top as these are exposed to higher velocities. Leading edge erosion can change the aerodynamic profile affecting the efficiency and creating delaminations on the leading edge [145].

- *Lightning*: Wind turbines are lightning magnets, especially the blades are vulnerable parts which can be completely destroyed by a lightning impact. Although blade manufacturers implement lightning protection, cracks are present where the lightning has hit the blade [32].

- *Icing*: The best locations for wind turbines are often windy areas in higher altitudes. In these locations, the blades experience low temperatures and are subjected to icing. Various problems occur due to icing on blades, e.g. interruption of energy generation, power reduction due to loss in aerodynamic efficiency [76] and increased fatigue in components such as the turbine tower, reducing the lifetime of a wind turbine [58].

- *Fatigue loads*: Wind turbines are continuously exposed to a stochastic wind field with various extreme events. This exposes the blade to cyclic loads and can introduced damage due to fatigue. If the damage is not timely identified, it can lead to failure and a collapse of the wind turbine.

Wind turbine failures can be subdivided based on the different damage types that lead to failure. These damage types are well defined and summarized by Debel [38] as follows:

- Type 1 - Skin/adhesive debonding: Damage in the adhesive layer bonding skin and main spar flanges

- Type 2 - Adhesive joint failure: Damage in the adhesive layer joining the up and downwind skins along the leading and/or trailing edges
- Type 3 - Sandwich debonding: Damage at the interface between face and core in sandwich panels in skins and main spar web
- Type 4 - Delamination: Internal damage in laminates in skin and/or main spar flanges, under a tensile or compression load
- Type 5 - Splitting along fibers: Splitting and fracture of separate fibres in laminates of the skin and main spar
- Type 6 - Skin/adhesive debonding by buckling: Buckling of the skin due to damage in the bond between skin and main spar under compressive load. This type of damage is a specific Type 1 case.
- Type 7 - Cracks in gel-coat: Debonding of the gel-coat from the skin

Delamination is potentially the most critical type of damage. It occurs in layered composite structures with poor bonding between layers. Delamination can develop from impact damage during the transport or service or from the manufacturing process. The most widely observed type of delamination is due to buckling, i.e. a high interlaminate stress that leads to the spreading of a crack. Generally, identifying this type of damage is crucial to avoid failures and accidents in WTBs.

1.2 Vibration-based structural health monitoring

Structural and mechanical systems are the backbone of modern societies. Guaranteeing the integrity of these systems is relevant for economic and safety reasons as

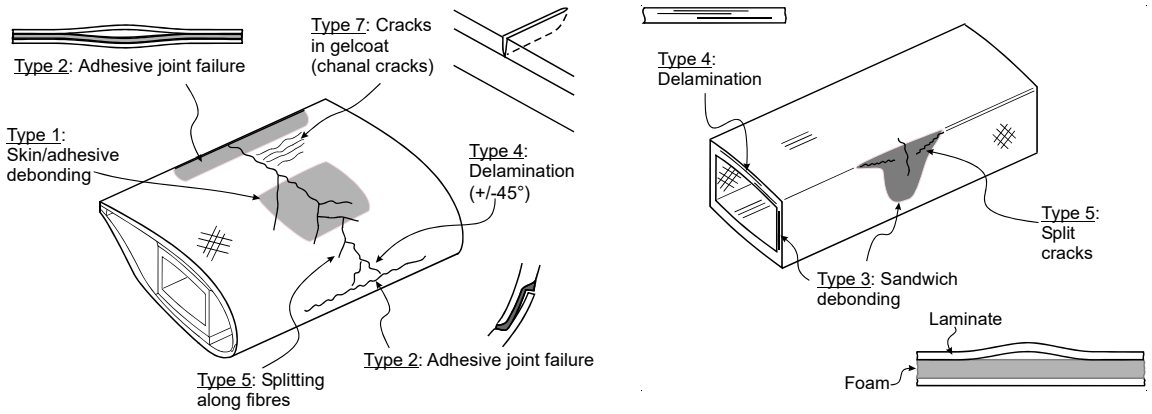


Figure 1.6: Common damage types in WTBs [146]

resilient infrastructure leads to fewer incidents and longer operating periods. Structural Health Monitoring (SHM) refers to the evaluation of the health or condition of a structure to identify damage. Traditionally, a change in the physical characteristics of the system, such as the geometry or material, is considered damage. This change affects the performance, safety, reliability, and the operational lifetime of the system. Therefore, detecting the presence of damage in civil structures has become highly important.

Initially, the structural integrity of a system was monitored through visual inspection. However, with the rapid development of sensing technologies, SHM has evolved in a complex and interdisciplinary research field ranging from experimental testing, through data acquisition and management and system identification to long-term measurement of environmental and operational conditions. With the development of sensing technologies, vibration-based SHM (VSHM) has enabled remote integrity monitoring. VSHM techniques are based on the measurement of vibration signals on structures which are further implemented in a monitoring process to detect damage.

VSHM systems have been successfully implemented in offshore oil platforms, aerospace structures and bridges [50]. In these applications, the damage location is unknown and there may be accessibility issues for inspections and monitoring. Thus,

remote integrity monitoring became an attractive workaround for these challenges. Wind turbines face similar challenges with damage appearing between inspections and the structures being in hard-to-access and/or remote locations. However, due to the recent emergence of the wind energy industry relative to the mature oil, gas and aerospace industries, VSHM has not yet become a widespread solution to monitor WTBs and has been mainly restricted to lab-scale experiments. VSHM systems have the potential to enable in-operation, continuous monitoring of the WTB structural integrity, thereby contributing to reduce failures and improving the reliability and cost competitiveness of what is expected to be one of the key energy technologies in low-carbon energy systems.

1.2.1 General monitoring framework

Monitoring structures such as WTBs involves the observation of the structure during a period of time and the collection of dynamic response measurements from sensors installed on it. This process also involves the extraction of damage sensitive features (DSF) from the measurements, which are further subjected to a statistical analysis to identify damage in the structure. Farrar et al. (2001) and Sohn et al. (2003) describe the problem of damage detection in a context of a statistical pattern recognition paradigm as a four-step process:

1. *Operational evaluation* aims to justify the implementation of an SHM system from the perspective of life-safety and economics. This first step defines damage and, in the case of different types of damage, it provides a prioritisation. This prioritisation could facilitate the management and repairs of failures in hard-to-access WTBs. Furthermore, during this step, the operational and environmental conditions under which the structure or system operates are identified and

considered as limitations when implementing the SHM system. This can be of particular interest for WTBs as they operate under harsh wind and temperature conditions.

2. *Data acquisition, cleansing and normalisation* are three sub-steps which come together to provide the measured data for further evaluation. Data Acquisition includes selecting the type and amount of sensors as well as the location for installation. Further, the excitation method is chosen which can be, e.g., an external excitation by an actuator and/or ambient excitation. Data cleansing is a process of detecting corrupted data which includes incomplete, irregular or inaccurate measurements. Data cleansing can be an automated process when, e.g., statistical techniques are used to fill in missing data or remove outliers. However, this can also require discretionary knowledge of the monitoring team which is aware of irregularities and/or interruptions during the measurement campaign (e.g., a loose sensor). In the case of WTBs, this can be, e.g., domain knowledge regarding the operation of the WT such as the cut in and cut out wind speeds. Finally, as variabilities in the measurements are expected due to environmental and operational conditions, normalising the data is important to ease the discrimination between the effect of the sources of variabilities and damage.

3. *Feature extraction.* A feature is a crucial element for pattern recognition containing information that characterises the state of a structure. Generally, a direct measurement of a system does not provide a condense description of the state or condition of the structure. Thus, a condensed representation of the measurements is used, also referred to as DSF in SHM applications. As the name suggests, DSFs are expected to hold information regarding the structural integrity and are further implemented in a statistical model to distinguish between a damaged and an undamaged state. DSFs are application specific and commonly identified through analytical approaches including finite element

models or numerical simulations which are known as modal parameters. DSFs also include non-modal based quantities which can include descriptive statistics or any type of data transformation of the vibration responses.

4. *Statistical and machine learning model development for damage identification.* Once DSFs have been selected, statistical or machine learning models are implemented to identify and quantify the damage state of the structure. These algorithms analyse the distribution of the selected features to better understand the condition of the structure.

   Fundamentally, and based on Rytter [135], Farrar et al. [52] defined the damage identification problem, and thus the SHM process, as a hierarchical structure with five levels:

   (i) *Detection*: Detecting the existence of the damage on the structure

   (ii) *Localisation*: Locating the damage

   (iii) *Classification*: Identifying the type of damage

   (iv) *Assessment*: Quantifying the severity of the damage

   (v) *Consequences*: Determining the actual safety of the structure after damage

In this hierarchical structure, the information of the previous level is a pre-requisite for the next level. Damage identification ultimately serves as a decision-making tool for the operator.

1.2.2 Approaches for damage identification

Damage identification has been extensively researched by the scientific community in the past decades. A variety of methods have been adopted to identify damage in

civil structures [3, 37]. These methods can be classified into two main approaches: *physics-based* and *data-driven*.

Physics-based SHM relies on the physical characteristics of a structure such as natural frequency, mode and curvature. These characteristics are modelled through analytical models that are constructed from the combination of first principles such as Newton's second law with additional closure models (e.g., a constitutive model). This is also known as a model-based approach with, e.g., a finite element model being updated or re-calibrated with newly measured data. The condition of the structure is evaluated based on the discrepancies between the calibrated and the updated model. This approach is well-established, nevertheless, still faces the challenge of high computational cost. A critical limitation of physics-based models are the modelling errors that result from model simplification and the complexity of the physical system [45].

On the contrary, the data-driven approach does not involve models that are based on physical laws. This approach commonly extracts DSFs from measurements and employs statistical pattern recognition or machine-learning techniques to identify damage in a structure. Compared with the physics-based approach for damage identification, the data-driven approach does not require a complex numerical model and has the potential of identifying damage in a structure regardless of the influence of operational and environmental variabilities such as wind and temperature. Not being based on physical laws and numerical models makes data-driven approaches more cost effective in terms of computational power yet these are often limited in terms of the interpretability of the results [45]. Also, data-driven pattern recognition techniques commonly require a large amount of data recorded over, e.g., different seasons in a year to identify a robust pattern, which introduces another limitation.

Despite the limitations of data-driven SHM, several studies have put their focus on this field e.g. [5, 9, 149, 169, 173]. As smart sensor networks become more affordable

and are a standard in many structures such as wind turbines, offshore platforms, etc., data-driven techniques become a promising approach to evaluate large amount of data in a timely manner.

1.2.3 Learning frameworks for data-driven damage detection

Generally, statistical and machine learning models can be trained in 3 different ways: supervised, unsupervised and semi-supervised [70]. The main differences between these approaches is the prior knowledge available. Supervised learning aims to learn the relationship between inputs and outputs given a sample of labelled features that a function approximates and maps inputs to outputs. For unsupervised learning, the features are not labelled, the algorithms find patterns in the data which allows the identification of discrepancies. Semi-supervised learning is a combination between supervised and unsupervised which involves partially labelled data for training.

Focusing on damage detection in structures, this can be transferred into prior knowledge about the existence of damage. Assuming a structure is monitored over a period of time, the learning frameworks for supervised, unsupervised and semi-supervised damage detection are introduced and defined as follows:

Supervised framework: Features are extracted from measurements while the structure in operation was undamaged and damaged. The knowledge about the damaged and undamaged structure is used (i.e., the features are labelled "damaged" and "undamaged") when designing a supervised damage detection framework. A framework that has fully labelled features as shown in Fig. 1.7 is considered a supervised framework.

Unsupervised framework: Features from measurements during operation are extracted not knowing if or when the structure is healthy or damaged. No information about

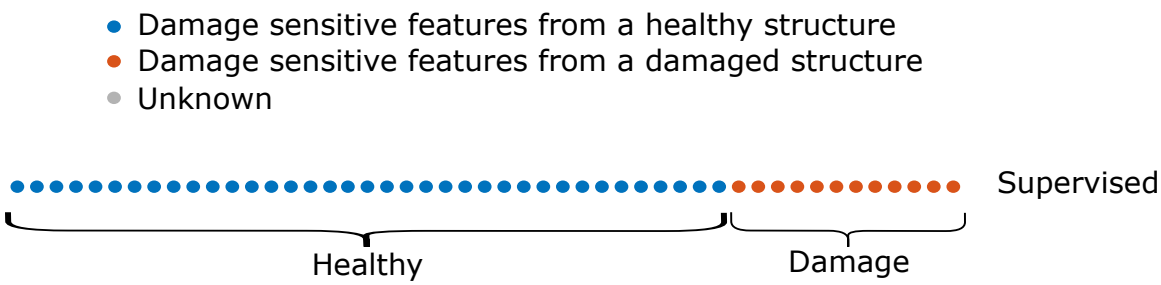


Figure 1.7: Supervised learning framework for damage detection. Each dot represents a DSF extracted from the measurements

the measured data with regards to the structural integrity is provided, i.e., no labels are provided as can be seen in Fig. 1.8.

Figure 1.8: Unsupervised learning framework for damage detection. Each dot represents a DSF extracted from the measurements.

Semi-supervised framework: Only measurements from a healthy structure are available (or what is considered to be healthy). When designing a semi-supervised framework, the knowledge about the healthy structure is used as a reference. Supervised modelling algorithms can be used in a semi-supervised framework as long as the labelled data contains only knowledge about the healthy structure as seen in Fig. 1.9.

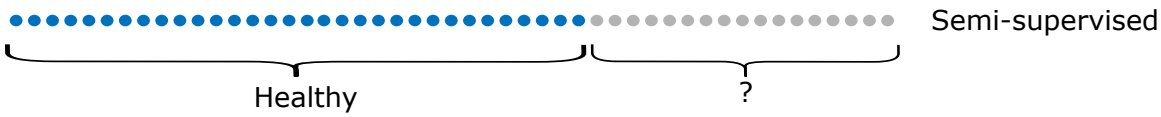


Figure 1.9: Semi-supervised learning framework for damage detection. Each dot represents a DSF extracted from the measurements.

It is worth highlighting that, e.g., a semi-supervised framework can adopt supervised

algorithms such as artificial neural networks as long as the knowledge of the damaged state is not used to train the algorithm to detect damage. In practice, recordings from a damaged structure are scarce, thus, making an unsupervised framework easier to implement as less information regarding the structural integrity is required [24]. Also, unsupervised frameworks are rarely applied for damage detection given that neither knowledge of a healthy nor damaged structure is available to train the algorithm. Commonly, an operator would like to prevent damage in a structure or detect it when it happens through continuous monitoring. In this situation, knowledge from the healthy state of the structure is available and the data can be used to create a reference state and train the algorithm to detect damage. This suggests that a semi-supervised framework is the most appropriate learning framework for online (continuous) damage detection. Nonetheless, challenges remain in the adoption of this learning framework. These challenges will be discussed in more detail in the next chapter.

1.2.4 Control charts

Control charts, also known as Shewhart charts (after Walter A. Shewhart) [138] or process-behaviour charts, are a statistical process control tool used to visualize variabilities in a process [40]. This can be done through monitoring one or multiple parameters in a process that have been summarised into a one-dimensional plot. In one of its simplest forms, this monitoring can be done by using the standard deviation and means of samples. By comparing the mean and standard deviation of a sample with those of the whole process, deviations can be identified and classified. The identified deviations can be utilised as warning signals.

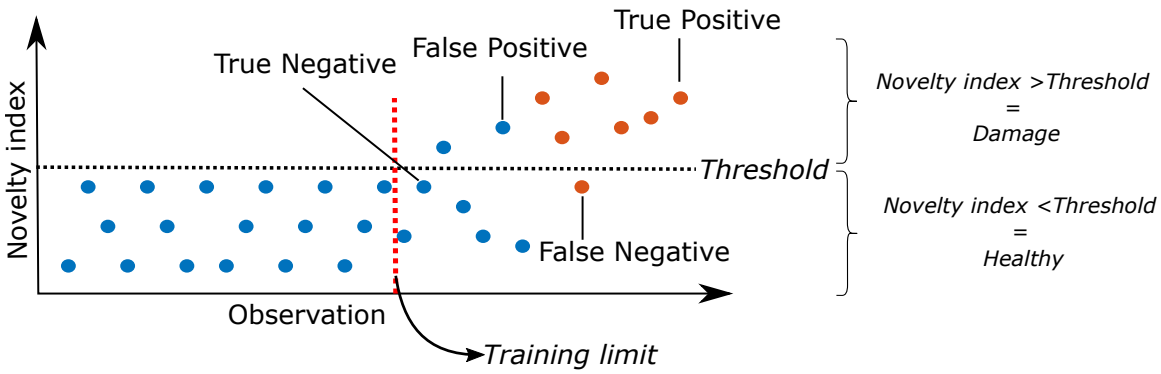


Figure 1.10: Data partitioning for an output-only damage detection methodology

For damage detection, a damage index (DI) (also known as a novelty index) is calculated. The DI can summarise the DSFs which can be plotted as a control chart as shown in Figure 1.10. A threshold is defined which allows to cluster data into a "healthy" and "damage" states. Correctly clustered data is referred to as true positive (damage) or true negative (healthy). Observations which are misclassified are referred to as false positive (incorrectly identified as damage) or false negative (incorrectly identified as healthy). The threshold can be defined through a distribution of the data and how much false positives (or percentage out of distribution) are acceptable for the operator. This often depends on the application and the implications of a false positive. The presence of false positives and false negatives are often caused by the influence of environmental and operational variabilities.

1.3 The challenge of EOVs in data-driven VSHM

The dynamic responses of a structure measured through a monitoring system can differ due to the influence of environmental and operational variabilities (EOVs). This represents one of the prevailing challenges in damage detection. It is crucial to discriminate between structural damage and EOVs as the latter can camouflage the

existence of damage or falsely identify damage when it does not exist.

Structures are generally subjected to multiple EOVs. Identifying these variabilities and assessing their influence becomes a critical task in damage detection. Common sources of these variabilities include environmental conditions such as ambient temperature, wind conditions, and moisture and humidity levels [49]. For instance, variations in temperature can affect virtually all material properties including elastic modulus, yield stress and mass density and result in changes in the geometry of a structure [49].

Farrar et al. [46] studied the influence of environmental variabilities such as thermal effects and operational variabilities such as vehicle weight and excitation source during a damage detection process in the Alamosa Canyon Bridge. On one hand, the authors performed modal tests and gathered temperature measurements at various locations on the bridge over a 24-hour period to examine the variabilities caused by thermal effects. They found that the modal frequencies vary approx. 5% during the examination period and these changes were related to the temperature differentials observed across the bridge deck. On the other hand, the authors carried out tests to identify differences in the modal parameters when using various excitation sources including roving hammer impact excitation, single point impact excitation, ambient excitation from traffic on an adjacent bridge, ambient excitation from a single car driven over the bridge, and from electrodynamic shakers. The authors found that it was challenging to identify the differences in the modal parameters measured during the tests from the variabilities caused by environmental effects, however, they were able to determine that there were statistically significant differences in the damping that were correlated to the excitation amplitude.

Similarly, Figueiredo et al. [56] found that the same bridge's frequencies and thermal environment changed over a 24-hour cycle in a study carried out nearly a decade

later. Ultimately, these studies highlighted the importance of quantifying the effects of EOVs on the modal-based parameters they used in the damage identification process. The issue of EOVs is also referred to as data normalisation and has become a key concern when implementing SHM systems for damage detection [48].

1.4 Conclusion

With the rapid development of the wind energy sector and its key role in the transition to low-carbon energy systems, reliability and cost competitiveness become crucial challenges that must be addressed to aid and strengthen this energy sector. As one of the main sources of accidents and failures, monitoring the integrity of WTBs is of particular interest. This chapter has described the potential synergy between VSHM systems and WTBs.

VSHM systems have already been widely adopted in other industries including aerospace and offshore oil & gas. In the wind energy industry, there is great interest in these systems. However, their implementation on WTBs has been often restricted to lab-scale experiments due to challenges such as instrumentation and high upfront costs. This thesis explores the adoption of a VSHM framework for damage detection in operating WTBs that addresses the challenge that EOVs pose to the successful detection of damage through a semi-supervised learning framework.

1.5 Outline of the thesis

The remaining chapters of this thesis are structured as follows:

- Chapter 2 provides an overview of the literature concerning approaches for damage sensitive feature extraction and data normalisation. An assessment of the strengths and limitations of these approaches is provided. Based on this assessment, the problem statement and research objectives for this thesis are defined.

- Chapter 3 introduces a semi-supervised damage detection framework based on artificial neural networks. This framework is tested on artificial data, numerical simulations and a lab scale experiment. Finally, a comparison between the proposed framework and state-of-the-art approaches such as PCA and cointegration is conducted.

- Chapter 4 introduces an interpretable novelty detection framework based on Gradient Boosted Decision Trees and Shapley Additive Explanation. This framework is implemented to interpret the widely used Mahalanobis distance-based novelty index. The framework is tested on a numerical simulation.

- In chapter 5, the proposed novelty detection framework introduced in chapter 3 and the interpretability framework introduced in chapter 4 are evaluated on an operating wind turbine blade with artificial damage. The strengths and limitations of these frameworks are discussed and a combined approach is presented.

- Finally, chapter 6 summarises the conclusions and recommendations derived from the research presented in this thesis. Additionally opportunities for further work based on the challenges and limitations encountered throughout the research are presented.

Chapter 2

Literature review and research objectives

This chapter provides an overview of available literature that discusses the prevailing challenge of EOVs in data-driven damage detection. EOVs can camouflage damage or falsely hint towards it. VSHM systems in WTBs can be hindered by multiple EOVs such as temperature, wind conditions, humidity, among others. Hence, addressing the influence of EOVs can improve reliability in the wind energy sector. The literature review discusses the main stages involved in a framework to mitigate EOVs in damage detection, namely the extraction of DSFs such as modal and non-modal parameters; the use of data normalisation techniques to mitigate the effect of EOVs in data-driven damage detection and the construction of a novelty index that facilitates damage detection and decision-making. Table 2.1 summarises the literature that has been reviewed in this dissertation within each stage. This is not intended to be a thorough review of the existing literature on the damage detection stages but to provide an overview of state-of-the-art methods and techniques that are relevant to the work presented in this dissertation.

Table 2.1
Summary of literature reviewed in this dissertation organised by stages of the damage detection and techniques

Stage of Damage Detection Framework	Technique	Reference
Damage Sensitive Feature Extraction	Natural frequencies	[19, 25, 67, 109, 124, 148, 174]
	Modal damping	[19, 25, 67, 90, 113]
	Mode shapes	[13, 67]
	AR parameters	[53, 55, 59, 87, 121, 124, 142, 148]
	CWT	[4, 7, 106, 134, 158]
	Empirical vibration parameters	[39, 64, 110, 152]
Data Normalisation	ARIMA, ARX , ARMX	[84, 85, 108, 125, 151]
	GMM	[4]
	Cointegration	[34, 127, 139, 155]
	Gaussian process regression	[139]
	PCA	[114, 126, 143, 144, 156, 157, 162, 166]
	MD	[164]
	AANN	[41, 71, 162]
	ANN	[12, 43, 69, 72, 75, 85, 85, 104, 115, 133, 165]
	LASTM	[30]
	K-means	[144]
Novelty Index	Euclidean Distance	[71, 156]
	Mahalanobis Distance	[4, 25, 152, 170]
	t-squared statistic	[31]

Additionally, given the recent uptake in the use of machine learning techniques for damage detection, their lack of interpretability is highlighted and discussed. Shapley

Additive exPlanations (SHAP) is investigated as a tool to add interpretability to so-called black-box models. Ultimately, the gaps and limitations found in the literature are selected as the research objectives of this thesis.

2.1 Damage sensitive feature extraction

The extraction of DSFs is a key step in the vibration-based damage detection process. This has become an area of research that has attracted significant attention [60, 95, 167, 172]. The ideal DSF should be sensitive to damage whilst being robust to all EOVs. Conceptually, DSFs condense a large volume of raw data, however, the extraction of these features should require low computational effort, particularly when considering wireless sensing networks and online monitoring systems. Given all these considerations, it is difficult to find a single feature that has all these properties and, at the same time, can be applied to different types of structures to detect damage under the influence of different EOVs [171]. Over the years, several techniques have been adopted to extract DSFs. These techniques can be divided into modal-based and non-modal-based. The most commonly used techniques will be briefly discussed in the coming sections.

2.1.1 Modal parameters

Commonly, early studies on VSHM were based on the premise that changes in a structure can be reflected in changes in natural frequency, mode shape, or modal damping [20]. These modal parameters have been widely adopted as DSFs. One of the first implementation of such DSFs dates back to 1975, when Vandiver [159] used changes in natural frequency to detect structural failures in an offshore platform.

Since then, numerous studies have adopted such DSFs in damage detection processes [13, 19, 25, 67, 90, 109].

However, modal parameters can be relatively insensitive to damage [39]. Particularly EOVs and changes in the material properties due to, e.g., aging can change the means, variances and covariances of vibration signals over time, thus violating the assumption of linear stationarity which are characteristic of standard modal techniques based on Fourier transforms [39]. Talebinejad et al. [150] compared modal-based damage detection techniques and found that only high intensity damages were detectable through the use of parameters such as natural frequencies, mode shapes and curvature. Furthermore, their study concluded that the selected DSFs and damage detection techniques were sensitive to noise contamination, which led to the identification of numerous false positives. Finally, the results suggest that the first modes, which are generally easily extracted, are less sensitive to damage and most influenced by EOVs whilst the higher modes show more sensitivity to damage yet have higher uncertainty when identifying them.

Similarly, Larsen et al. [92] found that the natural frequencies of the first few modes of a WTB show negligible changes after introducing significant artificial damage in the trailing edge. The authors found that mode shapes contained information that may help to detect and localise damage in the WTB when decomposed into flapwise, edgewise and torsional components. However, these conclusions were only valid after introducing 120 cm damage in the trailing edge and in an experimental setup (i.e., under negligible influence of EOVs).

Techniques and methods based on modal parameters or detailed models of a structure for damage detection have been effective. However, they face particular issues including the challenge of calibrating the models for non-linear and inelastic behaviour, the sensitivity of the parameters to EOVs, among others. Previous studies have shown

that damage, which is a local phenomenon, is difficult to identify through global properties of the structure such as modal frequencies [47]. For accurate damage detection through modal-based parameters, very precise measurements or large levels of damage are required. The local nature of damage requires registering the structure's higher frequency modes, which is a more challenging process in practice and, thus, limits the applicability of these techniques for online monitoring systems.

2.1.2 Non-modal parameters

To address the issues of stationarity, linearity and sensitivity to EOVs, several studies have proposed the use of alternative DSFs to modal parameters for damage detection which are referred to as non-modal parameters [21]. Some examples will be reviewed in this section.

To deal with the challenges of linearity and stationarity, Delgadillo and Casas [39] propose the use of empirical vibration parameters as DSFs for bridge damage identification including cumulative absolute velocity, cumulative absolute displacement, distributed vibration intensity, among others. The proposed DSFs were suitable for damage identification (detection, localisation and quantification), however, this study did not account for the effect of EOVs on the vibration parameters of the bridge examined.

Commonly used non-modal parameters adopted in damage detection studies include coefficients of an autoregressive (AR) model, continuous wavelet transform (CWT), and statistical parameters such as covariance. A brief overview of some of these studies is provided next.

AR models have been widely used as techniques to extract DSFs under the approach

of time series analysis [53, 59, 87, 142]. These models can be used in two ways to extract DSFs, namely using the residual errors of the model or directly using the coefficients of the AR model. AR models or specifically the AR moving average (ARMA) are also utilised to estimate modal parameters, where ambient vibration response of a system can be described by an AR model. Ultimately, the identification of the AR parameters are related to the eigenfrequencies of the system [124, 148]. Similarly, the CWT can be used as an abstract transformation to extract DSFs or an aimed analysis to estimate modal parameters such as eigenfrequencies and modal damping [113, 134].

Figueiredo et al. [55] constructed an AR model and use both the residual errors and the model's parameters (or coefficients) as DSFs to be able to distinguish the variations in the features that are caused by damage from those caused by EOVs. The authors used a lab-scale three-story frame structure excited with an electrodynamic shaker. A column suspended from the top floor of the structure was used to simulate nonlinear damage behaviour when it contacts a bumper mounted on the lower floor. EOVs were also simulated by replicating the variabilities observed in the real world through changes in mass and stiffness of the structure. Both the AR residual errors and the model parameters were able to detect damage even under the effect of EOVs and, thus, AR modelling was identified as an effective feature extraction technique.

Mollineaux et al. [106] sought to detect damage in a 34-meter WTB subjected to various configurations simulating different degrees of damage. In their approach, the authors use the AR coefficients as DSFs as was done by Figueiredo et al. [55]. However, Mollineaux et al. [106] also demonstrate the use of wavelet transforms of the input and output signals to address the non-stationary excitation, which aimed to simulate operational wind excitation. The continuous wavelet transform (CWT) has been also widely used as a feature extraction technique in other damage detection

applications [7, 158]. Albeit in an experimental setup, Mollineaux et al. [106] demonstrated an approach to detect damage on a full-scale WTB under non-stationary excitation while using low-power wireless accelerometers.

A feasibility study using AR parameters obtained from a single vibration sensor on an operating 12m WTB with artificial damage was performed by Panagiotopoulos et al. [121]. 15cm, 30cm and 45cm delaminations in the trailing edge were artificially introduced to simulate damage. The damage detection performance was assessed via Receiver operating characteristic (ROC) curves[1] and the results showed low accuracy for all three damage scenarios. By further implementing PCA, the authors dealt with the influence of EOVs and improve the performance of the damage detection process, obtaining adequate performance (64% true positive rate for 5% false positive rate) for the 15cm damage scenario, very good (100% true positive rate for 2.7% false positive rate) for the 30cm damage scenario and excellent (100% true positive rate for 0% false positive rate) for the 45cm damage scenario.

As opposed to modal-based parameters, non-modal or data-driven parameters do not require detailed analytical models such as Finite Element (FE) models. With non-modal parameters, naturally obtained, random dynamic responses accounting for inherent uncertainties such as EOVs can be treated through statistical tools to effectively detect damage [54]. As discussed earlier, non-modal parameters can be used in an abstract way and are also able estimate modal parameters through e.g. the CWT or an AR model. A substantial benefit of this approach is that the detection can take place without interrupting the normal operation of the structure and, thus, are suitable for continuous monitoring systems. However, this approach requires the construction of a baseline, i.e., the algorithms must be trained to differentiate between the patterns that will arise [54]. Also, this approach still requires the use of additional

[1]ROC curves plot the true positive rate (i.e., probability of detection) vs. the false positive rate (i.e., probability of false alarm) evaluated at different thresholds.

data normalisation techniques, as seen in the review presented thus far, to mitigate the influence of EOVs.

2.2 Data normalisation approaches

Several techniques have been adopted to mitigate the effects of EOVs on vibration-based damage detection processes. Some of these techniques will be briefly discussed in this section along with a review of their applications for damage detection in civil structures and, particularly, wind turbine blades.

Peeters et al. [125] used the Z24 bridge measured responses collected from different damage detection tests performed by Krämer et al. [86] to demonstrate and mitigate the effect of temperature on the eigenfrequencies (used as DSFs). The natural frequencies showed intra-day variations, which are correlated to the recorded temperature variations. The authors observed an increase in the identified natural frequencies when temperature decreased. This effect was amplified when the temperature was below 0°C, suggesting a nonlinear relationship. The authors used a linear single-input, single-output (SISO) autoregressive with exogenous terms (ARX) model to describe the relation between temperature and one eigenfrequency. Based on this methodology, damage is detected when a new measured eigenfrequency lies outside the confidence intervals estimated from the ARX model. The authors were able to successfully detect damage in the Z24 bridge despite the effect of temperature and demonstrated that an ARX model including the thermal dynamics of the bridge performed better than a 'static' regression model, thereby, being capable of mitigating the effects of EOVs.

A different approach is adopted by Moser and Moaveni [108], where the effect of temperature on the natural frequencies measured on the Dowling Hall Footbridge on the campus of Tufts University through continuous monitoring was studied. Moser and Moaveni [108] evaluated several alternative regression models including a linear and a fourth-order polynomial model, a quadratic model, a bilinear model, and an ARX model. Based on quality metrics such as the estimation error variance, the coefficient of determination R-Squared and objective criteria such as the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the fourth-order polynomial was selected. Confidence intervals were estimated for the identified natural frequencies as a function of temperature. The authors proposed using the ratio of observed outliers to the expected rate of outliers (based on the confidence level) as a damage detection index. In this way, if a future natural frequency falls outside the confidence intervals established for the corresponding temperature, it can be concluded that the bridge has experienced structural change and, thus, potential damage.

Tatsis et al. [151] proposed a framework for structural identification, damage detection and localization using SISO ARX models that enable the tracking of the operational variability and capturing the dependency on varying environmental and operational conditions. The proposed framework was applied on two numerical case studies of different complexities for its assessment, namely a spring-mass-damper system with six DOFs and a six-storey shear frame under base excitation. The authors demonstrated that their framework detected and localized damage at early stages. However, their approach was mostly set to treat faults resulting from events such as earthquakes or strong wind gusts. The detection of gradual degradation would require further research. The approach would also require to be subjected to experimental validation.

Other studies have normalised the DSFs used in their damage detection processes

for aircraft structures by means of different regression models including the Vector-dependent Functionally Pooled AutoRegressive with eXogenous excitation (VFP-ARX) model [84] and Gaussian Mixture model (GMM) [128]. GMM has found its use for damage detection in WTBs. Avendano-Valencia et al. [4] adopted a Gaussian Process time-series modelling approach to quantify the effect of EOVs such as wind and temperature on the dynamic properties of frequency vibration responses from an operating WTB subjected to different operating regimes. The authors found that changes in temperature have a subtle effect on the characteristics of the vibration response, whilst operational variations such as changes in the rotor azimuth were the main driver of the variations on the characteristics of the vibration response. The authors studied three cases, namely no separation of operational states and feature normalisation, separation of operational states but no feature normalisation and, finally, separation of operational states and GMM feature normalisation. The results show that the last two cases reduce the variations in the Mahalanobis Distance (MD)-based damage detection index, thus, enabling the distinction of damaged states of the blade, albeit with the presence of false negatives.

Cross et al. [34] introduced the cointegration technique to analyse the non-stationary dynamic and quasi-static responses in VSHM data to deal with the issue of EOVs in a damage detection process. In this approach, the cointegrated residuals of the monitored features have the common trends in the original dataset removed and can then be used as DSFs free of the effect of EOVs. Cointegration is an econometric technique based on the assumption that some linear combination of two non-stationary time-series is stationary and the stationary residuals will describe the long-run dynamic equilibrium of the process [26].

Tomé et al. [155] sought to detect simulated damage on the Corgo Bridge using an approach based on multivariate cointegration analysis and statistical process control. Different damage scenarios were numerically simulated and superposed to data

collected over a 3.5 years long monitoring campaign. The approach was able of suppressing EOVs and long-term effects and detecting different damage intensities in the cable-stayed bridge.

An extended framework based on cointegration and Gaussian process regression (GPR) was tested by Shi et al. [139] on data from the Z24 bridge to account for non-linear relationships between the extracted features. The combination of cointegration and GPR successfully dealt with EOVs in this case study. However, the authors acknowledged the limited capability of GPR to deal with underlying regime switches in structures, e.g., different structural response under cold and warm temperatures.

The cointegration approach has been tested on WTBs as well. Qadri et al. [127] employed the cointegration technique for a vibration-based damage detection process of a WTB under operating conditions. The selected approach was able to detect the 15 cm trailing edge opening introduced to simulate damage in the WTB, thus discriminating between observations with and without implemented damage.

A technique that has become widely used for data normalisation and dimensionality reduction is PCA. In SHM, this technique has been applied to vibration-based DSFs to eliminate the effects of EOVs in damage detection without the need for direct measurements of the environmental parameters. Yan et al. [166] adopted a PCA-based damage detection method that accounts for linear (and weakly non-linear) effects of EOVs on DSFs. In this application, the authors tested the approach on a simulated bridge model exposed to temperature gradients and a wooden bridge lab experiment. In both cases, the natural frequencies served as DSFs and were transformed by PCA. In the numerical example, one out of six principal components was used for damage detection. The results have shown that the simulated temperature field was successfully mitigated and damage was clearly detected. The authors also highlighted that the use of more principal components did not significantly change the results. In

the lab experiment, eight out of nine principal components were used for successful damage detection and mitigation of EOVs. Yan et al. concluded that the PCA-based damage detection method was simple and efficient.

