Elimination of traffic induced operational variability with Robust Principal Component Analysis (rPCA)

Snorri Þór Sigurðsson

Delft University of Technology

TU Delft

Elimination of traffic induced operational variability with Robust Principal Component Analysis (rPCA)

by

Snorri Þór Sigurðsson

Student number : 5129990
Thesis committee: Prof. dr. E. Lourens, TU Delft
Prof. dr. H. Wang, TU Delft
Ir. F. Besseling, Witteveen+Bos
Ir. J. de Bruijn Witteveen+Bos

Cover Image: ”An oil painting by Jacob van Ruisdael of a blue bascule bridge on the Haringvlietbrug” by Dall-E 2

TU Delft

Preface

A keen observer might have noticed that the cover image is, in fact, not the Haringvlietbrug. The image is generated by OpenAI's DALL-E 2, a text-to-image AI system. DALL-E 2 uses OpenAI's CLIP neural networks to learn visual concepts from natural language supervision. Fundamental to data-driven methods such as neural networks is the use of data. The Haringvlietbrug is not a bridge of international renown. There are few images of the Haringvlietbrug compared to recognisable bridges, such as the Erasmusbrug in Rotterdam. DALL-E 2 has difficulties in "knowing" what the Haringvlietbrug looks like due to a lack of data. This results in DALL-E 2 generating a generic bridge. Much like the example with the Haringvlietbrug and DALL-E 2, this thesis will face challenges regarding what data to use for analysis.

DALL-E 2 is just one of many inventions that take advantage of recent advances in data-driven science and engineering. The methods in the field have the potential to help solve problems in many engineering disciplines. There is no doubt that the field of Civil Engineering will benefit from these advancements.

Snorri Þór Sigurðsson
Reykjavík, March 2023

Abstract

Regular maintenance of civil engineering structures is essential for their safety. Current maintenance regimes involve periodic inspections at regular time intervals. In the time between inspections, there can be a critical development in the structural integrity of a structure, which can be expensive to repair or could even lead to structural failure. A more robust maintenance approach would involve continuous monitoring of the structure. This is the research area of Structural Health Monitoring (SHM). Vibration-based monitoring, which is a subset of SHM, aims to provide new cost-effective maintenance solutions that provide long-term life-safety benefits. Vibration-based monitoring uses vibration measurements from sensors to assess the ”Health” of a structure. Damage in a structure will alter the structure's stiffness, mass and damping characteristics, which in turn will change the dynamic properties of the system. This change can then be discovered in the vibration data. The growth of the field is partly due to the significant advances in data-driven science and engineering in recent decades.

One of the issues in vibration-based monitoring is the presence of operational and environmental variability in the vibration data. With this variability present, it is challenging to determine from vibration data the characteristics of the underlying dynamic system. The aim of this thesis is to use Robust Principal Component Analysis (rPCA) to reduce or eliminate the operational variability from the traffic to allow for environmental and damage detection. rPCA is a matrix factorisation method that decomposes a data matrix into a low-rank matrix L and a sparse matrix S. The reconstructed low-rank matrix L contains the main correlations in the data that are robust to outliers and corrupt data that are contained in the sparse matrix S. By applying rPCA to the frequency representation of the vibration data, it is hoped that the underlying coherent structure corresponding to the dynamic system can be recovered.

The vibration data used for this thesis is from two measurement campaigns conducted on the Haringvlietbrug. The Haringvlietbrug is a steel box girder bridge in the Netherlands, and there are several fatigue cracks present in the bridge. This presented an opportunity for damage detection. The goal of the first measurement campaign is to conduct damage detection and discover if there is a difference between vibration data from a damaged area with fatigue cracks and a ”healthy” reference area. In the second measurement campaign, the goal was to extract the underlying dynamics of the structure at different temperatures and see if it was possible to distinguish between the different structural states at different temperatures. After applying the rPCA on the vibration data, (regular) principal component analysis (PCA) is used to embed the data into the low-rank subspace of the PCs to distinguish between the different structural states.

The rPCA was successful in extracting the coherent structures in the vibration data corresponding to the underlying dynamic properties of the system. In the subsequent PCA, vibration data with underlying different structural states had different scores in the first three PCs. In other words, it was possible to distinguish between the different structural states based on the first three PCs, which correspond to the main correlation within the data. For the first measurement campaign, this meant it was possible to distinguish between vibration data in the damaged area and the ”healthy” reference area and detect ”damage”. However, there was a difference in the structural configuration between the two areas, so it was not possible to conclude that the differences in vibration data were due to damage caused by the fatigue cracks. In the second measurement campaign, the dynamic system properties at different temperatures were recovered. With the low-rank vibration data from the rPCA, it was possible to distinguish between vibration data with a 1°C difference in the first three PCs from the (regular) PCA. This was not possible without lowering the regularisation parameter in the rPCA. Another method, the Sparse Sensor Placement for Optimisation (SSPOC), was used to determine the locations in the frequency spectrum that contained the largest differences between structural states. For both the first and second measurement campaigns, these locations were at specific natural frequencies of the system.

Contents

1

Introduction

The design and construction of civil engineering structures are only a fraction of the life cycle of a structure. Most of a structure's lifetime will be spent in operational use and under various loads that can lead to structural deficiencies. In order to prevent structural degradation and maintain an acceptable level of safety, regular maintenance must be performed. As infrastructure becomes older and nears the end of its lifetime, the importance of maintenance and retrofitting becomes even more critical. In the United States of America, there are around 61700 bridges, 42 % of which are older than 50 years. Furthermore, 46,154 or 7.5 % of the total bridges are considered structurally deficient, with many elements nearing the end of their service life, as stated in a report by the American Society of Civil Engineers [1]. The report continues with that the backlog of bridge repair is around 125 billion U.S. dollars, and at the current level of investment into public infrastructure, it will take until 2071 to finish all the repairs that are currently necessary, and this is excluding the repairs needed due to structural deterioration over this period.

This problem isn't localised to the U.S.A., and Iceland faces a similar problem. Yearly public funds granted to the Icelandic Road and Coastal Administration towards the maintenance of infrastructure is lower than what is required, further increasing the already established backlog of repairs [22]. The Netherlands also has to maintain their infrastructure and might face similar problems. In order to maintain the safety and functionality of their infrastructure, these countries need to invest a substantial amount of capital into maintenance. Failure to do so could increase the risk of tragic events such as the collapse of the Morandi Bridge in Genoa, Italy. With the amount of funds that have to be invested in infrastructure and the life safety implications if a failure occurs, there is a considerable incentive to use new technology to make the maintenance of structures cheaper while ensuring a safer structure.

Current maintenance regimes of civil engineering structures usually involve periodic inspections at regular time intervals. Personnel inspect any visible damage and may employ non-destructive testing (NDE), such as acoustic or ultrasonic measurement techniques, to detect if there is any damage present in the structure. However, these methods require a priori information about the location of damage and require access to the section under investigation. This limits the effectiveness of these methods, as they can only be used where access to the structure is readily available, such as its surface and outer perimeter, and internal members of the structure might not be reachable. Another concern is that in the time between maintenance inspections, damage can develop to critical levels that could cause structural failure. However, these shortfalls are being tackled in Structural Health Monitoring (SHM), an emerging field over the last two decades. Although still primarily a research topic, SHM has the potential to offer cost-effective maintenance solutions and to ensure long-term life-safety benefits. The core of SHM involves using sensors to observe a structure over a given time. The data from these sensors can be used to extract information or features that are sensitive to damage and allow the identification of damaged and undamaged states of a structure. Thus, it is possible to perform damage detection of a particular structure and identify when there has been some deviation from a normal condition. This

method offers a continuous monitoring scheme and identifies damage when it occurs but not at the next inspection.

Vibration-based monitoring is a subsection of SHM that uses the dynamic properties of a structural system as a damage indicator, such as the natural frequencies and mode shapes. Damage changes the stiffness, mass, damping characteristics, and boundary conditions of the structural system. All of these affect the dynamic properties of the system [18]. By measuring the acceleration response of a structure over a period of time in which there is stiffness degradation in one of its members, one should be able to identify when damage occurs. However, this is not as straightforward. Although it is intuitive to use the dynamic properties of a structural system, there are a number of problems to overcome. Structures in operational use, such as bridges, experience a wide array of excitation due to the variability in the traffic. This variability in traffic is part of the operational variability. The dynamic properties of the structures are also affected by temperature changes during measurement; these changes are part of the environmental variability. Both Environmental and Operational Variability (EOV) combine to create a more complex problem.

In damage detection, the goal is to discern whether the state of the structure has deviated from an undamaged state to a damaged state. When working with vibration data, one has to compare the dynamic properties between the two different states. However, in most SHM applications, only the response of the system is measured, which is a product of the excitation and the dynamic properties of the system, which are affected by operational and environmental variability. Therefore, any deviation in the measured response could be attributed to the EOV but not actual damage. This can lead to false positives where the different environmental and operational conditions are classified as damage. Ultimately, the EOV masks the variable of interest, the dynamic properties. The subject of this thesis is to tackle the EOV in damage detection.

The field of Structural Health Monitoring has benefited from advances made in data-driven methods and improvements in sensor technology over the last two decades. Data-driven methods can ”learn” the relationship within the data and can be used to model and predict complex systems. For instance, these methods can identify a deviation in the normal condition of a system, which can enable damage detection. Structural Health Monitoring has seen an increase in research in data-driven methods by employing Pattern Recognition (PR), Machine Learning (ML), and Deep Learning (DL) [2, 37]. This thesis aims to explore emerging data methods for damage detection in Structural Health Monitoring.

1.1. Problem Definition

The structure that is the subject of this thesis is the Haringvliet bridge. It is a steel box-girder bridge over the Haringvliet in the southeast of the Netherlands. The bridge’s total length is 1220 m, with ten sections of around 106 m each. A moveable bridge in the north part of the bridge allows maritime traffic to travel up and down the Haringvliet. The bridge consists of a steel deck with a rectangular hollow section. Diagonal struts from the hollow section support the cantilever deck plate on either side of the bridge. The bridge is an important logistical connection between the south of Holland and Zeeland via the A29 highway. In 2019, the average daily number of vehicles that crossed the bridge was 64400[1]. Construction finished in 1964, and the bridge has been in service for nearly 60 years. Being a steel box-girder bridge, it is susceptible to fatigue damage under dynamic loads in the welding connections [42]. The bridge’s age, design and traffic load may have contributed to the formation of fatigue cracks in some sections of the bridge between the transverse beams and the longitudinal stiffeners. These fatigue cracks are a clear example of bridge damage and offer an opportunity to conduct damage detection. The fatigue cracks are present only in some of the sections of the bridge, and the sections have a similar structural configuration. Thus, it is possible to compare the vibration data of the damaged area with the vibration data from the ”healthy” reference area and perform damage detection between the two areas. This led to the Haringvlietbrug measurement campaign that installed accelerometers and temperature sensors in these areas. This measurement campaign was concerned with detecting damage from these fatigue cracks. After the conclusion of the first measurement campaign, a second

[1] Intensiteiten op Wegvakken - INWEVA Source

measurement campaign was conducted on the bridge with a more dispersed sensor layout in a single area to look at more global behaviour of the bridge. This thesis will use data from both campaigns but with different goals for each campaign. Data from the first measurement campaign will be used to detect damage from the fatigue cracks. However, as discovered in previous studies on the bridge, the damage due to the fatigue cracks is not the only difference between the two comparison areas. Thus the ”damage” detection is not solely detecting damage. Data from the second measurement campaign will be used to investigate the environmental variability.

1.2. Previous Research

A vibration-based damage detection study has already been conducted on the data from the first measurement campaign of the Haringvliet bridge [21]. In this study, a data-driven approach was taken to identify the presence of damage. Before damage detection was conducted, the vibration data was preprocessed with a method called similarity filtering. The method was first proposed by an SHM study on the Zwartewaterbrug [32] and later used by another study on the same bridge [15]. Compared to Haringvlietbrug, the Zwartewaterbrug is an arch bridge with an orthotropic bridge deck and a length of 104 m. In these two studies on the Zwartewaterbrug, damage detection was performed. However, there was no presence of actual damage. Instead, the damage was simulated with the introduction of small masses under the bridge deck. This would alter the dynamic profile of the bridge, similar to stiffness reduction with the formation of a crack. According to the two Zwartewaterbrug studies, similarity filtering was successfully used to eliminate operational variability. However, similar filtering was not successful in eliminating the operational variability in the previous study of the Haringvlietbrug. Alongside this research on the Haringvlietbrug, there was another study on the Haringvlietbrug [25]. That study concluded that similarity filtering does not work, and the successful results in the Zwartewaterbrug were due to false positives. For this research, instead of using similarity filtering, a promising method called Robust Principal Component Analysis (RPCA) will be used [9]. The question is whether it is possible with this method to reduce or eliminate the operational variability within the traffic, making damage detection possible for the Haringvlietbrug data.

1.3. Research Question

The main research question of this thesis is formulated as follows:

”Is it possible to reduce the operational variability of the traffic with Robust Principal Component analysis to enable the detection of either damage or changing environmental conditions.”

1.4. Thesis outline

The second chapter of this thesis is a general overview of SHM and data-driven methods, as well as an introduction to sparsity and compressed sensing. Chapter 3 briefly goes over the two measurement campaigns of the Haringvliet bridge. The theory for the methods used in the thesis is presented in chapter 4. Chapter 5 involves the general preprocessing of the data and the application of Robust principal component analysis (rPCA) on acceleration data. Chapter 6 is focused on detecting the damage due to fatigue cracks with data from the first measurement campaign. On the other hand, chapter 7 presents the analysis of the effect of environmental variability on the dynamic properties with the data from the second measurement campaign. Chapter 8 gives a more thorough analysis of the behaviour of the Robust principal component analysis. Finally, conclusions and recommendations are presented in chapter 9.

"*An oil painting by Jacob van Ruisdael of the Erasmus bridge in Rotterdam*" by Dall-E 2

2

Structural Health Monitoring & Data-Driven Science

This chapter presents background on Structural Health Monitoring (SHM), the advantages that the technology can bring, and the challenges faced within the field. It also presents an introduction to data-driven methods and a discussion of sparsity and compressed sensing.

2.1. Structural Health Monitoring: Advantages & Challanges

Structural health monitoring (SHM) is an emerging multidisciplinary field with aerospace, mechanical and civil engineering communities utilising SHM to solve problems in their respective fields. The civil engineering community has employed vibration-based monitoring for bridges and buildings since the 1980s [17]. SHM is a rapidly growing field, as seen in the number of publications of SHM literature in the past two decades in figure 2.1. Since 2010 there has been nearly a 100% increase in publications. Although still mainly a research topic, it has the potential to offer cost-effective solutions and life-safety benefits for the maintenance of civil engineering structures. The goal of any SHM project, as the name implies, is to observe a structure and monitor its structural integrity over time and diagnose potential issues with a structure before they become significant problems, allowing for timely and cost-effective maintenance and repair. This can help prevent structural failures and improve overall safety and reliability. With the use of sensors, it is possible to collect data about the state of the structure under investigation. Vibration-based monitoring is a subset of SHM that uses vibration data, such as acceleration recordings, for analysis. The idea of vibration-based monitoring is that damage will manifest itself as a change in a structure's stiffness, mass, damping properties, and boundary conditions. All of these aforementioned properties affect the dynamic properties of a structure, the natural frequencies and modal shapes. Thus the presence of damage will result in a different dynamic system than that of a system prior to the introduction of damage. By comparing the dynamic properties of the system before and after the introduction of damage, it is possible to identify if there has been a deviation in the normal condition of a structure.

The growth of the Structural Health Monitoring field has been partly due to significant advances in data-driven science and engineering, which has revolutionised how complex systems can be predicted and modelled. Driving this advancement of data-driven science is the vast availability of data enabled by the accessibility of affordable sensors, an increase in computational power, and improvements in data transfer and storage solutions. So promising is data-driven discovery that it has been hailed as the fourth paradigm of scientific discovery [23].

One of the major advantages of SHM is that it offers an alternative to time-based maintenance ap-

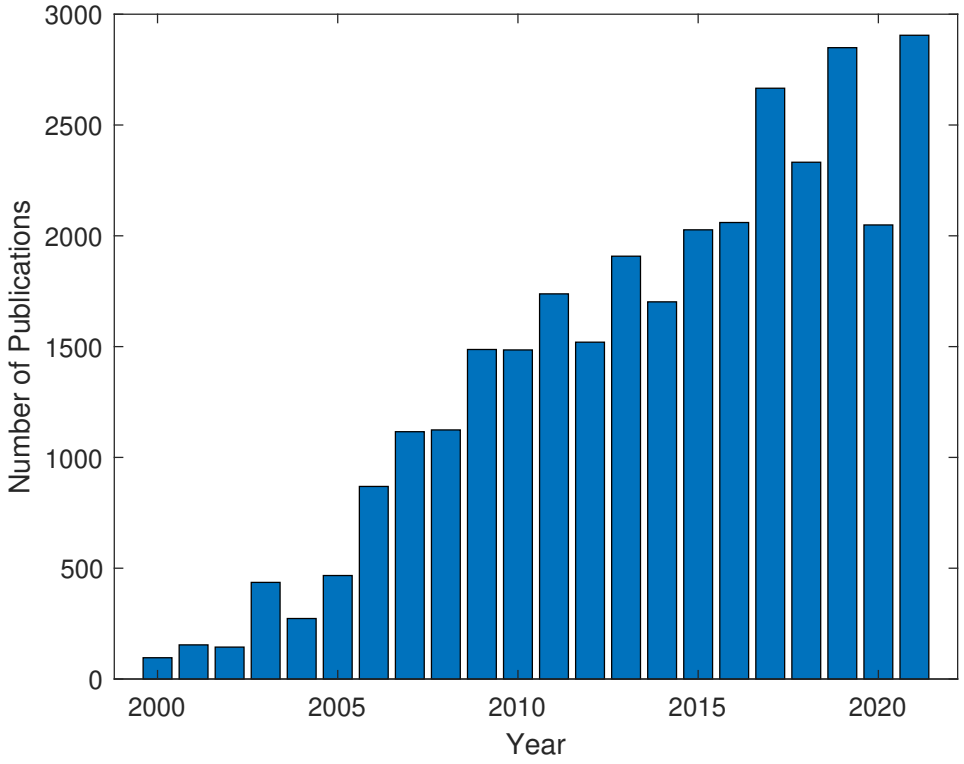


Figure 2.1: The number of Structural Health monitoring publications in the last two decades. Source : Scopus. Data gathered by searching ”Structural Health Monitoring”

proaches, where a structure is inspected at predetermined time intervals, regardless of the condition of the structure. On the other hand, condition-based monitoring is where the sensing system continually monitors the system and will notify operators of the structure if any damage has occurred. Compared to time-based maintenance schemes, a condition-based approach allows for the early detection of structural issues and the ability to take action before the issue becomes critical and potentially causes a major failure. This increases the structure's safety and reliability and reduces the need for costly and time-consuming inspections. Finally, much of the infrastructure built in the 20th century is reaching or exceeding its initial design life. Due to economic pressures, these structures are in use despite the increased risk associated with damage accumulation over the years. Therefore, monitoring the structural state of these at-risk structures becomes even more critical. [18]

While SHM for civil engineering structures promises great benefits for the maintenance of structures, there are major challenges for the field to overcome before the technology becomes commercially viable. The physical size of civil engineering infrastructure can pose a challenge to the implementation of an effective damage detection framework. Damage in a structure is generally a local phenomenon, and its influence can be restricted to high-frequency local mode shapes and their corresponding natural frequencies. The effect of damage might not be noticeable in the lower-frequency global modes of the structure. This essentially implies that to be able to capture the effect of potential damage. The sensors need to be optimised to higher-frequency local mode shapes. The local modes cover less area of the structure than the global modes, and thus more sensors are required to provide comprehensive coverage of the structure. A cost-effective SHM framework would minimise the number of sensors to reduce costs [17]. However, the aforementioned problems make this difficult.

Civil engineering structures are subject to operational and environmental variability that poses a challenge for SHM. A robust damage detection scheme has to take this variability into account. Changing operational and environmental conditions can alter the measured response of the system and can potentially result in false positives for the damage detection framework, where different operational and environmental conditions are misidentified as damage. Therefore it is imperative that the damage detection framework is robust to this variability to avoid misidentifying it as damage [39]. Temperature is

a typical example of an environmental condition that can influence the dynamic properties of a system. Different types of vehicles, with different masses, travelling at various speeds over a bridge is an example of operational variability. The excitation by these vehicles can be different, leading to a different measured response.

A final remark on the challenges of SHM for civil infrastructure is that nearly every structure is unique. Structures are not designed according to a single template. Even a group of similar structures, such as bridges, exhibit variation in their designs according to design specifications. The damage detection implementation has to be tailored to each structure. That means that there is no defined ”baseline” or normal condition that can be shared between structures. In contrast, SHM for aerospace structures can utilise data from different aircraft of the same model to determine a baseline for a given aircraft [18]. However, each civil engineering structure has to undergo long-term evaluation to determine its normal conditions under various operational and environmental conditions [6].

2.1.1. Statistical Pattern recognition

The SHM process can be approached from a statistical pattern recognition perspective. Pattern recognition is the process of finding regularities or patterns within data. By referencing an already-established database, it is possible to categorise data into groups. For example, it would be a trivial task for a person to classify a group of images if they either included a cat or a dog (given that they have seen both animals). That person would be referencing an internal database of past experiences to identify if the image had a cat or a dog. The same principle can be applied to SHM, but the problem would be to identify if there is damage present in a building. Similar to the cat and dog comparison, it would be necessary to reference a database of what constitutes an undamaged and damaged structure. This means that information of both states is required for damage detection. However, information on the undamaged state or the normal condition of the structure is more readily available.

Pattern recognition for SHM is achieved by working with measurement data from sensors, such as accelerometers, to identify the normal condition and potential damage. Unlike the cats and dogs example, a person is not looking at the data to determine if it represents a damaged or undamaged structure. Instead, pattern recognition for SHM can be done via machine learning algorithms. Machine learning is a principled set of mathematical methods that enables the extraction of features and patterns from data that can be exploited for decision-making. These methods ”learn” from data and make predictions based on the data, making them data-driven methods.

The statistical pattern recognition approach for SHM can be split into four distinct processes [27]

1. Operational evaluation
2. Data acquisition
3. Feature extraction and selection
4. Statistical modelling for feature discrimination

Operational evaluation

Operational evaluation is a preliminary phase in the SHM process to determine the benefits and possible problems with SHM on a given structure. The phase aims to answer four questions regarding the implementation of SHM technology for a given structure. The first question is about assessing the economic and life-safety benefits that SHM technology can bring. The second part is determining how damage will be defined, the possible damage scenarios, and which of them is the most critical. The third question is determining the possible operational and environmental variability present in the structure. The final question is assessing the limitations of data acquisition on in situ structures.

Data acquisition

This phase of the SHM process is to answer questions regarding the sensor configuration. It involves selecting sensor types, their location and how many should be used. It also includes all auxiliary equipment regarding data storage and transfer. To achieve a cost-effective damage detection framework, it is crucial to consider the economic aspect of the selection of sensors. Ideally, an effective SHM should minimise the number of sensors to limit costs.

Feature extraction and selection

Feature extraction involves selecting meaningful data from the raw measurement data that is sensitive to damage, which allows for the comparison between damaged and undamaged states in the next phase. There are many possible damage-sensitive features. For example, the features can be modal properties, such as the natural frequencies. Likewise, the mode shape of the corresponding natural frequencies can be used as features. One of the challenges in selecting a damage-sensitive feature is that features sensitive to damage are also sensitive to changing operational and environmental conditions.

Another issue with the selection of features is their dimensions. Machine learning is employed to analyse these features and identify potential damage. Crucial to the success of these machine learning methods is the quality and the dimension of the features. If the features are not correlated with damage, then the machine learning process will not be able to identify damage. Furthermore, the performance of machine learning algorithms depends on the feature dimension. This is sometimes referred to as the curse of dimensionality [4]. It can be stated as follows: As the number of features or dimensions grows, the required amount of data to generalise the problem accurately grows exponentially. Increasing the dimension leads to the volume of the space increasing faster, resulting in the data becoming sparse in this space. To achieve the same level of statistical robustness as for lower dimensions, the space needs to be sufficiently populated with data to fill the gaps, but that requires an unrealistic amount of data. In order to circumvent this problem, the data can be embedded into a low-rank subspace, reducing the dimensionality of the problem. For instance, instead of taking an entire frequency vector as a feature, only a single index of a natural frequency can be taken, reducing the dimension to a single scalar. Principal component analysis (PCA) is a common dimensionality reduction technique that represents the data according to the maximum correlation within the data set.

Statistical modelling for feature discrimination

The final process of SHM concerns identifying and quantifying the damage using the selected features. Machine learning algorithms can learn the relationships within the data and construct models that can predict and classify new data. In the case of damage detection, machine learning algorithms can create models based on selected features. This model can then predict whether new data is either part of a damaged or undamaged system, automating the discovery of damage. These machine learning algorithms fall into two categories: Supervised learning, where the model is trained on labelled data. Here the state of the features is known whether they are part of the damaged or undamaged systems. To be able to use supervised learning, data must be available from both the damaged and undamaged systems. However, information regarding the damaged state is not readily available. The goal of SHM is to detect when said damage occurs, which means that this data is unavailable when the machine learning model is created. On the other hand, unsupervised learning has no labels given to the data, and the goal is to find patterns within the data in a principled way to classify the data. However, in SHM, the undamaged state is known and therefore, novelty detection is a more appealing method for damage detection. Novelty detection involves detecting anomalies from a baseline condition. In the case of SHM, the anomalies would be features that represent damage, while the baseline condition is the normal condition of the structure.

The existence of damage is only a fraction of the damage state. Rytter [36] proposed a five-part hierarchical damage detection framework seen below. Each subsequent question requires an increasing level of knowledge of the damage state. The discussion in this thesis will be limited to the first question,

which is the existence of damage. However, an ideal SHM system would have answers to all of these questions.

1. Is there damage in the system (existence)?
2. Where is the damage in the system (location)?
3. What kind of damage is present (type)?
4. How severe is the damage (extent)?
5. How much useful (safe) life remains (prognosis)?

In summary, the statistical pattern recognition approach aims to extract and select damage-sensitive features from measurement data and create models that can distinguish between damaged and undamaged features. This chapter is only a short overview of the statistical pattern recognition paradigm for SHM. A more thorough discussion of the process can be found in [27].

2.1.2. Operational and Environmental Variability

One of the major challenges in implementing a successful damage detection framework for SHM for in-service structures is the presence of operational and environmental variability [39]. The changes in operational and environmental conditions can alter the measured response characteristics, which can be falsely interpreted as damage. A robust damage detection framework should reduce or remove the effect of this variability. Temperature is the most common environmental variability, and it can vary throughout the day and between seasons which can alter a system's dynamic properties.

Operational variability depends on the structure. For instance, wind turbines can have different operational conditions depending on the rotation speed of the blades, pitch angle and nacelle direction [24]. The operational variability encountered in bridges concerns the traffic loading on the bridge. Vehicles can cause different excitations depending on their mass and speed. The weight of the vehicles can also alter the system's dynamic properties [14, 29].

The problem of this variability can be tackled to some extent [18, 39] but is still an issue for SHM. The approach taken in this thesis will utilise the advancements in sparsity and compressed sensing [3, 16, 9] to limit this variability. The analysis of this thesis uses the acceleration recordings of several sensors. These are output measurements of the system. Equation 2.1 shows the frequency representation of a generalised dynamic system with input $F(\omega)$, output $R(\omega)$ and the transfer function $H(\omega)$. In the analysis of this thesis, only the output or response is known. Any change to either the excitation (input) or the system properties (transfer function) can lead to a different response. The introduction of damage can alter the stiffness, mass and damping properties of a given system. This will be reflected as a change in the transfer function. Operational and environmental variability can also potentially cause a change in the transfer function, either due to temperature or added mass. Damage detection is a comparison between two states; two different signal groups are compared. The environmental and operational conditions need to be the same or similar in order to avoid misidentifying damage as different operational and environmental conditions.

$$F(\omega) \cdot H(\omega) = R(\omega) \tag{2.1}$$

However, there is also a significant variance in the input, which can be attributed to the operational conditions that the bridge is subject to. Different vehicles have different masses and speeds that can lead to varying input. This makes it difficult to quantify the dynamic properties and its transfer function as only the measured output is available. Sparsity and Compressed sensing is a promising field in data driven science and engineering that offers potential solutions to existing engineering problems. This thesis will utilise concepts from this field, specifically the Robust Principal component analyis [11], to reduce or limit the operational variability in the input. Sparsity and compressed sensing will be discussed later in the chapter.

Literature studies

There have been many studies on the effects of operational and environmental variability on bridges. This section will review several papers on the topic.

Alamosa Canyon bridge in New Mexico, USA, exhibited daily variation in modal properties due to temperature changes [12]. The bridge is oriented in the north-south direction, which results in the sun heating the east side in the morning. This creates an east-west temperature differential across the deck, which results in a 6% daily variation in the modal frequencies.

The Z24 bridge was a pre-stressed concrete bridge in Switzerland. The bridge was the subject of a variety of experiments under the SIMCES project [35]. Over a span of a year, a monitoring campaign was in place to measure the modal parameters of the bridge and environmental variables such as temperature and humidity. During this time, it was closed to traffic and was only under ambient loading. Before it was demolished in 1998, it underwent a series of damage scenarios where damage was introduced to the structure. From the data gathered from the bridge, a bilinear relationship was discovered between ambient temperature and the natural frequencies [34]. When the ambient temperature dropped below freezing, there was a significant change in the natural frequencies. This was attributed to an increase in stiffness in the asphalt layer of the bridge.

Temperature variation can also alter the structural configuration of a bridge [20]. In this paper, the modal properties of a composite bridge change dramatically in colder weather. The expansion bearings of the bridge were partially constrained in cooler temperatures. This resulted in a nonlinear relationship between temperature and the natural frequencies. The first three resonance frequencies of the bridge decreased by 12.3 %, 16.8% and 9.0 %, respectively, when the temperature increased from -17.8 °C to 15.6 °C. There was little change in the natural frequencies at temperatures higher than 15.6 °C.

The weight of the vehicles on the bridge can alter the dynamic properties of the bridge. The Tamar bridge is a suspension bridge in the southwest of England. The bridge has been under long-term monitoring, and research was conducted on three years of data [14]. The traffic loading or the number of vehicle mass on the bridge was discovered to have a dominant effect on the daily fluctuation of the natural frequencies. Similarly, it was found that seasonal variations in temperature affected the natural frequencies. The second natural frequency of the bridge changed by 4.5 % over a temperature interval of 20 °C. Finally, the wind affected the modal frequencies at high wind speeds and when the wind was sideways to the bridge.

Another long-term monitoring campaign was on the D. Henrique Bridge in Porto, Portugal. It is a concrete arch bridge, and it was monitored over two years [29]. In the paper, the added vehicle mass during the morning rush hour caused a small shift of the first natural frequency of the bridge. Furthermore, during the morning rush hour, there was a detected increase in the damping ratio of the second natural frequency and was attributed to vehicle-bridge interaction.

There have also been literature reviews on the effect of environmental variability [41]. The research concluded that the majority of reviewed studies show that the natural frequencies show a negative correlation with temperature. As temperature increases, the resonance frequencies decrease. The variation in the material modulus under different temperatures was considered to be the main reason for the variation in the natural frequencies. The research also included a comparative laboratory study on the temperature effect on different structural materials; steel, aluminium and reinforced concrete. The natural frequencies of structures from different materials responded differently when the temperature increased by a single centigrade. The natural frequencies of steel, aluminium and reinforced concrete decreased by about 0.02 %, 0.03 % and 0.15 %, respectively. This change was independent of the structural type and which mode was under investigation. Finally, for most bridge structures, there was no significant change in the mode shapes, as there is little to no variation in the temperature in the longitudinal direction of a bridge.

These papers are just a fraction of the available literature on the challenge of operational and environmental variability faced in the field of SHM for civil engineering structures. However, they show

that damage-sensitive features, such as the natural frequencies, are also sensitive to operational and environmental variability and need to be considered in any damage detection framework.

2.2. Data Driven Science and Engineering

In recent decades there has been an increase in the usage of data-driven techniques to understand and model complex systems in engineering. This is happening in all fields of engineering, including civil engineering and SHM [2, 37]. This chapter is a brief introduction to machine learning and the concept of sparsity and compressed sensing.

Machine Learning

Machine learning is the use of optimisation methods on data. Using a set of principal mathematical methods makes it possible to create models that ”learn” from data and can subsequently make predictions based on the data. This can, for instance, be the classification of data or the discovery of inherent patterns in data. As mentioned before, there are two main categories of machine learning: Supervised machine learning and unsupervised machine learning. In supervised machine learning, algorithms are given information on the labels of the training data. The goal would then make predictions based on the relationship between features and the labels of the data. In unsupervised machine learning, the training data is not labelled. The goal is to find patterns or coherent structures in the data, which can then be used to generate labels and predict data.

Crucial to machine learning is understanding if a constructed model is either over-fitting or under-fitting the data. When a model is constructed, a data set must be split into training, validation and withholding data sets. Overfitting occurs when a model is too complex and fits the training data too well that it starts to follow detailed patterns in the data related to noise. The accuracy of the model on training data will be high, but it will generalise poorly and have poor accuracy on unseen data from the withhold set. Underfitting, on the other hand, is when the model complexity is simple and does not capture the underlying patterns within the data. To overcome the issue of over-fitting it is necessary to use cross-validation. Cross-validation is the act of splitting the data into random parts and training the model on a single part of the data while testing on the other parts. This is then repeated k number of times to obtain an average performance metric of the model. This leads to a more generalisation of the problem. Another aspect of machine learning models is the creation of accurate models while remaining parsimonious. That is, models with few parameters that are interpretable and generalise well while remaining accurate.

Sparsity & compressed sensing

The mathematics of sparsity and compressed sensing is the foundation of the main method used in this thesis, the robust principal component analysis. The viewpoint of sparsity is that complex systems and phenomena can be described in sparse terms. The motion of classical mechanical systems, for instance, can be described by a few major movements in the system. Sparsity can be used to promote parsimonious models from data that have a minimal number of terms. It can be used to add robustness with respect to outliers, as is done in the rPCA, via sparse optimisation. Compressed sensing aims to reconstruct complex systems and signals from a few sparse measurements.

Two methods in this thesis are based on this theory, The rPCA [11] and the sparse sensor placement for optimisation (SSPOC) [7]. These methods are promising and can potentially solve problems in SHM. There are many other interesting methods that utilise sparsity and compressing that could possibly be used in an SHM context. Sparse sensor placement for reconstruction [30] is an interesting method that aims to reconstruct a signal by exploiting known patterns in data. Recently, these methods have been successfully used in engineering applications [31]. In this research, an alternate approach to the maintenance of aircraft shims was proposed. rPCA and sparse sensor optimisation were used to optimise the collection of data. This reduced the number of measurements required to accurately predict shim

gaps in aircraft assembly. This approach was based on the fact that patterns in measurement data exist in the shim distribution in an airframe that can be recovered. There is potential to apply these methods to problems faced within SHM.

3

The Haringvlietbrug Measurement Project

3.1. Introduction

The Haringvlietbrug is a steel box-girder bridge. Its total length is 1220 metres, with ten similar sections that span 106 metres each and a section for a movable bridge in the north part of the bridge, allowing maritime traffic to pass the Haringvliet. The bridge consists of a hollow box section under the bridge that runs longitudinally along the bridge. Diagonal struts from the hollow section support the cantilever deck plate on either side. Transverse beams are equally spaced along the length of the bridge sections with stiffeners between them. Figure 3.2 shows a structural section of the bridge. The bridge has four lanes of traffic, two in each direction, with an extra lane designated for local low-speed traffic. The road surface consists of an asphalt layer on the bridge deck. The traffic, and thus the loading on the bridge, is not symmetric over its width. The southbound traffic is situated on the westward cantilever deck plate of the bridge, while the northbound traffic drives on the centre of the bridge. The local low-speed lane is located on the east cantilever deck plate.

The bridge is located south of Rotterdam and connects South Holland with Zeeland via the A29 highway. It is an important connection between the two regions for both daily commuters and cargo freight. The average daily weekday traffic was 64400 vehicles in 2019, with around 20 % of the traffic being heavy trucks [1]. Traffic has increased over the years. Figure 3.3 shows the increase in traffic in the past decade for both regular passenger cars and cargo trucks. In 2020 there was a decrease in passenger cars, but in August 2021 renovation of the movable bridge started, and a speed restriction of 50 km/h was put in place [2]. In 2020 there was also a global pandemic. This could explain the sudden decrease at that time. From 2014 to the peak traffic in 2019, there has been a 43 % increase in traffic with a 58 % increase in heavy trucks. The bridge was opened in 1964 it has been in service for nearly 60 years. It can be assumed that there has been a considerable increase in traffic from its opening, something it might not have been designed for. Given its age, increase in traffic load, and that steel box-girder bridges are prone to fatigue damage in welding connections [42], it is not surprising that fatigue cracks have been identified in the bridge.

In 2003 Rijkswaterstaat, which is responsible for the maintenance of main infrastructure facilities in the Netherlands, launched the project Risk: Rijdek Inspectie Stalen Kunstwerken (Bridge Deck Inspection of Steel Structures). The goal of the project was to monitor steel bridges in the Dutch national road network for fatigue damage. In 2017 and 2018, the structural state of the Haringvlietbrug was investigated by visual inspection. The work was carried out by the engineering consultancy Nebest, and a report

[1] Intensiteiten op Wegvakken - INWEVA Source
[2] A29: Renovation Haringvliet draw bridge - Rijkswaterstaat Source

Figure 3.1: The Haringvlietbrug. Looking north at the drawn up moveable bridge. Source : Rijkswaterstaat

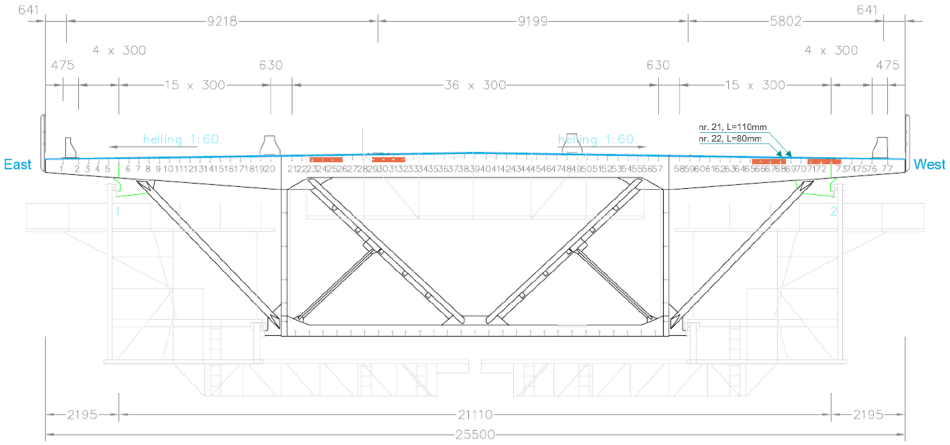


Figure 3.2: Crosssection of the Haringvlietbrug. Heavy traffic location highlighted in red.

was submitted to Rijkswaterstaat. In the report, there were 45 observations of fatigue cracks exceeding the critical length between the stiffeners or bulb profiles and the transverse beams. However, this amount of damage was not considered critical to the bridge's safety and presented an opportunity for an SHM project. The first step in such a project would be the acquisition of data for analysis through a monitoring campaign. As mentioned earlier, traffic over the width of the bridge is not symmetric, with southbound traffic mainly on the west deck plate cantilever. This is an area of concern, as it is a cantilever that carries a considerable dynamic load. Several fatigue cracks were discovered in this area between the stiffeners and the transverse beams. This area of the bridge is the subject of one of the vibration monitoring campaigns and was the basis of the previous study of the Haringvlietbrug [21]. There have been several different monitoring campaigns on the Haringvlietbrug at the time of writing, two of which are vibration-based monitoring campaigns that measure the acceleration response of the structure at various locations. The data used for this study are mainly from these two vibration monitoring campaigns. The accompanying chapters will go into the details of these measurement campaigns. There have also been campaigns that measure the strain and temperature of the bridge at different locations. These measurements were used in a study on the effect of temperature on the natural frequencies of the Haringvlietbrug [26]. The main finding from that research was that the temperature of

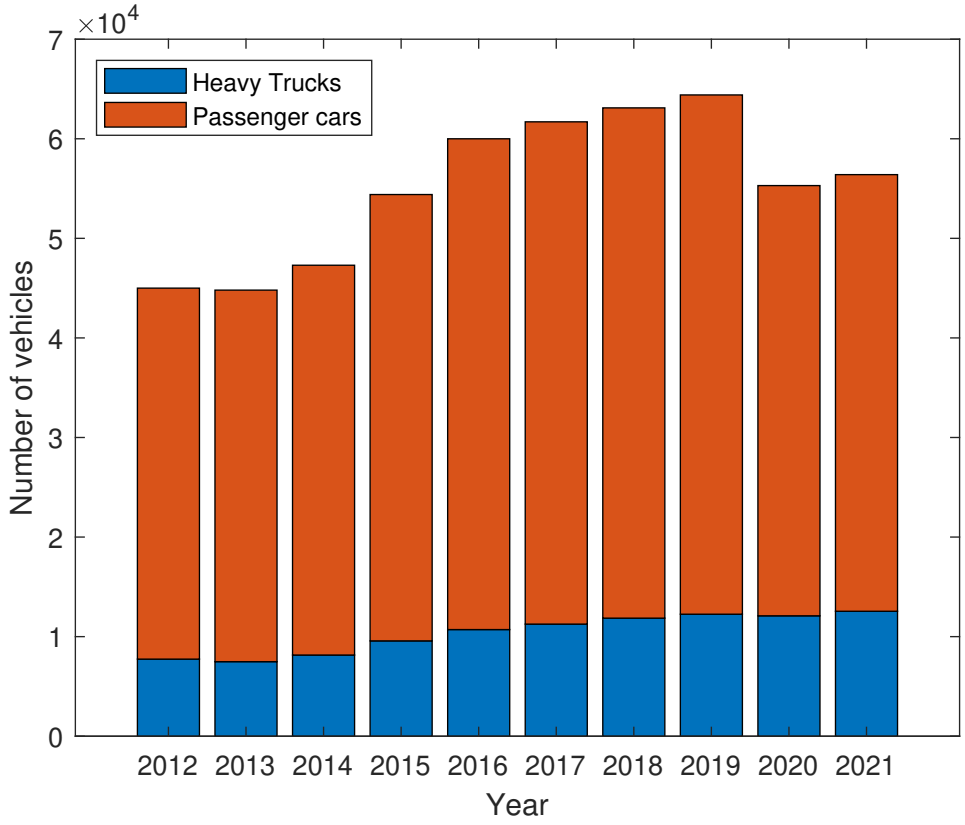


Figure 3.3: Traffic development of the Haringvlietbrug. Source INWEA

the asphalt had a major role in the dynamics of the bridge.

3.2. First Vibration Monitoring campaign

3.2.1. Goal

The first measurement campaign was initiated to acquire data to develop a robust damage detection framework. The accelerometers are located in the west cantilever of the bridge in the vicinity of fatigue cracks. To perform damage detection, comparable data had to be gathered in an undamaged or ”healthy” reference area to allow for comparison between undamaged and damaged vibration signals. Thus, in another undamaged section of the bridge that is structurally similar to the damaged area, the sensors are set in positions identical to those of the damaged area. This was the sensor setup in the previous damage detection study of the Haringvlietbrug [21] and will be used again for this study, but now a different preprocessing technique will be applied to reduce the operational variability. Instead of similarity filtering as was done in the previous research, Robust Principal Component Analysis will be used to reduce the operational variability.

3.2.2. Sensor Configuration

Sensors were installed in two areas, one damaged area in the presence of fatigue cracks and another undamaged reference area that is structurally similar. In both cases, the sensors are located on the west side of the bridge on the cantilever bridge deck that carries the traffic that is bound south to Zeeland and Brabant. The sensors are also in the same section of the bridge that is in the midspan of section 4, which is the second span after the moveable bridge from the north. Figure 3.5 shows the location of the two sensor areas. The baseline reference area is mirrored compared to the damaged area.

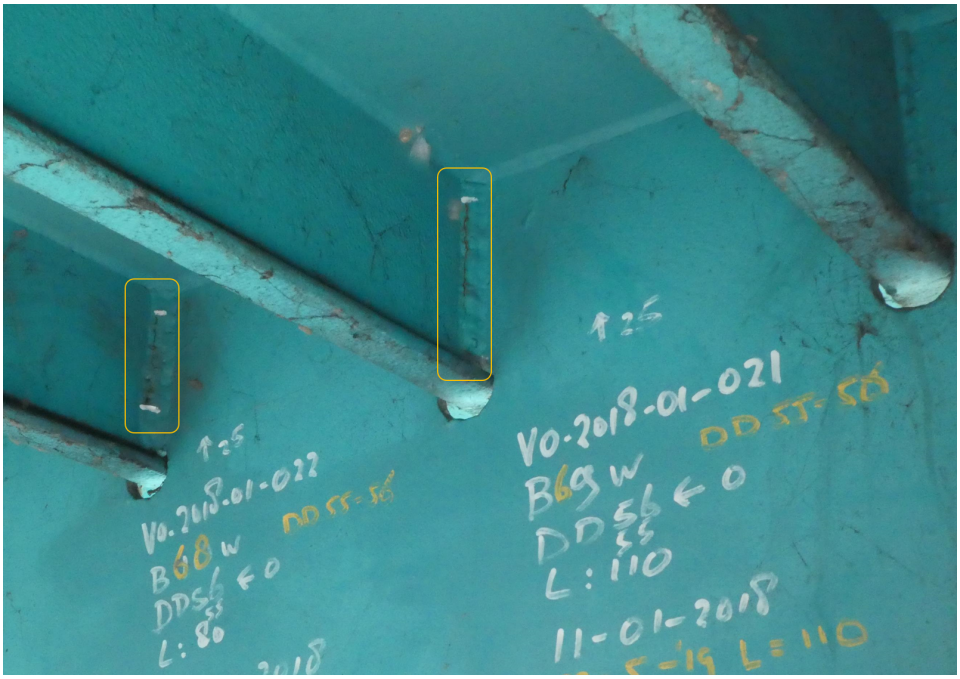


Figure 3.4: Fatigue Cracks Number 21 and 22 between the stiffeners and the transverse beam

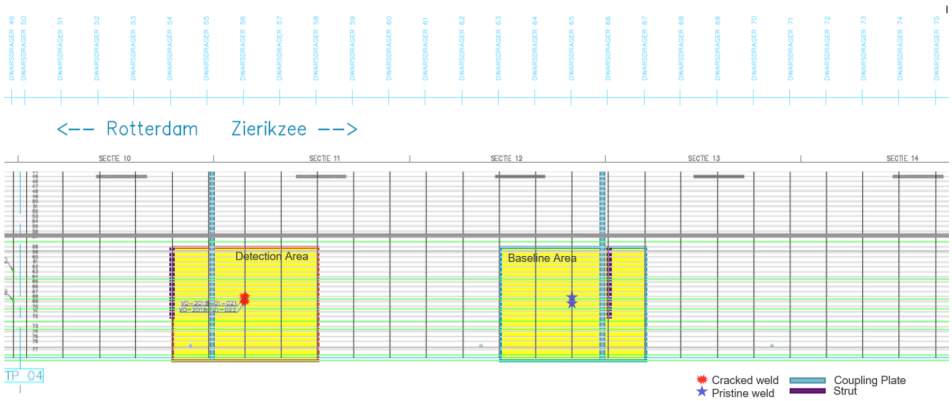


Figure 3.5: Baseline or ”healthy” area and the detection or damaged area. Bridge section 4. Taken from [21]

The goal of SHM is to create a functional and economically viable procedure to monitor the health of structures. Sensors are vital for any SHM project, and the number of sensors used in SHM projects should be minimised to reduce costs. However, to fully capture the nature of structural systems, many sensors may be required. A single sensor can never provide comprehensive coverage of an entire complex structure and cannot detect changes for every possible damage scenario. The number and position of sensors is an optimisation problem that has to be tackled in any SHM project. For this monitoring campaign, a total of 32 uni axial accelerometers are deployed for both areas, so 16 accelerometers for each zone. These accelerometers measure acceleration in the time domain in a single direction. They are positioned in a dense network around the fatigue cracks in the damaged area and in ”identical” positions in the baseline reference area. The proximity of the accelerometers to the fatigue cracks should increase the likelihood that the sensors would capture relevant data that is sensitive to damage. An alternative would be a sensor setup that would capture modal parameters of the bridge, such as the mode shapes. This would lead to a network of more distributed sensors to capture the vibrational modes. However, as the first method of sensor setup is used, there is limited information on the shapes of the vibration modes.

The accelerometers are uniaxial and thus can only measure acceleration in one direction. This makes

the orientation of the sensors as crucial as their position to capture vibration data sensitive to damage. Depending on the position and orientation of the accelerometer, the frequency content captured by the sensor can be vastly different. For both the damaged and undamaged reference areas, the sensor orientation is the same to allow direct comparison. Two accelerometers measure acceleration in the direction parallel to the bridge, defined as the X-direction. These accelerometers are positioned on the transverse beam and measure the out-of-plane movement of the beam and should capture the high-frequency modes that have this out-of-plane movement. Another nine accelerometers are on different stiffeners that measure the acceleration in the Y-direction, which is orthogonal to the bridge in the west-east direction. These sensors also measure the out-of-plane movement of the stiffeners and should capture the corresponding out-of-plane high-frequency modes of this element. Finally, five accelerometers under the bridge deck measure the Z-direction, the vertical motion of the bridge. The corresponding modes that these sensors capture should be low-frequency modes associated with the bending of the bridge deck. Figure 3.6 shows the position of the sensors as well as the orientation of each sensor.

