Improving sensor placement optimisation robustness to environmental variations and sensor failures for structural health monitoring systems.

A Thesis submitted to the University of Sheffield

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

by

Tingna Wang

Supervisors: David J Wagg, Robert J Barthorpe, Keith Worden

Department of Mechanical Engineering

University of Sheffield

March 2023

Dedicated to my beloved parents.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, David Wagg, Robert Barthorpe and Keith Worden. Thanks to Prof. Wagg for many valuable suggestions on how to do research, and some conversations about life that made me feel like family. Thanks to Dr Barthorpe for his assistance during the first two years of my PhD studies. His regular supervision not only improved my research ability but also helped me gradually learn how to be responsible for work and life. Thanks to Prof. Worden for his insightful comments and advice during the last two years of my PhD. I would not have completed my studies without their constant support, encouragement and patience over the past few years. I would also like to thank Robin Mills and Michael Dutchman for their tremendous help with the experiment for this research and their friendliness.

Thanks to Keith Dean for organising all the trips in the UK and Denmark, which made it possible for PhDs from other departments or universities and me – as a member of the Prosperity Partnership – to have wonderful memories together. I would like to extend my appreciation to all the members of the Dynamics Research Group for their kindness and enthusiasm that made my study and life in the UK a marvellous and unforgettable time. I would like to especially thank Sikai Zhang and Weijiang Lin for their company and trust in me. My appreciation also goes to my family and friends for their love, support and understanding throughout my studies. Thank you, Lord, for always being by my side.

I would like to acknowledge the support of Siemens Gamesa and the Engineering and Physical Sciences Research Council (EPSRC) via grant number EP/R004900/1. This research utilised the Laboratory for Verification and Validation (LVV), which was funded by the EPSRC (grant numbers EP/J013714/1 and EP/N010884/1).

Abstract

The installation of structural health monitoring (SHM) systems based on machine learning algorithms on structures has been of constant interest. The application of this kind of SHM system can facilitate decisions regarding maintenance and the remaining useful life of the structure in a more automatic and convenient way. As part of the SHM system to collect information, the sensor system can be optimally designed to improve the performance of the final system. In this thesis, the work focuses on how to consider the effects of environment and sensor failures during the sensor placement optimisation (SPO) to build a more robust and effective monitoring system. Since the availability of data during the design phase varies widely from project to project and there are no studies or specifications that provide specific guidance, not much research has been done on the design of such sensor systems, which require reliable simulated or measured data to be available during the design phase. Considering the different levels of data accessibility at the design phase, this thesis proposes a series of strategies for the optimal design of sensor systems for SHM systems from a machine-learning perspective.

The first main content of this thesis is hierarchical assessment criteria of designed-system performance to balance the computational feasibility and visualisation of the final system performance. At the stage after data is collected, machine learning model results are often used as a criterion, whose acquisition is usually time-consuming. At the same time, higher data accuracy is required. Therefore, the criteria used in the design of sensor systems are divided into different tiers. The criteria for the initial stage can be abstracted from the purpose of the applied machine learning model to significantly reduce the number of candidate designs. The criteria for the final stage can be similar to those used in the stage after data is collected. Whether or not to

use criteria from all tiers depends on the level of data availability.

It can be found that more work on the optimisation design of the sensor system can be done at the initial stage of the hierarchical design framework. Therefore, the other three main contents of this thesis are developed at this stage. Considering different levels of data availability, supervised and unsupervised correlation-based strategies to evaluate sensor combinations are proposed, including the evaluation criterion and the fast calculation methods of this criterion. Sensor combinations can be ranked even if only the healthy state data are accessible. To account for the effects of environmental variations, two SPO strategies based on approaches to extracting robust features are proposed, and an appropriate criterion that can be used is also introduced. These two strategies cover both situations where environmental change information is available and not. To consider the sensor-failure effect in the SPO process, another two strategies, namely fail-safe sensor optimisation or fail-safe optimisation with redundancy, are proposed in this thesis, both of which can take into account the performance of the designed system before and after the failure of some critical sensors. Different assessment criteria are adopted to demonstrate the generality of these strategies.

Publications

Conference papers:
T. Wang, D. J. Wagg, K. Worden, and R. J. Barthorpe, 2022. Assessment criteria for optimal sensor placement for a structural health monitoring system. The 13th International Workshop on Structural Health Monitoring, California, United States. (Presented in the 10th EWSHM (July 2022), Palermo, Italy.)

T. Wang, D. J. Wagg, R. J. Barthorpe, and K. Worden, 2022. On robustness of optimal sensor placement to environmental variation for SHM. The 13th International Workshop on Structural Health Monitoring, California, United States. (Presented in the 10th EWSHM (July 2022), Palermo, Italy.)

T. Wang, K. Worden, D. J. Wagg, E. J. Cross, A. E. Maguire, and W. Lin, 2023. Automatic Selection of Optimal Structures for Population-based Structural Health Monitoring. IMAC-XLI, (Feb. 2023), Austin, United States. (In press.)

Journal papers:
T. Wang, D. J. Wagg, K. Worden, and R. J. Barthorpe, 2021. On sensor optimisation for structural health monitoring robust to environmental variations. Wind Energy Science, 6(5), pp.1107-1116. (Presented in Wind Energy Science Conference (May 2021), Hannover, Germany.)

T. Wang, R. J. Barthorpe, D. J. Wagg, and K. Worden, 2022. On improved fail-safe sensor distributions for a structural health monitoring system. Data-Centric Engineering, 3. doi:10.1017/dce.2022.27

S. Zhang, T. Wang, K. Worden, and E. J. Cross, 2021. Canonical-correlation-based fast feature selection. arXiv preprint arXiv:2106.08247.

Contents

Chapter 1

An introduction to SHM and SPO

At present, the number of large or mechanical structures needing to be monitored is increasing every year. To avoid catastrophic damage and keep the structural system running normally during its lifetime, a key measure is to build a structural health monitoring (SHM) system to monitor the health state of a structure. SHM, in short, is an automated monitoring process with the aim of inferring the health state of a structure using data derived from a network of sensors placed upon it, as shown in Figure 1.1.

To build an effective SHM system, it is necessary to ensure that informative data can be collected by the monitoring system on the structure. For the acquisition of informative data, one approach is to perform the sensor placement optimisation (SPO) technique, by which the number and location of sensors are optimised for a specific objective to reduce instrument and processing costs without compromising on the effect of monitoring. On the other hand, SPO can also help avoid collecting data containing similar information, which can benefit the establishment of a more compact model and increase the effectiveness of real-time monitoring.

This chapter will survey a body of literature and present some current research advances in the field of SHM and SPO for SHM. Two main challenges in SHM and their available solutions will be highlighted in the first part, followed by a discussion of issues encountered when designing a sensor system in the second part.

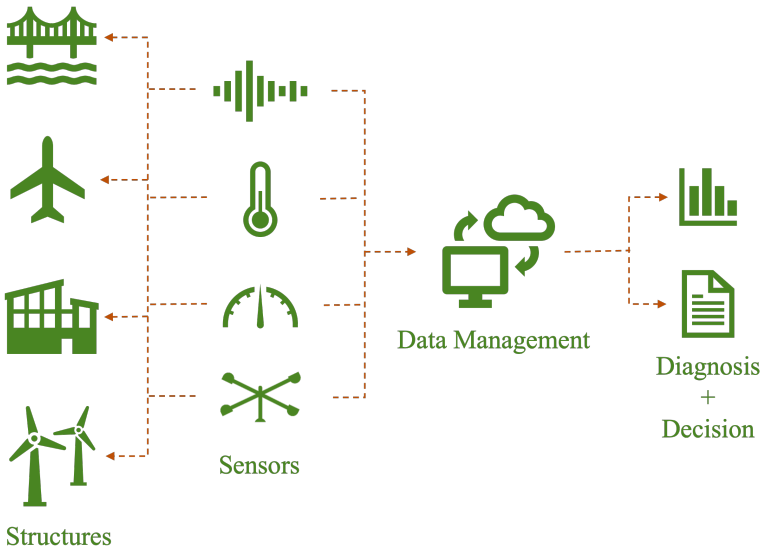


Figure 1.1: A schematic diagram of the workflow of an SHM system.

1.1 A general paradigm for SHM and its challenges

1.1.1 A statistical pattern recognition paradigm for SHM

The health state of a structure can be defined as a pattern, usually described by some features, so that the structural damage identification issue can be transferred to a quantitative pattern recognition problem. Because of various degrees of uncertainty prevailing in engineering problems, it is natural to use *a statistical approach* to give a description of a pattern. This process finally transfers the damage identification to statistical pattern recognition (SPR) for SHM.

A statistical approach here refers to one that is used either to infer the relationships within the data or to predict future values. The former goal is usually achieved by *statistical modelling*, which mainly includes the steps of finding explainable relationships between variables and the significance of those relationships. For the latter goal, *machine learning* (ML) is designed with the most accurate prediction as the main purpose. Note that predictions can only be based on interpolation and not extrapolation.

To demonstrate the above-mentioned damage identification idea more clearly, an SPR

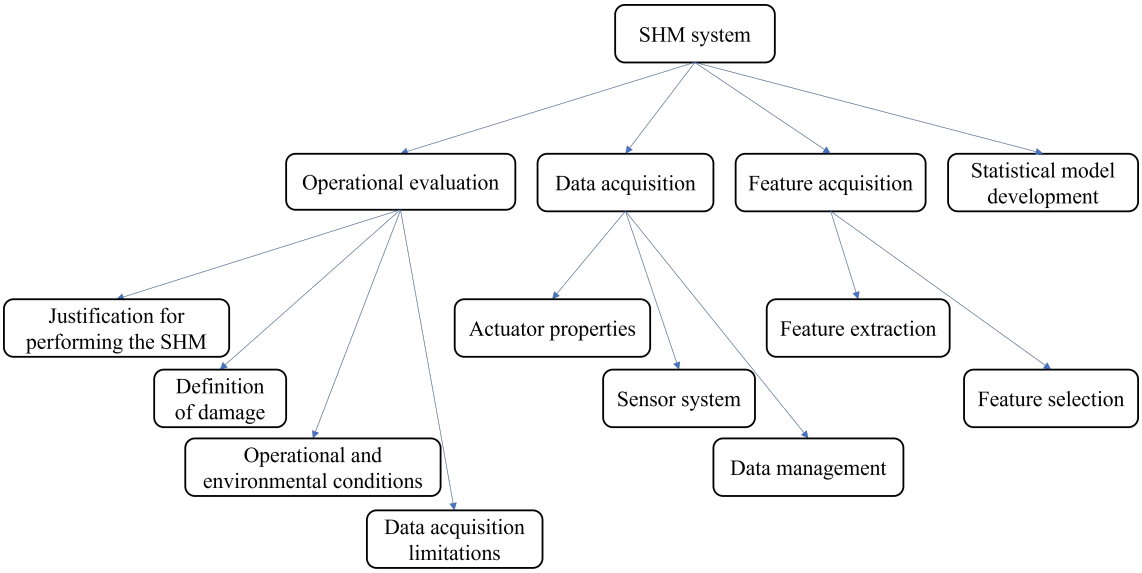


Figure 1.2: A statistical pattern recognition paradigm for an SHM system [1, 2].

paradigm for SHM, defined by Farrar and Worden [1, Ch. 1, pp. 7–14], is explained and discussed here (as shown in Figure 1.2), which contains four integrated steps: operational evaluation, data acquisition, feature acquisition and statistical model development for feature discrimination. The content of the feature acquisition step introduced here is slightly modified according to [2].