Trendafilova et al. [156] adopted a PCA-based damage detection approach and tested it on a scaled aircraft wing simulated by a finite element model (FEM). Two types of damage were simulated, namely cracks introduced by disbanding a number of elements in the FE model and distributed damage simulated by stiffness reduction in the whole volume of the wing. The authors used PCA to transform the frequency response functions (FRFs) of the simulated vibration responses and reduced their dimensionality. They retained the first two principal component (PC), which account for 92.11% of the variance, and used these as DSFs. For damage detection, the authors implemented the one nearest neighbour classifier on the basis of the Euclidean distance to differentiate between damage and no-damage. The methodology was proven to be suitable for structural damage detection, albeit limited to the provided numerical experiments. Ulriksen et al. [157] explored a similar PCA-based approach to detect damage in an operating WTB highlighting the applicability and suitability of such a methodology for practical applications.

More recently, Pereira et al. [126] proposed a combined application of weighted regression models and PCA to detect damage on the Baixo Sabor arch dam in Portugal. A numerical model of the dam and three years worth of monitoring data (including environmental and operational conditions) were used to simulate different types of damage in the dam. PCA was directly applied to the natural frequencies and produce control charts that facilitated damage detection. However, the authors found that the use of combined models such as multiple linear regressions and PCA were most efficient to detect damage simulated by stiffness reduction.

Another method used to normalise variabilities in DSFs due to EOVs through outlier

analysis is the Mahalanobis distance (MD) [164]. The MD evaluates the discordance of a single set of DSFs from a particular observation with respect to the rest of DSFs or a set of DSFs [49]. The MD is essentially the Euclidean distance coupled with data normalisation based on the covariance between DSFs. The MD has been used in several studies with various combinations of DSFs such as AR coefficients [28], SSA [63], or in a comparative study with multiple modal assurance criteria (MAC) values, coordinate modal assurance criteria (COMAC) and damping ratios [25]. Often, the MD is used to construct a novelty index for damage detection as seen in Trendafilova et al. [156] and Tcherniak and Mølgaard [152].

The three last approaches discussed - namely MD, PCA and cointegration - have gained popularity as data normalisation techniques to deal with the influence of EOVs in damage detection. Cross et al. [33] conducted a comparative study exploring the potential and limitations of the three approaches using benchmark data from the Brite-Euram project DAMASCOS, collected from a Lamb-wave inspection of a composite panel which was subjected to temperature variations. In the study, after selecting features that were insensitive to environmental conditions, the authors found that outlier analysis, i.e., MD, was able to detect damage yet warned the reader about a potential trade-off between insensitivity to environmental conditions and sensitivity to damage in the features selection process. If feature selection is omitted, a significant amount of false positives is present. The authors found that both PCA and cointegration enable the creation of environmentally insensitive features and the clear detection of damage. In the case of PCA, the selection of PCs is crucial for successful damage detection. This study found that the minor components are sensitive to damage whilst insensitive to temperature variations. However, the choice of components is case-specific and requires careful consideration as this can significantly alter the outcome of the damage detection process. Unlike PCA, in the case of cointegration, the selection of the most stationary cointegrating vectors can be carried out by means

of the Johansen procedure [8]. Nevertheless, the study showed that additional considerations were necessary to improve damage detection and avoid anomalies, thereby ensuring the stationarity of the residual during the normal response of a structure. Such considerations include either the use of a subset of the original variable set selected or a linear combination of cointegrating vectors not assessed by the Johansen procedure.

These three approaches have underlying assumptions that may limit their applicability and generalisation for real structures. The MD implicitly assumes a Gaussian distribution of the normal condition data. PCA and cointegration assume linearity in the data set. While these assumptions can often be a good approximation of real-world data, the random variability and influence of simultaneous EOVs encountered in real applications are often not adequately represented when using experimental or numerical examples as also highlighted by Cross et al. [33].

To avoid limiting assumptions about the distribution of the data, studies have explored the use of non-traditional techniques for damage detection and mitigation of EOVs [6, 56]. These techniques include machine learning algorithms such as Artificial Neural Networks (ANN), eXtreme Gradient Boosting Decision Trees (XGBoostDT), Support Vector Machines (SVM), among others. Particularly, ANNs served in several methodologies as a powerful tool for pattern recognition, specifically to learn the patterns of EOVs. Two main types of neural networks (NNs) architecture can be found in SHM applications: auto-associative neural network (AANN) and conventional ANNs for regression or classification problems. AANNs is a particular network architecture that is trained to approximate the mapping between a network inputs and outputs which have equivalent amount of neurons. In this way, the DSFs are reconstructed and a pattern can be identified. The AANNs are trained with a subset of healthy DSFs and the reconstruction error of new observations serves as novelty index which is expected to increase if a structure is damaged. The AANN has been

applied successfully in several studies. AANN is an unsupervised learning algorithm, yet its implementation when considering data from a healthy structure for training falls into the semi-supervised learning frameworks described in section 1.2.3

Gu et al. [71] proposed a two-step procedure, where, first, auto-associative neural networks (AANN) are used to reconstruct modal parameters and, second, the novelty index is calculated by the Euclidean norm to highlight the reconstruction error. The methodology was tested on a finite element model of a simple supported steel beam girder which was subjected to uniformly distributed temperature variations. The temperature varied from -15°C to 50°C and non-linear temperature dependent material properties were considered. The authors have shown that the AANN were capable of identifying the non-linear temperature variations in the DSFs and differentiate structural damage from temperature variations for different levels of noise. Several studies have highlighted the potential of AANN in various combinations for damage detection, e.g., in combination with PCA [114, 143] or directly with modal frequencies [174].

AANN can also contribute to damage detection in WTBs as shown by Dervilis et al. [41]. Vibration measurements were obtained from a 9m blade exposed to approx. 8.5 million loads cycles which created a visible crack in the WTB. Due to the laboratory environment, environmental conditions were rather stable. Nonlinear Principal Component Analysis (NLPCA) is applied to the FRF in order to transform the raw measurements into a lower dimensional representation and as a feature selection technique ultimately retaining 96% of the variance explained (10 PCs). FRF in combination with NLPCA and AANNs have shown to be applicable for damage detection in WTBs. Additionally, the Radial Basis Function (RBF) AANN adopted was found to be equivalent in performance to a 5-layer AANN yet leading to fast and efficient runs. Finally, the authors highlight the importance of the feature selection process and suggest further discussions on this subject.

A successful monitoring approach to identify impending blade breakages using AANNs was proposed by Wang et al. [162]. The authors used supervisory control and data acquisition (SCADA) data collected from a wind farm with 33 turbines, where 2 turbines experienced blade breakages. 10-min summary statistics of the SCADA data collected prior to the breakages were used to train a 5-layer AANNs. The bottleneck layer of the AANNs architecture was set to 10 neurons based on a PCA analysis which revealed that the first 10 PCs accounted for 90% of variance. The Euclidean norm of the reconstruction error served as novelty index in a control chart to detect the damage. The results show that the breakages were successfully identified. Furthermore, false alarms were not observed. Nevertheless, the authors acknowledged that the proposed approach needs to be further validated.

ANNs have mainly been applied for classification and regression problems. For examples, Kostić and Gül [85] employed an ANN regression model in a proposed framework for damage detection and localisation. For this, a FEM of a bridge is used with 2000 simulations influenced by a temperature field. The DSFs were obtained from the vibration response by an autoregressive moving average model with exogenous inputs (ARMAX). To mitigate the temperature from the DSFs, Kostić and Gül used ANNs to build a relationship between the simulated temperature field and the DSFs. The difference in prediction were used as new DSFs which were free from EOVs, yet were capable to detect the damage in the structure. The results showed that the proposed methodology was able to determine the presence, location and, furthermore, the severity of the damage. Due to the promising results in mitigating EOVs, the authors discussed further expansion of numerical simulation with different sources of EOVs and implementation in real structures. Multi-layer feedforward ANNs have been subjected to numerous studies in the past decade showing successful pattern recognition for modal parameters-based DSFs [12, 69, 72, 104, 115] and non-modal-based DSFs [43, 75, 133, 165].

In the field of wind energy, ML techniques have been researched and tested in field applications such as fatigue analysis of a WT tower [35, 96, 112], fatigue analysis of a WTB [89, 160] or predicting the power production in a wind farm [1, 154]. More recently, Choe et al. [30] proposed and tested a supervised damage detection framework for floating offshore wind turbine (FOWT) blades. For this, a numerical model is developed for the FOWT with a capacity of 5 MW and 61.5m blades including non-linear EOVs such as wind inflow, blade pitch, generator torque and yaw. The numerical blade model was divided into 10 sections, where damage was introduced by reducing the bending stiffness and the mass of the damaged materials ranging from 5% to 30% in several locations. A total of 12000 simulations were performed with different damage locations and intensity and 1200 simulations of an undamaged blade. To identify the pattern of different damage scenarios and an undamaged blade, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks were implemented. For this, the vibration response was wavelet transformed and was treated as an input for the LSTM and the different damage scenarios as prediction targets. The authors tested four different LSTM networks with an overall accuracy of 91.7% and up to 100% in special cases. The study has shown the applicability of LSTM NN and their application for highly non-linear EOVs in damage detection.

Apart from ANNs other ML techniques were successfully applied in SHM. Solimine et al. [144] used PCA and K-means clustering to identify outliers, i.e., deviations from the normal operation of a full-scale WTB, through the collection of audio signals from the blade cavity. Their method successfully detected structural and acoustic abnormalities when the WTB underwent fatigue testing. Furthermore, the authors concluded that their method facilitated the ability to distinguish between damage-related acoustic aberrations and EOVs.

Recently, some studies have adopted the population-based SHM methodology to address the sparsity of data in SHM and overcome the need for large sets of data required

before the start of a monitoring campaign to train ML algorithms [15]. This methodology considers the available data of structures from both physical and numerical systems, thereby expanding the number of observations available for diagnosis and classification of damage on a new structure at the start of its monitoring campaign and facilitating the transfer of knowledge [15]. Gardner et al. [65] demonstrate the applicability of this methodology for transferring localisation labels from a Gnat aircraft wing to an unlabelled Piper Tomahawk aircraft wing dataset in combination with multiple domain adaptation algorithms such as balanced distribution adaptation and a k-nearest neighbour classifier. This novel approach overcame the need for damage observations from the Piper Tomahawk wing. Based on this evidence, it can be said that the ML techniques in combination with transfer learning frameworks such as the population-based SHM methodology could enable a pathway for industrial applications of online monitoring systems for damage detection in structures.

Most civil structures are not equipped with an SHM system for continuous monitoring, therefore, data from real damage in a structure is limited. As such, the use of data-driven techniques and frameworks for damage detection including ML techniques are often limited to numerical simulations or lab-scale experiments. Commercially available SCADA systems in a WT do not include a monitoring system for WTBs, making it particularly hard to validate such damage detection frameworks on operating WTB. Nevertheless, ML techniques have shown in the various studies reviewed a potential for damage detection. Especially WTBs, which operate often under extreme and harsh environmental conditions, can benefit from these techniques.

As mentioned in section 1.2.4, control charts can be laid out to ease decision-making by constructing a one-dimensional novelty index. The novelty index can be a direct result from the data normalisation when, e.g., PCA is implemented and the dimensional space is reduced to the first principal component. In other cases, when implementing, e.g., ANN, a prediction error would need to be calculated to serve as

a one-dimensional novelty index [110]. Finally, when directly mitigating the EOVs in the feature space, the dimensionality may remain unchanged and a common way to summarise this information is the construction of a novelty index based on a distance metric such as the Euclidean distance [71], the Mahalanobis Distance [170] or the T-squared-Statistic [31]

2.3 Black-box models and their interpretability

ML techniques have been successfully implemented in experimental and some real damage detection applications. Nevertheless, the ML algorithms adopted in these studies for SHM and VSHM frameworks are often referred to as black-box models and they provide little to no information about the decision-making process. The lack of interpretability and understanding regarding how these models detect changes in structural integrity and how DSFs are used by these models to make a certain prediction is one of the reasons hindering the widespread adoption of these techniques for industrial applications. From Fig. 2.1, it can be seen that ANNs are a good example for models with high accuracy but rather low interpretability. Commonly, highly accurate models build a non-linear relationship and require long computational time. On the other hand, highly interpretable models often assume a linear and smooth relationship and are rather easy to compute.

Enabling interpretability in ML-based VSHM frameworks can help building evidence for reliable decision-making in SHM applications. In particular, differentiating between novelties caused due to the influence of EOVs or damage in a structure can help reducing false alarms.

The issue of interpretability when using ML techniques is not new and has been

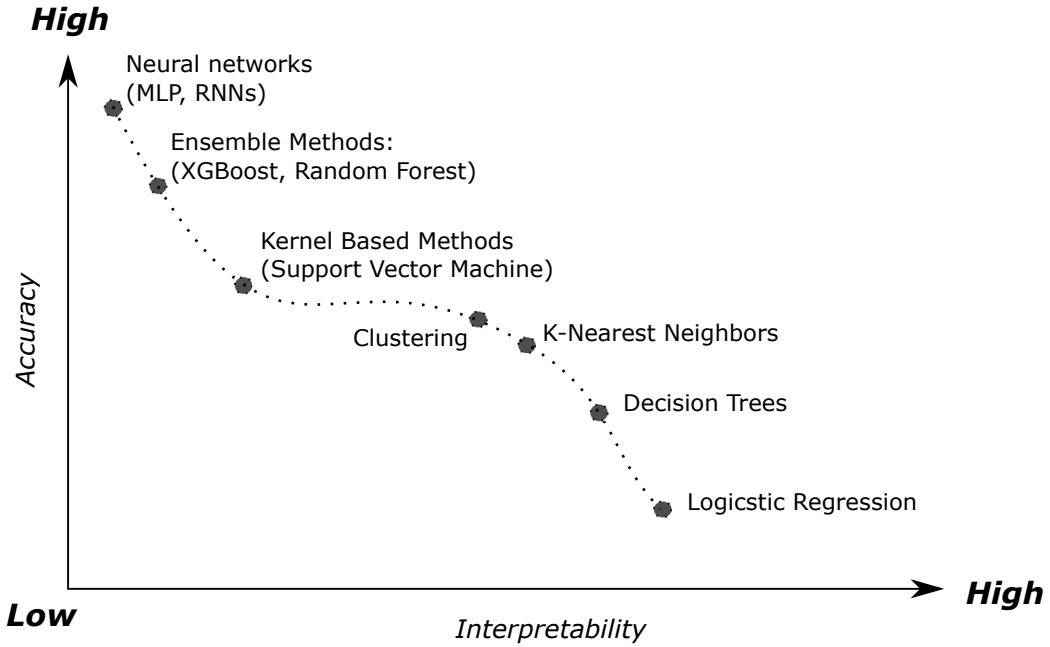


Figure 2.1: The trade-off between accuracy and interpretability when using ML techniques. The hierarchy is rather subjective and can vary with different applications [129].

previously addressed in fields outside engineering [161]. Interpretation of models can be addressed in several ways, one way is to simply use interpretable models such as a linear regression, which often yields a low accuracy but is intuitive for the user. Another way is to use model-agnostic tools, which can be ultimately applied to any supervised learning algorithm. These are often categorised into global and local methods, where global methods describe the overall behaviour of a feature and local methods are able to explain an individual prediction [107].

In particular, the Shapley Additive Explanations (SHAP) approach proposed by Lundberg et al. [99] has lately attracted attention in the research community. The authors adopted a model agnostic representation of feature importance which is estimated by Shapley values [132, 147] and developed an algorithm to estimate the Shapley values in a computationally efficient way. SHAP is an additive feature attribution method that defines the output of a model as the sum of the real values

attributed to each input feature. In other words, the Shapley values represent the marginal contribution of each feature to the prediction of the model. Lundberg et al. [99] have also shown that the Shapley values are the only way to assign feature importance while maintaining two important properties:

- *Local accuracy*: The sum of the feature attributions is equal to the output of the function one is seeking to explain
- *Consistency*: Changes in the model where the impact of a feature increases will never decrease the attribution assigned to that feature.

SHAP has been implemented in fields such as medicine and finance. For example, Lundberg et al. [100] used the SHAP technique to understand the outputs of a ML-based system predicting the risk of hypoxaemia during anaesthesia care from electronically recorded intraoperative data. Similarly, in the field of finance, Bussmann et al. [16] employed the SHAP technique to explain the risks associated to peer to peer lending credit in regulated financial services. The interpretable model enables lending entities to classify potential borrowers into risky and non-risky borrowers based on common financial characteristics identified by SHAP as highly influential including earnings before interest and profit before taxes for non-defaulted companies and total assets and shareholders funds for defaulted companies. In both cases, the added interpretability helped understanding the predictions made by the ML models and highlighted the potential of interpretable ML for reasoning in predictive models.

Interpretable ML techniques such as SHAP have also been adopted in engineering applications, albeit less frequently. For instance, Parsa et al. [123] used SHAP to analyse the causes for traffic accidents predicted by an XGboost classification model. By implementing SHAP, the study found that traffic related features such as the vehicle speed had a relative high impact on the prediction of an accident.

Furthermore, non-traffic related features such as demographic, network, land use and the weather conditions were identified as important predictors for the occurrence of an accident. This information can facilitate policy decisions related to e.g., speed limits, urban planning, among others.

For SHM applications, Lim et al. [94] estimated different damage levels for several bridges based on monitored conditions of their decks using an XGBoost classification model. Their data was retrieved from the Korean Bridge Management System and corresponded to 142,439 deck inspection records of 2388 bridges. For their analysis, the authors collated 53 variables including identification factors such as region, shape, vehicle weight limit, among others; structural factors such as deck material, thickness and strength, etc.; and traffic and inspection factors such as age, damage type and condition rating of the damage. Further, SHAP was adopted to capture the influence of the features used for the classification. Their results showed that major factors influencing damage to bridges include age, average daily truck traffic and vehicle weight limit. Onchis et al. [120] used Local Interpretable Model-agnostic Explanations (LIME) and SHAP in combination to characterise the location and depth of damage in cantilever beams. Through the added interpretability and a novel stability index, the authors provided trust in models prediction. The findings emphasised the benefits of using Shapley values for model interpretability in addition to ML-based techniques and their contribution to the decision-making process for preventive maintenance.

These studies have shown that interpretable ML techniques improved the pattern recognition properties of the predictive models providing supplementary information to select the most adequate predictors (i.e., predictors with the highest contribution). Enabling interpretability in ML-based VSHM frameworks helps building evidence for reliable decision-making in SHM applications.

2.4 Problem statement

Despite significant efforts to mitigate the influence of EOVs in damage detection and the various approaches adopted by the scientific community to face the challenge of EOVs potentially camouflaging damage in a structure, there is no one single approach that can be generalised and pin-pointed as the most accurate and reliable methodology for damage detection.

The approaches examined in the literature review mainly follow the general monitoring framework discussed in section 1.2.1, which implies that feature normalisation follows feature extraction and, finally, a novelty index is calculated for decision-making as shown in Fig. 2.2. When this conventional framework is followed and techniques such as MD, cointegration, PCA and AANN, among others, are adopted, assumptions about the nature of the EOVs or their relationship to the DSFs are made. The selection of the technique to be used for data normalisation must match the characteristics of the EOVs and DSFs or the results may be invalid for the particular problem being examined. For instance, the adoption of cointegration implies that there is collinearity between two features and the adoption of the MD-based approach implies that the features can be normalised by calculating their inverse covariance. In the presence of two features that do not exhibit collinearity, cointegration would provide biased results. These limitations will be further examined and discussed in sections 3.5 and 3.3.

Furthermore, one of the main limitations in damage detection through the use of this conventional framework to mitigate EOVs is the need for normalisation prior to the calculation of the novelty index, which requires unsupervised learning techniques given that the data is neither classified nor labelled at that stage. This directly hampers the accuracy of the damage detection process as the algorithms are not

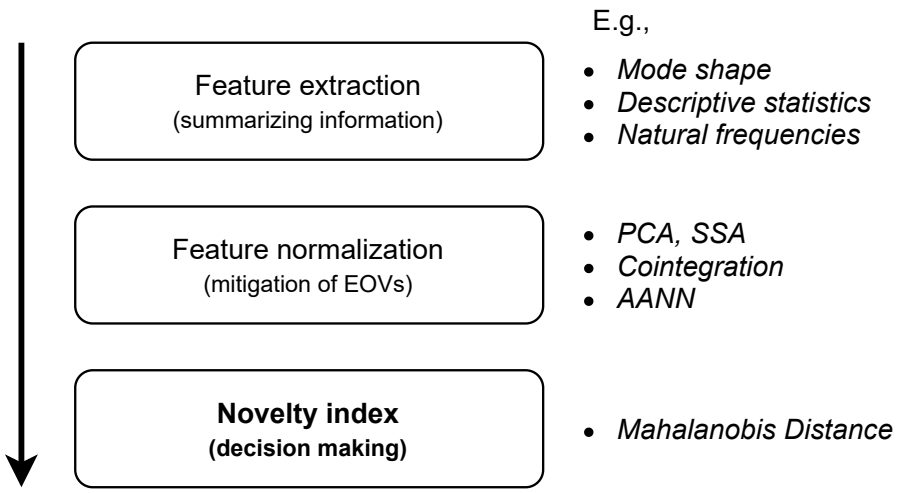


Figure 2.2: Conventional framework to mitigate EOVs in damage detection

guided towards a specific target during the learning process.

The literature review showed that, increasingly, ML techniques are being adopted to improve the accuracy of the conventional damage detection framework. Techniques such as AANN, SVM, among others have gained popularity as tools for feature normalisation through pattern recognition. Nonetheless, there is a lack of consistency when applying ML techniques for damage detection, particularly in terms of the architecture of ANNs which leads to reproducibility issues. The heuristic approach witnessed among the reviewed studies to design an ANN architecture can be limiting to industrial applications, especially in a semi-supervised learning framework, where validation data is scarce given that most studies are based on numerical simulations and lab-scale experiments. This situation undermines the confidence of an operator in ML techniques for damage detection.

Adding to this mistrust is the fact that most ML algorithms adopted in the reviewed studies for SHM or VSHM frameworks are often referred to as black-box models due to the lack of transparency in their decision-making process. The damage detection

process cannot be understood when an operator does not see, e.g., how DSFs contribute to a certain prediction. This lack of interpretability hinders the widespread implementation of ML techniques for industrial damage detection applications, where these could be particularly beneficial.

Ultimately, the replacement of feature normalisation techniques such as cointegration, PCA, etc. by ML techniques has shown promising results; however, this does not address the limitations of the conventional framework and the requirement to normalise the data prior to the calculation of the novelty index. This challenge remains unaddressed and, thus, will be the focus of this thesis.

2.5 Research objectives and contribution to knowledge

The main objective of this dissertation is to:

> **Develop a robust, interpretable, data-driven and semi-supervised damage detection framework that addresses the influence of environmental and operational variabilities in damage detection.**

The novelty of the proposed framework lays in the approach to data normalisation. As opposed to the conventional framework to damage detection, in this dissertation, the novelty index is calculated before the normalisation as can be seen in Fig. 2.3. In this approach, the novelty index showcases the influence of EOVs and their effect is mitigated through post-normalisation (or post-processing).

The developed framework provides transparency and useful knowledge for industrial

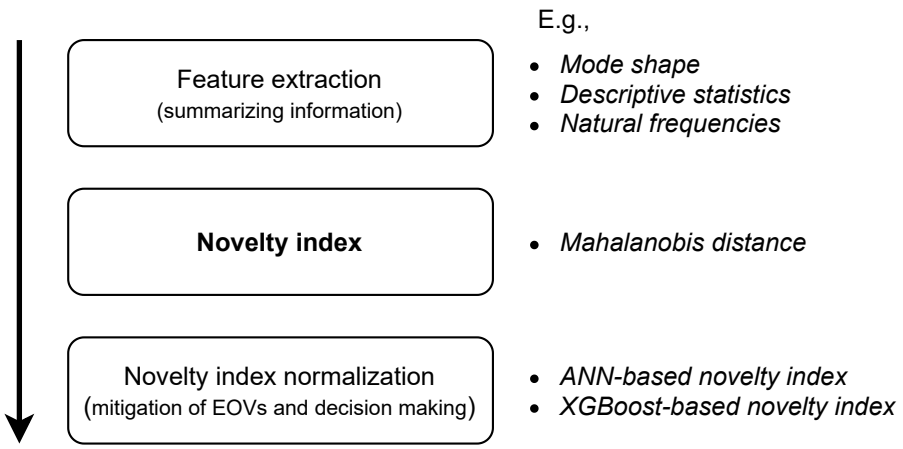


Figure 2.3: Proposed framework to mitigate EOVs in damage detection

VSHM applications, thereby contributing to online damage detection in WTBs. The framework addresses both the challenges of robustness, reproducibility and interpretability of the damage detection process encountered in previous studies.

To meet the general objective of this dissertation and address the research problem defined in section 2.4, specific objectives are defined:

- Examine the limitations of the use of the MD-based novelty index for data normalisation (section 3.3)
- Evaluate the adoption of ANN as a pattern recognition tool for the normalisation of the novelty index. This is achieved by embedding supervised ANNs in a semi-supervised framework to learn the relationship between DSFs and novelty index. (section 3.4)
- Guarantee reproducibility with fixed hyper-parameters and structure of ANN (section 3.4)
- Test the proposed framework on several applications including numerical simulations and laboratory experiments to ensure generalisation and its practical

use. (section 3.5)

- Compare the proposed framework with state-of-the-art techniques such as cointegration and PCA. (section 3.5 and 3.6)
- Introduce XGBoost and Shapley Additive Explanations for interpreting the influence of EOVs on the conventional MD-based novelty index and test this approach in a numerical simulation. (section 4.1)
- Incorporate interpretability into the proposed data driven damage detection framework shown in Fig. 2.3 (section 5.5).
- Implement the data driven and interpretable framework proposed in an operating WTB and evaluate its performance when detecting damage. (section 5.5)
- Assess the limitations of the proposed framework and identify opportunities for further research. (section 6)

Chapter 3

An artificial neural network methodology for damage detection

This chapter begins investigating the limitations that inherently come along the use of the MD-based novelty index for damage detection. To address some of these limitations, a novel ANN-based pattern recognition framework is proposed seeking to mitigate the influence of EOVs in DSFs. The outcome of the proposed framework is a robust novelty index where the influence of variabilities is accounted for. The framework was published in Movsessian et al. [110] and some of the work in this chapter is based on this publication.

This chapter also compares the performance of the proposed framework to state-of-the-art damage detection approaches. First, the ANN-based approach is compared to cointegration using inputs from a numerical simulation. Later, the proposed framework is compared to PCA in an experimental case study. Finally, the benefits and limitations of the proposed framework are discussed.

3.1 Semi-supervised damage detection framework

The suggested ANN-based methodology is introduced within a semi-supervised VSHM framework for statistical pattern recognition similar to the one described in Section 1.2.3. The methodology follows the flow illustrated in Fig. 3.1 and can be divided into the following four steps: (i) DSFs are extracted from the vibration responses collected when monitoring the dynamics of a structure. (ii) A subset of the collected DSFs, which belongs to an undamaged structure, is selected as a reference state and used to compute the MD for the remaining DSFs. The MD serves as a label for the DSFs. (iii) An ANN regression model learns the relationships between the DSFs and their corresponding MD. This model is used to predict the MD for DSFs from new observations. (iv) A prediction error is computed and serves as the new novelty index for damage detection.

Commonly, DSFs are normalised in some way where the distribution of the data is considered a priori, further a novelty index is calculated for damage detection. In the proposed framework, the novelty index is first calculated where variabilities are expected and afterwards the relationship between the DSFs and the calculated novelty index is learned.

The steps of the methodology are described more in detail in the following sections.

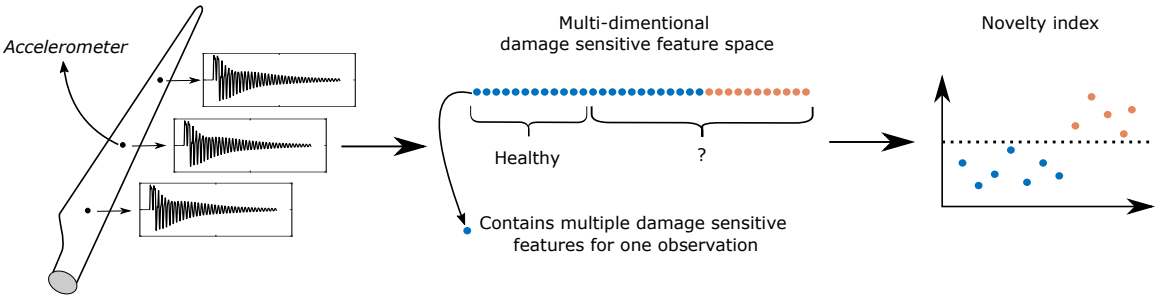


Figure 3.1: Data partitioning for a semi-supervised damage detection framework

3.1.1 Damage sensitive feature space

DSFs, which are application-specific, are obtained from the measured vibration responses and stored in a DSF vector $\boldsymbol{f}^{(k)} \in \mathbb{R}^{q\times 1}$, which contains q DSFs for observation k. The DSF matrix $\boldsymbol{F} \in \mathbb{R}^{q\times r}$ is defined as follows:

$$\boldsymbol{F} = \left[\boldsymbol{f}^{(1)} \quad \boldsymbol{f}^{(2)} \quad \dots \quad \boldsymbol{f}^{(k)} \quad \dots \quad \boldsymbol{f}^{(r)}\right], \tag{3.1}$$

where r is the number of observations. The split of the DSFs for training and testing the damage detection methodology is visualised in Fig. 3.2. The DSFs are split into training $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{q\times u}$ and testing $\boldsymbol{F}_{ts} \in \boldsymbol{F}^{q\times m}$, where u is the number of training and m the number of testing observations. The testing set $\boldsymbol{F}_{ts}$ contains a subset of healthy $\boldsymbol{F}_{hl}$ and damaged $\boldsymbol{F}_{dm}$ observations of the structure. For the output-only methodology, observations from the training set $\boldsymbol{F}_{hl,tr}$, that contain only healthy observations, are available to train the selected damage detection algorithm. The testing set $\boldsymbol{F}_{ts}$ is used to assess the performance of the algorithm by identifying new healthy data $\boldsymbol{F}_{hl}$ and further detecting damage in the structure with observations from $\boldsymbol{F}_{dm}$.

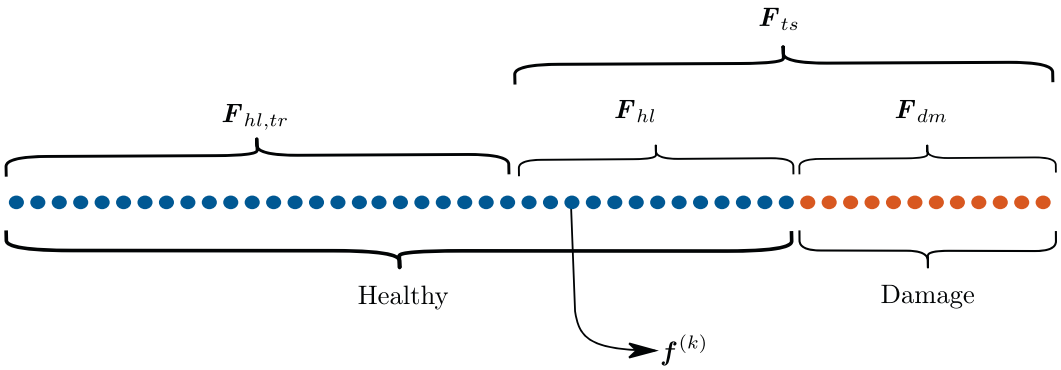


Figure 3.2: Data partitioning for an output-only damage detection methodology

3.1.2 The Mahalanobis distance

The distance between a point and a distribution in a multivariate space can be described by the Mahalanobis Distance (MD). As opposed to the Euclidean distance, the MD normalises the features by the inverse covariance matrix taking into account the features' spread. A common use for the MD is to cluster multivariate data and detect outliers which can be correlated to damage. The MD is calculated using the following equation:

$$d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) = \sqrt{(\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_{\mathcal{F}})^T \boldsymbol{\Sigma}_{\mathcal{F}}^{-1} (\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_{\mathcal{F}})}, \tag{3.2}$$

where $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ represents the MD between the feature vector $\boldsymbol{f}^{(k)}$ and the reference state created by the baseline matrix $\mathcal{F} \in \boldsymbol{F}^{q \times \hat{u}}$ with $\mathcal{F} \subset \boldsymbol{F}_{hl,tr}$ and $\hat{u} < u$. $\boldsymbol{\mu}_{\mathcal{F}}$ is the mean vector of the observations in $\mathcal{F}$ and $\boldsymbol{\Sigma}_{\mathcal{F}}$ is the covariance between the observations in $\mathcal{F}$.

The covariance matrix $\boldsymbol{\Sigma}_{\mathcal{F}}$ is positive semi-definite, therefore, its inverse exists and is also positive semi-definite. The inverse can be presented by the Cholesky decomposition as follows: $\boldsymbol{\Sigma}_{\mathcal{F}}^{-1} = \boldsymbol{U}^T\boldsymbol{U}$, where $\boldsymbol{U}$ is the upper triangular matrix, and the MD can be defined as follows [11]:

$$d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) = \sqrt{(\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_{\mathcal{F}})^T\boldsymbol{U}^T\boldsymbol{U}(\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_{\mathcal{F}})}, \tag{3.3}$$

$$d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) = \sqrt{(\boldsymbol{U}\boldsymbol{f}^{(k)} - \boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}})^T(\boldsymbol{U}\boldsymbol{f}^{(k)} - \boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}})}, \tag{3.4}$$

Thus, the MD is the Euclidean distance between the transformed DSFs space $\boldsymbol{U}\boldsymbol{f}^{(k)}$ and $\boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}}$.

3.1.3 Artificial Neural Networks

An ANN is a mathematical model widely used in the field of artificial intelligence due to its strong pattern recognition capability [136]. In this study, a two-layer feedforward ANN is adopted for the regression model to learn the existing patterns in DSFs which influence the MD-based novelty index. This structure has been used to approximate a nonlinear function in several studies [85, 88, 116]. The ANN structure is defined as:

$$g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W}) = \sigma\left(\sum_{j=1}^{N} w_{kj}^{(2)} h\left(\sum_{i=1}^{d} w_{ji}^{(1)} f_i^k + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right), \tag{3.5}$$

where $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W}) \in \mathbb{R}^{u\times 1}$ is the set of estimated variables and $\boldsymbol{f}^{(k)} \in \mathbb{R}^{d\times 1}$ the DSFs for observation k. The output layer consists of the weights $w_{kj}^{(2)}$ and the bias

$w_{k0}^{(2)}$, consequently $w_{ji}^{(1)}$ and $w_{j0}^{(1)}$ are the weights and the bias for the hidden layer. In the hidden layer, the hyperbolic tangent sigmoid activation function and a linear transfer function $\sigma(\cdot)$ for the output layer are implemented [93]. The least-square solution for training the weight matrix $\boldsymbol{W}$ in Eq. 3.5 is obtained by minimising E_D, which is defined as follows:

$$E_D = \frac{1}{u}\sum_{k=1}^{u}[d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) - g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})]^2, \tag{3.6}$$

where E_D is the mean-squared error between the prediction target $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ and the ANN prediction $g^{(k)}(\boldsymbol{f}^{(k)}, W)$. To avoid a local minimum when minimising Eq. 3.6, Bayesian regularisation is introduced for supervised training. The cost-function E_D in Eq. 3.6 is extended to $S(\boldsymbol{W})$ as defined in MacKay [101]:

$$S(\boldsymbol{W}) = \beta\left(\frac{1}{u}\sum_{k=1}^{u}[d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) - g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})]^2\right) + \alpha E_{\boldsymbol{W}}, \tag{3.7}$$

where $E_{\boldsymbol{W}}$ is the mean-square sum of the weight matrix $\boldsymbol{W}$. In a Bayesian regularisation framework, two objective functions α and β are added to penalise large weights. The ratio α/β between the objective functions is minimised leading the ANN function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ towards generalisation (i.e., avoiding over- and under-fitting).