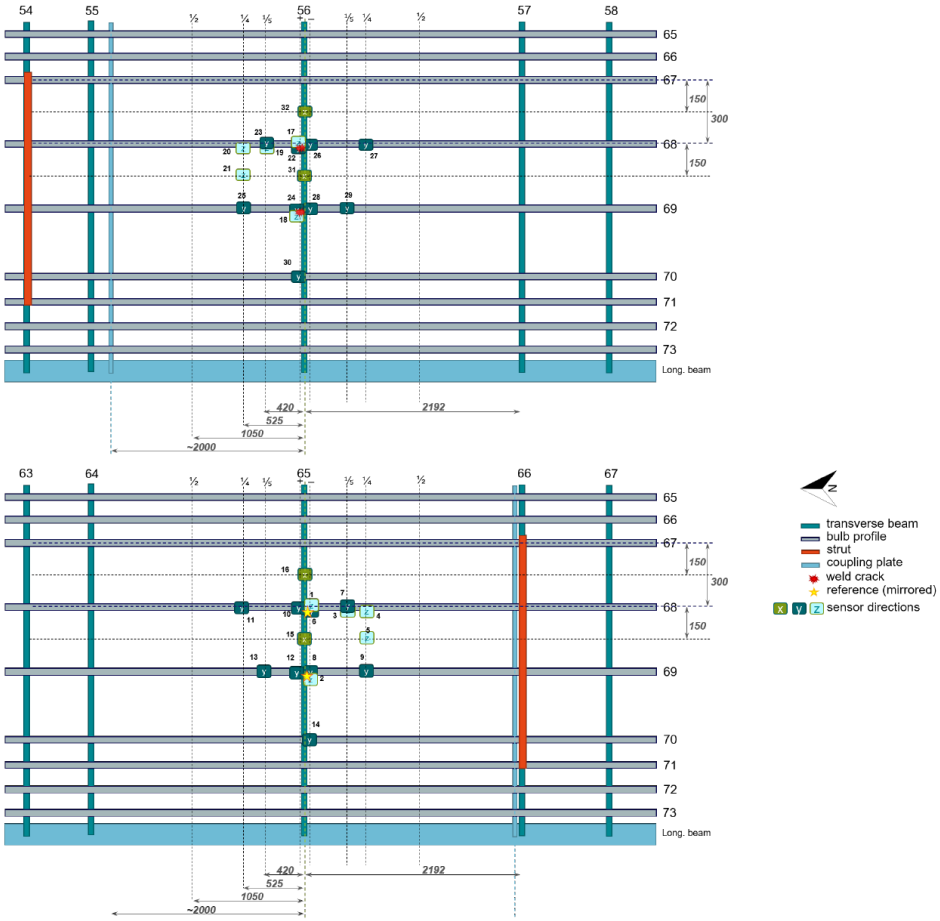


Figure 3.6: Sensor layout of the damage area (top figure) and the Healthy reference area (bottom figure) . Taken from [21]

The accelerometers were installed in April 2020, and the monitoring campaign was active for approximately one year. The accelerometers have a sampling rate of 1000 Hz. The sensors are active throughout the day but only collect data for 15 minutes each hour. The aforementioned sampling rate leads to a data vector of the size 21600000 for a single day for only a single sensor. After the sensors retrieve the data, the data needs to be stored somewhere and await analysis. A cloud-based solution was used to move the data to a remote server via a 4G antenna. Twente University oversaw the implementation of this system and the installation of sensors.

3.3. Second Vibration monitoring campaign

3.3.1. Goal

The second vibration monitoring campaign repurposed the acceleration sensors from the first measurement campaign. Instead of a comparison between two different areas, all sensors were installed in a single section of the bridge. Prior to the launch of this monitoring campaign, it was deemed unsatisfactory to continue with the previous arrangement of sensor layout. The presence of fatigue cracks between the undamaged and damaged areas is not the only difference between the two areas. The two areas have slightly different structural configurations, meaning that if there is any detectable difference in the vibration data of the two areas, it is not possible to conclude whether it is due to fatigue cracks or different structural configurations. This was the conclusion of the previous damage detection study on the Haringvlietbrug, and a different approach was taken with the subsequent second measurement campaign. The goal of the second monitoring campaign was to gain more insight into the global behaviour of the bridge by looking at the mode shapes while also tracking the variation in the individual modes due to environmental variability.

3.3.2. Sensor configuration

The accelerometers in the second measurement campaign are located in the third bridge section from the north. The same 32 uniaxial sensors from the first measurement campaign are used again but are now more dispersed in a single detection area. Of the 32 sensors, 30 measure the bridge deck's vertical acceleration. The remaining two measure the horizontal acceleration of the two separate stiffeners. Again, the sampling rate is 1000 Hz, and measurements are taken for 15 minutes every hour. These sensors are located at the east-mid cross-section of the bridge that carries the northbound traffic. This is different from the first measurement campaign, which had sensors under the west cantilever. Figure 3.7 shows the location of the sensors.

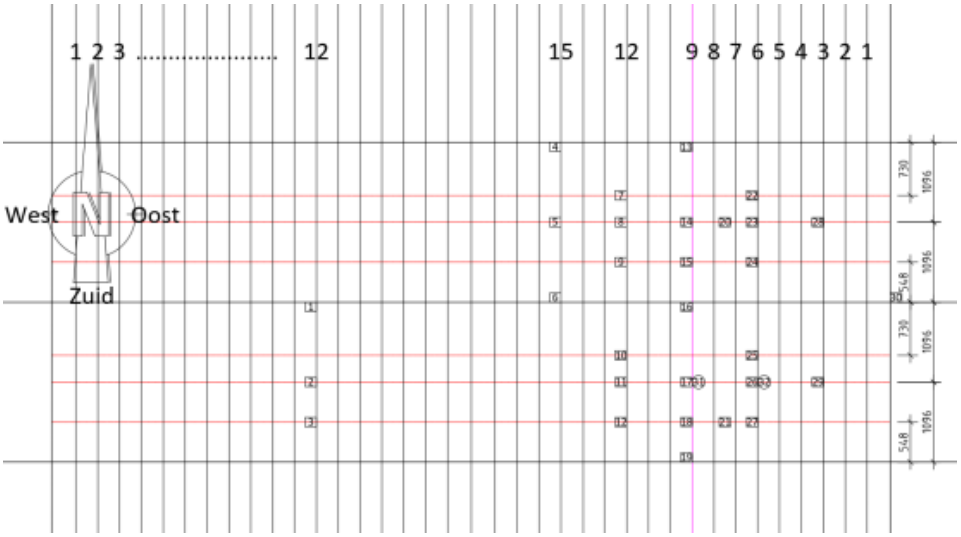


Figure 3.7: Top view of the accelerometer layout of the 2nd Measurement campaign. The vertical lines denote the stiffeners, and the black horizontal lines denote the transverse beams. The figure shows the area within the midsection, that is, not the cantilevers. The numbers in squares are the vertical accelerometers, while the numbers in circles are the horizontal accelerometers.

Alongside these 32 accelerometers, there are also 16 temperature sensors spread over the bridge. These temperature sensors are attached to the various structural components, such as the stiffeners and transverse beams and measure the temperature of said components. There are also sensors embedded into the asphalt layer of the road. Figure 3.8 shows the sensor layout of the temperature sensors. They are spread out over the vertical height of the bridge's midsection. The figure does not include the location of the temperature sensors in the asphalt.

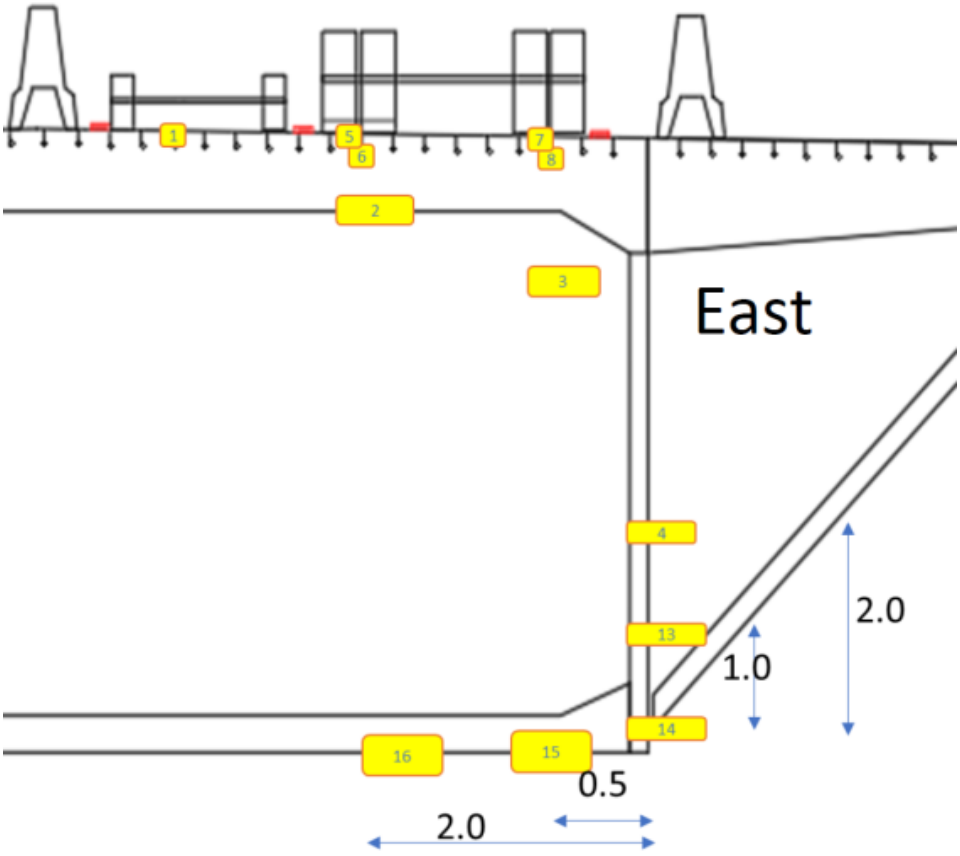


Figure 3.8: Temperature sensor layout

4

Theory

This chapter presents the theory behind the methods utilised in the thesis. The goal of these methods is to manipulate data in a meaningful way. The first half of these methods (Singular Value Decomposition (SVD), Principal Component Analysis (PCA), and Robust Principal Component Analysis (rPCA)) aim to recover the dominant patterns in the data. These patterns can contain relevant information about the underlying dynamic system properties of vibration signals. Robust Principal Component Analysis (rPCA) and Sparse Sensor Placement for Optimisation (SSPOC) are two methods that use sparsity-promoting methods. In the rPCA, sparsity is used to robustify an existing algorithm PCA with respect to outliers. Sparsity-promoting methods are used in the SSPOC to find the minimum number of points in the data to adequately distinguish between data belonging to different groups. Finally, the last methods discussed in the chapter enable the classification of data into distinct groups or classes (Linear Discriminant Analysis (LDA), Support Vector Machine (SVM)).

The usage of the rPCA on noisy vibration data in this thesis is motivated by its promising results in removing outliers in image data as was done in [11]. It is possible that the rPCA can be used to remove noise or outliers within the vibration data and recover the underlying system properties. Similarly, the use of the SPPOC on vibration data is motivated by its usage and effectiveness in other engineering disciplines, as can be seen in the optimisation of aircraft maintenance regimes [31]. Both of these methods utilise sparsity and compressed sensing which is a promising field that can benefit many engineering disciplines.

4.1. Singular Value Decompisation (SVD)

Singular value decomposition (SVD) is a matrix factorisation method that decomposes a matrix into dominant patterns that contain the most correlation within the data and is the foundation of many of the methods in the chapter. It is a generalisation of the eigendecomposition for a square normal matrix and is guaranteed to exist contrary to the eigendecomposition. The SVD can provide a low-rank approximation to a matrix X by only considering the most dominant patterns that contain the highest correlation within the data. High dimensionality is a common problem when working with data from complex systems. For example, image and video data can be highly dimensional depending on the image's resolution, with each pixel being a single index in a data matrix. However, much of the relevant information in an image may be represented in a lower-dimensional subspace. This enables the compression of images, reducing the storage size while retaining the majority of the image information. Complex physical systems, such as classical mechanical systems and fluid flows, can also exhibit high dimensional behaviour, but there may be dominant patterns that describe the system. For example, the majority motion of a classical mechanical system can be described by several mode shapes and their corresponding natural frequencies. Using the SVD, these low-dimensional patterns can be systemati-

cally uncovered from data. It is a powerful tool for dimensionality reduction. The SVD is a data-driven method [8]. The patterns that it uncovers are solely based on the inputted data. It provides a basis *tailored* to the inputted data. In contrast, the Fourier transform provides a *generic* basis, the frequency domain.

Defination of the SVD

Consider a large data set $X \in \mathbb{C}^{n\times m}$:

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix} \tag{4.1}$$

The matrix consists of data that may be obtained with measurements from simulations or sensor equipment. A single column $x_k \in \mathbb{C}^n$ represents a single measurement. Examples of these measurements would be images. An image can be reshaped into a column vector in which a single index represents a pixel's colour value. A data set of m images would form the data matrix X. These images could be of cats and dogs, and the goal with the SVD would be to find the dominant features that distinguish between these two types of animals. The column vectors can also represent the state of a physical system that changes with time. Each column vector is then a snapshot of the system at a given time. The SVD is a unique matrix decomposition that is guaranteed to exist for every complex-valued matrix $X \in \mathbb{C}^{n\times m}$:

$$X = U\Sigma V^* \tag{4.2}$$

Where $U \in \mathbb{C}^{n\times n}$ and $V \in \mathbb{C}^{m\times m}$ are unitary matrices with orthonormal columns, and $\Sigma \in \mathbb{R}^{n\times m}$ is a real-valued non-negative rectangular diagonal matrix. V^* is the conjugate transpose V. For real-valued matrices, the conjugate transpose defaults into a regular transpose. If X is real-valued, then U and V are guaranteed to be real-valued orthonormal matrices. In this case, the SVD can be written as $U\Sigma V^T$. The diagonal entries of Σ are the singular values of X and are ordered from the largest to the smallest, and the number of non-zero singular values determines the rank of X. The columns of U are the left singular vectors, and V are the right singular vectors of X.

The SVD offers a hierarchical low-rank approximation to a matrix X. The singular values in Σ and corresponding column vectors in U and V are ordered according to the value of the singular value in Σ. Keeping the most significant singular values and vectors and discarding the rest can retain most of the information of the data while significantly reducing the dimension of the data. It is possible that only a few singular values and vectors are needed adequately describe the data. This enables the reduction of the size and dimension of the data, giving a manageable basis for visualisation and analysis of the data. Furthermore, in the case of dynamical systems, the SVD provides a hierarchy of possible modes in the data where column vectors represent these modes.

The columns of U and columns V capture the correlation within X. These columns are the eigenvectors of the correlation matrices XX^* and X^*X respectively. The SVD is connected to an eigenvalue problem involving these two correlation matrices. The following equations are obtained if the SVD in equation 4.2 is inserted into these matrices:

$$XX^* = U \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} V^* \, V \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} U^* = U \begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} U^* \tag{4.3a}$$

$$X^*X = V \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} U^* U \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} V^* = V\hat{\Sigma}^2 V^* \tag{4.3b}$$

However, as U and V are unitary, U Σ and V will be solutions to the following eigenvalue problem:

$$\mathsf{X}\mathsf{X}^*\mathsf{U} = \mathsf{U}\begin{bmatrix} \hat{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{4.4a}$$

$$\mathsf{X}^*\mathsf{X}\mathsf{V} = \mathsf{V}\hat{\Sigma}^2 \tag{4.4b}$$

The non-zero singular values of X are a positive square root of an eigenvalue of $\mathsf{X}\mathsf{X}^*$ and of $\mathsf{X}^*\mathsf{X}$. Both of the correlation matrices have the same non-zero eigenvalues. The columns of U and V can then be interpreted as the eigenvectors of correlation matrices $\mathsf{X}\mathsf{X}^*$, and $\mathsf{X}^*\mathsf{X}$ respectively. Since the singular values are ordered from highest to lowest, the columns of U and V are similarly ordered by how much correlation they capture in the columns and rows of X. The SVD and the correlation of data that it captures will be utilised with the Principal component analysis (PCA).

4.2. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a data-driven method that transforms high-dimensional data into a hierarchical coordinate system of principal components (PCs). The principal components are uncorrelated or orthogonal to each other but are directions of maximum variance in the data. In addition, the principal components are ordered in the amount of variance they explain. This enables dimensionality reduction of data by looking at the leading PCs and discarding the rest, increasing the interpretability of data while preserving the maximum amount of variance within the data.

Definition

The PCA can be said to be a statistical interpretation of the SVD. The PCA pre-processes the data by subtracting the mean and setting the variance to unity before applying an eigendecomposition or Singular Value Decomposition. Consider a matrix $X \in \mathbb{C}^{n \times m}$ that consists of n row measurements of m length. These measurements could be images or acceleration recordings.

$$X = \begin{bmatrix} — & x_1 & — \\ — & x_2 & — \\ & \vdots & \\ — & x_n & — \end{bmatrix} \tag{4.5}$$

First, the mean of all rows is computed and subtracted from the data matrix X.

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij} \tag{4.6}$$

The mean matrix is:

$$\bar{\mathsf{X}} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \bar{\mathsf{x}} \tag{4.7}$$

The subtraction of the mean results in the mean-subtracted data B:

$$\mathsf{B} = \mathsf{X} - \bar{\mathsf{X}} \tag{4.8}$$

The covariance matrix is obtained by:

$$\mathsf{C} = \frac{1}{n-1} \bar{\mathsf{B}}^* \bar{\mathsf{B}} \tag{4.9}$$

The first principal component u_1 is:

$$\mathsf{u}_1 = \underset{\|\mathsf{u}_1\|=1}{\operatorname{argmax}} \, \mathsf{u}_1^* \, \mathsf{B}^* \mathsf{B} \mathsf{u}_1, \tag{4.10}$$

u_1 is an eigenvector of $\mathsf{B}^*\mathsf{B}$ that corresponds to the largest eigenvalue. Comparing this to the SVD and the correlation matrics in the previous section establishes that u_1 is a right singular vector of B

corresponding to the largest singular value. The principal components can be obtained by computing the eigendecomposition of C:

$$CV = VD \tag{4.11}$$

The eigendecomposition is guaranteed to exist, as C is a Hermitian matrix. D contains the eigenvalues and V the eigenvectors. The principal components obtained with:

$$T = BV \tag{4.12}$$

Where T are the principal components and here V are the eigenvectors or ”loadings”, which represent how much each principal component contributes to each measurement. This is a ”classical” approach to solving the PCA with eigendecomposition. However, the SVD can be used to compute the PCA. Applying the SVD on B yields:

$$B = U\Sigma V^{*} \tag{4.13}$$

As the right singular vectors are equivalent eigenvectors of $B^{*}B$, inserting the above equation into 4.12 yields:

$$T = U\Sigma \tag{4.14}$$

The principal component can be constructed from the left singular vectors of B and multiplied by a corresponding singular value. In practice the SVD is used to compute the PCA.

Example: Ovarian cancer

The PCA is a powerful tool to represent high-dimensional data in a low-rank subspace. From this low-rank subspace of the principal components, patterns or clusters of the data can emerge, which can be utilised for further analysis. To illustrate the power of the PCA, consider the ovarian cancer data set from Matlab®. This data set contains genetic data from 216 patients, 121 of whom have ovarian cancer, and 95 of whom are cancer free. The genetic data of each ”measurement” or patient consists of the expression of 4000 genes, resulting in a data size of 216x4000. The high dimensionality of the gene expression poses a problem for the interpretation of the data. However, the genetic data is suspected to be highly correlated; patients share an overlap in the expression of these genes. Thus, most of the gene expression variance within the population could be explained by some dominant patterns or principal components. This is shown in figure 4.1, with singular values on the left and the total variance explained by the leading r principal components on the right. A single principal component explains around 80 % of the variance of the data. This implies that the gene expression of the patients is highly correlated.

Recall that 121 of the patients have cancer, and it is expected that their expression in these 4000 genes should be different. Figure 4.2 shows the principal component score of the patients in the subspace of the first three principal components. It appears that the patients with cancer cluster separately from cancer-free patients. The clustering of these two groups in the low-rank subspace of the principal components presents an opportunity for machine-learning algorithms to diagnose new patients if they have ovarian cancer.

Limitations of the SVD and PCA

While PCA and SVD are powerful methods, they cannot be used indiscriminately. It is essential to prepare data with pre-processing before use. The SVD is based on the coordinate system in which the

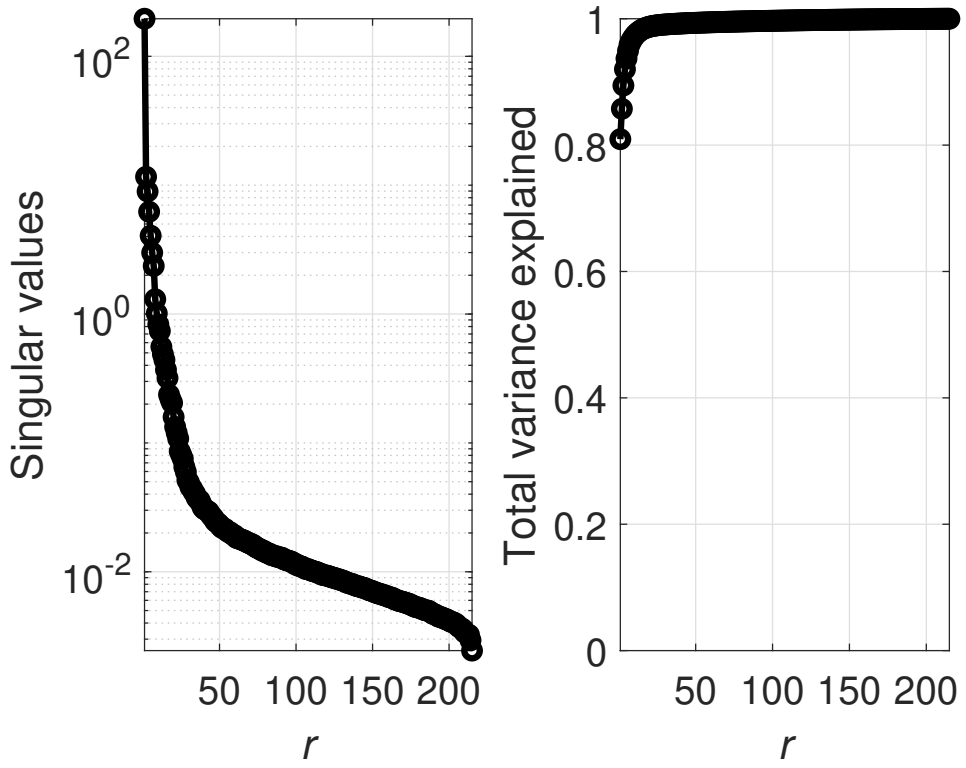


Figure 4.1: Singular values of the ovarian cancer data set and total variance explained of r leading principal components.

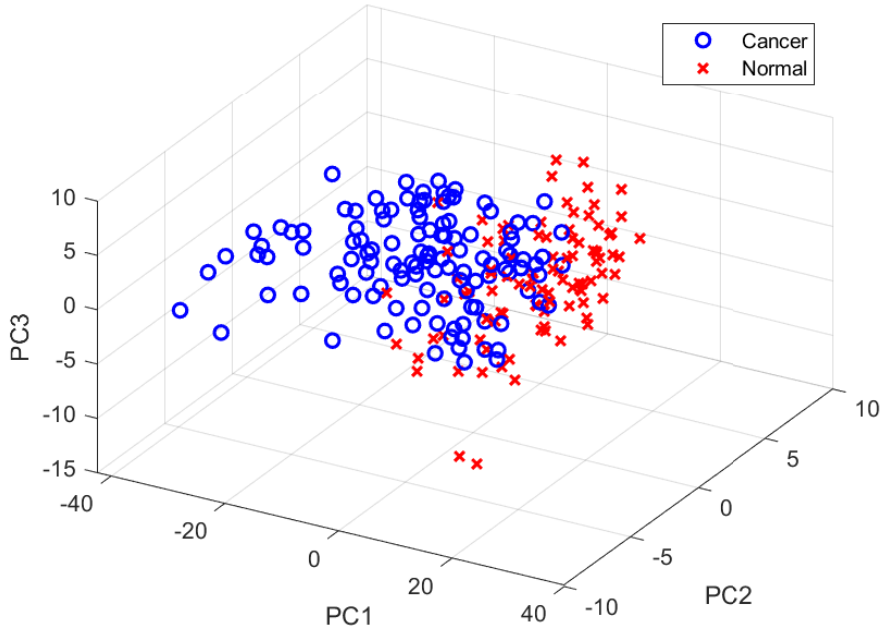


Figure 4.2: Low-rank repsentation of the patient gene data in the subspace of the first three Principal components.

data is represented. Data might require extensive pre-processing for the SVD to generate good results. An example where the SVD and PCA would break down is when there is translation or rotation of the data. For instance, a comparison between images of dogs, but all the mages are rotated by a random amount. The SVD would not be able to find the dominant patterns explaining the correlation within

dogs. This limitation can be seen in the example in the chapter on the Robust Principal Component Analysis.

4.3. Robust Principal Component Analysis (rPCA)

One of the flaws of the PCA is it is highly susceptible to outliers and corrupted data. However, this weakness has been addressed with the development of a robust principal component analysis (rPCA) [11]. The rPCA aims to decompose a data matrix X into a low-rank matrix L and a sparse matrix S containing the outliers:

$$\mathsf{X} = \mathsf{L} + \mathsf{S} \tag{4.15}$$

The underlying problem that the rPCA aims to solve is rather complex. This is an inverse problem, and it is not well-posed at all. It is completely ill-conditioned and underdetermined with n knowns and 2n unknowns. There are infinitely many solutions to this problem. However, this ill-posed problem can be solved with regularisation to promote a single solution. Mathematically the decomposition involves minimising the rank of L, so it is well described by principal components, while S is a sparse matrix by taking the ℓ_0 norm. The principal components of L are *robust* to the outliers and corrupt data in S. L and S can be found by satisfying the following equation:

$$\min_{\mathsf{L},\mathsf{S}} \text{rank}\,(\mathsf{L}) + \|\mathsf{S}\|_0 \quad \text{subject to} \quad \mathsf{X} = \mathsf{L} + \mathsf{S} \tag{4.16}$$

However, both the rank term and the ℓ_0 norm are non-convex. Thus this is not a manageable optimisation problem. There is no guarantee that a computer will find a solution to this problem. Nevertheless, it is possible to solve for the optimal L and S with *high probability* using convex relaxation [10, 16]:

$$\min_{\mathsf{L},\mathsf{S}} \|\mathsf{L}\|_* + \lambda\|\mathsf{S}\|_1 \quad \text{subject to} \quad \mathsf{X} = \mathsf{L} + \mathsf{S} \tag{4.17}$$

Here proxies are introduced for the non-convex terms. The proxy for rank is the nuclear norm $\|\cdot\|_*$ or the sum of singular values, and the ℓ_0 norm is changed to the ℓ_1 norm. $\lambda = \sqrt{max(n.m)}$ is the regularisation parameter, where n and m are the dimensions of X. The solution of 4.17 converges to the solution of 4.16 with *high probability* if $\lambda = \sqrt{max(n.m)}$, given that L and S satisfy the following conditions:

- L is not sparse
- S is not Low-rank; we assume that the entries are randomly distributed so that they do not have low-dimensional column space.

The optimisation problem in 4.17 is known as the *principal component pursuit* (PCP). The augmented Lagrange multiplier (ALM) algorithm may be used to solve this problem. The Augmented Lagrangian can be constructed as follows:

$$\mathcal{L}(\mathsf{L},\mathsf{S},\mathsf{Y}) = \|\mathsf{L}\|_* + \lambda\|\mathsf{S}\|_1 + \langle \mathsf{Y}, \mathsf{X} - \mathsf{L} - \mathsf{S}\rangle + \frac{\mu}{2}\|\mathsf{X} - \mathsf{L} - \mathsf{S}\|_F^2 \tag{4.18}$$

Where $\|\mathsf{M}\|_F = \sqrt{\sum_{ij} |m_{ij}|^2}$ is the Frobenius norm. A general solution would solve for the L_k and S_k that minimises $\mathcal{L}$, update the Lagrange multipliers $\mathsf{Y}_{k+1} = \mathsf{Y}_k + \mu(\mathsf{X} - \mathsf{L}_k - \mathsf{S}_k)$ and iterate until the solution converges. The code provided by the authors of the rPCA [11] uses the Alternating Directions Method (ADM) to solve this optimisation problem [28, 43].

Regularisation parameter

The authors of the rPCA recommend a value of $\lambda = 1/\sqrt{max(n,m)}$ for the regularisation parameter for practical problems. However, the authors state that the regularisation parameter can be changed according to prior knowledge of the solution to improve the performance of the rPCA. For instance, if it is known that S will be very sparse, then a higher regularisation parameter can be taken, which will allow the recovery of a larger rank of L. Likewise, a lower regularisation parameter would yield a lower rank of L.

The Robust Principal Component Analysis has been used to remove corruption in fluid flows and recover coherent patterns [38]. In this study, fluid flow around a cylinder was simulated and corrupted by salt-and-pepper corruption. By using the rPCA, the researchers were able to extract the dominant coherent structures from corrupted fluid flows around a cylinder. They also experimented with different values of the regularisation parameter λ. Lowering the parameter results in more aggressive filtering of the rPCA, and higher-order coherent structures or eigen flows are incorrectly filtered out, and only the first three modes remain. However, using a too-high value of the regularisation parameter λ leads to the corruption not being filtered out. The researchers conclude that there needs to be a better understanding of the selection of regularisation parameter λ for different scenarios in fluid mechanics.

Sparsity

The key to the rPCA is the concept of sparsity and how the ℓ_0 and ℓ_1 norms promote a sparse solution to an undetermined system. As stated earlier, the SVD and PCA are fragile with respect to outliers as they minimise the ℓ_2 norm. The difference between least squares or the ℓ_2 norm, and the ℓ_1 norm is relatively simple. Consider a vector:

$$\mathsf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{4.19}$$

The ℓ_2 norm is:

$$\|\mathsf{x}\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2} \tag{4.20}$$

While the ℓ_1 norm is:

$$\|\mathsf{x}\|_1 = \sum_{k=1}^{n} |x_k| \tag{4.21}$$

The ℓ_2 norm is a squared summation of elements in vector x, while the ℓ_1 norm is the summation of absolute values in x. The ℓ_2 norm is fragile to outliers because it computes the power of two for each element, amplifying its relevance if it is an outlier. The ℓ_0 norm is a pseudo-norm given by the number of non-zero entries. It does not fulfil the definition of a norm. Consider now the system equations:

$$\mathsf{y} = \Theta\mathsf{s} \tag{4.22}$$

This is an undetermined system of equations with y and Θ being known and s unknown. There are infinitely many potential values of vectors s that satisfy this equation. A sparse solution means that only

a few components in s would contribute to a solution. This sparsity criterion would also promote a single solution to these systems of equations. To visualise the effect of sparsity, consider that s is only two variables, s_1 and s_2. Figure 4.3 shows a plane of these two variables and a blue line representing the infinitely many solutions, or combinations of s_1 and s_2, to the system of equations. The orange curves represent the minimum required norm levels that intersect this blue line. These intersection points are minimum norm points that are solutions to the system equations for different lp norm levels. The ℓ_0 norm returns either a value of s_1 or s_2; it is always sparse. The ℓ_1 norm is a summation of s_1 and s_2. The ℓ_2 norm shape is a circle, and the radius is always the minimum required norm level to satisfy the solution. The ℓ_0 and ℓ_1 norm produces the sparsest solution with only one coordinate, s_2 contributing to the solution. In contrast, the ℓ_2 norm uses all available coordinates; it does not produce a sparse solution. An early version of this geometrical interpretation of the ℓ_1 norm and how it promotes a sparse solution was presented in Tibhirani's 1996 paper on the Lasso algorithm (least absolute shrinkage and selection operator) [40].

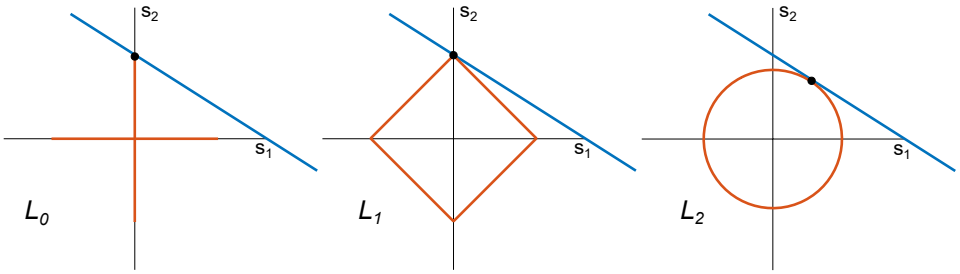
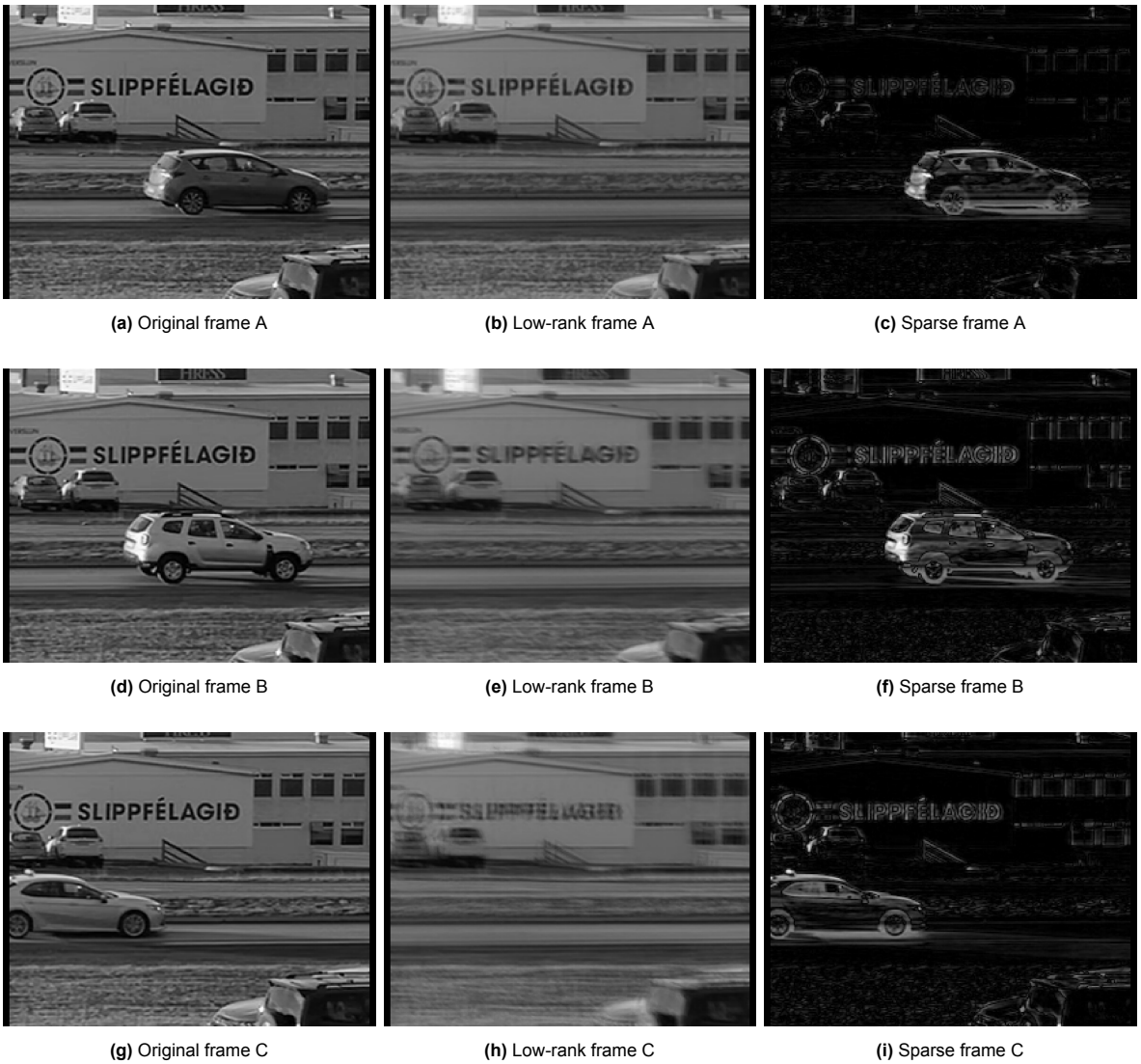


Figure 4.3: The minimum norm point on a line for different ℓ_n nroms. The blue line represents the solution set of an under-determined system of equations. The orange line represents the minimum norm level sets that intersect the blue line. The ℓ_0 and ℓ_1 produce the sparsest solution with only one coordinate active.

This sparsity-promoting effect of the ℓ_1 norm scales well with increasing dimensions as its corners "stick out" more [33]. The corners represent the sparse solutions, as was seen in figure 4.3.

Example - Video Surveillance

The decomposition of the rPCA into low-rank and sparse components has various applications. Depending on the problem, either the low-rank L or the sparse matrix S could be of interest. For example, consider video surveillance of traffic, where cars appear in the video's foreground while the background is unchanged throughout the video. The majority of the video is characterised by the background. Any deviation in the background image could be considered an outlier, such as the traffic. Said another way, the cars in the foreground are not part of the statistical similarity of the background. Figure 4.4 shows the rPCA decomposition of the frames of a traffic video. Each frame is 224x288 pixels and is reshaped into a column vector before applying the rPCA. The length of the video is 30 seconds at 30 frames per second, producing a total of 900 frames. The size of the data matrix X is 900x65412. In the video, cars are travelling along a road. The rPCA effectively filters out these cars into the sparse matrix S; they are not present in any of the frames in the low-rank matrix L. The low-rank L is reconstructed based on the statistical similarity between frames: the background while excluding the outliers, the cars. The low-rank frames appear to be at a lower resolution. This is because the camera was not completely stationary during recording, and thus the background slightly changes between frames. The PCA and the rPCA are fragile to the misalignment of the data. They depend on the coordinate system in which the data is being represented. In this problem, it could be of interest to track the objects in the foreground; the sparse matrix S is the variable of interest. A clear image of the background could also be desired, such as getting a good image of a crowded tourist destination by removing undesired tourists.



(a) Original frame A **(b)** Low-rank frame A **(c)** Sparse frame A

(d) Original frame B **(e)** Low-rank frame B **(f)** Sparse frame B

(g) Original frame C **(h)** Low-rank frame C **(i)** Sparse frame C

Figure 4.4: Comparison between the Original, low-rank and Sparse data frames.

4.4. Sparse sensor placement optimisation for classification (SSPOC)

Sparse sensor placement optimisation for classification (SSPOC) [7] uses sparsity-promoting techniques to determine sensor locations that contain the most discriminating information between n categories of data. The method is inspired by advances in compressed sensing [3, 16, 9], a framework that enables the reconstruction of signals with few measurements. Consider a data matrix X with two categories of data, $c = 2$, with r number of features Ψ_r (e.g. PCA modes). It is possible to identify a discrimination vector $w \in \mathbb{R}^r$ (e.g. LDA directions) in the subspace of Ψ_r that is most informative in discriminating between the two different categories in X. A Sparse vector s that finds the measurements that best reconstruct w is found by:

$$\mathbf{s} = \underset{\mathbf{s}'}{\operatorname{argmin}} \|\mathbf{s}'\|_1 , \quad \text{subject to } \Psi_r^T \mathbf{s}' = \mathbf{w} \tag{4.23}$$

This is a convex optimisation problem, and the ℓ_1 minimisation is used to find a sparse solution. s is a sparse vector with most entries being zero, and it contains only r non-zero entries. These non-zero entries in s are sensor locations that best recapture the discriminant projection vector Ψ_r. The method can be expanded for data with more than two categories with the projection X to decision space, then

independently solving equation X for each column in w. This approach, on the other hand, scales poorly with c, with at most $q = r(c-1)$ learned sensor locations. Another approach is to solve for the columns of s $\in \mathbb{R}^{n\times(c-1)}$ simultaneously, where each column of s projects to a column w in feature space. A penalty term λ is coupled with a norm that penalises the total number of non-zero rows in s to reconstruct the $c-1$ columns of w. This yields the following equation:

$$\begin{aligned} \mathbf{s} = &\underset{\mathbf{s}'}{\operatorname{argmin}} \left\{ \|\mathbf{s}'\|_1 + \lambda \|\mathbf{s}'\mathbf{v}\|_1 \right\}, \\ &\text{subject to } \left\| \Psi_r^T \mathbf{s}' - \mathbf{w} \right\|_F \leq \varepsilon, \end{aligned} \tag{4.24}$$

where v is a column vector of $(c-1)$ ones, $\|\mathrm{M}\|_1 = \sum_{ij} |m_{ij}|$, $\|\mathrm{M}\|_F = \sqrt{\sum_{ij} |m_{ij}|^2}$ is the Frobenius norm, and ϵ is the error tolerance.

An example method is shown in figure 4.5 and is generated by code provided by [7]. In this example, the data matrix X contains images of cats and dogs, leading to two categories c = 2. The SSPOC algorithm aims to find pixel locations in the images that most explain the difference between cats and dogs. The figure shows the discriminant projection vector $\Psi_r w$ and the optimised sensor locations in red in a vector format and a matrix image format. The most discriminating features between cats and dogs, according to SSPOC, appear to be the eyes, nose, mouth and ears.

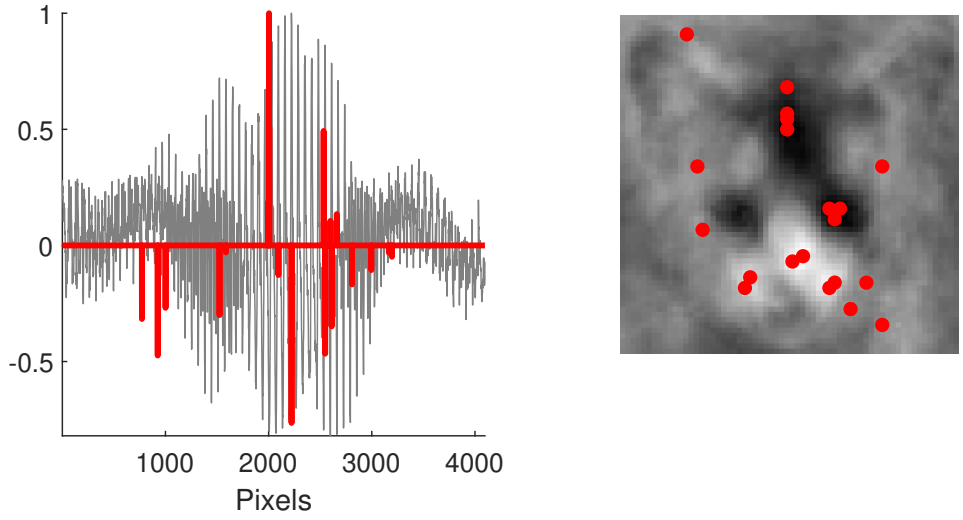


Figure 4.5: (a) A visualisation of the discriminant projection vector $\Psi_r w$ and sparse approximations s. (a) The discriminat projection vector $\Psi_r w$ reorganised into an image format and optimised sensor locations in red. Produced from code provided by [7]

The researchers also conducted experiments on the Extended Yale Face Database B. The sensor locations computed by the SSPOC algorithm that discriminated between three different faces were around the eyes, nose, corners of the mouth and eyebrows. These locations are similar to those in which humans look at faces to identify a person.

4.5. Mahalanobis Distance

Mahalanobis distance is a measure of the distance between a point to the centroid of a distribution in a multivariate space. The squared Mahalanobis distance from a vector x to a distribution with mean μ and covariance Σ is as follows:

$$\Delta^2 = (\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \tag{4.25}$$

The distance represents how far vector x is from the mean or centroid of the distribution in standard deviations. The Mahalanobis distance essentially constructs a new coordinate system based on the distribution of the data. The method works best for data that are approximately multivariate normal. The Mahalanobis distance can be used in multivariate anomaly detection by reducing the dimensionality of the problem.

4.6. Linear Discrimiant Analysis (LDA)

Linear Discriminant Analysis (LDA) was introduced by Fisher in 1936 [19]. It is one of the oldest supervised methods for classification problems. The aim of LDA is to find an optimal low-dimensional subspace that has a clear separation between different groups of labelled data. There are many possible subspaces in which the data can be represented. However, the goal of the LDA is to solve an optimisation problem that finds an optimal subspace that exhibits a clear separation between the different classes of data. This optimal subspace will make classification easier. Figure 4.6 shows the general idea of LDA. In this case, two possible subspaces or projections onto new bases are shown. In the projection on the left, there is considerable overlap between the two groups of data. This projection does not produce a clear separation between the data. However, the projection on the right produces a better separation of data. Said another way, the means of the two distributions μ_1 and μ_2 have considerable distance between them in this projection. The LDA aims to maximise this distance between distributions of the different classes while minimising the intra-class data.

The distributions seen in the figure follow that of a normal distribution. The projection assumes the features of the low-dimensional subspace follow a multivariate normal distribution. This is a fair assumption, as the linear combination of the original features is approximately a normal distribution, according to the central limit theorem.

Two Class LDA

A data set with a feature vector $x \in \mathbb{R}^D$ can be projected to low-dimensional feature $z \in \mathbb{R}^L$ with a linear projection vector **w**:

$$z = \mathbf{w}^T \mathbf{x} \tag{4.26}$$

For an LDA with two classes $\mathcal{C}_1$ and $\mathcal{C}_2$ that have N_1 and N_2 number of points in the respective classes, the mean vectors of the two classes are given by:

$$\mu_1 = \frac{1}{N_1} \sum_{n \in \mathcal{C}_1} \mathbf{x}_n, \quad \mu_2 = \frac{1}{N_2} \sum_{n \in \mathcal{C}_2} \mathbf{x}_n, \tag{4.27}$$

The projection of the mean of the classes onto the line **w** provides a simple measure of separation between classes and is given by:

$$m_k = \mathbf{w}^T \mu_k \tag{4.28}$$

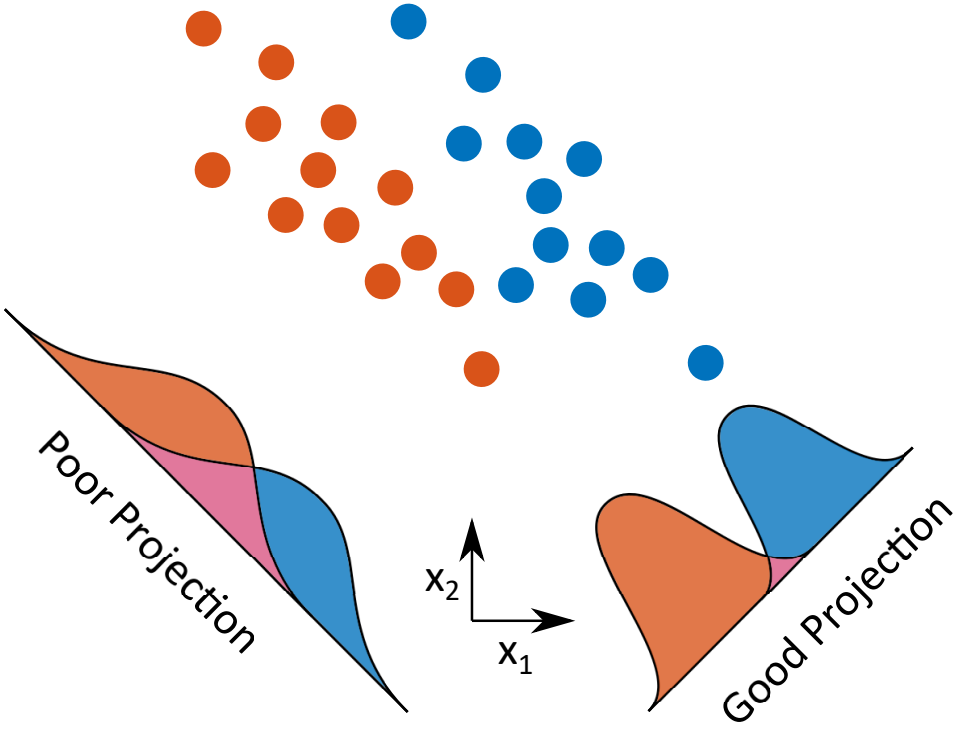


Figure 4.6

This projection can also be made for every data point, resulting in $z_n = \mathbf{w}^T\mathbf{x}_n$, which can be used to compute the variance of the projected points:

$$s_k^2 = \sum_{n \in \mathcal{C}_k} (z_n - m_k)^2 \tag{4.29}$$

The goal of LDA is to find a projection **w** that maximises the distance between data from different classes while minimising the intra-class data. This is done by maximising the distance between the means:

$$J(\mathbf{w}) = \frac{(\mu_2 - \mu_1)^2}{s_1^2 + s_2^2} \tag{4.30}$$

This equation can be rewritten in terms of **w**:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}} \tag{4.31}$$

Where $\mathbf{S}_B$ is the between-class scatter matrix and $\mathbf{S}_W$ is the within-class scatter matrix and are given by:

$$\mathbf{S}_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T \tag{4.32}$$

$$\mathbf{S}_W = \sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mu_1)(\mathbf{x}_n - \mu_1)^T + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mu_2)(\mathbf{x}_n - \mu_2)^T \tag{4.33}$$

When the function 4.31 is maximised, it yields a criterion commonly known as the generalised Rayleigh quotient [8], and the solution can be found with a generalised eigenvalue problem:

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w} \tag{4.34}$$

The maximum eigenvalue lambda and the corresponding eigenvector give the projection basis. From this projection basis, a discriminant can be constructed by choosing a threshold z. New data from the test set will be classified depending on where it falls into the low-dimensional subspace and the chosen threshold value. For example, data fulfilling $z(x) \geq z_0$ could be classified as belonging to class $\mathcal{C}_1$ while data that does not fulfil this condition would belong to class $\mathcal{C}_2$. The class-conditional densities are assumed to be normally distributed, and the parameters of said distributions can be extracted and used to find the optimal threshold.

Multiclass and multidimensional case

The LDA can be extended to include more than two classes $K > 2$ and higher dimensions. It is assumed that the dimensionality of the input space D is larger than the number of classes. The goal is to find a projection matrix **W** that maps from D, the input feature dimension, to L, the low-dimensional feature. $D' > 1$ linear "features" $z_k = \mathbf{w}_k^T\mathbf{x}$ are introduced where $k = 1, ..., D'$. Both the features and the projections can be grouped together to form vector **z** and matrix **W** respectively:

$$\mathbf{z} = \mathbf{W}^T\mathbf{x} \tag{4.35}$$

Similar to the previous case, a projection matrix can be found by maximising:

$$J(\mathbf{W}) = \frac{|\mathbf{W}^T\boldsymbol{\Sigma}_B\mathbf{W}|}{|\mathbf{W}^T\boldsymbol{\Sigma}_W\mathbf{W}|} \tag{4.36}$$

Where the between-class and within-class covariance matrices are, respectively:

$$\boldsymbol{\Sigma}_B = \sum_k \frac{N_k}{N}(\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \tag{4.37a}$$

$$\boldsymbol{\Sigma}_W = \sum_k \frac{N_k}{N}\boldsymbol{\Sigma}_k \tag{4.37b}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n\in\mathcal{C}_k}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \tag{4.37c}$$

N_k is the number of data points in a given class $\mathcal{C}_k$, and $N = \sum_k N_k$ is the total number of data points in the whole data set. The means are defined as:

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n\in\mathcal{C}_1}\mathbf{x}_n, \quad \boldsymbol{\mu} = \frac{1}{N}\sum_{k=1}^{K} N_k\boldsymbol{\mu}_k, \tag{4.38}$$

The solution to equation 4.36 can be found by an eigenvalue problem and yields the projection:

$$\mathbf{W} = \boldsymbol{\Sigma}_W^{-\frac{1}{2}}\mathbf{U} \tag{4.39}$$

Where U are the L leading eigenvectors of $\boldsymbol{\Sigma}_W^{-\frac{1}{2}}\boldsymbol{\Sigma}_B\boldsymbol{\Sigma}^{-\frac{1}{2}}$. The projection **W** can thus be computed.

4.7. Support Vector Machines (SVM)

Support vector machines (SVM) is a supervised machine learning algorithm that finds a hyperplane that splits data into distinct categories based on training data. Many possible hyperplanes exist that can separate the data, and the SVM is an optimisation to find the optimal hyperplane that separates the data groups according to their labels. This optimisation problem involves finding a decision line which minimises the labelling errors of the data while maximising the margin between the data. The method was developed in the 20th century, and the current ”standard” approach (soft margin) was introduced by Cortes and Vapnik [13].