Among them, the operational evaluation step is summarised via four aspects, which are: the justification for performing the SHM, the definition of damage to be detected, the operational and environmental conditions, and data acquisition limitations. Subsequently, the data acquisition step involves three aspects: actuator properties, the applied sensor system and data management. Moreover, feature acquisition can be achieved via *feature extraction* and *feature selection* [2]. Here, 'feature extraction' is the process of extracting candidate functional features from collected data; while 'feature selection' refers to selecting the most effective features for a specific task. The last step, statistical model development, is to operate on the extracted features to establish the relationship between the features and health states of a structure. Different approaches can provide the relationship with various degrees of interpretability.

1.1.2 A challenge caused by environmental or/and operational variations

It is not difficult to realise that each step in the SPR paradigm shown in Figure 1.2 presents some challenges. These challenges can be described by three questions: what can be measured to identify a structural state, how to measure it, and how to infer from the measured data [3]. A more comprehensive review of these challenges can be found in [1, 4–7]. Among them, the interference of operational and environmental factors is one of the main problems in the field of SHM.

Damage is typically indicated via changes in the material properties and at structural boundaries [1, Ch. 1, pp. 4–7]; these can be revealed via dynamic response properties of a structure, thus realising quantitative damage detection. However, since environmental or/and operational changes can produce similar effects as damage on the collected structural response data, distinguishing the impact of damage from the impact of condition changes is a major challenge in SHM. Many studies in the literature have focussed on this issue – often called confounding influences.

Considering the presence of environmental variables, the approaches to dealing with the confounding effects can be divided into three types. When the measurements of environmental variables are accessible, the behaviours of the structure under each environmental condition can be explicitly modelled and the relationship between them can be established, as demonstrated by [8–11]. In contrast, some approaches, such as those demonstrated in [12] and [13], can derive damage-sensitive features that are robust to environmental variations from the frequency-domain data without using environmental measurements. In addition, a cointegration technique or an approach based on a Gaussian process was also developed to remove environmental and operational trends from monitored time-domain features without the need for labels [14, 15]. In the case when the environmental measurements are partially available, how to make full use of data to eliminate environmental effects was discussed in [16].

Potential strategies for dealing with the problem induced by the operational and environmental variations have been proposed, either by thinking about what can be measured [8–11, 16–18], or about how to realise the most promising inference [12–15]. Whether there should be some action to address this challenge has been almost overlooked when considering how to measure quantities in the context of structural health monitoring. The only example found in the literature is that Li

et al. [19] proposed a load-dependent sensor placement method to account for the changing load conditions, more specifically, the variation of excitation location. The possible reasons for this phenomenon will be given in Section 1.2.2.

This thesis will make an initial attempt to develop strategies that can be employed to account for the confounding influence of changing environmental and operational conditions in terms of how to measure quantities.

1.1.3 An introduction to feature selection

Another challenge in SHM belongs to the question of how to infer from measurement data, which is, more specifically, feature acquisition. As shown in Figure 1.2, there are two main steps to obtain more informative and lower-dimension features from immediate measurements, including feature extraction and feature selection. This section focuses on feature selection because the emphasis of this thesis is on SPO, and feature selection is similar to SPO for SHM based on statistical pattern recognition, in that it selects the subset of candidates most advantageous to the target realisation. Later in this thesis, some feature selection methods will be applied to develop design strategies for sensor systems.

Depending on whether output variables are used, feature selection methods can be classified as supervised or unsupervised methods [20]. *Supervised feature selection* methods are more commonly used, and can be divided into filter, wrapper and embedded methods [21]. Filter methods use a statistical score, such as correlation or dependence, to evaluate the relationship between input variables and output variables and rank the input features by these scores. For wrapper methods and embedded methods, the performance metric of an ML model is used to rank subsets of features, and a penalty for a large number of variables is involved in embedded methods simultaneously.

Unlike wrapper and embedded methods, filter methods do not involve model training, which makes them faster than the other two methods. However, filter methods are also different from the other two types of methods in that they usually evaluate a single feature at a time without considering redundancy and interactions between features [22]. The problem with not considering redundancy, is that some selected features are highly ranked but provide similar information. The problem with ignoring interaction is that some features that work well with others are ignored

because they are not favourable on their own.

There are two existing approaches to overcome the two issues mentioned above when using a filter method; one is to define a new statistical score that is penalised or compensated by feature redundancy and interaction [23–25]. The other approach is to use a criterion that has been defined, but not used in feature selection. These criteria that have been defined in the field of mathematics can deal with multiple features, so are naturally immune to redundancy and interaction issues [25, 26].

When the response variables/labels are unavailable, *unsupervised feature selection* techniques can be employed, which have attracted significant interest in different fields, such as text mining, bioinformatics, social media and intrusion detection [27]. Unsupervised feature selection strategies can also be classified into three main types: filter, wrapper, and hybrid.

Similar to supervised filter methods, unsupervised methods can also be classified into univariate and multivariate. For univariate methods, features are scored individually by an entropy-based criterion [28–30], or the correlation-based criterion [31], or by the ability to preserve the graph structure induced by the similarity of a set of pairwise instances [32, 33]. For multivariate methods, feature subsets are evaluated using information or statistical theory [34–39]. It can be seen that, similar to supervised feature-selection methods, unsupervised feature-selection strategies also require a target, which is not related to categories or responses, but to general characteristics of the data.

Moreover, with the exception of studies in [31], [34] and [39], which all used principal component analysis, the other studies were based on, or designed only for, clustering and classification. As stated in [34], variable selection can be achieved by using principal components, not only for clustering and classification tasks, but also for regression tasks. Zou *et al.* [31], selected features with non-zero coefficients for each sparse principal component. The coefficients indicate the pairwise relationship between each feature and each principal component. Cadima and Jolliffe [39] defined two correlation-based criteria to select a subset of original variables to represent the space spanned by all principal components or a subset of principal components. However, as stated in [39], searching for the optimal subset using the proposed criteria is computationally difficult.

Some excellent treatments of feature extraction and selection are in [2, 27, 40], with

Worden *et al.* [2] focusing more on the application for mechanical systems research. At this point, it can be realised that an attractive research direction is on the filter feature selection method, which can choose a bundle of features at a time but as fast as possible. In this thesis, one supervised-filter method and one unsupervised-filter method will be introduced, using an existing statistical score as a ranking criterion for feature combinations. According to the characteristics of the data, fast calculation methods of this criterion are developed to improve the calculation efficiency.

1.2 SPO for SHM literature review

1.2.1 SPO research based on a statistical model

In this section, the topic shifts from the SHM system to a part of the SHM system, the sensor system (as shown in Figure 1.2). In order for the collected data to provide more information for subsequent monitoring tasks, it is necessary to study the optimal design of a sensor system, which includes the selection of the sensor type, the number of sensors and the distribution of sensor spatial locations. The sensor type and the number of sensors are usually determined by the project budget, and the optimal distribution of sensors is usually obtained via an SPO technique. When trying to optimise sensor placement, two key aspects need to be considered, namely the optimisation algorithm and the optimisation objective function (OF). A schematic diagram of the SPO process is given in Figure 1.3, which is further explained later in this section.

A statistical model is usually established to provide a quantitative OF based on model parameters or the functionality of a model. According to the assumptions of model parameters, i.e. whether a prior is placed on the parameters, the notions of OFs can be divided into classical and Bayesian ones. Kammer [41] and Yao *et al.* [42] used a linear model to describe the relationship between sensor outputs and target mode shapes and optimised sensor combinations to maximise the estimation quality of target modal coordinates. The difference between the classical optimisation objectives adopted by [41] and [42] is the scalarisation approach to the covariance of variable errors. Further studies of mode shape-based SPO, taking into account modelling errors and measurement noise can be found in [43–45]. OFs may use quantities from information theory, such as entropy and mutual information, or may

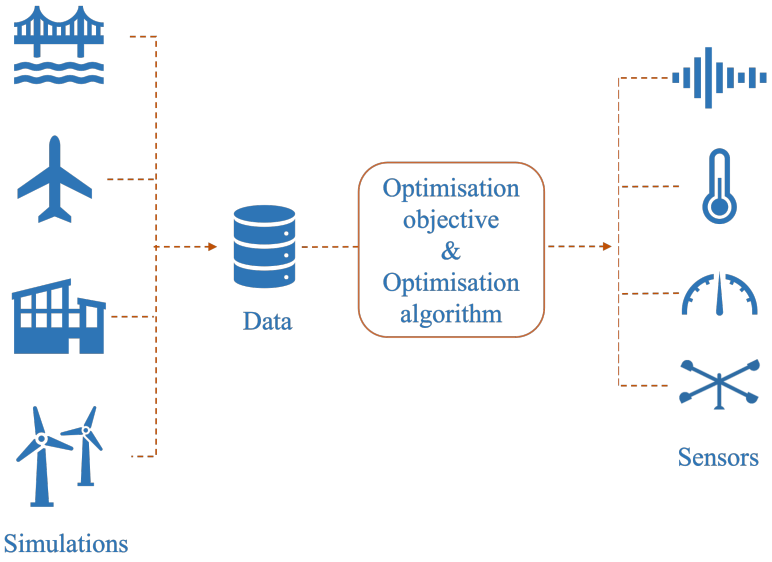


Figure 1.3: A schematic diagram of the optimal design process of a sensor system.

use concepts from Bayesian statistics. Papadimitriou [46] parameterised a structural model to describe the input-output behaviour of a structure, and information entropy was used to measure the uncertainty of the parameters updated by the measurements from a set of sensors. The sensor combination corresponding to the smallest entropy, in other words, the smallest uncertainty of the estimated parameters, was chosen as the optimal result. Krause *et al.* [47] took a model used in spatial statistic – a Gaussian process – and used mutual information to measure the effect of selected sensors on the predictive uncertainty of variables in unobserved locations, so that the prediction quality over the whole space of interest was considered. In this case, the sensor combination corresponding to the largest mutual information, or rather to the maximum reduction in the uncertainty of responses from unobserved locations, was chosen as the optimal one.