The structure of the ANN shown in Table 3.1 consists of q input nodes, which refers to the amount of DSFs and 2 nodes in the hidden layer. The simple 2-nodes architecture in the hidden layer is chosen as an additional step to avoid over-fitting of the ANN model and streamline generalisation.

A total number of 2000 iterations is set for learning and generalising the weights in the Bayesian backpropagation. The learning rate is set to 0.005. The training stops

Table 3.1
Artificial neural network architecture and activation functions

Network Architecture	Hidden Layer Activation Function	Output Layer Activation Function
q-2-1	Hyperbolic Tangent Sigmoid Activation Function	Linear transfer function

when the optimal combination of errors and weights is found that minimises the mean squared error. The hyperparameters and the network architecture is kept consistent throughout all case studies. As the weights are randomly assigned before the training and are drawn from a uniform distribution, the results can vary when the training is repeated, therefore, the ANNs are trained 1000 times and the average results are presented with the corresponding standard deviation.

3.1.4 ANN-based novelty index for damage detection

A regression model is built by training the ANN with the DSF vector $\boldsymbol{f}^{(k)}$ and the MD $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$, where $k : [1 < k \leq r]$. For the output-only methodology, the ANN function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ is trained exclusively with DSFs from a healthy structure $\boldsymbol{F}_{hl,tr}$ and their corresponding MD. The proposed novelty index for damage detection is obtained by the prediction error $\hat{d}^{(k)}$ between the predicted and calculated MD. This ANN-based novelty index is defined as follows:

$$\hat{d}^{(k)} = \log \frac{g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W}) + \varepsilon}{d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F}) + \varepsilon} \tag{3.8}$$

where ε is a small scalar to avoid negative infinity due to DSFs that lay in the vicinity of the reference state, in this case $\varepsilon = 10^{-2}$. The ANN mapping function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ is trained with a unique relationship between the DSFs from the healthy structure and its target calculated by the MD metric. If an unlearned relationship between the features and the target occurs, due to, e.g., damage in the structure, an increased

prediction error is expected.

3.2 Evaluation of performance

Once the novelty index for damage detection is calculated, the evaluation criteria need to be defined. For this, three classification metrics are used to evaluate the performance of the novelty index. The first is *Accuracy*, which is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3.9}$$

with $Accuracy : [0 \leq Accuracy \leq 1]$ where true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) represent a ratio of correctly predicted observations to the total number of observations. The accuracy measure is more appropriate for balanced data sets (i.e., equal number of healthy and damaged observations). Therefore, the F_1-score metric is introduced (as it is more suitable for unbalanced data sets). It is defined as follows [77]:

$$F_1 = \frac{2P_{pre} \times R_{rec}}{P_{pre} + R_{rec}} \tag{3.10}$$

with $F_1 : [0 \leq F_1 \leq 1]$, which is formed from the precision P_{pre}:

$$P_{pre} = \frac{TP}{TP + FP} \tag{3.11}$$

and the recall R_{rec}:

$$R_{rec} = \frac{TP}{TP + FN} \tag{3.12}$$

Despite the popularity of the accuracy and F_1-score as measures of performance used in many studies [66, 130, 140], Matthews Correlation Coefficient (MCC) has been included to provide more reliable and truthful scoring when evaluating binary classification as shown by Chicco and Jurman [29]. MCC is the third metric and is defined as:

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{3.13}$$

with $MCC : [-1 \leq MCC \leq 1]$. The maximum values for the three classification metrics are reached for perfect classification.

The cutoff value (or threshold) to select the TP, FP and the corresponding TN and FN is the following [2, 78, 91]:

$$T(t) = \arg\min_{t} \sqrt{(1 - TP(t))^2 + (FP(t))^2}, \tag{3.14}$$

where t ranges over all possible thresholds. The threshold is set, when $T(t)$ value is the minimum.

All three performance metrics are used for binary classification where the MD- and ANN-based damage detection are evaluated based on $\boldsymbol{F}_{hl}$ as healthy and $\boldsymbol{F}_{dm}$ as damaged.

3.3 Artificial data set with 2D feature space

To illustrate the advantage and the limitations of the MD-based novelty index, an artificial 2D feature space is generated with 6000 observations, where $\boldsymbol{F} \in \mathbb{R}^{2\times 6000}$ is the DSF matrix, which is populated by $\boldsymbol{f}^{(k)} \in \boldsymbol{F}^{2\times 1}$ DSFs for 6000 observations. The generated features are referred to as DSFs from a healthy or damaged state. This is for convenience, discussion and illustration purposes and are by no means related to real damage in a structure. As described in section 3.1.1, the data set is split into training data $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{2\times 2000}$ and testing data $\boldsymbol{F}_{ts} \in \boldsymbol{F}^{2\times 4000}$. The testing set $\boldsymbol{F}_{ts}$ consists of $\boldsymbol{F}_{hl} \in \boldsymbol{F}_{ts}^{2\times 2000}$ and $\boldsymbol{F}_{dm} \in \boldsymbol{F}_{ts}^{2\times 2000}$. The partition of the data is summarised in Table 3.2.

Table 3.2
Partition of data for a semi-supervised learning framework: Simulated data set

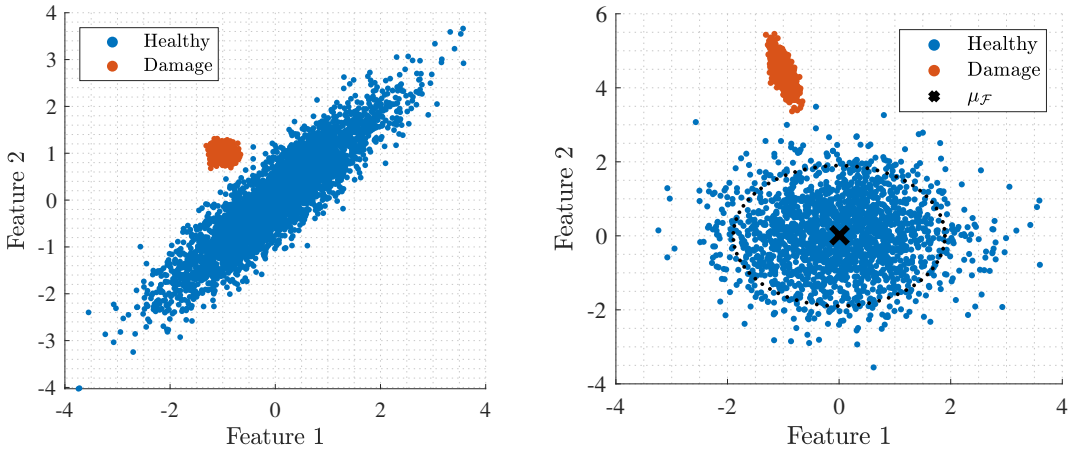
	Healthy		Damage $\boldsymbol{F}_{dm}$	Total $\boldsymbol{F}$
	Training $\boldsymbol{F}_{hl,tr}$	Testing $\boldsymbol{F}_{hl}$		
Observations	2000	2000	2000	6000

To illustrate the effect of variabilities in the DSFs and the capabilities of the MD-based novelty index, two cases are shown.

Case 1

The DSFs from the healthy set are generated by a linear function with added Gaussian noise as shown in Fig. 3.3(a). Observations from the damaged state are placed near the healthy class set and are clearly separated in the 2D feature space.

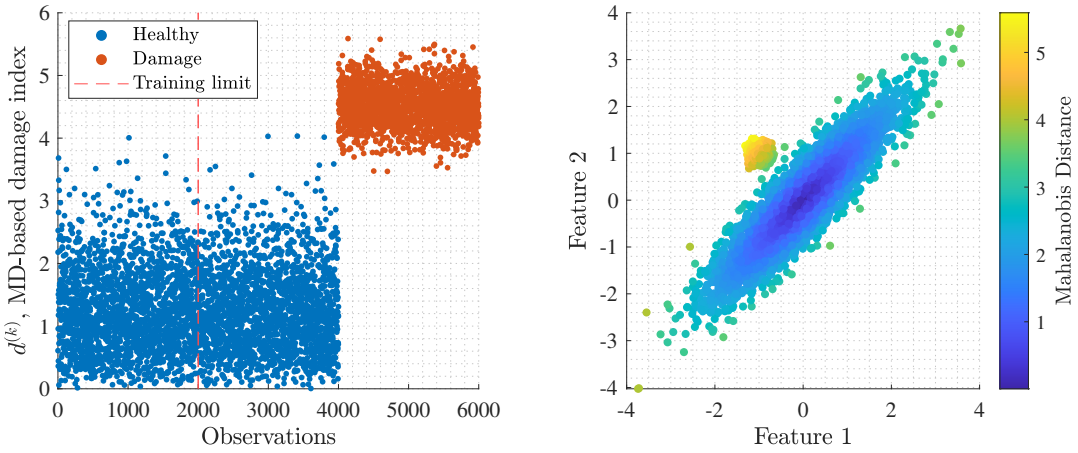
The MD-based novelty index $d^{(k)}$ presented in Eq. 3.2 is calculated with the inverse



(a) Features before normalisation

(b) Features after normalisation

Figure 3.3: Case 1: Artificial example (a) 2D plot of DSFs from a healthy state $\boldsymbol{F}_{hl,tr} \cup \boldsymbol{F}_{hl}$ and simulated damage DSFs $\boldsymbol{F}_{dm}$. (b) Transformed DSFs by $\boldsymbol{U}\boldsymbol{f}^{(k)}$ and $\boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}}$, where the dotted circle is the threshold.



(a) MD-based novelty index of the artificial 2D data set with an elliptic trend

(b) 2D heatmap

Figure 3.4: Case 1: Artificial example - Successful implementation of the MD-based novelty index

covariance matrix $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{2\times 2000}$ by the reference $\boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}}$. The features from the healthy set appear to vary together diagonally, which suggests that the inverse covariance matrix is able to capture this trend as shown in the transformed feature space in Fig. 3.3(b). The MD-based novelty index in Fig 3.4(a) shows the successful damage

detection, where two classes are well separated in a 1D space. The heatmap in Fig. 3.6 highlights how distance increases when features are not within the healthy observations.

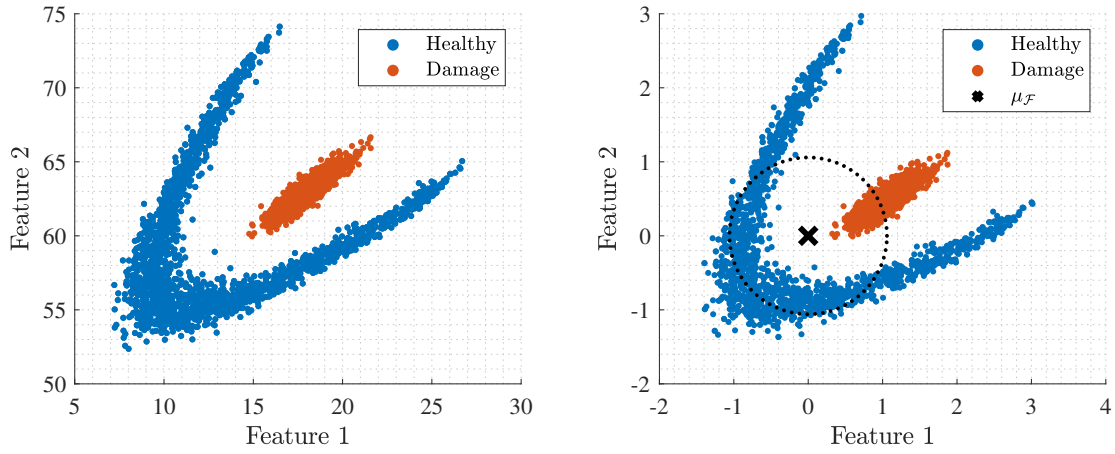
Table 3.3
Case 1: Classification results with $\boldsymbol{F}_{hl}$ as healthy and $\boldsymbol{F}_{dm}$ as damaged with artificial 2D DSFs

Performance metric	MD-based
Accuracy	0.9982
F_1-score	0.9982
MCC	0.9965

The performance matrices confirm the identified damage in novelty detection with an accuracy, F_1-score and MCC close to 1 shown in Table 3.3

Case 2

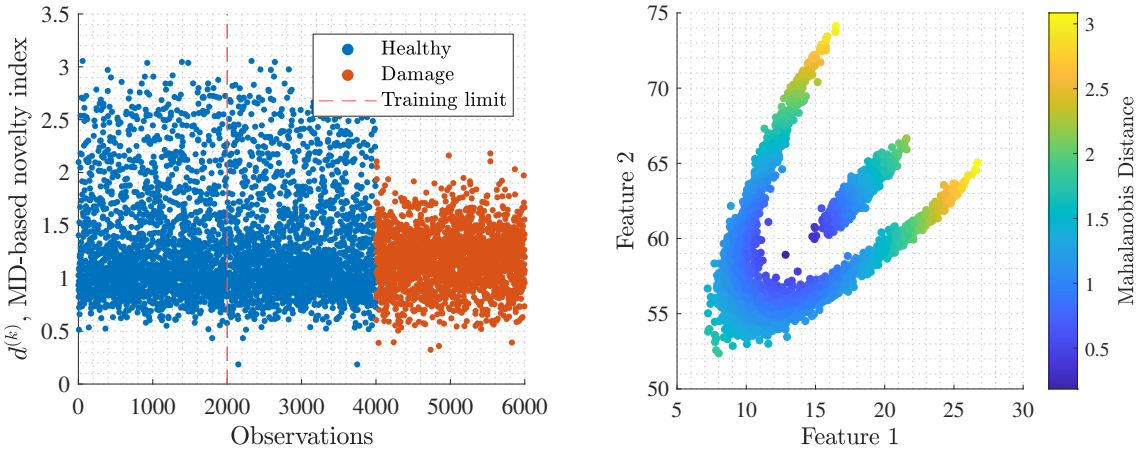
Additional variabilities are introduced in case 2. The data from the healthy set is generated by a quadratic function with added Gaussian noise as shown in Fig. 3.5(a). Observations from the damaged state are within the distribution of the healthy class set but are clearly separated and do not overlap in the 2D feature space. It can be observed that the transformed DSFs in Fig. 3.5(b) have very little difference to the original DSF set in Fig. 3.5(a) as the introduced variabilities are still present. Due to the nonlinear relationship of the variables it is not sufficient to normalize the DSF set by the inverse covariance. The DSFs plotted in a 2D space show a clear separation between the healthy and damaged data sets. Nevertheless, once the Mahalanobis distance is calculated, the information regarding the location of the clusters is eliminated resulting in an overlapping novelty index. All data points on the threshold will have the same distance with different DSFs to the reference $\boldsymbol{\mu}_{\mathcal{F}}$ as shown in Fig. 3.6(a) and the heatmap in Fig. 3.6(b). Discrimination between the two data sets is not possible due to the properties of the MD and the nature of this



(a) Features before normalization

(b) Features after normalization

Figure 3.5: Case 2: (a) Artificial example - 2D plot of DSFs from a healthy state $\boldsymbol{F}_{hl,tr} \cup \boldsymbol{F}_{hl}$ and simulated damage DSFs $\boldsymbol{F}_{dm}$ and (b) shows the transformed DSFs by $\boldsymbol{U}\boldsymbol{f}^{(k)}$ and $\boldsymbol{U}\boldsymbol{\mu}_{\mathcal{F}}$ where the dotted circle is the threshold highlighting the overlap



(a) MD-based novelty index of the artificial 2D data set with a parabolic trend

(b) Heatmap

Figure 3.6: Highly overlapping MD-based novelty index

artificial set of DSFs with introduced variabilities.

For the MD-based novelty index, the accuracy and F_1-score presented in Table 3.4 is 0.54 and 0.56 respectively. The MCC confirms these results with a value close to 0. These results show that detecting novelties is possible in around 50% of the cases.

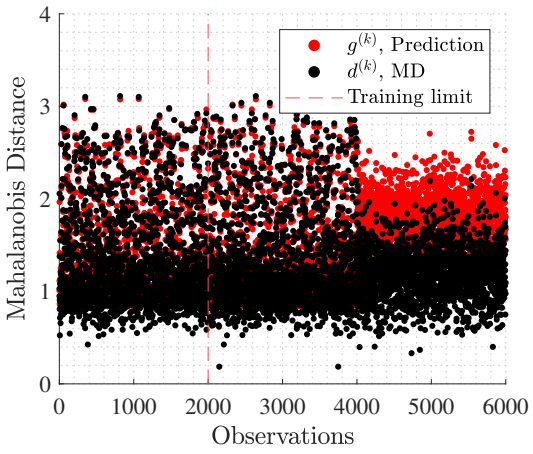
Table 3.4
Case 2: Classification results with $\boldsymbol{F}_{hl}$ as healthy and $\boldsymbol{F}_{dm}$ as damaged with artificial 2D DSFs

Performance metric	MD-based
Accuracy	0.54
F_1-score	0.56
MCC	-0.03

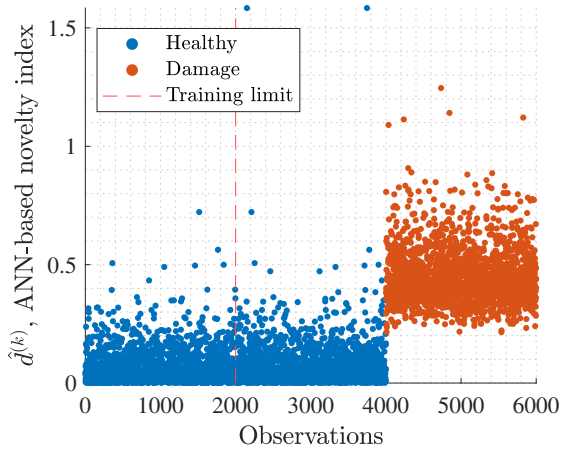
3.4 ANN-based novelty index with artificial data

Using the DSF set from case 2 in section 3.3, the proposed damage detection methodology is applied to learn the relationship between the DSFs and their influence on the MD-based novelty index. The complete data set is labelled by the MD calculated as per Eq. 3.2 with a reference matrix $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{q\times 200}$. The ANN regression function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ is trained with $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{2\times 2000}$, with DSFs $\boldsymbol{f}^{(k)}$ and the MD label $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ with $k : [1 \leq k \leq 2000]$. The calculated and predicted MD are presented in Fig. 3.7(a). The novelty index calculated as per Eq. 3.8 is presented in a control chart in Fig. 3.7(b). As only observations from the $\boldsymbol{F}_{hl,tr}$ are used to train the ANN, this results in an increment of the error when DSFs from the damaged set are used to predict the MD as this relationship was not learned by the ANN regression model.

The performance metrics are presented in Table 3.5. To demonstrate the reproducibility of the ANN-based novelty index, the procedure was performed 1000 times for which each time the results are evaluated using the same performance metrics. The mean values of the performance metrics are shown in Table 3.5 with the corresponding standard deviation in parentheses. The application of the suggested methodology was able to increase the accuracy of identifying novelties by predicting variabilities present in the MD-based novelty index and resulting in an improved separation of the two classes. This can be also observed on the heatmap in Fig. 3.8(a). Fig.

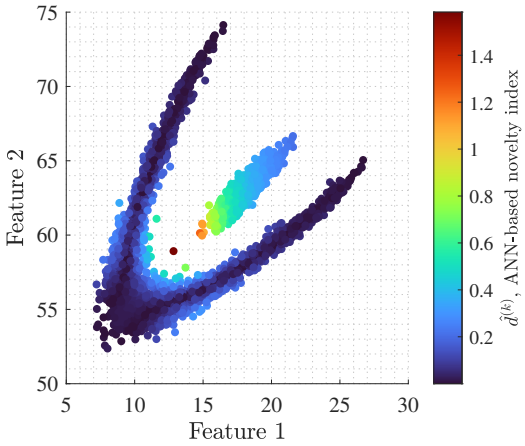


(a) Calculated MD $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ and the predicted $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ by ANN.

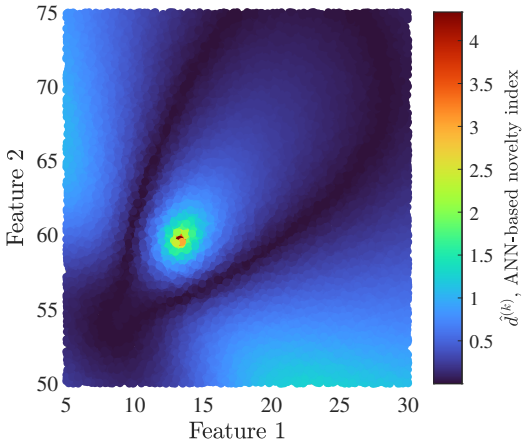


(b) Prediction error $\hat{d}^{(k)}$ obtained by Eq. 3.8 for the artificial data set used as a new novelty index (ANN-based novelty index) for damage detection

Figure 3.7: Constructing the ANN-based novelty index



(a) Heatmap of the ANN-based novelty index based on existing features and observations



(b) Heatmap of the ANN-based novelty indices resulting from all potential feature combinations and observations

Figure 3.8: Constructing the ANN-based novelty index

3.8(b) shows the ANN-based novelty index space, i.e., all potential feature combinations with values from feature 1 ranging between 5 and 30 and values from feature 2 ranging from 50 to 75 and the resulting novelty indices.

The performance confirms the improvement in novelty detection with an improved accuracy, F_1-score and MCC close to 1. The information that was lost when calculating the MD is recovered by the ANN-based novelty index. For the MD-based novelty index, the accuracy and F_1-score presented in Table 3.5 is 0.54 and 0.56 respectively. The MCC confirms these results with a value close to 0. These results show that detecting novelties is possible in around 50% of the cases.

Table 3.5
Classification results with $\boldsymbol{F}_{hl}$ as healthy and $\boldsymbol{F}_{dm}$ as damaged using the artificial 2D DSFs

Performance metric	MD-based	ANN-based*
Accuracy	0.54	0.98 (0.02)
F_1-score	0.56	0.97 (0.03)
MCC	-0.03	0.96 (0.04)

*ANN-based results computed by 1000 repetitions with the corresponding standard deviation in parentheses

3.5 A comparison of ANN and cointegration

3.5.1 Simulated six DOF mass-spring system

This section intends to compare between the proposed ANN-based novelty index and Cointegration using the six DOF mass-spring system presented Fig. 3.9. The system consists of six springs that are connected to six masses and fixed in one of the ends.

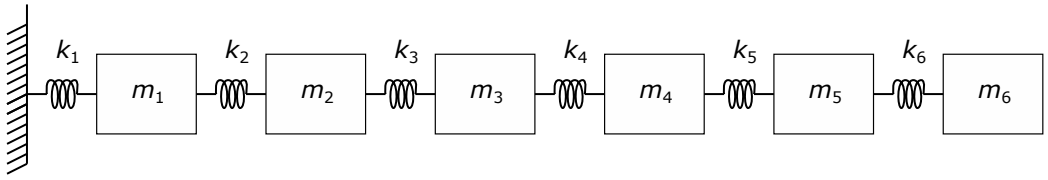


Figure 3.9: A six DOF mass-spring system

In this case study, the six natural frequencies of the system are considered as the DSF vector $\boldsymbol{f}^{(k)} = [\boldsymbol{s}_1(t), \boldsymbol{s}_2(t), ..., \boldsymbol{s}_6(t)]^T$, where $\boldsymbol{s}_i = \frac{\sqrt{\boldsymbol{\lambda}_i}}{2\pi}$, with the $\boldsymbol{\lambda}$ being the eigenvalues in Eq. 3.15:

$$(\boldsymbol{K}_\alpha - \boldsymbol{\lambda}\boldsymbol{M})\boldsymbol{\theta} = \boldsymbol{0} \tag{3.15}$$

with a constant mass matrix where $m_i = 2$ and i ranges from 1 to 6. To simulate the effect of EOVs on the DSFs, the stiffness of the system is exposed to the temperature field $\boldsymbol{u}(t)$ shown in Fig. 3.10. Additionally, stiffness k_2 and k_5 are exposed to random changes. The $\boldsymbol{K}_\alpha$ is the affected stiffness matrix expressed as:

$$\boldsymbol{K}_\alpha = (-\boldsymbol{u}(t) + C)\boldsymbol{K} \tag{3.16}$$

where C is an arbitrary constant value that maps the expression $(-\boldsymbol{u}(t) + C)$ to a positive value. For this study, $C = 80$. $\boldsymbol{K}$ is the stiffness matrix that holds the k_i coefficients defined as:

$$k_1 = k_3 = k_4 = k_6 = 2 \tag{3.17}$$

$$k_2 = 2r_2(t) \tag{3.18}$$

$$k_5 = 2r_5(t) \tag{3.19}$$

where $r_2(t) \neq r_5(t)$ are independent and uniformly distributed random values with $r_{2,5}(t) : [1 - \alpha > r_{2,5}(t) < 1 + \alpha]$ and $0 \leq \alpha \leq 1$. α represents random variabilities which were previously highlighted as a limitation by Cross et al. [33]. The eigenvalue problem is then solved M times, where M is the total amount of observations.

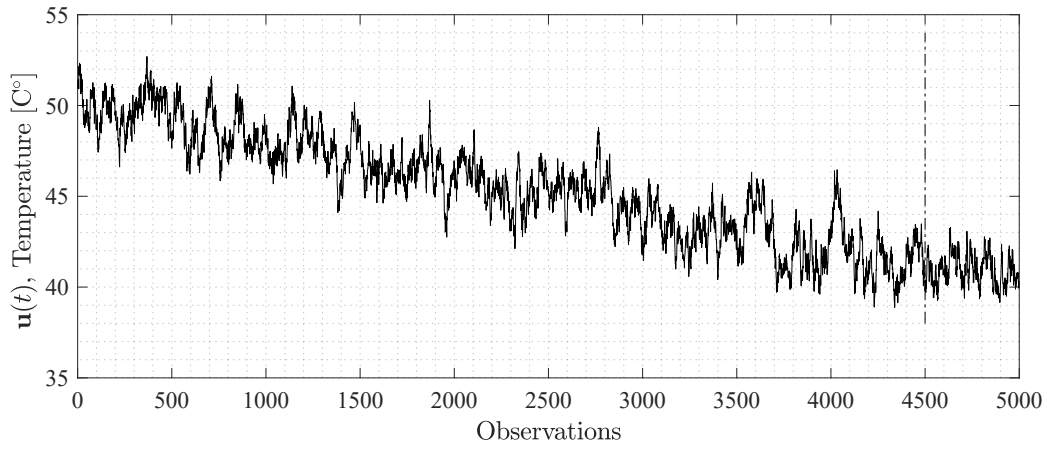


Figure 3.10: Simulated temperature field $\boldsymbol{u}(t)$. The dashed vertical line indicates when damage is introduced.

In this case study, the eigenvalue problem is solved for a total of 5000 observations: The first 4500 observations are from a system in a healthy state and the remaining 500 refer to a system in a damaged state. Having as much data available as possible for training is an important aspect to consider to maintain high accuracy for damage detection. Therefore, approx. 90% of the data was used to identify and learn possible patterns in the data. The summary of the number of observations used for semi-supervised learning is presented in Table 3.6.

The damage is emulated by decreasing k_3 by 2%. Initially, the stiffness reduction

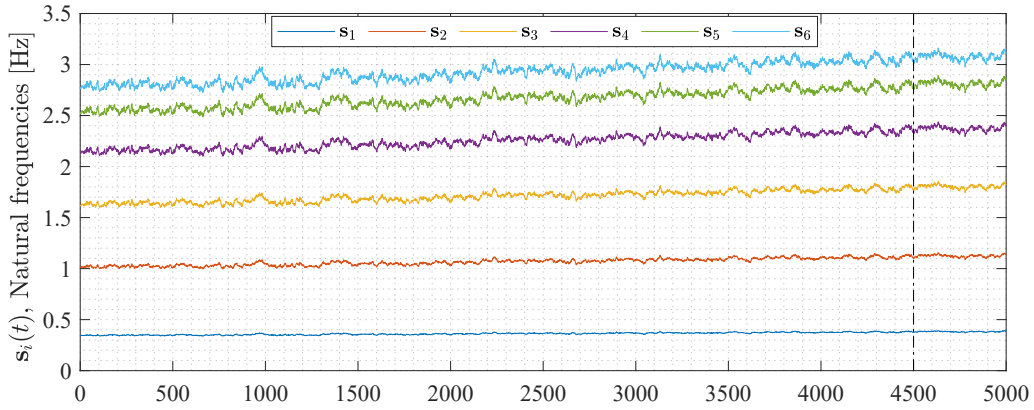


Figure 3.11: Six extracted natural frequencies $\boldsymbol{s}_i$. The dashed vertical line indicates when damage is introduced.

was set to 5%, however, this led to damage that was too clearly identifiable. Thus, a more conservative value was selected to allow for an overlap between damaged and undamaged novelty indices when implementing the MD-based approach. It can be expected that choosing a lower value for stiffness reduction than the selected 2%, the damage detection accuracy will decline.

Table 3.6
Summary of observations used for training, testing and damage detection in the six DOF mass-spring system

Scenario	Stiffness reduction	Stiffness parameter	Number of observations
Healthy			
Training $\boldsymbol{F}_{hl,tr}$		-	1 to 4000
Testing $\boldsymbol{F}_{hl}$		-	4000 to 4500
Damage $\boldsymbol{F}_{hl,dm}$	2%	k_3	4501 to 5000
Total			5000

To simulate the measurement noise, white Gaussian noise with 90dB signal-to-noise ratio is added to the natural frequencies. Observations 1 to 4000 were used for training. Both methods, ANN- and cointegration-based damage detection are tested using observations 4001 to 5000; of which the first 500 are from the healthy state and the remaining from the damaged state. The performance, by means of accuracy, F1 score and MCC, of the damage detection methods is evaluated based on observations 4501 to 5000 where the damage is present.

3.5.2 Mitigating the effect of EOVs by ANN and Cointegration

To illustrate the effects of EOVs, the MD-based damage index $d^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{\mathcal{F}})$ is calculated, where $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{6\times 4000}$ and $\boldsymbol{\mathcal{F}}$ is based on the first 4000 observations from a

healthy system, see Fig. 3.12. The damage detection threshold is assigned based on 2% allowed outliers for the first 4000 observations. It can be seen that the MD-based damage index follows a trend which hinders the discrimination between damage and EOVs.

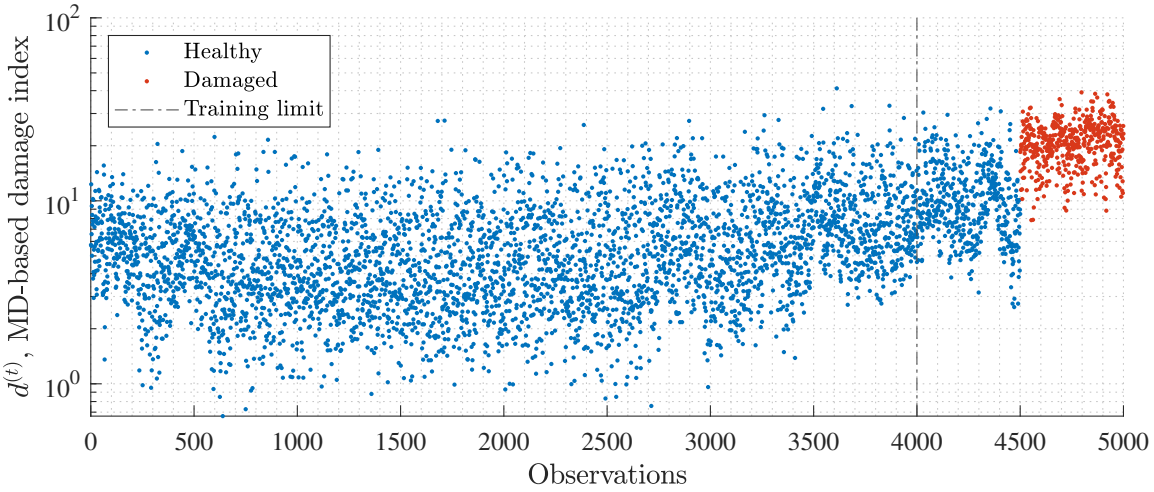


Figure 3.12: MD-based damage index for $\alpha = 0.15$

The complete data set is labelled by the MD calculated as per Eq. 3.2 with a reference matrix $\boldsymbol{\mathcal{F}} \in \boldsymbol{F}_{hl,tr}^{6\times500}$. The ANN regression function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ is trained with $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{6\times4000}$, with DSFs $\boldsymbol{f}^{(k)}$ and the MD label $d^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{\mathcal{F}})$ with $k : [1 \leq k \leq 4000]$. The ANN architecture presented in section 3.1.3 (with identical hyperparameters) is implemented. The results of the ANN prediction and the MD-based novelty index are presented in Fig. 3.13. While sacrificing on accuracy during the training in order to avoid over-fitting, the model resulted in a higher novelty index as shown in Fig. 3.14. The benefits can be observed between observation 4000 and 4500 which were not used to train the ANNs yet show a lower novelty index. An increase in the prediction error can be seen from observations 4500 to 5000, where the stiffness reduction is implemented. A clear difference between the validation set and the damaged observations can be seen in Fig. 3.14, where the ANN-based novelty index is calculated as per Eq. 3.8.

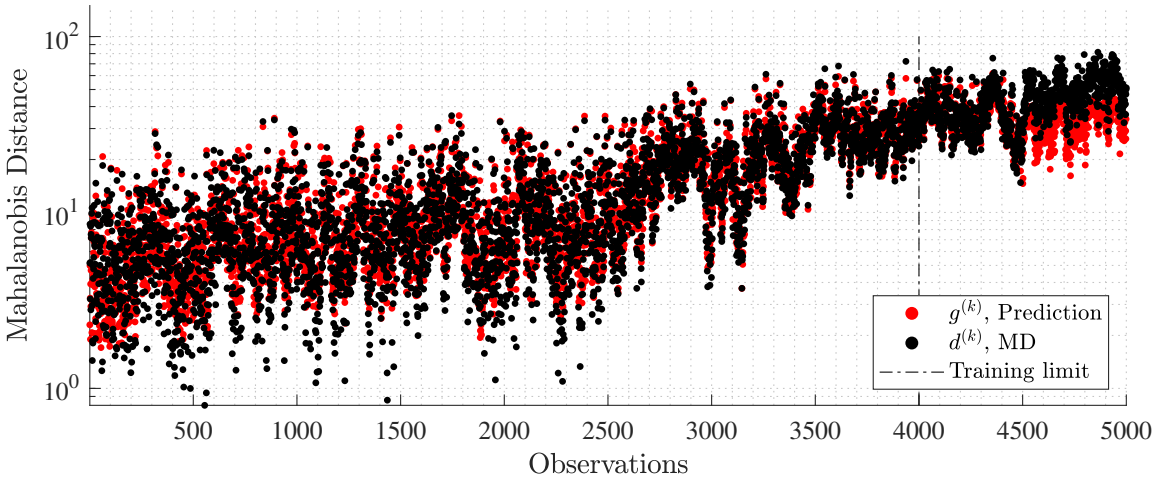


Figure 3.13: MD and predicted MD by ANN

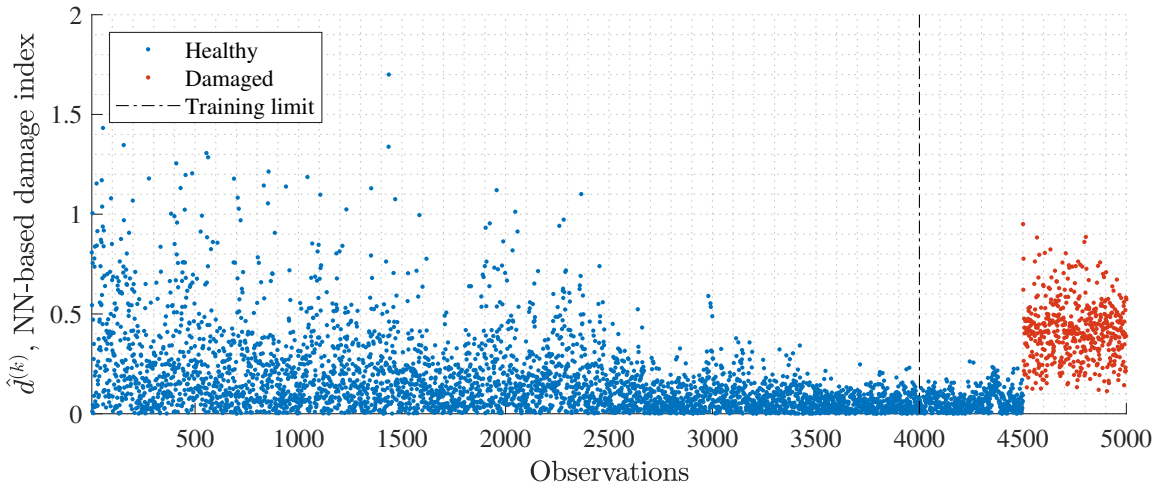


Figure 3.14: ANN-based novelty index

When observing the Cointegration residuals in Fig 3.15, the variabilities caused by the temperature field are removed and the observations from the damaged simulations appear with a slight shift in trend. Nevertheless, a significant overlap between the healthy and damaged data can be observed as also summarised in Table 3.7.