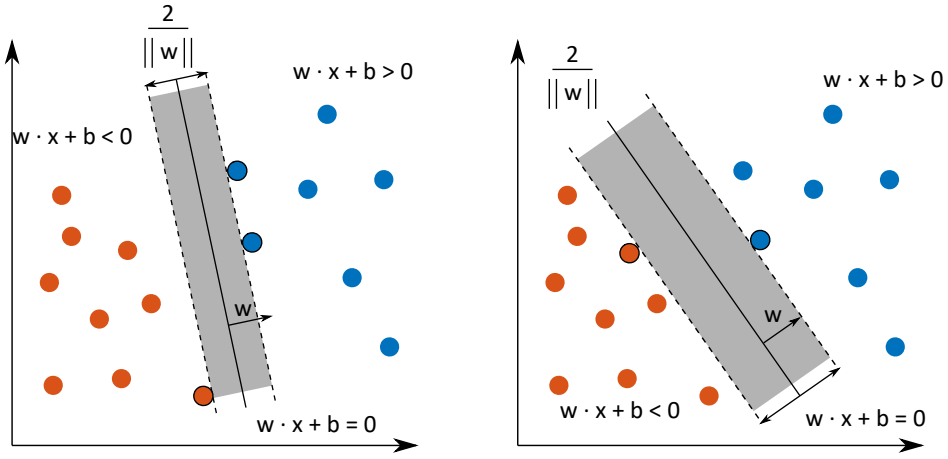


Figure 4.7: Two possible hyperplanes for the SVM.

The aim of the linear SVM method is to construct a hyperplane:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{4.40}$$

The vector w and constant b parameterise the hyperplane. Figure 4.7 shows two possible hyperplanes with different values of w and b to split the same data set. The SVM is an optimisation to find the optimal hyperplane that separates the data. This optimisation problem involves finding a decision line which minimises the labelling errors of the data while maximising the margin between the data. The margin is shown in grey in the figure. The support vectors are the vectors that determine the boundaries of the margin, located at the edge of the margin in figure 4.7. After defining a hyperplane, new data points x_j can be classified based on which side the points are to the hyperplane by computing the sign of $(w \cdot x_j + b)$. This will produce classification labels $y_j \in \pm 1$ depending on which side the data points are in relation to the hyperplane. Now the optimisation objective can be assembled. The goal of the optimisation is to minimise labelling errors. This can be done by defining a loss function:

$$\ell\left(\mathbf{y}_j, \bar{\mathbf{y}}_j\right) = \ell\left(\mathbf{y}_j, \text{sign}\left(\mathbf{w} \cdot \mathbf{x}_j + b\right)\right) = \begin{cases} 0 \text{ if } \mathbf{y}_j = \text{sign}\left(\mathbf{w} \cdot \mathbf{x}_j + b\right) \\ +1 \text{ if } \mathbf{y}_j \neq \text{sign}\left(\mathbf{w} \cdot \mathbf{x}_j + b\right) \end{cases} \tag{4.41}$$

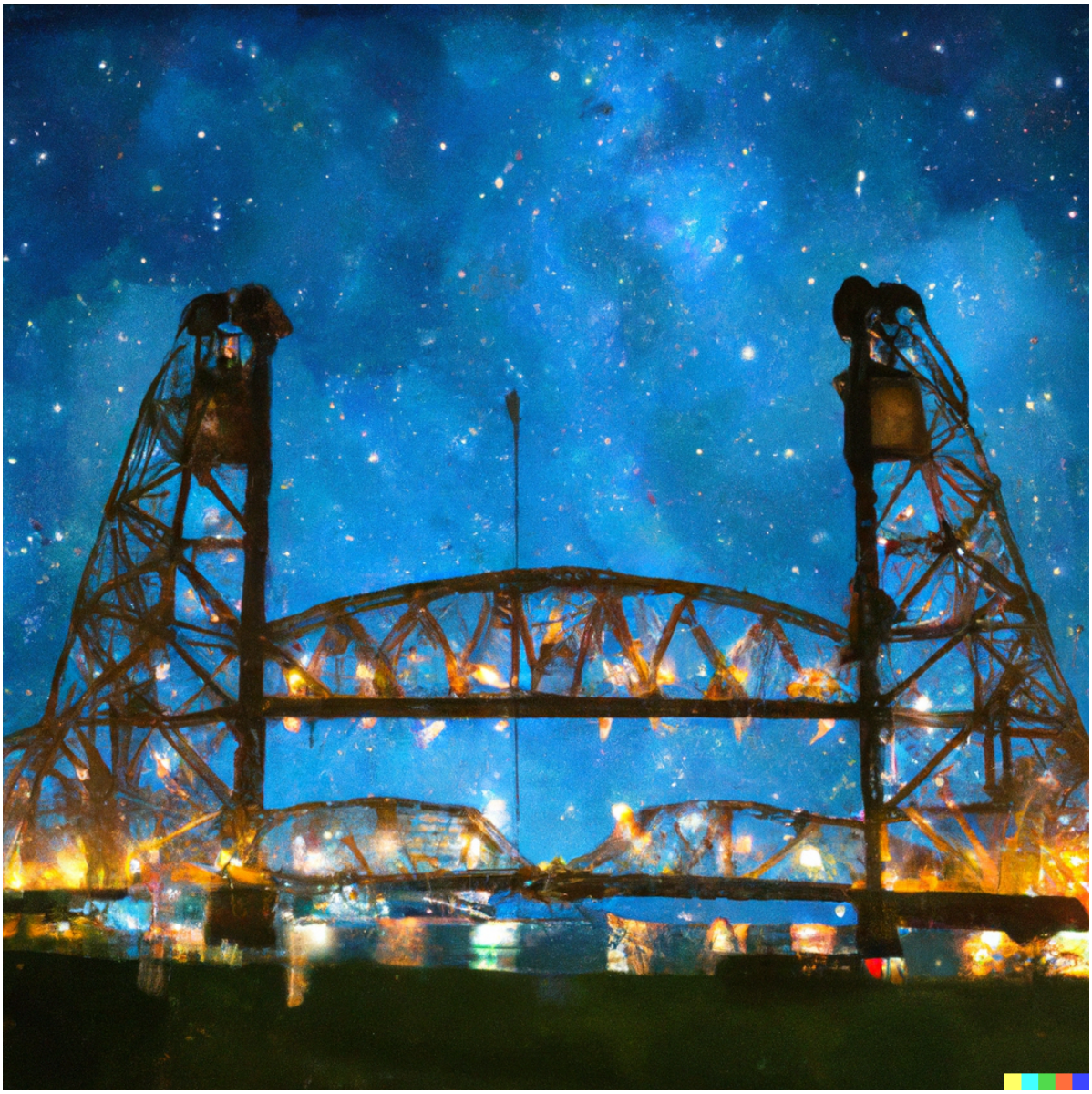
That is the loss function, either 0 if the data is correctly labelled or +1 if the data is incorrectly labelled. Each mislabelled point produces a loss of unity. The total training error is the sum of the loss function $\ell(y_j \bar{y}_j)$. The optimisation problem also involves maximising the margin. The optimisation objective function can then be framed as follows:

$$\underset{\mathbf{w},b}{\text{argmin}} \sum_{j=1}^{m} \ell\left(\mathbf{y}_j, \bar{\mathbf{y}}_j\right) + \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } \min_j |\mathbf{x}_j \cdot \mathbf{w}| = 1. \tag{4.42}$$

However, the loss function is discrete and assembled from binary values. This makes it difficult for optimisation algorithms based on gradient descent to function as they require a smooth objective function in order in order to update the solution. A Hinge loss function $H(z) = max(0, 1 - z)$ can be introduced to create a smooth objective function.

$$\underset{\mathbf{w},b}{\text{argmin}} \sum_{j=1}^{m} H\left(\mathbf{y}_j, \overline{\mathbf{y}}_j\right) + \frac{1}{2}\|\mathbf{w}\|^2 \text{ subject to } \min_j |\mathbf{x}_j \cdot \mathbf{w}| = 1 \tag{4.43}$$

The Hinge loss function allows for piecewise differentiation, which enables optimisations based on gradient descent to be used. The optimal linear hyperplane can now be found to separate two classes of data. A more thorough discussion of SVM and an explanation of the multiclass SVM can be found at [33, 5].



“*An oil painting in the style of starry night of De Hef bridge*” by Dall-E 2

5

Preprocessing

5.1. Data Examination

Data Sampling

Before the data from the accelerometers is manipulated in any meaningful way, it is beneficial to inspect these acceleration signals. Table 5.1 shows the data used for this research. A single day of data is available for the first measurement campaign, while data from the second measurement campaign consists of two days under different temperature conditions. The data consists of acceleration recordings from 32 sensors in each campaign. The accelerometers are active during a 15-minute period each hour. Figure 5.1 shows a 15-minute time series of two vertical sensors, each from the two different measurement campaigns. In both measurements, many ”spikes” are present in the time series signal. It seems as if there is an impulse in the measured response. It is a reasonable assumption that these ”spikes” are due to the traffic on the bridge. The bridge is under both loading from both the traffic and ambient vibrations. However, it can be assumed that the traffic produces a higher amplitude in the measured response. The occurrence of these ”spikes” also seems to be characteristic of traffic. The time between spikes seems to be in the range of 2-3 seconds. This can be seen in figure 5.2. Two to three seconds is the recommended clearance between vehicles [1]. However, there are instances where the time between impulses goes as low as one second. This could indicate a reckless driving culture on the Haringvlietbrug or the presence of multiple lanes. Vehicles can drive parallel to one another on different lanes, resulting in a lower time between impulses than the recommended clearance between vehicles. There are a total of 5 lanes of traffic on the Haringvliet bridge. Two are northbound to Holland, and two are southbound to Zeeland. There is also a lane for local low-speed traffic. In the first measurement campaign, the accelerometers are located on the west cantilever that carries two lanes of traffic headed south. The sensors are positioned between these two lanes and should mainly capture the acceleration caused by vehicles on these two lanes. They could also be capturing the acceleration of the northbound traffic on two lanes at the centre of the bridge.

Table 5.1: Measurement data used in the analysis of this thesis.

Campaign	ID	Date	Start time	End time	Temperature range
1st Campaign	516 - 539	21-05-2020	00:25:37	23:25:37	XXX C
2nd Campaign	09762 - 09785	18-10-2021	00:14:57	23:29:57	11.5-17.5 C
	10147 - 10170	03-11-2021	00:17:01	23:32:01	7.9 - 12.5 C

[1] Samgöngustofa Source

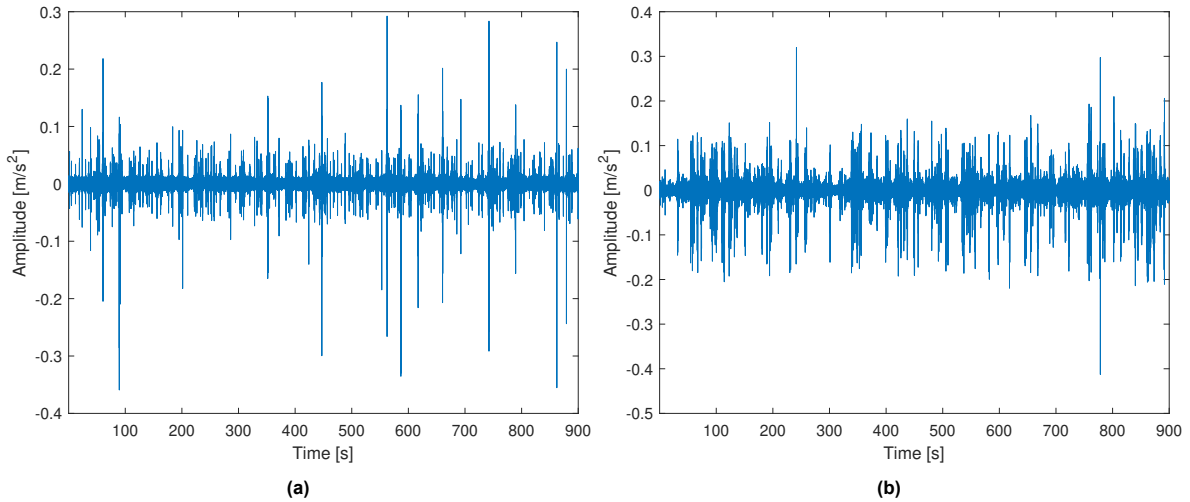


Figure 5.1: Time series acceleration output of vertical sensors over a 15-minute interval during the afternoon rush hour. (a) Vertical sensor 3, located in the ”Healthy” area from the 1st measurement campaign. Measurement number: 00532 (b) Vertical sensor 17 from the second measurement campaign. Measurement number: 09775

These observations provide a foundation for the data segmentation necessary for further analysis. At each of these impulses caused by a possible vehicle, energy is inputted into the structural system, and the accelerometers capture the response due to this force. Each vehicle can potentially excite the structural system differently. However, the input force is amplified by the dynamic properties of the bridge at the location of a sensor. The dynamic properties can potentially be the same between different vehicle passings, given that the operational and environmental conditions do not change. Thus a data set of different vehicle passings can contain information about the same system, while the vehicles can all potentially cause a different excitation. The excitation due to the traffic can be considered a random process, while the dynamic system properties are a deterministic quantity in this scheme. The measured response is a product of the random process of the traffic and the dynamic properties, and thus the response is also a random process. However, each realisation of the response contains information regarding the dynamic properties. It contains a piece of the puzzle on the characteristics of the dynamic system. By sampling the excitation caused by many different vehicles, it is possible to obtain statistical knowledge of the coherent structures in the response, which correspond to the dynamic properties. It is the goal of the robust principal component analysis (rPCA) to discover the coherent patterns within the data.

The goal now is to create samples that capture the random process of the traffic. It is important to create many samples of the random process to encompass its entire behaviour. The lowest denominator of the traffic is the excitation caused by a single vehicle. A simple and elegant approach would be to create samples with excitation from a single vehicle. As mentioned before, the impulses seen in the time series seem to follow guidelines regarding the clearance between vehicles of two seconds. Thus a two-second time window can be taken around these peaks. Each time window or sample would then ideally contain energy from a single vehicle passing. However, there are multiple lanes, and vehicles can drive side by side and in opposite directions. Each sample does not necessarily contain an excitation from a single vehicle. This condition depends on the traffic intensity of the bridge. With more traffic, there is a higher probability that vehicles are driving parallel to each other, and there are more ”meetings” between vehicles driving in opposite directions.

Figure 5.2 shows how a sample is created in the time domain. The red box is a time window around peaks in the time series, and each of them represents a single sample. The two figures show different scenarios regarding traffic intensity. In figure 5.2a, there is low traffic, and the peaks are well separated. In contrast, the traffic intensity in figure 5.2b is much higher, and the time window between different samples intersects on multiple occasions. In these cases, the distance between peaks is less than 2 seconds. This choice of a time window of 2 seconds for each sample leads to a frequency resolution of 0.5 Hz in the frequency domain.

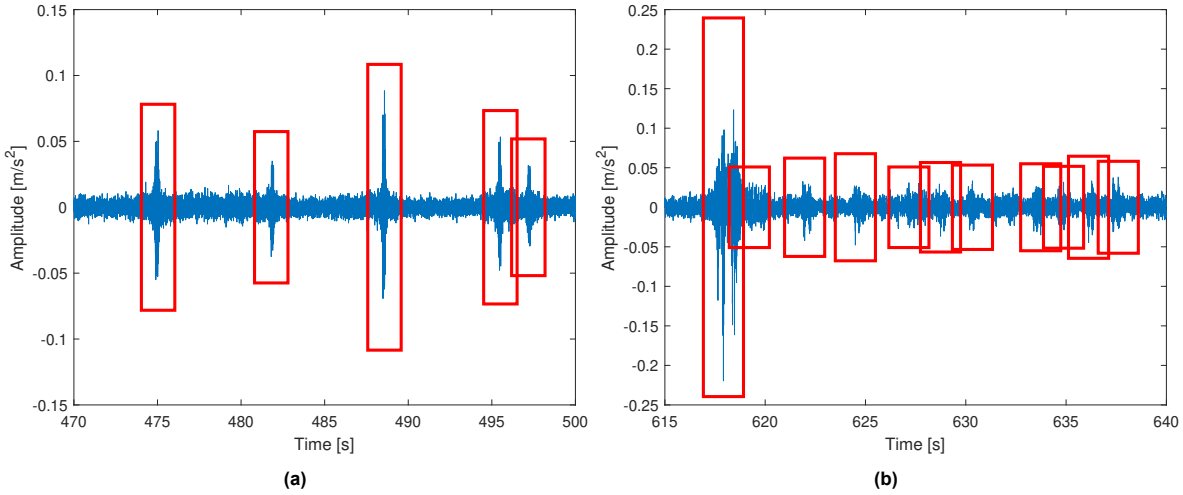


Figure 5.2: Creation of samples of the time series data. The red box indicates the time window taken around each sample, which is 2 seconds. Note that the samples can share time series data with adjacent samples. This happens when the assumption distance between the maxmium amplitude of each sample is not less than 2 seconds. (a) Vertical sensor 3, located in the ”Healthy” area from the 1st measurement campaign. Measurement number: 00532 (b) Vertical sensor 17 from the second measurement campaign. Measurement number: 09775

Truncation

This approach to data segmentation also enables the filtering of samples. There is a considerable variation in the amplitude of these impulses in the time series. This could indicate that vehicles driving on the Haringvliet bridge are driving at various speeds and have different masses. It is known that around 20 % of traffic consists of cargo trucks, as seen in figure 3.3 according to data from INWEVA. It seems likely that these trucks can potentially cause a higher excitation. To better visualise that the samples have varying amplitude, a histogram of the ℓ_2 norm or the energy of each sample is plotted in figure 5.3a. The figure contains data from both days in the second measurement campaign, specifically from sensor 31. Higher amplitude signals should have higher energy. The ℓ_2 norm or energy of the sample is used here as, according to Parseval's theorem, the energy is preserved when performing the Fourier transformation B.3. The ℓ_2 norm is invariant to the Fourier transformation. This is important when the samples are transformed to the frequency domain.

The histogram reveals that there seem to be two distinct distributions of the energy of each sample. These distributions seem to follow that of normal distribution. These two distributions can possibly be correlated with the different types of vehicles on the road. The distribution with the lower mean energy could consist of excitation from smaller passenger cars, while the higher mean energy distribution could be connected to cargo trucks on the road. To better support this claim, a comparison can be made by comparing the cumulative distribution function in figure 5.3b with the ratio of passenger cars and cargo trucks in figure 3.3, which contains data from INWEA. The tail end of the low-energy mean normal distribution ends around 1 m/s^2. The cumulative number of samples at this point is around 80 %, so the passenger cars should correspond to 80 % of the total traffic. This is in accordance with the data from INWEVA, where the percentage of passenger cars is also around 80 %.

As the goal is to reduce or eliminate operational variability within the traffic, there is an incentive to discard samples from either distribution to simplify the data set. The ”passenger car” distribution has much more samples than the ”cargo truck” distribution and has less variance. Thus the ”cargo truck” samples can be discarded to simplify and reduce the variance within the data set. This truncation is done per sensor for both measurement campaigns. **For each sensor 20 % of the samples with the highest energy are discarded, similar to what was described above**. The truncation is performed on the time domain data, but most of the subsequent analysis is done in the frequency domain. However, this is not an issue as, according to Parseval's theorem, the energy is invariant to the Fourier transformation, which means that the truncation of samples according to their energy is the same in the time and frequency domains.

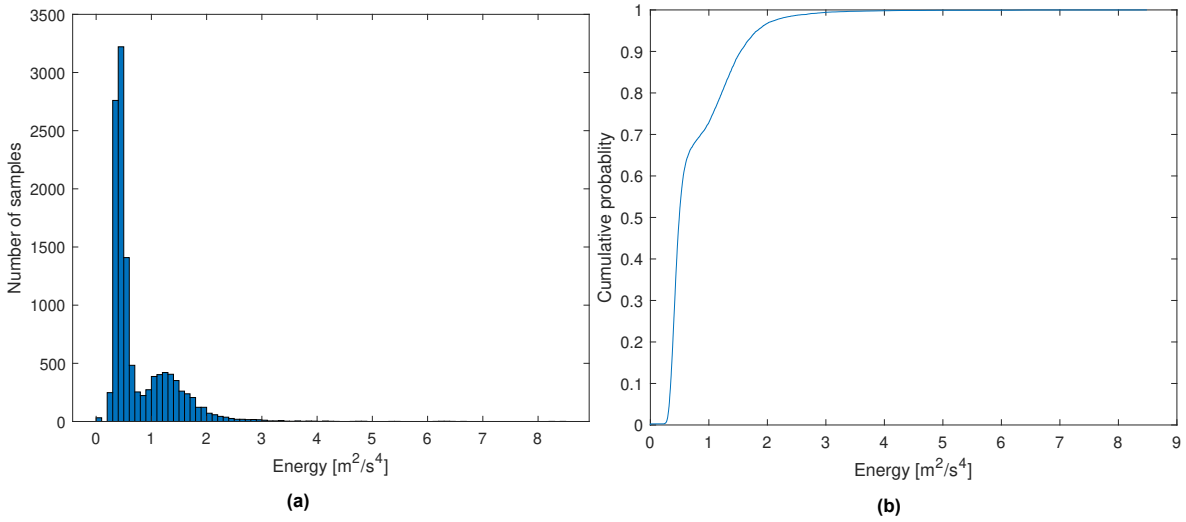


Figure 5.3: Distribution of the ℓ_2 norm of each sample or energy within each sample. (a) Histogram of ℓ_2 norm of each sample. Notice the two different normal distributions with different means and variances. The lower amplitude distribution should represent the passenger cars while the other should correlate to the cargo trucks (b) Cumulative probability function of the Histogram.

A note should be made about the differences in traffic between the two measurement campaigns. In August 2021, maintenance of the draw bridge started, and the maximum speed limit on the bridge was reduced from 90 km/h to 50 km/h, and the width of the lanes was reduced [2]. Therefore the speed limit is 90 km/h for the first measurement campaign and 50 km/h for the second measurement campaign. While the speed limit decreased, that does not necessarily mean the daily number of vehicles decreased. It can be assumed that during this lower speed limit period, there will be higher traffic intensity and congestion as daily commuters before and after this change can potentially be the same. This means that vehicles are potentially more ”packed”, and the clearance between them is generally lower. This translates to more densely populated peaks in the measured acceleration response. An indication of this can be seen in figure 5.2b. There is a higher probability that samples will intersect, containing energy from different vehicle passings. The lower speed limit also means that the vehicles' speed is lower, and thus the energy input into the system by these vehicles might be lower.

The speed limit was not the only aspect that was changed in the operational conditions of the bridge during this maintenance period. The road width was reduced with concrete barricades, thereby altering the traffic path, so the excitation is spatially different. This also changed the mass distribution of the structure as concrete barriers on the bridge deck were moved to redirect the traffic, which can alter the dynamic characteristics of the bridge. Both of these points can potentially lead to different response measurements, which can be identified as damage as the ”normal” conditions have changed. However, in the case of this study, there is no comparison between data from the first and second measurement campaigns, so this operational variability is not present in any of the analyses. Still, it illustrates the difficulties of implementing a robust SHM framework.

Visualisation

In order for the robust principal component analysis to be of any use, the data needs to be on a Fourier basis. This means moving the time domain samples to the frequency domain via a discrete Fourier transform or, more specifically, the Fast Fourier transform. The sampling rate is 1000 Hz or 1000 samples per second, and according to the Nyquist–Shannon sampling theorem, the perfect signal recovery requires a sampling rate twice that of the highest frequency present in a signal. Therefore the highest detectable frequency for a sampling rate of 1000 Hz is 500 Hz. The ”default” time window for each sample in the time domain is two seconds, which translates to a frequency resolution of 0.5 Hz.

[2] A29: Renovation Haringvliet draw bridge - Rijkswaterstaat Source

This time window size will be mainly used in the upcoming analysis, but there are occasions where a higher frequency resolution is required, which requires a larger time window.

Figure 5.4 shows the frequency spectrum of six random samples from horizontal sensor 31 from the second measurement campaign. This sensor is located on one of the stiffeners and captures the out-of-plane motion of the stiffener. The absolute value of the Fourier representation of the data is taken, which is the complex magnitude. Even at this stage, there is some hint of a dynamic system with peaks at around 190 Hz, 270 Hz and 450 Hz, although the signals are particularly noisy. Given that we see these peaks, there has to be energy from the excitation at these frequencies. There is considerable variance in the amplitude of the peaks between samples, both in the total amount of energy present in the frequency domain and at what frequencies the ”energy” is present. This can be seen in the varying amplitudes of the different peaks between the samples. A notable example is a comparison between the top middle and top right plots, where the peak at 200 Hz is the largest in the top middle plot, but all the peaks in the top right figure have equivalent amplitudes. These two plots also have the most ”energy” with much higher amplitudes than the other four samples. These samples are part of the truncated data set where 20 % of the highest energy samples are filtered out. Even then, there is still some variance present in the energy of the samples. Broadly speaking, it seems that the random process of the traffic is a broadband phenomenon. However, the ”energy” over this frequency band seems to vary in magnitude between samples and the distribution of ”energy” over this frequency band varies between samples. This seems to indicate that the random process of the traffic is broadband, but it is definitely not as simple as white noise.

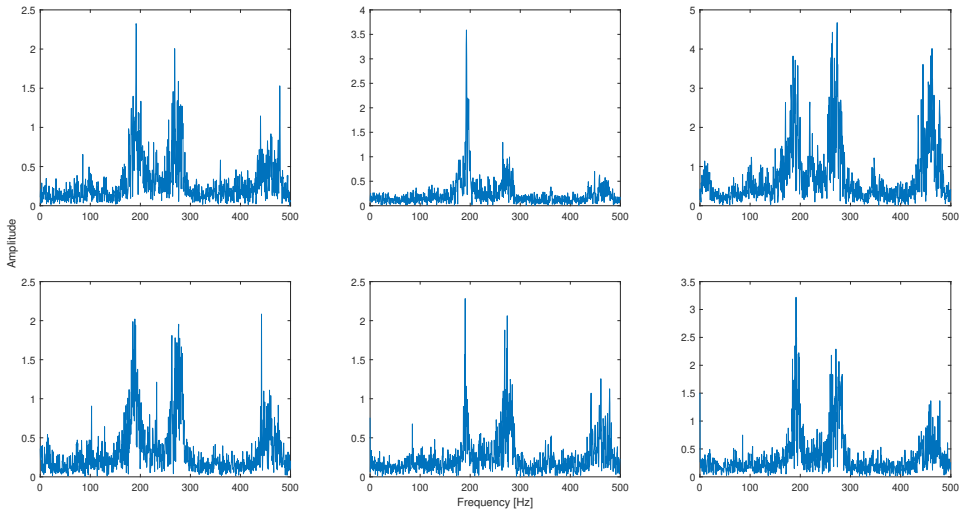


Figure 5.4: Frequency spectrum of 6 random samples from horizontal sensor 31 from the 2nd Measurement Campaign. There seems to be peaks located at 190 Hz, 270 Hz and 450 Hz. Note that the amplitude between the samples is different.

Each of the accelerometers capture different modal properties and corresponding natural frequencies of the bridge depending on their location. For instance, there is quite some difference between the vertical and horizontal sensors. Figure 5.5 shows samples from vertical sensor number 17 from the second measurement campaign. This sensor is located under the bridge deck and captures its vertical motion. The ”energy” content of this sensor is much different from that of sensor 31, with the ”energy” concentrated at the lower frequencies. These sensors are, in fact, capturing the different mode shapes and their corresponding frequencies. Vertical sensor number 17 on the bottom of the bridge deck is possibly capturing a low-frequency mode of the bending motion of the bridge deck. Horizontal sensor number 31 captures higher-frequency modes associated with the out-of-plane bending motion of the stiffener. It also reveals that ”energy” from the random process of the traffic is present at both low and high frequencies.

In both cases, the samples have a high amount of noise due to the random process of the traffic. A straightforward filtering technique to remove the noise is simply to take the average of the samples in the frequency domain. Caution should be taken in which samples are averaged. Temperature alters

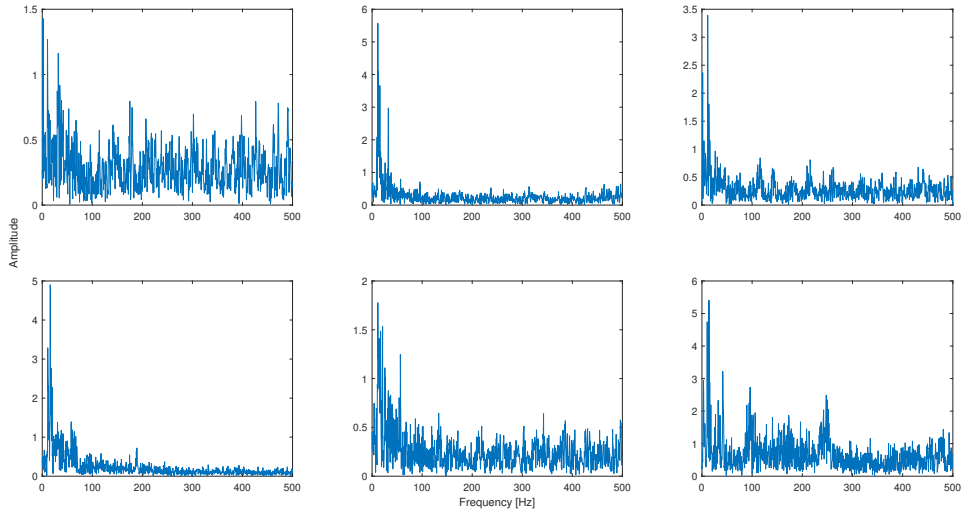


Figure 5.5: Frequency spectrum of 6 random samples from horizontal sensor 17 from the 2nd Measurement Campaign. There seems to be a peak located at 10 Hz. Note that the amplitude between the samples is different.

the dynamic properties of the structural system and is reflected as a change in the frequency spectrum. Thus to obtain the mean response for a given system, the average sample set should have minimal variance in the environmental conditions. The temperature should be the same for every sample. **The approach taken here is to group all samples within a single centigrade, so there is minimal variance in the environmental conditions and the underlying dynamic properties.**

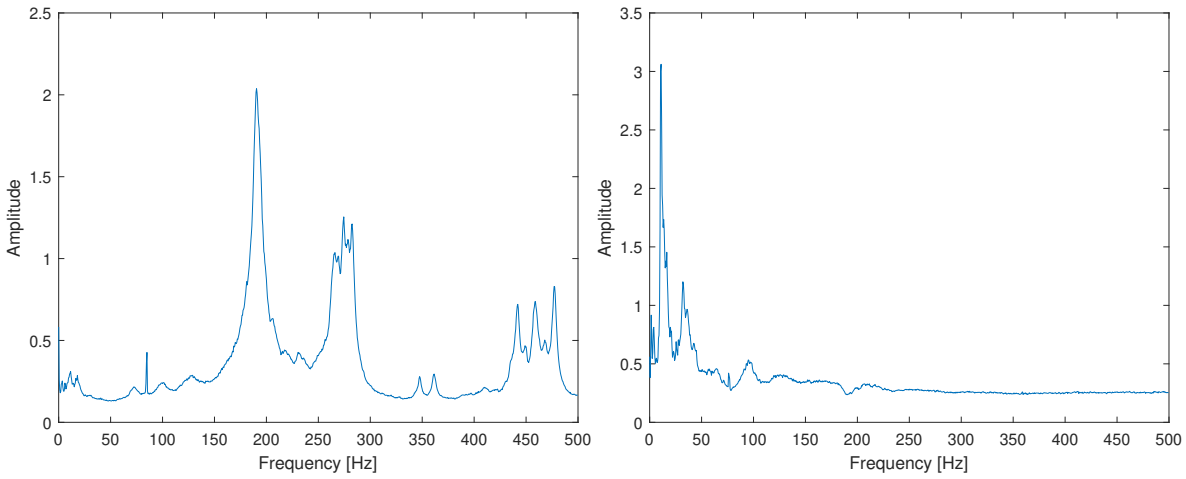
The samples contain not only energy from the random process of the traffic but also from the ambient vibrations. In figure 5.2, there is a constant low amplitude in the acceleration around 0.15 m/s^2 in the time series. These two input forces are two different stochastic processes that can have different characteristics in their frequency content. For instance, the ambient vibrations are generally approximated as white noise. The amplitude of these two processes is also completely different. There is hardly any variation in the ambient vibrations. However, as seen in figure 5.3a, there is considerable variance in the amount of energy in the random process of the traffic. Lower energy samples should have a higher relative contribution from the ambient vibrations as there is variation in the amplitude of the random process of the traffic but not so much in the ambient vibrations.

Another aspect to consider is potential non-physical ”modes” or ”natural frequencies” due to the input force. The stochastic processes can be potentially narrow-banded at specific frequencies. There could be a peak in the frequency spectrum that is not connected to the system’s dynamics but rather due to the similarity in the input force between samples. As only the response is measured, it can be difficult to discern whether a given peak is an operational mode or physical mode. In this research, an anomalous peak is present in most signals that is not part of the system dynamics. Details are given in appendix A, and these peaks will be discussed further in upcoming chapters.

Figure 5.6 shows the average frequency spectrum of the samples that fall within a temperature range of 12 ° C to 13 ° C over a period of two days. There are a total of 2270 samples for each sensor. There is little noise in the mean response, revealing a dynamical system with clear natural frequencies. Even with simple averaging, it is possible to get a decent picture of the dynamic system. This could indicate that the response data follows some ”simple” statistical distribution. Figure 5.7a shows the histogram of the amplitude over several frequency indexes. It shows that the amplitude of this sample set at these frequency indexes seems to adhere to a skewed normal distribution. Each frequency index has a similar distribution which changes depending on the location of the frequency index. At the location of the natural frequencies, for example, at 190 Hz, there is a higher degree of variance compared to the other frequency indexes. It seems that the natural frequencies are amplifying the excitation at these locations, which leads to a higher degree of variance of the response at the location of the natural frequencies. This can be better seen in figure 5.8, which shows the variance of the sample set for each frequency index for both sensors. Again, there is a higher degree of variance at the natural frequencies.

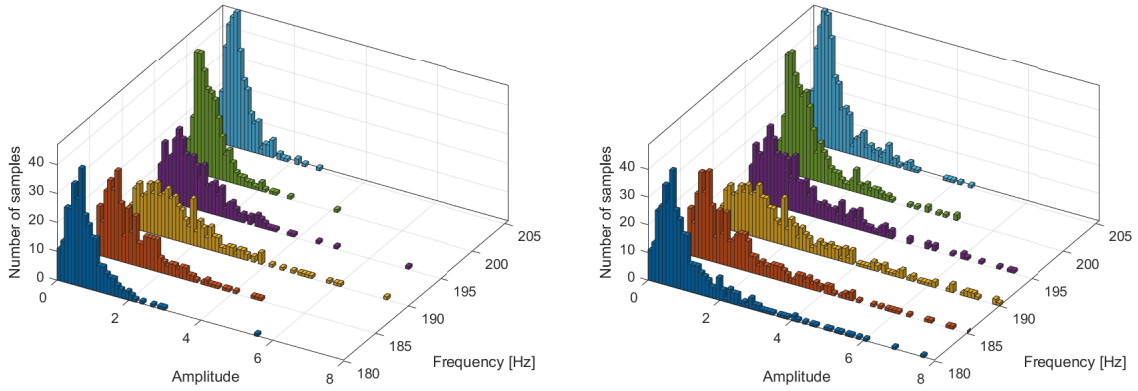
The variance itself can serve as an indication of the location of natural frequencies.

Another detail to consider is that the truncation of the higher energy samples could be simplifying the distribution by removing the samples at the tail end of the distribution. This is shown in figure 5.7 with the comparison between the two figures. It appears that the frequency spectrum is composed of normal variables with this data. It could explain why simply taking the mean provides a decent picture of the dynamic system.



(a) **(b)**

Figure 5.6: Average frequency spectrum of the samples within temperature range of 12 °C to 13 °C. (a) Average of samples from horizontal sensor 31. (b) Average of samples from vertical sensor 17.



(a) Truncated data set **(b)** UnTruncated data set

Figure 5.7: Distribution of the amplitude of specfic frequency indexes with a 5 Hz interval. There is a natural frequency around 190 Hz (a) Truncated data set. The tail end has been removed from the distributions. (b) Untruncated data set. There is a large tail end on all distributions.

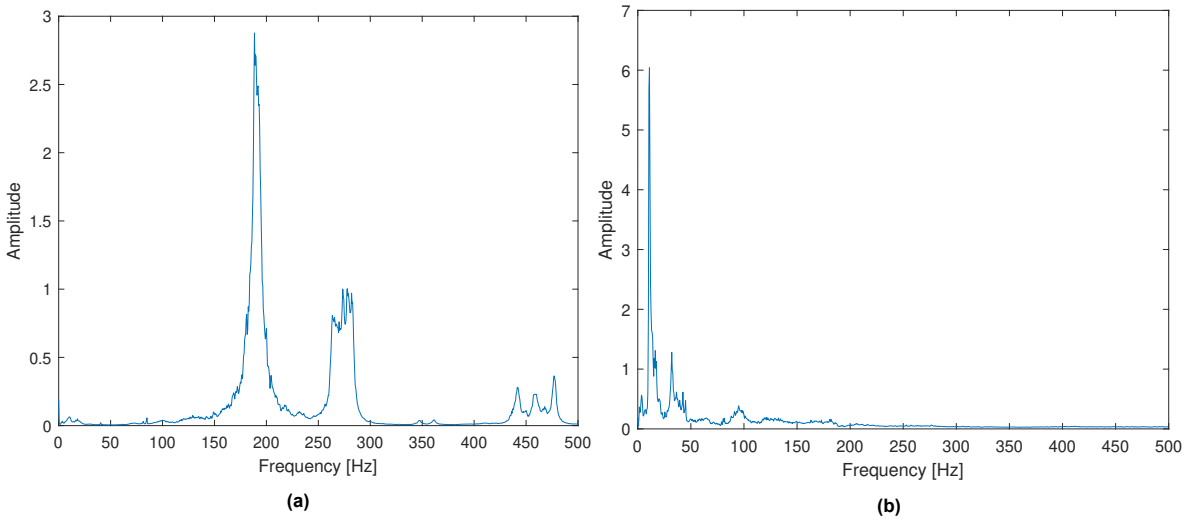


Figure 5.8: Variance at each frequency index of data within temperature range of 12 °C to 13 °C. (a) Variance at each frequency from horizontal sensor 31. Variance at each frequency from vertical sensor 17.

5.2. Application of Robust Principal Component Analysis

Now that a basic understanding of the data has been established, the next step is to explore how robust principal component analysis (rPCA), which was detailed in chapter 4.3, performs on the vibration data.

In chapter 4 on the rPCA, there was an example illustrating the power of rPCA by removing objects in the foreground in video surveillance. In this example, the data consists of many frames from a video. Most of the frames consist of a background, but some frames have cars moving in the foreground. These cars are outliers to the statistical similarity of the background between frames. Thus the rPCA can ”remove” them and reconstruct the background without any cars with the low-rank matrix. Here the problem with the vibration data is different, but the principles are the same. Every sample of the vibration data has some ”corruption” or noise within them. However, the samples from a single sensor all share the same underlying dynamic system, given that operational and environmental conditions are the same. They share the statistical knowledge of the dynamic system even with the variance present in the traffic. Therefore, the rPCA can reconstruct a low-rank matrix that corresponds to the main correlation within the data, which is robust towards outliers. The low-rank matrix would then contain the dynamic properties of the structure, while the sparse matrix contains the noise and outliers. With the operational variability reduced, it is then possible to proceed with damage detection and compare damaged and undamaged samples to see if there are statistical differences between the set of samples.

The rPCA needs to be applied to data in the correct domain. The vibration data needs to be ordered so that statistical similarities fall into the same indexes between sample vectors. In the frequency domain, a natural frequency at 20 Hz will be at the same index for all samples, given that the sampling rate and frequency resolution between samples remains the same. The same cannot be said for time-domain vibration data. Thus the rPCA needs to be on a Fourier basis in order to yield adequate results. Representing the vibration data on a Fourier basis also eliminates any difficulties with data alignment. Consider the example in chapter 4.3 of application rPCA on the video surveillance data. In this example, a video is taken of traffic on the road. During the video, the camera is not perfectly still. It translates and rotates, resulting in the background being at different locations for the pixels of each frame. The bottom frame in figure 4.4 has moved to the right compared to the other two frames. This misalignment of the data results in the background becoming ”blurry”. This is because the SVD, which the rPCA and PCA use, depends on the coordinate system in which the data is represented. The SVD cannot capture translations or rotations in the data. However, representing the data on a Fourier results in a ”stable” coordinate system that cannot translate or rotate.

Another criterion to fulfil for effective usage of rPCA is that the dynamic system should be the same

between the samples. This means that samples should be grouped according to similar operational and environmental conditions before rPCA is applied. In this research, that means grouping samples together according to their temperature.

With these criteria in place, a general procedure can be established when using rPCA to reduce the operational variance of the traffic and perform subsequent damage detection. The data assembly procedure continues from that of the previous chapter, with samples created around the excitation in the time domain and 20 % of the highest energy samples removed from the sample set. The rPCA is then applied to the frequency representation of the data with the same operational and environmental conditions. This process is then repeated for another data set to allow for comparison between two different structural states and perform damage detection.

To demonstrate the power of the method, rPCA is applied to a data set of samples from sensor 31 from the second measurement campaign. The data set is a subset of the data set from table 5.1, and it only contains samples within the temperature range of 10 ° C - 11 ° C. The time window is two seconds for each sample, which gives a frequency resolution of 0.5 Hz. With this sampling and the truncation previously described, this data set contains a total of 2532 samples. Grouping samples within a single centigrade should ensure that the dynamic properties are roughly similar between the samples. Thus the rPCA should find the dynamic system as the samples share that information.

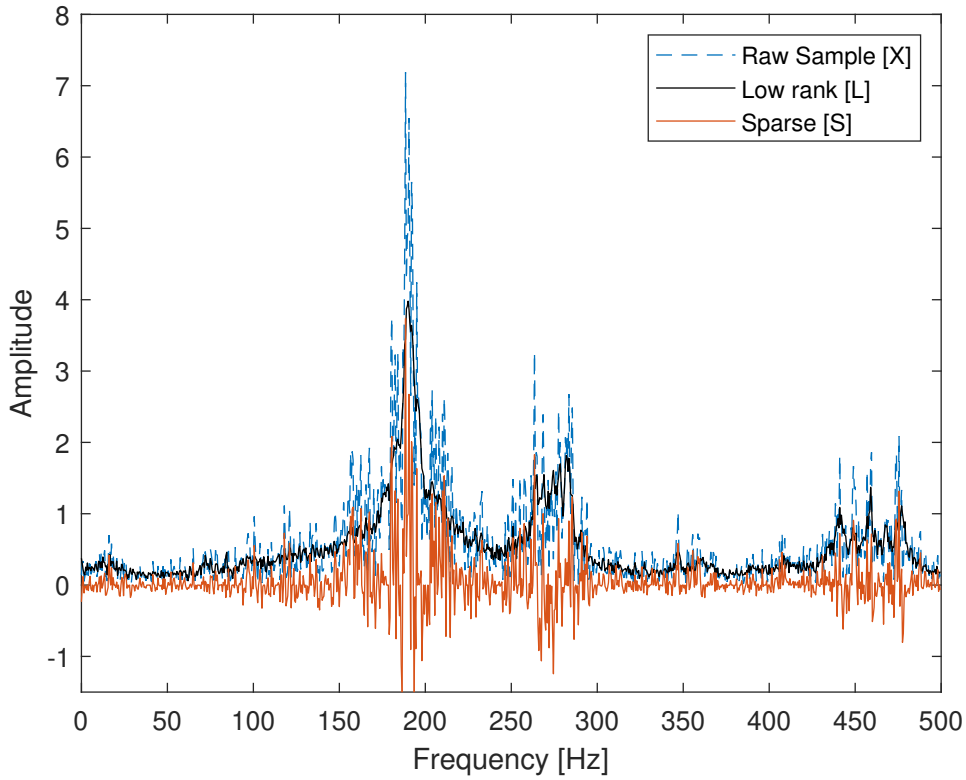


Figure 5.9: Decomposition of a single sample vector into its low-rank and sparse vectors.

Figure 5.9 shows a single sample and its sparse and reconstructed low-rank vectors. The low-rank vector has a more distinct shape similar to that of a transfer function with clearly defined natural frequencies, more so than the ”raw” discrete Fourier transform. While it was possible to make a rough estimate of the natural frequencies with only the discrete Fourier transform, such as in Figure 5.4, it is much more clear where the natural frequencies lie with the low-rank vector. Essentially, the noise within this sample has been reduced. The low-rank vector now contains mostly information regarding the coherent structure of the dynamical system due to the statistical similarity between samples. The low-rank vector contains information about the entire data set, all of the 2532 samples used in this rPCA. The low-rank vector is reconstructed from the principal components of the data set, which contain information about the fundamental aspect of the data set. This fundamental aspect is the similarity

between samples. The measured response signal is a product of the random process of the traffic load and the deterministic physical dynamic system. All of the samples then have information about the physical system, but each of the samples only has a piece of the puzzle. With enough samples, it is possible to get a picture of the dynamic system and reconstruct it with rPCA. On the other hand, the sparse component contains the deviation in the sample to this shared coherent structure of the samples. The amplitude of the sparse part appears to be highest around the natural frequencies.

In this analysis, the default parameters recommended by the authors of the rPCA [11] were used. The most notable among them is the regularisation parameter. This is a penalty term in the optimisation problem of 4.17 and is a key to promoting a single unique solution for ill-conditioned problems. The recommended value is $\lambda = 1/\sqrt{max(n,m)}$, where n and m are the dimensions of the original data matrix X. The solution of equation 4.17 should converge with textithigh probability to equation 4.16 for this recommended regularisation value. This recommended value will be primarily used throughout this study unless stated otherwise.

However, changing this regularisation parameter can yield interesting results, as discussed in chapter 4.3 on the rPCA. In the research on the removal of corruption from the fluid flow with rPCA [38], changing the regularisation parameter had a major effect on the results. The regularisation parameter λ can also be changed for the vibration data. In the vibration data in this thesis, there should only be a single coherent structure in the data. Only one frequency vector corresponds to the system, given that operational and environmental conditions remain the same. Looking at figure 5.9, there is still some ”noise” present in the low-rank frequency vector. There are some low amplitude fluctuations in the low-rank frequency vector. By lowering the regularisation parameter, this noise can be potentially removed.

Figure 5.10 shows the comparison between two rPCAs with different regularisation parameters, one with a default value of $\lambda = 1/\sqrt{max(n,m)}$ and another rPCA with a regularisation value of $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. The maximum dimension is multiplied by two within the square root. This value is arbitrarily chosen. In this figure, the same data set was used as in the previous example; a total of 2532 samples are in this data set. The dimension of the data matrix is X, then 2532x1000, as the length of the frequency vector is 1000. The resulting regularisation values are then $\lambda = 0.0199$ and $\lambda = 0.0141$ for the default value regularisation and the ”lower” value regularisation, respectively. What can be seen in the figure is that the low-rank frequency vector with the lower regularisation value has an even more ”smoother” profile. This is because it has a lower rank; it consists of fewer principal components. The difference in rank between the low-rank matrices is 594 and 111. The more ”noisy” low-rank term is being described by more principal components, some of which do not describe the dynamic system properties but rather some noise within the data set. Essentially, the more ”smooth” low-rank frequency vector has a higher concentration of information regarding the dynamic system with its 111 principal components. Keep in mind that the figure only shows a single sample, but the low-rank frequency vector in the figure has information from the entire data set. A final remark on the lower regularisation parameter is that the rPCA took significantly longer to converge to a solution than with the default regularisation parameter.

More aggressive filtering is possible with the rPCA by lowering the regularisation parameter even more and potentially removing even more noise. However, caution should be taken in not removing information about the system. There should be a single coherent structure or a single principal component in vibration data that characterises the system. Most of the variance within the data set can be explained by the first principal component. However, it is not known if higher-order principal components beyond the first one contain information on the dynamic system or just noise. These can potentially be filtered out with a lower regularisation parameter. To stay on the safe side, the regularisation parameter used in the analysis in this thesis will not go lower than the previous example. However, just like in the research of the rPCA on corrupted fluid flows, there needs to be an understanding of the selection of the regularisation parameter for vibration data. This thesis will primarily use the default regularisation value, which proved to be adequate in most problems. In some cases, a lower regularisation parameter will be used.

In the previous example, only a single sample was observed and examined, but here it is the perfor-

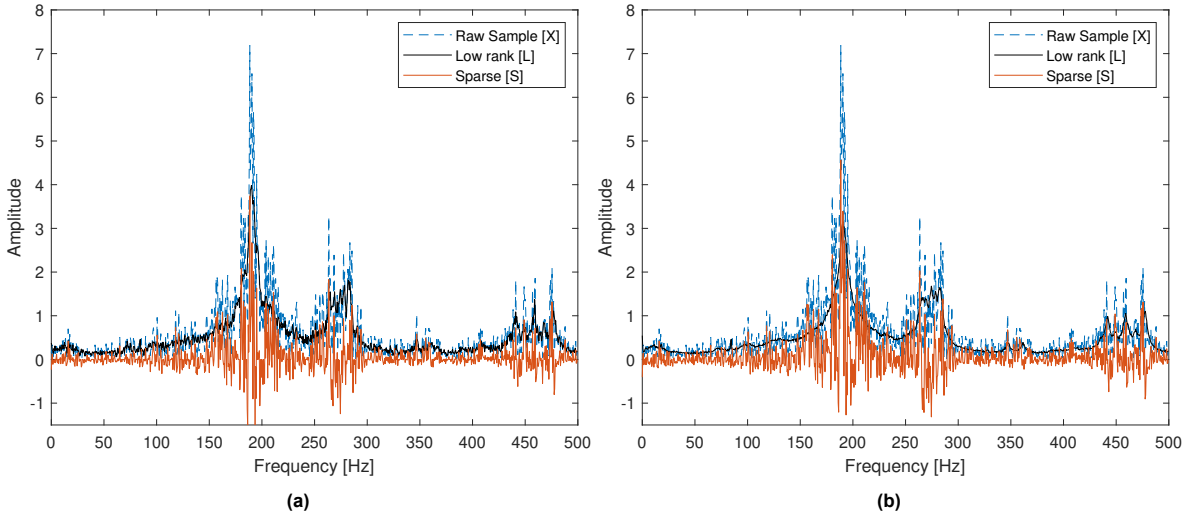
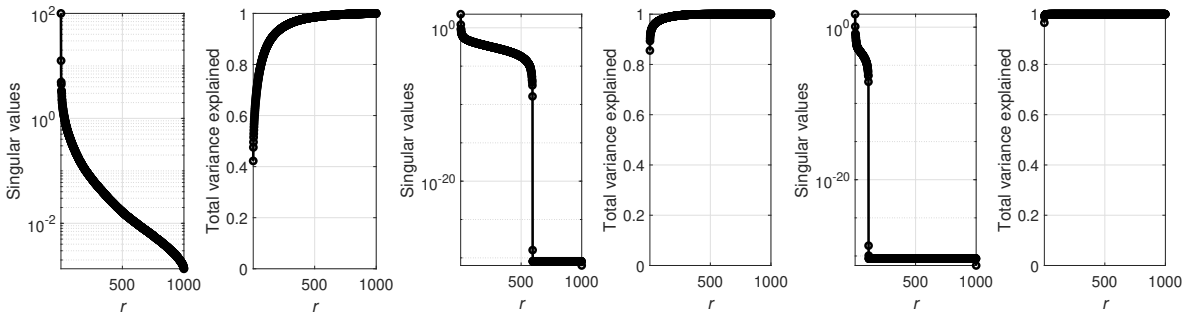


Figure 5.10: Effect of the regularisation parameter on the low-rank and sparse components. The Low-rank with a lower regulation value is smoother as there are fewer principal components (a) Default value of the regularisation parameter $\lambda = 0.0199$. Rank of low rank: 594 (b) "Double" value of the regularisation parameter $\lambda = 0.0199$. Rank of low-rank : 111

mance of the whole data set that is of interest. The goal of rPCA is to reduce the variance within a data set. The variance explained by the principal components can be used to quantify the variance in a data set. If most of the variance within the data set is explained by a few principal components, then a few coherent structures explain the majority of the data. To measure the effectiveness of the rPCA, a comparison can be made on the total variance explained by the principal components of the raw and low-rank data sets. Figure 5.11 shows the singular values or, more precisely, the eigenvalues of the covariance matrix of the raw truncated data X and the low-rank matrix L for two different regularisation parameters as previously defined. More importantly, it shows the total variance explained by the principal components for both the raw truncated data and low-rank data. Comparing the low-rank data and the raw data reveals a massive difference in the total variance explained of the principal components, particularly in the variance explained of the first principal component of both data sets. The total variance explained by the first principal component is 42% for the raw truncated data set, while it is 85 % for the low-rank data set with the default regularisation parameter. The variance explained by the first principal component doubled with the application of rPCA. This means that most of the underlying data within the low-rank data set can be described by this principal component, a single vector. This vector represents the similarity between samples and should correspond to the dynamic system. Higher-order principal components explain the remaining variance within the data, and those PCs can correspond to noise within the data or other minor coherent structures in the data. The second PC of the raw truncated and low-rank data sets explain much less than the first PC or around 5.3 % for the raw truncated and 3.8 % for the low-rank data. While they explain a low amount of variance in the data compared to the first PC, they could contain information about the dynamic system. When using a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n, m)}$, then the variance explained by the first PC becomes even higher or around 95 %. This single principal component can explain nearly the entirety of the data as the rPCA has discarded higher-order PCs with this lower regularisation parameter.