When it comes to optimisation objectives using the model functionality as the basis, most of them are related to supervised approaches, in which data and labels for different structural states are involved to train a model. Worden and Burrows [48] adopted an established supervised learning algorithm – a neural network – to rank an SHM sensor layout according to the normalised mean-square error between the desired responses and estimated responses and the probability of misclassification. The sensor deployment with the minimum weighted fitness value was treated as the optimal result. Considering the ability to classify a large amount of multivariate data, Eshghi *et al.* [49] used a Mahalanobis distance classifier to carry out a reliability-

based sensor system design to minimise the area of sensor patches while satisfying the detectability requirement for different health states. If a prior is placed on the probability of the potential outcomes of various structural states, the Bayes risk can be used as an OF to minimise the cost of each outcome, which is demonstrated in [50].

Some comprehensive references on SPO for SHM are available in [51–53]. Ostachowicz *et al.* [51] discusses the optimisation OFs corresponding to each SHM technique more systematically by dividing the SHM techniques into three types: vibration monitoring, strain monitoring, and wave propagation-based monitoring. Tan and Zhang [52] puts more effort into classifying and comparing the optimisation algorithms, which are treated as consisting of five different types. In particular, the development of evolutionary optimisation methodologies and sequential sensor placement algorithms is clearly illustrated. Barthorpe and Worden [53] provides a more complete discussion of optimisation objectives and specifies emerging trends and future directions of SHM SPO.

1.2.2 Research on the design stage of a sensor system

At the design stage of an SHM system, a physics-based model can be used to provide predictions of dynamic response of the structure in the different health states of interest, as shown in Figure 1.3. To make the predictions from a model more reliable, approaches incorporating the uncertainties of input parameters into model simulation have received a great deal of interest in the research literature [54]. The most intuitive way to implement uncertainty propagation is via Monte Carlo sampling [55]. However, for large-scale models, the computational cost of the process can be prohibitive. Some surrogate modelling methods have been developed to overcome this limit, such as the polynomial chaos expansion, Gaussian process regression (also referred to as Kriging) and neural networks [56–58].

Based on the above research, some studies on SPO technologies by using features derived from model predictions have been carried out. Guratzsch and Mahadevan [59] proposed a probabilistic finite-element model analysis to provide measurements for the subsequent SPO process. The spatial and temporal distribution of model parameters was considered via a random process or random field. Monte Carlo sampling was adopted to transfer the uncertainty of input variables to outputs. Castro-Triguero

et al. [60], also used Monte Carlo simulation to realise the propagation of parametric uncertainties, but focussed on investigating the influence of these uncertainties on the SPO methodologies based on the Fisher information matrix and energy matrix rank. Eshghi *et al.* [49], developed a reliability-based method to optimise the sensor network design, considering the uncertainty of structural geometry and input-force amplitude. Kriging combined with Latin hypercube sampling was applied to reduce the computational cost of generating training and testing data for cases with different health states, sensor locations and sensor sizes. It can be found that the focus of current research on sensor-system design is on how to consider the model uncertainties in optimisation and to find the influence of these uncertainties. The study of system-assessment criteria for the design process has currently received little attention.

The reason for this phenomenon is likely related to the researchers' low confidence in the accuracy of the data obtained during the design phase. This may also explain why actions to address problems caused by operational and environmental variations were all but ignored when considering how to measure quantities of interest. If the data were not trusted, no further processing would be conducted based on those data.

As said before, an accurate simulation model plays a key role in providing the data at the design stage of an SHM system. Model verification, model validation, and uncertainty quantification to account for errors and uncertainties are inevitable for a reliable simulation. A comprehensive survey of these three model-based activities can be found in [61]. This review compared and summarised the advantages and disadvantages of different approaches to these three activities, and provided a method to support approach selection. For more information on model verification, validation and uncertainty quantification, interested readers can refer to [55, 62, 63].

It can be seen that reliable simulation data can be obtained, at least for some simple simulations. Marks *et al.* [18] illustrated and discussed the feasibility of using simulations to provide data for SPO. Furthermore, data availability for optimisation largely depends on the degree to which extrapolation is required. This is because a significant challenge for obtaining data from a verified and validated simulation is the high extrapolation uncertainties caused by the difference between experimental and actual conditions [61]. In this thesis, data availability is divided into four levels to consider the situation of different projects, which will be explicitly presented in Chapter 6.

For some projects that cannot obtain reliable simulation data but have a large amount of measured data, the SPO strategies proposed in this thesis can also be used to select data to build a more compact model and help increase the possibility of real-time detection. Nowadays, due to the current price of sensors, it is not uncommon to find hundreds of sensors on civil infrastructure with terabytes of collected data, especially for long-span bridges [64, 65].

1.2.3 SPO considering sensor redundancy or sensor failure

The study of another factor – sensor failure – that affects the performance of damage identification will be discussed next. Different from the solutions to dealing with effects of environmental variability that are usually carried out during thinking about what can be measured, or about how to realise the promising inference (as described in Section 1.1.2), the effects of sensor failure are usually considered when determining how to measure quantities, especially during sensor system optimisation steps. The following discussion on SPO involving sensor failure is based on two general types of SHM sensing systems: wired and wireless systems.

For wired sensor networks, Side *et al.* [66] developed a novel optimisation idea for fail-safe sensor distributions. By maximising the worst performance of $N-1$-sensor distributions generated by a N-sensor distribution, this optimisation approach can provide a sensor set that can maintain an acceptable level of damage identification performance after a sensor failure. However, the OF was only composed of the *child* fitness, which led to a conflict between the fail-safe *mother* distribution and optimal mother distribution.

For wireless sensor networks, Bhuiyan *et al.* [67] proposed a backup or redundant sensor placement algorithm by incorporating the requirements of SHM and the limitations of wireless sensor networks to ensure a given level of fault tolerance. The case study showed that if the effects of sensor failures were not handled properly, it is meaningless to apply wireless sensor networks to SHM.

Note that, for wireless sensor networks, sensor nodes within a sensor field will cooperate along designed routes via wireless connections, to transmit data to the base station; therefore, the failure of one sensor may affect the acquisition of data from other sensors. More information about wireless sensors and the wireless SPO can be found in [68, 69]. One of the main differences between assumptions about these

two types of sensor networks is whether the failures of each sensor are independent of each other. If they are dependent, as in the wireless sensor network, except the effects of sensor failures, solutions to deal with failures, such as placing backup sensors, must be considered in the optimisation process, as demonstrated by [67].

There is some other SPO research involving sensor failure. Kullaa [15] tried to discriminate the variability caused by sensor faults, environmental or operational changes or structural damage. Kullaa [70] provided an approach to detect, identify the type of and quantify the faulty sensor. Note that at most, one sensor is assumed to be failed in [70]. Furthermore, before the identification and quantification of the failed sensor, the information from the faulty sensor should be ignored. As can be seen, it is attractive and useful to consider sensor failures when designing a sensor system. In this thesis, some research will be done in this regard.

1.3 Summary

This chapter first introduced the general concept of SHM based on statistical approaches. An SPR paradigm for SHM was summarised to show the components used to achieve monitoring goals. In particular, two challenges in SHM were described, including one caused by environmental or operational changes, and the other arising from the requirement of informative features to indicate the structural state. It turns out that strategies for tackling the first challenge are either to consider what can be measured or how to infer from the data collected. When considering how to collect data, whether some strategies can be employed to cope with this challenge should be considered as well. For the second challenge, selecting a subset of features at a time, which can account for redundancy and interactions between features, in a computationally efficient manner is an appealing research direction.

The second main content of this chapter is the study of SPO based on statistical models. As part of the optimal design of a sensor system, SPO consists of two main components: the optimisation algorithm and the optimisation objective. A review, particularly on the optimisation objective, was given. Furthermore, two issues in the SPO area were mentioned and discussed. First, there is a lack of research on the optimisation objective (or the assessment criterion) considering the characteristics of SPO for SHM. One possible reason for this phenomenon was given, which was a lack of confidence in the accuracy of the data obtained from the

simulations. Regarding this concern, on the one hand, model verification, validation, and uncertainty quantification can be used to improve the reliability of data at the design stage. On the other hand, assessment criteria can be designed according to the data availability level of different projects. The second issue mentioned is how to consider the impact of sensor failure in SPO. It is also worth investigating whether there is a need to consider solutions for handling failed sensors during the design process.

Chapter 2

SPO robust to environmental variations and sensor failures

This chapter mainly consists of the motivation for the research, the research aim and the research objectives. The structure of the whole thesis is introduced at the end of the chapter to highlight the content of chapters offering research novelties on robust SPO in the context of SHM.

2.1 Motivations for this research

From the previous chapter, it can be seen that numerous studies have been conducted to establish a statistical model to provide a quantitative OF based on model parameters or the functionality of a model. According to the assumptions of model parameters, the OFs can be probabilistic or non-probabilistic [41–47]. Still, most of them are not directly related to the damage identification performance of an SHM system. When it comes to optimisation objectives based on model functionality, most of them are related to supervised methods, which involve data of different structural states and corresponding labels to train a model [48–50]. In this case, when a large number of models with hyperparameters need to be trained, computational feasibility will become a non-trivial issue. Therefore, there is a demand to segment the optimisation process and establish the assessment criteria appropriate to the systematic design of a sensor system for SHM, which can balance the speed of sensor-system

design while considering the detection performance of the designed sensor system.

Damage is typically indicated via changes in the material properties and at structural boundaries [1]; these can be revealed via dynamic response properties of a structure, thus realising quantitative global damage detection. However, there are often *confounding effects*, caused by variations in the environmental and operational conditions, which can mask the changes from the actual damage. Therefore, a critical step in a damage detection method is to attempt to identify features which are sensitive to variations of material and geometric properties of the structure, but robust to environmental disturbances. Currently, there are some studies on methods to filter out the environmental effects [8–16]. However, very few studies in the literature consider confounding effects as part of the SPO process and compare the effects of robust feature extraction techniques on the optimisation results. Besides, the failure of a sensor may also lead to the mis-identification of damage and significantly affect the performance of the entire system. Therefore, it is very important to study optimal sensor placement considering the possibility of sensor failure.

2.2 Scope and objectives

The existing SPO research based on statistical models can be divided into classical optimisation design and Bayesian optimisation design according to the type of OF. The scope of this research is within the classical optimisation design. At the design stage of a sensor system, most of the available data for all states of interest come from a validated model of a structure. However, this research does not involve the issue of how to acquire an accurate model to provide the required data sets but focuses on how to use the data from a validated model or feasible tests. This means that the SPO technology studied in this thesis can also be applied to select the most critical sensor data among a large amount of collected data, thereby speeding up data processing, which is helpful for the realisation of real-time monitoring. The distribution of sensors collecting the initial dataset can be determined by insights and engineering experience, such as placing them at the anti-nodes to maximise the amplitudes of vibrations.