Overall, the ANN-based novelty index outperformed the cointegration-based novelty

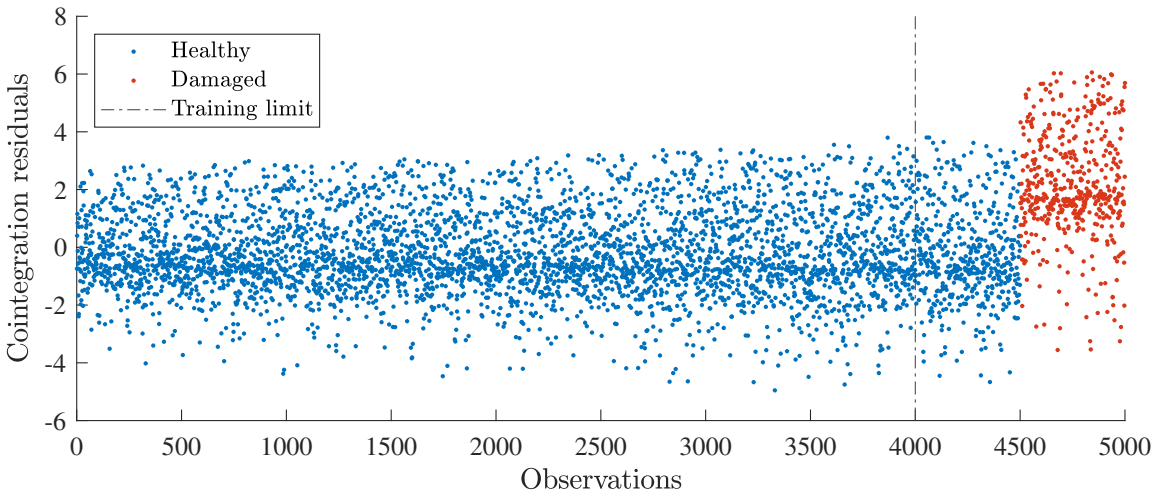


Figure 3.15: Cointegrating residuals

index in the provided performance metrics under additional consideration of repeatability by the given standard deviation.

Table 3.7
Classification results with $\boldsymbol{F}_{hl}$ as healthy and $\boldsymbol{F}_{dm}$ as damaged in the six DOF mass-spring system

	α	Accuracy	F1_score	MCC
ANN-based*	0	1	1	1
	5	1	1	1
	10	0.99 (0,00)	0.99 (0.00)	0,99 (0,01)
	15	0.96 (0,04)	0.96 (0.04)	0.91 (0.07)
	20	0.90 (0.06)	0.89 (0.06)	0.79 (0.13)
	25	0.79 (0.07)	0.79 (0.07)	0.59 (0.14)
Cointegration-based	0	1	1	1
	5	0.99	0.99	0.98
	10	0.82	0.82	0.64
	15	0.71	0.71	0.41
	20	0.73	0.73	0.45
	25	0.60	0.57	0.19

*ANN-based results computed by 1000 repetitions with the corresponding standard deviation in parentheses

Nevertheless, Fig. 3.15 shows clearly that the trend introduced by the temperature

field has been removed. While the observations from the damaged cases are overlapping with the healthy cases a clear shift can be identified. To highlight the advantages of cointegration the novelty index can be post-processed by calculating the moving average across 20 neighboring elements. The improved novelty index is shown in Fig. 3.16 where the damage can be clearly identified.

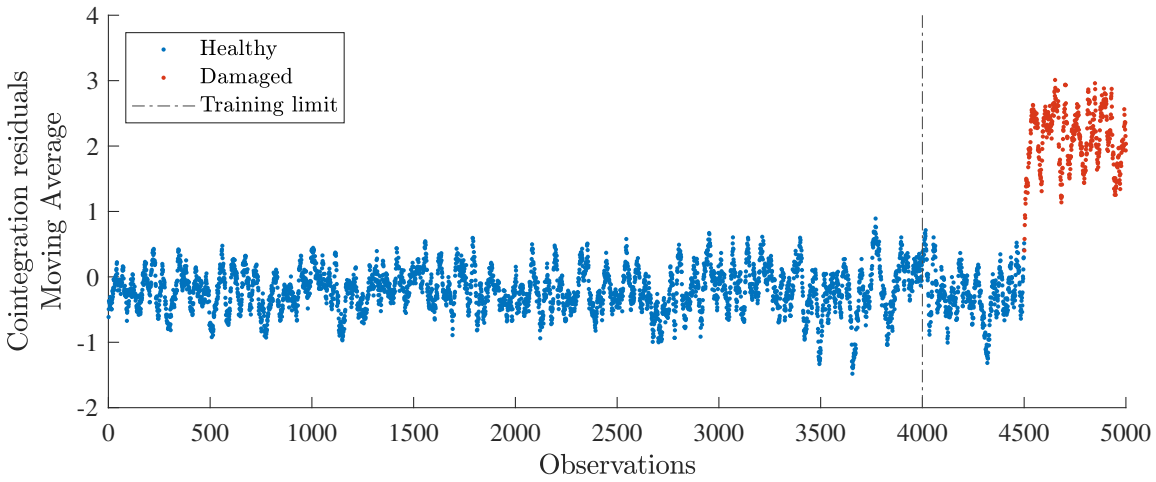


Figure 3.16: Cointegrating residuals - Moving average across 20 neighboring elements

3.6 A comparison of ANN and PCA

3.6.1 Lab scale experiment

The proposed method is tested on the 14.3m WTB manufactured by Olsen Wings [119] shown in Fig. 3.17. The blade weighs approx. 750 kg [22] and is mounted on a test rig and equipped with an electromagnetic actuator such as the one in Fig. 5.1(c) and placed close to the root of the blade.

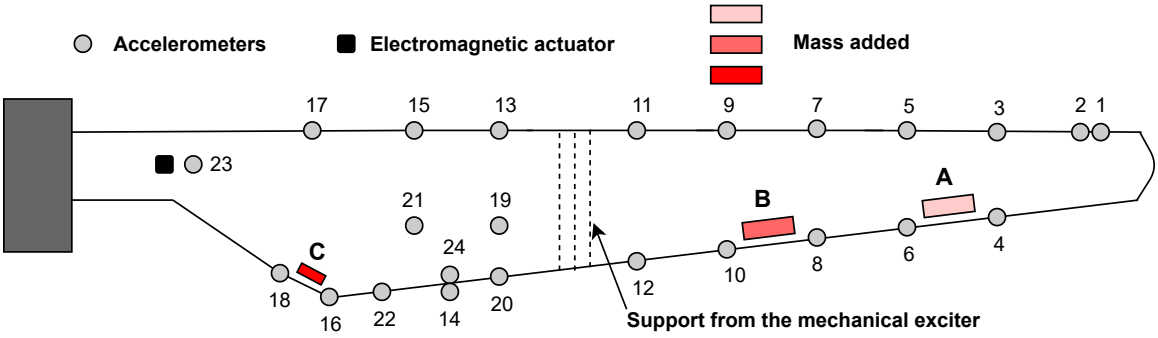
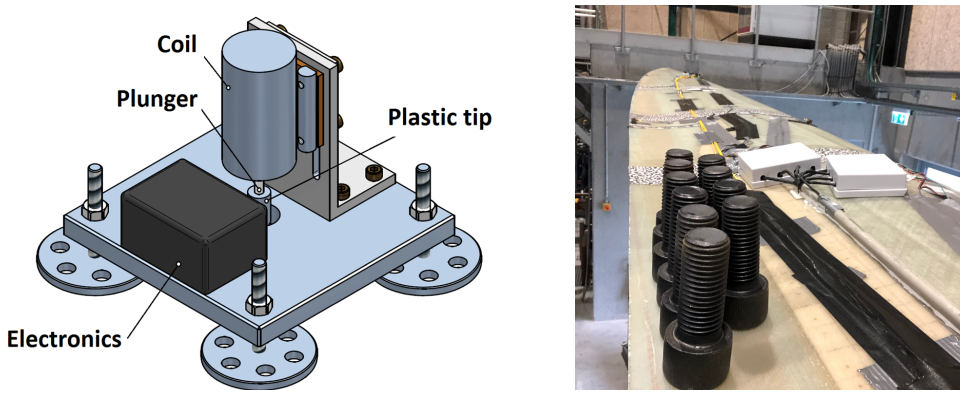


Figure 3.17: Schematic representation of the experimental set up of the 14.3m wind turbine blade. The WTB is mounted to an additional shaker which is not used for this experiment.

To measure the vibration response of the blade excited by the actuator hit, a total amount of 24 mono-axial accelerometers, Brüel & Kjær Type 4507, were mounted along the blade as shown in Fig. 3.17. The accelerometers 2 to 22 measure the vibration response in flap-wise direction, accelerometer 1 in edgewise direction. The accelerometers were connected to a data acquisition system Brüel & Kjær Type 3660-C with two LAN-XI modules, a 12-channel input module Type 3053-B-120 and 4-channel input/output module Type 3160-A-042.

The measured vibration response was recorded by a data acquisition system with a

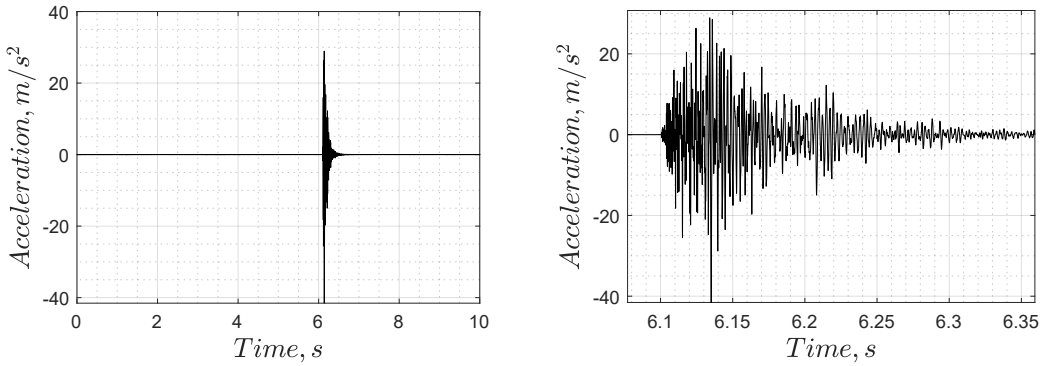


(a) Electromechanical actuator (b) Example of mass added

Figure 3.18: Experimental set up. Photo courtesy of DTU Large Scale Facility.

sampling frequency of 16,384Hz. Using the actuator as a trigger signal, the vibration responses from all accelerometers were aligned and trimmed to 200 samples around the maximum amplitude of accelerometer 23. This sample range was identified as the most sensitive to changes in dynamic response as shown in Tcherniak and Mølgaard [152]. Since only 200 samples are used for damage detection, it can be expected that reducing the sampling frequency can lead to features which are less sensitive to damage. An example of the vibration response from accelerometer 8 is shown in Fig 3.19(a) and 3.19(b).

To simulate novelties, mass in form of heavy bolts weighing approx. 1kg each, shown in Fig 3.18(b), were placed in different locations marked in red in Fig. 3.17. The blade was excited 2000 times without mass to create a baseline for novelty detection. Then, in location A, five bolts (representing approx. 0.67% of the mass of the blade) were place on the blade and the blade was excited ten times. Next, the amount of bolts was increased to 10 (representing 1.33% of the mass of the blade) and 20 bolts (representing 2.66% of the mass of the blade) repeating the procedure for each variation of mass and for locations B and C. Masses for each location were placed and measured separately. Table 3.8 summarises the damage scenarios simulated with



(a) 10s vibration response (b) Trimmed vibration response

Figure 3.19: Vibration response measured by accelerometer 8

the corresponding mass added and locations.

Table 3.8
Simulated novelties according to location and mass added

Scenario	Mass	Location	Number of Excitations
Blade without mass			
Training $\boldsymbol{F}_{hl,tr}$		-	1500
Testing $\boldsymbol{F}_{hl}$		-	500
Blade with mass $\boldsymbol{F}_{dm}$	5kg	A	10
	10kg	A	10
	20kg	A	10
	5kg	B	10
	10kg	B	10
	20kg	B	10
	5kg	C	10
	10kg	C	10
	20kg	C	10
Total $\boldsymbol{F}$			2090

3.6.2 MD-based novelty detection

Prior to calculating the MD-based novelty index, DSFs are obtained by calculating the cross-covariance between the trimmed sensor response for each observation. Cross-covariance between sensor has shown promising results in experimental studies, e.g., Parker [122], and was further adopted by Tcherniak and Mølgaard [152] showing promising results for damage detection in operating WTBs. By obtaining the cross-covariance from 22 sensors for each observation, the DSF matrix is populated by $\boldsymbol{f}^{(k)} \in \boldsymbol{F}^{253\times 1}$ and a total feature space of $\boldsymbol{F} \in \mathbb{R}^{253\times 2090}$ for 2090 observations.

The MD-based novelty index $d^{(k)}$ is calculated with the inverse covariance matrix $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{253\times 1500}$. A threshold is set by allowing 1% of outliers based on the first 1500 observations. Deviations after the training limit of 1500 observations can be observed leading to false positives in the test set $\boldsymbol{F}_{hl,tr}$.

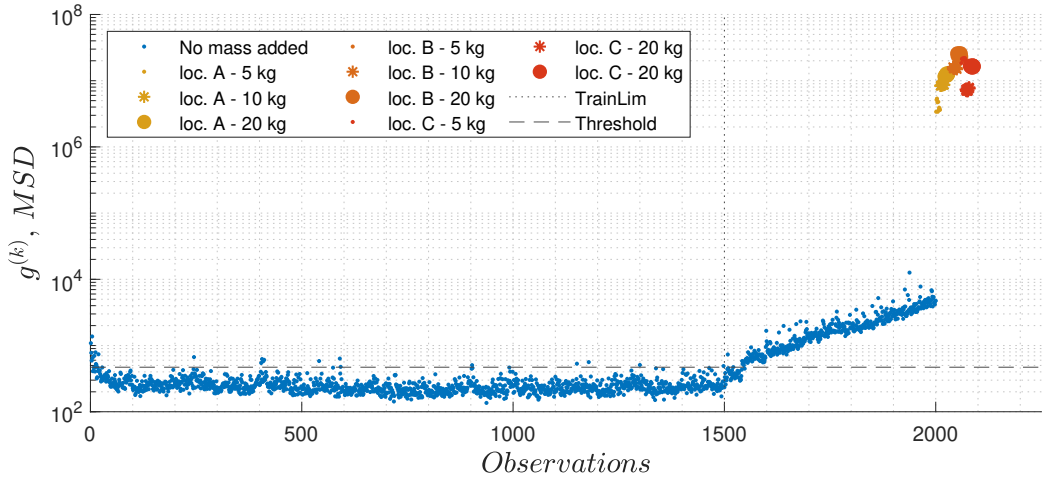


Figure 3.20: MD-based novelty index without dimensional reduction

Nevertheless, a clear separation between observations without mass and with the added mass can be seen in Fig. 3.20. Rather then evaluating the overall performance of the damage detection, which can be misleading due to a clear separation

of damage, the focus lies here on the false positives occurring in the test set $\boldsymbol{F}_{hl,tr}$. By allowing 1% of outliers during the training set, a total of 5 outliers is expected when observing the testing set, which is clearly exceeded with a total of 456 indices above this threshold. Due to the repeatability of the experiment, high collinearity between the sensor responses is present. This causes numerical errors when computing the inverse covariance matrix for the MD-based novelty index. The effect can be observed in an increased novelty index after observation 1500 as new observations are not normalised.

As discussed in the literature review, PCA is a popular tool and considered as state-of-the-art for mitigating the effects of EOVs on DSFs for damage detection. Before applying the proposed ANN-based novelty index, the application of PCA will be implemented and discussed in the next section.

3.6.3 PCA and the MD-based novelty detection

PCA is an orthogonal transformation applied to a set of potentially correlated values into linearly uncorrelated values which are known as the principal components (PC) [163]. The first PCs contain most of the variability from the original dataset leaving each succeeding component with a share of the remaining variability. This procedure allows to reduce the dimension of a dataset while keeping most of the information possible expressed in the variability of the data. In this analysis, PCA is applied to reduce the dimension of the feature matrix and account for collinearity between the feature vectors, thereby highlighting its advantages and limitations.

Identifying the number of PC required for data normalisation can be a challenging task. As mentioned in the literature review, there is no general rule which guarantees

robust results. To get a closer insight, the PC are reduced one by one and the MD-based novelty index is calculated with the first 1500 observations. For each set of PC, the MD-based novelty index is calculated and the false positives are evaluated as shown in Fig. 3.21. Additionally, to avoid numerical error by computing the inverse of the covariance matrix $\mathbf{\Sigma}_{\mathcal{F}}^{-1}$ for the MD, the condition number of $\mathbf{\Sigma}_{\mathcal{F}}^{-1}$ is examined along with the influence of dimensional reduction by PCA.

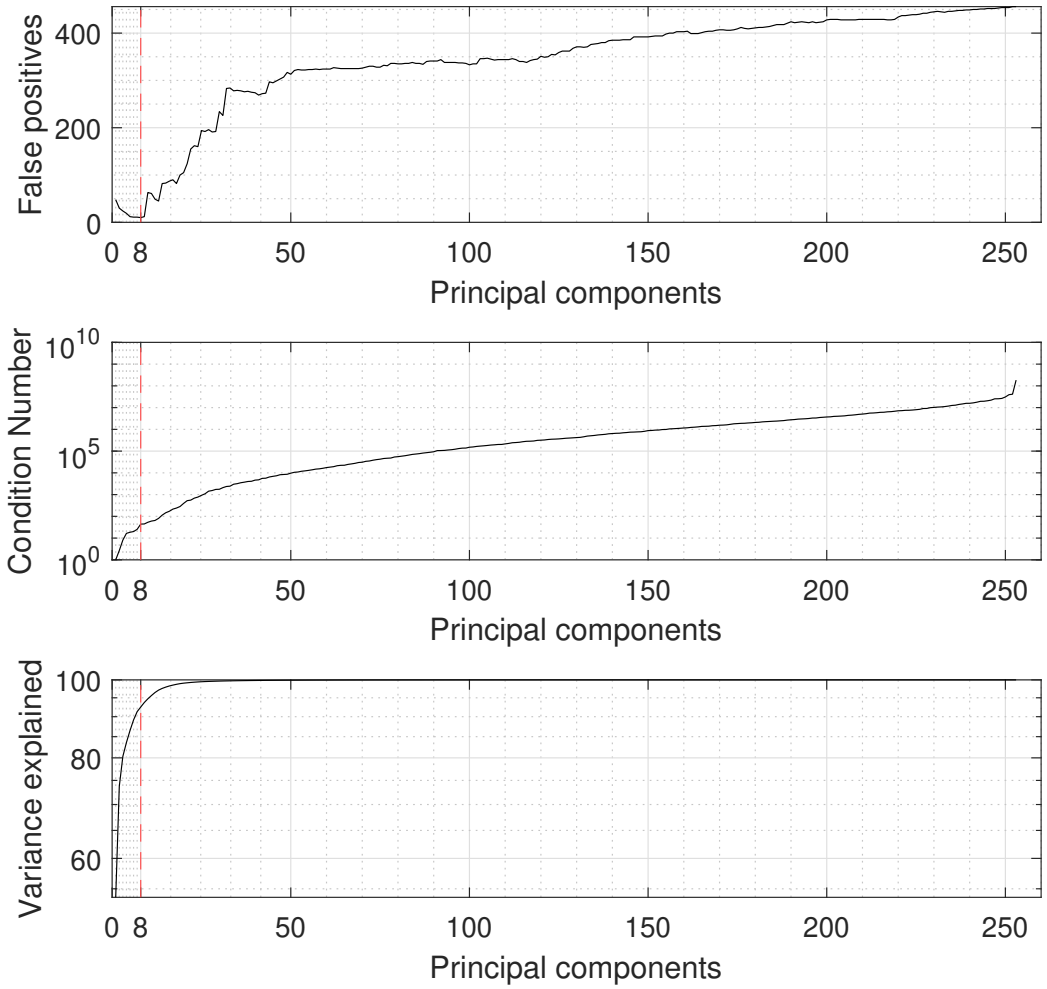


Figure 3.21: Top plot: False alarm rate. Middle plot: Condition number of DSFs. Bottom plot: Remaining variance explained.

By reducing the total amount of 253 features to 8 PC, this results in a total of 10 (2%) false positives in the test set $\boldsymbol{F}_{hl}$, as shown in Fig. 3.22, with around 92.5%

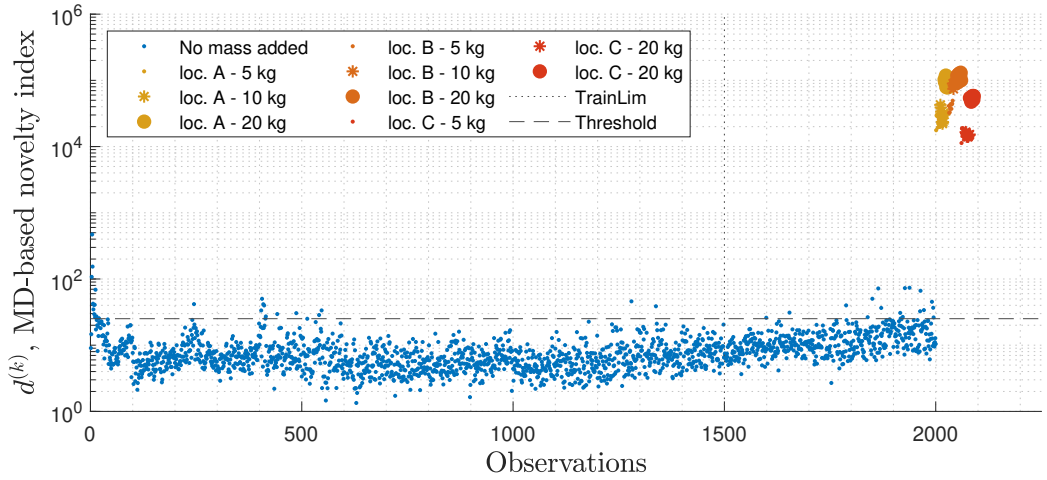


Figure 3.22: MD-based novelty index with dimensional reduction

of the remaining variance explained. Significant reductions are also observed for the condition number which is reduced from around 10^7 to 44. If the condition number is not much larger than 100, the matrix can be considered as well conditioned and non-singular. Given the fact that the blade was excited by the electromechanical actuator for 2000 times under the same conditions, small deviations in the excitation or measurement noise will amplify noise in the test data by the factor of the condition number. Therefore, the condition number is a potential guide for selecting the number of PC.

In the top plot of Fig. 3.21, it can be clearly seen how sensitive the results are to the choice of the PC. By using 10 PC (2 PC more), with a remaining variance explained of around 94%, the false positives increase to 63 which shows an insufficient transformation with remaining collinearity. This can make the application of PCA in real structures, where information regarding the damage is not known for validation, quite limiting.

3.6.4 ANN-based novelty index

The procedure to calculate the ANN-based novelty index is identical to the one described in section 3.1.4 and implemented in section 3.5. The MD as the prediction target is calculated with $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{6\times200}$ as shown in Fig. 3.23. It can be observed that the deviations increase directly after observation 200 which amplifies the variabilities when observing the novelty index. This allows the ANN to learn the existing patterns.

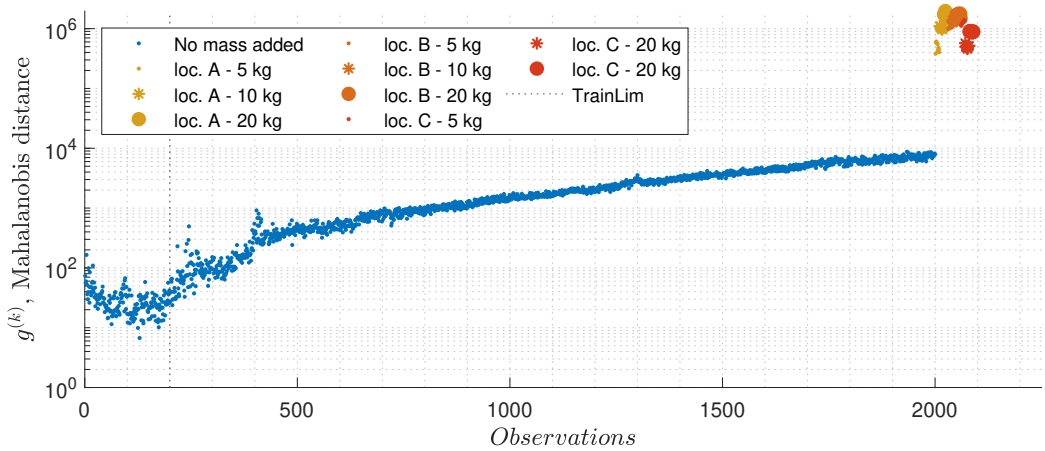


Figure 3.23: MD-based novelty index with $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{6\times200}$

The ANN regression function $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ is trained with $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{6\times1500}$, with DSFs $\boldsymbol{f}^{(k)}$ and the MD label $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ with $k : [1 \leq k \leq 1500]$. The calculated and predicted MD is presented in Fig. 3.24. Increased deviations between $g^{(k)}$ and $d^{(k)}$ can be observed between observation 1 to 1500 which were used to train the ANN. This is a result of generalising the ANN and avoiding over-fitting. The ANNs are not able to predict the artificial damage introduced to the blade as can be seen in Fig. 3.24 after observation 2000. This is also confirmed in Fig. 3.25 which shows the ANN-based novelty index according to Eq. 3.7.

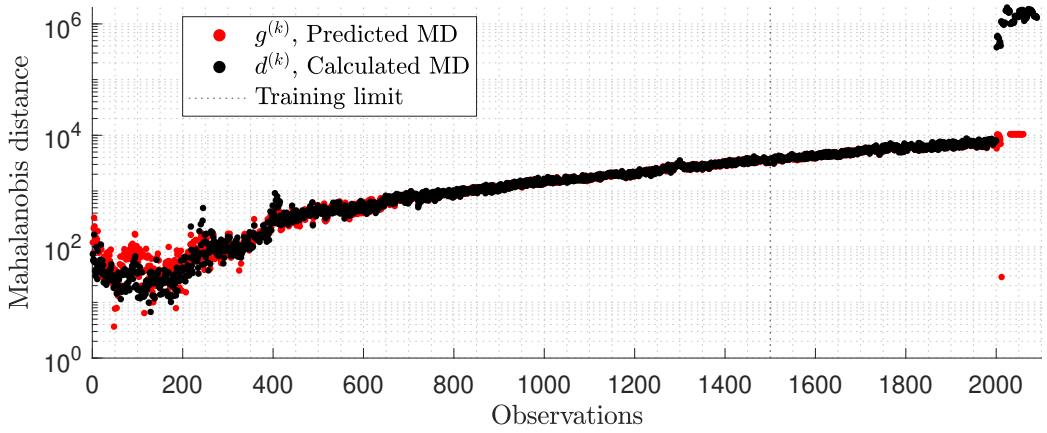


Figure 3.24: MD-based novelty index without dimensional reduction

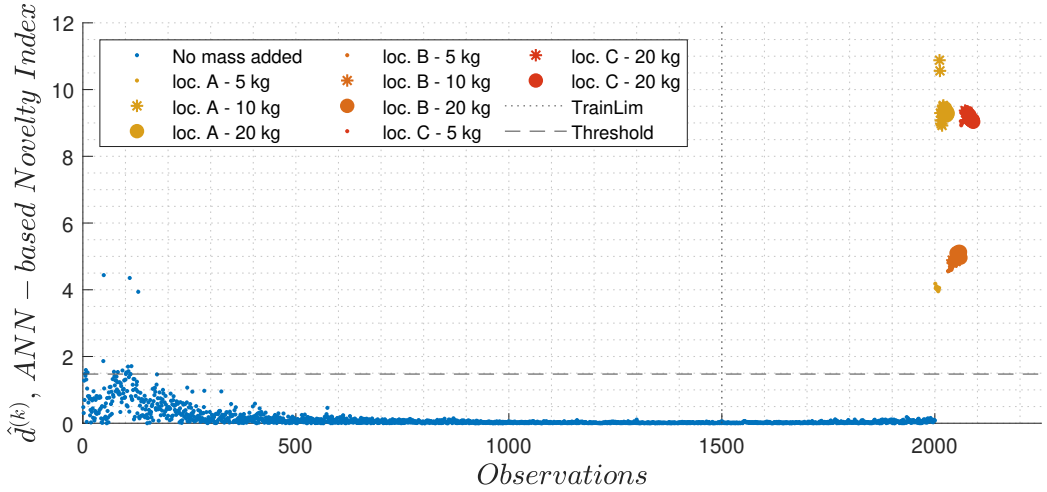


Figure 3.25: ANN-based novelty index

3.7 Discussion

The MD-based novelty index has long been used for damage detection given its low computational cost and its reproducibility. To some extent, the MD-based novelty index can normalise and mitigate variabilities in data as shown in 3.3, Case 1. Nonetheless, clear limitations appear when the variabilities in the data do not follow a linear trend as shown in 3.3, Case 2, as this results in overlapping indices. These limitations

of the MD-based novelty index have been discussed by Farrar and Worden [51]. The authors highlight the use of auto-associative neural network (AANN) to cope with non-Gaussian distributions in DSFs. This approach, which is based on unsupervised learning, was successfully implemented and demonstrated by Dervilis et al. [41].

The main challenge with the adoption of unsupervised learning is the lower accuracy in the results. An unsupervised algorithm is provided with non-labelled information to find patterns, i.e., categories or clusters, to enable the organisation of the data [17]. This implies that the algorithm decides whether a pattern is relevant or not without a reference. For this reason, the analysis of the results may require the intervention of the operator and domain knowledge, making this approach costlier.

The framework introduced and demonstrated in this chapter is a semi-supervised framework regardless of the implementation of ANN, which is a supervised regression model. In other words, the detection of novelties is performed by taking advantage of supervised learning whilst using only a subset of data labelled as healthy.

The proposed framework was demonstrated with a simple two-dimensional example of an artificial dataset in section 3.4. The proposed ANN structure consists of one hidden layer with only two neurons. While the choice of the ANN hyperparameters and architecture can be modified to improve the results, it is most crucial to have these fixed across different applications for the generalisation of the proposed approach. This is due to the fact that, in real applications such as operating wind turbines or civil structures, the damage state is not known beforehand and, therefore, it is not possible to calibrate the model towards a higher detection accuracy.

The proposed framework was compared to well established approaches such as PCA (in section 3.6.1) and cointegration (in section 3.5). In both cases, the proposed framework outperformed the alternative approaches. Cointegration was challenged

by the random variabilities introduced to the stiffness parameter in the 6-DOF mass-spring system. Since the ANN is able to recognise non-linear patterns, it coped well with the randomness that caused non-linear behaviour. PCA was able to improve the results from the MD-based novelty index, producing significantly less false alarms. PCA normalised the data and reduced the variabilities after training the model with 1500 observations. Nonetheless, the results were highly sensitive to the choice of PC, which is a major limitation for a generalised solution.

Throughout the two experiments, the size of the training set is significantly larger than the size of the testing set. This is important as machine learning techniques require large datasets to generalise their predictions and be able to observe as many patterns as possible. When implementing an online monitoring system, it is recommended to retrain the algorithm including the test data once these observations have been identified as healthy. This will enhance the prediction, and thus damage detection capability of the algorithm, over time. The damage detection accuracy is likely to decline when the training set size is reduced.

To be able to generalise the propose framework, two objectives need to be addressed, namely a fixed ANN structure with fixed hyperparameters and ANN architecture through all case studies and reproducibility, i.e., obtaining consistent results after re-training the ANN. In this chapter, the artificial numerical and lab-scale examples used a fixed ANN structure with identical hyperparameters. Additionally, reproducibility was ensured by repeating the training and prediction process with the ANN 1000 times. In this way, both of the previously mentioned objectives were achieved.

The proposed framework does not address the remaining challenge of interpretability. Although damage is successfully detected by mitigating the variabilities in the data, the relationship between DSFs and the variabilities in the novelty index is unknown. The issue of interpretability will be addressed in the next chapter and evaluated on

an in-operation wind turbine in Chapter 5.

Chapter 4

Shapley Additive Explanations for Interpretable Damage Detection

As discussed in the literature review, one of the primary challenges facing ML in SHM is the lack of interpretability of the prediction models hindering the broader implementation of these techniques. For this purpose, this chapter introduces a novel data-driven framework to distinguish novelties due to the influence of environmental and operational variabilities (EOVs) from those caused by damage in a structure. This framework, published in Movsessian et al. [111], addresses the lack of interpretability and transparency in ML techniques used for vibration-based structural health monitoring (VSHM).

The novel framework incorporates the Shapley Additive exPlanations (SHAP) method into a ML-based damage detection process, thereby enabling a tool for interpretability and, thus, building evidence for reliable decision-making in SHM applications. The SHAP method is based on coalitional game theory and adds global and local interpretability to ML-based models by computing the marginal contribution of

each feature to the model's prediction. The contributions are used to understand the nature of damage indices (DIs). The proposed framework is evaluated on a 6-degrees-of-freedom lumped mass-springer-damper system with simulated temperature variabilities. Finally, the applicability, benefits and limitations of this interpretable damage detection framework are discussed.

4.1 Methodology on interpretable novelty detection

This section elaborates on the steps proposed to determine the cause of novelties identified through an MD-based damage detection approach. First, extracted DSFs are used to calculate the MD-based DI. For this, under a semi-supervised framework, only a subset of healthy data is required. Then, an XGBoost regression model is trained considering the DSFs and measured environmental variabilities as predictors and the MD-based DI as the prediction target. Finally, SHAP is included in the analysis to interpret the XGBoost model and thus interpret the contribution of each DSF and measured environmental parameter to the DI. A schematic representation of the proposed framework is illustrated in Fig. 4.1.