Another observation is that the singular values suddenly drop around vector numbers 594 and 111 for the low-rank data set with the default and lower regularisation parameters, respectively. This corresponds to the rank of these low-rank matrices or the number of linearly independent vectors. The rank of the raw truncated data is equal to the vector's length, which is 1000. This means that the vectors of the raw truncated data span the entire vector space. With the rPCA, the data is reconstructed with PCs that explain most of the data within the data set. The PCs are orthogonal to each other but are directions of maximum variance in the data. The optimisation problem of the rPCA is to minimise the rank of the low-rank L. This results in the aforementioned rank of the low-rank matrices. Keep in mind that the number of samples in the low-rank matrices is the same as in the raw truncated data set. In this example, it is 2532 samples. The difference is that only 594 or 111 linearly independent vectors

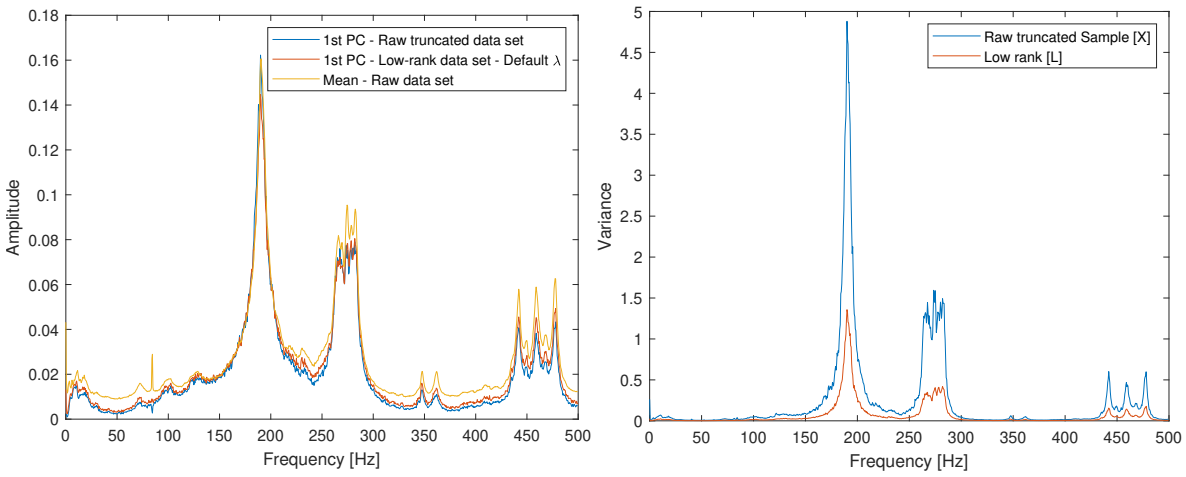


(a) Raw Data set **(b)** Low rank data set. Default λ **(c)** Low rank data set. Lower λ

Figure 5.11: Singular values and total variance explained of each of the principal components. The total variance explained of the first principal component is first entry in the graph to right in each respective data set.

are needed to describe these 2532 vectors instead of 1000, as is the case with the raw truncated data.

Figure 5.12a shows the first principal component for both raw truncated and low-rank data sets as well as the mean value of the raw truncated data set. The mean has been normalised to allow for better comparison with the PCs by setting the value of the natural frequency at 190 Hz equal to the 1st PC of the raw data. Interestingly, the principal components and the mean are all fairly similar. The similarity between the PCs and mean could be connected to the fact that the raw truncated data follows a normal distribution. Another question to consider is why there is such a difference in the total variance that the PCs explain in their corresponding data set. A potential answer is that redundant information related to traffic noise was filtered out in the low-rank data set, thus increasing the relative variance of the first principal component in the low-rank data set. The information regarding the dynamic system is present in all data sets, but its relative contribution is much more in the low-rank data set.



(a) **(b)**

Figure 5.12

The variance within the data can be visualised with indicators other than the variance explained by the principal components. For example, computing the variance of the samples for each frequency index yields figure 5.12b. Here, the variance at each frequency is much less for the low-rank data set with the default regularisation parameter than the original ”raw” data set.

6

1st Measurement campaign - Damage detection

It has now been established that the rPCA can reduce the variance due to the operational variability of the traffic. The next step is now to perform damage detection. This requires comparing two different data sets, each representing a different structural state. The rPCA would need to be applied to each data set independently before being compared. For data from the first measurement campaign, that would mean comparing the measured response from sensors in the undamaged and damaged areas. The damaged area has two fatigue cracks between the stiffeners and a transverse beam, while the undamaged area is a reference area that is structurally similar. The rPCA can reduce the operational variability within the data sets to allow for comparison to measure if there is any significant difference between the two that could be attributed to the difference in the damage states between the two areas.

There are 16 sensors in each of the areas. For both areas, the sensors are located on the west side of the bridge on the cantilever bridge deck that carries the traffic southbound. The two areas are nearly identical in their structural configuration. This makes it feasible to compare the output of sensor pairs, which occupy the same spot in the ”healthy” reference area and damaged area. In total, there are then 16 sensor pairs which translates to 16 possible damage comparisons. The underlying dynamic system of the sensor pair should be the same, except for the damage due to the fatigue cracks. However, this is not entirely true, as there are other differences in the structural configuration, such as the presence of a flange discontinuity and a coupling plate in the ”healthy” reference area. A more thorough discussion on this topic will be at the end of the chapter.

A single day of data from the first measurement campaign is used for the damage detection analysis. The data set will be further reduced in order to minimise the effect of environmental variability. **This is achieved by only using a single 15-minute time series recording for the analysis**. It is not expected that there will be significant temperature shifts in 15 minutes. However, this does not produce a significant amount of samples, but it is necessary to maintain the low variability in the temperature so as not to classify temperature change as structural damage

After applying the pre-processing in the previous chapter, the rPCA can be applied. Damage detection is a comparison between two structural states, so two different rPCA are conducted, one on the data from the damaged area and another on the data from the healthy area. This gives two low-rank matrices for each area. The default regularisation parameter of $\lambda = 1/\sqrt{max(n,m)}$ is used for all analyses in the damage detection analysis. Afterwards, the (regular) PCA is applied to the combined data set containing both of these Low-rank matrices. The PCA will find the largest correlation or Principal Components (PCs) within the new combined data set, and the data can be represented in the subspace of these PCs. The difference between the two data sets from the damaged and ”Healthy” areas can correspond to the PCs, and data from the two areas can form distinct clusters in the subspace of these

PCs, which allows distinguishing between the different data from both areas.

Applying the PCA to the combined data reduces the dimensionality of the problem. The ”default” length of the frequency vector is 1000. This is due to the 1000 Hz sampling rate and time window of 2 seconds for each sample, which gives a 0.5 Hz frequency resolution and a maximum frequency of 500 Hz. This, in turn, yields a 1000-length frequency vector. It would be difficult for machine learning algorithms to work with data with such a high dimension. As the dimension increases, the required amount of data to generalise the problem accurately grows exponentially. Data will be sparse in this high-dimensional space without an unrealistic amount of data. However, to overcome this problem, the data can be embedded into a low-rank subspace, as is done with the PCA. Machine learning algorithms can then be applied with data embedded into the subspace of the PCs.

Another question is, what portion of the frequency spectrum should be used? Damage will manifest itself as changes in the natural frequencies of the system. Should only portions of the frequency spectrum where there is resonance activity be taken and used? The frequency content of the sensors can vary considerably. Some sensors have resonance activity over a broad frequency band, as seen in the previous chapter. There are also sensors that mainly have low-frequency content. These are mainly sensors capturing the vertical motion of the bridge. The approach taken in this research is to take the entire frequency vector when there is high-frequency content over a wide band. When the majority of the frequency content is situated in a localised frequency band, as is the case with the low-frequency content sensors, then only that narrow band is considered for analysis.

This brings up another issue regarding the low-frequency content. There needs to be an adequate frequency resolution to be able to spot changes in the natural frequencies due to damage. This is always an issue regardless of the value of the frequency, but assuming damage will lead to a percentage change in the natural frequencies, the absolute value of shifts in lower-value frequencies will be less than those of higher frequencies. Thus higher frequency resolution might be required to spot changes in the lower frequencies. When working with sensors that primarily consist of low-frequency content, a frequency resolution of 0.125 Hz is used, which translates to a time window of 8 seconds for each sample.

These points and the discussion in the previous chapter give the foundation of the damage detection framework, and is visualised in figure 6.1 and can be summarised as follows:

1. Create samples by taking a time window around peaks in the time series of a single sensor
2. Discard 20 % of the highest energy samaples for each sensor
3. Apply discrete Fourier transform to the samples
4. Take the absolute of the fourier transform
5. Obtain a low-rank matrix by performing rPCA on a sample set within a particular temperature interval. Default regularisation parameter of $\lambda = 1/\sqrt{max(n, m)}$ is used for the analysis.
6. Compute (regular) PCA on the combined low-rank with data from both damaged and ”Healthy” areas
7. Conduct feature discrimination with the score of the samples in the PCs as features.

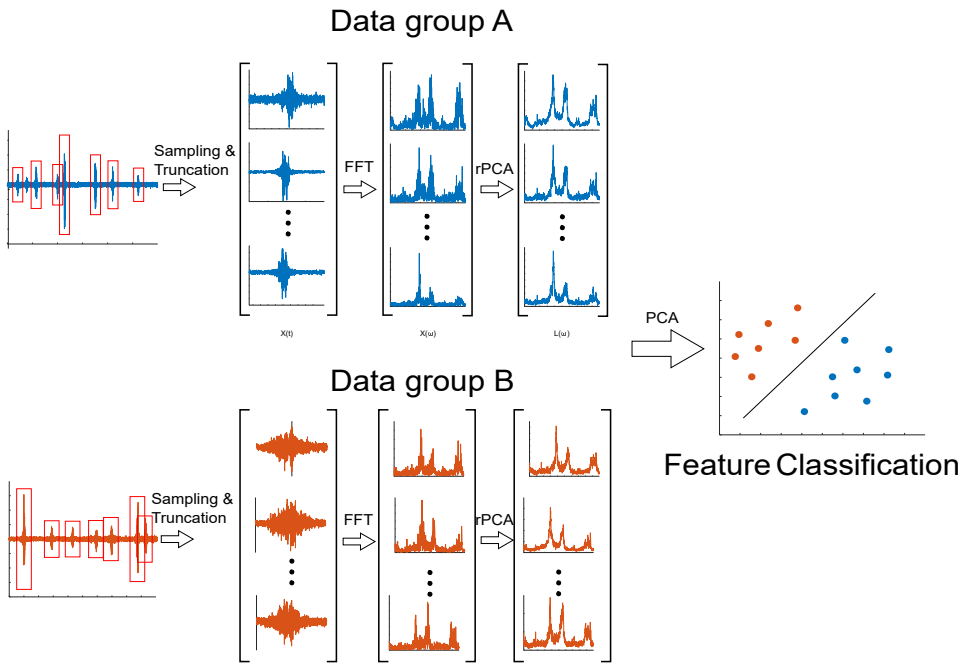


Figure 6.1: General preprocessing procedure. Samples are created around excitation in the time domain and the Fourier transformed applied. Afterwards samples are grouped according to temperature and the rPCA is applied

6.1. Sensor Comparison

6.1.1. Sensors 7 and 23

The first sensor pair under analysis are sensors 7 and 23. Sensor 7 is located in the healthy reference area, while sensor 23 is in the damaged area. These accelerometers measure the horizontal acceleration or the out-of-plane movement of one of the stiffeners. The damaged area has a crack between this stiffener and the transverse beam. This damage should result in a different dynamical system compared to that of the reference area. The data for these sensors consists of **413 samples** that are generated with the sampling and truncation scheme defined in the previous chapter. Each sample is a 1000-length frequency vector with a frequency resolution of 0.5 Hz and a range of up to 500 Hz. Like all the samples used for the damage detection analysis, they are generated during a 15-minute period during the evening rush hour. This ensures that environmental variability is kept to a minimum. **190 of the samples are from sensor 7, while the remaining 223 are from sensor 23**. Interestingly the number of samples generated by the sample algorithm is not the same for the sensors. The sensors are both located at the west cantilever of the bridge that carries southbound traffic but at different sections of the bridge. This means the same vehicles drive over both sensors. However, this does not mean the vehicles will generate the same excitation. For example, the vehicles can be at different speeds, and the clearance between vehicles can differ. The excitation due to the traffic is a random process that is dependent on many various unknowns.

Nevertheless, there is a consistent theme for nearly all sensor pairs regarding the number of samples in healthy and undamaged areas. Nearly all sensors in the damaged area have a higher number of samples than the corresponding pair in the undamaged area. A possible explanation is that the damaged area has reduced stiffness, and therefore the measured response is greater given the ”same” excitation. This would mean more samples are created with the sample generation algorithm as more vehicles generate an excitation above the minimum threshold. The minimum threshold is based on the amplitude of the ambient vibrations. Without the minimum threshold, the samples could consist of just ambient vibrations. This brings in another topic that not all vehicles generate a signal above the amplitude of the ambient vibrations. According to INWEA, the average daily number of vehicles that

drive over the Haringvliet bridge is 55000 vehicles in 2020 [1]. However, the number of samples from a single day is 2273 from sensor 23. When normalised with the measurement period of 15 minutes, the number of ”detected” vehicles is 9092 over the day. Even if the detected vehicles would only come from a single lane, which is probably not the case, then there is a deficit of around 4658 vehicles compared to the average daily vehicles from INWEA. This is an extremely crude estimation as there is only data from a single day, but it hints that not all vehicles cause an excitation higher than the amplitude of the ambient vibrations.

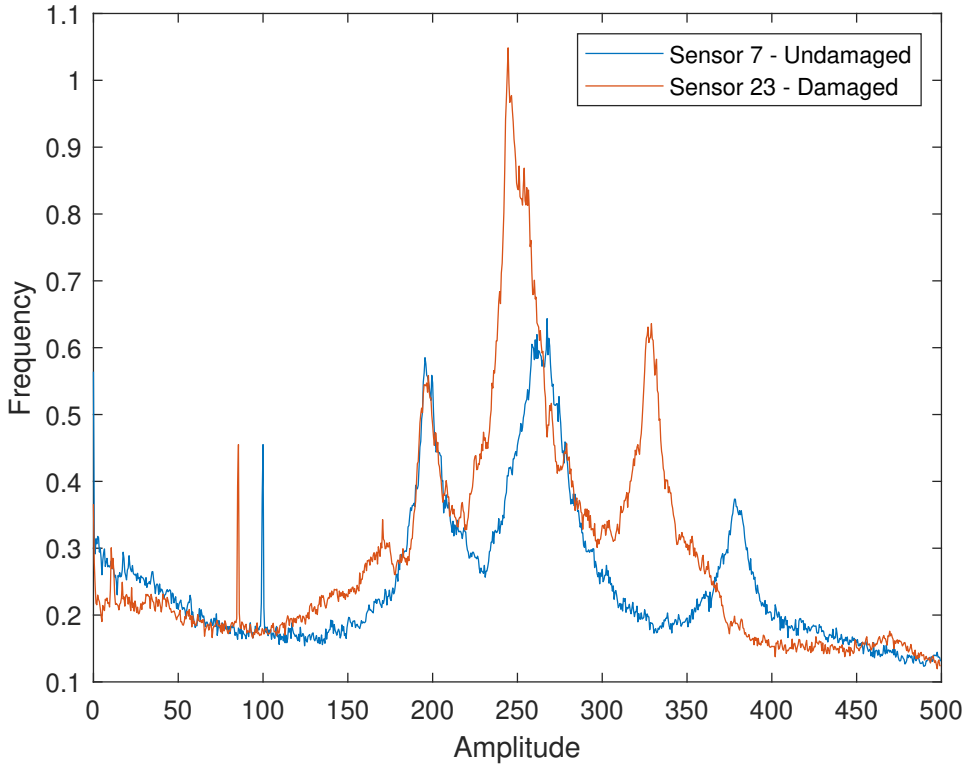


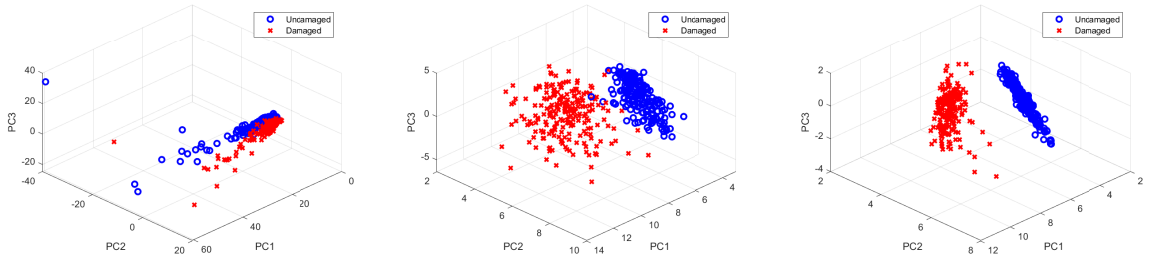
Figure 6.2: Mean frequency vectors of sensors 7 in the undamaged area and sensor 23 in the damaged area.

Before performing the damage detection, the mean vectors of the raw truncated samples from the two sensors can be compared. They are on display in figure 6.2. Even by looking at mean vectors, it is clear that these two sensors have two different responses. It seems unlikely that changes in the stochastic process of the traffic would be responsible for this magnitude of a difference in the mean response. This difference is likely due to the different underlying dynamic systems that the sensors capture. That would mean that by simply looking at the mean response of the raw truncated samples, the underlying dynamics can be seen. For both sensors, there appear to be three primary natural frequencies. Sensor 7 in the ”healthy” area has resonance frequencies at 190 Hz, 260 Hz and 380 Hz, while sensor 23 in the damaged area has natural frequencies at 190 Hz, 240 Hz and 330 Hz.

The differences in the natural frequencies between the sensors aligns with the idea that damage, such as the fatigue cracks, will result in a reduction of stiffness. This, in turn, will lower the natural frequencies due to the reduction in stiffness. Two of the resonance frequencies of sensor 23 at 240 Hz and 330 Hz are lower than the ”corresponding” frequencies of sensor 7 at 260 Hz and 380 Hz. It appears that the natural frequency at 190 Hz is unaffected by the differences in stiffness characteristics of the two areas, as this frequency is the same for the two sensors. Another indication that the damaged area has reduced stiffness is the higher amplitude of the mean response of sensor 23. A less stiff structure would respond with a higher amplitude to the same level of force. It is possible that this difference is due to the random process of the traffic. However, the exact same vehicles drive over the two areas and generate the samples. It seems unlikely that the random process of the traffic is responsible for

[1] Intensiteiten op Wegvakken - INWEVA Source

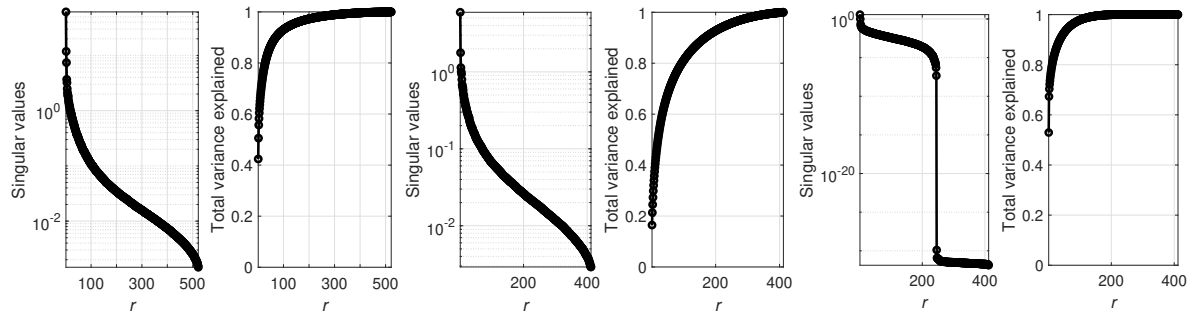
this change. This difference in amplitude is a consistent theme in the comparison between sensors in the ”healthy” and damaged areas. Another possible reason for this difference in amplitude could be due to the different damping characteristics of the two zones. Increased damping would lead to less amplification. The amplitude of the response would also lower as a result.



(a) Raw Data set **(b)** Raw truncated data set **(c)** Low rank data set

Figure 6.3: Principal component score of samples from sensors 7 and 23 for the first three principal components.

The next step in the damage detection process is to apply the rPCA independently on the samples from the two sensors. That is, it is applied once on the 190 samples from sensor 7 and again on the 223 samples from sensor 23. This produces two sets of low-rank matrices, each containing the original amount of samples. Each sample has been reconstructed according to coherent structures in their respective sample set. The low-rank samples from sensor 7 should now contain minimum noise and correspond to the underlying dynamic system that the sensor captures. The same applies to low-rank samples from sensor 23 but now with a different dynamical system. Now, these two low-rank sample sets from both sensors are combined, and PCA is applied to the data. This creates a new ”basis” for the data. The samples are then represented in the first three principal components (PCs), which are shown in figure 6.3. The figure shows the three PCAs with different data, raw untruncated samples, raw truncated samples and the low-rank truncated samples. The ideal condition here is that the data will form separate clusters in the low-rank representation of the first three PCs to allow for subsequent damage detection. In all cases, there seem to be two separate clusters forming, and the separation between them becomes more evident with truncation and the application of the rPCA. Even without performing any truncation, there is some degree of separation between the clusters. By applying the truncation scheme defined in the previous chapter, there is nearly a clear separation between the two clusters. The separation becomes even better with the low-rank data from the rPCA. What is happening here is that the data from sensors 7 and 23 have different ”scores” or contributions from the first three PCs. The PCs are the direction of maximum variance within the data and are ordered in the amount of variance they explain. This implies there is some fundamental difference between the two data sets that form these clusters.



(a) Raw data set **(b)** Raw truncated data set **(c)** Low rank data set

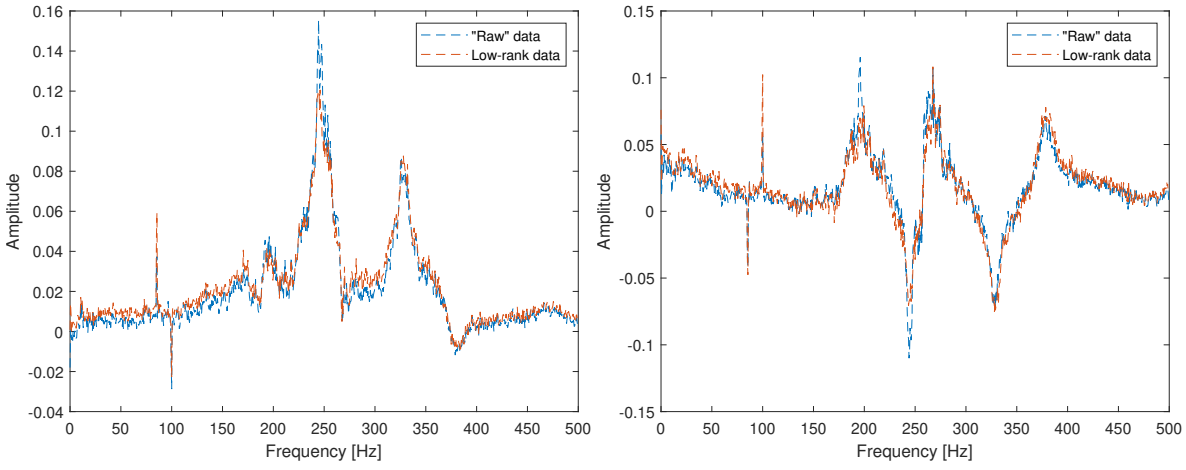
Figure 6.4: Singular values and total variance explained of each of the principal components. Data from sensors 7 and 23.

Another way to see the effectiveness of rPCA is to look at the principal components and the variance that they explain. Figure 6.4 shows the singular values and the total variance explained by each PC. Comparing the total variance explained between the three data sets reveals quite a difference. The total

variance explained by the first PC is 42 %, 18 % and 53 % for the raw untruncated, raw truncated and truncated low-rank datasets, respectively. There is a considerable increase in the variance explained by the first PC with the application of rPCA, as seen in the difference between the raw truncated and low-rank data. The rPCA has filtered out much of the noise within the data resulting in a higher relative variance explained by the first PC.

It can be seen in figure 6.3 that the clusters are separating the PC 1 and PC 2 plane. It is possible to draw a line in this plane between the two clusters and completely separate them for the low-rank data. PC 1 and PC 2 explain the most variance within the data and contain information regarding the difference between the two sample groups. These PCs are shown in figure 6.5. They can be compared to the mean vector of the sensors in figure 6.2. There is some similarity in the 1st PC with the mean response of sensor 23 with the same natural frequencies at 240 Hz and 330 Hz. However, the similarities end there. The PCs are rather the subtraction of two mean vectors. PC 1 appears to be the mean vector from sensor 23, with the mean from sensor 7 subtracted from it. Likewise, PC 2 seems to be the opposite; the mean vector from sensor 7 has the mean from sensor 23 subtracted from it. This can be seen in the ”valleys” at the natural frequencies for sensor 23, at 240 Hz and 330 Hz. It is the ”score” of the samples in these two PCs that separates the two groups. It also seems that these PCs represent the variance in the data that corresponds to the difference between the ”healthy” and damaged samples. So the data is being separated into distinct clusters based on the variance that corresponds to the difference in the ”healthy” and damaged data sets.

This ”subtraction” between the data of the two sensors becomes apparent when looking at the anomalous ”spikes”. These are the sudden spikes in amplitude in the frequency spectrum around 85 and 100 Hz. They can be seen in the mean frequency vector in 6.2. The spike at 85 Hz is part of the data from sensor 23, and the spike at 100 Hz is part of the data from sensor 7. These spikes are present in the PCs and are always opposite of each other in either PC. This gives an additional indication that these two PCs explain the variance that corresponds to the difference between data in sensors 7 and 23. These ”spikes” are not part of the structural response of the system. The hypothesis is that these ”spikes” are due to the sensor system and auxiliary equipment. A more detailed analysis of these peaks can be found in Appendix A. These ”spikes” pose a challenge for the damage detection process as they add a false-positive bias, as will be revealed later in the chapter.



(a) 1st Principal Component **(b)** 2nd Principal Component

Figure 6.5: Second and first principal components of the raw truncated and low-rank data sets. Data from sensors 7 and 23.

Another observation on the first two principal components is that there is hardly any difference in the principal components between the raw truncated and low-rank data sets, even though for the low-rank data, these principal components explain much more of the total variance within their respective data sets. Both the raw truncated and low-rank data sets are separated based on the same variation within the data. In the case of the raw truncated data set, this variation is only a fraction of the total variance within the data set. For the low-rank data set, most of the variance can be explained by these two principal components. The two low-rank PCs have a higher relative information content of the data set

as much of the noise has been removed. This could explain why the degree of separation of the two groups in these two PCs is much better for the low-rank data, as seen in figure 6.3c.

To summarise the comparison between data from sensors 7 and 23, there is a clear separation between samples from the ”healthy” and damaged areas in the low-rank subspace of the first three PCs. This separation seems to be based on the variation in the data that corresponds to the difference between the two data sets. The ”score” of the samples in these PCs is an ideal feature to work for automated damage detection.

6.1.2. Sensors 13 and 29

The next step is to conduct comparisons between other sensor pairs to see if this degree of separability between clusters of sensors 7 and 23 is present in other sensor pairs. Sensors 7 and 23 were located on the stiffener that had a fatigue crack between the stiffener and the transverse beam. The influence of the damage on these sensors is likely to be the highest of all sensors in the campaign. Sensors 13 and 29 are positioned on the stiffener on the opposite side of the crack. There is no crack between the transverse beam and this stiffener. **The effect of damage on the dynamical properties at this sensor’s location is likely less than for sensors 7 and 23 as it is further from the damage**. There are a total of 411 samples for this analysis, 202 from sensor 13 in the ”healthy” area and 209 from sensor 29 in the damaged area. The frequency resolution of the samples is the same as before at 0.5 Hz. Figure 6.6 shows the mean frequency vectors of samples from sensors 13 and 29. It is clear that the signals from these sensors are different, but there is considerable similarity between the signals. According to the mean response, sensor 13 seems to have natural frequencies at 220 Hz, 275 Hz, 345 Hz and 375 Hz. Sensor 29 appears to have natural frequencies at 220 Hz, 270 Hz and 355 Hz. It also appears that the mean amplitude of sensor 29 in the damaged area is higher than that of sensor 13. This could be due to the damaged area being less stiff.

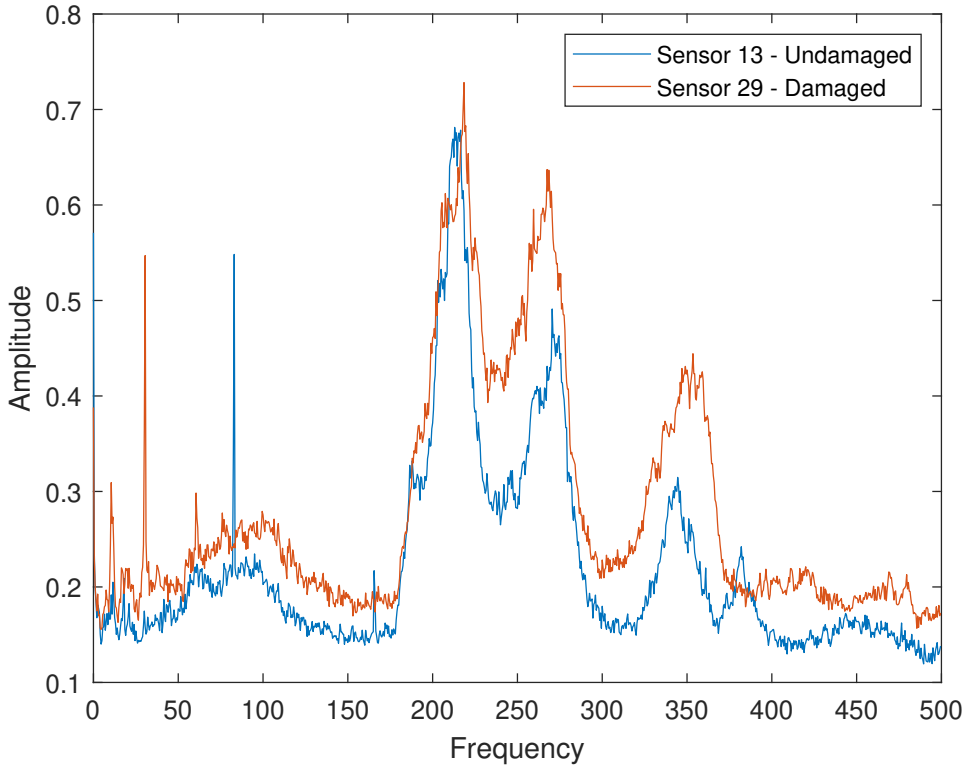
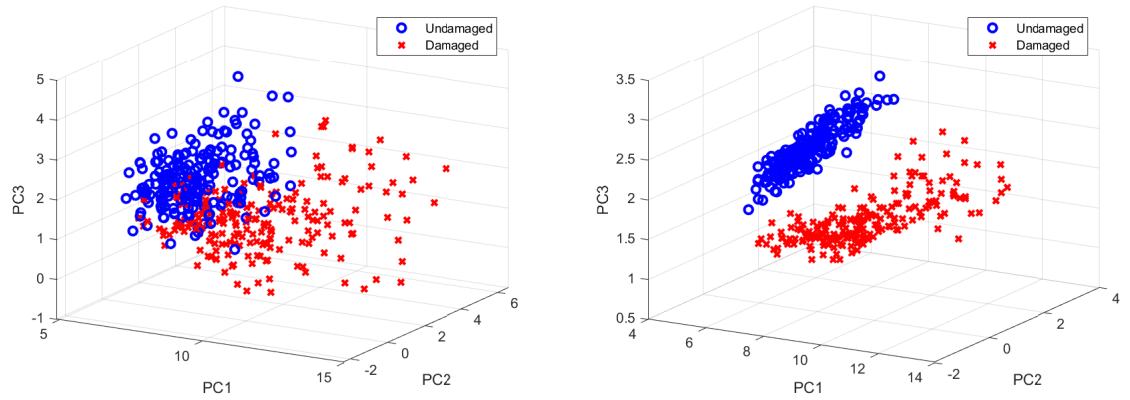


Figure 6.6: Mean frequency vectors of sensors 13 in the undamaged area and sensor 29 in the damaged area.

Figure 6.7 shows data from sensors 13 and 29 represented in the first three PCs. The truncated raw data seems to form two clusters. However, there is an intersection of the two clusters, making it difficult

to classify which group the samples belong to. On the other hand, there is a clear separation between groups with the low-rank data from the rPCA, making it easy to distinguish which group the samples belong to.

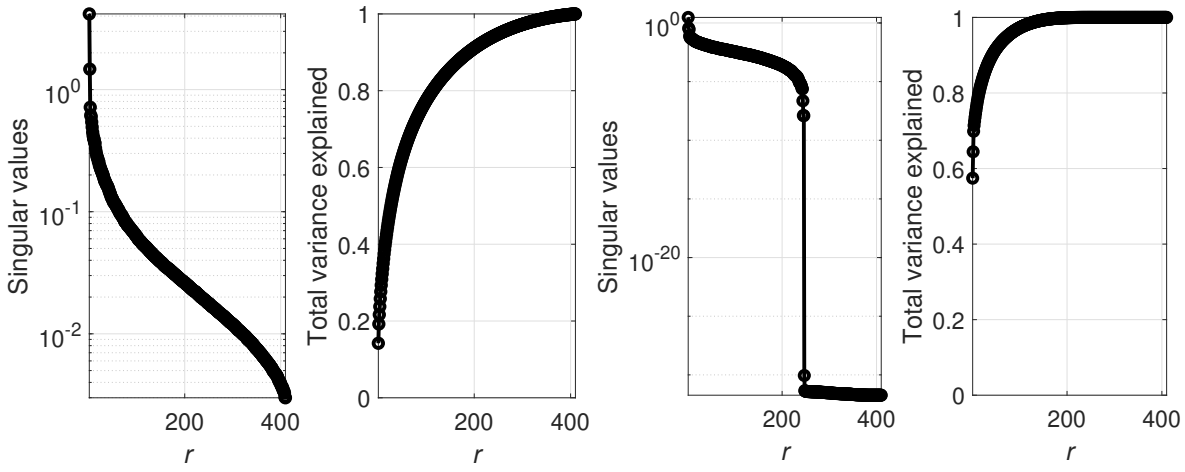


(a) Raw truncated data set

(b) Low rank data set

Figure 6.7: Principal component score of samples from sensors 13 and 29 for the first three principal components.

Similar to before, the first PCs of the low-rank data explain more of the variance within the data set compared to the raw truncated data, as seen in figure 6.8. The first three PCs of the low-rank data explain most of the variance within the data set. After the third principal component, the remaining PCs seem to explain the same amount of variance. The first three PCs appear to be significant in the amount of variance they explain. These PCs explain variance that most likely corresponds to the differences between data sets as there is a separation between the groups in the subspace of these PCs. A plane can be drawn in the space of these three PCs to separate the two groups. The PCs are shown in figure 6.9, and the PCs exhibit the same behaviour as in the previous analysis of sensors 7 and 23. The PCs are the difference between the signals of sensors 13 and 29. There are corresponding peaks and valleys where the natural frequencies are located. This difference between the signals of the two sensors is best seen when looking at the spikes; they are always opposite as they belong to different sensors.



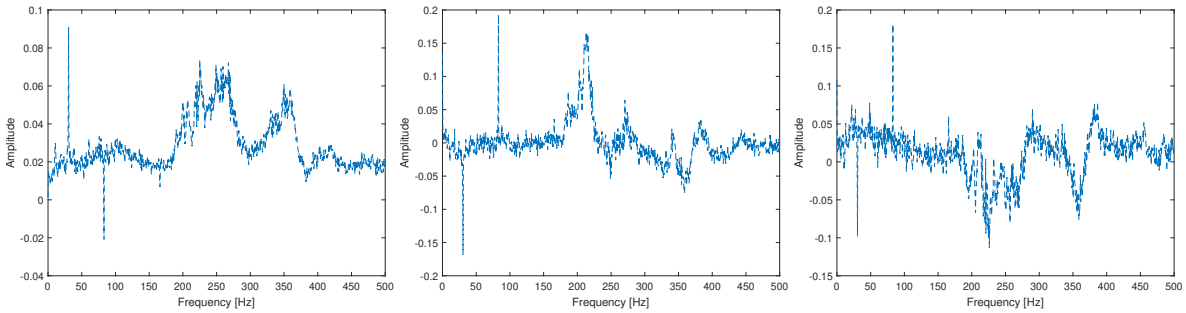
(a) Raw truncated Data set

(b) Low rank data set

Figure 6.8: Singular values and total variance explained of each of the principal components. Data from sensors 13 and 29.

In the case of this sensor pair, the raw truncated data set performed worse than for sensor pairs 7 and 23, as the two sample groups are not completely separated in the subspace of the first three PCs. This could be due to the difference between the samples of sensors 13 and 29 being less than in sensors 7 and 23. However, there was a separation between the sample groups for the low-rank data for both

sensor pairs.



(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 6.9: First three Principal Component. Low-rank data set

6.1.3. Sensors 3 and 19

Entire frequency vector - frequency resolution 0.5 Hz

Up to this point, two sensor pairs have been considered. In both cases, the sensors have been measuring the horizontal motion of the stiffeners and are capturing high-frequency content with natural frequencies at 200 Hz and higher. The vertical motion can also be analysed. Consider sensors 3 and 19, which are positioned on the bottom of the deck plate and measure the vertical motion of the deck. The natural frequencies of the signals of these sensors are lower compared to the horizontal sensors. This can be seen in figure 6.10, which shows the mean frequency vector of all observations. The frequency content of these sensors seems to be more ”chaotic”. The natural frequencies are not as well defined as was the case with the other two sensor pairs. However, as was seen in the two previous comparisons, the mean amplitude of the samples from the sensor in the damaged area is higher.

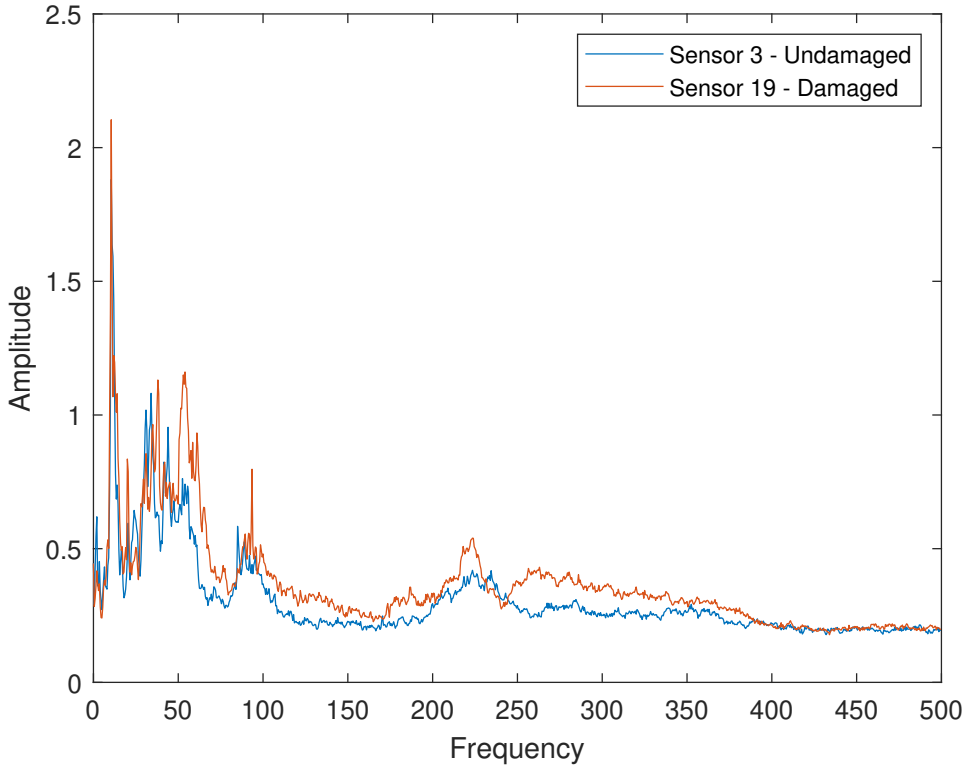
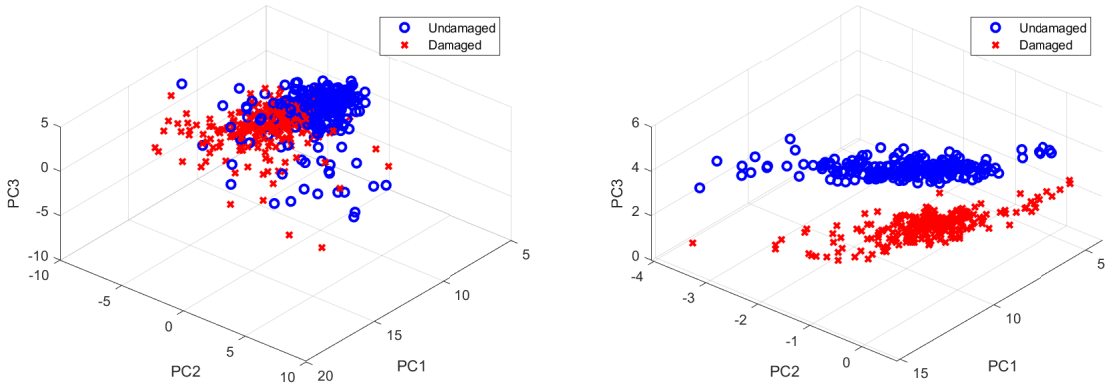


Figure 6.10: Mean frequency vectors of sensors 3 in the undamaged area and sensor 19 in the damaged area.

A frequency resolution of 0.5 Hz is chosen for the analysis. A lower frequency resolution could be required for the low-frequency content, especially for low frequencies such as the natural frequencies at 10 Hz in figure 6.10. However, given the level of difference between the data groups below 100 Hz, as seen in figure 6.10, a high-frequency resolution is not required to detect the differences between the data groups. Increasing the frequency resolution lowers the amount of data available, which can have a negative effect on the performance of the rPCA and PCA. With a frequency resolution of 0.5 Hz, the total number of samples for this analysis is 495, with 244 from sensor 3 and 251 from sensor 19.

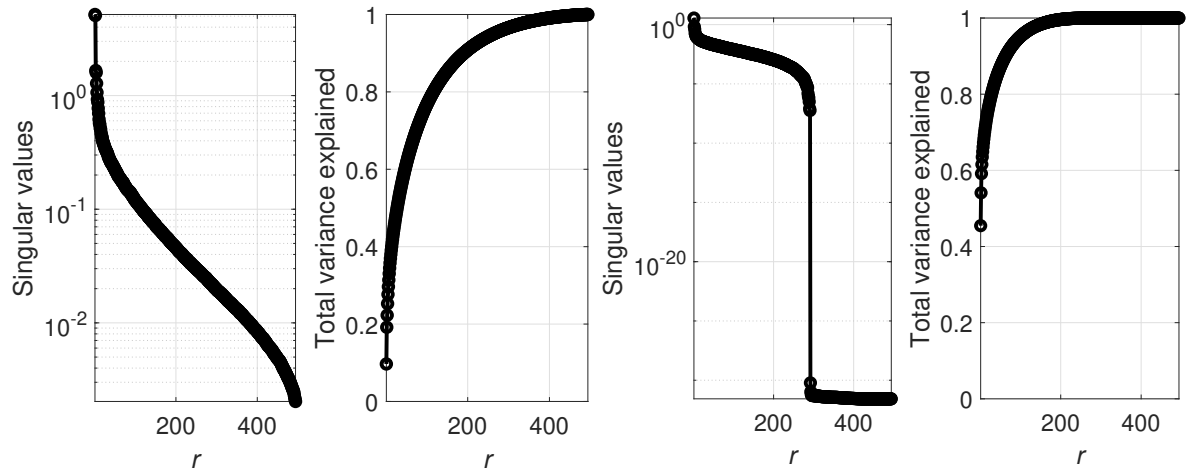
The same procedure as before is performed, that is computing the PCA of the entire data set of both sensors and visualising the data in the first three PCs, which can be seen in figure 6.11. Here, the separation of the raw truncated data from the two sample groups is worse than before. There is a significant overlap of the data in the PCs. These PCs do not have complete information that distinguishes between these two groups. However, for the low-rank data, there is a clear separation of the two groups with the first three PCs. It seems that there is no trouble in separating the groups of signals with the low-rank data set. Again, the total variance explained by the first principal components of the low-rank data is higher than for the raw data set as indicated by figure 6.12.



(a) Raw truncated data set

(b) Low rank data set

Figure 6.11: Principal component score of samples from sensors 3 and 19 for the first three principal components.



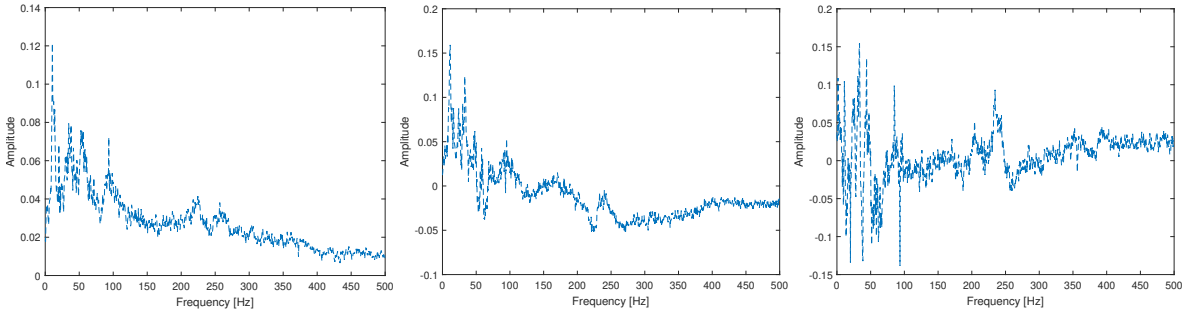
(a) Raw truncated data set

(b) Low rank data set

Figure 6.12: Singular values and total variance explained of each of the principal components. Data from sensors 3 and 19.

For the low-rank data, the two signal groups are separated by the first three PCs, but their separation is mainly in the 2nd and 3rd PCs. A line can be drawn in the 2nd and 3rd PC to separate the two data groups. Figure 6.13 shows the first three PCs. It is harder to interpret these PCs compared to previous PCs from the other analysis of the other sensor pairs. The 1st PC seems to be a combination of the

mean frequency vectors of the two sensors. No particular pattern can be observed with the 2nd and 3rd PCs. Similar to the other analysis, these two PCs contain information on the difference between the signal groups as the groups are being separated by these PCs. However, it is more difficult to see that this difference is due to the differences in structural systems compared to the other two sensor pairs.

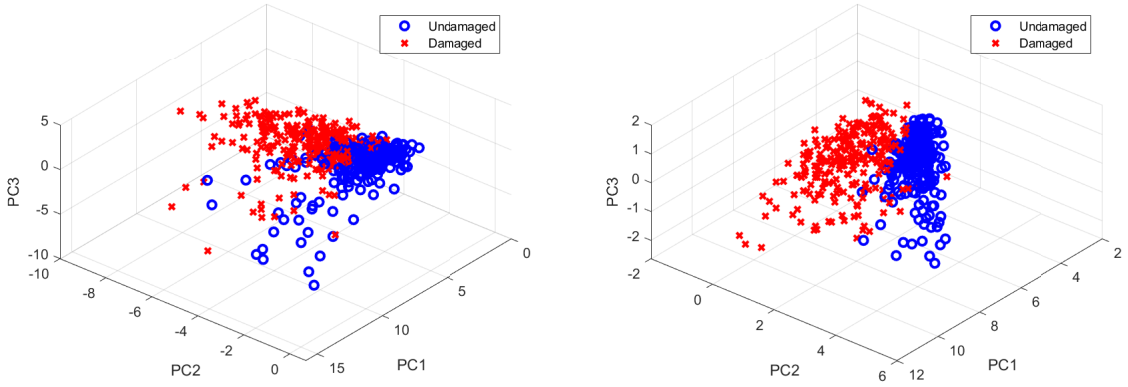


(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 6.13: First three Principal Component. Low-rank data set

Below 100 Hz - frequency resolution 0.5 Hz

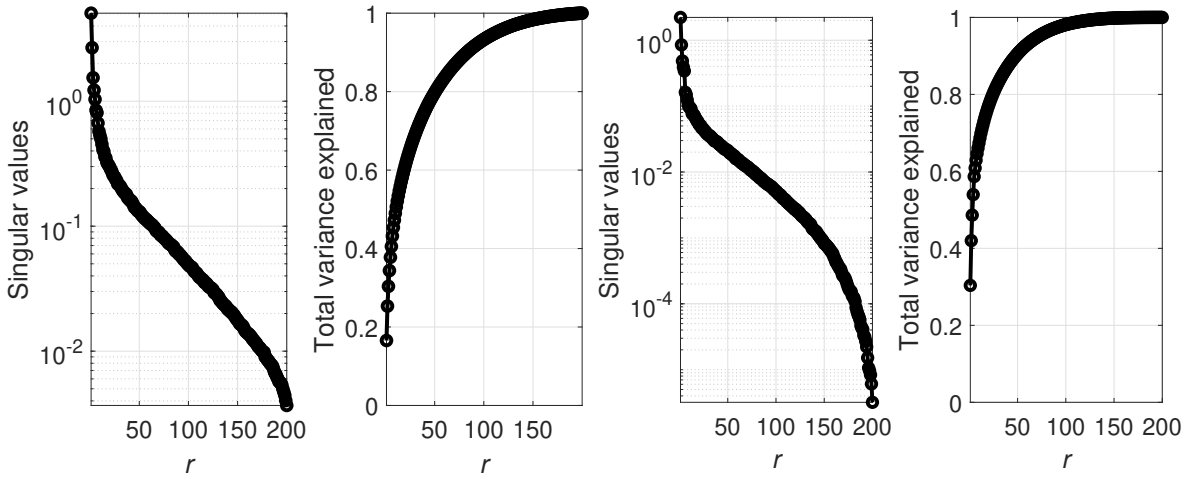
For all the analyses conducted, the entire frequency vector has been taken. This is suitable when natural frequencies occur over the entire spectrum or when changes in the signal can occur over the entire spectrum between the two sensors. However, in the case of sensors 3 and 19, most of the frequency content is below 100 Hz. If the dynamic system changes due to damage, it would most likely be at these frequencies. It could be redundant or even harmful for the analysis to use the entire frequency spectrum for a signal with low-frequency content like these sensors. There is little to no information on the dynamic system at these higher frequencies. A simple solution would be only to take the natural frequencies below 100 Hz. The resulting principal score of the observations in the first three PCs is shown in figure 6.14. Once more, there is a separation of the two groups. The signals of the two sensors are different below 100 Hz. The separation between the groups is not due to some irrelevant data at high frequencies but rather due to changes in the dynamic system at low frequencies. However, the separation between the data groups for low-rank data is better when the entire frequency vector is considered, but that could be due to irrelevant information in the higher frequencies.