This thesis aims to propose a general SPO framework for SHM to systematically design an effective sensor system that is robust to environmental changes and sensor failures. The specific objectives are as follows:

1) Investigate SPO techniques and corresponding optimisation objectives systematically for three data types commonly used in the SPO process: time-domain data, frequency-domain data, and modal parameters.
2) Analyse the robustness of SPO techniques to environmental variations and evaluate whether it is necessary to take account of these environmental factors as part of the optimisation process.
3) Develop strategies for considering environmental variations in the optimum design of the sensor system, and assess the effectiveness of the developed approaches.
4) Modify the optimisation objectives to take the sensor failure into account and evaluate the impact of the modified objectives by analysing the performance of the sensor system when all sensors are working normally and after any sensor fails.

2.3 Organisation of this thesis

Chapter 3 introduces the concepts of machine learning, two supervised machine learning algorithms, and two linear techniques for dealing with confounding effects.

Chapter 4 shows two types of optimisation algorithms, including deterministic and stochastic algorithms, and five commonly used optimisation objectives to consider the different performance requirements of sensor systems.

Chapter 5 provides information on the experimental setup and design. The data sets for the three data types collected are also introduced, which will be used to demonstrate and validate the approaches developed in this thesis.

Chapter 6 suggests a correlation-based optimisation criterion which is able to evaluate a feature set. Eight real-world datasets from open data platforms are used to demonstrate the effectiveness of the criterion. According to the characteristics of the feature set, fast calculation methods of this criterion are developed to improve the calculation efficiency.

Chapter 7 establishes a hierarchical evaluation system related to machine learning algorithms to design a sensor system and puts forward corresponding evaluation indicators suitable for different stages of the design process. A numerical case study of a plate is used to show the design process.

Chapter 8 investigates the robustness of SPO techniques to environmental changes.

Strategies to improve the robustness of SPO results are provided in situations where the technique is susceptible to such changes.

Chapter 9 studies SPO considering sensor failures and improves the existing fail-safe SPO strategy. In addition, a solution for handling failed sensors is considered, and the value of the applied solution is assessed.

Chapter 10 concludes the thesis and identifies the benefits and limitations of this study and topics for future work.

Chapter 3

SHM via machine learning

This chapter will focus on existing research into classical machine learning models used in the development of an SHM statistical framework and methods to eliminate the effects of condition changes in SHM tasks.

The basic theories of two supervised-learning algorithms are explicitly presented, namely linear discriminant analysis (LDA) and support vector machines (SVM), to facilitate structural health-state discrimination in the following chapters. In addition, canonical correlation analysis (CCA) is introduced, and the relationship between the LDA and CCA is explored to demonstrate the advantages of CCA-based criteria. A study of the relationship between SVM and CCA can be found in [71].

Linear techniques to filter out such environmental effects exist, are simple to apply in practice and are computationally efficient [13]. Therefore, the second half of this chapter introduces two existing linear techniques from a theoretical viewpoint to deal with the environmental and operational variations to make an SHM system more effective, and then discusses their connections. These two linear techniques will also be used in later chapters.

3.1 Machine learning algorithms for statistical pattern recognition

When it is difficult to explicitly define a statistical model between the selected features and structural states, ML techniques can be adapted to derive the statistical model from the empirical data. One theoretical basis of ML is the *statistical learning theory*, which is a field at the intersection of statistics, functional analysis and computer science. Simply speaking, ML is the study of computer algorithms to find a predictive function based on data.

There are four main types of machine learning: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Supervised learning algorithms build a model using data containing both the input data and the corresponding labels to deal with regression and classification problems. In contrast, the data without labels are used in unsupervised learning algorithms to explore the underlying patterns to cluster the data sets. Learning algorithms combined with data having partially missing labels refer to semi-supervised learning. Reinforcement learning is a learning process that interacts with feedback from the environment to maximise the expected reward.

The focus of this thesis is on supervised learning. The general process of supervised learning includes two significant steps (refer to Figure 3.1): model creation and model prediction. The model-creation process of supervised learning can be divided into three stages: training, validation and testing.

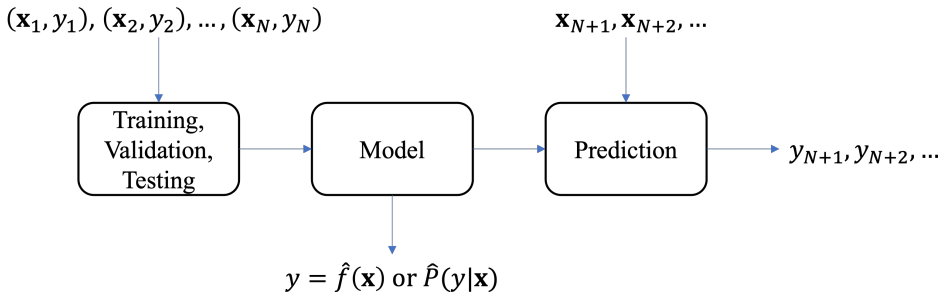


Figure 3.1: Schematic representation of the supervised learning process.

In the process of model creation, input and output data will be expressed here by letters $\mathbf{X}$ and $\mathbf{y}$. One instance of the n input variables in input space χ is usually a

vector, written as,

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top \tag{3.1}$$

where $\top$ indicates transpose. The vectors in this thesis are all column vectors. One instance of the output variable in output space γ is represented by y, which is usually a scalar. Note that x_i in equation (3.1) is different from $\mathbf{x}_i, i = 1, 2, \ldots, N$, in Figure 3.1. $\mathbf{x}_i$ represent the i^{th} observation of the n input variables, given by,

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})^\top \tag{3.2}$$

One basic assumption of supervised learning is that there is a joint probability distribution between random variables $\mathbf{x}$ and y. The model obtained by supervised learning can be a non-probabilistic model or a probabilistic model, represented by a decision function $y = \hat{f}(\mathbf{x})$ or a conditional probability distribution $\hat{P}(y|\mathbf{x})$, depending on the specific learning algorithm. Correspondingly, in the prediction process, the output y_{N+1} corresponding to the input $\mathbf{x}_{N+1}$ can be determined by the model $y_{N+1} = \hat{f}(\mathbf{x}_{N+1})$ or $y_{N+1} = \text{argmax}_{y_{N+1}} \hat{P}(y_{N+1}|\mathbf{x}_{N+1})$.

3.1.1 Linear discriminant analysis

To explain the basic idea of LDA, the demonstration is limited to binary class discrimination. Furthermore, since LDA seeks to maximise the scalar projection of the distance between different classes, the samples will be projected onto one dimension.

For LDA, the data of each category can be described by the sample mean and sample covariance, that is,

$$p(\mathbf{X}_1) \sim \mathcal{N}(\bar{\mathbf{x}}_1, \boldsymbol{\Sigma}_1), \quad p(\mathbf{X}_2) \sim \mathcal{N}(\bar{\mathbf{x}}_2, \boldsymbol{\Sigma}_2) \tag{3.3}$$

where $\mathbf{X}_1 \in \mathbb{R}^{N_1 \times n}$, $\mathbf{X}_2 \in \mathbb{R}^{N_2 \times n}$ and $n \geq 2$. N_1 and N_2 are the number of observations in each class. $\bar{\mathbf{x}}_1 \in \mathbb{R}^{n \times 1}$ and $\bar{\mathbf{x}}_2 \in \mathbb{R}^{n \times 1}$ are the sample mean vectors of class 1 and 2, and $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the two sample covariance matrices.

A projection direction $\mathbf{w}$ for the samples in these two classes is expected to be found such that the projected distributions have minimal overlap, which is designed to be achieved by maximising the ratio of the between-class scatter $\mathbf{S}_\text{b}$ and within-class

scatter $\mathbf{S}_\mathrm{w}$ [72]. The objective function is,

$$F(\mathbf{w}) = \max_{\mathbf{w}\in\mathbb{R}^{n\times 1}} \frac{\mathbf{w}^\top \mathbf{S}_\mathrm{b}\mathbf{w}}{\mathbf{w}^\top \mathbf{S}_\mathrm{w}\mathbf{w}} \tag{3.4}$$

where $\mathbf{S}_\mathrm{w}$ and $\mathbf{S}_\mathrm{b}$ are given by,

$$\mathbf{S}_\mathrm{b} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top \tag{3.5}$$

and,

$$\mathbf{S}_\mathrm{w} = \frac{(N1-1)}{(N1+N2-2)} \times \mathbf{\Sigma}_1 + \frac{(N2-1)}{(N1+N2-2)} \times \mathbf{\Sigma}_2 \tag{3.6}$$

This problem can be seen that the objective function $F(\mathbf{w})$ is invariant under rescaling of $\mathbf{w}$, so the denominator in equation (3.4) can be assumed to be equal to unity, i.e. $\mathbf{w}^\top \mathbf{S}_\mathrm{w}\mathbf{w} = 1$. The objective function (3.4) then becomes,

$$\max_{\mathbf{w}\in\mathbb{R}^{n\times 1}} \mathbf{w}^\top \mathbf{S}_\mathrm{b}\mathbf{w} \tag{3.7a}$$

$$\text{s.t.} \quad \mathbf{w}^\top \mathbf{S}_\mathrm{w}\mathbf{w} = 1 \tag{3.7b}$$

It can be solved by using a Lagrange multiplier λ, such that,

$$F(\mathbf{w}) = \mathbf{w}^\top \mathbf{S}_\mathrm{b}\mathbf{w} - \lambda(\mathbf{w}^\top \mathbf{S}_\mathrm{w}\mathbf{w} - 1) \tag{3.8}$$

followed by,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{w}}\big(F(\mathbf{w})\big) = 2\mathbf{S}_\mathrm{b}\mathbf{w} - 2\lambda\mathbf{S}_\mathrm{w}\mathbf{w} = \mathbf{0} \tag{3.9}$$

Hence,

$$(\mathbf{S}_\mathrm{w}^{-1}\mathbf{S}_\mathrm{b} - \lambda\mathbf{I})\mathbf{w} = \mathbf{0} \tag{3.10}$$

and thus λ is an eigenvalue of $\mathbf{S}_\mathrm{w}^{-1}\mathbf{S}_\mathrm{b}$, and $\mathbf{w}$ is the corresponding eigenvector. According to Equations (3.9) and (3.7b), it is obvious that,

$$F(\mathbf{w}) = \max_{\mathbf{w}\in\mathbb{R}^{n\times 1}} \mathbf{w}^\top \mathbf{S}_\mathrm{b}\mathbf{w} = \mathbf{w}^\top \lambda\mathbf{S}_\mathrm{w}\mathbf{w} = \lambda \tag{3.11}$$

Note that the above derivation requires the existence of $\mathbf{S}_\mathrm{w}^{-1}$. A necessary condition for $\mathbf{S}_\mathrm{w}$ to be invertible is N_1, $N_2 \geq (n+1)$. For the case when $\mathbf{S}_\mathrm{w}$ is not invertible, please refer to [73, Ch. 16, pp. 363–364] for modification details. Furthermore, for this case, since the rank of $\mathbf{S}_\mathrm{w}^{-1}\mathbf{S}_\mathrm{b}$ is one, the number of eigenvalues is only one.