4.1.1 Gradient boosted becision trees

XGBoost is a scalable ML technique for tree boosting developed by Chen and Guestrin [27]. The algorithm uses nested decision trees to improve efficiency and learning in the model [57]. Whilst decision trees use simple conditional statements to classify the features into target labels based on particular attributes of the features, XGBoost uses

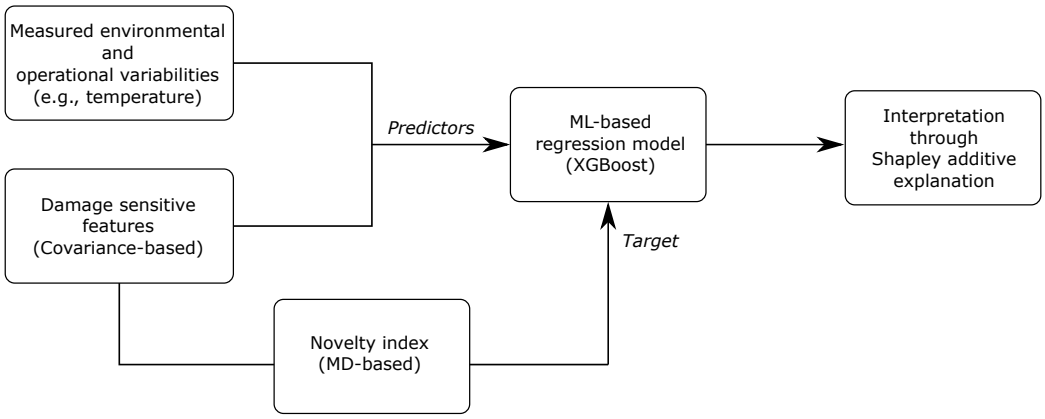


Figure 4.1: Proposed flow for interpreting the MD-based DI

multiple decision trees in parallel, thereby optimising performance and improving the accuracy of the classification. Each decision tree within the XGBoost algorithm learns from the previous decision tree gradually improving the model and building a strong learner as shown in Fig. 4.2

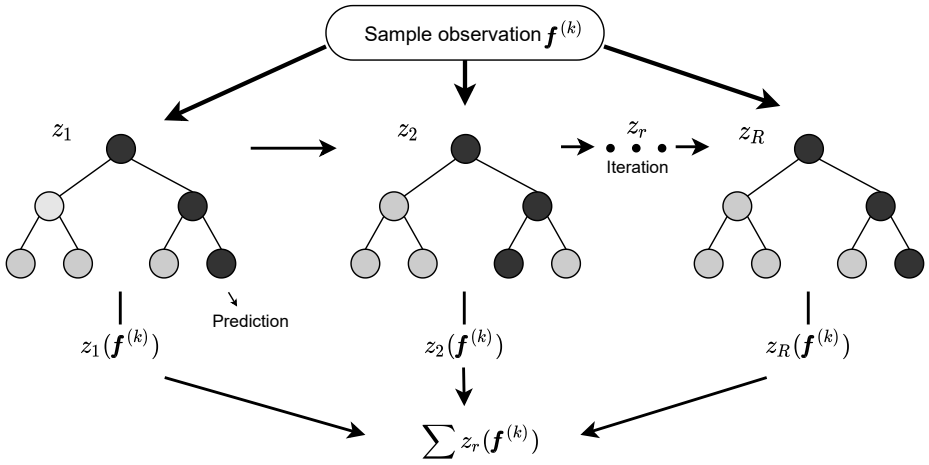


Figure 4.2: Basic structure of gradient boosted decision trees

Let $(\boldsymbol{f}^{(1)}, d^{(1)}), (\boldsymbol{f}^{(2)}, d^{(2)}), ..., (\boldsymbol{f}^{(k)}, d^{(k)}), ..., (\boldsymbol{f}^{(M)}, d^{(M)})$ be a data set with M observations containing the DSFs including measured EOVs $\boldsymbol{f}^{(k)} \in \mathbb{R}^{\mathcal{L}}$ and the corresponding prediction target $d^{(k)} \in \mathbb{R}$. The tree ensemble model is the summation of R additive

functions with the estimated prediction $g^{(k)}$ defined as follows:

$$g^{(k)} = \sum_{r=1}^{R} z_r(\boldsymbol{f}^{(k)}), \quad z_r \in \mathcal{K}, \tag{4.1}$$

where z_r is the independent tree structure of r trees and $\mathcal{K}$ the tree space. Different sets of functions are trained by minimising the following objective function:

$$J^{(r)} = \sum_{i=1}^{M} l(g^{(i)}, d^{(i)}) + \sum_{z=1}^{t} \Omega(r_z) \tag{4.2}$$

where the first regularisation term l is a loss function which measures the difference between the prediction $g^{(i)}$ and the prediction target $d^{(i)}$ at iteration step t (or t^{th} tree). The second term Ω is the complexity of the model and is defined as:

$$\Omega(r_z) = \gamma P + \frac{1}{2}\lambda \sum_{j=1}^{P} w_j^2, \tag{4.3}$$

where P is the number of leaves in the tree, w_j the score of a leave, γ the complexity scalar of each leaf and λ the scaling parameter to penalise the complexity. Unlike in general gradient boosting decision trees, for XGBoost the second-order Taylor expansion is considered for the loss function. Using the mean squared error (MSE) as a loss function, the t^{th} tree can be derived as follows:

$$\begin{aligned} J^{(r)} &\approx \sum_{i=1}^{M} [b_i w_{q(\boldsymbol{f}^{(i)})} + \frac{1}{2}(h_i w^2_{q(\boldsymbol{f}^{(i)})})] + \gamma P + \frac{1}{2}\lambda \sum_{j=1}^{P} w_j^2 \\ &= \sum_{j=1}^{P} [(\sum_{i \in I_j} b_i) w_j + \frac{1}{2}(\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma P. \end{aligned} \tag{4.4}$$

by defining $I_j = \{i | q(\boldsymbol{f}^{(i)}) = j\}$ as the instance set assigned to the j^{th} leaf. I_j represents all the data samples in leaf node j; b_i and h_i are the first and second derivatives of the MSE loss function and data points are assigned to the corresponding

leaf with the function $q(\cdot)$. Further, by letting $G_j = \sum_{j \in I_j} b_i$ and $H_j = \sum_{j \in I_j} h_i$, optimising the objective function can be transformed to finding the minimum of a quadratic function. The optimal weight w_j^* for the j^{th} leaf is computed by:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{4.5}$$

for a given structure $q(x_i)$

$$J^* = -\frac{1}{2} \sum_{1}^{T} \frac{G_j}{H_j + \lambda} + \gamma R \tag{4.6}$$

where Eq. 4.6 is used as a scoring function to measure for a given tree structure q. To evaluate the best split during the training and determine when the splitting is stopped, the following gain score is calculated:

$$Gain = \frac{1}{2} \left[\frac{{G_L}^2}{H_L + \lambda} + \frac{{G_R}^2}{H_R + \lambda} - \frac{{G_L + G_R}^2}{H_L + H_R + \lambda} \right] - \gamma, \tag{4.7}$$

where the first two terms refer to the summation of the left and right tree respectively and are compared to the score of the original leaf in the third term. The score of the generated leafs is penalised by γ. If the gain is negative, the tree stops growing and returns to the previous split, otherwise it continues improving the model.

In other words, after every split, the decision tree (DT) model is evaluated based on the two regularisation terms: The first term in Eq. 4.2 is responsible for the prediction accuracy during the training and the second term considers the complexity of the DT. As high tree complexity can cause over-fitting, this is penalised.

4.1.2 Shapley values

The Shapley Additive Explanations (SHAP) method is used in this study to interpret the output of the model described in section 4.1.1. The SHAP method is an additive feature attribution method that defines the output of a model as the sum of the real values attributed to each input feature. Additive feature attribution methods have an explanation model a defined as a linear function of binary features as in the following equation:

$$a(z) = \phi_0 + \sum_{l=1}^{L} \phi_l ß_l \tag{4.8}$$

where $ß_l \in \{0,1\}^L$, L is the length of the DSF vector and $\phi_l \in \mathbb{R}$ is the feature attribution values. The variables $ß_l$ represent a feature being observed ($ß_l = 1$) or unknown ($ß_l = 0$).

The SHAP method defines $z_{\boldsymbol{f}}(S) = E[z(\boldsymbol{f})|\boldsymbol{f}_S]$, where S contains a set of non-zero indexes in ß. To compute the feature attribution value ϕ_l for each input feature as illustrated in Fig. 4.3, SHAP uses the classic Shapley values from game theory as in the following equation:

$$\phi_l = \sum_{S \subseteq N \setminus \{l\}} \frac{|S|!(L - |S| - 1)!}{L!} [z_{\boldsymbol{f}}(S \cup \{l\}) - z_{\boldsymbol{f}}(S)], \tag{4.9}$$

where N is the set of all input features.

While the feature attribution value ϕ_l is theoretically optimal and can be used to interpret any model, it requires an algorithm of the order $O(LR2^M)$, where R is the

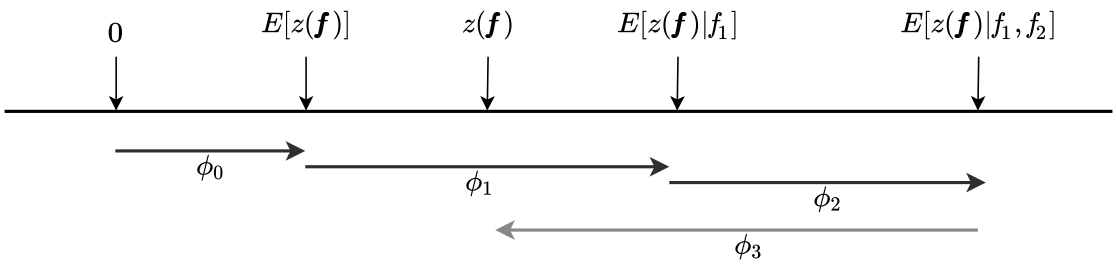


Figure 4.3: Additive atribution of SHAP values ϕ_l to explain a model's output

number of trees in the ensemble model, L is the maximum number of leaves in any tree and M is the number of features. This can be impractical for a large feature space. Lundberg et al. [98] proposed a modified algorithm to compute the interaction effect from pairwise feature attributions based on both the classic Shapley values from game theory and the concept of interaction indices [61]. The SHAP interaction values $\Phi_{i,j}$ are calculated through an algorithm with a computational complexity of $O(LTD^2)$, where D is the maximum depth of the tree, as shown in Equation 4.10:

$$\Phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M-1)!} \nabla_{ij}(S), \tag{4.10}$$

where,

$$\begin{aligned} \nabla_{ij}(S) &= z_{\boldsymbol{f}}(S \cup \{i,j\}) - z_{\boldsymbol{f}}(S \cup \{i\}) - z_{\boldsymbol{f}}(S \cup \{j\}) + z_{\boldsymbol{f}}(S) \\ &= z_{\boldsymbol{f}}(S \cup \{i,j\}) - z_{\boldsymbol{f}}(S \cup \{j\}) - [z_{\boldsymbol{f}}(S \cup \{i\}) - z_{\boldsymbol{f}}(S)] \end{aligned} \tag{4.11}$$

Equation 4.10 shows that the SHAP interaction value of the i-th feature with respect to the j-th feature corresponds to the difference between SHAP values of the i-th feature with and without the j-th feature.

The SHAP interaction effect $\Phi_{i,j}$ between the i-th and j-th is split equally (i.e. $\Phi_{i,j}$=$\Phi_{j,i}$) and the total interaction effect is $\Phi_{i,j} + \Phi_{j,i}$. The main effect for the prediction is then the difference between SHAP value and sum of SHAP interaction values for a feature as seen in Eq. 4.12:

$$\Phi_{i,i} = \phi_i - \sum_{j \neq i} \Phi_{i,j} \tag{4.12}$$

To maintain the integrity of the SHAP values and avoid the computational complexity of Eq. 4.9, the SHAP algorithm for tree-based models, i.e., TreeShap, as proposed by Lundberg et al. [97], is adopted to interpret the XGBoost model described in section 4.1.1.

4.2 Simulated lumped six degree of freedom system

To evaluate the proposed framework in an application to VSHM, a simple dynamic system is considered. Figure 4.4 shows a six DOF lumped mass-spring-damper system consisting of six masses m_1 to m_6 connected to each other by springs with stiffnesses k_1 to k_6 and dampers with the coefficients c_1 to c_1. The system is fixed to the ground on one of the sides. The mass and damper coefficients are given the same value of $2kg$ and $0.01\frac{Ns^2}{m}$ respectively.

To model the effect environmental variability on the system's dynamic, the temperature T_i is assume to affect the spring stiffness. For simplicity, the following rule is used:

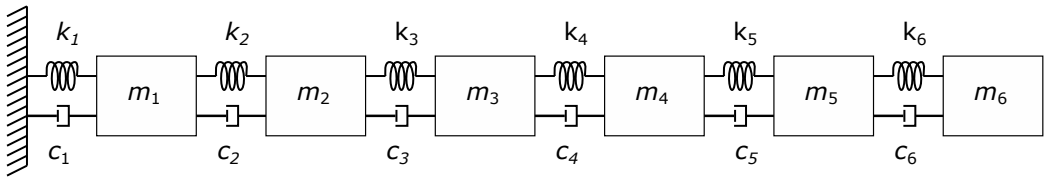


Figure 4.4: A six DOF mass-spring-damper system

$$k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = 0.1 \times T_i, \tag{4.13}$$

The system is excited by an impulse force applied to the first mass and, after the impulse, the masses vibrate freely. The resulting time series with 4350 samples with a sampling frequency of $1321Hz$ are used for calculating the cross-covariance which serves as DSFs.

In total, 2000 observations were simulated. At each observation, the temperature is set according to Fig. 4.5. It ranges from approximately 7.5°C to 15.5°C, where extreme weather events are implemented as sudden temperature drops of 4°C between observations $\boldsymbol{F}_{hl,tmp_1}$ 1100 - 1150 and $\boldsymbol{F}_{hl,tmp_2}$ 1400 - 1450. To simulate damage, after 1950 observations, the stiffness k_3 is reduced by 5%. The value for the stiffness reduction was selected to allow an overlap between novelty indices from the damage state and the extreme weather events. For each observation, the stiffnesses are set according to Eq. 4.13, and the DSFs are calculated. The number of observations per simulated state is summarised in Table 4.1.

4.2.1 MD-based novelty index

The control chart in Fig. 4.6 is obtained by calculating the MD-based novelty index $d^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{\mathcal{F}})$ according to Eq. 3.2. $\boldsymbol{\mathcal{F}} \in \mathbb{R}^{6\times 1250}$ and $\boldsymbol{\mathcal{F}}$ is based on the first 1250

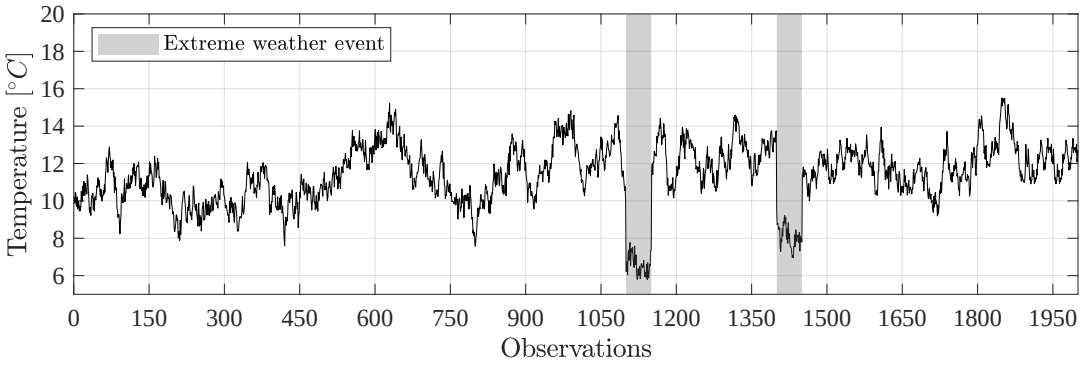


Figure 4.5: Simulated temperature field with sudden temperature drops highlighted with a grey background

Table 4.1
Summary of observations used for training, testing and damage detection in the six DOF mass-spring-damper system

Scenario	Stiffness reduction	Stiffness parameter	Number of observations
Healthy			
Training $\boldsymbol{F}_{hl,tr}$		-	1 to 1250
Extreme Tmp$_1$. $\boldsymbol{F}_{hl,tmp_1}$		-	1100 to 1150
Extreme Tmp$_2$ $\boldsymbol{F}_{hl,tmp_2}$		-	1400 to 1450
Damage $\boldsymbol{F}_{dm}$	5%	k_3	1951 to 2000
Total			2000

observations from a healthy system (without stiffness reduction).

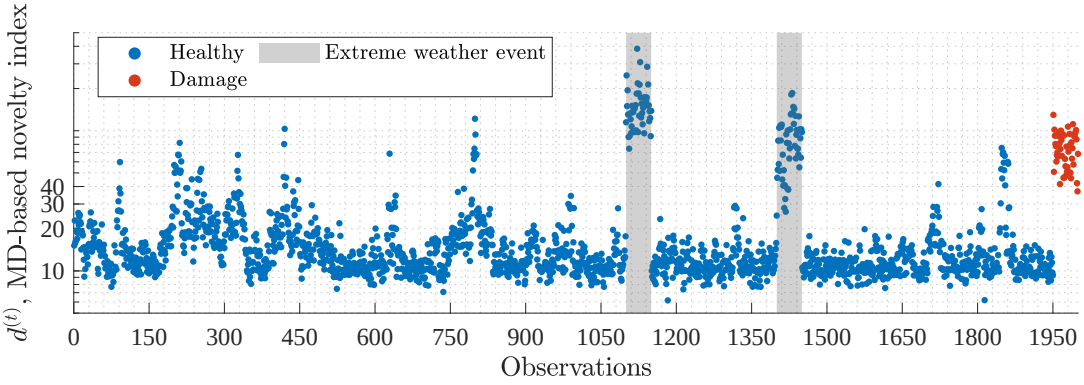


Figure 4.6: MD-based damage index

The sudden temperature drops by 4 ° C implemented are clearly visible between observations 1100 - 1150 and 1400 - 1450. These are overlapping with the simulated damage and hinder damage detection by causing false positives. The size of the training dataset is determined by the number of observations that would include the first extreme weather event yet purposely exclude the second extreme weather event.

The previous limitation discussed in section 3.3 and Fig. 3.6(a) are highlighted for damage detection in this numerical simulation.

4.2.2 Interpretability through SHAP

To understand the difference between novelties caused by the stiffness reduction and the influence of temperature effects, a regression model is built taking the temperature and DSFs as predictors and the MD-based DI as the prediction target. The XGBoost regression model is trained as described in section 4.1.1 using observations from the healthy and damaged states of the system. The aim of adopting a regression model in this study is to maintain a high prediction accuracy of the regression model to reproduce the MD-based DI. For this, the hyperparameters (i.e., maximum depth, minimum child weight, gamma, subsample, colsample by tree, alpha for L1 regularisation, and learning rate) are identified by a random search. The search is performed until a performance accuracy $R^2 > 0.99$ is reached, where R^2 is the square of the sample correlation coefficient between the predicted and the calculated MD as per Eq. 3.2. The calculated and predicted MD-based novelty indices are shown in Fig. 4.7. The parameters are summarised in Table 4.2. For each randomly selected value in a given range a maximum of 1000 iterations is set and the training stops if the mean squared error does not improve for 50 iterations to avoid over-fitting.

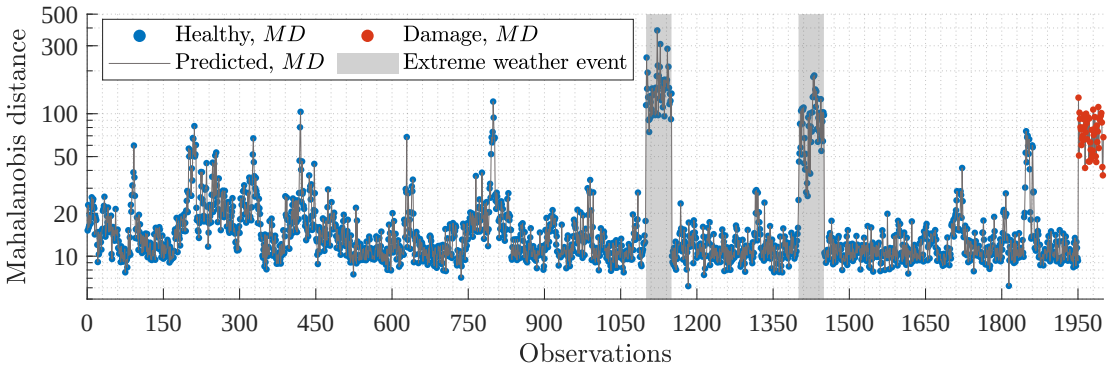
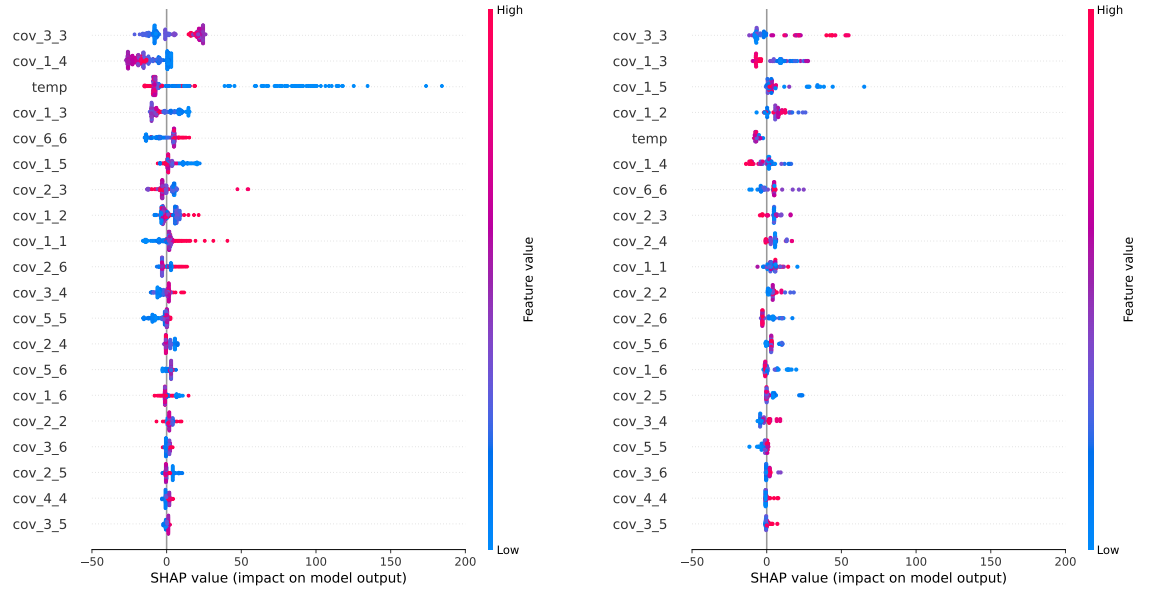


Figure 4.7: Predictions from the XGBoost regression model for the 6-DOF mass-spring-damper system

Table 4.2
Summary of hyperparameters and ranges chosen for the random search

Hyperparameter	Description	From	To	Interval	Best parameters
Maximum depth	The maximum depth of which each tree is built	1	10	1	7
Minimum child weight	The minimum sum of instance weight of all the observationsrequired in a child	1	5	1	3
Gamma	Controls the minimum loss reduction required to make a node split	0	1	0.1	0
Subsample	The number of observations randomly sampled at each tree	0.1	1	0.1	0.4
Colsample by tree	The number of features selected to build each tree	0	1	0.1	0.6
Regularisation alpha	L1 regularisation value on weights	0.01	0.3	0.01	0.19
Learning rate	Step-size for updating the weights	0.01	0.3	0.01	0.22

Figure 4.8(a) summarises the Shapley values obtained for the 6-DOF lumped mass-spring-damper system XGBoost model for observations 1 to 1950, while Figure 4.8(b) refers to the summary of observations 1950 to 2000. The features are ranked in descending order. It can be seen that the variance of sensor 3 has the largest marginal contribution to the model's prediction and the covariance of sensor 3 and 5 the lowest. On the horizontal axis, the contribution of the feature can be observed indicating a positive or a negative impact on the model's prediction. For example, the temperature feature in Figure 4.8(a) shows that low values (coloured blue) have a positive impact on the model's prediction, i.e., these increase the model's predicted index, whilst lower values have a negative impact, decreasing the model's predicted index. This suggests that low temperatures cause a high MD-based DI and vice-versa. Looking at the last 50 observations in Figure 4.8(b), where the stiffness reduction is introduced,



(a) Observations 1 to 1950 without implemented stiffness reduction

(b) Observations 1950 to 2000 with stiffness reduction

Figure 4.8: Summary contribution plots

the contribution of the temperature to these predictions drops from rank 3 to 5. This indicates that the predictions of these observations are not dominated for the influence of the temperature field on a global scale.

Looking at the local Shapley values ϕ_{temp} for the temperature in Figure 4.9, it can be observed that observations 1100 - 1150 and 1400 - 1450 have particularly high Shapley values. This indicates that these observations are caused by temperature. Conversely, observations 1950 to 2000 have a negative impact on the model's prediction indicating that these novelties did not occur due to the influence of temperature.

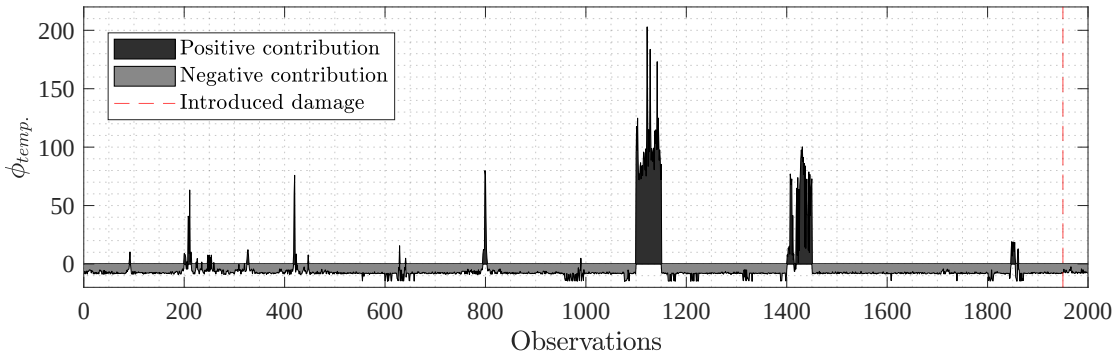


Figure 4.9: Shapley explanation for temperature in the 6-DOF mass-spring-damper system XGBoost regression model

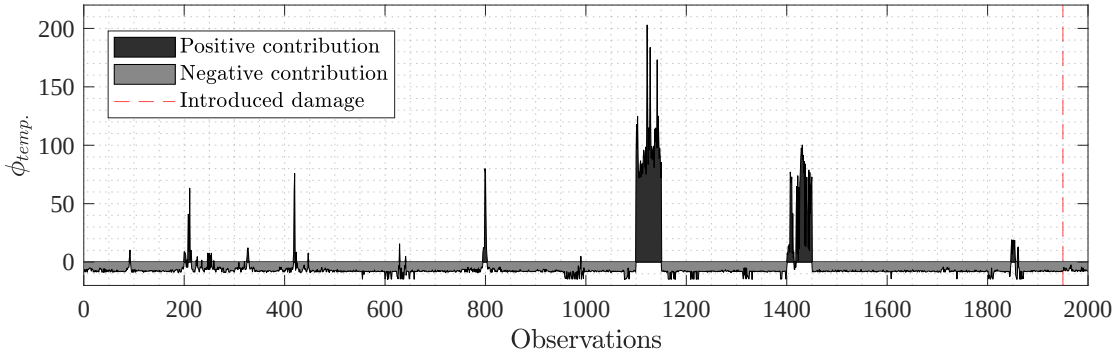
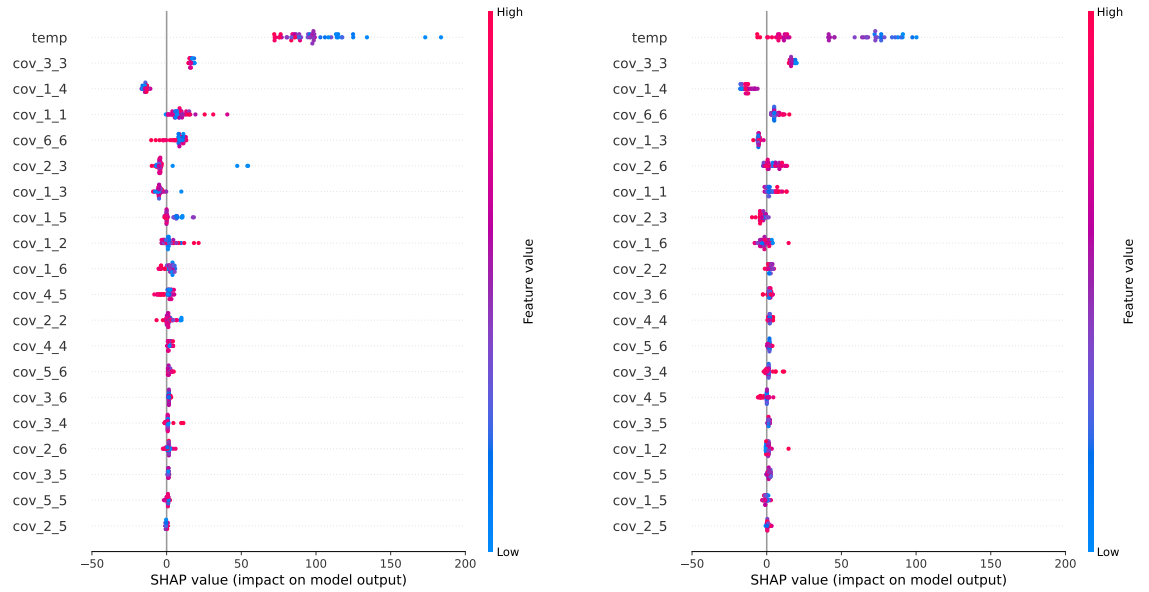


Figure 4.10(a) and 4.10(b) show the Shapley values for observations 1100 - 1150 and 1400 - 1450 respectively in a global impact overview of the features contribution. In these figures, it is clear that the temperature is the most influential feature for the predictions. The temperature drops of 4°C introduced into the 6-DOF simulated system, which were responsible for the increased DI, can be identified as false alarms by observing the Shapley values.



(a) Observations 1100 - 1150

(b) Observations 1400 - 1450

Figure 4.10: Summary contribution plot for extreme weather events

4.3 Discussion

This chapter addresses the challenge of interpretability in data-driven damage detection, particularly when ML techniques are implemented. Based on a numerical simulation of a 6-DOF lumped mass-spring-damper system, the use of the XGBoost model to build the relationship between DSFs, EOVs and MD-based damage indices is demonstrated. Additionally, SHAP was used complementarily to calculate the marginal contribution of each of the features to the model's outputs.

This process provided transparency and eased the understanding of the influence of temperature on the damage indices. The SHAP method estimated the marginal contributions of EOVs such as temperature and enabled the identification of false positives.

The results suggest that the proposed framework can strengthen the confidence in the decision-making process and contribute to the widespread adoption of ML techniques for damage detection in industrial applications. However, this framework requires further testing and demonstration beyond numerical simulations. Testing the framework on a lab-scale experiment or real application is highly advisable.

Chapter 5

Application example on a full-scale wind turbine

The two frameworks for robust damage detection and interpretability through XGBoost and SHAP presented in Chapter 3 and 4 respectively were evaluated on an operating wind turbine blade from a Vestas V27 wind turbine. This chapter presents the results from this evaluation and describes the experiment in detail. Ultimately, the combination of these two frameworks is presented for a robust ML-based and interpretable novelty index.

5.1 Vestas V27 wind turbine

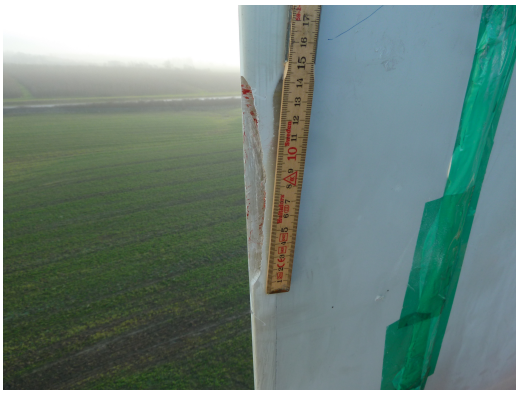
Between November 28$^{\text{th}}$, 2014 and March 12$^{\text{th}}$, 2015 a Vestas wind turbine (WT) model V27 was subjected to an experiment set up for damage detection in one of its WTBs. The WT, with a nominal power of 227 kW and a 13m blade length, has two operational regimes: 32 and 43rpm. Tcherniak and Mølgaard [152] equipped

one of the blades of the WT shown in Fig. 5.1(a) with 11 monoaxial piezoelectric accelerometers (Brüel & Kjær Type 4507B). The accelerometers and their cables were placed on the downwind side of the blade. The accelerometers were mounted using plastic mounting clips and the cables were secured with silicone and helicopter tape. The accelerometers were connected to a data acquisition system Brüel & Kjær Type 3660-C with two LAN-XI modules, a 12-channel input module Type 3053-B-120 and 4-channel input/output module Type 3160-A-042. For the experiment, the blade was excited by an electro-mechanical actuator mounted 1m away from the root on the surface of the blade for practical reasons including more efficient cabling management and reduced exposure to centrifugal force during operation. Controlled by the signal from the signal generator built into one of the data acquisition modules, the actuator launches a plunger towards the surface of the blade. During each pulse, a 100 µF capacitor discharges through the coil and is charged again to 48 V using a DC/DC converter before the next impulse. The blade's vibrations caused by the actuator's impact were measured by the accelerometers mounted on the blade.

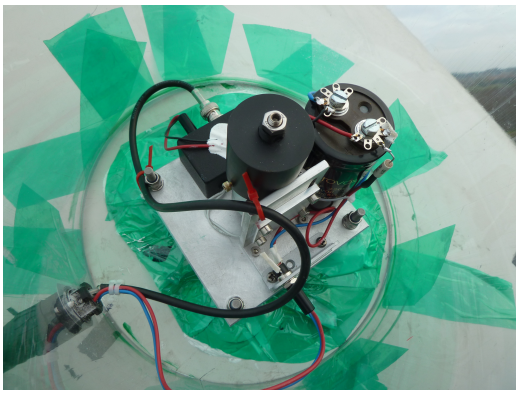
The data acquisition system and the electronics were placed in a waterproof box (dimensions 60 x 45 x 20 cm^3 and weight 25 kg), which was mounted to the inner surface of the spinner. The equipment was powered by 24 V from the nacelle via a slip ring. The recorded data were transmitted from the rotating part to the nacelle through two Cisco wireless access points. One access point was located inside the waterproof box and another in the nacelle. To address the possible blockage of the signals by the steel parts of the hub during the operation of the turbine, two pairs of antennas were employed, namely two omni-directional antennas attached to the hub and two directional antennas mounted inside the nacelle.

Meteorological data were simultaneously collected from a weather mast located nearby. The weather data averaged within 1-min intervals included temperature,

(a) Vestas V27 wind turbine with a hub height of 33.5 m and a 27m rotor diameter



(b) First damage implemented by a 15cm trailing edge opening around 6m from the tip of the blade



(c) Electromagnetic actuator mounted close to the root of the blade for mechanical excitation

Figure 5.1: Full-scale experimental setup

wind speed and direction, wind turbulence at different altitudes, atmospheric pressure and precipitation. The power production data and yaw angle were also available from the wind turbine system.

During the monitoring campaign, three levels of damage severity were artificially introduced. The damage takes the form of trailing edge delaminations and the levels of severity were openings of 15cm, 30cm and 45cm respectively. Figure 5.1(b) represents the first level of severity introduced on December 9^{th}, 2014, i.e., roughly after 1.5 weeks of monitoring the undamaged blade. The next two levels of severity were

introduced on December 15th, 2014, when the delamination was extended from 15cm to 30cm and January 6th, 2015, when the delamination was extended from 30cm to 45cm. Finally, the blade was repaired on January 19th, 2015 and monitored for the remaining duration of the campaign.

The data acquisition system controlled by Brüel & Kjær PULSE LabShop software started recording the responses 10 seconds before the actuator hit the WTB and recorded for 20 seconds thereafter with a sampling frequency of 16384 Hz. Thus, the number of samples corresponding to a hit from the actuator is approx. 500,000 for the 30 seconds of recording. One actuator hit corresponds to one observation. During the monitoring campaign, a total of 24,693 observations (recorded actuator hits) were recorded considering a healthy, damaged and a repaired WTB. To exemplify the vibration response, Fig. 5.2 shows the signal from a sensor during operation where the blade is excited after 10s of recording. Each measured sensor response recorded was filtered by a band-pass filter with cut-off frequencies of 700 and 1200 Hz [152]. The filtered response was then trimmed and aligned to the 300 samples that correspond to the highest amplitudes during the actuator hit.