(a) Raw truncated Data set

(b) Low rank data set

Figure 6.14: Principal component score of samples from sensors 3 and 19 for the first three principal components. Only includes data below 100 Hz. Frequency resolution: 0.5 Hz



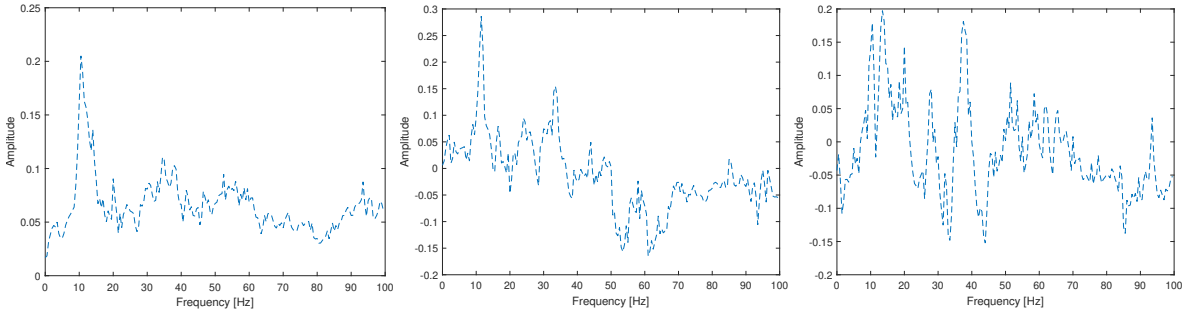
(a) Raw truncated data set

(b) Low rank data set

Figure 6.15: Singular values and total variance explained of each of the principal components. Data from sensors 3 and 19. Only includes data below 100 Hz. Frequency resolution: 0.5 Hz

The first three PCs are shown in figure 6.16. The two data groups are separated mainly in their different scores in the 2nd and 1st PCs. It is difficult to interpret these PCs, but they contain information about the differences between the data groups, which should correspond to the underlying differences in

dynamic properties. Another observation is in figure 6.15, which shows the total variance explained by the PCs. In the previous analysis, after the third PC, there was a significant drop in the variance explained by the PCs. However, here the variance drops after the fifth PC. The fourth and fifth PCs could also contain relevant information regarding the differences in the dynamic properties between the data groups.

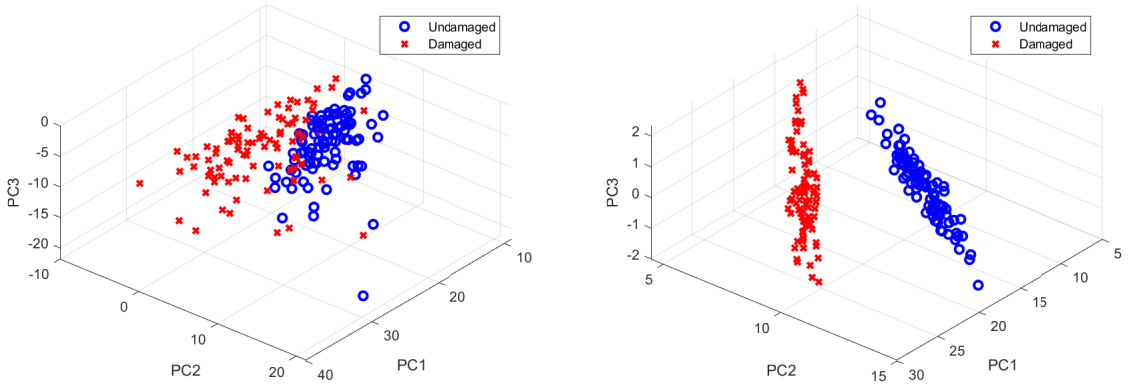


(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 6.16: First three Principal Component. Low-rank data set

Below 100 Hz - frequency resolution 0.125 Hz

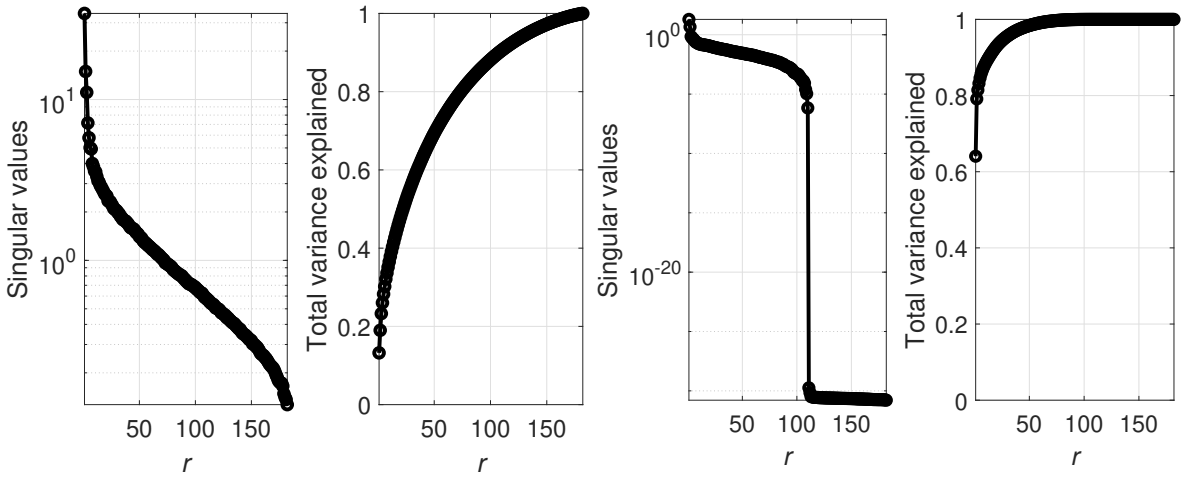
Due to the level of difference between the data groups, it seems sufficient to use a frequency resolution of 0.5 Hz to separate the two different data groups in the subspace of the first three PCs. However, the analysis can be repeated with a higher frequency resolution of 0.125 Hz to compare with the previous analysis. As previously stated, this will lower the number of samples in the analysis and also increase the size of the frequency vector. The sample-to-feature-length ratio has an effect on the performance of the rPCA. Too few samples and the rPCA cannot discover the underlying dynamic system properties. With the frequency resolution of 0.125 Hz, the number of samples for the analysis drops to 183, 93 from sensor 3 and 90 from sensor 90.



(a) Raw truncated Data set **(b)** Low rank data set

Figure 6.17: Principal component score of samples from sensors 3 and 19 for the first three principal components. Only includes data below 100 Hz. Frequency resolution: 0.125 Hz

Figure 6.17 shows the data in the subspace of the first three PCs. For both the raw truncated samples and the low-rank there is a separation of the data groups in the PCs. It is an improvement compared to the previous analysis. For the low-rank data, the separation is exclusively in the 1st and 2nd PCs. Looking at figure 6.18, these two PCs explain most of the variance in the data set, and after the second PC, there is a sharp decrease in the total variance explained by the PCs. The PCs of the low-rank data are shown in figure 6.19. PCs 1 and 2 contain information regarding the differences between the two data groups. It can be seen that these PCs contain anomalous ”spikes” at around 90 Hz, and they are

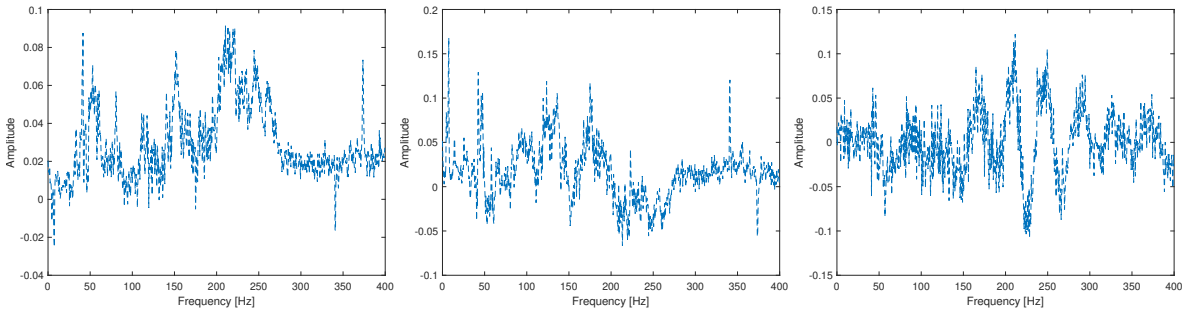


(a) Raw truncated data set

(b) Low rank data set

Figure 6.18: Singular values and total variance explained of each of the principal components. Data from sensors 3 and 19. Only includes data below 100 Hz. Frequency resolution: 0.125 Hz

opposite to each other in these PCs, which is similar to what has been seen in the previous analyses. It is difficult to interpret these PCs in relation to the dynamic system properties. It could be that these Pcs contain information regarding the underlying differences in the dynamic system, but it could also be due to differences in the level of noise between the two data sets. Caution should be taken when interpreting these results.



(a) First Principal Component

(b) Second Principal Component

(c) Third Principal Component

Figure 6.19: First three Principal Component. Low-rank data set

Three sensor pairs have been considered, but there are a total of 16 sensor comparisons possible. For all valid sensors (sensors that generate a detectable vibration), there is a statistical difference between all sensor pairs for the low-rank data set. There is a separation of the two signal groups in the first three PCs. The rPCA is able to remove the noise within the samples and reveal the underlying dynamic properties.

6.2. Points of maximum difference

With the PCA, the combined data of two sensors have been embedded into a low-rank subspace of the PCs. The PCs explain the correlation within the data and are orthogonal to each other. They contain information about the entire feature or frequency vector, and the samples are clustering according to information in the entire vector. However, it could be of interest to find which frequency is contributing most to the separation between the two sample groups. Here the Sparse sensor placement optimisation for classification (SSPOC) [7] can be used to find the points in the frequency vector that contains the most discriminating information between two categories of the data. Out of a 1000-length frequency vector which points contain the most discriminatory information between two groups of data? When

comparing two categories, the method can find locations in the frequency vector up to the number of features. Here in the previous analysis, the number of features is 3, corresponding to the three PCs. There is a requirement that the categories of data form distinct clusters in the subspace of the features to allow classification algorithms, in this case, the Linear Discriminant Analysis, to function correctly.

Here any N number of features or PCs can be selected and used with the method. However, as was discovered in the previous analysis, the two sample groups of damaged and undamaged samples are separating in the first three PCs. These PCs are selected and used with the method. Figure 6.20 shows the three locations in the frequency vector that contains the most discriminatory information between low-rank data from sensors 7 and 23. The left figure contains the discriminant projection vector, and the red line shows the three sparse locations. The value of the sparse locations dictates the relative discrimination of each of the three points. A better view is to look at the sparse locations overlayed on the mean frequency vectors of sensors 7 and 23, which is shown in the right figure.

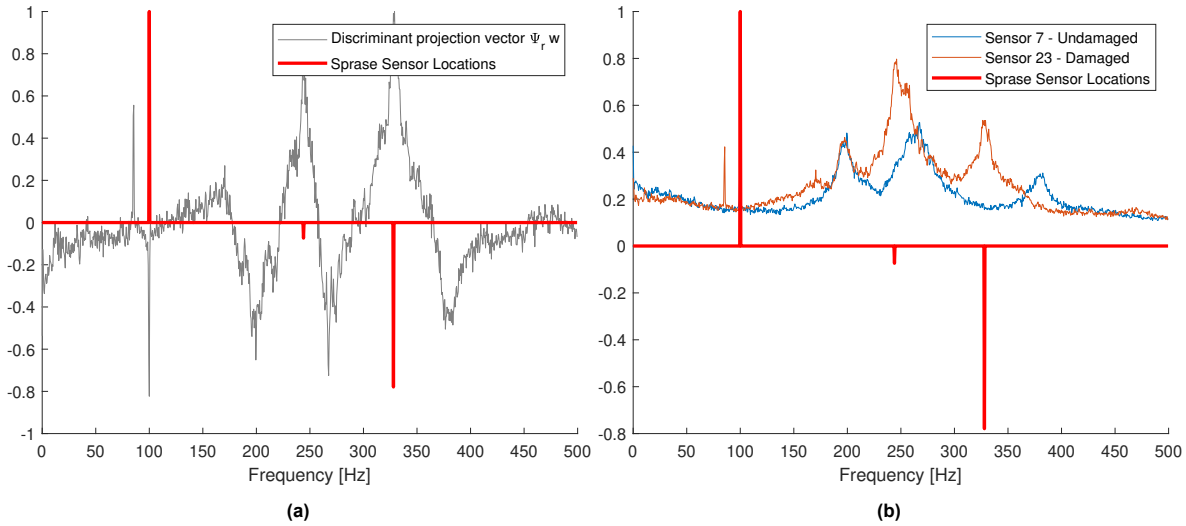
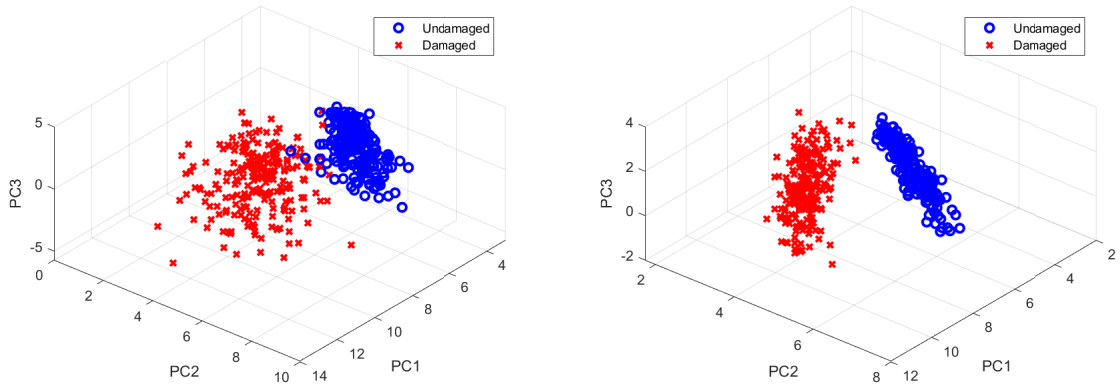


Figure 6.20: Sparse sensor locations for Low-rank data of sensors 7 and 23. (a) Discriminant projection vector with the sparse sensor locations marked in red. (b) Mean vector of sensors 7 and 23

Here it can be seen that the locations that contain the most discriminatory information between the sample groups are at the natural frequencies. However, the point in the frequency vector that holds the maximum amount of difference between the sample groups is the anomalous ”spike”. This is not that unexcepted, as these spikes are different for the two sample groups and inhabit only a single frequency index. Furthermore, these anomalous ”spikes” are always the same within their respective sample group. These anomalous ”spikes” practically function as a dimensionality reduction for the problem. Only these two points can be followed, and that would give enough information to categorise the data. However, the anomalous ”spikes” are not part of the dynamics of the system but are likely due to the sensors and auxiliary equipment. These anomalous ”spikes” are then adding a false-positive bias to the data. In fact, the ”spikes” are the location of the maximum difference between the data sets, at least for data from sensors 7 and 23.

This leads to the question of whether the separation in the data in the first three PCs is due to the false positive of the anomalous ”spikes”. The PCA of the combined data set of two sensors can be performed again, but this time excluding the frequency range where the anomalous ”spikes” appear. Figure 6.21 shows the low-rank samples from sensors 7 and 23 in the first three PCs but only in the frequency range of 110 Hz to 500 Hz. **Here there is still a separation between the two sample groups, even when excluding the anomalous ”spikes”**. The separation of the data in the first PCs is not solely due to these anomalous ”spikes”. The same occurs for the other sensor pairs. However, it is not possible to take this approach to filter out the anomalous ”spikes” when natural frequencies are in the same band as these anomalous ”spikes”.

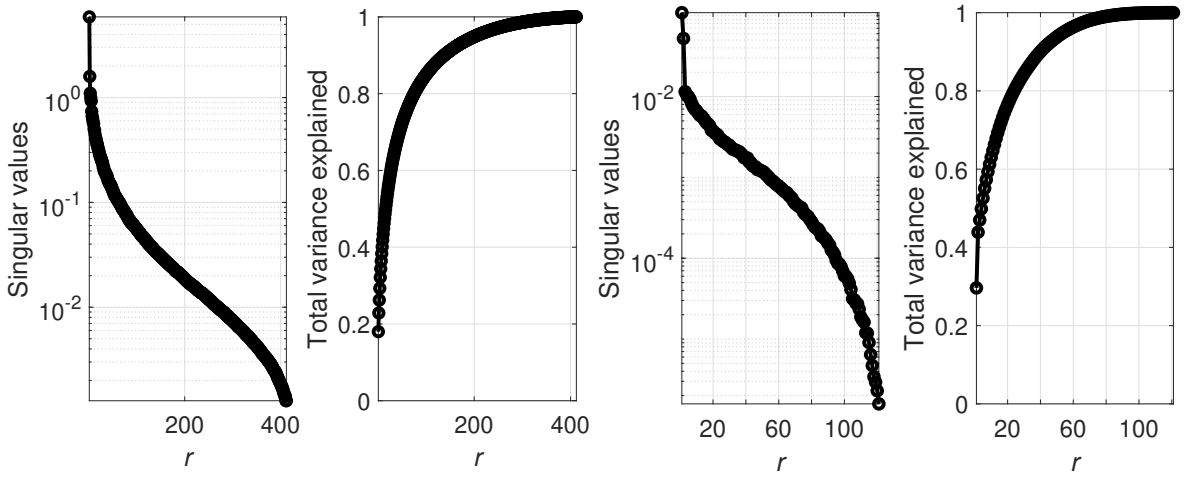
While the anomalous ”spikes” are locations that contain the most discriminatory information between



(a) Raw truncated data set

(b) Low rank data set

Figure 6.21: Total variance explained for the ”raw” and low-rank data sets excluding the anmalous ”spikes”. Reduced length of the frequency vector from 110 Hz to 500 Hz.



(a) Raw truncated data set

(b) Low rank data set

Figure 6.22: PCA score in the first three principal components excluding the anmalous ”spikes”. Reduced length of the frequency vector from 110 Hz to 500 Hz.

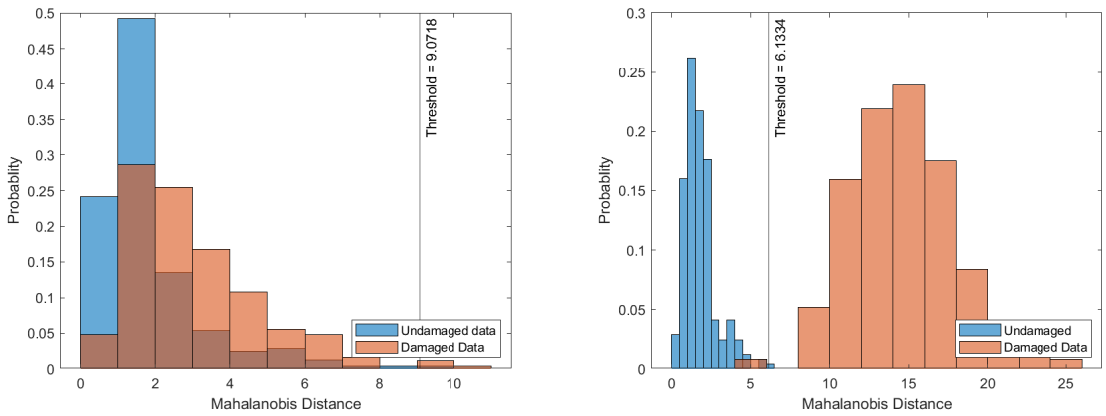
the sample groups, that does not mean that the data only depends on these locations. The anomalous ”spikes” occur just at a single frequency index. There are also other points in the frequency vector, such as the natural frequencies, that contain information regarding the difference between the two groups. However, this information on the natural frequencies is spread out over more frequency indexes than the anomalous ”spikes”. This leads to each of the indexes containing less relative information about the difference between the two groups compared to the anomalous ”spikes”. While the anomalous ”spikes” contain the most discriminatory information between the sample groups, that does not necessarily mean that they dictate the entire difference of the vectors. However, the problem that these anomalous ”spikes” pose to the damage detection process becomes more critical when the overall difference between two dynamical systems becomes less. Then the relative contribution of the anomalous ”spikes” to the difference between sample groups becomes more and can start dictating the difference between sample groups.

6.3. Novelty detection

The next step in the damage detection process is to identify if there is actual damage present in the signals. In the previous analyses, only a visual confirmation has been made so far on the division between undamaged and damaged sample groups in the subspace of the first PCs. However, a more robust approach than visual identification is needed to automate the damage detection process. Setting up the damage detection problem as unsupervised novelty detection would be ideal. The problem here is to answer whether there is a presence of damage in the structure, that is, whether there has been a deviation in the normal ”healthy” condition of the structure. The damage comparison is always a comparison between two states. Here this is the comparison between sensors in the damaged and undamaged areas for each of the sensor pairs.

In the previous analysis, it was found that the low-rank data from the different sample groups formed distinct clusters in the first PCs. The features are the ”score” of the samples in these PCs. There is a considerable distance between the features belonging to the damaged and undamaged sample groups. This can be exploited to automatically detect a deviation in the normal condition. If the ”distance” is too great, that would indicate deviation in the normal condition. This assumes that the subspace of these PCs corresponds to variance related to the differences in the underlying structural system of the two sample groups. It could be that the sample groups are separating the subspace of the PCs, but those PCs would not be connected to the variance in the different underlying structural systems but rather differences in the level of noise in the sample groups. For instance, the data could be forming clusters in the subspace of the PCs due to the anomalous ”spikes”. Therefore, it has to be certain that the PCs are connected to the variance that explains the difference in dynamic systems.

Mahalanobis distance, outlined in chapter 4.5, will be used to measure the distance between samples. It measures the distance between the features and a given distribution. Here the distance of the samples can be measured relative to the distribution corresponding to the undamaged state, which represents the normal condition of the bridge. Since there is a separation between the undamaged and damaged samples in the subspace of the PCs, that means that the damaged samples are further away from the undamaged distribution. If the distance between the two distributions is significant, that would mean that the two distributions have different ”scores” in the PCs and are therefore different. This would imply that there is damage present causing this deviation. If the distance is the same between the two distributions, that would mean that the two distributions have the same ”scores” in the PCs, and there is no difference between them, at least according to these PCs.



(a) Raw Data set

(b) Low rank data set

Figure 6.23: Mahalnobis distance

Figure 6.23 shows the distribution of Mahalanobis distance for the undamaged or ”training” data and the damaged or ”test” data. Data is from sensors 3 and 19. When the raw truncated data is used, there was no separation of the data in the first three PCs, as seen in figure 6.11. The samples from both groups inhabit the same area in the PC subspace; there is no difference between them according to

the first three PCs. The distribution of the Mahalanobis distance is nearly the same for the two groups. However, when the low-rank data is used, there is a separation of the sample groups in the subspace of the first PCs. This leads to the distribution of the Mahalanobis distance for the two groups being completely different, as the damaged sample group is far from the ”healthy” sample group.

This then enables the identification of whether samples deviate from the normal condition of the bridge by only looking at the Mahalanobis distance. For the low-rank data, the maximum Mahalanobis distance in the undamaged data can be taken as a threshold. Any sample that has a higher Mahalanobis distance to the undamaged distribution than this threshold could be said to be deviating from the normal condition of the bridge. This is the case for nearly all samples from the damaged data set for the low-rank data. They can be classified as a novelty.

This is a rather crude approach to picking this threshold value. Ideally, it should be based on an extensive statistical analysis of the data. However, since there is such a large difference in the score of the samples in the PCs, simply taking the maximum Mahalanobis distance in the undamaged data is enough to classify the data. This brings another point regarding comparing the undamaged and damaged signals. The damage signals are from an area where damage has already developed to quite some extent with the presence of two fatigue cracks. The damaged area also has a different structural configuration compared to the ”healthy” reference area. Both of these points mean that the damaged signals differ considerably from the undamaged ones. The goal of robust damage detection would be to detect damage before it reaches this level of damage development, such as at the early stages of fatigue crack formation. This would mean a more challenging problem as the underlying difference in dynamical systems between the healthy and damaged areas would be less, and it is not certain if the damaged and undamaged signals would separate in the PCs.

6.4. Conclusions

rPCA significantly improves the results of the damage detection process. Using the rPCA on vibration data reduces the noise, and coherent structures within the data are revealed. The low-rank data has information regarding the underlying dynamics that the sensor captures. When the PCA is performed on the combined low-rank data set of two sensors, there is a clear separation between the damaged and undamaged samples. Without the use of the low-rank data from the rPCA, it would not be possible to determine that there has been a deviation in the normal condition of the structure.

For all valid comparisons between sensor pairs, there is a statistical difference between damaged and undamaged signals beyond that of the randomness of the traffic. This indicates that the sensors in the damaged area have a different dynamical system than those in the undamaged area. The difference in the dynamical system can be rather significant, with entire natural frequencies shifting considerably in the frequency spectrum. **However, as alluded to at the beginning of the section, the difference in dynamical systems is not necessarily due to the damage caused by the fatigue cracks**. The structural configuration of the undamaged and damaged areas is not completely the same. In the damaged area on one of the stiffeners, there is a flange discontinuity. In addition, there is a coupling plate and a strut in the vicinity. These two structural components are located slightly differently than the equivalent coupling plate and strut in the undamaged area. Essentially the structural configuration of the two areas is different, and it is not possible to claim the difference is actually due to the damage caused by the fatigue cracks.

However, there are indications that the difference in the signals between the two areas is due to stiffness reduction. The distribution of energy within each sample differs between sensors in the undamaged and the corresponding sensor in the damaged area. The mean energy of the samples is higher for sensors located in the damaged area. It is unlikely this is due to the randomness of the traffic, as the exact same vehicles drive over both areas. Reduced stiffness will result in a higher response given the same amount of force. Damage such as the fatigue cracks can reduce the stiffness of the structure and could be responsible for this behaviour. However, the same could be said for the different structural configurations; it can lead to a reduction in stiffness compared to the baseline area.

Even though it is not possible to determine that the difference in signals between the sensors of the two areas is due to damage, it is comforting that the difference between the signals is that great. Consider the alternative. If there was no fundamental difference between signals of the two areas, even with the presence of fatigue cracks and a different structural configuration, it would bode ill for SHM. How would then damage or structural distortions of less magnitude be detected? The fatigue cracks have already developed to a significant length. The goal of SHM would be to discover these cracks before they become an issue.

The rPCA is not required to produce a division between data from sensors 7 and 23 in the subspace of the PCs. For these two sensors, the performance of the raw truncated data set was on par with the low-rank data set. The aim of the rPCA is to create a low-rank matrix that is robust to outliers within the original data. Regular PCA, which is used when two sensors are compared, is not robust to outliers and is best suited for data that follow a Gaussian distribution. If the performance of the comparison is the same for the raw truncated data set and low-rank data set, that could indicate that there are no or few outliers within the raw truncated data set. Several pre-processing steps have been taken to simplify the data. Samples with high energy were removed and thus simplified the distribution of samples. This could explain the good performance of the raw truncated data set. However, this does not explain why the raw truncated data performs so particularly well for sensors 7 and 23. It could be due to the significant difference between the two signals. This sensor pair exhibits the largest difference in the mean frequency vectors of all sensor pairs. Only the first three PCs are considered, and the PCs are ordered in the amount of variance that they explain. The first PCs explain more of the variance in the data relative to other higher PCs. It could be that the difference between the signals is so large for sensors 7 and 23 that it results in these first PCs capturing this variance through all the noise within the data.

The number of samples used in the analysis, or rather the ratio between samples and the length of the frequency vector, is something to consider when working with the rPCA. This was observed in the analysis of the low-frequency content below 100 Hz. When the frequency resolution was increased to 0.125 Hz, it resulted in fewer samples compared to the previous analysis. The separation of the damaged and undamaged samples in the first PCs was better in the data set with fewer samples. This was the case for both the raw truncated data and the low-rank data. Introducing more samples means more realisations of the random process of the input force, which in turn leads to more measurement noise in the data set. It could be that due to the lower noise within the data, the differences between the damaged and undamaged samples have a higher relative significance in the data set, and thus, fewer samples give a better separation between the damaged and undamaged samples in the PCs. However, this would also mean fewer amount of samples would not encompass the entire random process of the input force. The comparison between damaged and undamaged samples could then be comparing different levels of noise in the damaged and undamaged sample sets in addition to the differences in the dynamic properties.

7

2nd Measurement Campaign

After it was discovered that there was a difference in the structural configuration between the undamaged and damaged area in the first measurement campaign, a different approach was taken with the second measurement campaign. Instead of comparing two areas, all 32 sensors are installed in the same area. In addition to these accelerometers, several temperature sensors were installed on the various bridge components, such as in the asphalt, under the bridge deck, and on the stiffeners. This allows for studying the effect of temperature on the dynamic system of the bridge. Compared to the damage detection in the previous measurement campaign, environmental detection has several advantages. Firstly, there is no need for actual damage. Temperature changes will always occur throughout the day. It should always be possible to perform environmental detection. Furthermore, the output of the same accelerometer can be compared at different temperatures. Thus, only environmental changes take place. The structural configuration of this sensor will not change, as was the case in the previous measurement campaign when comparing two sensors in two different areas. The intended goal of SHM is to measure damage but is knowledgeable to see if it is possible to detect changing environmental conditions and how the system properties look at a given temperature.

Accelerometer data from 32 sensors over two days will be used for this analysis. The measurements' id, date, and temperature range are displayed in table 7.1. Note that the temperature range of these two days barely intersects. Of the 32 accelerometers, 30 measure the vertical acceleration of the deck plate, while the remaining two measure the horizontal acceleration at the stiffeners. As was seen in the previous measurement campaign, these two types of accelerometers have fairly different frequency content. The vertical accelerometers on the deck mainly capture low-frequency content, while it is the opposite for the horizontal accelerometers on the stiffeners, which capture mostly high-frequency content. These two dynamic systems might respond to changing environmental conditions in different ways.

Table 7.1: Measurement data - 2nd Measurement Campaign

ID	Date	Start time	End time	Temperature range
09762 - 09785	18-10-2021	00:14:57	23:29:57	11.5-17.5 C
10147 - 10170	03-11-2021	00:17:01	23:32:01	7.9 - 12.5 C

The rPCA can be used again to reduce operational variability within the data. However, rPCA should be used when the underlying dynamic system is the same within a data set. It should only be applied to samples that share the same temperature. Here, rPCA will be applied to samples within a single centigrade. This should limit the environmental variability within the rPCA so that the underlying dynamic system is similar for all samples while also providing an adequate amount of samples for each rPCA. The distribution of samples over the temperature range is shown in figure 7.1. The samples are generated from accelerometer number 31 with a frequency resolution of 0.5 Hz and with data from the

two days. The samples are categorised according to temperature sensor 6, which is located on the same stiffener as accelerometer number 31. The uneven distribution of samples can be attributed to the varying traffic flow throughout the day, with more traffic during the rush hours and assuming that the temperature is a function of the time of the day.

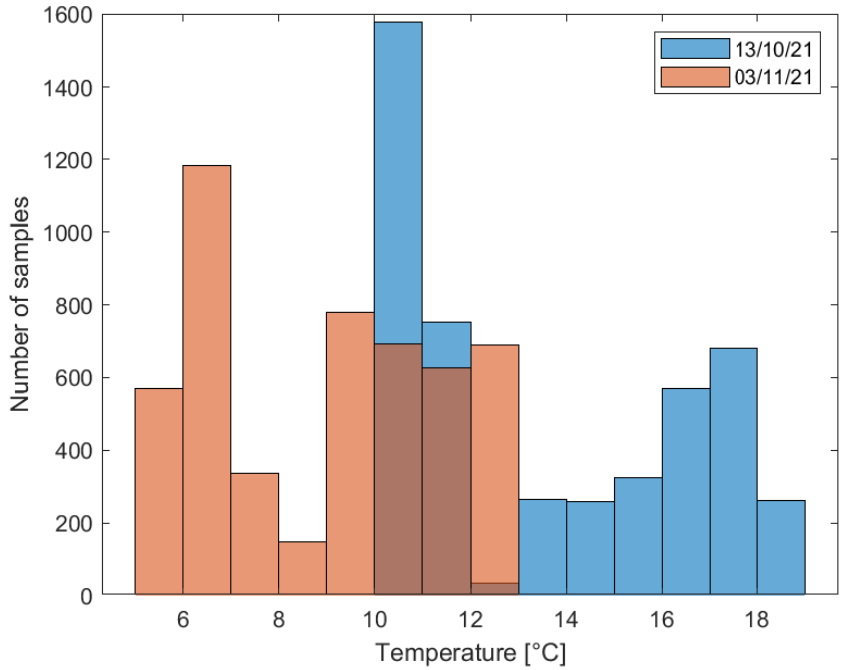


Figure 7.1: Histogram of the number of samples in each of centigrade for the two measurement days. Data from sensor 31. Samples have a frequency resolution of 0.5 Hz

Overall the procedure is similar to what was done in the damage detection in chapter 6. The procedure can be summarised as follows:

1. Create samples by taking a time window around peaks in the time series of a single sensor
2. Discard 20 % of the highest energy samaples for each sensor
3. Apply discrete Fourier transform to the samples
4. Take the absolute of the fourier transform
5. Obtain a low-rank matrix by performing rPCA on a sample set within a particular temperature interval. Default regularisation parameter of $\lambda = 1/\sqrt{max(n,m)}$ and a lower regularisation parameter of $\lambda = 1/\sqrt{2 * max(n,m)}$ is used for the analysis.
6. Compute (regular) PCA on the combined low-rank with data from two different temperature intervals.
7. Conduct feature discrimination with the score of the samples in the PCs as features.

The chapter is divided into three parts. The first part investigates the temperature of the bridge over the day and the temperature gradients in the bridge. The second part investigates the temperature effect on the dynamic properties of a single resonance frequency. Finally, in the third part, the whole frequency vector is analysed in relation to temperature changes. The Sparse Sensor Placement Optimisation for Classification (SSPOC) will also be used to find the locations in the frequency vector that contain the largest difference between the data for different temperatures.

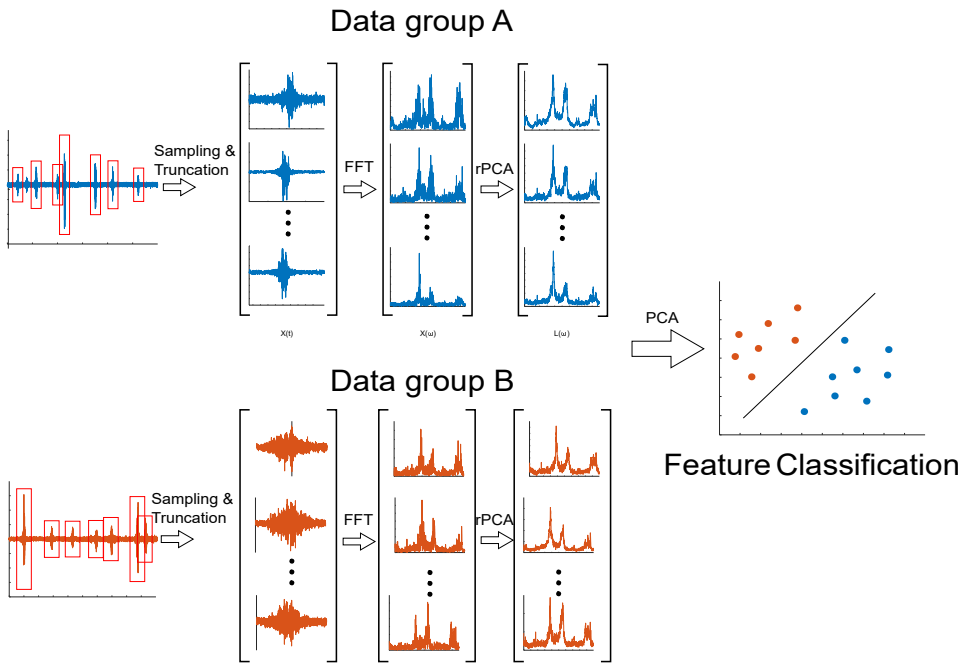


Figure 7.2: General preprocessing procedure. Samples are created around excitation in the time domain and the Fourier transformed applied. Afterwards samples are grouped according to temperature and the rPCA is applied

7.1. Temperature Sensors

A total of 16 temperature sensors are present on the bridge. These sensors are positioned at different locations of the bridge and measure the temperature of the different structural components. Figure 3.8 in chapter 3.3 shows the layout of the sensors (Note that the temperature sensors in the asphalt are missing, sensors 9 - 12). These sensors span the entire height of the bridge with temperature sensors in the asphalt, under the bridge deck, on the stiffeners, on the east wall, and finally on the lower transverse beam. These temperature sensors give an idea of the overall temperature profile of the bridge.

Each day the temperature will rise over the day and then fall in the evening and afternoon. Figure 7.3 shows the temperature of four sensors over four and half days. All of the sensors follow the same fundamental pattern: a rise in temperature followed by a decrease in temperature, and this process repeats each day. However, there is a difference in the temperature of each sensor, indicating that there is a temperature gradient in the bridge. Figure 7.3 contains temperature measurements from sensors that form a vertical line in the bridge. It can be seen that during the day, when the temperature is at its highest, the temperature at the top part of the bridge in the asphalt is highest. The temperature progressively drops as you go down the height of the bridge. The opposite happens during the evening and at night. Then the temperature at the top is lowest while it is highest at the bottom of the bridge. This would suggest that there is a vertical temperature gradient in the bridge.

Another observation is the phase lag between the peaks of the temperature of the asphalt and the stiffener. The asphalt warms up and cools down faster than the stiffener; it is more sensitive to temperature changes than other parts of the bridge. The inner layer where the stiffener is located retains its thermal energy for a longer duration. There is also a difference in the absolute value of temperature that the sensors measure. Asphalt undergoes the most extreme temperature shifts; it has both the highest and lowest temperature measurements. The lower transverse beam, on the contrary, undergoes minimal temperature changes. The top part of the bridge is exposed to solar radiation during the day. The lower side of the bridge could be exposed to warmer air from the Haringvliet.

There should also be a temperature gradient in the transverse direction of the bridge. However, the temperature sensors here are insufficient to capture the bridge's entire transverse length. There are

only sensors within the box girder, but none on the cantilevers. Fortunately, this has already been researched by Kortendijk [26] on the Haringvlietbrug. More temperature sensors were available for his research, and they covered a wider area of the bridge. It was discovered that the temperature of the outer part of the bridge deck, that is, the cantilevers was lower than in the inner part of the bridge deck when the temperature was at its highest. During the night, when the temperature is at its lowest, the outer cantilevers have a lower temperature than the inner part. The inner part of the bridge retained its heat better at night than the outer part. Kortendijk also found no evidence of a temperature gradient along the longitudinal length of the bridge.

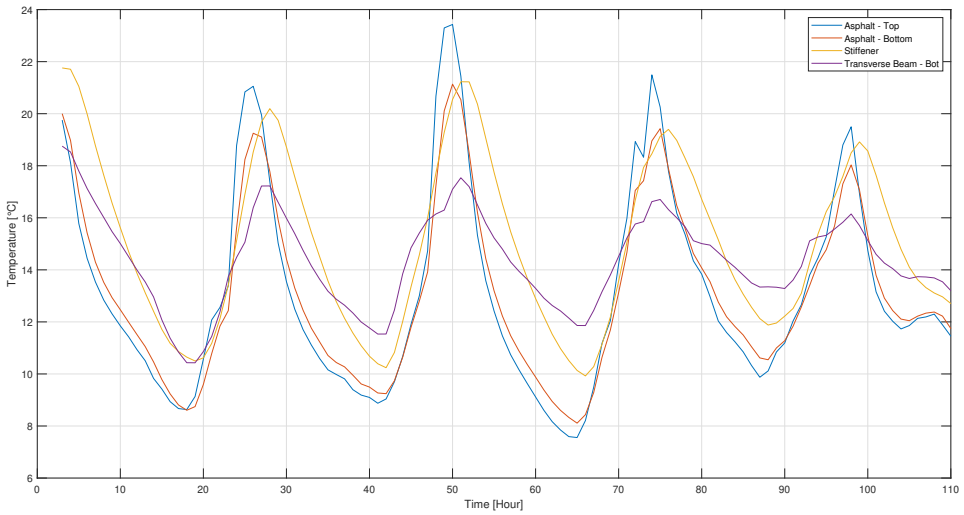


Figure 7.3: Temperature measurements of four sensors. Over a period of four and half days.

There are then two major temperature gradients in the bridge, which occur daily. This makes it more challenging to assess the effect of the temperature change on the dynamic properties of the system. A single temperature sensor cannot capture the entire temperature profile of the bridge; many sensors would be required to do so. However, connecting the information of many different temperature sensors to the acceleration data is much more complicated than just using a single temperature sensor. The simplest approach is to see how dynamic properties change with a single temperature sensor. Using many sensors would create many more temperature scenarios compared to using a single sensor.

However, as can be seen in figure 7.3, the temperature follows a reoccurring pattern. It could be that the temperature profile of the bridge undergoes the same pattern each day with slight deviations. It might be close enough to follow a single sensor, as this sensor is part of the profile. The sensors all have the same relation between the different days. A single temperature measurement should be part of the whole.

The most relevant temperature sensors should be those closest to the accelerometers. The accelerometers capture natural frequencies and corresponding vibration modes that vibrate at the sensor location. These dynamic quantities gain their characteristics from the surrounding material. This also includes the temperature of said material. Temperature sensors near the accelerometers should be more important than those farther away from the accelerometers. However, this does not mean that the temperature of material further away has no effect on the dynamic properties but rather that the relative contribution should be less. The temperature sensors of interest are then temperature sensors in the asphalt, under the bridge deck, and on the stiffeners. These temperature sensors are closest to the accelerometers. The temperature from these sensors will be mainly used in the analysis.

7.2. Single resonance peak

In order to better understand how temperature affects the dynamic properties of the system, a single resonance frequency can be investigated in relation to environmental variability. A single index of

the resonance frequency will be used, that is, the peak of the natural frequency where the highest amplification occurs. This greatly simplifies the problem as the whole frequency vector is reduced to a single variable. It also has the added benefit of limiting the analysis to a single natural frequency and how it responds to changing environmental conditions.

Data from two accelerometers will be considered. One measures the deck plate's vertical acceleration, and another measures the horizontal acceleration of one of the stiffeners. These are sensors 17 and 31, respectively. The frequency content of these sensors varies greatly, with the vertical sensor capturing mostly low-frequency content and the horizontal sensor that captures mainly high-frequency content. A single resonance frequency from both of these sensors will then be analysed with respect to the changing environmental conditions. In both cases, the accelerometers will be paired with the temperature sensor on the stiffener, that is, temperature sensor number 6.

7.2.1. Horizontal sensor - High natural frequency

Sensor 31 has high-frequency content, which can be seen in figure 7.4. The figure shows the mean frequency vector of all the temperature intervals, which was previously established in figure 7.1. There are a total of 9749 samples in the 14 temperature intervals available for analysis. The frequency resolution is 0.5 Hz for all samples. Here, the only major preprocessing step is the removal of the heavy-truck observations, as was outlined in chapter 5, and no rPCA has been applied to the data. The shape of the spectrum is almost the same between the different temperature intervals. However, there is a slight difference in amplitude, and the overall shape seems to shift with temperature. As mentioned before, instead of making a comparison between the whole frequency vectors, only the peak around 190 Hz is considered. That is, only this natural frequency is examined and how it changes with temperature.

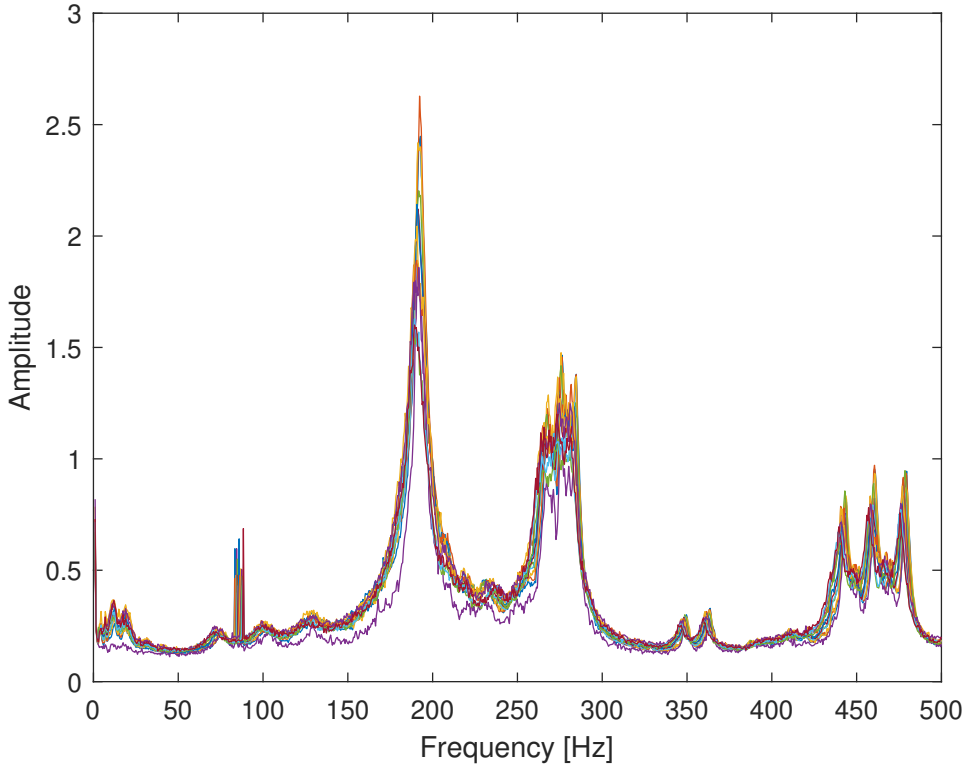


Figure 7.4: Mean frequency vectors of each of the temperature intervals. Raw truncated data from sensor 31. Frequency resolution of 0.5 Hz.

A simple way to visualise the effect of temperature on the natural frequency is to use the mean value of each temperature interval. Figure 7.5 shows the mean amplitude and frequency value for each temperature interval with respect to temperature. The data is from the raw truncated samples. There

seems to be a negative linear correlation between the temperature and the other two variables. If the temperature is lowered, the mean amplitude and frequency of the natural frequency increase. The negative temperature correlation to these variables is visible without using the rPCA.

Instead of only working with the mean, the whole distribution of the samples can be considered. Figure 7.6a shows a histogram with the number of samples for each given temperature interval and frequency of the resonance frequency. The natural frequency at around 190 Hz is being tracked for each sample in the data set. As the number of samples is uneven for the different temperature bins, it is necessary to normalise each temperature interval with respect to the number of samples in each interval. This allows for a comparison between distributions of different temperature intervals.

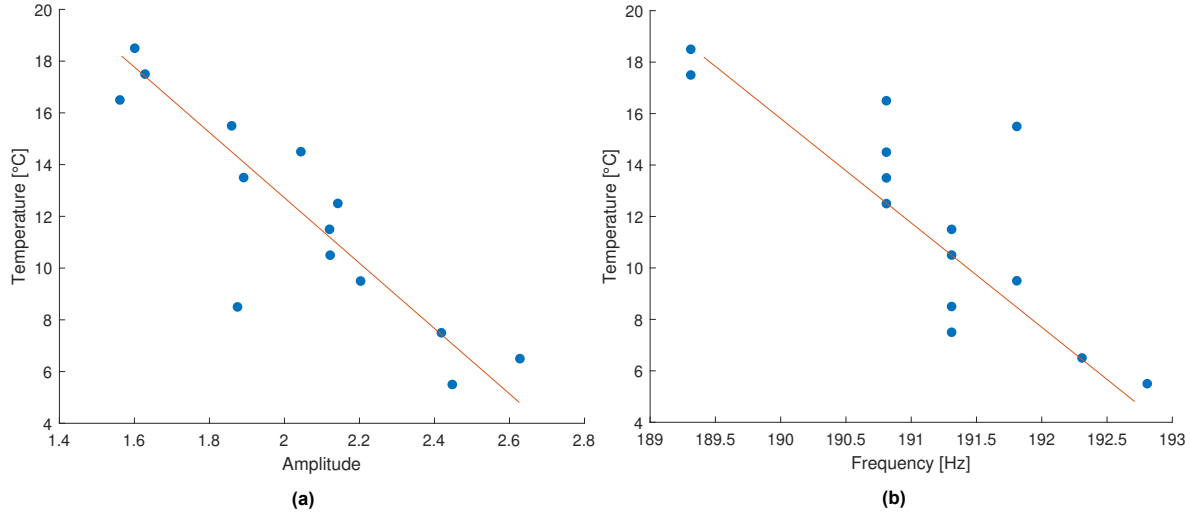


Figure 7.5: (a) Relationship between temperature and amplitude. (b) Realtionship between temperature and frequency. In both cases the variables have a negative linear correlation with temperature. Raw truncated data from sensor 31.

Within each temperature interval, there is a hint of a normal distribution. This distribution shifts with temperature, which can be better seen in the boxplot in figure 7.8. There seems to be a negative correlation between temperature and the distribution. The mean in the previous example serves as a decent indicator of the correlation between temperature frequency as the distribution within each temperature bin follows that of a normal distribution. Figure 7.6b shows a data set in which rPCA has been applied to samples in each temperature interval. The variance within each temperature interval has been reduced. However, the centre of mass of these distributions is similar to before, and thus the mean remains similar.