When there are more than two classes, the between-class scatter $\mathbf{S}_\mathrm{b}$ and within-class scatter $\mathbf{S}_\mathrm{w}$ for C classes can be defined as [74, 75],

$$\mathbf{S}_\mathrm{b} \equiv \sum_{c=1}^{C} N_c \times (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})(\bar{\mathbf{x}}_c - \bar{\mathbf{x}})^\top \tag{3.12}$$

and,

$$\mathbf{S}_\mathrm{w} \equiv \sum_{c=1}^{C} N_c \times \mathbf{\Sigma}_c \tag{3.13}$$

where $\bar{\mathbf{x}}_c$ and $\mathbf{\Sigma}_c$ are the mean vector and covariance matrix of $\mathbf{X}_c$ respectively, and $\bar{\mathbf{x}}$ is the mean of the whole data set $[\mathbf{X}_1; \cdots ; \mathbf{X}_C]$. In this condition, the number of the eigenvalues will be $\min(n, (C-1))$.

The eigenvalue of $\mathbf{S}_\mathrm{w}^{-1}\mathbf{S}_\mathrm{b}$ giving $F(\mathbf{w})$ is the largest eigenvalue, and the corresponding eigenvector can be used to obtain the linear boundary. For a new observation, $\mathbf{x}_{N+1}$, based on the distances between each $\bar{\mathbf{y}}_c = \mathbf{w}^\top \bar{\mathbf{x}}_c$ and $\mathbf{y}_{N+1} = \mathbf{w}^\top \mathbf{x}_{N+1}$, it can be assigned to one of C classes corresponding to the smallest distance.

3.1.2 Connection between linear discriminant analysis and canonical correlation analysis

A brief introduction to CCA is given first, to facilitate the comparison of LDA and CCA. CCA aims to find the maximal linear association between two co-occurring multivariate random variables with N observations, which are collected as matrices $\mathbf{X} \in \mathbb{R}^{N \times n}$ and $\mathbf{Y} \in \mathbb{R}^{N \times m}$ [76]. n and m are the number of variables in $\mathbf{X}$ and $\mathbf{Y}$ respectively.

CCA can be carried out by finding pairs of projection directions $\mathbf{U} \in \mathbb{R}^{n \times L}$ and $\mathbf{V} \in \mathbb{R}^{m \times L}$, so that the Pearson's correlation coefficient between $(\mathbf{X} - \overline{\mathbf{X}})\mathbf{u}$ and $(\mathbf{Y} - \overline{\mathbf{Y}})\mathbf{v}$ for a pair of projection directions is maximised [77], i.e.

$$\begin{aligned} F(\mathbf{u}, \mathbf{v}) &= \max_{\mathbf{u},\mathbf{v}} r(\mathbf{X}\mathbf{u}, \mathbf{Y}\mathbf{v}) \\ &= \max_{\mathbf{u},\mathbf{v}} \frac{\mathbf{u}^\top \mathbf{X}_\mathrm{M}^\top \mathbf{Y}_\mathrm{M} \mathbf{v}}{\sqrt{\mathbf{u}^\top \mathbf{X}_\mathrm{M}^\top \mathbf{X}_\mathrm{M} \mathbf{u}} \sqrt{\mathbf{v}^\top \mathbf{Y}_\mathrm{M}^\top \mathbf{Y}_\mathrm{M} \mathbf{v}}} \end{aligned} \tag{3.14}$$

where $\mathbf{X}_\mathrm{M} = \mathbf{X} - \overline{\mathbf{X}}$ and $\mathbf{Y}_\mathrm{M} = \mathbf{Y} - \overline{\mathbf{Y}}$. $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are the means of columns of $\mathbf{X}$ and $\mathbf{Y}$ respectively. Note that $r(\cdot)$ refers to the *Pearson's correlation coefficient*.

It can be seen that $F(\mathbf{u}, \mathbf{v})$ does not depend on the scale of $\mathbf{u}$ and $\mathbf{v}$. Therefore, the $\mathbf{X}_{\mathrm{M}}\mathbf{u}$ and $\mathbf{Y}_{\mathrm{M}}\mathbf{v}$ can be standardised to have unit variance. Then, equation (3.14) can be expressed as,

$$\max_{\mathbf{u},\mathbf{v}} \mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v}, \quad \text{s.t. } \mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\mathbf{u} = 1,\ \mathbf{v}^{\top}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} = 1 \tag{3.15}$$

Lagrange multipliers α and β can be used to find $\mathbf{u}$ and $\mathbf{v}$, such that,

$$F(\mathbf{u}, \mathbf{v}) = \mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} - \frac{1}{2}\alpha(\mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\mathbf{u} - 1) - \frac{1}{2}\beta(\mathbf{v}^{\top}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} - 1) \tag{3.16}$$

followed by,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}}\big(F(\mathbf{u}, \mathbf{v})\big) = \mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} - \alpha\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\mathbf{u} = \mathbf{0} \tag{3.17}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{v}}\big(F(\mathbf{u}, \mathbf{v})\big) = \mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{u} - \beta\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} = \mathbf{0} \tag{3.18}$$

Premultiplying equations (3.17) and (3.18) by $\mathbf{u}^{\top}$ and $\mathbf{v}^{\top}$ respectively, gives

$$\mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} - \alpha\mathbf{u}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\mathbf{u} = \mathbf{0} \tag{3.19}$$

$$\mathbf{v}^{\top}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{u} - \beta\mathbf{v}^{\top}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\mathbf{v} = \mathbf{0} \tag{3.20}$$

It can be found that $F(\mathbf{u}, \mathbf{v}) = \alpha = \beta$.

By a series of transformations based on equations (3.17) and (3.18), the values of $F(\mathbf{u}, \mathbf{v})$ and the corresponding projection directions $\mathbf{u} \in \mathbb{R}^{n\times 1}$ and $\mathbf{v} \in \mathbb{R}^{m\times 1}$ can be obtained by solving the eigenvalue equations given by [78, 79],

$$\left[(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}} - F(\mathbf{u}, \mathbf{v})^2\mathbf{I}\right]\mathbf{u} = \mathbf{0} \tag{3.21a}$$

$$\left[(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}} - F(\mathbf{u}, \mathbf{v})^2\mathbf{I}\right]\mathbf{v} = \mathbf{0} \tag{3.21b}$$

Thus, $F(\mathbf{u}, \mathbf{v})$ is the square root of an eigenvalue of $(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}$ or $(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}$, which is known as the *canonical correlation coefficient*. If $\mathbf{X}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{M}}$ have full column rank, the number of the non-zero eigenvalues is not greater than $\min(n, m)$, that is, L is not greater than $\min(n, m)$.

In order to explore the connection between LDA and CCA, the description and notation of data need to be unified and clarified. $\mathbf{X} \in \mathbb{R}^{N\times n}$ is equal to $[\mathbf{X}_1; \cdots; \mathbf{X}_C]$, and $\mathbf{Y} \in \mathbb{R}^{N\times m}$ is the set of labels, represented by m variables, for all classes. In

addition, dummy coding should be adopted here to encode the label matrix $\mathbf{Y}$. Note that the column rank of the matrix is reduced by one after each column subtracts its column mean. This is the case because the sum of all matrix columns is now equal to 0, which means that one column can be rewritten as a linear combination of all other columns. Dummy encoding uses $C-1$ features to represent C labels/classes, which can make the label matrix $\mathbf{Y}$ remain full column rank after subtracting its column mean to satisfy the definition of CCA. Thus, the number of variables m in $\mathbf{Y}$ equals $C-1$.

Equations (3.10) and (3.21a) can be rewritten as,

$$\mathbf{S}_{\mathrm{b}}\mathbf{w} = F(\mathbf{w})\mathbf{S}_{\mathrm{w}}\mathbf{w} \tag{3.22}$$

and

$$\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\mathbf{u} = \mathbf{F}(\mathbf{u},\mathbf{v})^2(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})\mathbf{u} \tag{3.23}$$

where $\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}$ is the total scatter matrix and equals $\mathbf{S}_{\mathrm{b}}+\mathbf{S}_{\mathrm{w}}$, and according to [80], $\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}$ is equal to $\mathbf{S}_{\mathrm{b}}$. Therefore, equation (3.23) can now be written as,

$$\mathbf{S}_{\mathrm{b}}\mathbf{u} = \frac{F(\mathbf{u},\mathbf{v})^2}{1-F(\mathbf{u},\mathbf{v})^2}\mathbf{S}_{\mathrm{w}}\mathbf{u} \tag{3.24}$$

where $F(\mathbf{u},\mathbf{v})^2/(1-F(\mathbf{u},\mathbf{v})^2)$ increases monotonically for $F(\mathbf{u},\mathbf{v}) \in (0,1)$, which means the largest $F(\mathbf{w})$ corresponds to the largest $F(\mathbf{u},\mathbf{v})$.

By comparing Equations(3.22) and (3.24), it is reasonable to say that the two analysis methods are equivalent when a scalar projection is expected and labels are encoded using dummy encoder. It can also be seen that CCA generalises LDA to consider multiple projected dimensions to define the linear discriminant.

3.1.3 Support vector machines

This section also uses binary classification to demonstrate the basic idea of the SVM algorithm. A decision function will be provided by an SVM. The inputs for a linear SVM training can be represented by,

$$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\} \tag{3.25a}$$

$$\mathbf{x}_i \in \boldsymbol{\chi} = \mathbb{R}^n \quad i = 1, 2, \ldots, N \tag{3.25b}$$

$$y_i \in \gamma = \{+1, -1\} \tag{3.25c}$$

where N is the number of training examples, $\mathbf{x}_i$ is the i^{th} feature vector, $(\mathbf{x}_i, y_i)$ is the i^{th} training sample and y_i is either 1 or -1, which indicates the class to which the $\mathbf{x}_i$ belongs.

A separating hyperplane can be described by,

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{3.26}$$

where $\mathbf{w}$ is the normal vector to a hyperplane. The parameter $b/||\mathbf{w}||$ determines the offset of the hyperplane from the origin along the normal vector $\mathbf{w}$, as shown in Figure 3.2.