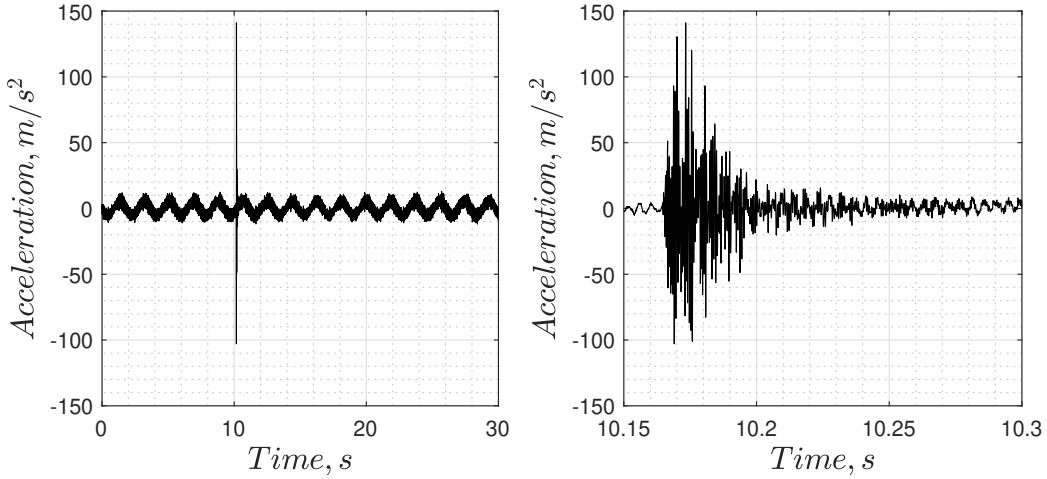


Figure 5.2: V27 signal response during operation including an actuator hit. This signal is uncut and unfiltered.

In this study, 4 accelerometers from the leading edge and 4 accelerometers from the trailing edge (see Figure 5.3) are considered. Due to the limited availability of recordings during the undamaged state of the blade, the recordings from the repaired blade are considered as healthy for further evaluation. The repaired blade was monitored for approx. two months. This extended monitoring period (relative to the 1.5 weeks of monitoring the undamaged blade) includes various environmental and operational variabilities. Furthermore, the analysis is performed considering the recordings during the 32rpm operational mode, which totals 2927 observations.

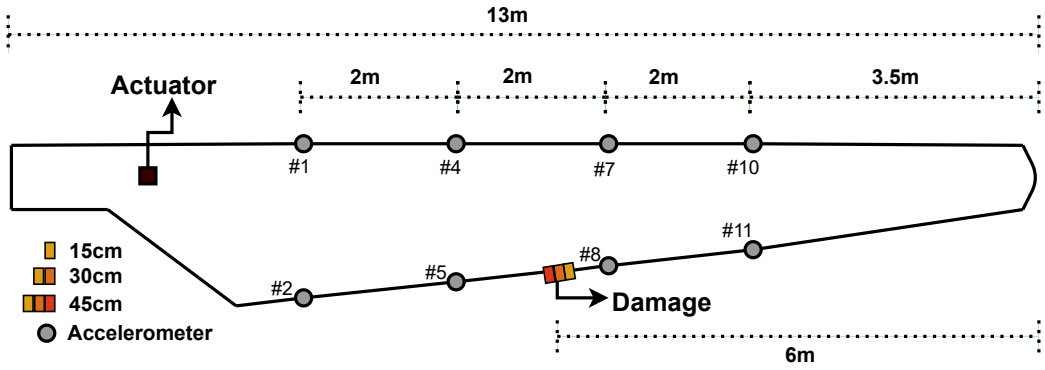


Figure 5.3: Sketch of the equipment on the blade of the Vestas V27 wind turbine

5.1.1 Data partitioning and DSF extraction

As DSFs, the cross-covariance from each actuator hit between the accelerometer responses is calculated. The covariance is used as a measure of similarity between two signals, thus if a structural defect is acquired, the energy propagation from the actuator hit to the accelerometer will change its pattern. The cross-covariance matrix $\mathbf{\Sigma}_{\mathcal{F}} \in \mathbb{R}^{a \times a}$ is symmetric, therefore, unique values from the diagonal and upper triangle are obtained resulting in DSFs vectors with dimension $\frac{a(a+1)}{2}$, where a is the

number of accelerometers used in the analysis. Considering the 8 accelerometers previously mentioned and shown in Figure 5.3, the resulting DSF vector is $\boldsymbol{f}^{(k)} \in \mathbb{R}^{36\times 1}$ with 36 unique covariance entries. The DSF matrix consists of 2927 observations with $\boldsymbol{F} \in \mathbb{R}^{36\times 2927}$. The partition of the data for training and testing is summarised in Table 5.1.

Table 5.1
Partition of data for an output-only methodology: Vestas V27 data set

	Healthy		Damage			Total $\boldsymbol{F}$
	Training $\boldsymbol{F}_{hl,tr}$	Testing $\boldsymbol{F}_{hl}$	$\boldsymbol{F}_{dm15}$	$\boldsymbol{F}_{dm30}$	$\boldsymbol{F}_{dm45}$	
Observations	2000	639	66	117	105	2927

The observations are divided into training data $\boldsymbol{F}_{hl,tr} \in \boldsymbol{F}^{36\times 2000}$ and testing data $\boldsymbol{F}_{ts} \in \boldsymbol{F}^{36\times 927}$. The testing set contains a set of healthy data $\boldsymbol{F}_{hl} \in \boldsymbol{F}_{ts}^{36\times 639}$ and three states of damage, namely 15cm damage $\boldsymbol{F}_{dm15} \in \boldsymbol{F}_{ts}^{36\times 66}$, 30cm damage $\boldsymbol{F}_{dm30} \in \boldsymbol{F}_{ts}^{36\times 117}$ and 45cm damage $\boldsymbol{F}_{dm45} \in \boldsymbol{F}_{ts}^{36\times 105}$. The training dataset $\boldsymbol{F}_{hl,tr}$ (i.e., the first 2000 observations) corresponds to recordings at the beginning of this period, whilst the healthy testing dataset $\boldsymbol{F}_{hl}$ (i.e., the remaining 639 observations) corresponds to the end of this period and has no influence on the training.

5.2 MD-based novelty index

First, the MD-based novelty index $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ is calculated and presented in Fig. 5.4 with a reference set $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{36\times 2000}$. Several deviations can be observed around observations 500, 1200 and, more specifically, at the end of the training set after 2000 observations. Damage 1 with 15cm trailing edge opening is overlapping with $\boldsymbol{F}_{hl}$ and some parts of the training data. Damage 2 with 30cm and damage 3 with 45cm trailing edge opening are well separated from the healthy data as well as from each

other. The damage extension could be related to the increase of the MD.

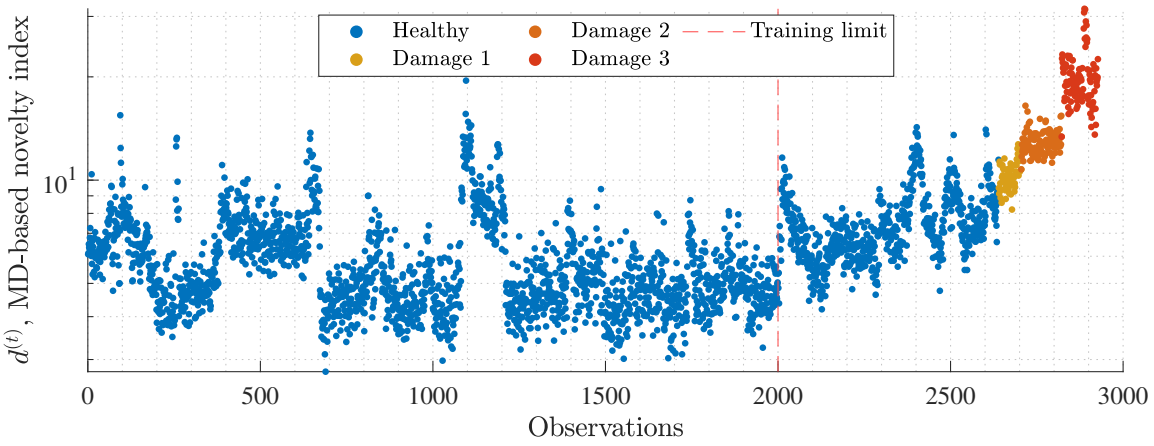
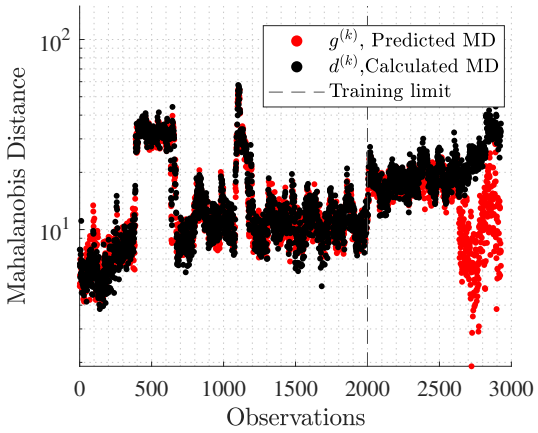


Figure 5.4: MD-based damage detection V27 calculated with healthy data (obtained from a repaired WTB) and three introduced damages

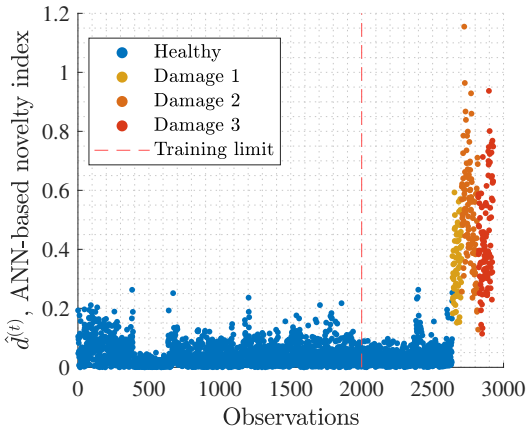
5.3 ANN-based novelty index

The ANN-based novelty index is calculated in the same fashion as described and demonstrated in Chapter 3. The DSFs are labelled by the MD $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ with $\mathcal{F} \in \boldsymbol{F}_{hl,tr}^{36\times200}$, which refers to the first 200 observations. The ANN regression model is built with the DSFs $\boldsymbol{f}^{(k)}$ as predictors and their label (or learning target) $d^{(k)}$, where $k : [1 \leq k \leq 2000]$. Fig. 5.5(a) shows the calculated MD and the ANN prediction. The shift after 2000 observations is well-predicted by the proposed methodology as highlighted in Fig. 5.5(b). Significant deviations in the predictions occur after observation 2639, which corresponds to DSFs from a damaged WTB.

The overlap between the novelty indices from healthy and damaged states of the WTB was reduced, increasing the performance for all evaluation metrics as shown in Table 5.2. Significant improvement can be seen for the smallest damage with 15cm trailing edge opening, which increased close to 1. These results highlight the high



(a) Calculated MD $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ and the predicted $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ by ANN



(b) Prediction error $\hat{d}^{(k)}$ obtained by Eq. 3.8 for experimental data used as new novelty index (ANN-based novelty index) for damage detection

Figure 5.5: ANN-based novelty index

Table 5.2
Classification results with $\boldsymbol{F}_{hl}$ as healthy and the three damage cases $\boldsymbol{F}_{dm15}$,$\boldsymbol{F}_{dm30}$,$\boldsymbol{F}_{dm45}$ for damage detection in the V27 wind turbine

	MD-based			ANN-based*		
Performance metric	Damage 1	Damage 2	Damage 3	Damage 1	Damage 2	Damage 3
Accuracy	0.86	0.97	0.99	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
F_1-score	0.56	0.90	0.98	0.94 (0.03)	0.99 (0.01)	0.97 (0.02)
MCC	0.56	0.89	0.93	0.93 (0.04)	0.99 (0.01)	0.97 (0.02)

*ANN-based results computed by 1000 repetitions with the corresponding standard deviation in parentheses

prediction accuracy for the test-set $\boldsymbol{F}_{hl}$, where variabilities in the MD-based novelty index were well-predicted by the ANN. On the other hand, a bigger overlap in the different damage cases is observed, limiting the identification of damage propagation. For the MD-based novelty index, the accuracy decreases to 0.86, the F_1-score to 0.56 and the MCC to 0.56 as presented in Table 5.2 for the smallest damage case.

5.4 Interpretable MD-based novelty index

To understand the potential cause of these novelties and how they differ from the identified damage after observation 2639, an XGBoost regression is built. DSFs and measured EOVs such as temperature, wind direction, wind speed and rotational speed at the high and low speed shaft are included in the feature space. The XGBoost regression model from section 4.1.1 is adopted with the hyperparameters shown in Table 5.3. The model's predictions with an $R^2 > 0.99$ are shown in Fig. 5.6.

Table 5.3
Summary of hyperparameters and ranges chosen for the random search

Hyperparameter	Description	From	To	Interval	V27 - Best parameters
Maximum depth	The maximum depth of which each tree is built	1	10	1	8
Minimum child weight	The minimum sum of instance weight of all the observationsrequired in a child	1	5	1	1
Gamma	Controls the minimum loss reduction required to make a node split	0	1	0.1	0
Subsample	The number of observations randomly sampled at each tree	0.1	1	0.1	0.6
Colsample by tree	The number of features selected to build each tree	0	1	0.1	0.5
Regularisation alpha	L1 regularisation value on weights	0.01	0.3	0.01	0.08
Learning rate	Step-size for updating the weights	0.01	0.3	0.01	0.14

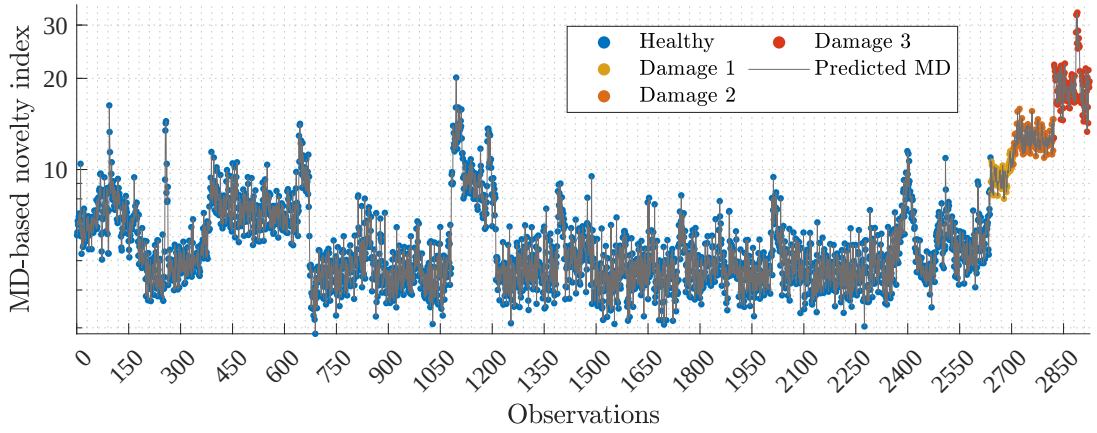
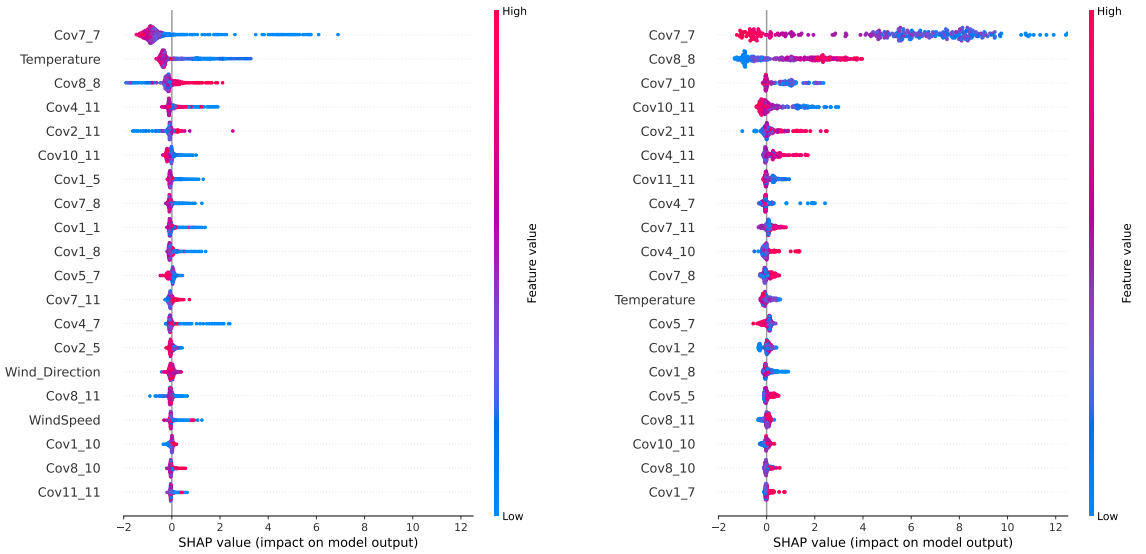


Figure 5.6: Predictions from the XGBoost regression model for the V27 MD-based DI



(a) Observations 1 - 2639 for the healthy state of the WTB

(b) Observations 2639 - 2927 with artificially introduced damage

Figure 5.7: Summary contribution plots

The global influence, i.e., the Shapley values, of the DSFs and EOVs are shown in Figures 5.7(a) and 5.7(b). In these figures, only the 20 highest contributing DSFs to the prediction of the model are displayed and, thus, further discussed in this section. The Shapley values presented for the healthy state of the WTB in Figure 5.7(a) show that the variance of accelerometer 7 has the highest influence on the model's predictions, followed by temperature. This indicates that the DI is highly influenced by temperature while the turbine has no artificial damage introduced. This is consistent with the SHAP results obtained from the simulation in section 4.2, where the drops in temperature led to an increase in the MD-based DI. EOVs such as wind direction and wind speed have a rather small contribution to the model's prediction. Conversely, the significance of temperature drops for observations 2639 - 2927 as shown in Fig. 5.7(b). The majority of Shapley values corresponding to temperature for observations during the damaged state of the WTB fluctuate around or below zero indicating a low contribution to the model's prediction. This suggests that these novelties are not caused by the influence of temperature.

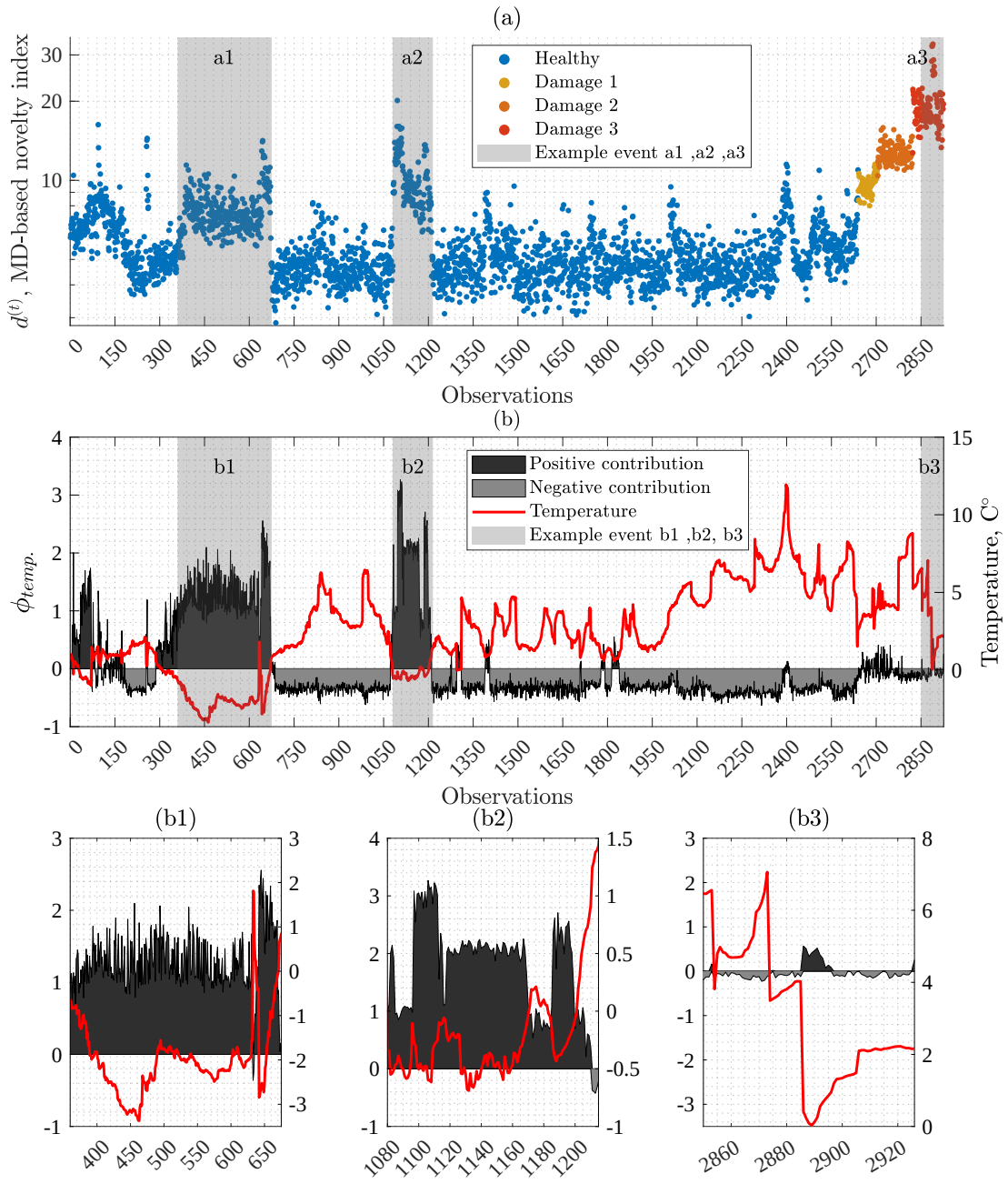
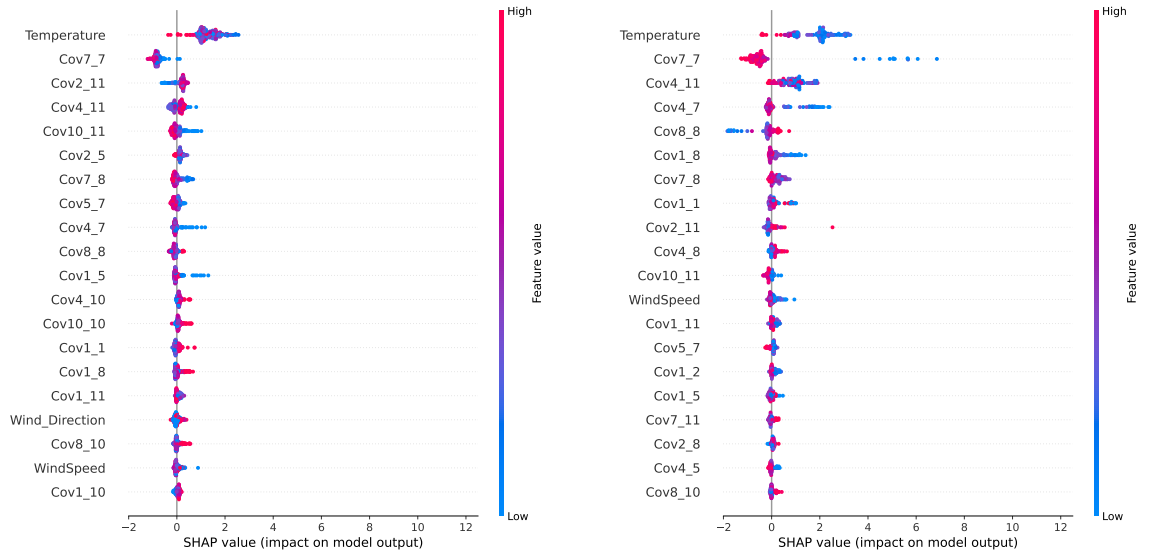


Figure 5.8: SHAP based interpretability of environmental parameters. (a) MD-based novelty index. (b) Recorded temperature during the measurement campaign and the corresponding Shapley values. (b1-b3) Zoomed view of potential areas of correlation between the MD, temperature and Shapley values



(a) Observations 360 - 675 where the first significant outliers were identified

(b) Observations 1080 - 1215 where second outliers were identified

Figure 5.9: Summary contribution plots

For local interpretability and to gain insight into the contribution of the temperature measurements to the model's predictions, Fig. 5.8(a) and Fig. 5.8(b) can be compared. It can be seen that the temperature values that are close to or below zero have a positive contribution towards the prediction suggesting that low temperatures lead to a high MD-based DI. Looking into the previous false alarms identified between observations 360 - 675 and 1080 - 1215 in Fig. 5.8(a), it is clear that these outliers are influenced by the positive contribution, i.e., Shapley values, of temperature to the model's outputs. The Shapley values for temperature within these ranges of observations, i.e., 360 - 675 and 1080 - 1215, are the highest among all features as seen in Fig. 5.9(a) and Fig. 5.9(b). Furthermore, when looking at the observations after 2639, when the artificial damage was introduced, the temperature contribution is virtually negligible with Shapley values fluctuating around zero.

False alarms are also present at around observation 2400 as can be seen in Fig. 5.8(a). In Fig. 5.8(b) can be seen that the temperature peak at 12°C correlates with

these false alarms. However, in Fig. 5.9(b), the contribution of temperature to the predictions is rather minor relative to, e.g., the contributions of observations 1200 - 1300. Fig. 5.10 shows that there were four other DSFs with higher contributions than temperature for the cases of observations 2380 - 2420. This suggests that EOVs which were not measured or explicitly considered in the model may have had an influence on the novelty index. This represents a source of uncertainty and thus a limitation in the proposed framework. A temperature of 12°C occurred only once during the measurement campaign. ML techniques are based on pattern recognition, i.e., a particular behaviour or instance has to occur multiple times before the algorithm is able to recognise a pattern. Thus, a longer measurement campaign encompassing a wider range of environmental conditions may be necessary to address this uncertainty and provide clearer interpretations of the predictions.

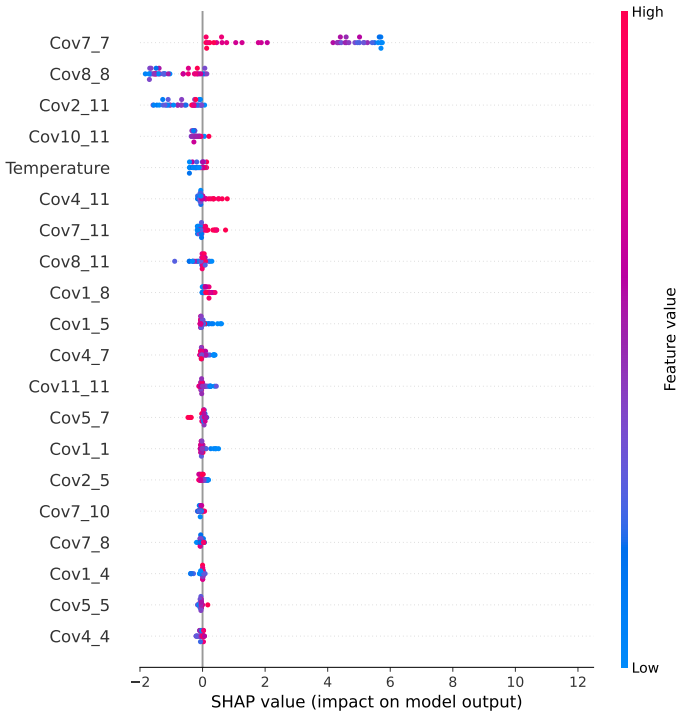


Figure 5.10: Summary contribution plot for observations 2380 - 2420

Further, observing less significant environmental measurements such as the wind

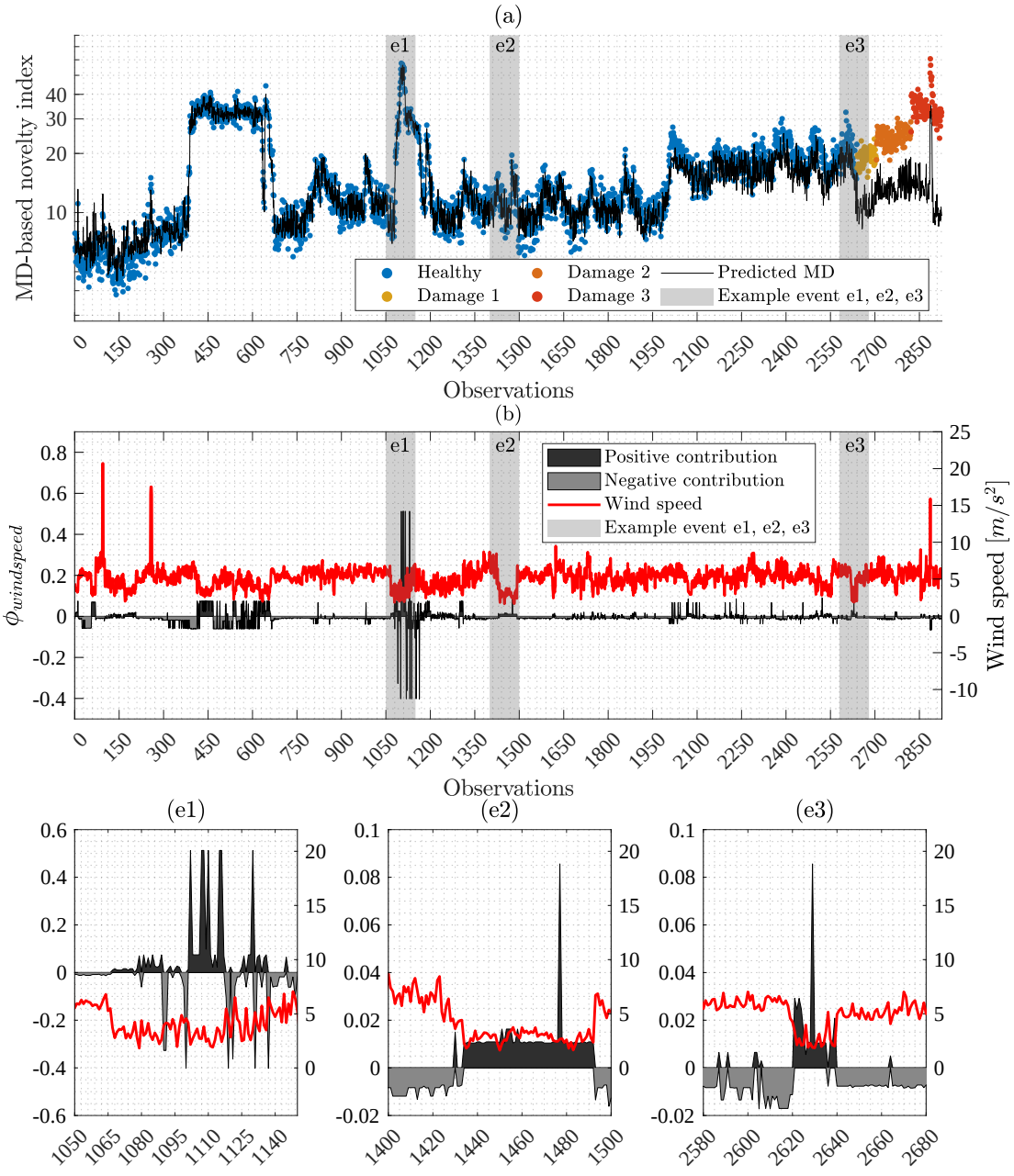


Figure 5.11: SHAP-based interpretability of environmental parameters. (a) MD-based novelty index. (b) Recorded wind speed during the measurement campaign and the corresponding Shapley values. (b1-b3) Zoomed view of potential areas of correlation between the MD, wind speed and Shapley values.

speed shown in Fig. 5.11, it can be observed that extreme wind speeds such as the one highlighted in the event $e1$ can lead to outliers. Wind speed, in general, is not a high contributor to the model's prediction on its own. However, extreme wind speeds can cause both turbulence and lower temperature at the blade and, thus, can influence the vibration response leading to the observed outliers. This correlation between high wind speed and the positive contribution towards the model's prediction can be seen at around observation 95 in Fig. 5.8(e1).

Outliers can also be caused by wind speeds close to the cut-in speed of the turbine (i.e., 3.5 m/s) as can be observed in the event $e2$, where an increased contribution of the wind speed can be observed between observations 1070 and 1130 in Fig. 5.8(e2). A similar phenomenon is visible in event $e3$, where a wind speed drop from 5 to 3 m/s leads to a positive contribution towards the prediction of the MD-based novelty index.

Interpreting the contribution of wind direction to the model's prediction is not so straightforward relative to the cases of temperature and wind speed. No consistent relationship can be observed between the measured wind direction and the contribution to the model $\phi_{winddir.}$ throughout the measurement campaign. Changes over time in the effect on the novelty index are observed. For instance, looking at the event $e1$ in Fig. 5.12(b), a decrease in novelty index can be observed while the wind direction changes from approx. 210°to 175°. A spike in positive contribution $\phi_{winddir.}$ can be found at around observation 96, which correlates with the spike in the novelty index observed in the same area in Fig. 5.12(a). Yet, it was previously discussed that this spike in the novelty index is mainly caused by the spike in wind speed observed in Fig. 5.11(e1).

When examining the event $e2$ in Fig. 5.12(a), the spike in the novelty index at approx. observation 1100 is caused by a positive contribution of the wind direction $\phi_{winddir.}$

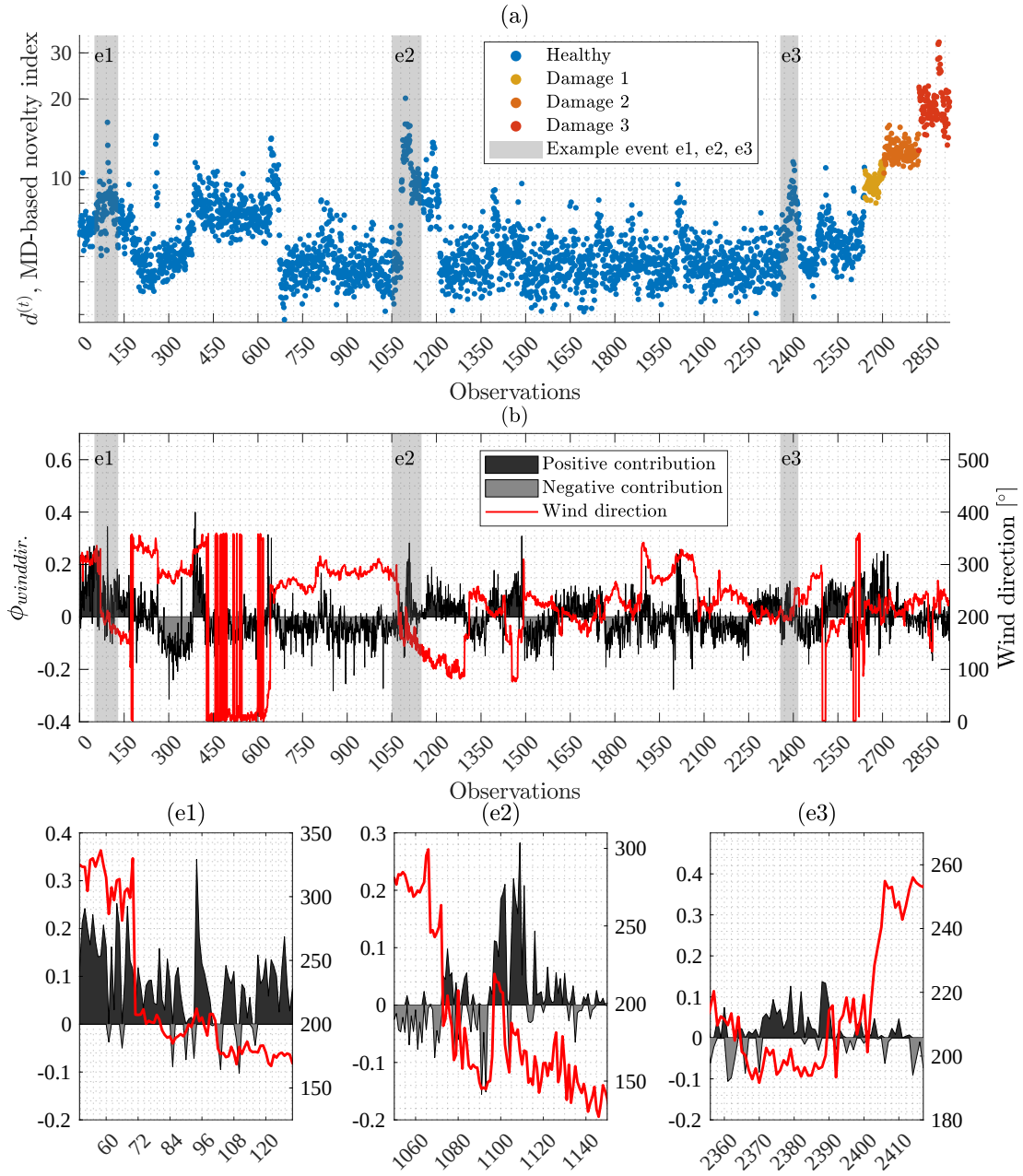


Figure 5.12: SHAP-based interpretability of environmental parameters. (a) MD-based novelty index. (b) Recorded wind direction during the measurement campaign and the corresponding Shapley values.

observed in Fig. 5.12(b). However, as discussed previously, this spike in the novelty index was also influenced by a wind speed close to 3 m/s, i.e., the cut-in speed of

the turbine. Further, in the example event $e3$ in Fig. 5.12(a) at approx. observation 2400, a spike in the novelty index correlates to the sudden change in wind direction from approx. 200°to 270°observed in Fig. 5.12(b). This correlation, however, is not reflected in $\phi_{winddir.}$. In this case, the spike in the novelty index is rather explained by the highest temperature recorded during the monitoring campaign (as seen in Fig. 5.8(b)).