The distribution of the amplitude over the temperature intervals can also be examined and is shown in figure 7.7 for both the raw truncated data and the low-rank data from the rPCA. Here the data within each temperature interval seems to follow that of a skewed normal distribution. The shift of the distribution over the temperature interval can be seen when looking at the centre of mass of the distribution. Since this distribution is not a simple normal distribution, then the mean will not be at the centre of mass of the distribution; the mean, median and mode are not the same. There is then the question of how sound it is to interpret the shift in the mean as in figure 7.5b. Like before, the low-rank data has less variance than the raw truncated data set.

The negative correlation between temperature and frequency of a resonance frequency is established in the literature, as was discussed in chapter 2. However, the same cannot be said for the negative correlation between temperature and amplitude of a natural frequency. One speculation is that the vibration mode of this natural frequency is shifting in space. The accelerometer occupies the same location throughout all the measurements. However, the vibration mode could shift in space due to temperature changes. This could lead to the sensor no longer measuring the same spot of the vibration mode, leading to a different amplification level. Another possible reason could be that the damping characteristics change with temperature, which leads to a different level of amplification of the natural

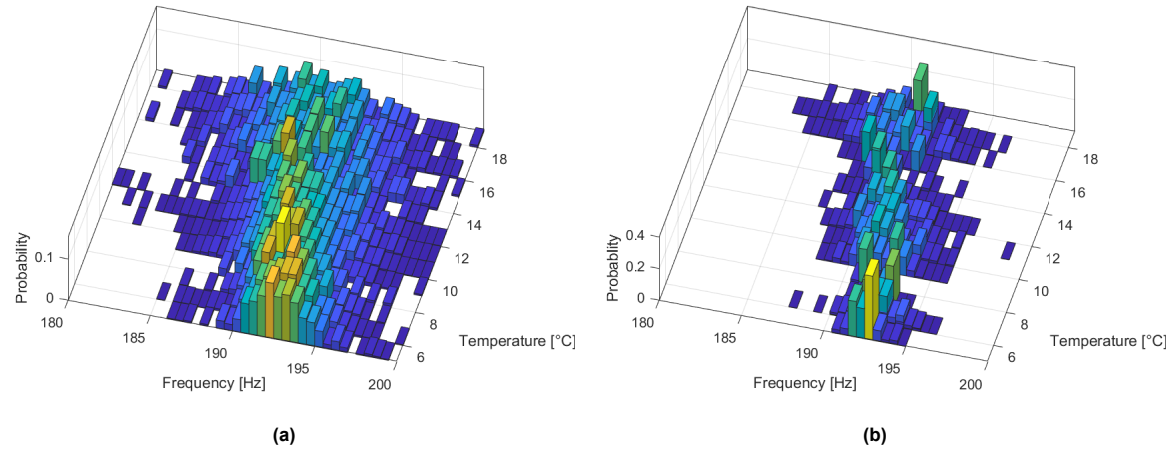


Figure 7.6: Distribution of the frequency value of the natural frequency for each sample. (a) Raw truncated data set (b) Low-rank data set. Data from sensor 31.

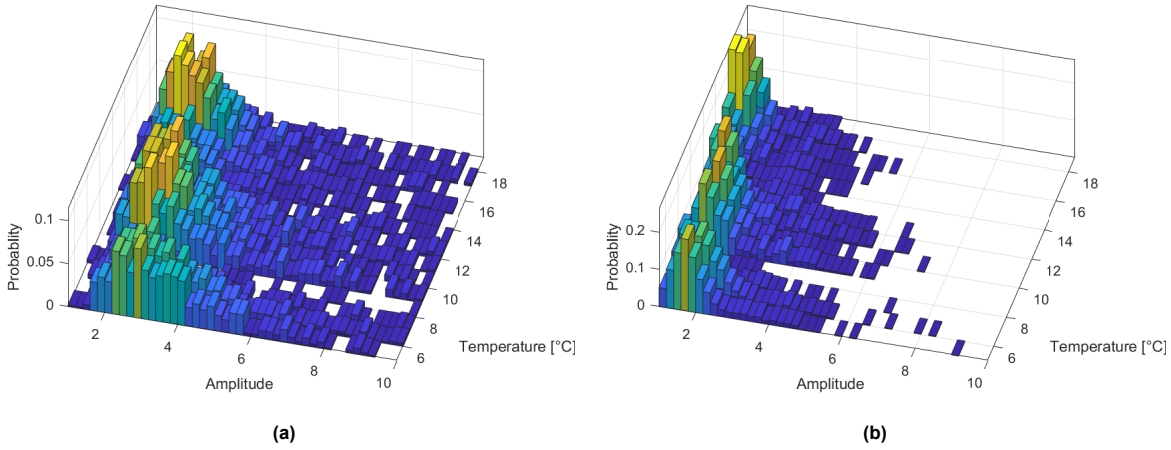


Figure 7.7: Distribution of the amplitude value of the natural frequency for each sample. (a) Raw truncated data set (b) Low-rank data set. Data from sensor 31.

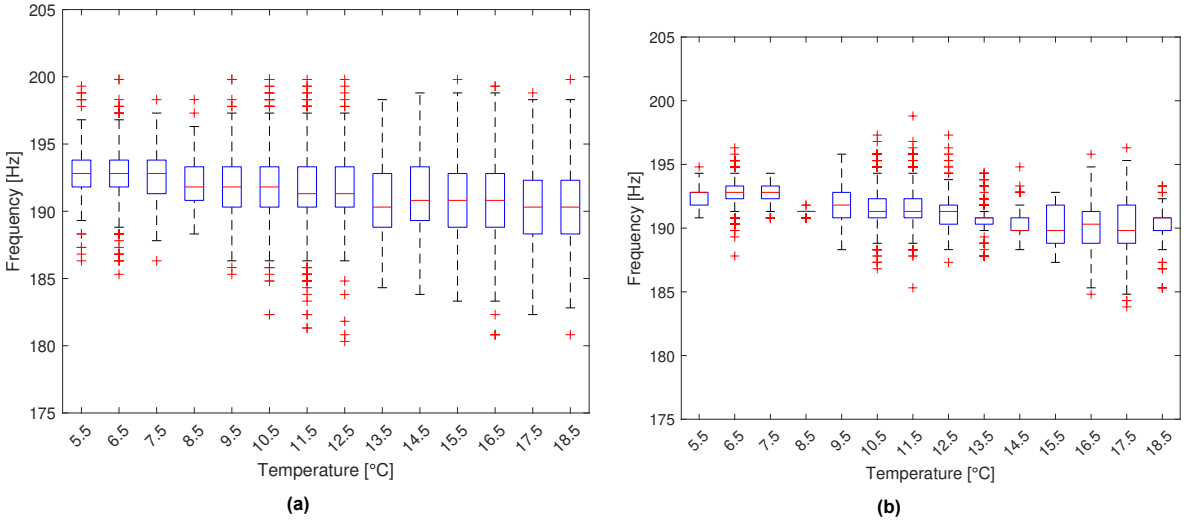


Figure 7.8: Boxplot (a) Raw truncated data set (b) Low-rank data set. Data from sensor 31.

frequencies.

7.2.2. Vertical sensor - Low natural frequency

So far, the analysis of the effect of environmental variability has only been carried out on horizontal sensor number 31. This sensor captures higher-order modes at a relatively high frequency. The resonance frequency at 190 Hz is well defined with apparent rise and fall in amplification; it is a textbook natural frequency. However, the natural frequencies of the accelerometers capturing the vertical acceleration of the deck plate are quite different, as seen in figure 7.9b, which shows the mean frequency vector of each of the temperature intervals for sensor 17. The sensor's frequency content is at low frequencies, below 100 Hz. The resonance frequencies are also more "sharp". There is a sudden rise and drop in the natural frequencies. The effect of temperature on the natural frequency at around 10 Hz will be analysed. The temperature sensor from the previous example will be used again to categorise the vibration data, that is, temperature sensor six, which is located on the adjacent stiffener.

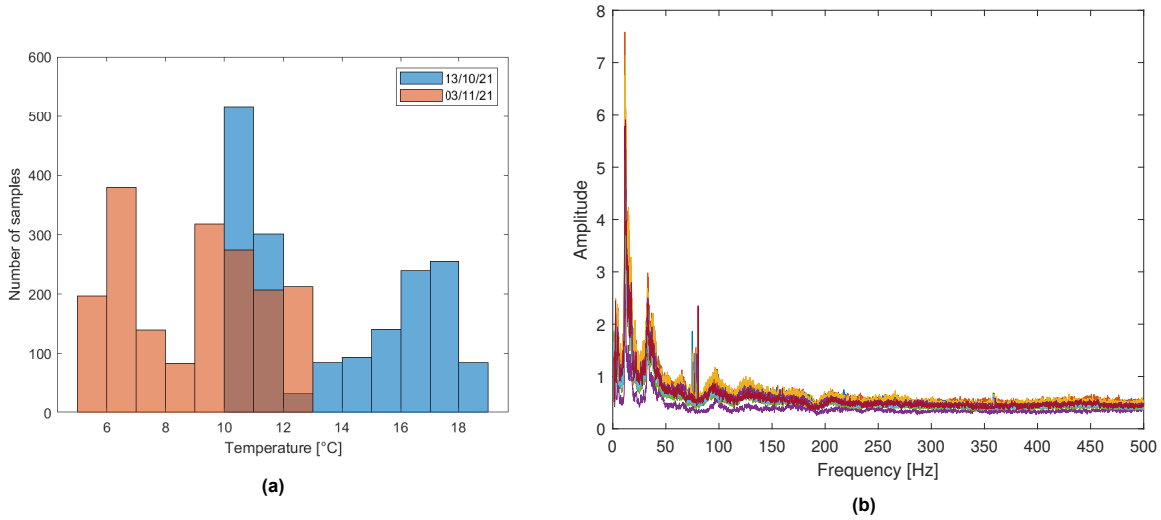
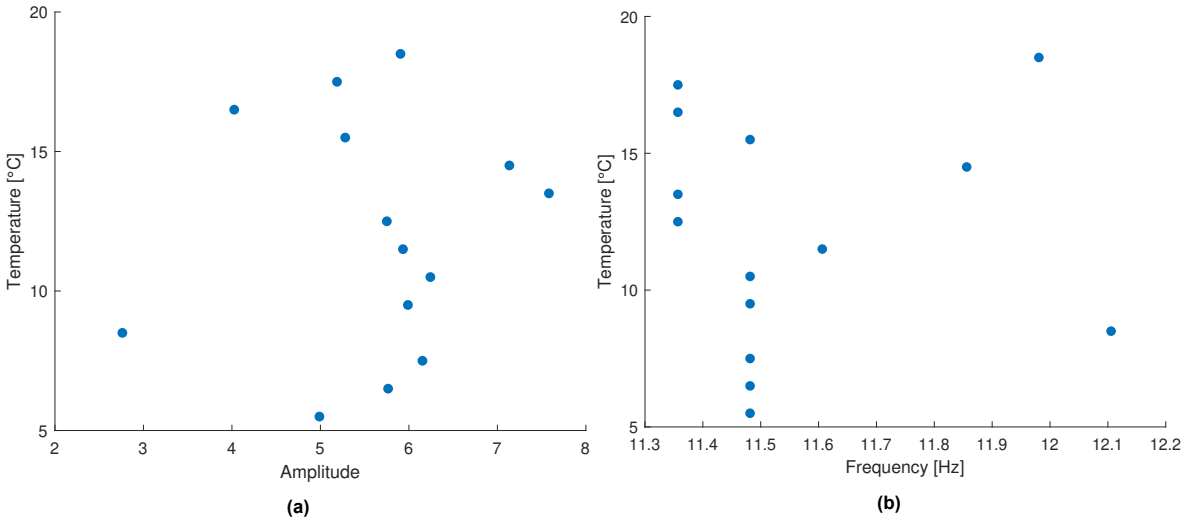


Figure 7.9: Histogram and mean of the temperature intervals for sensor 17. Data categorised according to temperature sensor 6. Samples have a 0.125 Hz frequency resolution (a) Histogram of the number of samples in each of centigrade for the two measurement days. (b) Mean frequency vectors of each of the temperature intervals.

There are some challenges with using low-frequency content data compared to high-frequency content, mainly the frequency resolution. To be able to capture the shift in natural frequencies due to temperature, the frequency resolution has to have sufficient accuracy. This is more important in the case of low frequencies, as the shift in temperature is based on the value of the frequency [41]. So, a one % shift for a 100 Hz frequency is 1 Hz but 0.1 Hz for a 10 Hz frequency. Therefore the frequency resolution in this analysis is taken as 0.125 Hz. It was previously 0.5 Hz. However, this requires that the time window of the samples increases to eight seconds. This results in a lower amount of samples or 3562 samples for all 14 temperature intervals. The distribution of the samples over the temperature intervals can be seen in figure 7.9a

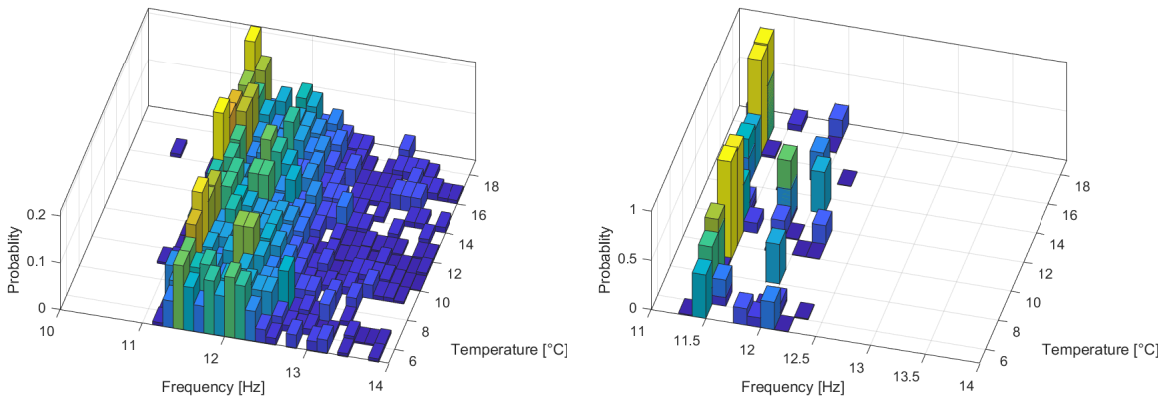
The same procedure is followed as in the previous analysis. First, the mean value of the natural frequency is calculated for each temperature interval. Figure 7.10 shows the mean value of the frequency and the amplitude of the resonance frequency at 10 Hz in figure 7.9b. The linear regression line has been omitted due to the presence of outliers. Compared to the high natural frequency of the previous analysis, there is less of a clear relationship between the variables. This could suggest that this low natural frequency is less sensitive to temperature changes compared to the high natural frequency from the previous analysis. However, there is a hint of a possible negative linear correlation between temperature and frequency. There are two "lines" of data points at 11.35 and 11.48 Hz. They have the same value of frequency over a wide temperature interval. Although completely ignoring the other data points is inaccurate, it reveals a negative correlation between temperature and frequency. The



(a) (b)

Figure 7.10: (a) Relationship between temperature and amplitude. (b) Relationship between temperature and frequency. Raw truncated data from sensor 17.

data is from two separate days and these two days had completely different temperatures, as seen in table 7.1. These two ”lines” of data points are from two separate days. Within each day, the frequency values are the same (ignoring the other data points). It is almost as if the frequency is a function of the day of measurement. There could be a difference in the environmental conditions between these two days that are not captured by a single temperature sensor. For instance, it could be that the temperature profile between these two days is different, which leads to this change in frequency. A single temperature sensor cannot capture the entire temperature profile of the bridge. This is a relatively liberal interpretation of the data. However, in the coming sections, more evidence will be presented to reinforce this claim that there is a higher difference in the environmental conditions of the bridge that measurements of a single sensor cannot capture.



(a) (b)

Figure 7.11: Distribution of the frequency value of the natural frequency for each sample. (a) Raw truncated data set (b) Low-rank data set. Data from sensor 17.

The distribution of the samples for the temperature interval and frequency are shown in figure 7.11. This distribution seems to follow that of a skewed normal distribution. There is a massive difference between the raw truncated data set and the low-rank data set, with the variance within the data set greatly reduced. For the raw truncated data, the peak of the natural frequency can take many different frequency values, but for the low-rank data, most of the samples fall within a single bin. There is a shift in the distributions when the temperature bins contain data from different days at around 12 °C. This is better seen with the low-rank data in figure 7.11b and mirrors what was discovered when investigating

the mean values of the frequency. The frequency value of the natural frequency seems to depend on which day the data is from.

Looking at the distribution of the amplitude of the samples in figure 7.12, there doesn't seem to be any particular pattern. However, the variance with the data set is significantly reduced with the application of rPCA.

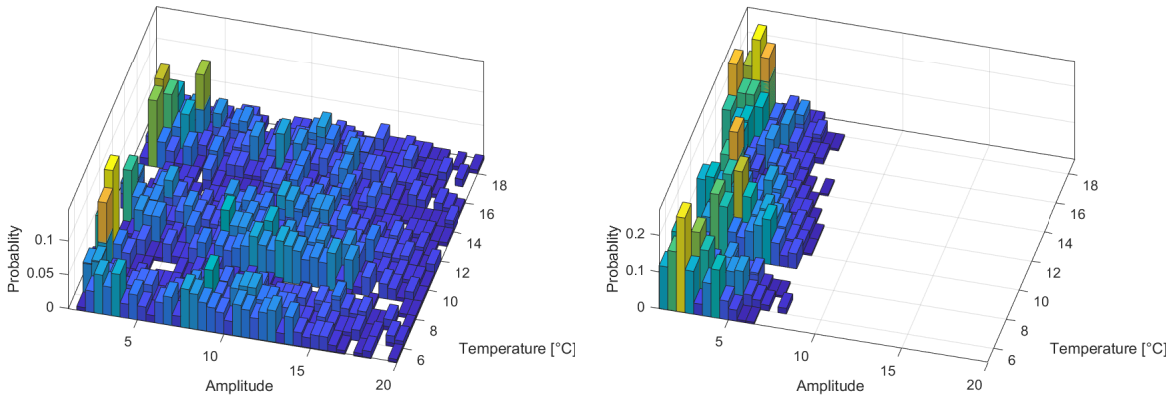


Figure 7.12: Distribution of the amplitude value of the natural frequency for each sample. (a) Raw truncated data set (b) Low-rank data set. Data from sensor 17.

It is more challenging to capture the change in the natural frequency at 10 Hz due to temperature compared to the higher natural frequency at 190 Hz. This could be due to the temperature having less effect on the lower natural frequencies. Another reason could be that the frequency resolution of 0.125 Hz is insufficient to capture the change in the natural frequency due to temperature. The frequency resolution needs to be 0.02 Hz for the 10 Hz natural frequency to have the same level of accuracy as in the analysis of the 190 Hz natural frequency with a frequency resolution of 0.5 Hz. However, this frequency resolution would result in a time window of 40 seconds for each sample, which would produce a few samples for analysis. This frequency resolution would be impractical with the data size used in this analysis, which is only two days. Nevertheless, even if a frequency resolution of 0.125 Hz, there is some indication of a negative linear correlation between temperature and the location of the natural frequency at 10 Hz.

7.3. Full vector analysis

7.3.1. Horizontal sensor 31

In the previous investigation, the effect of environmental variability on the vibrational data was explored by looking only at a single data point for each sample frequency vector. By only looking at a single natural frequency and how it changes with temperature, it was possible to reduce the dimensionality of the problem and see a relationship between temperature change and frequency shift. Another approach is to follow the same procedure as in the damage detection from the first measurement campaign, as was outlined at the beginning of chapter 6. That is, use principal component analysis (PCA) and only look at data the in the relevant principal components but now the different structural states are connected to the different environmental conditions instead of damage. The whole frequency vector is considered with PCA, so the information of the entire dynamic system is present, not just one natural frequency, as was the case in the previous investigation.

However, if the raw data set were used, the operational variability due to the traffic would still be present in the data. Here, Robust Principal Component Analysis (rPCA) can be used to reduce the operational variability for samples within the same temperature interval and help discover any patterns or relation-

ships concerning the environmental variability. Note that there is an uneven amount of samples in each of the temperature intervals, which means the performance of the rPCA can vary between temperature intervals. The ability of the rPCA to extract the coherent patterns that correspond to the dynamic system can differ between temperature intervals. After the application of the rPCA, the PCA is applied to data from multiple temperature intervals. There needs to be an even representation of the data from the different temperature intervals to ensure that the PCA is capturing the correlation between the data from the different temperature intervals and not the correlation of the data within each temperature interval. This means that the same amount of low-rank samples from the rPCA are taken from each temperature interval or 200 for the upcoming analyses.

In the damage detection of the first measurement campaign, there was a comparison between two states, damaged and undamaged. The same can be done for environmental detection, that is, comparing samples from different temperatures to see if there is any statistical difference between the two groups. However, the environmental state of the bridge is not binary. The structure undergoes a wide range of temperatures. Therefore, instead of comparing just two temperature intervals, several or all of the temperature intervals can be compared to see if there is any separation of the samples of the different temperature intervals in the PC space.

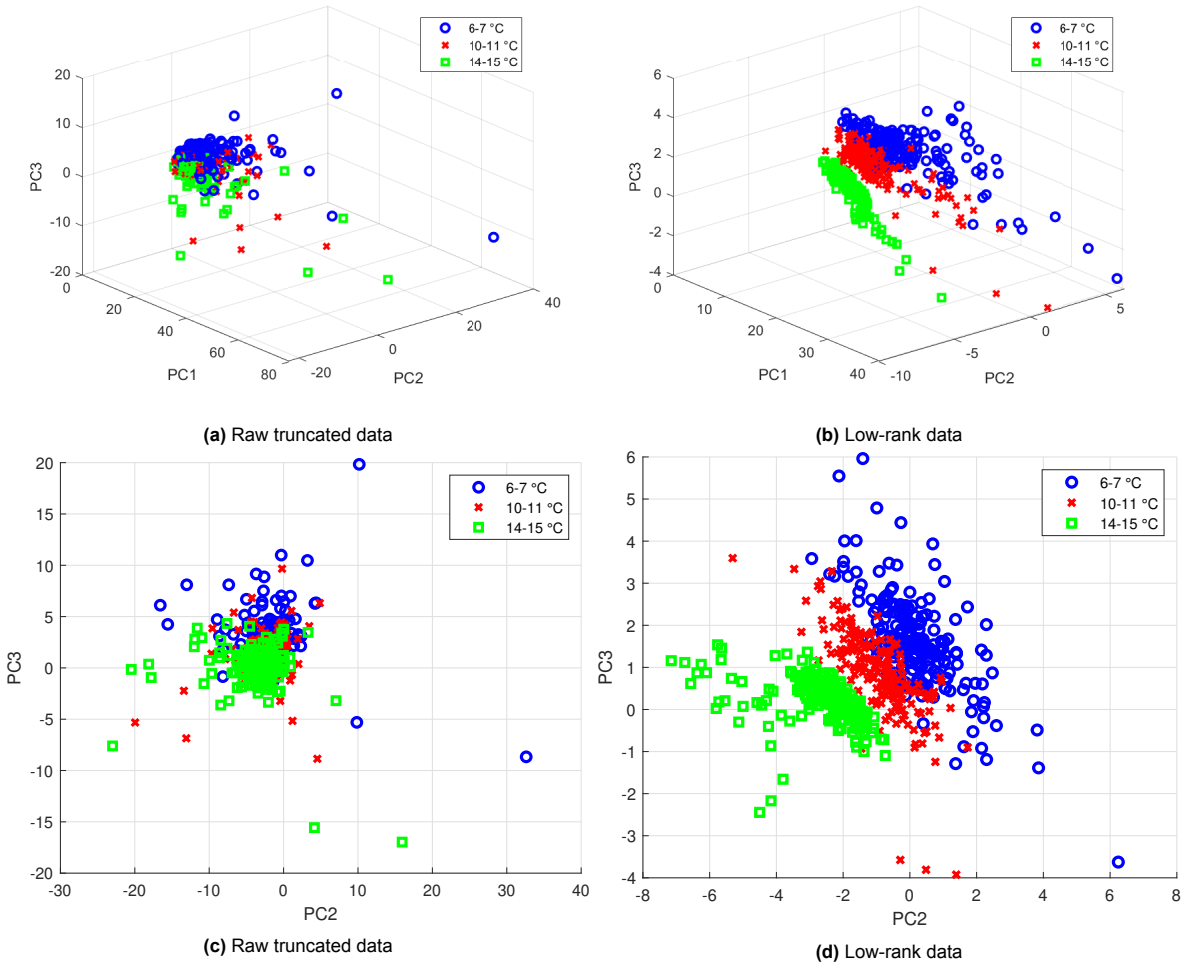
Also, in the damage detection in the previous measurement campaign, there was a considerable difference between the damaged and undamaged signals, which made the rPCA unnecessary in some cases to measure the difference. Entire natural frequencies shifted considerably in the frequency domain between the damaged and undamaged signals. The same degree of change is not present in the mean frequency vectors of the different temperature intervals, as seen in figure 7.4. It should be more challenging to distinguish between the dynamic properties of the system under different temperatures compared to the damage detection in the previous chapter.

Three classes, or in this case, samples from three temperature intervals, will be compared. Three comparison scenarios based on how close the temperature intervals are to each are considered. These are a 4°C difference, 2°C and finally, a single centigrade. It should be more challenging to distinguish between the temperature groups the closer they are together. This process will be done for both low and high-frequency sensors.

4°C Difference

Figure 7.13 shows the PC score of 600 samples, 200 from each temperature interval, in the first three PCs. The data is from sensor 31, which is located on one of the stiffeners and measures the horizontal acceleration. The three temperature intervals are 6-7 °C, 10-11 °C and 14-15 °C. A four-degree Celcius difference is between the samples. Temperature measurements are from temperature sensor number 6 at the stiffener. Two hundred samples are selected to provide an even representation for all temperature intervals. The figure shows the PC score of the samples for both the raw truncated data set and the low-rank data set. With the raw truncated data set, there seems to be no particular division of the samples for the different temperature groups. Most samples of the temperature groups intersect in the PC space, but there appears to be a correlation between the temperature groups and PC3. The lower the temperature, the higher the score of the samples. Similar to the previous measurement campaign, the PCA of the low-rank data performs much better. There is a clear division between the temperature groups. Figures 7.13c and 7.13d show only the score of the samples in principal components 2 and 3. Here the division of the data according to the temperature groups is evident for the low-rank data. There also appears to be a negative linear correlation between temperature with these two PCs.

Figure 7.15 shows a notched boxplot for PCs 2 and 3 for the different temperature intervals. This gives a better visualisation of how these PCs change with temperature. Both PCs show a negative correlation with temperature. Furthermore, the notches in the boxplot display the median between samples. If the notches of the different temperature groups do not overlap, then those groups have different medians at the 5 % significant levels. Comparing the medians in such a way is comparable to a visual hypothesis test. Finally, the total variance explained by the first PCs of the low-rank data is much higher than that of the raw truncated data set, which indicates that there is less variance in the low-rank. This can be



(a) Raw truncated data

(b) Low-rank data

(c) Raw truncated data

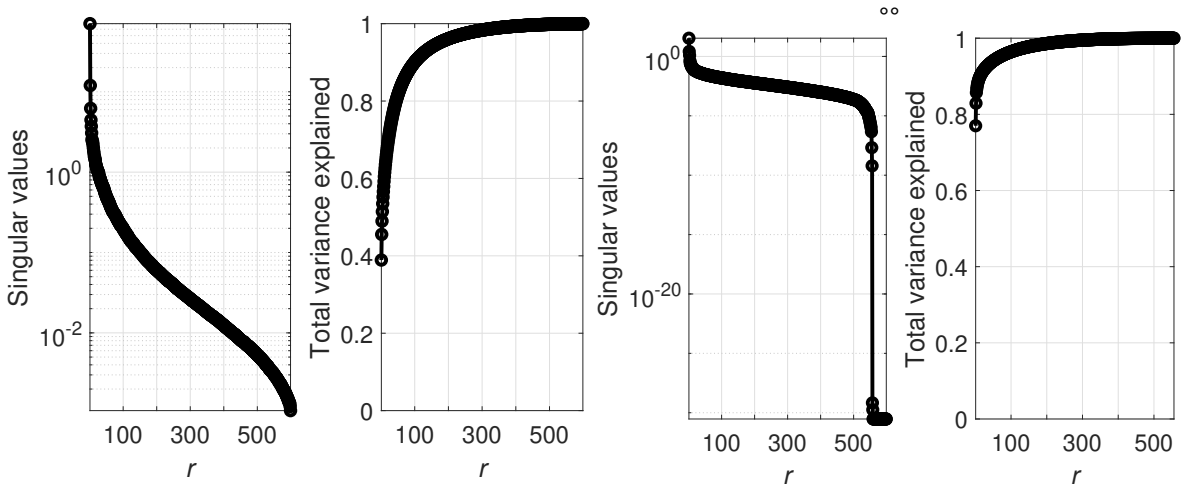
(d) Low-rank data

Figure 7.13: Principal component score of the samples in the PCs for data in three temperature intervals with a 4° difference. Data from sensor 31. (a) Raw truncated data. First three PCs. (b) Low-rank data. First three PCs. (c) Raw truncated data. PCs 2 and 3. (d) Low-rank data. PCs 2 and 3.

seen in figure 7.14. This is in agreement with the previous analysis when the rPCA was used in the damage detection.

The first three PCs of the low-rank data are shown in figure 7.16. These PCs explain most of the variance within the low-rank data. The first PC resembles that of the mean frequency vector of one of the temperature intervals of sensor 31 in figure 7.4. This PC is the dynamic properties of the system at this sensor location. However, it is PCs 2 and 3 that are of interest as these are the PCs in which the data is being separated. PCs 2 and 3 contain information about the differences between samples from the different temperature intervals. Both PCs exhibit a sawtooth pattern at the location of the higher natural frequencies at 190 Hz, 250 Hz and 460 Hz. The sawtooth pattern arises due to the shift in natural frequencies with temperature. In the previous analysis, it was discovered that the natural frequency at 190 Hz had a negative linear correlation with temperature. It seems reasonable that this is also the case for other the higher natural frequencies. Subtraction between two dynamic systems with shifted natural frequencies yields the sawtooth pattern. Figure 7.17 shows the subtraction of the mean frequency vector of the temperature interval 6-7 °C with the mean frequency vector of temperature interval 10-11 °C, which shows the saw tooth pattern. This figure shows the difference between these two intervals.

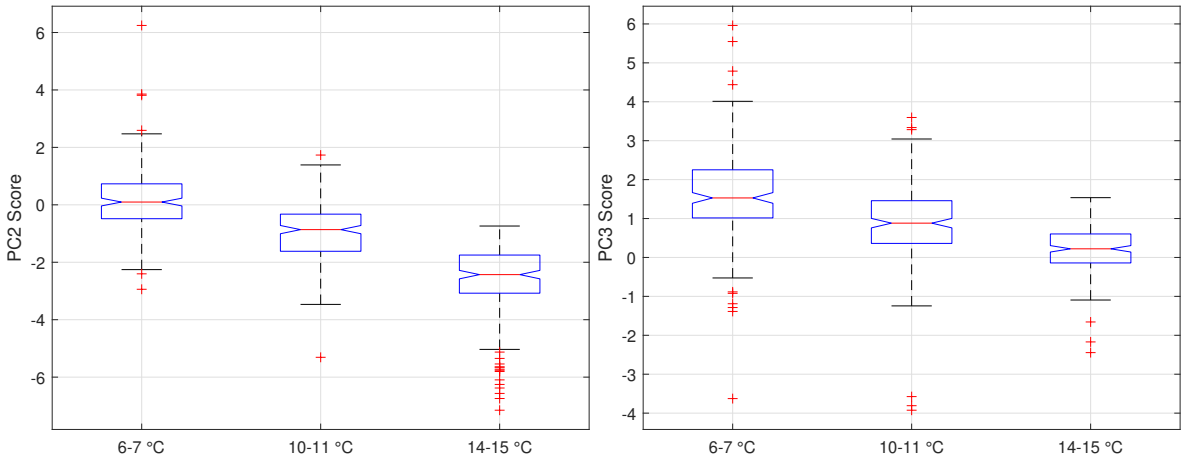
The anomalous ”spikes” are also present in PCs 2 and 3 at around 80 Hz. They were also present in the PCs in the damage detection in the first measurement. These anomalous ”spikes” change with



(a) Raw truncated data.

(b) Low-rank data.

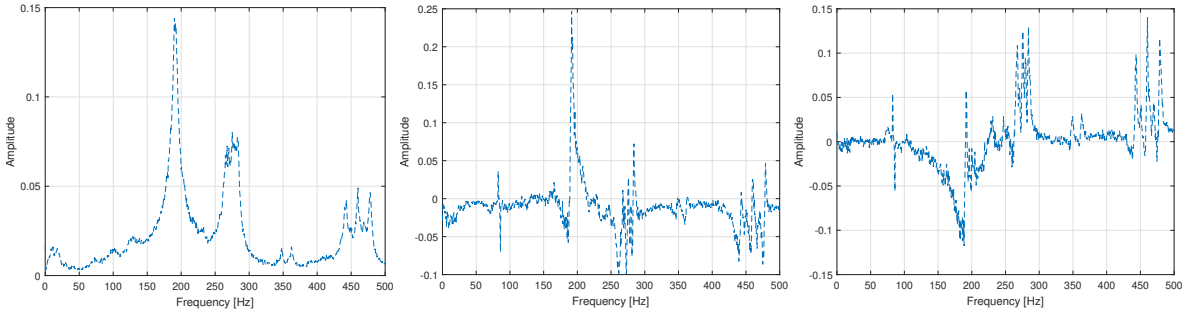
Figure 7.14: Singular values and total variance explained by the PCs.



(a) PC 2

(b) PC 3

Figure 7.15: Boxplox of the PC score of the low-rank data from temperature intervals with a 4°C difference. Data from sensor 31.



(a) First Principal Component

(b) Second Principal Component

(c) Third Principal Component

Figure 7.16: First three Principal Components. Data from three temperature intervals with a 4 °C difference. Low-rank data set

temperature but are not part of the dynamic properties but rather due to the sensors and auxiliary equipment. The location of the anomalous ”spikes” is the same within each temperature interval but different between intervals. The PCs capture this difference between the data from the different temperature intervals as the signs of the ”spikes” is opposite in these two PCs. A more thorough discussion of the anomalous ”spikes” can be found in appendix A.

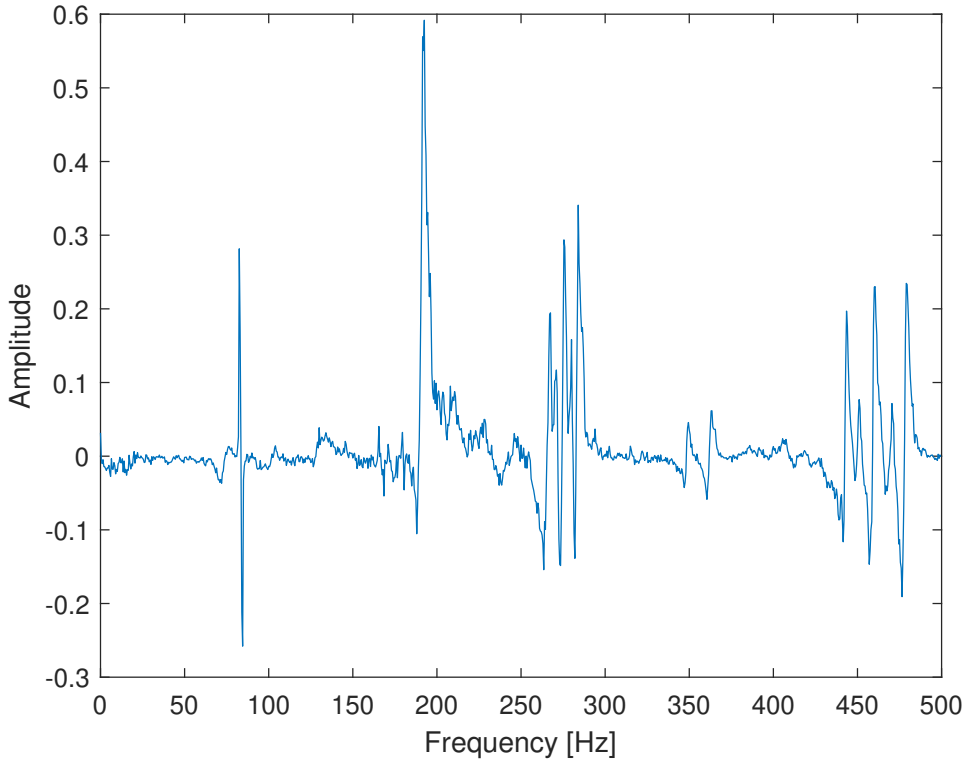


Figure 7.17: Difference in mean frequency vectors of samples from temperature interval 6-7 °C and 10-11°C. Raw truncated data from sensor 31. Resembles the patterns seen in PCs 2 and 3.

PCs 2 and 3 contain the underlying differences between data from the different temperature intervals. The different combinations of PCs 2 and 3 are enough to separate data according to their temperature group. It seems that this difference is connected to the differences in dynamic properties but also connected to the anomalous ”spikes”.

2°C Difference

Now that it has been established that it is possible to distinguish between structural systems that have a 4-degree Celcius difference, the next step is then to lower the difference and see if there is still a division or a relationship between the temperature groups in the PC space. The temperature difference between the temperature groups is now 2°C and the temperature intervals under investigation are now 10-11 °C, 12-13 °C and 14-15 °C. Temperature sensor 6 is used again for this analysis. Since the results of the raw truncated data were not satisfactory in the previous analysis, it is a fair assumption that the quality will not improve by lowering the temperature difference and thereby increasing the difficulty of the problem. Thus only the low-rank of the rPCA is considered in the remaining analysis.

Figure 7.18 shows the PC score of the samples in the first PCs for three temperature groups with a 2°C difference. Again 200 samples for each temperature interval are used for analysis. Here the division of the temperature groups in the PC space is worse than before. The two temperature groups of 10-11 °C and 12-13 °C occupy most of the same area in the PC space. On the other hand, the 14-15 °C temperature group is completely separated from the other two temperature groups. The data is from two separate days and the temperature on these days only intersects in the temperature interval of 10-13 °C. Thus the two lower temperature intervals of 10-11 °C and 12-13°C consist of data from both 18th of October and the 3rd of November, while the temperature interval 14-15°C consists only of data from the 18th of October. This can be seen by looking at the distribution of samples in figure 7.1.

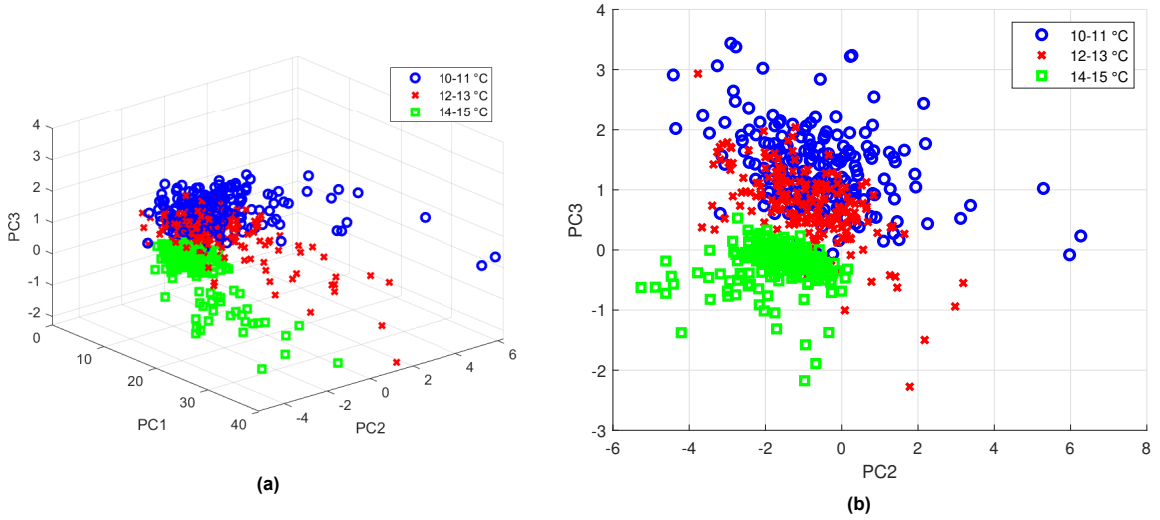
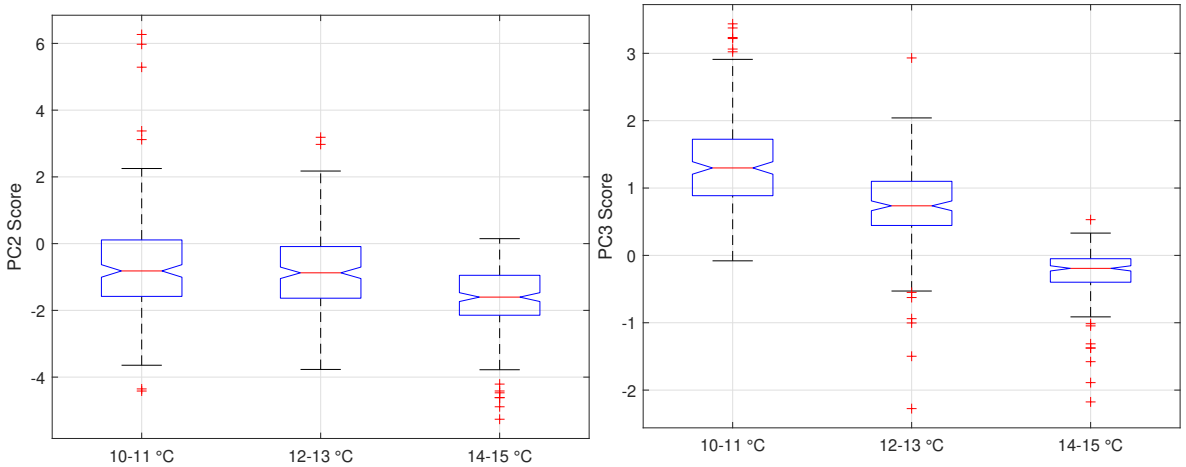


Figure 7.18: Principal component score of the samples in the PCs for data in three temperature intervals with a 2°C difference. The 10-11°C and 12-13°C temperature intervals contain data from both measurement days but the temperature interval 14-15°C only contains data from 18th of October. Data from sensor 31. (a) Low-rank data. First three PCs. (b) Low-rank data. PCs 2 and 3.

A possible explanation for this behaviour is that the acceleration samples are grouped into these temperature intervals by a single temperature sensor. This single temperature sensor does not capture the full temperature profile of the bridge. So while there is a 2°C difference between the temperature intervals according to this sensor, the actual temperature profile of the bridge could be potentially different between the different days. This, in turn, means that there is a difference in the dynamic properties, which is not explained by this single temperature sensor. The ”extra” difference in the dynamic properties between the two different days could also be due to changes in other operational and environmental conditions. For instance, it might have been raining on one of the days, which led to an increase in the mass of the structure, thereby altering the dynamic properties. During the measurement period, there was ongoing maintenance of the drawbridge. The position of concrete barriers on the road might have been moved between the days due to maintenance, thus altering the dynamic properties. There is also the consideration that the relationship between the dynamic properties and temperature is nonlinear. However, at least it seems that a single temperature sensor is not capturing the cause of this difference between the days.

Even though it is hard to distinguish between the two lower temperature groups, there still appears to be a correlation between the temperature and the 2nd and 3rd PC scores of the samples. Figure 7.19 shows a notched boxplot of the samples for the different temperature intervals. There is no distinct difference in the score of the samples in the 2nd PC for the two lower temperature intervals. However, there is a change in the median of the highest temperature interval. For the score in the 3rd PC, there is a negative shift in the median of all the temperature groups. This would suggest a negative correlation between temperature and this PC.

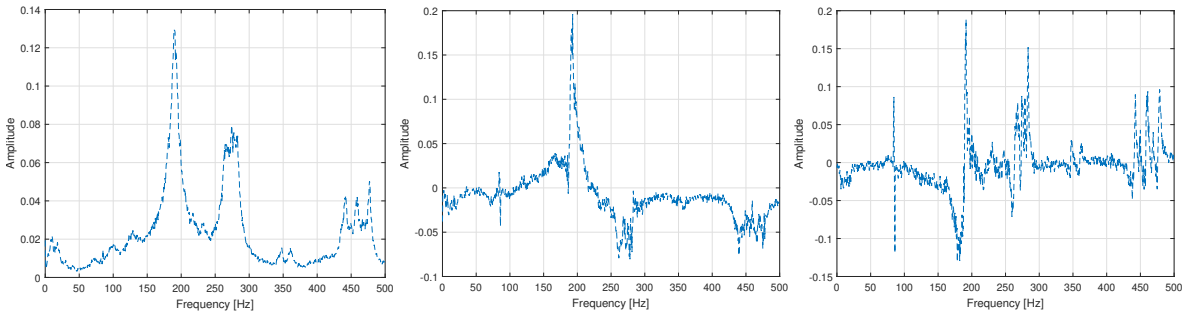


(a)

(b) Boxplox of the PC score of the low-rank data from temperature intervals with a 2°C difference. Data from sensor 31.

Figure 7.19: Boxplox of the PC score of the low-rank data from temperature intervals with a 2°C difference. The 10-11°C and 12-13°C temperature intervals contain data from both measurement days but the temperature interval 14-15°C only contains data from 18th of October. Data from sensor 31.

The first three PCs are shown in figure 7.20. The PCs are slightly different compared to the previous analysis. Here we see the sawtooth pattern again as the natural frequencies shift with temperature.



(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 7.20: First three Principal Component. Data from three temperature intervals with a 2 °C difference. Low-rank data set

1°C Difference

The final test is to see if it is still possible to distinguish between the temperature intervals that are 1°C from one another. In the previous analysis, it was discovered that there is a higher degree of difference in the dynamic properties between the samples from different days that is not accounted for by the temperature measurement of a single sensor. Figure 7.21 shows the PC score of samples from 3 temperature intervals. These are temperature intervals 11-12°C, 12-13°C and 13-14°C. The two lower temperature intervals contain data from both measurement days, while the highest temperature interval contains only data from the 18th of October. Even with 1°C difference between the temperature intervals, there is a separation between the samples according to their temperature in the PC space, which can be better seen with the boxplot in figure 7.22 with the data mostly separating in PC 3. Similar

to the 2°C comparison case, there is a better degree of separability of the data for the temperature interval that only has data from the 18th of October. The other two lower temperature intervals contain data from both days. Again there is a greater difference in the underlying data between the different days that is not explained by the measurement of a single sensor. This difference could be linked to different operational and environmental conditions between the days, which influences the dynamic properties of the system.

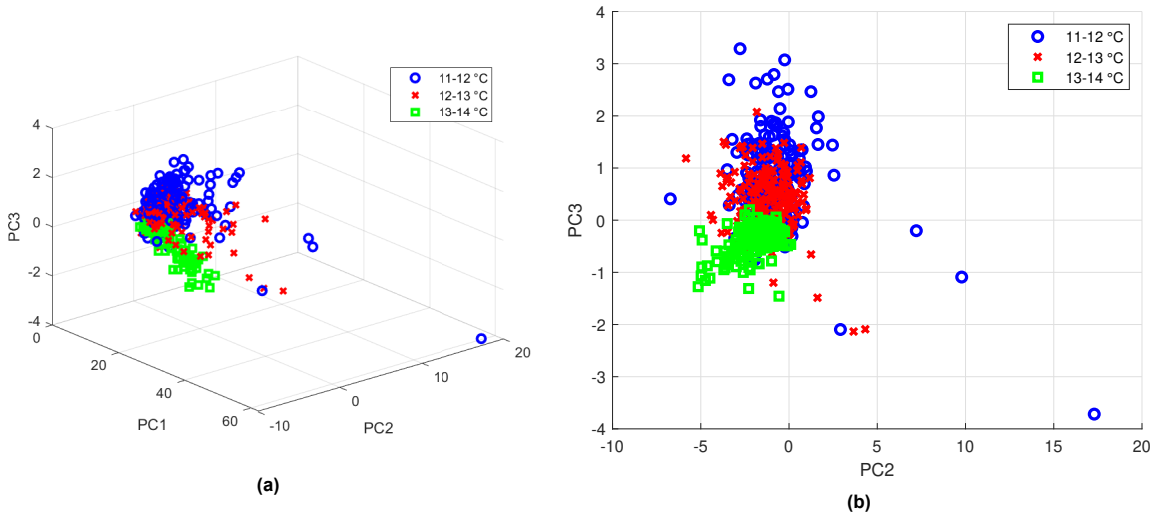


Figure 7.21: Principal component score of the samples in the PCs for data in three temperature intervals with a 1°C difference. The 11-12°C and 12-13°C temperature intervals contain data from both measurement days but the temperature interval 12-14°C only contains data from 18th of October. Data from sensor 31. (a) Low-rank data. First three PCs. (b) Low-rank data. PCs 2 and 3.

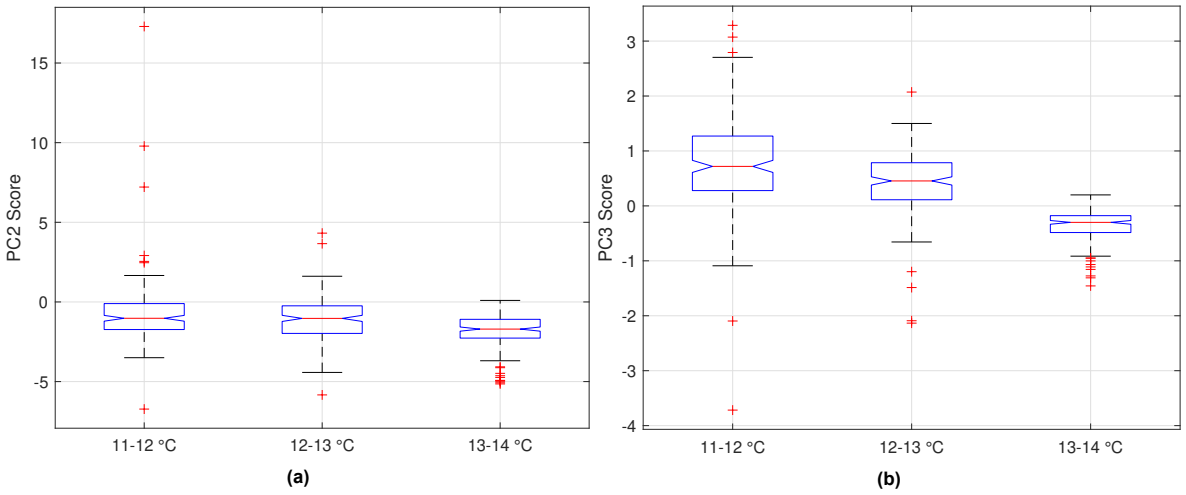


Figure 7.22: Boxplox of the PC score of the low-rank data from temperature intervals with a 1°C difference. The 11-12°C and 12-13°C temperature intervals contain data from both measurement days but the temperature interval 12-14°C only contains data from 18th of October. Data from sensor 31. (a) Low-rank data. PC 2 (b) Low-rank data. PC 3

However, when all the temperature intervals have data from both days, there is no particular separation of the data in the first three PCs according to temperature. This can be seen in figures 7.23 and in the boxplot in 7.24. Here temperature intervals 9-10 °C, 10-11 °C and 11-12 °C are used for the analysis. The first three PCs do not contain information to distinguish between the underlying dynamic properties at different temperatures. It seems that without the ”extra” difference between days, it is not possible to distinguish between data with a 1°C difference.