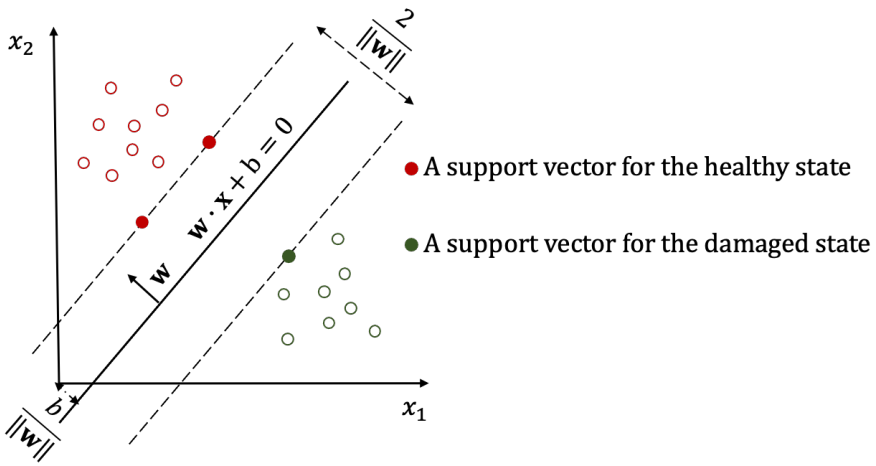


Figure 3.2: The maximum-margin hyperplane for an SVM.

Finding the best hyperplane means maximising the smallest signed distance. The smallest signed distance from a sample point $(\mathbf{x}_i, y_i)$ to a hyperplane can be calculated by,

$$\gamma_i = y_i \left(\frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x}_i + \frac{b}{||\mathbf{w}||} \right) \tag{3.27a}$$

$$\gamma = \min_{i=1,\ldots,N} \gamma_i \tag{3.27b}$$

Maximising γ gives the optimisation problem,

$$\max_{\mathbf{w},b} \gamma \tag{3.28a}$$

$$\text{s.t.} \quad y_i \left(\frac{\mathbf{w}}{||\mathbf{w}||} \cdot \mathbf{x}_i + \frac{b}{||\mathbf{w}||} \right) \geq \gamma, \quad i = 1, 2, \ldots, N \tag{3.28b}$$

This constraint means that the signed distance from any feature vector in the training set to the max-margin hyperplane is at least equal to γ. An important consequence of this constraint is that the max-margin hyperplane is completely determined by those feature vectors that lie on the margins. These feature vectors are called *support vectors*. However, changing the values of $\mathbf{w}$ and b proportionally will not change the hyperplane. Therefore, when the ratio of $\mathbf{w}$ and b is constant, $\mathbf{w}$ can be selected according to the demand. For the convenience of calculation, the $||\mathbf{w}||$ can be set to $1/\gamma$. In this case, equation ((3.28)) can be rewritten as,

$$\max_{\mathbf{w},b} \frac{1}{||\mathbf{w}||} \tag{3.29a}$$

$$\text{s.t.} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \quad i = 1, 2, \ldots, N \tag{3.29b}$$

As shown in Figure 3.2, a maximum margin width is thus equal to,

$$D_{\text{margin}} = \frac{2}{||\mathbf{w}||} \tag{3.30}$$

It is worth noting that there is no direct relationship between SVM and CCA, but referring to the idea of SVM, [71] developed a method to assign soft labels to samples in each class to improve the classification performance using features selected by CCA.

3.2 Approaches to confounding-effect elimination

3.2.1 Principal component analysis

The first approach introduced here is principal component analysis (PCA), a classical dimension-reduction technique from multivariate data analysis. The theory behind PCA is briefly recalled bellow to illustrate why it can be used to create damage-sensitive features using data collected under changing environmental conditions.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ denote the original feature matrix with n variables, and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]$ denotes the derived multivariate matrix after performing PCA on $\mathbf{X}$, such that,

$$\mathbf{y}_j = a_{1j}\mathbf{x}_1 + a_{2j}\mathbf{x}_2 + \cdots + a_{nj}\mathbf{x}_n, \quad j = 1, \ldots, n \tag{3.31a}$$

$$\mathbb{E}(\mathbf{y}_i, \mathbf{y}_j) = \mathbb{E}(\mathbf{y}_i)\mathbb{E}(\mathbf{y}_j), \quad i \neq j \tag{3.31b}$$

$$\sigma^2_{\mathbf{y}_1} \geq \sigma^2_{\mathbf{y}_2} \geq \ldots \geq \sigma^2_{\mathbf{y}_n} \tag{3.31c}$$

where $\mathbf{x}_j \in \mathbb{R}^{N\times 1}$ is the j^{th} feature vector and $\mathbf{y}_j \in \mathbb{R}^{N\times 1}$ is the j^{th} principal component, which is a linear combination of the original feature vectors. It can be seen that the first p derived $\mathbf{y}$ vectors with $p < n$ can be used to represent original n variables in $\mathbf{X}$ while maintaining the most variation in the original data.

Note that to distinguish the vector of an observation for all features ($\mathbf{x}_i$) from the vector of a feature with N observations, $\mathbf{x}_j$ is used here to facilitate the comparison of PCA and another linear technique in the next section.

Let $\mathbf{a}_j = (a_{1j}, \cdots, a_{nj})^\top$. The first principal component $\mathbf{y}_1 = \mathbf{X}\mathbf{a}_1$ can be obtained by choosing $\mathbf{a}_1$ with the scaling $\mathbf{a}_1^\top \mathbf{a}_1 = 1$ to maximise the variance of $\mathbf{y}_1$, which is equivalent to,

$$V(\mathbf{a}_1) = \max \sigma^2_{\mathbf{y}_1} = \max \sigma^2_{\mathbf{X}\mathbf{a}_1} = \max_{\mathbf{a}_1 \in \mathbb{R}^{n\times 1}} \mathbf{a}_1^\top \mathbf{\Sigma}_{\mathbf{X}} \mathbf{a}_1, \quad \text{s.t. } \mathbf{a}_1^\top \mathbf{a}_1 = 1 \tag{3.32}$$

By following a similar procedure used to solve equation (3.7), one gets,

$$V(\mathbf{a}_1) = \mathbf{a}_1^\top \boldsymbol{\Sigma}_\mathbf{X} \mathbf{a}_1 - \lambda(\mathbf{a}_1^\top \mathbf{a}_1 - 1) \tag{3.33}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{a_1}}\big(V(\mathbf{a}_1)\big) = 2\boldsymbol{\Sigma}_\mathbf{X}\mathbf{a}_1 - 2\lambda\mathbf{a}_1 = \mathbf{0} \tag{3.34}$$

Hence,

$$(\boldsymbol{\Sigma}_\mathbf{X} - \lambda\mathbf{I})\mathbf{a}_1 = \mathbf{0} \tag{3.35}$$

This equation means that, except for $\mathbf{a}_1 = \mathbf{0}$, a solution is available if and only if λ is an eigenvalue of $\boldsymbol{\Sigma}_\mathbf{X}$. Thus, equation (3.32) becomes,

$$V(\mathbf{a}_1) = \max \mathbf{a}_1^\top \lambda \mathbf{I} \mathbf{a}_1 = \max \lambda \mathbf{a}_1^\top \mathbf{a}_1 = \max \lambda \tag{3.36}$$

For the positive semi-definite matrix $\boldsymbol{\Sigma}_\mathbf{X}$, there are n non-negative eigenvalues, labelled as $\lambda_1, \lambda_2, \ldots, \lambda_n$, where $\lambda_1 > \lambda_2 > \ldots > \lambda_n \geq 0$. Thus, the maximum variance of $\mathbf{y}_1$ is the largest eigenvalue λ_1 and $\mathbf{a}_1$ is the corresponding eigenvector of $\boldsymbol{\Sigma}_\mathbf{X}$.

Intuitively, the maximum variance of the second principal component $\mathbf{y}_2$ should equal the second largest eigenvalue of $\boldsymbol{\Sigma}_\mathbf{X}$. But a concise theoretical demonstration of finding $\mathbf{y}_2$ is also given here to explain this prevalent dimension reduction technique more clearly.

Similar to the first principal component, $\mathbf{y}_2$ is equal to $\mathbf{X}\mathbf{a}_2$, with $\mathbf{a}_2^\top \mathbf{a}_2 = 1$. From equation (3.31b), the covariance of $\mathbf{y}_1$ and $\mathbf{y}_2$ should be zero, that is,

$$cov_{\mathbf{y}_1,\mathbf{y}_2} = cov(\mathbf{X}\mathbf{a}_2, \mathbf{X}\mathbf{a}_1) = \mathbf{a}_2^\top \boldsymbol{\Sigma}_\mathbf{X} \mathbf{a}_1 = \lambda_1 \mathbf{a}_2^\top \mathbf{a}_1 = 0 \tag{3.37}$$

which works if and only if $\mathbf{a}_2^\top \mathbf{a}_1 = 0$. Now with two constraints,

$$V(\mathbf{a}_2) = \mathbf{a}_2^\top \boldsymbol{\Sigma}_\mathbf{X} \mathbf{a}_2 - \lambda(\mathbf{a}_2^\top \mathbf{a}_2 - 1) - \delta(\mathbf{a}_2^\top \mathbf{a}_1 - 0) \tag{3.38}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{a_2}}\big(V(\mathbf{a}_2)\big) = 2\boldsymbol{\Sigma}_\mathbf{X}\mathbf{a}_2 - 2\lambda\mathbf{a}_2 - \delta\mathbf{a}_1 = \mathbf{0} \tag{3.39}$$

Premultiplication of equation (3.39) by $\mathbf{a}_1^\top$ results in $\delta = 0$. The last equation then becomes,

$$(\boldsymbol{\Sigma}_\mathbf{X} - \lambda\mathbf{I})\mathbf{a}_2 = \mathbf{0} \tag{3.40}$$

which means that the maximum variance of $\mathbf{y}_2$ is also an eigenvalue of $\boldsymbol{\Sigma}$. According to equation (3.31c), it should be the second-largest eigenvalue λ_2 and $\mathbf{a}_2$ is the

corresponding eigenvector. The remaining principal components can be obtained by following the same analysis pattern.

Let $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$, where $\mathbf{a}_i$ are the eigenvectors of $\mathbf{\Sigma_X}$. Then the relationship between $\mathbf{Y}$ and $\mathbf{X}$ can be written as,

$$\mathbf{Y} = \mathbf{XA}, \quad \text{s.t. } \mathbf{\Sigma_Y} = \mathbf{A}^\top \mathbf{\Sigma_X} \mathbf{A} \tag{3.41}$$

In addition, the relationship between the sum of variances of $\mathbf{Y}$ and the sum of variances of $\mathbf{X}$ is as follows,

$$\sum_{j=1}^{n} \sigma_{\mathbf{y}_j}^2 = \sum_{j=1}^{n} \lambda_j = \mathrm{tr}(\mathbf{A}^\top \mathbf{\Sigma_X} \mathbf{A}) = \mathrm{tr}(\mathbf{\Sigma_X} \mathbf{A}\mathbf{A}^\top) = \mathrm{tr}(\mathbf{\Sigma_X}) = \sum_{j=1}^{n} \sigma_{\mathbf{x}_j}^2 \tag{3.42}$$

Therefore, $\sum_{j=1}^{p} \lambda_j / \sum_{j=1}^{n} \lambda_j$ can represent the proportion of the total variance of the original variables captured by the first p principal components.