Using wind direction as a DSF without knowing the position of the nacelle can be misleading as no clear and consistent relationship could be observed throughout the measurement campaign. Nonetheless, the XGBoost model prioritises the global contribution of the wind direction over the wind speed in some cases as observed in Fig. 5.7(a). The expertise and domain knowledge of the operator is required to evaluate the contribution of such DSF and to question whether such behaviour is senseful from a physics point of view. However, the proposed approach will serve as a support to an operator for the the decision-making process.

5.5 Combining frameworks: Improved and interpretable novelty index

To provide an interpretable and improved novelty index, the first framework presented in section 3.1 is modified by replacing the ANN regression model with the XGBoost regression model and adding the interpretability approach through SHAP as presented in section 4.1. This modification entails a trade off between accuracy and interpretability as discussed in section 2.3 and illustrated in Fig. 2.1.

In this modified framework, the implementation follows the procedure described in section 5.3 for the calculation of the ANN-based novelty index in the case of the V27

WT with the difference that the ANN-based novelty index is now the XGBoost-based novelty index. Additionally, as in the implementation presented in section 5.5, the feature space is extended by including measured EOVs such as rotational speed, wind direction and temperature as DSFs.

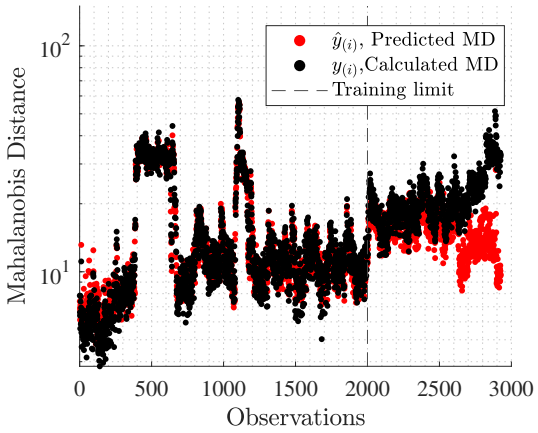
The XGBoost regression model is trained to build the relationship between the DSFs and the novelty index for the first 2000 observations. The hyperparameters found previously in Table 5.3 were not suitable for the prediction model. To be able to predict observations 2000 to 2639, which correspond to the healthy WTB, the search of the hyperparameters shown in Table 5.4 was performed in a heuristic manner. It was found that reducing the maximum depth of the tree to 3 and increasing the learning rate to 0.3 during the training, the model is less accurate yet was performing better when predicting observations which were not used for the training.

Table 5.4
Hyperparameters for the XGBoost model

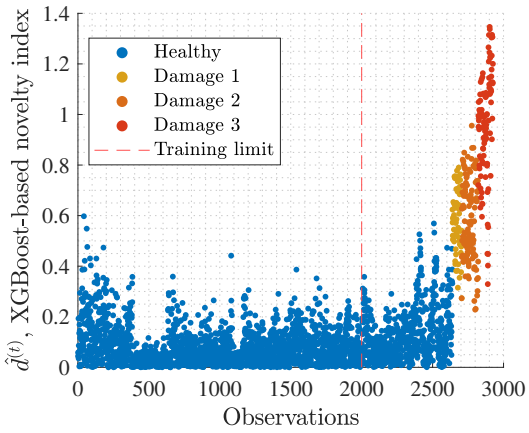
Hyperparameter	Description	Best parameters
Maximum depth	The maximum depth of which each tree is built	3
Minimum child weight	The minimum sum of instance weight of all the observations required in a child (control over-fitting)	0
Gamma	Controls the minimum loss reduction required to make a node split	0
Subsample	The number of observations randomly sampled at each tree	0,6
Colsample by tree	The number of features selected to build each tree	0,5
Regularisation alpha	L1 regularisation value on weights	0,3
Learning rate	Step-size on updating the weights	0,14

The calculated MD-based novelty index and the prediction error can be observed in Fig. 5.13(a). An increased prediction error occurs when the damage is present, similar to the implementation of the ANN-based framework shown in section 5.3 and Fig. 5.5(b). As shown in Table 5.5, the performance metrics show similar or slightly decreased results for damage types 2 & 3 compared to the MD-based novelty index. However, the performance improved significantly for the smallest damage of 15cm (damage 1).

Unlike when training the XGBoost model on all observations to interpret the novelty



(a) Calculated MD $d^{(k)}(\boldsymbol{f}^{(k)}, \mathcal{F})$ and the predicted $g^{(k)}(\boldsymbol{f}^{(k)}, \boldsymbol{W})$ by ANN



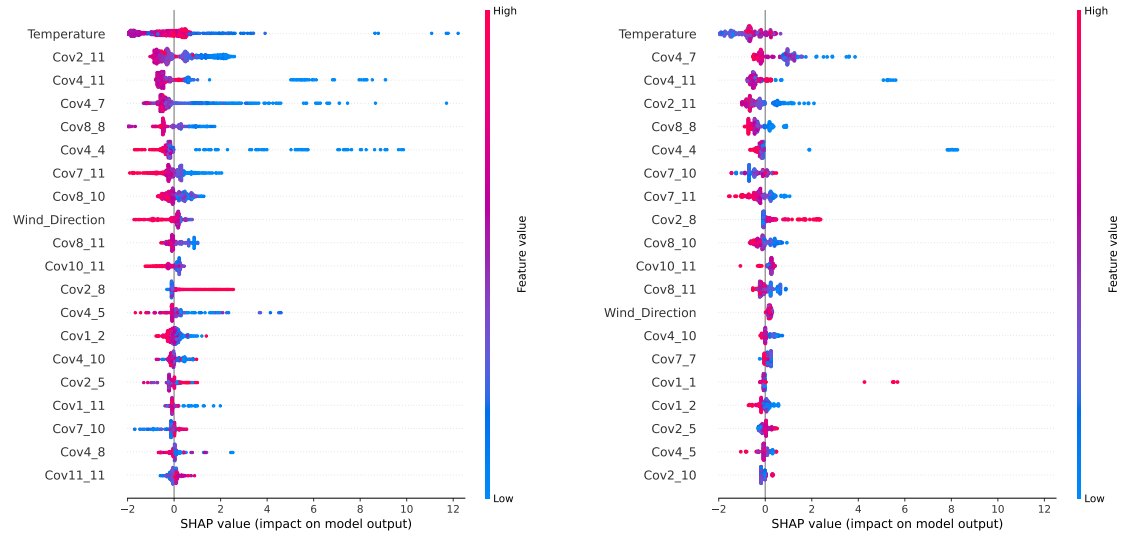
(b) XGBoost prediction error $\hat{d}^{(k)}$ obtained by Eq. 3.8, identical as for the ANN-based novelty index

Figure 5.13: Construction of ML-based novelty index

Table 5.5
Classification results with $\boldsymbol{F}_{hl}$ as healthy and the three damage cases $\boldsymbol{F}_{dm15}$,$\boldsymbol{F}_{dm30}$,$\boldsymbol{F}_{dm45}$ for damage detection in V27 wind turbine based on the XGBoost prediction error

	MD-based			XGBoost-based		
Performance metric	Damage 1	Damage 2	Damage 3	Damage 1	Damage 2	Damage 3
Accuracy	0.86	0.97	0.99	0.95	0.95	0.99
F_1-score	0.56	0.90	0.98	0.80	0.86	0.95
MCC	0.56	0.89	0.93	0.79	0.84	0.95

index as was done in section 5.5, training the model with a subset of 2000 observations influences the hierarchy of the DSFs as shown in Fig. 5.14. In Fig. 5.14(a), it can be observed that the temperature is the highest contributor to the model's prediction in the healthy state. Not surprisingly, this condition remains the same after damage is introduced since the model did not significantly change the weight of the features when predicting observations that had not been used for training. This consistent weighting results in misprediction when potential damage is present, which is the objective of this framework. This shows that the underlying idea of the semi-supervised framework presented in 3.1 that predicts the MD-based novelty index can be adopted with the XGBoost regression model.



(a) Observations 1 - 2639 for the healthy state of the WTB

(b) Observations 2639 - 2927 with artificially introduced damage

Figure 5.14: Summary contribution plots

The results in Fig. 5.14 show that low temperature values have a positive contribution to the model's prediction whilst high temperature values are close to zero or have a negative contribution. This is consistent with the implementation in section 5.5.

Examining the temperature field and its contribution over all observations, a few interesting findings can be observed. In the first 2000 observations, only a few temperature readings can be found at around 6°C. This can be limiting to a model that is seeking to identify a pattern and extrapolate this information for new observations. In Fig. 5.15(c), event $e1$ shows the highest recorded temperature reading at approx. 12°C. The contribution $\phi_{temp.}$ of this peak is negative, leading to a model prediction below the target value as can be seen in Fig. 5.15(a). Ultimately, this results in a higher novelty index that overlaps with the damage scenarios as can be seen in Fig. 5.15(b), which decreases the performance of this combined framework. A similar phenomenon can be observed in events $e2$ and $e3$.

Overall, the model was capable of adequately using the higher temperature values

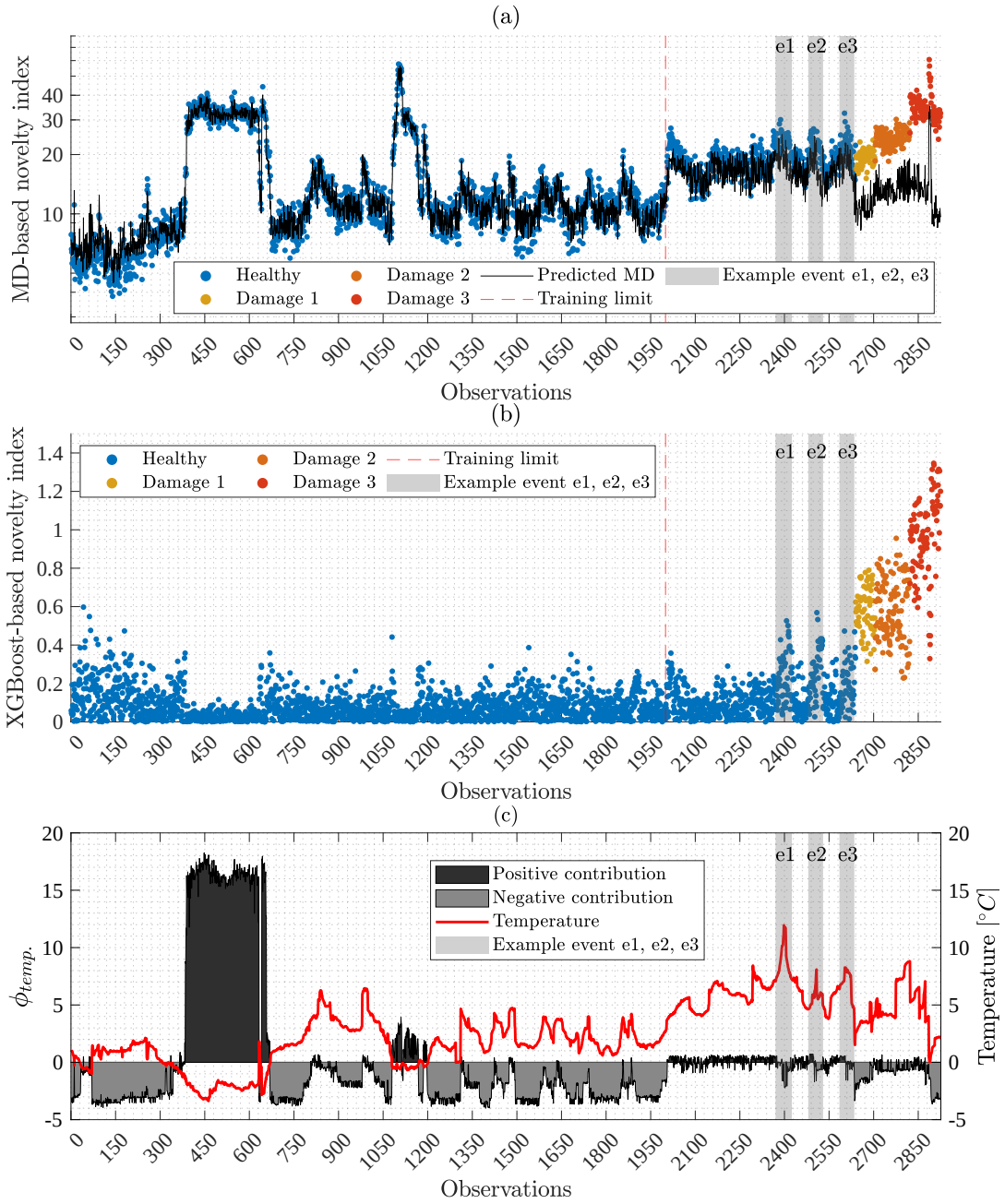


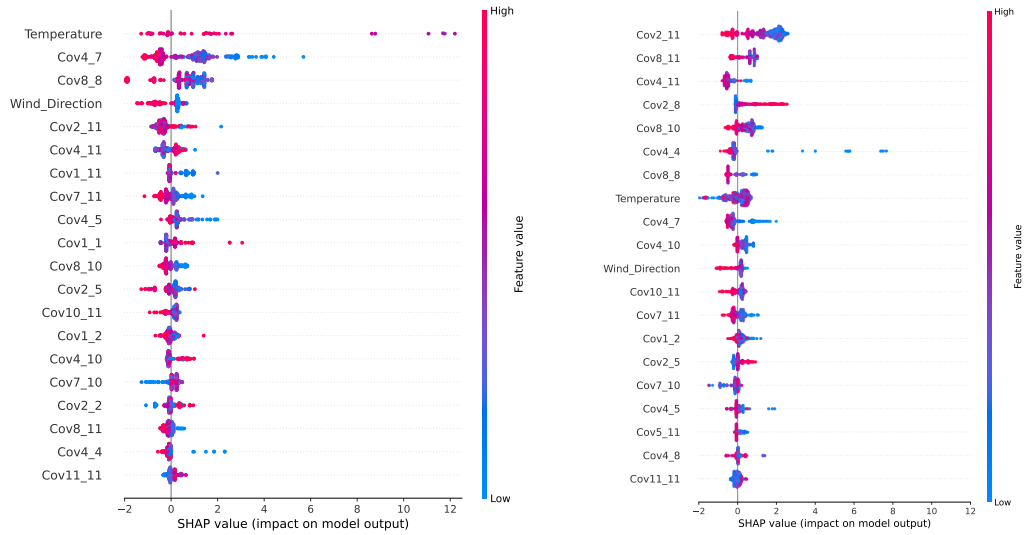
Figure 5.15: SHAP-based interpretability of environmental parameters. (a) MD-based novelty index and predicted MD. (b) XGBoost-based novelty index. (c) Recorded temperature during the measurement campaign and the corresponding Shapley values.

that were not used for training to predict new observations adjusting the contribution of this feature to the model's prediction. Between observations 1200 and 2000, the temperature readings range between -1°C and approx. 3°C and there is a predominantly negative contribution. Between observations 2000 and 2600, temperature rises up to the previously mentioned peak of 12°C and the contribution of this feature to the model's prediction increases.

Fig. 5.16(a) shows the contribution of the 20 most influencing features to predict the MD-based novelty index between observations 360 and 375. From this, it can be concluded that temperatures below the freezing point are mainly responsible for a high MD-based novelty index. An increase in the MD-based novelty index can be observed for observations 2000 - 2639, where a different range of temperatures from those registered during the training set were observed. In this case, as can be seen in Fig. 5.16(b), the contribution of temperature to the model's prediction decreases. This suggests that this feature is not properly taken into account to predict the MD-based novelty index and the model is not able to extrapolate values of the temperature that have not been used for training.

Wind speed has a relatively insignificant contribution to the model's prediction. This feature is not present in any of the summary contribution plots (Fig. 5.14 and Fig. 5.16). However, hints of similar behaviour to that described in section 5.5 are visible. Particularly, the similarities can be seen in events $e1$ - $e3$ in Fig. 5.17, where a positive contribution $\phi_{windspeed}$ is present when the wind speed drops close to the cut-in speed of the turbine (3.5 m/s).

For consistency, wind direction has been included as a DSF in this implementation. The previous findings of a lack of consistency in the relationship between this feature's contribution and the MD-based novelty index found in section 5.5 are also found here as can be seen in Fig. 5.18.



(a) Observations 360 - 675 where the first significant outliers were identified

(b) Observations 2000 - 2639 where the second outliers were identified

Figure 5.16: Summary contribution plots

During the training, between observations 0 - 400, the wind direction of approx. 300°mainly contributed to adjusting the prediction of the MD-based novelty index down towards the target. After training the model with the first 2000 observations, the previously described relationship is maintained. However, a higher novelty index is observed between observations 2000 - 2060 with a similar wind direction of 300°. This suggests that the negative contribution seen before between observations 0 - 400 should not be replicated and leads to a higher prediction error, as observed between observations 2000 - 2060 in Fig. 5.18.

5.6 Discussion

This chapter demonstrates the implementation of the frameworks presented in Chapter 3 and Chapter 4 on an operating wind turbine with artificially introduced damage

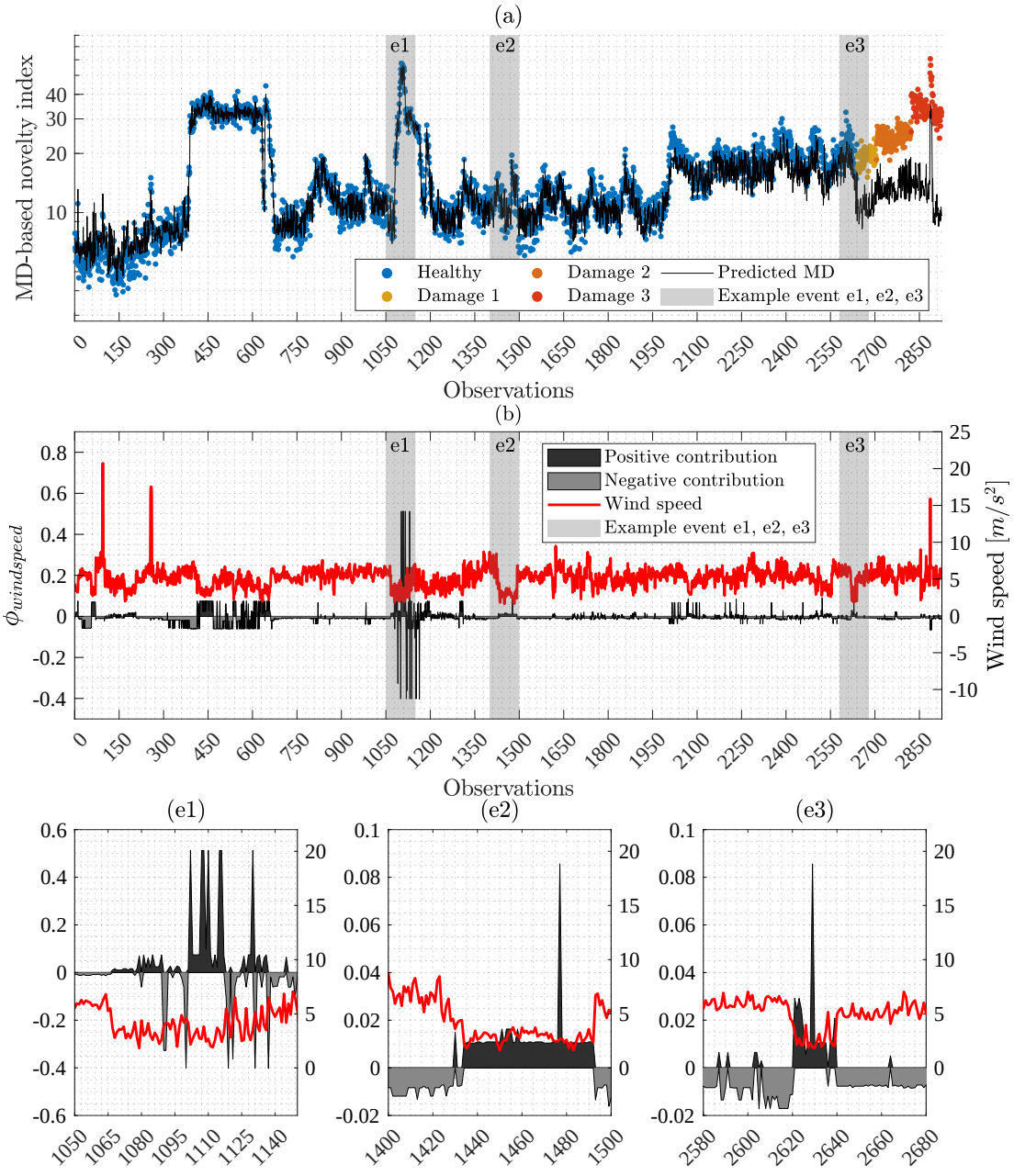


Figure 5.17: SHAP-based interpretability of environmental parameters. (a) MD-based novelty index and predicted MD. (b) Recorded wind speed during the measurement campaign and the corresponding Shapley values. (e1-e3) Zoomed view of potential areas of correlation between the MD, wind speed and Shapley values.

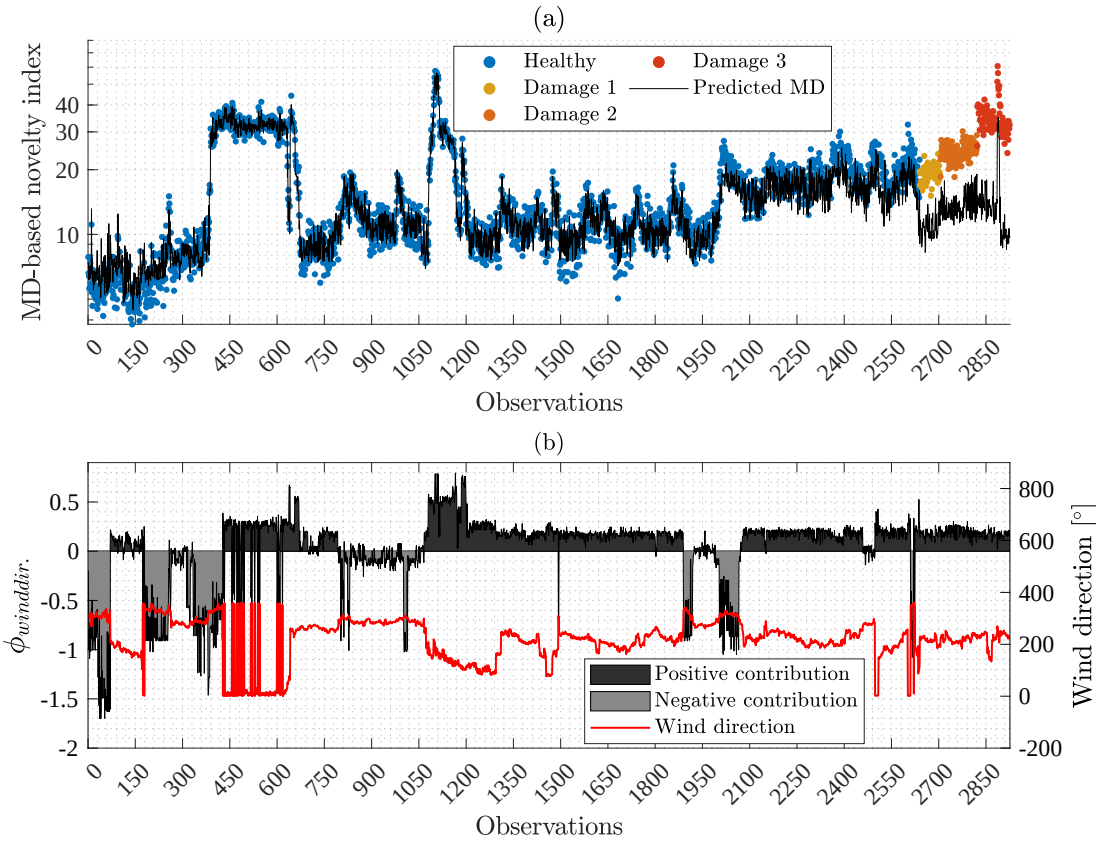


Figure 5.18: SHAP-based interpretability of environmental parameters. (a) MD-based novelty index. (b) Recorded wind direction during the measurement campaign and the corresponding Shapley values.

and under the presence of harsh EOVs. Furthermore, the results from these demonstrations are compared to the results from the application of the MD-based novelty index.

The first application adopts the ANN-based novelty index approach to mitigate EOVs such as temperature and wind speed. The overall damage detection results improved relative to the MD-based approach. Particularly the smallest damage of 15cm had a significant improvement based on the judgement of the implemented performance metrics. More importantly, the implementation of the ANN-based framework on an

operating wind turbine confirmed that the proposed ANN architecture with fixed hyperparameters that was tested in artificial numerical and experimental data was equally accurate in a real application. This confirms the robustness and generalisation of the proposed ANN-based framework.

The ANN-based approach demonstrated provided high accuracy yet lacked interpretability. The issue of interpretability was first addressed by implementing the interpretable ML approach proposed in Chapter 4. The implementation of XGBoost and SHAP was able to identify outliers caused by the influence of temperature and wind speed rather than by the presence of damage in the MD-based novelty index. Nonetheless, EOVs such as wind speed, rotational speed and pitch angle were expected to have a more significant marginal contribution as identified by SHAP. A more thorough analysis based not only on data-driven findings but on domain knowledge is necessary. It must be bore in mind that the ML techniques do not consider physical relationships between EOVs and the novelty index. The ML model is successful not when physical relations are represented but when the objective prediction error is minimised. Therefore, a critical analysis of the contributions is required.

The interpretable ML-based approach has not improved the damage detection accuracy in terms of the performance metrics chosen but rather provides the operator with the understanding of the novelties identified. As such, a combination of the ANN-based approach and the interpretable ML-based approach is a potential solution for a robust and interpretable novelty index. This combination was demonstrated in section 5.5. Essentially, this implies replacing the ANN model with the XGBoost model.

The results show a slight decrease in the performance metrics for the 30cm and 45 cm damage. However, the accuracy remains high and the results are robust. In the case of the 15cm damage, which had been poorly identified by the MD-based approach,

the results show a significant improvement. Therefore, overall, the XGBoost model can be considered as an improvement to the MD-based novelty index but less accurate than the ANN model.

On the one hand, the main limitation of the XGBoost model is that the hyperparameters of the model were identified in a heuristic manner as the hyperparameters chosen in section 4.2 were found not to be applicable for the V27 case study. On the other hand, the main advantage of using XGBoost is that it enables the implementation of SHAP for added interpretability. XGBoost provides a computationally inexpensive calculation of Shapley values, which represent the marginal contribution of the features used in the model. The results resemble those obtained from the implementation in section 4.2.2, where most non-damage related novelties are mainly caused by the variations of temperature.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

Wind turbines are exposed to harsh environmental and operational conditions such as turbulent wind and abrupt temperature changes. Under these conditions, structural health monitoring can be a crucial yet challenging task. These conditions can add uncertainty in the damage detection process and camouflage the existence of damage in a wind turbine blade (WTB). Well researched and established data normalisation and damage detection techniques such as the Mahalanobis distance (MD), cointegration and Principal Component Analysis (PCA) have been adopted for data normalisation prior to calculating the novelty index. This can require the establishment of assumptions about the distribution of the data and/or the adoption of unsupervised learning techniques. The accuracy of the damage detection process can be directly impacted by the choice of unsupervised learning techniques.

To overcome this challenge, this dissertation proposed a novel approach to data normalisation, where the novelty index is calculated before normalisation takes place. In this way, the influence of environmental and operational variabilities (EOVs) can be visualised in the novelty indices and mitigated through post-processing (or post-normalisation). Referenced on this novel approach, three different frameworks have been proposed and evaluated in this dissertation.

6.1.1 Artificial Neural Network-based framework

An Artificial Neural Network (ANN) regression model was implemented, drawing on this technique's pattern recognition capabilities, to learn the relationship between the identified damage sensitive features (DSFs) and the MD-based novelty index. For this, only a subset of healthy data was used and the prediction error served as a new novelty index with a compensation for the influence of EOVs.

The method was illustrated on an artificial dataset with a non-Gaussian distribution, where the ANN-based novelty index was able to successfully distinguish between two clusters. The sole use of the MD-based novelty index did not yield a clear classification of the two clusters. This suggests that the pattern recognition capabilities of the ANN contributed to overcoming the limitation of identifying damage in a non-normally distributed dataset through post-normalisation, which has been a significant drawback of the MD-based approach. Commonly, this drawback has been addressed through unsupervised learning techniques such as cointegration and PCA. However, the proposed approach enabled the use of a semi-supervised learning framework with a supervised ANN-based regression model, thereby, leading to high accuracy in damage detection.

The proposed ANN-based framework was further tested on a 6-degrees-of-freedrom (DOF) mass-spring system. For comparison, the same 6-DOF system was evaluated by the well-established cointegration approach. To challenge both methods, a temperature field and additional random stiffness changes in the spring were introduced. The accuracy of both methods decreased with increased random variabilities in the stiffness. Nevertheless, when introducing more than 10% random stiffness changes in the spring, the ANN-based novelty index outperformed the cointegration-based approach with a significantly higher accuracy, F1-score and MCC results.

Cointegration is an easily reproducible approach since the technique is a straightforward mathematical procedure where no parameters must be trained. This can be considered an advantage as this is not the case for the ANN-based approach. By retraining the ANN, different results can occur. To assess the reproducibility of the ANN-based approach, the training was repeated 1000 times. Considering the performance metrics and their standard deviation, the results from the ANN-based framework were better than the results from the cointegration-based approach.

Further, the proposed ANN-based framework was compared to the PCA-based approach in a laboratory experiment, where a blade was mounted on a test rig and damage was simulated by adding a mass on several locations. When implementing the MD-based novelty index, clear deviations or outliers can be observed after the training threshold. It was possible to remove this trend by reducing the principal components and, thus, reducing the collinearity between the feature vectors. With this, the false alarm rate was reduced and the accuracy of damage detection increased relative to the MD-based approach. The main challenge with the PCA-based approach was the choice of the number of principal components. The balance between condition number and the variance explained can be an indication of an adequate choice of principal components, however, the balance is subjective to the modeller's choice. The results are highly sensitive to the number of principal components chosen and,

therefore, this is a key factor in the damage detection process.

The implementation of the ANN-based novelty index led to high accuracy and, thus, a low false alarm rate. To address the added uncertainty from, e.g., the subjective choice of principal components, the semi-supervised framework with the ANN regression model was implemented with fixed architecture and hyperparameters. The predictions were consistent with a small standard deviation after 1000 repetitions.

Finally, the semi-supervised framework with the ANN regression model was tested on an operating Vestas V27 wind turbine with artificial damage introduced to one of its blades. As previously seen in the results from the 6-DOF system and the lab experiment, the ANN-based approach outperformed the MD-based approach yielding higher results in all performance metrics. The ANN model was able to represent the MD-based novelty index trained on a subset of healthy data up to the point where the damage was introduced and where the prediction error increased, thus, indicating potential damage to the blade. However, despite the high accuracy and reproducibility of the semi-supervised ANN-based framework, the issue of interpretability remained unaddressed.

It is important to mention that machine learning techniques such as ANN require a large amount of data to identify patterns for further extrapolation. As such, it was not possible to use only the available healthy data as it is limited to around 800 observations. The repaired data of the blade was required for training to achieve the aforementioned results, i.e., approx. 2000 observations were used to properly predict the remaining 500 observations in the healthy state. This is very likely to affect the prediction accuracy in longer monitoring campaigns when damage occurs later.

6.1.2 Interpretable MD-based framework

To address the lack of interpretability in the MD-based novelty index, a novel framework was introduced. The framework adopts the gradient boosted (XGBoost) decision trees regression model, which learns the relationship between DSFs and environmental and operational parameters (EOPs). To explain the predictions of the regression model, which ultimately represents the MD-based novelty index, the Shapley Additive exPlanations (SHAP) method was used. The combination of both XGBoost and SHAP leads to an interpretable MD-based novelty index.

This framework was tested on a numerical 6-DOF mass-spring system influenced by a temperature field and affected by stiffness reduction in one of its springs to simulate damage. The hyperparameters for the regression model were identified by a random search and were fixed after an R^2 above 0.99 was reached. This was considered a good representation of the MD-based novelty index by the XGBoost regression model. The interpretability by SHAP has clearly shown the marginal contribution of the temperature field to the novelty index and could be directly correlated to the false alarms generated by the temperature field, which are different from the novelties generated by the stiffness reduction in the system.

This interpretable framework was also tested on the operating V27 wind turbine. The same training approach as in the 6-DOF system was adopted to represent the MD-based novelty index. Measured EOPs such as ambient temperature, wind speed and wind direction, among others, were included as features in the prediction model. SHAP enabled the correlation of extreme temperature values such as around or below 0°C to an increase in the novelty index, thereby, explaining false alarms. This contributed to dealing with uncertainties in the damage detection process. A similar conclusion can be reached when dealing with wind speed. Wind speeds that are

close to the cut-in wind speed of the wind turbine were found to be correlated to an increase in the novelty index. Nevertheless, in the case of wind direction, it was not possible to identify a clear and consistent relationship between the measurements and the novelty index. This is due to the lack of knowledge regarding the position of the nacelle. Therefore, trying to interpret parameters that have virtually no meaning can be misleading without domain knowledge.

Overall, it can be said that the interpretable framework does not answer all questions or explain all novelties yet can be a good addition and support for the operator in the decision-making process. This framework is considered a general framework that can be implemented on any novelty index where interpretability is missing.

6.1.3 Combined, data-driven and interpretable framework

The damage detection framework proposed results from the combination of the previously described semi-supervised damage detection framework and the interpretable framework. The combined, semi-supervised and interpretable damage detection framework provides accurate and insightful results. Nonetheless, this combination implies a trade-off between accuracy and interpretability. The damage detection accuracy is lower relative to the use of the ANN-based novelty index yet superior relative to the use of the MD-based novelty index.

Although the proposed framework combining high accuracy and interpretability aids the understanding of the relationship between EOVs, DSFs and novelty index, comprehending the marginal contribution of some EOVs was not straightforward. For instance, there were cases where low temperature values were directly correlated to high novelty indices leading to false alarms. Meanwhile, there were other cases where

high novelty indices due to high temperature were not clearly reflected on the relationship between novelty index and their marginal contribution (or Shapley values). As the training took place with only the first 2000 observations, it is very likely that not all patterns were identified, particularly as the high temperature occurred the first time after the first 2000 observations.