Fortunately, steps can be taken to improve the results. So far, only the default regularisation parameter $\lambda = 1/\sqrt{max(n,m)}$ has been used in the rPCA. By selecting a lower regularisation parameter, the

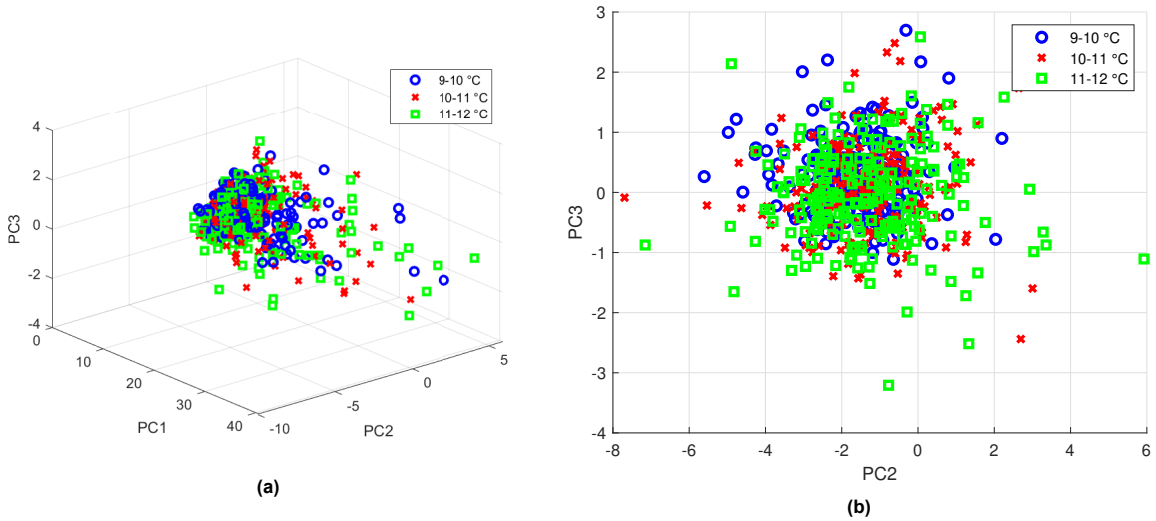


Figure 7.23: Principal component score of the samples in the PCs for data in three temperature intervals with a 1°C difference. The 9-10°C, 10-11°C and 11-12°C temperature intervals contain data from both measurement days. Data from sensor 31. (a) Low-rank data. First three PCs. (b) Low-rank data. PCs 2 and 3.

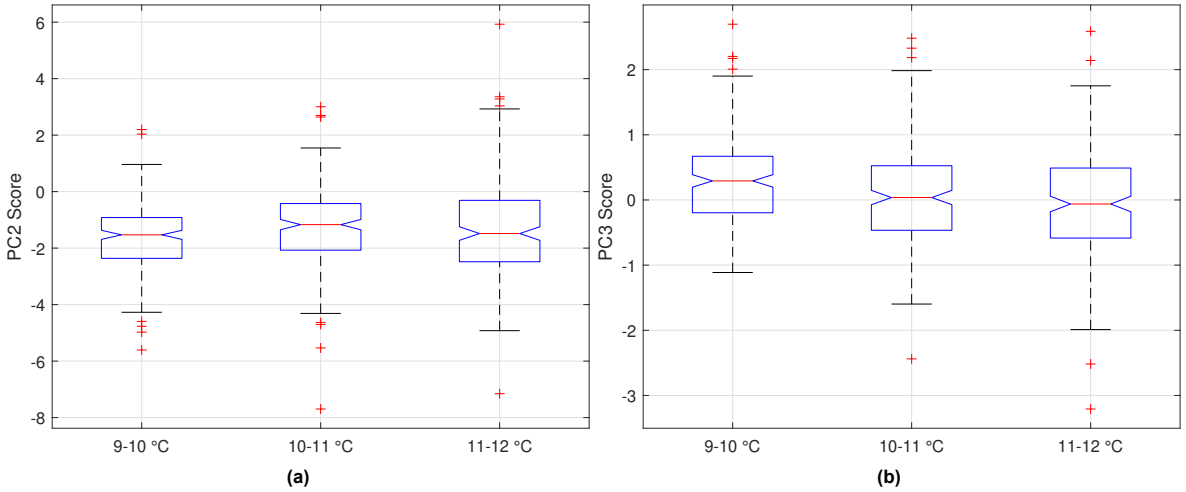


Figure 7.24: Boxplox of the PC score of the low-rank data from temperature intervals with a 1°C difference. The 9-10°C, 10-11°C and 11-12°C temperature intervals contain data from both measurement days. Data from sensor 31. (a) Low-rank data. PC 2 (b) Low-rank data. PCs 3

filtering of the rPCA becomes more aggressive, as was discussed in chapter 4.3. More of the noise within the data will be discarded into the sparse component of the rPCA, and the low-rank matrix of the rPCA will be reconstructed from even fewer PCs which will result in an even lower rank of the low-rank. This essentially means that the low-rank data will more closely resemble the coherent structures in the data, that is, the dynamic properties. However, lowering the regularisation parameter too much runs the risk of the rPCA identifying coherent patterns in the data as outliers [38]. An arbitrary value of $\lambda = 1/\sqrt{2 \cdot max(n,m)}$ will be utilised for the next analysis.

The previous analysis can now be repeated but now with the new lower regularisation parameter. Figure 7.25 shows the PC score of the samples for the low-rank data with the aforementioned new regularisation parameter. There is a substantial improvement compared to figure 7.23 with the default regularisation parameter. Now the data is separated in the first three PCs according to temperature, and there is a negative linear correlation between temperature and the score of the samples in the 3rd PC. The data is mostly being separated in the 3rd PC. Figure 7.27 shows the first three PCs. PC 3 has a sawtooth pattern at the natural frequencies, similar to what was seen in the 4°C difference case. Even at 1°C difference, the frequencies are shifting, and the rPCA can recover the coherent patterns

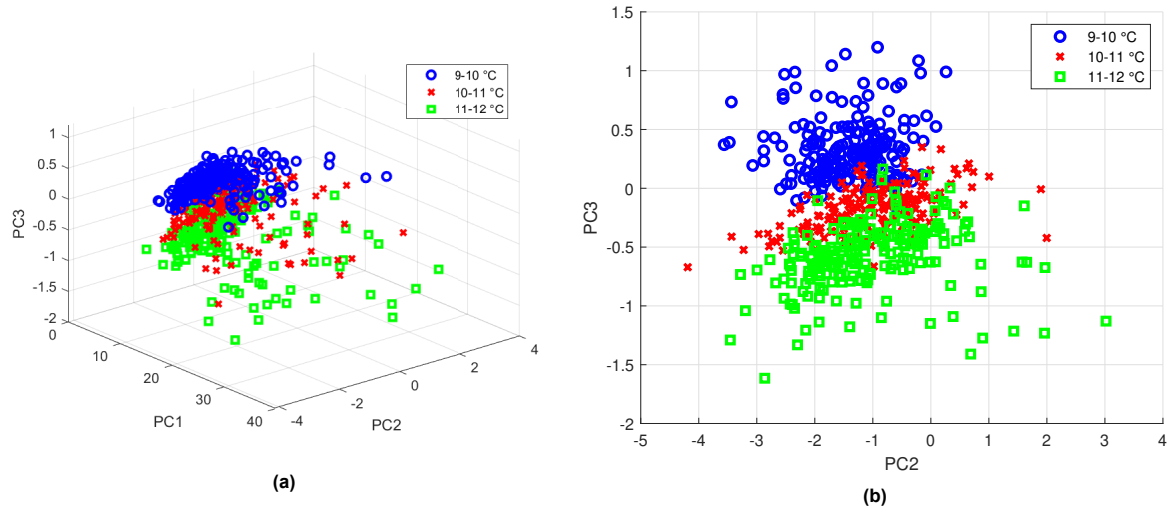


Figure 7.25: Principal component score of the samples in the PCs for Low-rank data with a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$ in three temperature intervals with a 1°C difference. The 9-10°C, 10-11°C and 11-12°C temperature intervals contain data from both measurement days. Data from sensor 31. (a) Low-rank data. First three PCs. (b) Low-rank data. PCs 2 and 3.

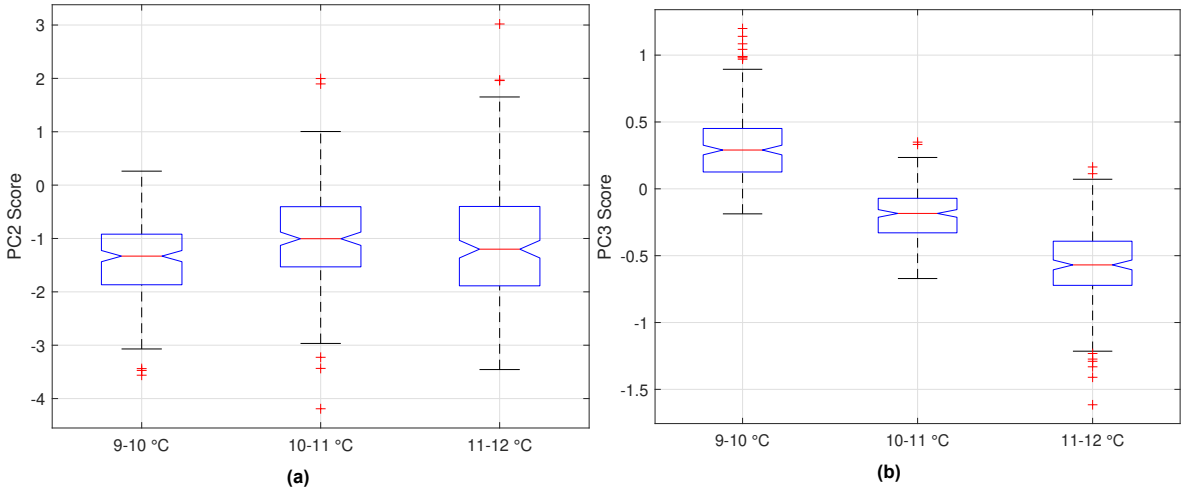
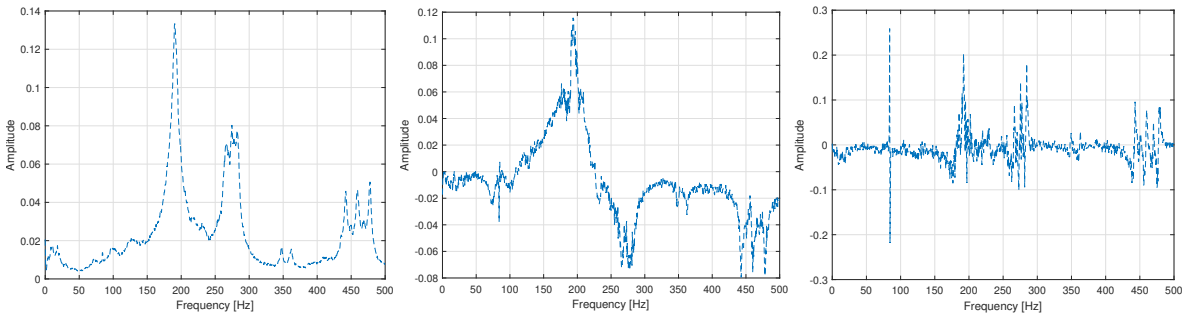


Figure 7.26: Boxplox of the PC score of the low-rank data with a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$ from temperature intervals with a 1°C difference. The 9-10°C, 10-11°C and 11-12°C temperature intervals contain data from both measurement days. Data from sensor 31. (a) Low-rank data. PC 2 (b) Low-rank data. PC3

at each corresponding temperature when a lower regularisation parameter is used.



(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 7.27: First three Principal Component. Low-rank data set with a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$

Points of maximum difference

In the damage detection with the data from the first measurement campaign, the sparse sensor placement optimisation for classification (SPPOC) [7] was used to find the locations in the frequency vector that contained the most significant difference between the signals in the undamaged and damaged areas. The SPPOC can be used again but now in relation to the temperature change to find the frequency locations that contain the most discriminating information between the data in the different temperature intervals. These frequency locations should be the natural frequencies that change the most with temperature.

Data from the three temperature intervals in the previous analysis are used in the SPPOC method. These are temperature intervals 9-10°C, 10-11°C and 11-12°C. Temperature sensor 6 is used again to categorise the samples. The low-rank data with the lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$ will be used with the SSPOC method, and 200 samples are taken from each temperature interval. There is a requirement with the SPPOC that the data is embedded into a coordinate system that allows for classification algorithms to work, in this case, the Linear Discriminant analysis (LDA). This means that the data has to cluster in the subspace of the PCs according to temperature, which is the case as seen in figure 7.26. There are three categories of the data or three temperature intervals, and the number of features is three, corresponding to the first three PCs. This gives a total of six sparse sensor locations.

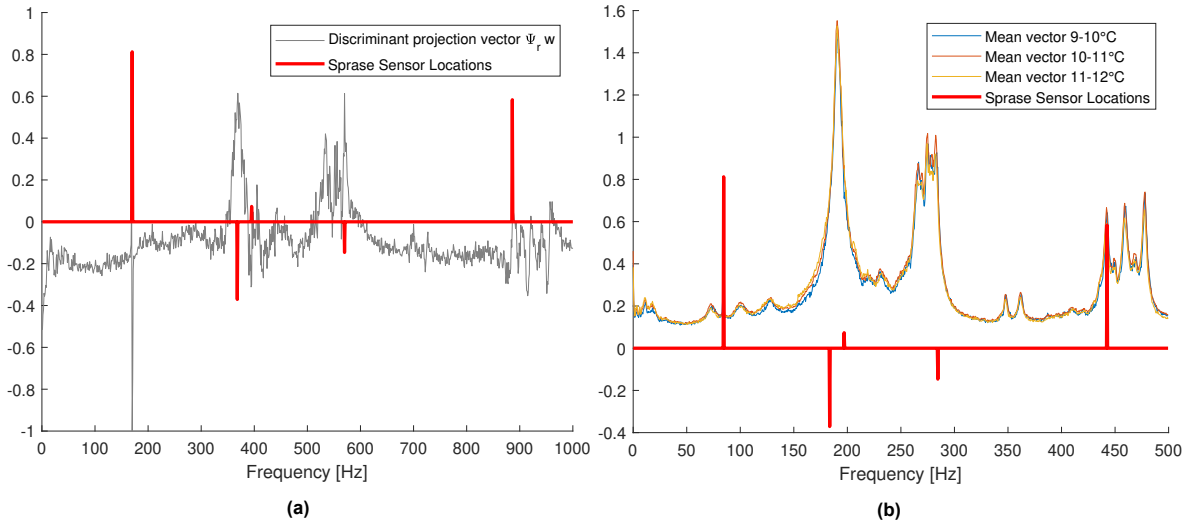


Figure 7.28: Sparse sensor locations for Low-rank data with a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. Data from random 200 samples from temperature intervals 9-10°C, 10-11°C and 11-12°C (a) Discriminant projection vector with the sparse sensor locations marked in red. (b) Mean vectors of the three temperature intervals.

Figure 7.28 shows the sparse sensor locations that contain the most discriminatory information between the data in the different temperature intervals. Here the calculated sensor locations are at the natural frequencies but also at the anomalous ”spike” at around 100 Hz. This is similar to what was seen when the SSPOC was used in damage detection. The anomalous ”spike” is not part of the dynamics of the structure but holds the most significant difference in the frequency vector connected to the temperature change. In fact, these anomalous ”spikes” have a clear positive correlation with temperature, which can be seen in appendix A. These anomalous ”spikes” add a false positive bias to the data as they are not part of the system dynamics and should be filtered out of the signals.

The other sparse sensor locations are at the natural frequencies. There is a sparse sensor location at each of the ”major” natural frequencies. The location that contains the largest difference, after the anomalous ”spike”, is the natural frequency at 440 Hz. A possible reason why these higher natural frequencies contain the largest difference between the vibration data at different temperatures is that the absolute frequency shift of these higher natural frequencies is the largest. Another thing to consider is that the SPPOC depends on the inputted data. Therefore, any variation in the input can produce different sparse sensor locations. Consider figure 7.29, which has a different set of 200 random samples

from each of the temperature intervals. Here, there are new sparse sensor locations, such as the natural frequency at 480 Hz. However, most of the sensor locations remain the same. A better approach here would be to conduct multiple iterations of the SPPOC using random samples and construct a distribution of the sensor locations as was done in [7]

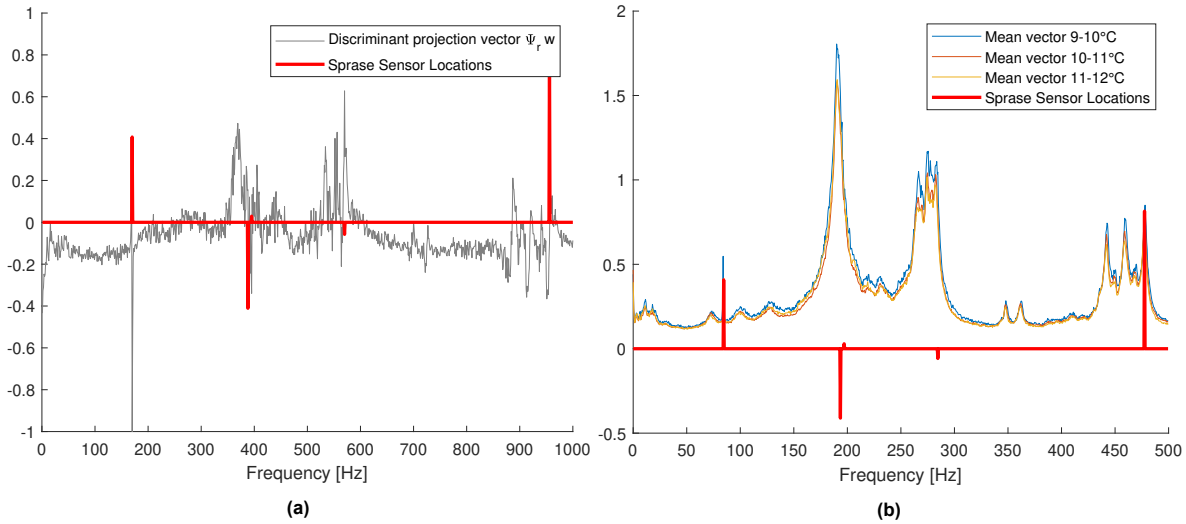


Figure 7.29: Sparse sensor locations for Low-rank data with a lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. Data from 200 random samples from temperature intervals 9-10°C, 10-11°C and 11-12°C. The random samples are different from figure 7.29 (a) Discriminant projection vector with the sparse sensor locations marked in red. (b) Mean vectors of the three temperature intervals.

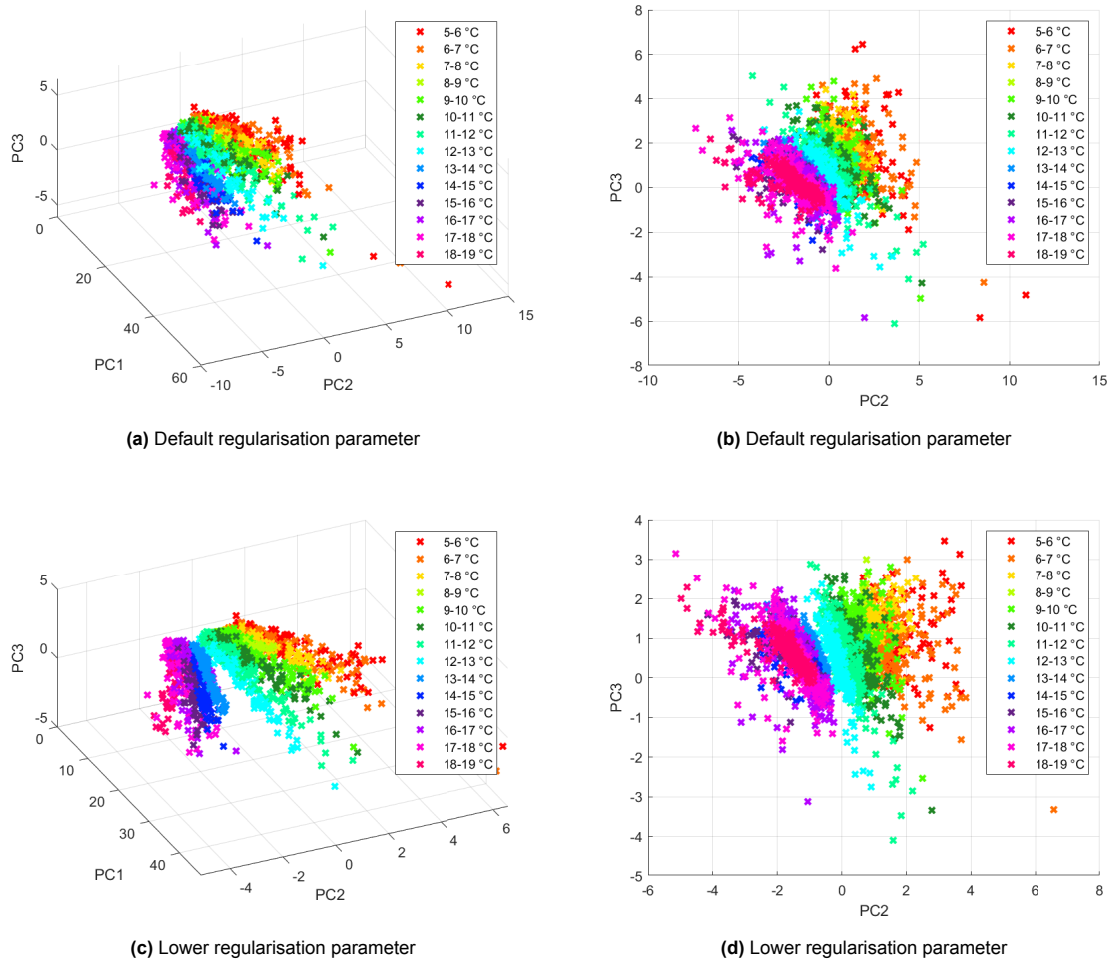
1°C Many intervals

The previous analyses can be repeated, but instead of just inputting data from three temperature intervals into the (regular) PCA, data from all available temperature intervals can be used in the PCA. The PCA would then have information about the system properties under all available measured temperatures, given that the rPCA can find the dynamic properties for each temperature interval. Two data sets are considered with different regularisation parameters in the rPCA. One data set with the default regularisation parameter and another with the lower regularisation parameter of $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. This allows for the comparison of their performance.

As was discussed at the beginning of the chapter, there is an uneven distribution of samples for the different temperature intervals, as seen in figure 7.1. For example, some temperature intervals have over 2000 samples, while other intervals have only 148 samples. This means that the performance of the rPCA in extracting the coherent patterns of each temperature interval can vary between the different intervals.

After applying the rPCA to the data in each of the temperature intervals, 200 random low-rank samples are taken from each temperature interval, and the PCA is then performed. There is an exception to this rule for temperature interval 8-9°C, which only has 148 samples. In this case, all samples are used from this temperature interval. This leads to a total of 2748 samples in the PCA. Figure 7.30 shows the low-rank data in the first three PCs for all the temperature intervals. There is an intersection of the data from different temperature intervals in the PC space. This makes it difficult to distinguish between data from different temperature intervals based on the PC score. However, data from each temperature interval is surrounded by data in corresponding adjacent temperature intervals in the PC space. There is also a correlation between the temperature and the score of the samples in the PCs. For the case with the default regularisation parameter, there is a correlation between temperature with the 2nd and 3rd PCs. While in the case of the lower regularisation parameter, the correlation is mainly with the 2nd PC. This correlation is better seen when viewing the boxplot in figure 7.32.

The boxplot also reveals that there is a considerable shift in the score of the samples in the 2nd PC at



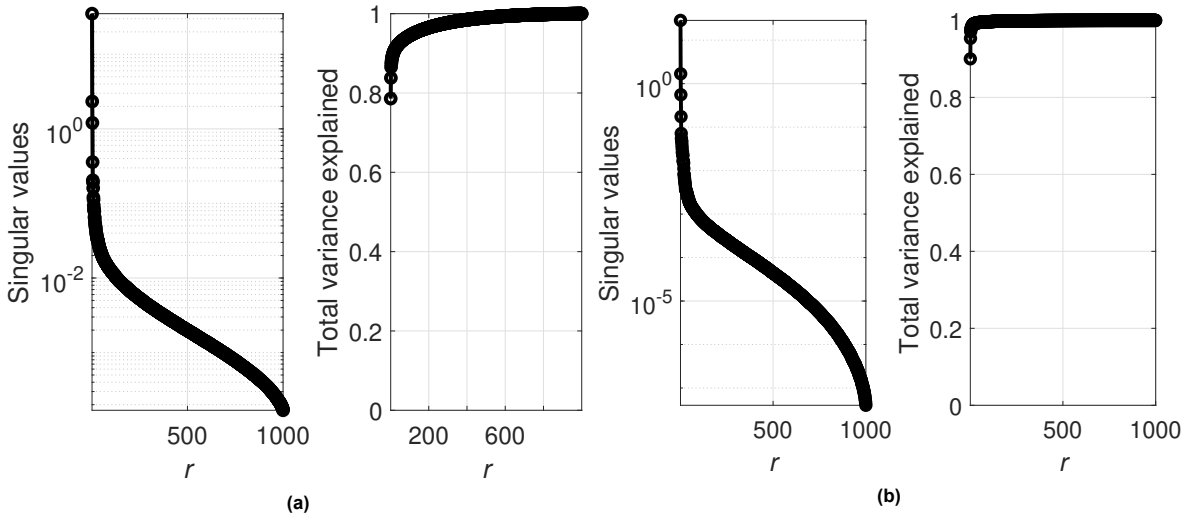
(a) Default regularisation parameter

(b) Default regularisation parameter

(c) Lower regularisation parameter

(d) Lower regularisation parameter

Figure 7.30: Principal component score of the samples in the PCs for data in All temperature intervals Data from sensor 31. (a) Low-rank data with default regularisation parameter. First three PCs. (b) Low-rank data with default regularisation parameter. PCs 2 and 3. (c) Low-rank data with lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. First three Pcs. (d) Low-rank data with lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$. PCs 2 and 3.



(a) **(b)**

Figure 7.31: Singular values and total variance explained by the PCs. (a) Low-rank data with default regularisation parameter. (b) Low-rank data with lower regularisation parameter $\lambda = 1/\sqrt{2 \cdot max(n,m)}$.

13°C. This shift is the largest between two adjacent temperature intervals, and the shift is maintained for all temperature intervals above 13°C. At temperatures above 13°C, the data is only from a single day, the 18th of October. This is similar to what was encountered in the previous analysis when there was a greater difference in the PC score of the samples between days than in temperature intervals within the same day. Furthermore, this shift can also be seen in a large gap between the samples at this temperature boundary in the PC space for the lower regularisation parameter case. However, this behaviour is not observed for the data at the other lower temperature boundary between days at 10 °C. The total variance explained by the 1st PC is higher for the low-rank data with the lower regularisation parameter compared to the low-rank data with the default regularisation parameter. This can be seen in figure 7.31. By lowering the regularisation parameter, there is more aggressive filtering of the rPCA, which means more noise is filtered out, which in turn means the major correlations in the data explain more of the variance within the data. Finally, the PCs for the two different regularisation cases are shown in figure 7.33. It is similar to the PCs of the previous analyses with a sawtooth pattern around the natural frequencies for the 2nd and 3rd PCs. The PCs are nearly the same for the two regularisation parameter cases.

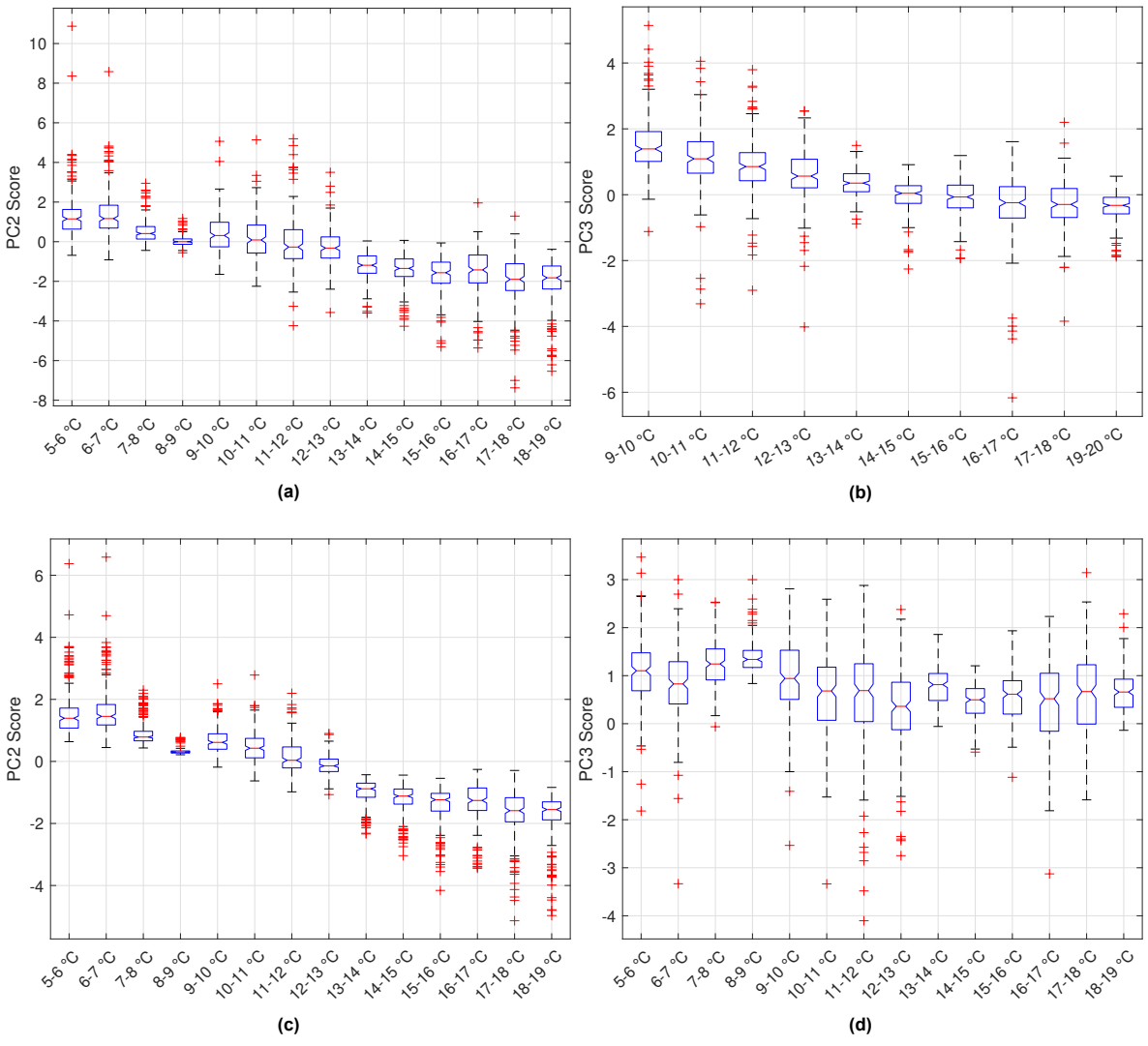
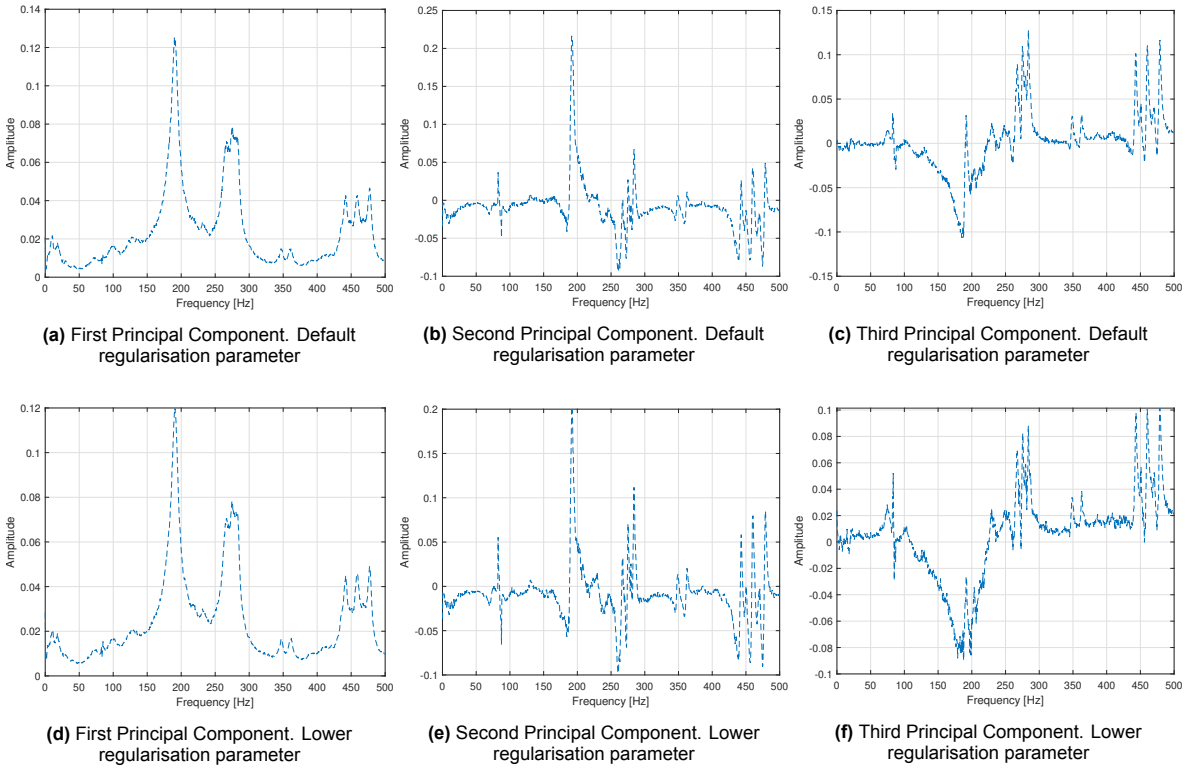


Figure 7.32: Boxplox of the PC score of the low-rank data from all temperature intervals. Data from sensor 31. (a) Low-rank data with default regularisation parameter. PC 2 (b) Low-rank data with default regularisation parameter. PC 3 (c) Low-rank data with lower regularisation parameter. PC 2 (d) Low-rank data with default regularisation parameter. PC 3

In the previous analysis with three data groups with a 1°C difference, the lower regularisation parameter performed better as the data was separated in the first three PCs. At the same time, there was no separation with the default regularisation parameter. However, in this case, it is more challenging to



(a) First Principal Component. Default regularisation parameter

(b) Second Principal Component. Default regularisation parameter

(c) Third Principal Component. Default regularisation parameter

(d) First Principal Component. Lower regularisation parameter

(e) Second Principal Component. Lower regularisation parameter

(f) Third Principal Component. Lower regularisation parameter

Figure 7.33: First three Principal Component for the different regularisation parameters. Low-rank data set

assess the performance difference between the two different regularisation parameter cases by only visually looking at the data in the PC space. The performance can be measured by comparing the accuracy of two supervised classification models. Two models are trained, one with the low-rank data with the default regularisation parameter and another with the low-rank data with the lower regularisation parameter. The aim of the models is to predict which classes the features belong to based on training data. In this problem, the classes are the different temperature intervals, and the features are the PC score of the samples in the first three PCs. The models themselves serve no purpose. There is no need to evaluate the temperature from vibration data. This information is already known for the entire vibration data. However, it is the performance of these models for the data with different regularisation parameters that is of interest. The ”performance” here is how the data is embedded into the PCs or how well separated the data from different temperature intervals is in the subspace of the PCs. The model with well-separated data should perform better.

Table 7.2: Comparison between accuracy of different classifiers between low-rank data with different regularisation parameters.

	Linear discriminant	Support Vector Machine (SVM)
Default Regularisation	31.1 %	42.9 %
Lower Regularisation %	41.0 %	72.5 %
Difference	9.9 %	29.6 %

Two supervised classification methods are considered, Linear discriminant analysis (LDA) and Support Vector Machines (SVM). These methods were outlined in chapters 4.6 and 4.7, respectively. For both methods, five-fold cross-validation is applied, and for both models, the linear version is used. The features are the PC score of the samples in the first three PCs. For the SVM, a One-vs-One multiclass method is used. The models are constructed with Matlab's Classification Learner App [1].

Table 7.2 shows the accuracy of the classification models for the two methods and the two different cases of the regularisation parameter. The accuracy of the models with the lower regularisation pa-

[1] Matlab Classification Learner - Source

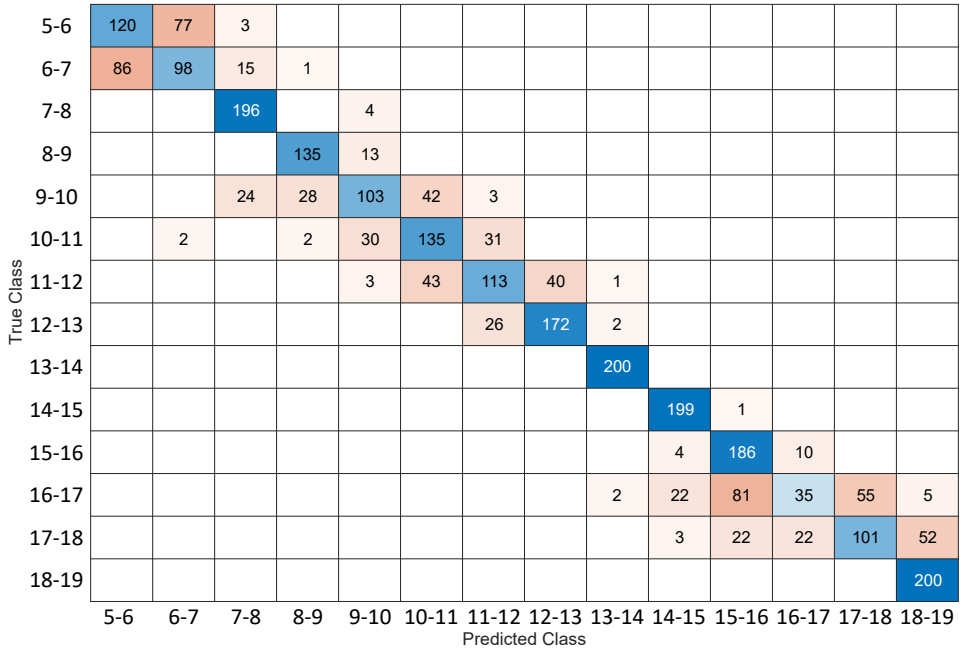


Figure 7.34: Confusion Matrix for the SVM model with the lower regularisation parameter

rameter is higher for both the Linear discriminant and the SVM. There is nearly a double increase in accuracy for the SVM model. This would indicate that there is a higher degree of separability between data from each of the temperature intervals for the data with the lower regularisation parameter. Figure 7.34 shows the confusion matrix of the SVM model with the lower regularisation parameter data. The error mainly occurs between adjacent bins. One of the highest locations of accuracy is at the boundary between days at 13°C. There is a high degree of difference between the data at that point.

Interestingly there is hardly any difference in the PCs between the two regularisation cases, but there is a substantial amount of difference in the accuracy of the classification models. A likely explanation is that with a lower regularisation parameter, more of the noise within the data is discarded into the sparse matrix of the rPCA. As more noise in the data has been removed, the main correlations in the data will explain more of the variance in the data as can be seen in figure 7.31. The samples are characterised by less noise and thus correspond to the main correlations in the data seen in the PCs.

This analysis shows that by changing the value of the regularisation parameter, more of the noise and outliers are removed from the data. This, in turn, makes it easier to distinguish between the underlying coherent structures at different temperatures, making it possible to differentiate between data with a 1°C difference.

7.3.2. Vertical Sensor 17

The next task is to evaluate how the low-frequency content of the accelerometers that measure the vertical acceleration of the bridge changes with temperature. The vibration data from accelerometer number 17 will now be analysed. As previously mentioned, there are several challenges when working with low-frequency content. A single centigrade temperature difference leads to a percentage shift of the natural frequencies of a system [41]. The absolute value of this shift is less for the lower natural frequencies than for the higher natural frequencies. Thus a higher frequency resolution is required to be able to detect the shift of the lower frequencies compared to the higher frequencies for the same level of temperature difference. Increasing the frequency resolution of the Fourier transform means taking a larger time window in the time domain for the samples. The samples are taken around the excitation due to traffic in the time domain, and the excitation due to the traffic can be densely packed

in the time domain. Increasing the time window ”merges” many of the samples together. This lowers the number of available samples for analysis, which is detrimental to the effectiveness of the rPCA.

For the low-frequency content, a frequency resolution of 0.125 Hz is selected, which translates to a time window of 8s. The amount of samples available for analysis with this frequency resolution is shown in figure 7.9a. It is considerably less than the samples when the frequency resolution was taken as 0.5 Hz, as seen in figure 7.1.

Most of the dynamic amplification in these sensors occurs below 100 Hz. As such, it seems reasonable to limit the analysis to this frequency band and discard the higher frequencies. It is not desirable to categorise samples based on the high-frequency content that is not related to the actual dynamics of the system. Therefore, only the frequency band below 100 Hz is used for analysis.

The same analysis is carried out as was done for the horizontal sensor 31. The rPCA is applied to all the data for each temperature interval. Three temperature intervals with a 5°C difference are selected with 200 random low-rank samples taken from each temperature interval, and the PCA is then computed. The three temperature intervals are 6-7 °C, 11-12 °C and 16-17 °C. There was no particular separation of the samples in the subspace of the first three PCs according to temperature. The analysis was then repeated with a lower regularisation parameter of $\lambda = 1/\sqrt{2 \cdot max(n, m)}$. However, this did not significantly improve the results. The only visible separation of the data in the subspace of the first three PCs was between data from the different days. This is similar to what was seen in the analysis of the horizontal accelerometer.

There is barely any separation between data from the different temperature intervals. This will hardly improve when the temperature intervals are more closely spaced, which is a more challenging problem. Thus no more analysis will be conducted with the data from the low-frequency vertical sensors.

7.4. Conclusion

The goal of the analysis of the vibration data from the 2nd measurement campaign was to see if it was possible to distinguish between vibration data at different temperatures. Two analyses were conducted. First, a single resonance peak was tracked in relation to temperature. It revealed a negative linear correlation between temperature and the frequency value of a high natural frequency, which is in agreement with what is seen in the literature as was discussed in chapter 2. There was also a negative linear correlation between temperature and the amplitude of the natural frequency. It could be an indication that the damping properties of the structure change with temperature. Another possible reason is that the position of the mode shape that corresponds to the natural frequency is shifting within the structure with temperature. This would mean the sensor is capturing a different part of the mode shape with potentially different amplification.

In the second analysis, the entire frequency vector was investigated in relation to the temperature and PCA was used to distinguish between the samples belonging to the different temperature intervals. This proved to be a more challenging problem compared to the ”damage” detection with the data from the first measurement campaign. The temperature has a minor effect on the system properties compared to the ”damage” present in the first measurement campaign. This means that the underlying dynamic properties at a given temperature interval are similar to those at adjacent temperature intervals. Any noise present in the signals will make it difficult to distinguish between the underlying dynamic properties at different temperatures. However, with the rPCA the noise within the vibration data can be reduced, and the underlying coherent structures can be obtained. By lowering the regularisation parameter in the rPCA, even more of the noise within the data can be filtered out, enabling the distinction between signals with a 1°C temperature difference which was otherwise impossible with the default regularisation parameter. The rPCA, with a lower regularisation parameter, is able to extract the underlying coherent patterns that correspond to the dynamic properties at each temperature interval.

The analysis of this chapter has revealed that temperature affects the system and that it is possible to extract the underlying coherent structures of the data at different temperatures. These coherent

structures correspond to the dynamic properties at different temperatures. Thus it is possible to obtain information about the system under different environmental conditions. A complete understanding of how the system responds to all possible environmental conditions is necessary for a robust structural health monitoring program.

This was only achieved for vibration data from a horizontal sensor with high-frequency content. It was more challenging to distinguish between the dynamic properties at different temperatures for low-frequency content from a vertical sensor. It was not possible to distinguish between signals with a temperature difference lower than 5°C. A possible reason for this is the frequency resolution, which needs to be higher for the low-frequency content to capture the shifts in the lower natural frequencies. However, increasing the frequency resolution comes at a cost by increasing the time window around the samples in the time domain. This, in turn, decreases the number of samples in the data, which has a negative effect on the performance of the rPCA. Increasing the frequency resolution also decreases the number of samples per length of the frequency vector. This could have adverse effects on the performance of the PCA and rPCA, as is discussed in chapter 8.

A frequency resolution of 0.125 Hz was used for the analysis of the low-frequency content of sensor 17. This might have been insufficient. A 0.5 Hz frequency resolution was used for the high-frequency content of sensor 31, which translates to a ratio of 380 between the natural frequency at 190 Hz and the frequency resolution. If the same ratio was used for the natural frequency at 10 Hz, then the frequency resolution would be 0.026 Hz. However, this frequency resolution would translate to a time window of around 40 seconds, which would yield few samples in a 15-minute time frame. Using this frequency resolution would be impractical with the amount of data available for this research.

There was a higher degree of difference between the vibration data from two different days that could not be explained by the temperature measurement of a single sensor. A possible explanation is the operational and environmental conditions between the two days are different, which a single temperature sensor cannot capture. It could be that the temperature profile of the bridge is different between days, for instance.

Finally, the sparse sensor placement optimisation for classification (SSPOC) revealed that the frequency location that contains the largest difference between different temperature intervals is the anomalous ”spikes”, which are not part of the structural system. This is the same result that was discovered in the first measurement campaign. These anomalous ”spikes” must be filtered out as they add a false positive to the data. However, it also revealed that the natural frequencies contain the largest difference between vibration data at different temperatures.

8

rPCA in detail

8.1. Truncation unnecessary for the rPCA

The rPCA has shown good performance for data in the 1st and 2nd measurement campaigns. It is able to reduce the operational variability of the traffic, thereby allowing for comparison between structural states and identifying if the structural state has changed. However, caution should be taken when interpreting the results. It has been observed that the low-rank of the rPCA produces, in most cases, a visible division of data groups in the first three PCs when comparing different structural states. The separation between data groups in these examples is believed to be based on actual differences in the structural condition.

Figure 8.1 shows the clustering of temperature groups with a 4°C difference in the low-rank representation of the first three PCs. Here there is a comparison between a 20 % truncated sample set and a non-truncated sample set when comparing different temperature groups, as was done in chapter 7. In both cases, the rPCA has been applied, and low-rank L is used. **There does not seem to be any significant difference between the truncated and non-truncated sample sets**. Similarly, as seen in figure 8.3, there does not appear to be any particular difference between the two sample sets in their PCs. However, there is one noticeable difference in the PC score of the sample. The samples in the non-truncated sample set can have a potentially larger score in the first PC. The maximum score in the first PC for the non-truncated sample set is 80, while it is only 40 for the truncated sample set. This seems logical as the non-truncated sample set has a higher energy content, leading to a higher response. Furthermore, since there is more variation in the score of the first PC, it explains more of the total variance within the sample set compared to the truncated sample set. This can be seen in figure 8.2. The second and third PCs also have a greater extent for the non-truncated sample set compared to the truncated sample set.

To better assess how the data from the two cases are embedded into the first three PCs, a comparison can be made with classifiers, as was done in the previous chapter. Table 8.1 compares different classifiers for the truncated and non-truncated sample sets. There is a slight difference in the accuracy of the classification models. However, it is not possible to discern if this difference is due to the truncation, the selection of random samples or the model creation of the classifiers as the difference is so insignificant.

While it seems unnecessary to apply the truncation for the performance of the rPCA, it likely helps as it simplifies the data. The preprocessing step is simple and intuitive as it reduces the variance within data and makes the data adhere to a normal distribution in relation to the energy of each sample. The only ”cost” of the truncation is the removal of samples, but these samples are already undesirable as they deviate from the normal condition of the traffic.

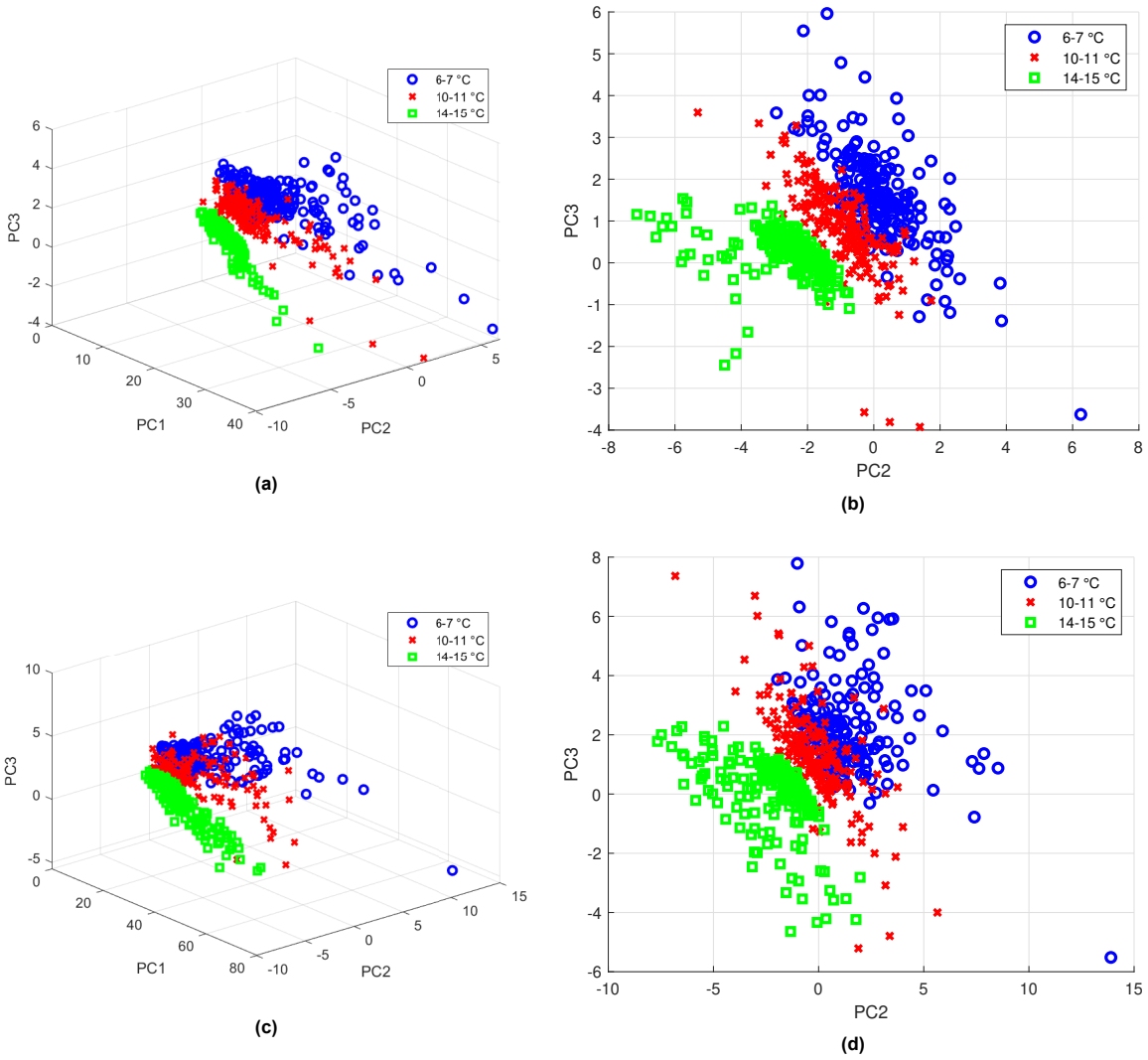


Figure 8.1: Principal component score of the samples in the PCs for data in three temperature intervals with a 4° difference. Data from sensor 31. (a) Truncated low-rank data. First three PCs. (b) Truncated low-rank data. First three PCs. (c) Untruncated low-rank data. PCs 2 and 3. (d) Untruncated low-rank data. PCs 2 and 3.