When deriving features that are insensitive to environmental changes but sensitive to structural changes (including material changes and geometrical changes), the essence of this method is to identify a linear subspace of a transformed feature space to account for the environmental conditions. The remaining subspace of the transformed feature vectors then makes contributions to the damage sensitivity, which will be explained more clearly in Section 3.2.3.

3.2.2 Mahalanobis squared-distance

The second linear method to deal with environmental effects is based on the Mahalanobis squared-distance (MSD). It can naturally eliminate temperature effects as long as the normal condition data include measurements under an appropriate range of environmental conditions. The MSD is given by,

$$D_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{\Sigma}_\mathbf{X}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \tag{3.43}$$

where $\mathbf{x}_i$ is the i^{th} row of a feature matrix, which refers to the i^{th} observation of n variables; $\bar{\mathbf{x}}$ is the mean value vector for a set of observations $\mathbf{X}$ of n variables and $\mathbf{\Sigma}$ is the corresponding covariance matrix. Note that $\mathbf{x}_i$ and $\bar{\mathbf{x}}$ are column vectors because, as specified in Section 3.1, all vectors are written as column vectors.

Following the idea of PCA, variable vectors are going to be represented by their principal components, whose covariance matrix is diagonal; this helps to interpret the composition of the MSD. Firstly, the eigenvalue decomposition of the covariance matrix $\mathbf{\Sigma_X}$ is given by,

$$\mathbf{\Sigma_X} = \mathbf{A\Lambda A}^{-1} \tag{3.44}$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and since $\mathbf{A}$ is orthogonal here, $\mathbf{A}^\top\mathbf{A} = \mathbf{I}$. The following transformation is performed,

$$\mathbf{y}_i = \mathbf{A}^\top\mathbf{x}_i \tag{3.45}$$

and by inverting this transformation with $\mathbf{A}^\top = \mathbf{A}^{-1}$,

$$\mathbf{x}_i = \mathbf{A}\mathbf{y}_i \tag{3.46}$$

The MSD in equation (3.43) is then calculated by,

$$D_i^2 = (\mathbf{y}_i - \bar{\mathbf{y}})^\top\mathbf{A}^\top\mathbf{\Sigma}_\mathbf{X}^{-1}\mathbf{A}(\mathbf{y}_i - \bar{\mathbf{y}}) = (\mathbf{y}_i - \bar{\mathbf{y}})^\top\mathbf{\Lambda}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}}) \tag{3.47}$$

Since $\lambda_j = \sigma_{\mathbf{y}_j}^2$, equation (3.47) can be written as,

$$D_i^2 = \sum_{j=1}^{n} \frac{1}{\sigma_{\mathbf{y}_j}^2}(y_{ij} - \bar{y}_j)^2 \tag{3.48}$$

where y_{ij} is the i^{th} observation of the j^{th} transformed variable (or the j^{th} principal component) and $\bar{y}_j$ is the mean value of the j^{th} transformed variable.

Here it is shown that the MSD is the sum of the contributions of independent variables transformed from the original variables. The contributions are scaled by dividing the variances of the associated new variables.

3.2.3 Connection between Mahalanobis squared-distance and principal component analysis

By comparing Equations (3.31a) and (3.45), it is clear that the calculation process to obtain an element in a transformed variable instance $\mathbf{y}_i$ is the same as the calculation process to obtain an element in a principal component vector $\mathbf{y}_j$; Elements in both vectors can thus be called principal component scores. If it is assumed that most of

the variance in the n healthy condition features is from environmental effects, the first p components can be used to explain it.

For a new instance of the transformed feature variables $\mathbf{y}_{N+i}$, the MSD can be decomposed into two parts:

$$D_{N+i}^2 = \sum_{j=1}^{p} \frac{1}{\sigma_{\mathbf{y}_j}^2}(y_{(N+i)j} - \bar{y}_j)^2 + \sum_{j=p+1}^{n} \frac{1}{\sigma_{\mathbf{y}_j}^2}(y_{(N+i)j} - \bar{y}_j)^2 \tag{3.49}$$

This equation shows that this distance is insensitive to environmental changes because the variances of transformed variables mainly responsible for environmental conditions are large, which makes the contributions of the corresponding variables to the distance calculation small.

Suppose PCA is firstly used to eliminate the environmental effects, and the MSD is subsequently used to derive the features. In that case, the first p components accounting for the environmental conditions will be abandoned. Then, the MSD for $\mathbf{y}_{N+i}$ reduces to computing,

$$D_{N+i}^2 = \sum_{j=p+1}^{n} \frac{1}{\sigma_{\mathbf{y}_j}^2}(y_{(N+i)j} - \bar{y}_j)^2 \tag{3.50}$$

It can be seen that if $\sum_{j=1}^{p} \lambda_j / \sum_{j=1}^{n} \lambda_j$ is larger than a certain threshold, say 99.9%, the result of the MSD calculation is equivalent to the result of the MSD of features projected onto the subspace of the principal components. Furthermore, both linear techniques can be used in situations where temperature measurements corresponding to feature vectors are not available. For more details, please refer to [13].

3.3 Summary

This chapter began with an introduction to the main role of machine learning and its main types, which could be used for SHM based on statistical pattern recognition. Among them, supervised learning algorithms can provide decision functions or conditional probability distributions to predict output values for new input values by using previous input and output data. A linear boundary or a hyperplane is obtained for LDA and linear SVM to separate the different classes. Note that,

strictly speaking, the LDA and CCA (or canonical variate analysis) mentioned in this chapter are not supervised learning algorithms, but supervised linear dimension reduction techniques. More information can be found in [73, Ch. 16, 17]. The equivalence of LDA and CCA shows that, for classification problems, the larger the first canonical correlation coefficient, the larger the distance between different classes in the projection direction.

Furthermore, two existing approaches, PCA and MSD, developed to eliminate confounding effects were explained and compared in this chapter. They are practical methods for extracting effective features. The similarity of them is that both approaches are valid under the premise that a linear subspace exists to account for changing environmental or operational conditions. They are different in that, by discarding that subspace, PCA can provide damage-sensitive features, while the MSD technique uses the variances of the variables in that subspace to reduce the effects of changing conditions.

Chapter 4

SPO based on statistical models

The previous chapter discussed how to obtain damage-sensitive features for SHM and how to use machine learning techniques to describe different structural state patterns. In this chapter, how to design a sensor system to collect information-intensive data for SHM is investigated, including optimisation strategies and optimisation objectives for SPO. Those applied later in this thesis are explained in detail.

4.1 Optimisation algorithms

Since the SPO here does not address the case of describing sensor locations in continuous space, selecting the optimal sensor combination can be regarded as a combinatorial optimisation of an OF. The number of combinations S for selecting I sensors from Z candidate sensors in different locations is given by,

$$S = {}_{Z}C_{I} = \frac{Z!}{I!(Z-I)!} \tag{4.1}$$

On this basis, this section will introduce two types of optimisation strategies, including deterministic and stochastic, to find the optimal sensor combination that leads to the minimum or maximum value of an OF. Numerous studies have explained and compared the specific characteristics of various optimisation algorithms. However, the following review is intended to help readers clearly understand the framework of this field and where different algorithms fit within the framework without going into

too much detail.

4.1.1 Deterministic optimisation algorithms

Exhaustive search, also known as brute-force search, is a deterministic optimisation algorithm that examines all possibilities, to find a result that satisfies all statements. This algorithm can indeed provide a global solution to an optimisation problem and be used as a benchmark when evaluating other algorithms. However, according to equation (4.1), S can be prohibitively large; for example, when it is applied to a finite element model with even a relatively small number of degrees of freedom. Therefore, there will be some compromises for the optimal solution; a sequential sensor placement strategy is another option to provide a deterministic solution.

A sequential sensor-placement strategy can be achieved by adding candidate sensors to the selected sensor group or removing sensors from the candidate sensor group. The former one is known as a *greedy search*. This search strategy is computationally efficient, but it is susceptible to local optima. For more detailed discussions regarding the possible approaches to mitigate this problem, please refer to [81–83]. Here, they are collectively referred to as two-stage methods.

4.1.2 Stochastic optimisation algorithms

It is common to say that *stochastic search* refers to optimisation algorithms that involve randomness in the measurements of the OF and/or in the search direction [84, Ch. 1, p. 7]. Note that 'randomness' in stochastic optimisation algorithms does not mean that the optimal results are entirely random, but rather that some decisions made in the search process contain some degree of randomness.

'Metaheuristics' constitute a significant subfield of stochastic optimisation. Like heuristics, metaheuristics do not guarantee an optimal global solution, but they use randomness to increase the possibility of escaping the local optima. Furthermore, given the elusive definition of metaheuristics in many papers, here is a quote from [85] to clarify the definition of metaheuristics and its relationship to heuristics:

> In its general sense, a metaheuristic is not an algorithm. Rather, it is a consistent set of ideas, concepts, and operators that can be used to

design heuristic optimisation algorithms.

There are three representative metaheuristics, which are simulated annealing [86], evolutionary computation and ant colony optimisation [87]. See [88] and [89, Ch. 1, 2 and 8] for the comparison and pseudocodes of these three techniques. Evolutionary computation is the collective name of evolutionary programming, evolutionary strategies and *genetic algorithms* (GA), and the last two are both algorithms. Because a GA will be adopted to handle SPO in this thesis, a brief introduction to this algorithm is given as follows.

The GA is a computational approach that simulates the natural selection of chromosomes with high efficiency. It starts from an initial population consisting of several random chromosomes representing the possible solutions. A *fitness function* is used to quantify the adaptation ability of a chromosome. Through a series of operations, including selection, crossover and mutation, new chromosomes will then be produced. Chromosomes in the updated population that can make the fitness function value close to the target limit will be selected to generate the next population. This process will repeat until some termination conditions are met. Refer to [89] for more details on the GA.

4.2 Optimisation objectives

Another non-trivial step for the SPO technique is to select an appropriate optimisation objective to evaluate the performance of a sensor system. A quantitative OF based on model parameters or model roles is usually adopted by establishing a statistical model. It is worth noting that other common names for the OF are loss function, fitness function or performance assessment criterion.

In this section, five commonly used optimisation objectives, also called assessment criteria, are discussed: the effective independence (EI), the determinant of Fisher information matrix (DFIM), the average driving point residue (ADPR), the root-mean-squared error (RMSE) and the parameters or derivations of a confusion matrix.