Another challenge of the interpretable framework is the choice of hyperparameters for the XGBoost model. The hyperparameters were initially defined based on the random search from the interpretable MD-based approach developed in sections 4.2.2 and 5.5. However, this random search was not suitable for the semi-supervised framework to accurately extrapolate the MD-based novelty index. A heuristic search, through trial and error, was conducted to achieve the presented results. This is a major drawback if the proposed combined framework and adds uncertainty in an online monitoring process.

Ultimately, the choice of approach for damage detection will be application-specific and requires domain knowledge and expertise. The operator should decide whether the trade-off between accuracy and interpretability is acceptable for the monitored structure. A summary of the advantages and disadvantages of the four approaches tested on the operating V27 wind turbine blade for damage detection in this thesis are presented in Table 6.1.

Table 6.1
Comparison of frameworks implemented for online monitoring of an in-operation V27 wind turbine blade

Approach	**MD-based (Section 5.2)**	**ANN-based (Section 5.3)**	**Interpretable MD-based (Section 5.4)**	**Combined [XGBoost + SHAP] (Section 5.5)**
Pros	- Low computational cost - Easy to implement - High reproducibility	- High damage detection accuracy through efficient mitigation of EOVs - High reproducibility - Low computational cost due to simple ANN architecture	- Low computational cost - Easy to implement - High reproducibility - Added interpretability - Potential mitigation of EOVs through interpretability	- Low computational cost - High reproducibility - Added interpretability - Adequate mitigation of EOVs
Cons	- Poor mitigation of EOVs - Limited interpretability in a multi - dimensional space	- Limited interpretability - Extensive monitoring campaign required	- Limited mitigation of EOVs - Extensive monitoring campaign required	- Trade-off between accuracy and interpretability - Extensive monitoring campaign required

6.2 Recommendations

Considering the advantages and disadvantages of the four different approaches for damage detection investigated in this dissertation, the following recommendations can be offered:

- The MD-based novelty index can be implemented as a first step for developing a damage detection framework as it provides an overview of the nature of the data and can aid the selection of DSFs that contain useful information for the detection of damage. The low computational cost and complexity of this approach makes it fast and easy to implement.
- The ANN-based novelty index yielded the highest accuracy and, therefore, is the most suitable approach for damage detection in wind turbine blades under the condition that a black-box model is acceptable to the operator.

- Much longer monitoring campaigns are required to provide the ANN model sufficient data and hidden patterns for training. Ideally, the monitoring campaign would cover two years. In this way, each season has been experienced twice and seasonal patterns can be learned and validated.

- It is recommended to retrain the ANN model as new data is collected to increase the robustness and accuracy. This can be done, for example, in steps of 500 observations to include new patterns that may arise rather than to extrapolate to unknown patterns.

- Unfortunately, since the semi-supervised approach relies on historical data from an undamaged wind turbine, it is recommended to install such a monitoring system early in the operating lifetime of the wind turbine. In the case of newly installed wind turbines, the chances for damage are low. Therefore, the initial operating years can be used for data collection. When sufficient data is gathered, the monitoring system can start providing damage-related recommendations.

- The interpretable framework did not provide high accuracy compared to the ANN-based approach. However, it is recommended as a supplementary framework to understand the reasons for false alarms through measured EOPs.

- It is important to evaluate the outcomes of the interpretable framework with aid from an operator with domain knowledge as results could be misinterpreted if not analysed with a critical view.

- The combined semi-supervised and interpretable approach presented a trade off between accuracy and interpretability. Whenever the trade off is not desirable, the parallel use of the ANN-based and the interpretable MD-based approaches is suggested, i.e., using mitigated EOVs with a non-interpretable novelty index in the ANN-based approach and non-mitigated EOVs with an interpretable novelty index in the second approach.

6.3 Further work

Considering all limitations, mitigating EOVs through an interpretable ML framework proved to be a successful approach for reliable and robust damage detection in an operating V27 wind turbine blade. Nonetheless, further work is required on structures monitored through different seasons over several cycles (or years). The limited patterns identified over the few months of the monitoring campaign of the V27 wind turbine did not provide a consistent explanation to the interaction between extreme temperature values and their contribution to the novelty index. This may add bias in the results since the algorithm could potentially be over fitting its predictions for a particular season. This hypothesis can only be tested when more data is available. Further monitoring campaigns should be conducted, monitoring a structure for at least 2 years so that seasonal patterns can repeat and appropriate testing of the proposed framework can be conducted. This would benefit the ANN-based approach, which could be then tested for retraining and continuous online monitoring.

Additionally, whilst a particular structure of the ANN with fixed architecture was identified and provided promising results, the accuracy of the results could change when new seasonal data is available in an operating wind turbine. New seasonal patterns may require a more complex ANN architecture, which could be found through a grid search similar to the one performed for the XGBoost model. Thus, more extensive monitoring campaigns and data collection is fundamental to continue testing and/or verifying the proposed ANN-based framework. Similarly, a longer monitoring campaign and additional data would serve the XGBoost model to develop a method for hyperparameters identification that can be more consistent in a combined framework.

Following up the work developed in this dissertation, it is advisable to further develop

the combined framework through a structured approach for hyperparameter identification in an XGBoost decision tree model using data from an operating wind turbine monitored for at least 2 years.

Bibliography

[1] Aghajani, A., Kazemzadeh, R., and Ebrahimi, A. (2016). A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. *Energy conversion and management*, 121:232–240.

[2] Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta paediatrica*, 96(5):644–647.

[3] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., and Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical systems and signal processing*, 147:107077.

[4] Avendano-Valencia, L. D., Chatzi, E. N., and Tcherniak, D. (2020). Gaussian process models for mitigation of operational variability in the structural health monitoring of wind turbines. *Mechanical Systems and Signal Processing*, 142:106686.

[5] Azim, M. R. and Gül, M. (2021). Data-driven damage identification technique for steel truss railroad bridges utilizing principal component analysis of strain response. *Structure and Infrastructure Engineering*, 17(8):1019–1035.

[6] Azimi, M., Eslamlou, A. D., and Pekcan, G. (2020). Data-driven structural health

monitoring and damage detection through deep learning: State-of-the-art review. *Sensors*, 20(10):2778.

[7] Balafas, K. and Kiremidjian, A. S. (2014). Extraction of a series of novel damage sensitive features derived from the continuous wavelet transform of input and output acceleration measurements. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, volume 9061, page 90611L. International Society for Optics and Photonics.

[8] Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford university press.

[9] Bao, Y. and Li, H. (2021). Machine learning paradigm for structural health monitoring. *Structural Health Monitoring*, 20(4):1353–1372.

[10] Barber, S., Wang, Y., Jafari, S., Chokani, N., and Abhari, R. S. (2011). The impact of ice formation on wind turbine performance and aerodynamics. *Journal of Solar Energy Engineering*, 133(1).

[11] Bellet, A., Habrard, A., and Sebban, M. (2015). Metric learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 9(1):1–151.

[12] Betti, M., Facchini, L., and Biagini, P. (2015). Damage detection on a three-storey steel frame using artificial neural networks and genetic algorithms. *Meccanica*, 50(3):875–886.

[13] Bhowmik, B., Tripura, T., Hazra, B., and Pakrashi, V. (2020). Real time structural modal identification using recursive canonical correlation analysis and application towards online structural damage detection. *Journal of Sound and Vibration*, 468:115101.

[14] Bortolotti, P., Berry, D. S., Murray, R., Gaertner, E., Jenne, D. S., Damiani, R. R., Barter, G. E., and Dykes, K. L. (2019). A detailed wind turbine blade cost model. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States). NREL/TP-5000-73585.

[15] Bull, L., Gardner, P., Gosliga, J., Rogers, T., Dervilis, N., Cross, E., Papatheou, E., Maguire, A., Campos, C., and Worden, K. (2021). Foundations of population-based shm, part i: Homogeneous populations and forms. *Mechanical Systems and Signal Processing*, 148:107141.

[16] Bussmann, N., Giudici, P., Marinelli, D., and Papenbrock, J. (2020). Explainable ai in fintech risk management. *Frontiers in Artificial Intelligence*, 3:26.

[17] Caballé-Cervigón, N., Castillo-Sequera, J. L., Gómez-Pulido, J. A., Gómez-Pulido, J. M., and Polo-Luque, M. L. (2020). Machine learning applied to diagnosis of human diseases: A systematic review. *Applied Sciences*, 10(15):5135.

[18] Caithness Windfarm Information Forum 2021 (2021). Summary of Wind Turbine Accident data to 31 March 2021. `http://www.caithnesswindfarms.co.uk/AccidentStatistics.htm`.

[19] Cao, M., Sha, G., Gao, Y., and Ostachowicz, W. (2017). Structural damage identification using damping: a compendium of uses and features. *Smart Materials and structures*, 26(4):043001.

[20] Casas, J. R. and Aparicio, A. C. (1994). Structural damage identification from dynamic-test data. *Journal of Structural Engineering*, 120(8):2437–2450.

[21] Casas, J. R. and Moughty, J. J. (2017). Bridge damage detection based on vibration data: Past and new developments. *Frontiers in Built Environment*, 3:4.

[22] Castro, O., Belloni, F., Stolpe, M., Yeniceli, S. C., Berring, P., and Branner, K. (2021). Optimized method for multi-axial fatigue testing of wind turbine blades. *Composite Structures*, 257:113358.

[23] Cattin, R. (2013). Icing of wind turbines: Vindforsk projects, a survey of the development and research needs, elforsk report 12: 13.

[24] Chandrasekhar, K., Stevanovic, N., Cross, E. J., Dervilis, N., and Worden, K. (2021). Damage detection in operational wind turbine blades using a new approach based on machine learning. *Renewable Energy*, 168:1249–1264.

[25] Chang, K.-C. and Kim, C.-W. (2016). Modal-parameter identification and vibration-based damage detection of a damaged steel truss bridge. *Engineering Structures*, 122:156–173.

[26] Chen, Q., Kruger, U., and Leung, A. Y. (2009). Cointegration testing method for monitoring nonstationary processes. *Industrial & Engineering Chemistry Research*, 48(7):3533–3543.

[27] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

[28] Cheung, A., Cabrera, C., Sarabandi, P., Nair, K. K., Kiremidjian, A., and Wenzel, H. (2008). The application of statistical pattern recognition methods for damage detection to field data. *Smart Materials and Structures*, 17(6):065023.

[29] Chicco, D. and Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):6.

[30] Choe, D.-E., Kim, H.-C., and Kim, M.-H. (2021). Sequence-based modeling of deep learning with lstm and gru networks for structural damage detection of floating offshore wind turbine blades. *Renewable Energy*, 174:218–235.

[31] Comanducci, G., Magalhães, F., Ubertini, F., and Cunha, Á. (2016). On vibration-based damage detection by multivariate statistical techniques: Application to a long-span arch bridge. *Structural health monitoring*, 15(5):505–524.

[32] Cotton, I., Jenkins, N., and Pandiaraj, K. (2001). Lightning protection for wind turbine blades and bearings. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 4(1):23–37.

[33] Cross, E., Manson, G., Worden, K., and Pierce, S. (2012). Features for damage detection with insensitivity to environmental and operational variations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468(2148):4098–4122.

[34] Cross, E. J., Worden, K., and Chen, Q. (2011). 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 Sciences*, 467(2133):2712–2732.

[35] d N Santos, F., Noppe, N., Weijtjens, W., and Devriendt, C. (2021). Two-tier model for wind turbine fatigue assessment based on scada-dependent neural networks. In *10th InternationalConferenceon StructuralHealth Monitoringof IntelligentInfrastructure*.

[36] Dao, C., Kazemtabrizi, B., and Crabtree, C. (2019). Wind turbine reliability data review and impacts on levelised cost of energy. *Wind Energy*, 22(12):1848–1871.

[37] Das, S., Saha, P., and Patro, S. (2016). Vibration-based damage detection techniques used for health monitoring of structures: a review. *Journal of Civil Structural Health Monitoring*, 6(3):477–507.

[38] Debel, C. (2004). Identification of damage types in wind turbine blades tested to failure. In Somers, M., editor, *Materialeopførsel og skadesanalyse*, pages 123–127. Dansk Metallurgisk Selskab. Dansk Metallurgisk Selskabs vintermøde 2004 ; Conference date: 07-01-2004 Through 09-01-2004.

[39] Delgadillo, R. M. and Casas, J. R. (2020). Non-modal vibration-based methods for bridge damage identification. *Structure and Infrastructure Engineering*, 16(4):676–697.

[40] Deming, W. E. (1975). On probability as a basis for action. *The American Statistician*, 29(4):146–152.

[41] Dervilis, N., Choi, M., Taylor, S., Barthorpe, R., Park, G., Farrar, C., and Worden, K. (2014). On damage diagnosis for a wind turbine blade using pattern recognition. *Journal of sound and vibration*, 333(6):1833–1850.

[42] Dutton, A. (2004). Thermoelastic stress measurement and acoustic emission monitoring in wind turbine blade testing. In *European Wind Energy Conference London*, pages 22–25.

[43] Eftekhar Azam, S., Rageh, A., and Linzell, D. (2019). Damage detection in structural systems utilizing artificial neural networks and proper orthogonal decomposition. *Structural Control and Health Monitoring*, 26(2):e2288.

[44] Energy, G. R. (2019). Haliade-X offshore wind turbine. `https://www.ge.com/renewableenergy/wind-energy/offshore-wind/haliade-x-offshore-turbine`.

[45] Erge, O. and van Oort, E. (2022). Combining physics-based and data-driven modeling in well construction: Hybrid fluid dynamics modeling. *Journal of Natural Gas Science and Engineering*, 97:104348.

[46] Farrar, C. R., Cornwell, P. J., Doebling, S. W., and Prime, M. B. (2000). Structural health monitoring studies of the alamosa canyon and i-40 bridges.

[47] Farrar, C. R. and Doebling, S. W. (1997). An overview of modal-based damage identification methods.

[48] Farrar, C. R., Sohn, H., and Worden, K. (2001). Data normalization : A key for structural health monitoring.

[49] Farrar, C. R. and Worden, K. (2012a). Data normalisation. In *Structural health monitoring: A machine learning perspective*, chapter 12, pages 403–437. John Wiley & Sons.

[50] Farrar, C. R. and Worden, K. (2012b). Historical overview. In *Structural health monitoring: A machine learning perspective*, chapter 2, pages 17–44. John Wiley & Sons.

[51] Farrar, C. R. and Worden, K. (2012c). Unsupervised learning – novelty detection. In *Structural health monitoring: A machine learning perspective*, chapter 10, pages 321–330. John Wiley & Sons.

[52] Farrar, C. R., Worden, K., and Dulieu-Barton, J. (2009). Principles of structural degradation monitoring. *Encyclopedia of structural health monitoring*.

[53] Farrar, C. R., Worden, K., Todd, M. D., Park, G., Nichols, J., Adams, D. E., Bement, M. T., and Farinholt, K. (2007). Nonlinear system identification for damage detection. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM.

[54] Fassois, S. D. and Kopsaftopoulos, F. P. (2013). Statistical time series methods for vibration based structural health monitoring. In *New trends in structural health monitoring*, pages 209–264. Springer.

[55] Figueiredo, E., Figueiras, J., Park, G., Farrar, C. R., and Worden, K. (2011a). Influence of the autoregressive model order on damage detection. *Computer-Aided Civil and Infrastructure Engineering*, 26(3):225–238.

[56] Figueiredo, E., Park, G., Farrar, C. R., Worden, K., and Figueiras, J. (2011b). Machine learning algorithms for damage detection under operational and environmental variability. *Structural Health Monitoring*, 10(6):559–572.

[57] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

[58] Frohboese, P. and Anders, A. (2007). Effects of icing on wind turbine fatigue loads. In *Journal of Physics: Conference Series*, volume 75, page 012061. IOP Publishing.

[59] Fugate, M. L., Sohn, H., and Farrar, C. R. (2001). Vibration-based damage detection using statistical process control. *Mechanical systems and signal processing*, 15(4):707–721.

[60] FUGATE, M. L., SOHN, H., and FARRAR, C. R. (2001). Vibration-based damage detection using statistical process control. *Mechanical Systems and Signal Processing*, 15(4):707–721.

[61] Fujimoto, K., Kojadinovic, I., and Marichal, J.-L. (2006). Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55(1):72–99.

[62] Garcia, D., Palazzetti, R., Trendafilova, I., Fiorini, C., and Zucchelli, A. (2015). Vibration-based delamination diagnosis and modelling for composite laminate plates. *Composite Structures*, 130:155–162.

[63] García, D. and Tcherniak, D. (2019). An experimental study on the data-driven structural health monitoring of large wind turbine blades using a single accelerometer and actuator. *Mechanical Systems and Signal Processing*, 127:102–119.

[64] García Cava, D., Avendaño-Valencia, L. D., Movsessian, A., Roberts, C., and Tcherniak, D. (2022). *On Explicit and Implicit Procedures to Mitigate Environmental and Operational Variabilities in Data-Driven Structural Health Monitoring*, pages 309–330. Springer International Publishing, Cham.

[65] Gardner, P., Bull, L., Gosliga, J., Poole, J., Dervilis, N., and Worden, K. (2022). A population-based shm methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings. *Mechanical Systems and Signal Processing*, 172:108918.

[66] Gardner, P., Liu, X., and Worden, K. (2020). On the application of domain adaptation in structural health monitoring. *Mechanical Systems and Signal Processing*, 138:106550.

[67] Gillich, G.-R., Furdui, H., Wahab, M. A., and Korka, Z.-I. (2019). A robust damage detection method based on multi-modal analysis in variable temperature conditions. *Mechanical Systems and Signal Processing*, 115:361–379.

[68] Global Wind Energy Council (2021). GWEC— Global Wind Report 2021. `https://www.eqmagpro.com/wp-content/uploads/2021/03/GWEC-I-Global-Wind-Report-2021_compressed-1-10.pdf`.

[69] Goh, L., Bakhary, N., Rahman, A., and Ahmad, B. (2013). Prediction of unmeasured mode shape using artificial neural network for damage detection. *Jurnal Teknologi*, 61(1).

[70] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

[71] Gu, J., Gul, M., and Wu, X. (2017). Damage detection under varying temperature using artificial neural networks. *Structural Control and Health Monitoring*, 24(11):e1998.

[72] Hakim, S., Razak, H. A., and Ravanfar, S. (2015). Fault diagnosis on beam-like structures from modal parameters using artificial neural networks. *Measurement*, 76:45–61.

[73] Hau, E. (2013). *Wind turbines: fundamentals, technologies, application, economics*. Springer Science & Business Media.

[74] Hellenkamp, O. (2020). Großes Windrad an der B58 in Haltern verliert einen Fluegel [Wind turbine loses blade]. `https://www.lokalkompass.de/dorsten/c-blaulicht/grosses-windrad-an-der-b58-in-haltern-verliert-einen-fluegel_a1352778`.

[75] Hester, D. and González, A. (2012). A wavelet-based damage detection algorithm based on bridge acceleration response to a vehicle. *Mechanical Systems and Signal Processing*, 28:145–166.

[76] Hochart, C., Fortin, G., Perron, J., and Ilinca, A. (2008). Wind turbine performance under icing conditions. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 11(4):319–333.

[77] Hripcsak, G. and Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, 12(3):296–298.

[78] Hwang, Y.-T., Hung, Y.-H., Wang, C. C., and Terng, H.-J. (2018). Finding the optimal threshold of a parametric roc curve under a continuous diagnostic measurement. *REVSTAT–Statistical Journal*, 16(1):23–43.

[79] International Energy Agency (2020). World Energy Outlook 2020. https://www.iea.org/reports/world-energy-outlook-2020.

[80] IPCC (2018). Summary for Policymakers. In Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.), editor, *Global Warming of 1.5°C. An IPCC Special Report*. Intergovernmental Panel on Climate Change.

[81] IRENA (2020). Renewable Power Generation Costs in 2019. https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Jun/IRENA_Power_Generation_Costs_2019.pdf.

[82] Jensen, F. M. and Branner, K. (2013). Introduction to wind turbine blade design. In *Advances in wind turbine blade design and materials*, pages 3–28. Elsevier.

[83] Jørgensen, E. R., Borum, K. K., McGugan, M., Thomsen, C., Jensen, F., Debel, C., and Sørensen, B. (2004). Full scale testing of wind turbine blade to failure-flapwise loading.

[84] Kopsaftopoulos, F. and Fassois, S. (2013). A functional model based statistical time series method for vibration based damage detection, localization, and magnitude estimation. *Mechanical Systems and Signal Processing*, 39(1):143–161.

[85] Kostić, B. and Gül, M. (2017). Vibration-based damage detection of bridges under varying temperature effects using time-series analysis and artificial neural networks. *Journal of Bridge Engineering*, 22(10):04017065.

[86] Krämer, C., De Smet, C., and De Roeck, G. (1999). Z24 bridge damage detection tests. In *IMAC 17, the International Modal Analysis Conference*, volume 3727, pages 1023–1029. Society of Photo-optical Instrumentation Engineers.

[87] Krishnan Nair, K. and Kiremidjian, A. S. (2006). Time Series Based Structural Damage Detection Algorithm Using Gaussian Mixtures Modeling. *Journal of Dynamic Systems, Measurement, and Control*, 129(3):285–293.

[88] Kudva, J., Munir, N., and Tan, P. (1992). Damage detection in smart structures using neural networks and finite-element analyses. *Smart Materials and Structures*, 1(2):108.

[89] Kulkarni, P. A., Dhoble, A. S., and Padole, P. M. (2019). Deep neural network-based wind speed forecasting and fatigue analysis of a large composite wind turbine blade. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 233(8):2794–2812.

[90] Kullaa, J. (2003). Damage detection of the z24 bridge using control charts. *Mechanical Systems and Signal Processing*, 17(1):163–170.

[91] Kumar, R. and Indrayan, A. (2011). Receiver operating characteristic (roc) curve for medical researchers. *Indian pediatrics*, 48(4):277–287.

[92] Larsen, G. C., Berring, P., Tcherniak, D., Nielsen, P. H., and Branner, K. (2014). Effect of a damage to modal parameters of a wind turbine blade. In *EWSHM-7th European Workshop on Structural Health Monitoring*.

[93] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

[94] Lim, S. and Chi, S. (2019). Xgboost application on bridge management systems for proactive damage estimation. *Advanced Engineering Informatics*, 41:100922.

[95] Lin, Y.-z., Nie, Z.-h., and Ma, H.-w. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 32(12):1025–1046.

[96] Lind, P. G., Vera-Tudela, L., Wächter, M., Kühn, M., and Peinke, J. (2017). Normal behaviour models for wind turbine vibrations: Comparison of neural networks and a stochastic approach. *Energies*, 10(12):1944.

[97] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):2522–5839.

[98] Lundberg, S. M., Erion, G. G., and Lee, S.-I. (2018a). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

[99] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*, pages 4765–4774.

[100] Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-F., Kim, J., et al. (2018b). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering*, 2(10):749–760.

[101] MacKay, D. J. (1992). Bayesian interpolation. *Neural computation*, 4(3):415–447.

[102] Martinez-Luengo, M., Kolios, A., and Wang, L. (2016). Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm. *Renewable and Sustainable Energy Reviews*, 64:91–105.

[103] Matthew, C. H. (2005). Impacts and Causes of Icing on Wind Turbines. `https://www.yumpu.com/en/document/read/20800745/impacts-and-causes-of-icing-on-wind-turbines`.

[104] Mehrjoo, M., Khaji, N., Moharrami, H., and Bahreininejad, A. (2008). Damage detection of truss bridge joints using artificial neural networks. *Expert systems with applications*, 35(3):1122–1131.

[105] Mishnaevsky, L., Branner, K., Petersen, H. N., Beauson, J., McGugan, M., and SÃ¸rensen, B. F. (2017). Materials for wind turbine blades: An overview. *Materials*, 10(11).

[106] Mollineaux, M., Balafas, K., Branner, K., Nielsen, P., Tesauro, A., Kiremidjian, A., and Rajagopal, R. (2014). Damage detection methods on wind turbine blade testing with wired and wireless accelerometer sensors. In *EWSHM-7th European workshop on structural health monitoring*.

[107] Molnar, C. (2019). *Interpretable Machine Learning*.

[108] Moser, P. and Moaveni, B. (2011). Environmental effects on the identified natural frequencies of the dowling hall footbridge. *Mechanical Systems and Signal Processing*, 25(7):2336–2357.

[109] Moughty, J. J. and Casas, J. R. (2017). A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions. *Applied Sciences*, 7(5):510.

[110] Movsessian, A., Cava, D. G., and Tcherniak, D. (2021a). An artificial neural network methodology for damage detection: Demonstration on an operating wind turbine blade. *Mechanical Systems and Signal Processing*, 159:107766.

[111] Movsessian, A., Cava, D. G., and Tcherniak, D. (2021b). Interpretable machine learning in damage detection using shapley additive explanations.

[112] Movsessian, A., Schedat, M., and Faber, T. (2021c). Feature selection techniques for modelling tower fatigue loads of a wind turbine with neural networks. *Wind Energy Science*, 6(2):539–554.

[113] Murtagh, P. and Basu, B. (2007). Identification of equivalent modal damping for a wind turbine at standstill using fourier and wavelet analysis. *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 221(4):577–589.

[114] Nazarko, P. and Ziemianski, L. (2016). Damage detection in aluminum and composite elements using neural networks for lamb waves signal processing. *Engineering Failure Analysis*, 69:97–107.

[115] Ng, C.-T. (2014). Application of bayesian-designed artificial neural networks in phase ii structural health monitoring benchmark studies. *Australian Journal of Structural Engineering*, 15(1):27–36.

[116] Nguyen, D. and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *1990 IJCNN International Joint Conference on Neural Networks*, pages 21–26. IEEE.

[117] Nilsson, J. (2006). *Maintenance management of wind power systems: Cost effect analysis of condition monitoring systems.* PhD thesis, Royal Institute of Technology KTH.

[118] Oh, C. K. and Sohn, H. (2009). Damage diagnosis under environmental and operational variations using unsupervised support vector machine. *Journal of Sound and Vibration*, 325(1):224–239.

[119] Olsen Wings A/S (2020). Manufacturer of small wind turbine blades. `http://olsenwings.dk/`. Accessed: 2020-02-25.

[120] Onchis, D. M. and Gillich, G.-R. (2021). Stable and explainable deep learning damage prediction for prismatic cantilever steel beam. *Computers in Industry*, 125:103359.

[121] Panagiotopoulos, A., Tcherniak, D., and Fassois, S. (2020). Damage detection on an operating wind turbine blade via a single vibration sensor: A feasibility study. In *European Workshop on Structural Health Monitoring*, pages 405–414. Springer.

[122] Parker, D. L. (2011). *Multi-objective design optimization framework for structural health monitoring*. Mississippi State University.

[123] Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., and Mohammadian, A. K. (2020). Toward safer highways, application of xgboost and shap for real-time accident detection and feature analysis. *Accident Analysis & Prevention*, 136:105405.

[124] Peeters, B. and De Roeck, G. (2001). Stochastic system identification for operational modal analysis: a review. *J. Dyn. Sys., Meas., Control*, 123(4):659–667.

[125] Peeters, B., Maeck, J., and Roeck, G. D. (2001). Vibration-based damage detection in civil engineering: excitation sources and temperature effects. *Smart Materials and Structures*, 10(3):518–527.

[126] Pereira, S., Magalhães, F., Gomes, J. P., Cunha, Á., and Lemos, J. V. (2021). Vibration-based damage detection of a concrete arch dam. *Engineering Structures*, 235:112032.

[127] Qadri, B. A., Ulriksen, M. D., Damkilde, L., and Tcherniak, D. (2020). Cointegration for detecting structural blade damage in an operating wind turbine: An experimental study. In *Dynamics of Civil Structures, Volume 2*, pages 173–180. Springer.

[128] Qiu, L., Fang, F., Yuan, S., Boller, C., and Ren, Y. (2019). An enhanced dynamic gaussian mixture model-based damage monitoring method of aircraft structures under environmental and operational conditions. *Structural Health Monitoring*, 18(2):524–545.

[129] Rane, S. (2018). The balance: Accuracy vs. Interpretability. https://towardsdatascience.com/the-balance-accuracy-vs-interpretability-1b3861408062.

[130] Reich, C., Mansour, A., and Van Laerhoven, K. (2018). Embedding intelligent features for vibration-based machine condition monitoring. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 371–375. IEEE.

[131] REN21 (2020). Renewables 2020 Global Status Report. https://www.ren21.net/wp-content/uploads/2019/05/gsr_2020_full_report_en.pdf.

[132] Roth, A. E. (1988). *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press.

[133] Ruffels, A., Gonzalez, I., and Karoumi, R. (2020). Model-free damage detection of a laboratory bridge using artificial neural networks. *Journal of Civil Structural Health Monitoring*, 10(2):183–195.

[134] Ruzzene, M., Fasana, A., Garibaldi, L., and Piombo, B. (1997). Natural frequencies and dampings identification using wavelet transform: application to real data. *Mechanical systems and signal processing*, 11(2):207–218.

[135] Rytter, A. (1993). *Vibrational based inspection of civil engineering structures*. PhD thesis, Dept. of Building Technology and Structural Engineering, Aalborg University.

[136] Samarasinghe, S. (2016). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. Crc Press.

[137] Schroeder, K., Ecke, W., Apitz, J., Lembke, E., and Lenschow, G. (2006). A fibre bragg grating sensor system monitors operational load in a wind turbine rotor blade. *Measurement science and technology*, 17(5):1167.

[138] Shewhart, W. A. (1926). Quality control charts. *The Bell System Technical Journal*, 5(4):593–603.

[139] Shi, H., Worden, K., and Cross, E. (2016). A nonlinear cointegration approach with applications to structural health monitoring. In *Journal of Physics: Conference Series*, volume 744, page 012025. IOP Publishing.

[140] Sobie, C., Freitas, C., and Nicolai, M. (2018). Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing*, 99:403–419.

[141] Soerensen, B. F., Lading, L., and Sendrup, P. (2002). Fundamentals for remote structural health monitoring of wind turbine blades - a pre-project.

[142] Sohn, H., Farrar, C., Hunter, N., and Worden, K. (2001a). Applying the lanl statistical pattern recognition paradigm for structural health monitoring to data from a surface-effect fast patrol boat. Technical report, Los Alamos National Lab., NM (US).

[143] Sohn, H., Worden, K., and Farrar, C. (2001b). Novelty detection using auto-associative neural network. Technical report, Los Alamos National Lab., NM (US).

[144] Solimine, J., Niezrecki, C., and Inalpolat, M. (2020). An experimental investigation into passive acoustic damage detection for structural health monitoring of wind turbine blades. *Structural Health Monitoring*, 19(6):1711–1725.

[145] Sørensen, B. F., Joergensen, E., Debel, C. P., Jensen, F., Jensen, H., Jacobsen, T., and Halling, K. (2004a). Improved design of large wind turbine blade of fibre composites based on studies of scale effects (phase 1) - summary report.

[146] Sørensen, B. F., Joergensen, E., Debel, C. P., Jensen, F., Jensen, H., Jacobsen, T., and Halling, K. (2004b). Improved design of large wind turbine blade of fibre composites based on studies of scale effects (phase 1)-summary report.

[147] Štrumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3):647–665.

[148] Su, W., Huang, C.-S., Chen, C., Liu, C., Huang, H., and Le, Q. (2014). Identifying the modal parameters of a structure from ambient vibration data via the stationary wavelet packet. *Computer-Aided Civil and Infrastructure Engineering*, 29(10):738–757.

[149] Sun, L., Shang, Z., Xia, Y., Bhowmick, S., and Nagarajaiah, S. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Journal of Structural Engineering*, 146(5):04020073.

[150] Talebinejad, I., Fischer, C., and Ansari, F. (2011). Numerical evaluation of vibration-based methods for damage assessment of cable-stayed bridges. *Computer-Aided Civil and Infrastructure Engineering*, 26(3):239–251.

[151] Tatsis, K., Dertimanis, V., Ou, Y., and Chatzi, E. (2020). Gp-arx-based structural damage detection and localization under varying environmental conditions. *Journal of Sensor and Actuator Networks*, 9(3):41.

[152] Tcherniak, D. and Mølgaard, L. L. (2017). Active vibration-based structural health monitoring system for wind turbine blade: Demonstration on an operating vestas v27 wind turbine. *Structural Health Monitoring*, 16(5):536–550.

[153] Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*.

[154] Ti, Z., Deng, X. W., and Zhang, M. (2021). Artificial neural networks based wake model for power prediction of wind farm. *Renewable Energy*, 172:618–631.

[155] Tomé, E. S., Pimentel, M., and Figueiras, J. (2020). Damage detection under environmental and operational effects using cointegration analysis–application to experimental data from a cable-stayed bridge. *Mechanical Systems and Signal Processing*, 135:106386.

[156] Trendafilova, I., Cartmell, M., and Ostachowicz, W. (2008). Vibration-based damage detection in an aircraft wing scaled model using principal component analysis and pattern recognition. *Journal of Sound and Vibration*, 313(3):560–566.

[157] Ulriksen, M. D., Tcherniak, D., and Damkilde, L. (2015). Damage detection in an operating vestas v27 wind turbine blade by use of outlier analysis. In *2015 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS) Proceedings*, pages 50–55.

[158] Ulriksen, M. D., Tcherniak, D., Kirkegaard, P. H., and Damkilde, L. (2016). Operational modal analysis and wavelet transformation for damage identification in wind turbine blades. *Structural Health Monitoring*, 15(4):381–388.

[159] Vandiver, J. K. (1975). Detection of structural failure on fixed platforms by measurement of dynamic response. In *Offshore Technology Conference*. OnePetro.

[160] Vera-Tudela, L. and Kühn, M. (2017). Analysing wind turbine fatigue load prediction: The impact of wind farm flow conditions. *Renewable Energy*, 107:352–360.

[161] Vilone, G. and Longo, L. (2020). Explainable artificial intelligence: a systematic review. *arXiv preprint arXiv:2006.00093*.

[162] Wang, L., Zhang, Z., Xu, J., and Liu, R. (2016). Wind turbine blade breakage monitoring with deep autoencoders. *IEEE Transactions on Smart Grid*, 9(4):2824–2833.

[163] Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.

[164] Worden, K., Manson, G., and Fieller, N. R. (2000). Damage detection using outlier analysis. *Journal of sound and vibration*, 229(3):647–667.

[165] Yam, L., Yan, Y., and Jiang, J. (2003). Vibration-based damage detection for composite structures using wavelet transform and neural network identification. *Composite Structures*, 60(4):403–412.

[166] Yan, A.-M., Kerschen, G., De Boe, P., and Golinval, J.-C. (2005). Structural damage diagnosis under varying environmental conditions – part i: A linear analysis. *Mechanical Systems and Signal Processing*, 19(4):847–864.

[167] Yan, Y., Cheng, L., Wu, Z., and Yam, L. (2007). Development in vibration-based structural damage detection technique. *Mechanical Systems and Signal Processing*, 21(5):2198–2211.

[168] Yao, R. and Pakzad, S. N. (2012). Autoregressive statistical pattern recognition algorithms for damage detection in civil structures. *Mechanical Systems and Signal Processing*, 31:355–368.

[169] Ye, X., Jin, T., and Yun, C. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Struct Syst*, 24(5):567–586.

[170] Yeager, M., Gregory, B., Key, C., and Todd, M. (2019). On using robust mahalanobis distance estimations for feature discrimination in a damage detection scenario. *Structural Health Monitoring*, 18(1):245–253.

[171] Ying, Y., Garrett Jr, J. H., Oppenheim, I. J., Soibelman, L., Harley, J. B., Shi, J., and Jin, Y. (2013). Toward data-driven structural health monitoring: application of machine learning and signal processing to damage detection. *Journal of Computing in Civil Engineering*, 27(6):667–680.

[172] Zhang, C., Mousavi, A. A., Masri, S. F., Gholipour, G., Yan, K., and Li, X. (2022). Vibration feature extraction using signal processing techniques for structural health monitoring: A review. *Mechanical Systems and Signal Processing*, 177:109175.

[173] Zhang, Z. and Sun, C. (2021). Structural damage identification via physics-guided machine learning: a methodology integrating pattern recognition with finite element model updating. *Structural Health Monitoring*, 20(4):1675–1688.

[174] Zhou, H., Ni, Y., and Ko, J. (2011). 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.