Table 8.1: Comparison between accuracy of different classifiers between truncated and non-truncated data.

	Linear discriminant	Support Vector Machine (SVM)
20 % Truncation	86.5 %	91.7 %
No Truncation	79.5 %	88.7 %
Difference	7 %	3 %

8.2. Dataset size and false positives

The rPCA has shown good performance for data in the 1st and 2nd measurement campaigns. It is able to reduce the operational variability of the traffic, thereby allowing for comparison between structural states and identifying if the structural state has changed. However, caution should be taken when interpreting the results. It has been observed that the low-rank of the rPCA produces, in most cases, a visible division of data groups in the first three PCs when comparing different structural states. The separation between data groups in these examples is believed to be based on actual differences in the structural condition.

However, there is a possibility that the rPCA produces a false positive, a division of two data groups

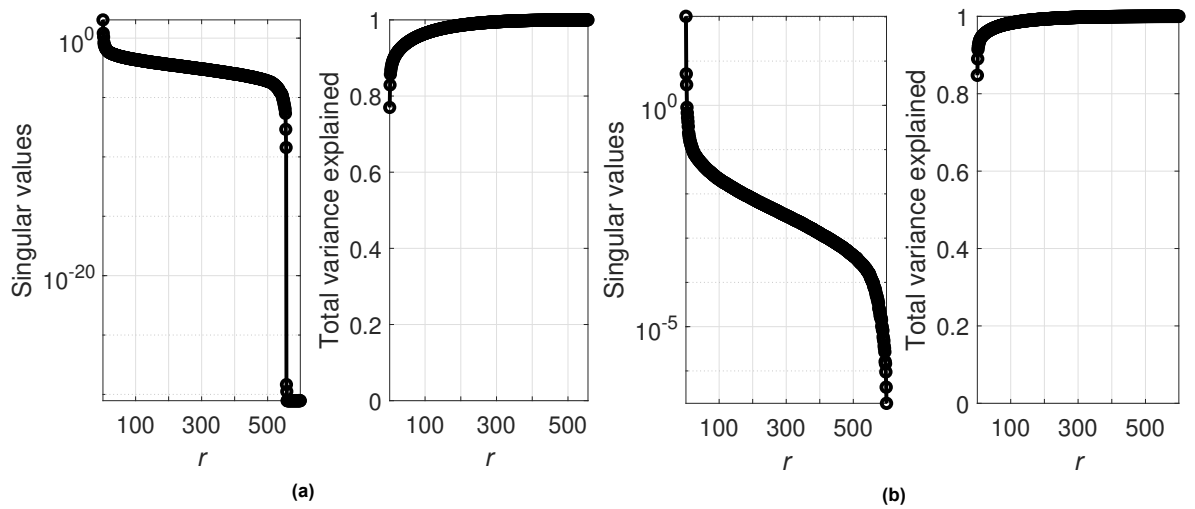
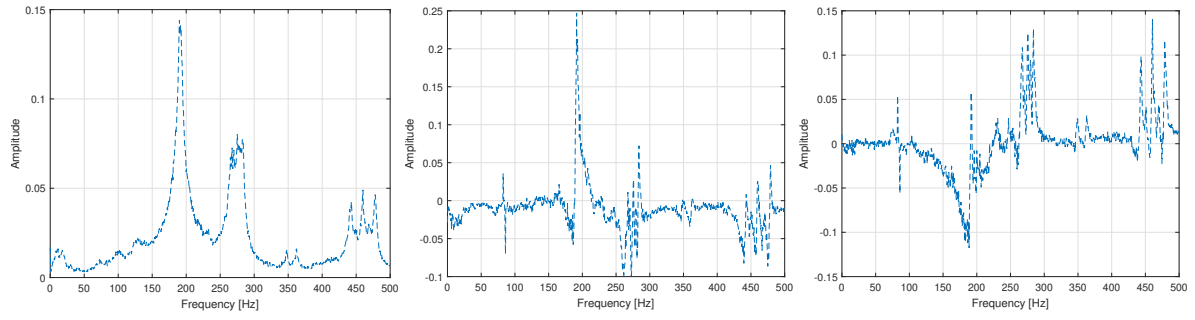


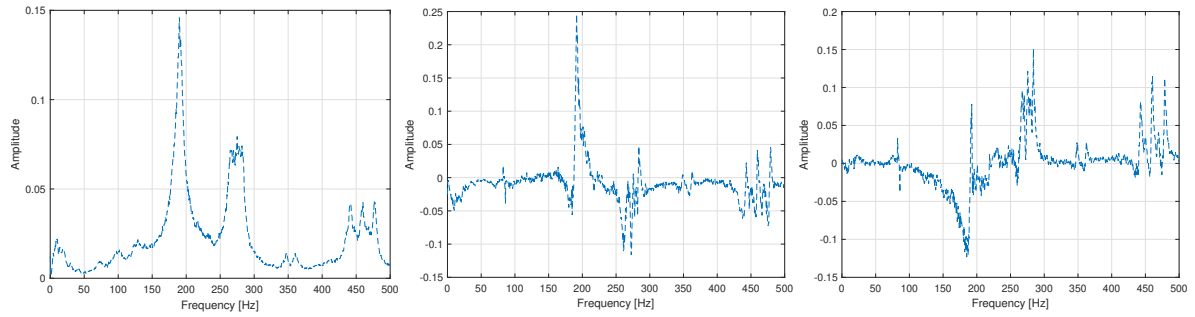
Figure 8.2: Singular values and total variance explained by the PCs. (a) Truncated Low-rank data. (b) Untruncated Low-rank data.



(a) First Principal Component. Untruncated low-rank data

(b) Second Principal Component. Untruncated low-rank data

(c) Third Principal Component. Untruncated low-rank data



(d) First Principal Component. Truncated low-rank data

(e) Second Principal Component. Truncated low-rank data

(f) Third Principal Component. Truncated low-rank data

Figure 8.3: First three Principal Component. Truncated and untruncated data sets

that have no underlying difference. For example, consider figure 8.4. It shows the PC score of the data in the first three PCs and for three different scenarios. All the data is from sensor 31 in the 2nd measurement campaign and within the same temperature interval of 6-7. Therefore, there is no underlying difference in the structural state of these samples. Two sets of 100 random samples are taken from the aforementioned data set in all scenarios. PCA is applied to the 200 samples from raw truncated data in the first scenario in figure 8.4a. This does not produce any visible division in the space of the first three PCs. The second scenario consists of low-rank data from the rPCA. The rPCA was computed on the entire data set of the temperature interval 6-7 for sensor 31. This data set is 1183 samples. Two sets of 100 low-rank samples were subsequently taken randomly from this data set, and the PCA was then computed to produce figure 8.4b. Again, there is no distinct separation between

the two randomly chosen groups. Finally, the rPCA is computed twice separately on the two random sample groups from the same temperature interval in the third scenario. Then the PCA on the two groups is computed. Here there is a visible separate clustering of the two groups of samples, even if there is no underlying difference in the structural state between these two groups.

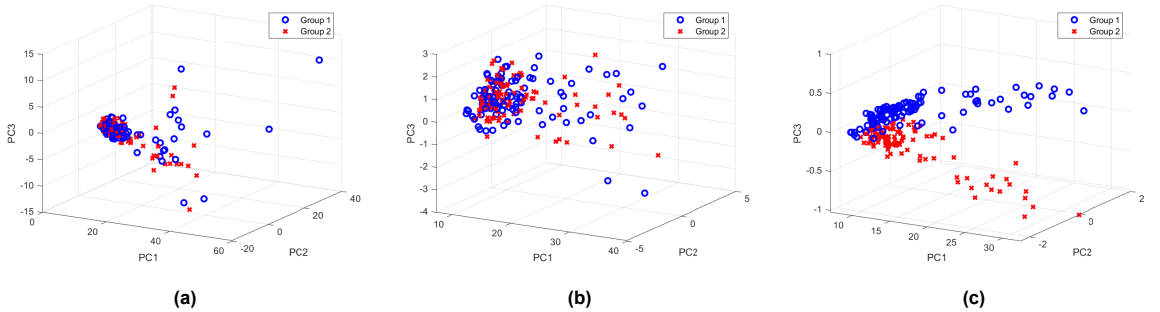


Figure 8.4: Principal component score of the samples in the PCs for data randomly sampled from temperature interval 6-7°C. (a) First three PCs of two groups of 100 raw truncated samples. (b) First three PCs of two groups of 100 low-rank samples. (c) First three PCs of two groups of 100 samples and the rPCA applied on them.

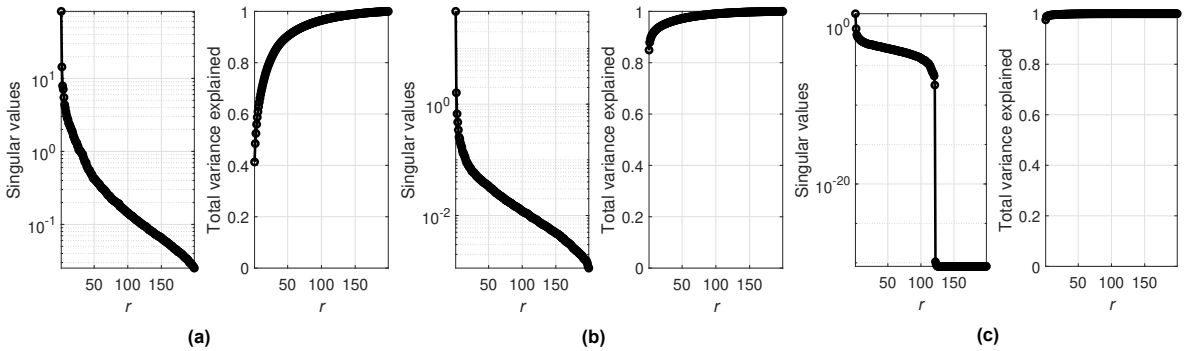
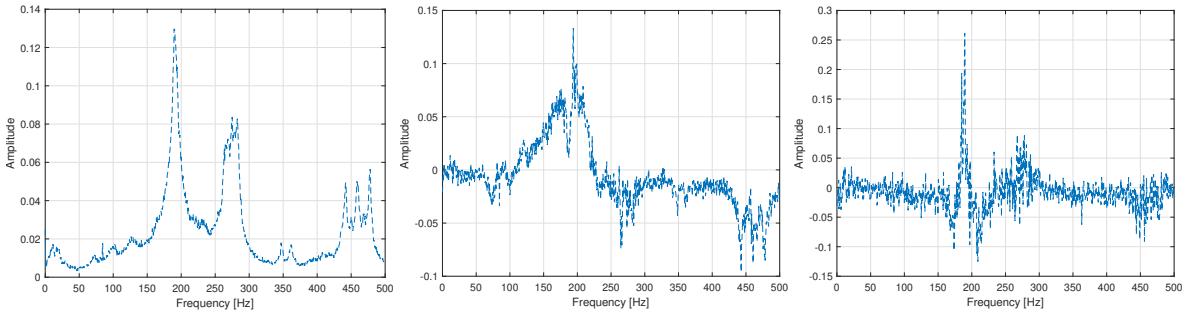


Figure 8.5: Singular values and total variance explained for data randomly sampled from temperature interval 6-7°C. (a) First three PCs of two groups of 100 raw truncated samples. (b) First three PCs of two groups of 100 low-rank samples. (c) First three PCs of two groups of 100 samples and the rPCA applied on them.

What is happening here is that the rPCA is a data-driven method; it creates a basis that is tailored to the input data. In this example, two sets of 100 random samples were gathered from a single sensor and the same temperature interval. The rPCA is then applied to these two sets separately. The two bases that are created from the rPCA are not the same. Their differences lie in the noise that they capture. Neither two of these bases encapsulate the entire random process of the traffic. This means that the reconstructed low-rank will be governed partially by the level of noise in the data, as there is a similarity of the noise within this small sample set. The difference in the noise between the two sample sets becomes one of the data's major variances. When the low-rank of these two signal groups is subsequently compared with the PCA, they will be separated by the different levels of noise in which the low-rank was reconstructed. The third PC separates the two groups in figure 8.4c. The first three PCs are shown in figure 8.6, and it appears that the third PC resembles noise.

Another observation to make is the variance explained by the PCs and the rank of these data sets. This can be seen in figure 8.5. For the third comparison scenario, the first PC accounts for 95 % of the variance, the second PC 3.1 % and the third PC 0.3 %. So the two groups are separated by a PC, which explains 0.3 % of the correlation in the data set. This variance explained is much less than has been observed when there was an actual underlying difference in the structural state. The high variance explained by the first PC can be attributed to the small data size of 100 samples for each rPCA. This small basis is not enough to capture the entire random process of the traffic. It only captures a small slice of the total variance of the random process. However, it does have information on the underlying coherent structure or the system properties, which is a larger part of the total variance within the data as there is less variance of the random process in the data. This can be compared to the variance

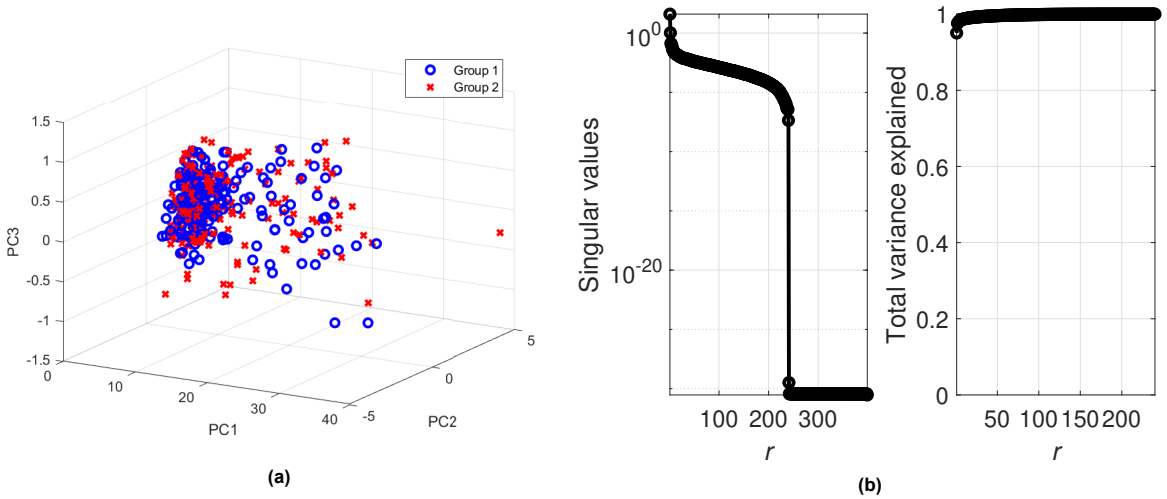


(a) First Principal Component **(b)** Second Principal Component **(c)** Third Principal Component

Figure 8.6: First three Principal Component. Two groups of 100 samples are taken from tempature interval 6-7°C and the rPCA applied on them.

explained by the first PC of the second comparison scenario, which has a lower variance as there is more information about the random process of the traffic in that data set. The data in the second comparison scenario contains information about the entire data even if only two random 100 samples are taken as the rPCA is applied on all 1883 samples.

A simple solution to this potential false positive is to increase the amount of data when the rPCA is applied. Increasing the sample size to two sets of 200 samples yields figure 8.7. There is no division between the data groups. However, the total variance of the 1st PC is still really high. It is not at comparable levels to the first PC with the rPCA applied on the entire sample size of 1883 samples in the second scenario in figure 8.5b. 200 samples are not enough to cover the variance of the entire random process, but it is enough so that the data is not separated based on noise in the first three PCs.



(a) **(b)**

Figure 8.7: Low-rank Data from two groups of 200 samples taken from tempature interval 6-7°C and the rPCA applied on them. (a) First three PCs (b) Singular values and total variance explained by the PCs.

8.3. Convergence

The rPCA is a data-driven method; it creates a basis tailored to the specific data. The PCs of the low-rank, which are robust towards outliers in the data, form this basis. This means that the amount of data inputted into the rPCA is essential to its performance. The question is, how much data is required to enable accurate comparisons between two different structural states?

The measured response of the accelerometers is a function of the dynamic properties at the sensor's location and the stochastic process of the traffic. The dynamic properties can be considered a determin-

istic quantity, given that the system does not change in the observed time frame (The environmental and operational conditions are the same). However, the traffic and energy it inputs into the system is a stochastic process in the time and frequency domain. The energy inputted into the system by each vehicle will be different. To be able to compare two sets of random variables from this process, they ideally need to span the same sample space; otherwise, there is a comparison between the differences in the random process. However, in all these random realisations, there is always information about the same dynamic system.

To better understand how the amount of data affects the performance of the rPCA consider figure 8.8a, which shows the development of the variance explained by the first PC with an increasing amount of samples. The samples are from temperature interval 10-12°C. Two cases are considered: the variance explained by the 1st PC of raw truncated data and the variance explained by the 1st PC of low-rank data. For the raw truncated data, the variance explained by the first PC converges early at a comparatively ”low” variance. The variance explained by the first principal component of the low-rank follows an entirely different pattern. It starts high, around 95 % for 100 samples. Then the variance steadily decreases until it reaches 1000 samples. This is the length of the feature vector/frequency vector. From this point and beyond, the variance of the first PC increases steadily, and it seems it will converge at some point. Figure 8.8c shows the rank of the same low-rank at different sample sizes. The rank increase with a nearly constant slope of $y = 0.6x$ until it reaches 1000 samples. This means that with the introduction of 100 new samples, around 60 samples will be linearly independent, while the remaining 40 samples will be linearly dependent. After reaching 1000 samples, the rate at which the rank increases with the introduction of new samples goes down considerably. Figure 8.8b shows the percentage of zero entries in the sparse matrix of the rPCA. The percentage steadily increases with an increasing number of samples, but at 1000 samples, there is a shift, and the percentage starts to decrease.

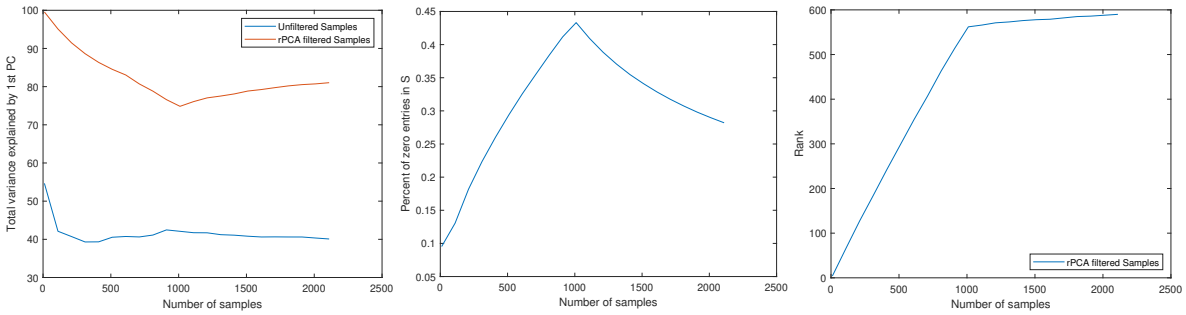
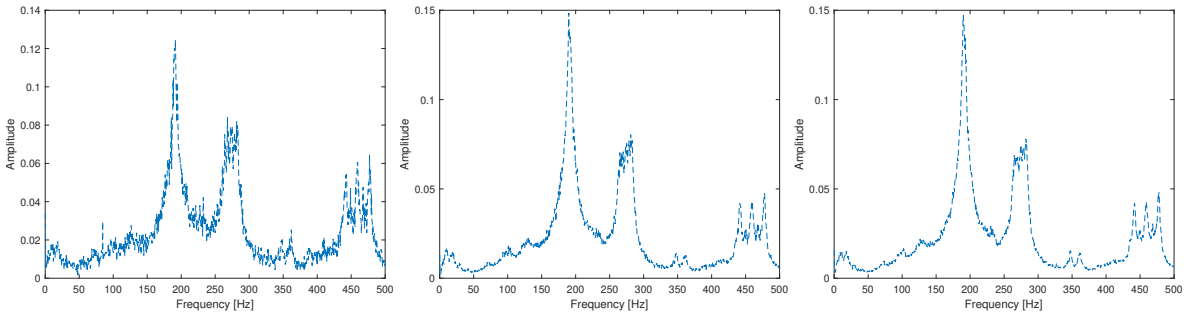


Figure 8.8: Development of the rPCA with an increasing amount of samples. Data from sensor 31 from temperature interval 10-11°C (a) Relationship between the number of samples and the total variance explained by the first PC for the low-rank and raw-truncated samples. (b) Relationship between the number of samples and the percentage of zero entries in the sparse matrix S of the rPCA. (c) Relationship between the number of samples and the rank of the low-rank.

Increasing the number of samples means more realisations of the random process of the traffic. This brings in more measurement noise, but the main correlation of the data becomes clearer as all samples have information on the dynamic properties. With more noise in the data, the total variance explained by the first PC will be lower, as seen in figure 8.8a. However, the PC with a low amount of samples does not resemble the dynamic system properties. This can be seen in figure 8.9. It is full of noise. Increasing the number of samples makes the first PC resemble the dynamic properties of the system, but the total variance of the PC is lower. The correlation connected to the noise is now explained by higher PCs and the 1st PC only explains the correlation connected to the same dynamic properties between samples. At 1000 samples, when the number of samples becomes the same as the length of the frequency vector, the regularisation parameter $\lambda = 1/\sqrt{max(mxn)}$ in the rPCA starts to decrease as the maximum dimension of the data matrix starts to increase. This means the filtering of the rPCA will be more aggressive. More of the noise in the data will be discarded into the sparse matrix of the rPCA. This can be seen with the percentage of zero entries in the sparse matrix decreasing with the increasing amount of samples after 1000 samples. It can also be seen with increasing variance explained by the first PC. The number of samples used with the rPCA has an effect on its performance.



(a) 100 Samples **(b)** 1000 Samples **(c)** 2270 Samples

Figure 8.9: The 1st PC with different amount of samples.

“*An oil painting by Aelbert Cuyp of the Skinny bridge in Amsterdam*” DALL-E 2

9

Conclusion and Recommendations

9.1. Conclusions

The main objective of this research was to remove or reduce the operational variability to allow for comparison and identification between structural states. The rPCA has proven to be an effective method for discovering the underlying dynamics of a given sensor. All the signals from a sensor share the same underlying dynamic system, given that the environmental and operational conditions are the same but are obstructed by noise due to the operational variability of the traffic. The reconstructed low-rank L of the rPCA is assembled with PCs that are robust to the outliers in the data, that is, the noise. These robust PCs contain information about the similarity in the process that generated the signals, which corresponds to the dynamic properties. Thus the low-rank L resembles that of the underlying dynamics of the system.

The low-rank of two different structural states, such as damaged or undamaged, can be compared to measure if damage is present. This involves computing the PCA on a combined data set of both structural states to find the PCs that explain the differences between the two states, in other words, finding a low-rank representation that separates the signals according to their structural states. This allows for automatic detection if the structural states are significantly different.

Applying the above process, ”damage” was successfully identified. In the first measurement campaign, two measurement areas were considered, one damaged area with the presence of fatigue cracks and another undamaged reference area. Sensor pairs from the two areas were compared, and it was possible to see a difference in the signals of the two areas and identify ”damage”. The damaged and undamaged signals clustered separately in the first three PCs. Computing the Mahalanobis distance of the samples reveals if there is a deviation in the normal condition of the bridge. The Mahalanobis distance of the ”damaged” samples was much greater than the ”Healthy” samples. A simple decision boundary according to the distance of the samples could be taken to distinguish between the two groups. Thus, it was possible to distinguish between the signals based on their score in the first three PCs. In the analysis, a difference could be found for every valid sensor pair. However, it is not possible to assert that the fatigue cracks caused this difference as the structural configuration of the two areas that were compared was different.

In the second measurement campaign, the effect of temperature on the structural state was investigated. Two analyses were conducted. First, a single resonance peak was tracked in relation to temperature. It revealed a negative linear correlation between temperature and the frequency value of a high natural frequency, which is in agreement with what is seen in the literature as was discussed in chapter 2. There was also a negative linear correlation between temperature and the amplitude of the natural frequency. It could be an indication that the damping properties of the structure change

with temperature. Another possible reason is that the position of the mode shape that corresponds to the natural frequency is shifting within the structure with temperature. This would mean the sensor is capturing a different part of the mode shape with potentially different amplification.

In the second analysis of the second measurement campaign, the entire frequency vector was investigated in relation to the temperature and PCA was used to distinguish between the samples belonging to the different temperature intervals. This proved to be a more challenging problem compared to the ”damage” detection with the data from the first measurement campaign. The temperature has a minor effect on the system properties compared to the ”damage” present in the first measurement campaign. This means that the underlying dynamic properties at a given temperature interval are similar to those at adjacent temperature intervals. Any noise present in the signals will make it difficult to distinguish between the underlying dynamic properties at different temperatures. However, with the rPCA the noise within the vibration data can be reduced, and the underlying coherent structures can be obtained. By lowering the regularisation parameter to $\lambda = 1/\sqrt{2 \cdot max(n, m)}$ in the rPCA, even more of the noise within the data can be filtered out, enabling the distinction between signals with a 1°C temperature difference which was otherwise impossible with the default regularisation parameter. The rPCA, with a lower regularisation parameter, is able to extract the underlying coherent patterns that correspond to the dynamic properties at each temperature interval.

The analysis of the data from the second measurement campaign revealed that temperature affects the system, and it is possible to extract the underlying coherent structures corresponding to the dynamic properties at different temperatures with the rPCA. A complete understanding of the dynamic properties under all possible environmental conditions is necessary for a robust structural health monitoring program.

This was only achieved for vibration data from a horizontal sensor with high-frequency content. It was more challenging to distinguish between the dynamic properties at different temperatures for low-frequency content from a vertical sensor. It was not possible to distinguish between signals with a temperature difference lower than 5°C. A possible reason for this is the frequency resolution, which needs to be higher for the low-frequency content to capture the shifts in the lower natural frequencies. However, increasing the frequency resolution comes at a cost by increasing the time window around the samples in the time domain. This, in turn, decreases the number of samples in the data, which has a negative effect on the performance of the rPCA.

There was a higher degree of difference between the vibration data from two different days that could not be explained by the temperature measurement from a single sensor. This was seen when comparing vibration data from the horizontal accelerometer with a 2°C difference. The data from the different days had a higher degree of a difference than data within a single day, even if the temperature difference was the same. A possible explanation for this behaviour is that the vibration data is grouped into temperature intervals by a single temperature sensor. This single temperature sensor does not capture the full temperature profile of the bridge. So while there is a 2°C difference between the temperature intervals according to this sensor, the actual temperature profile of the bridge could be potentially different between the different days. This, in turn, means that there is a difference in the dynamic properties, which is not explained by this single temperature sensor. The ”extra” difference in the dynamic properties between the two different days could also be due to changes in other operational and environmental conditions. For instance, it might have been raining on one of the days, which led to an increase in the mass of the structure, thereby altering the dynamic properties. During the measurement period, there was ongoing maintenance of the drawbridge. The position of concrete barriers on the road might have been moved between the days due to maintenance. Thus the mass distribution would be different, altering the dynamic properties. This might also have changed the position of the traffic on the bridge deck, which would also change the mass distribution. However, at least it seems that a single temperature sensor is not capturing the cause of this difference between the days.

The SSPOC also revealed the locations in the frequency vector that contained the largest difference between the different structural states. Aside from the anomalous ”spikes”, the location of these sparse sensor locations was at the natural frequencies for data from both measurement campaigns. Understanding which natural frequencies change the most between two structural states is beneficial, and

these natural frequencies can be analysed further. The SSPOC also revealed that the changes in the vibration data are due to changes in the natural frequencies.

9.2. Recomendations

Higher frequency resolution for low-frequency content

When working with vibration data with a low-frequency content, a frequency resolution of 0.125 Hz was used. This proved to be insufficient in capturing the change in the dynamic properties of the low-frequency content due to temperature. The frequency resolution has to be increased for the low-frequency content to have the same level of accuracy as the analysis with the high-frequency content. A 0.5 Hz frequency resolution was used for the high-frequency content of sensor 31 from the second measurement campaign, which translates to a ratio of 380 between the natural frequency at 190 Hz and the frequency resolution. This ratio is even higher for the natural frequencies at 400 Hz. If the same ratio or level of accuracy was used for the natural frequency at 10 Hz of sensor 17, then the frequency resolution would be 0.026 Hz. However, this frequency resolution would translate to a time window of around 40 seconds, which would yield few samples in a 15-minute time frame. Using this frequency resolution would be impractical with the amount of data available for this research, which was only two days. More data is needed in order to see if there is any change in the lower natural frequencies with temperature. Future analyses with access to a more extensive data set can investigate the low-frequency content in more detail.

Regularisation parameter optimisation

The default value of the regularisation parameter in the Robust Principal Component Analysis is $\lambda = 1/\sqrt{(max(mxn)}$ as recommended by the authors of the rPCA [11]. It was arbitrarily lowered to $\lambda = 1/\sqrt{2 \cdot (max(mxn)}$ in order to reduce the noise in the vibration data, and it enabled distinction between vibration data with a 1°C difference. Lowering the regularisation parameter leads to more aggressive filtering of the rPCA as more of the data is put into the sparse matrix of the rPCA. Thus the low-rank of the rPCA is reconstructed from even fewer PCs leading to an even lower rank which should contain only the major coherent structures in the data. However, selecting a too-low regularisation parameter can lead to the rPCA filtering out coherent structures in the data. This was seen when the rPCA was used to remove corruption in fluid flows [38]. The researchers of the paper also discussed that there needs to be an understanding of the selection of an optimal regularisation parameter for different scenarios in fluid mechanics. The same can be said for the vibration data in civil engineering structures. There needs to be an understanding of which regularisation parameter best suits the vibration data.

Temperature profile

It was discovered that there was more difference in the vibration data between days that was not explained by the temperature measurement of a single sensor. It is plausible that the temperature profile between these two days was different, which leads higher degree of difference in the underlying system and, subsequently, in the vibration data. This should be investigated. Instead of just working with temperature data from a single sensor, the entire temperature profile of the bridge should be used for analysis. There is a temperature gradient present in the bridge, in the vertical direction and in the transverse direction of the bridge. This underlying temperature pattern can be potentially extracted. This would provide more comprehensive information on the environmental condition of the bridge, and it could be linked to the vibration data. PCA or rPCA could be used to find the coherent structures within the temperature profile.

Remove the anomalous spikes

The anomalous ”spikes” present in the vibration data are not part of the dynamics of the structure, but it is theorised that the ”spikes” are due to the sensors themselves and auxiliary equipment. These

”spikes” contained the most significant difference between the structural states of both measurement campaigns, according to SSPOC. The ”spikes” add a false positive bias to any damage detection involving the entire frequency vector. They need to be filtered in order to avoid misclassifying damage as change these anomalous ”spikes”.

Expand the spatial domain of the analysis

In this thesis, the rPCA was used only on the output of a single accelerometer for each analysis. Only the frequency spectrum and its natural frequencies were used to determine if the two dynamic systems were different. However, these natural frequencies have corresponding vibration modes that can help distinguish between different structural states. These vibration modes hold information about the dynamic system that might be valuable in damage detection. The rPCA can reduce the noise within each sensor and reveal the underlying dynamic system in which each sensor is measuring. The next step with the rPCA should be to expand its usage in the spatial domain by incorporating more sensors into the analysis and experiment in extracting the modal shapes of the structure from the rPCA-filtered data.

Stochastic input force

In this thesis, the rPCA managed to discover the underlying patterns in the vibration data corresponding to the dynamic properties of the system. The goal of using the rPCA on the vibration data is to obtain the system properties by removing the operational variability in the input force, that is, the traffic. The input force is a stochastic process, with each realisation of the traffic being potentially different. On the other hand, the system properties can be thought of as a deterministic quantity as it does not change between the different realisations of the response, given that environmental and operational conditions do not change.

The performance of the rPCA in extracting the system properties depends on the stochastic input force. For instance, if the random process of the input force does not have energy at a particular frequency, then it will not be possible to determine the system properties at that frequency. The success of the rPCA in obtaining the system properties for the entire frequency spectrum depends on if the random process is broadband or narrowband. Broadband processes such as white noise should perform well in this regard. The vibration data in this thesis seems to be broadband, as there are natural frequencies over the entire frequency band. Another factor to consider is how many samples are required for the rPCA to obtain the underlying dynamic properties of the signals. The number of required samples is most likely tied to the characteristics of the random process of the input force. Simple processes such as white noise could require fewer samples to discover the underlying dynamics with the rPCA.

The vibration data in this thesis consists of ambient vibrations, traffic and the amplification due to the sensors. With this data, the rPCA was successful in finding the underlying patterns corresponding to the dynamic properties. Can the rPCA be used with other types of input forces and discover the underlying patterns for other structures? In order to understand how rPCA performs for vibration data for the different input forces, there needs to be a good understanding of the statistics of these stochastic input forces.

Sparsity and compressed sensing in SHM for civil engineering structures

The Mathematics of sparsity and compressed sensing, which are the foundation of rPCA and the SSPOC, is an exciting field and can be potentially applied to other problems in SHM for civil engineering structures. With sparsity and compressed sensing, it is possible to reconstruct complex systems from sparse measurements. One question in vibration-based monitoring is where to place the sensors to measure damage in a structure. The goal should be to minimise the amount of the sensor for an economical solution. The mathematics of sparsity and compressed sensing can prove useful in finding the optimal sparse sensor locations for a given structure.

References

[1] ASCE. *ASCE's 2021 Infrastructure Report Card*. 2021.

[2] Mohsen Azimi et al. "Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review". *Sensors* 20.10 (2020), p. 2778.

[3] Richard G. Baraniuk. "Compressive Sensing [Lecture Notes]". *IEEE Signal Processing Magazine* 24.4 (July 2007), pp. 118–121.

[4] Richard Bellman. "Dynamic Programming". *Science* 153.3731 (July 1966), pp. 34–37.

[5] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*. Vol. 4. Springer, 2006.

[6] J.m.w Brownjohn. "Structural health monitoring of civil infrastructure". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365.1851 (Feb. 15, 2007), pp. 589–622.

[7] B. W. Brunton et al. "Sparse Sensor Placement Optimization for Classification". *SIAM Journal on Applied Mathematics* 76.5 (Jan. 2016), pp. 2099–2122.

[8] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, May 5, 2022. 615 pp.

[9] E.J. Candes, J. Romberg, and T. Tao. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information". *IEEE Transactions on Information Theory* 52.2 (Feb. 2006), pp. 489–509.

[10] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements". *Communications on Pure and Applied Mathematics* 59.8 (2006), pp. 1207–1223.

[11] Emmanuel J. Candès et al. "Robust principal component analysis?" *Journal of the ACM* 58.3 (June 9, 2011), 11:1–11:37.

[12] P. Cornwell et al. "Environmental Variability of Modal Properties". *Experimental Techniques* 23.6 (1999), pp. 45–48.

[13] Corinna Cortes and Vladimir Vapnik. "Support-vector networks". *Machine Learning* 20.3 (Sept. 1, 1995), pp. 273–297.

[14] E. J. Cross et al. "Long-term monitoring and data analysis of the Tamar Bridge". *Mechanical Systems and Signal Processing* 35.1 (Feb. 1, 2013), pp. 16–34.

[15] Janno De Bruijn. "Vibration-based Monitoring of the Zwartewaterbrug: A Machine Learning Approach". Msc thesis. Tu Delft, 2019.

[16] D.L. Donoho. "Compressed sensing". *IEEE Transactions on Information Theory* 52.4 (Apr. 2006), pp. 1289–1306.

[17] Charles R Farrar and Keith Worden. "An introduction to structural health monitoring". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365.1851 (Feb. 15, 2007), pp. 303–315.

[18] Charles R. Farrar and Keith Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, Nov. 19, 2012. 667 pp.

[19] R. A. Fisher. "The Use of Multiple Measurements in Taxonomic Problems". *Annals of Eugenics* 7.2 (1936), pp. 179–188.

[20] Yongda Fu and John T. DeWolf. "Monitoring and Analysis of a Bridge with Partially Restrained Bearings". *Journal of Bridge Engineering* 6.1 (Feb. 1, 2001), pp. 23–29.

[21] Marie-Louise Greijmans. “Vibration-based damage detection of the Haringvlietbrug: A data-driven approach”. Msc thesis. Tu Delft, 2020.

[22] Guðmundur Valur Guðmundsson. “Sundabrautin”. Reykjavík, Iceland, 2022.

[23] Tony Hey et al. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Oct. 2009.

[24] Wei-Hua Hu et al. “Vibration-based structural health monitoring of a wind turbine system Part II: Environmental/operational effects on dynamic properties”. *Engineering Structures* 89 (Apr. 15, 2015), pp. 273–290.

[25] Tycho Kockelkorn. “Dynamic Response of an Orthotropic Bridge Deck Subjected to EOVs”. Msc thesis. Tu Delft, 2022.

[26] Coen Kortendijk. “Influence of temperature on natural vibrations of steel bridges”. Msc thesis. Tu Delft, 2020.

[27] N. a. J. Lieven et al. “Vibration–based structural damage identification”. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 359.1778 (Jan. 15, 2001), pp. 131–149.

[28] Zhouchen Lin, Minming Chen, and Yi Ma. “The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices”. *Journal of Structural Biology* 181.2 (Feb. 2013), pp. 116–127.

[29] F. Magalhães, A. Cunha, and E. Caetano. “Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection”. *Mechanical Systems and Signal Processing*. Interdisciplinary and Integration Aspects in Structural Health Monitoring 28 (Apr. 1, 2012), pp. 212–228.

[30] Krithika Manohar et al. “Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns”. *IEEE Control Systems Magazine* 38.3 (June 2018), pp. 63–86.

[31] Krithika Manohar et al. “Predicting shim gaps in aircraft assembly with machine learning and sparse sensing”. *Journal of Manufacturing Systems*. Special Issue on Smart Manufacturing 48 (July 1, 2018), pp. 87–95.

[32] Marko Milosevic. “Structural Health Monitoring of the Zwartewaterbrug Bridge”. Msc thesis. Tu Delft, 2018.

[33] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Sept. 7, 2012. 1102 pp.

[34] Bart Peeters and Guido De Roeck. “One-year monitoring of the Z24-Bridge: environmental effects versus damage events”. *Earthquake Engineering & Structural Dynamics* 30.2 (2001), pp. 149–171.

[35] Guido De Roeck. “The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience”. *Journal of Structural Control* 10.2 (2003), pp. 127–134.

[36] Anders Rytter. *Vibrational Based Inspection of Civil Engineering Structures*. Fracture and Dynamics. Aalborg: Dept. of Building Technology and Structural Engineering, Aalborg University, 1993.

[37] Hadi Salehi and Rigoberto Burgueño. “Emerging artificial intelligence methods in structural engineering”. *Engineering Structures* 171 (Sept. 15, 2018), pp. 170–189.

[38] Isabel Scherl et al. “Robust principal component analysis for modal decomposition of corrupt fluid flows”. *Physical Review Fluids* 5.5 (May 28, 2020), p. 054401.

[39] Hoon Sohn. “Effects of environmental and operational variability on structural health monitoring”. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365.1851 (Feb. 15, 2007), pp. 539–560.

[40] Robert Tibshirani. “Regression Shrinkage and Selection Via the Lasso”. *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[41] Yong Xia et al. “Temperature effect on vibration properties of civil structures: a literature review and case studies”. *Journal of Civil Structural Health Monitoring* 2.1 (May 1, 2012), pp. 29–46.

[42] Samol Ya, Kentaro Yamada, and Toshiyuki Ishikawa. “Fatigue Evaluation of Rib-to-Deck Welded Joints of Orthotropic Steel Bridge Deck”. *Journal of Bridge Engineering* 16.4 (July 1, 2011), pp. 492–499.

[43] Xiaoming Yuan and Junfeng Yang. “Sparse and low rank matrix decomposition via alternating direction method”. *Pacific Journal of Optimization* 9 (Jan. 1, 2009).

A

Anomalous Spikes

During analysis, anomalous spikes were encountered. These spikes are peaks in the DFT that are contained to a single frequency index. They are present in each sensor but at different frequencies and vary with temperature. The spikes are present in signals from both the first and second measurement campaigns, but both campaigns use the same sensors. Even without the excitation from the traffic, the spikes are still present. This section will go into detail on the behaviour and effect of these spikes.

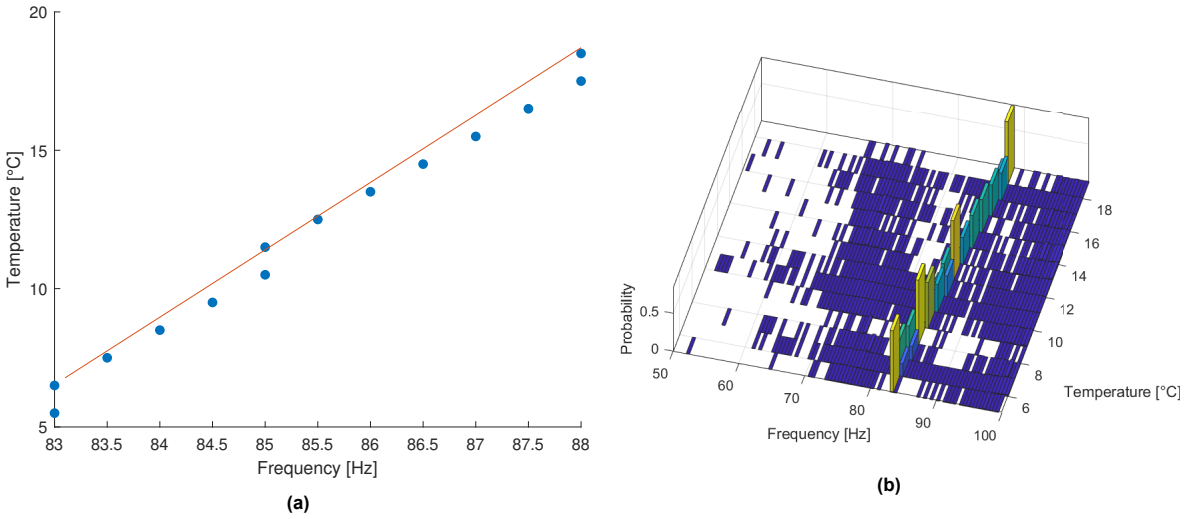
Location

The spikes appear at various locations below 110 HZ in the frequency domain for the various sensors. In some instances, they are also isolated from any other resonance activity. The location of the spikes does not change drastically between sensors measuring vertical and horizontal acceleration. These two types of sensors have completely different frequency contents. Horizontal sensors capture high-frequency content, while vertical sensors capture low-frequency content, but the anomalous spikes do not follow this change to the same degree.

These ”curious spikes” were first discovered by Marie-Louise [21] in the data from the first measurement campaign. However, these anomalous spikes were also present in the data from the second measurement campaign. Both the first and second measurement campaigns used the same sensors. The difference was in the locations where the sensors were deployed, possibly suggesting that the spikes were bound to the sensors and the auxiliary equipment.

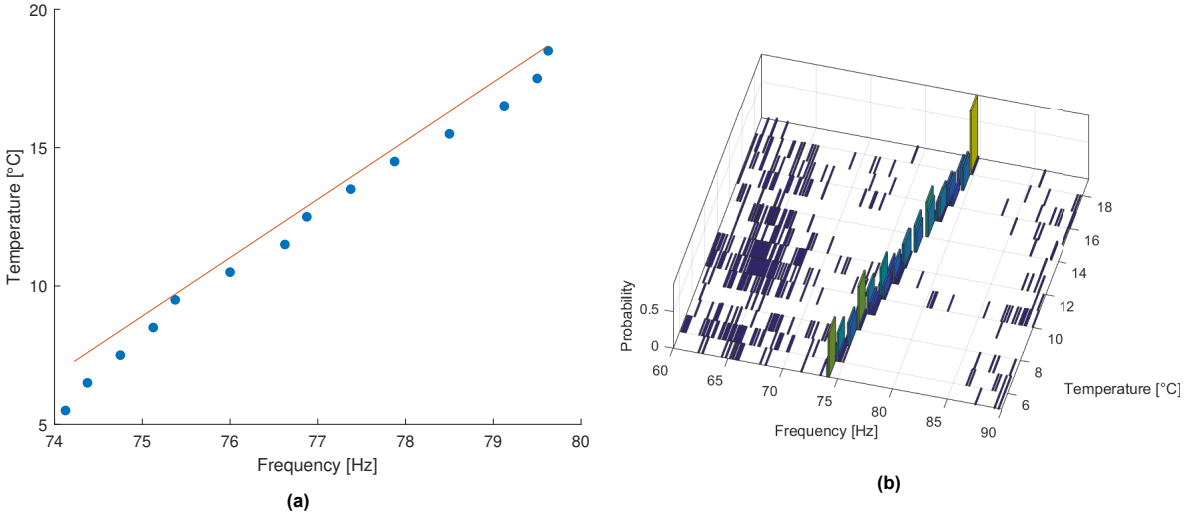
Temperature

The frequency location of the spikes shows a positive linear correlation with temperature. This can be seen in figure A.1, which shows how a single anomalous spike changes with temperature for horizontal sensor 31 from the second measurement campaign. Figure A.1a shows how the mean value of the spike, while figure A.1b shows a histogram where the largest peak is located in a given frequency interval. According to the histogram figure, the spike is present in nearly all samples. This behaviour of the spikes is the same for a vertical sensor, as seen for sensor 17 from the second measurement campaign in figure A.2. The spikes seem to be much more responsive to temperature than the natural frequencies; a clear shift in frequency is visible between all temperature intervals. There is no sudden shift in the relationship between the frequency of the spike and temperature between the two days. The temperature measurement of a single sensor is enough to capture the relationship between the spike and temperature. This could indicate that the sensor equipment is more sensitive to temperature changes than the dynamic system. This seems plausible as all the sensor equipment is exposed to the air.



(a) (b)

Figure A.1: Frequency location of anomalous spikes with temperature. There is a positive linear correlation of the spike with temperature. Data from sensor 31 (a) Mean value of each temperature interval. (b) Histogram of the location of the spike at different temperatures.



(a) (b)

Figure A.2: Frequency location of anomalous spikes with temperature. There is a positive linear correlation of the spike with temperature. Data from sensor 17 (a) Mean value of each temperature interval. (b) Histogram of the location of the spike at different temperatures.

Present without traffic excitation

In her research, Marie-Louise discovered that these spikes are present in the frequency domain even if there was no traffic on the bridge. Only the spikes were visible in the frequency domain; not even the system's natural frequencies were visible under the ambient load. This could suggest that the spikes are not a resonance phenomenon but rather some force. It also revealed other higher-frequency spikes around 190 Hz, 210 Hz and 290 Hz. These spikes have not been detected in the analysis of this thesis as all analysed signals contain excitation from the traffic, thereby masking these higher frequency spikes.

Summary

The spikes appear in a similar frequency range for sensors with widely different frequency content. They do not look like any other resonance frequencies in the system. The spikes are present even if a sensor is moved to a different location and also appear when there is no traffic on the bridge. This

suggests that the spikes are not part of the structural system of the bridge but rather some input force that could be due to the installed hardware of the sensor equipment. The variation in temperature could be explained by the resistance of the circuitry changing with temperature and thus changing the frequency of this ”force”. It could also explain why the spikes are so responsive to temperature, as all the electrical equipment is exposed.

The anomalous spikes are part of the overall frequency spectrum and can potentially affect any analysis of the frequency spectrum. Specifically, the goal of the analysis of this thesis is to identify the difference between data groups. In the case of the first measurement campaign, the goal was to distinguish between damaged and undamaged signals, while in the second measurement campaign, the analysis was focused on the effect of temperature on the structural system. In both cases, the spikes introduce a false positive bias to the analysis. The spikes are different between sensors and change with temperature. However, they do not represent the actual dynamic properties of the system. Thus a false positive bias is added as the spikes contribute to the difference between data groups but are not the system under analysis. The SSPOC algorithm revealed that the spikes are also potentially the highest contributing point in the frequency spectrum that discriminates between the data groups. This is shown in figure A.3, which shows three sensor locations that maximally discriminate between signals from sensors 7 and 23. The highest scoring point of the three sparse locations is one of the anomalous spikes, indicating that this frequency index contains the most discriminating information between the two groups. Ideally, this false positive bias should be removed either with some processing of the data or by reconsidering the choice of hardware.

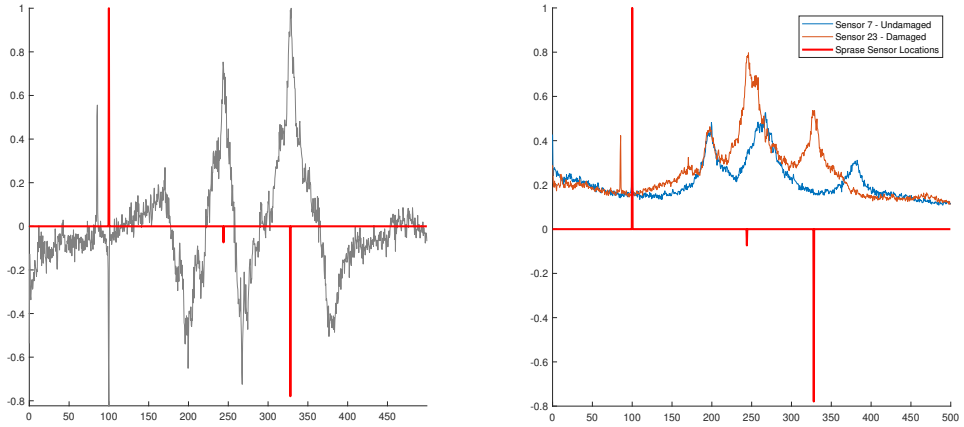


Figure A.3

B

Fourier transform

The Fourier transform is a mathematical operation that transforms a function into its frequency components. This transformation moves a function into the generic basis of the frequency domain. For a time domain function x(t), the Fourier transform is:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t}\, dt \tag{B.1}$$

The corresponding Fourier pair or the inverse Fourier transform is:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t}\, d\omega \tag{B.2}$$

Parsevals Theorem

The total energy within a signal is the same in its time-domain and frequency-domain representation. The Fourier transform preserves the ℓ_2 norm or energy between transformation. This is Parseval's theorem and can be stated follows:

$$\int_{-\infty}^{\infty} |X(\omega)|^2\, d\omega = 2\pi \int_{-\infty}^{\infty} |x(t)|^2\, dt \tag{B.3}$$