4.2.1 Effective independence

One trace-based criterion used for the SPO based on mode shapes is referred to as *effective independence* (EI), which transfers the accurate estimation of the target modal coordinates into the requirement of the linear independence of mode shapes. Assume that there are Z candidate sensor locations and J modes to be identified. Suppose I sensors will be selected from the Z candidate sensors. For the s^{th} combination of the I sensor locations, the outputs of these chosen sensors are described by a vector $\mathbf{y} \in \mathbb{R}^{I\times 1}$, which is given by,

$$\mathbf{y}_s = \mathbf{\Phi}_s \mathbf{q} + \boldsymbol{\epsilon}_s \tag{4.2}$$

where $\mathbf{\Phi}_s \in \mathbb{R}^{I\times J}$ is the s^{th} target modal matrix, $\mathbf{q} \in \mathbb{R}^{J\times 1}$ is the vector of the target modal coordinates and $\boldsymbol{\epsilon}_s \in \mathbb{R}^{I\times 1}$ is the vector of measurement noise at these selected sensors, which is assumed to be zero-mean Gaussian noise and of equal variance σ^2. The least-squares estimate of the target modal coordinates for the s^{th} sensor combination can be expressed as [41],

$$\hat{\mathbf{q}}_s = (\mathbf{\Phi}_s^\top \mathbf{\Phi}_s)^{-1} \mathbf{\Phi}_s^\top \mathbf{y}_s \tag{4.3}$$

The covariance matrix of the estimation error of the $\hat{\mathbf{q}}_\mathbf{s}$ is given by [41],

$$\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}} = \sigma^2 (\mathbf{\Phi}_s^\top \mathbf{\Phi}_s)^{-1} \tag{4.4}$$

where $\mathbf{\Phi}_s^\top \mathbf{\Phi}_s$ is termed the *Fisher information matrix* (FIM). According to the Gauss–Markov theorem [90], minimisation of this covariance matrix can help to obtain the best linear unbiased estimator of $\hat{\mathbf{q}}_s$. Referring to equation (4.4), this minimisation can be realised by maximising the FIM. Considering the contribution of each sensor location, the FIM can be written as,

$$\mathbf{\Phi}_s^\top \mathbf{\Phi}_s = \sum_{i=1}^{I} \boldsymbol{\phi}_s^{i\top} \boldsymbol{\phi}_s^i \tag{4.5}$$

where $\boldsymbol{\phi}_s^i$ is the i^{th} row of the target modal matrix $\mathbf{\Phi}_s$, corresponding to the i^{th} sensor location. It can be seen that each sensor location contributes some information to the FIM. In order to quantitatively evaluate the contribution of each location, the

following eigenvalue equation can be solved,

$$(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s - \lambda_s \mathbf{I})\boldsymbol{\Psi}_s = \mathbf{0} \tag{4.6}$$

where $\mathbf{I}$ is an identity matrix. A relationship between the sensor locations and each eigenvalue can then be acquired by calculating the square of $\boldsymbol{\phi}_s^i$ in terms of the basis formed by the columns $\boldsymbol{\Psi}_s$, which can be expressed as,

$$\mathbf{G} = (\boldsymbol{\Phi}_s \boldsymbol{\Psi}_s) \otimes (\boldsymbol{\Phi}_s \boldsymbol{\Psi}_s) \tag{4.7}$$

where the symbol $\otimes$ represents the element-wise product and the summation of the j^{th} column of $\mathbf{G}$ is equal to the j^{th} eigenvalue of $\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s$. To make it possible to compare the contribution of each sensor location to the identification of J selected modes, the matrix $\mathbf{G}$ is right-multiplied by the inverse of the diagonal matrix $\boldsymbol{\lambda}_s = \boldsymbol{\Psi}_s^\top (\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)\boldsymbol{\Psi}_s$ to rescale the columns, and then written as,

$$\mathbf{F}_{\mathrm{E}} = (\boldsymbol{\Phi}_s \boldsymbol{\Psi}_s) \otimes (\boldsymbol{\Phi}_s \boldsymbol{\Psi}_s)\boldsymbol{\lambda}_s^{-1} \tag{4.8}$$

The sum of each column of $\mathbf{F}_{\mathrm{E}}$ is equal to 1, which means that the contribution of one sensor location to identifying any selected mode can be quantitatively evaluated on the same scale, that is, within a range from 0 to 1. Thus, the sum of the i^{th} row can be used to evaluate the contribution of the i^{th} sensor location to the identification of J mode shapes.

On this basis, the vector composed of the sum of each row of $\mathbf{F}_{\mathrm{E}}$ can be defined as the independence distribution vector $\mathbf{E}_{\mathrm{d}}$, which can be formulated as the diagonal of an idempotent matrix,

$$\begin{aligned} \mathbf{E} &= \boldsymbol{\Phi}_s \boldsymbol{\Psi}_s \boldsymbol{\lambda}_s^{-1} \boldsymbol{\Psi}_s^\top \boldsymbol{\Phi}_s^\top \\ &= \boldsymbol{\Phi}_s (\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)^{-1} \boldsymbol{\Phi}_s^\top \end{aligned} \tag{4.9}$$

where $\mathbf{E}$ is known as the projection matrix. Because the trace of an idempotent matrix is equal to its rank [91], the diagonal terms of the projection matrix $\mathbf{E}$ represent the contributions of the corresponding sensor locations to the linear independence of its columns or the mode shapes.

The values in the independence distribution vector $\mathbf{E}_{\mathrm{d}}$ corresponding to sensor locations will be updated for each iteration of the SPO, which can indicate the importance of a sensor location to the independent information. Locations with

small values in $\mathbf{E}_{\mathrm{d}}$ can be removed at each iteration until a desired number of sensors are left. In this thesis, only one sensor is removed by default at each iteration.

For each iterative calculation of $\mathbf{E}$ with a different number of sensor locations, the trace of this matrix is constant, equal to the number of linearly independent columns in it. Furthermore, if target mode shapes with the candidate degrees of freedom are linearly independent, the trace of the matrix $\mathbf{E}$ is thus always equal to the number of target mode shapes. Based on this, the trace-based criterion cannot be used to compare the different sensor combinations with the same number of sensors to search out the globally optimal results.

4.2.2 Determinant of Fisher information matrix

Another criterion used for the SPO based on mode shapes is called the *determinant of the Fisher information matrix* (DFIM), which is directly related to the accuracy of the target modal-coordinate estimate. The brief derivation process is as follows.

According to [42], it can be assumed that the probability density function of $\hat{\mathbf{q}}_s$ in equation (4.3) is Gaussian, and the corresponding confidence region can be treated as the interior of the hyperellipsoid,

$$(\hat{\mathbf{q}}_s - \mathbf{q})^{\top} \mathbf{F}_{\hat{\mathbf{q}}_s}^{-1} (\hat{\mathbf{q}}_s - \mathbf{q}) = c^2 \tag{4.10}$$

where $\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}} = (\mathbf{\Phi}_s^{\top}\mathbf{\Phi}_s)^{-1}\mathbf{\Phi}_s^{\top}\mathbf{E}(\boldsymbol{\epsilon}_s\boldsymbol{\epsilon}_s^{\top})\mathbf{\Phi}_s(\mathbf{\Phi}_s^{\top}\mathbf{\Phi}_s)^{-1}$ is the covariance matrix of the estimation error of $\hat{\mathbf{q}}_s$ and c is a constant related to the confidence level. The volume D_s of the hyperellipsoid can be obtained as [92],

$$D_s = \pi^{I/2} c^I [\det(\mathbf{F}_{\hat{\mathbf{q}}_s})]^{1/2} [\mathbf{\Gamma}(I/2+1)]^{-1} \tag{4.11}$$

where $\mathbf{\Gamma}(\cdot)$ refers to the gamma function and $\det(\cdot)$ refers to the determinant. To obtain the best linear unbiased estimator, that is, the most accurate estimation of the target modal coordinates, the volume of this confidence region should be minimised. Because $\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}}$ is the only variable in equation (4.11), the key to the optimisation problem is to minimise the determinant of $\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}}$. As it was assumed that the measurement noise of sensors was uncorrelated and of equal variance σ^2 in [42],

they showed that,

$$\begin{aligned} \det(\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}}) &= \det\left[\sigma^2(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)^{-1}\right] \\ &= \sigma^{2n} \det\left[(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)\right]^{-1} \end{aligned} \tag{4.12}$$

Because the ${\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}}}^{-1}$ was defined as the Fisher information matrix (FIM) in [93], $\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s$ was subsequently referred to as the FIM in [41]. Thus, minimising the determinant of $\mathbf{F}_{\hat{\mathbf{q}}_\mathbf{s}}$, as discussed earlier, is equivalent to maximising the DFIM, i.e. maximising $\det(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)$. Note that when the lower–upper decomposition is applied, the computational complexity of calculating the DFIM is $\mathcal{O}(J^3)$.

Based on the above introduction, this FIM-based criterion can be applied to compare the different sensor combinations with the same number of sensors. It is beneficial to find the optimal sensor deployment beneficial to estimate the target modal coordinates most accurately.

4.2.3 Average driving-point residue

There is an obvious disadvantage in using the values in the independence distribution vector $\mathbf{E}_\mathrm{d}$ or DFIM as an optimisation objective, in that sensor locations with low signal strength can be selected. In order to overcome this issue and improve the signal-to-noise ratio, the *average driving-point residue* (ADPR) can be combined with these two criteria [94].

For the EI method, the vector $\mathbf{E}_\mathrm{d}^{\mathrm{EI-DPR}}$ can be defined as in [95]. One entry in $\mathbf{E}_\mathrm{d}^{\mathrm{EI-DPR}}$ for the i^{th} degree of freedom is given by,

$$E_{\mathrm{d}_i}^{\mathrm{EI-DPR}} = E_{\mathrm{d}_i} \sum_{j=1}^{J} \frac{(\phi_s^{ij})^2}{\omega_j} \tag{4.13}$$

where E_{d_i} is the i^{th} entry in $\mathbf{E}_\mathrm{d}$. ϕ_s^{ij} is the entry corresponding to the i^{th} degree of freedom of the j^{th} target mode in the s^{th} target modal matrix. ω_j is the natural frequency corresponding to the j^{th} target mode.

Similarly, for the DFIM, the $\det(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)_{\mathrm{ADRP}}$ corresponding to one sensor combination

can be constructed, which is defined as,

$$\det(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s)_{\text{ADRP}} = \det(\boldsymbol{\Phi}_s^\top \boldsymbol{\Phi}_s) \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\phi_s^{ij})^2}{\omega_j} \tag{4.14}$$

Similar to the EI criterion, the EI-ADPR can help find sub-optimal results by sequentially removing the sensors that contribute the least. Another criterion, the DFIM-ADPR, should be maximised to find the optimal sensor deployment.

4.2.4 Root-mean-squared error

The *root-mean-squared error* (RMSE) is a common metric for assessing the quality of regression, which is a supervised machine-learning technique to predict a continuous variable based on several independent variables. It works by measuring the average squared difference between the predicted value $\hat{y}_n$ and the observed value y_n, which is given by,

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{N}} \tag{4.15}$$

where N is the number of data points. It can be found that the closer the RMSE is to zero, the better the quality of the regression model.

To facilitate the comparison between data sets with different scales, the RMSE can be normalised by using various characteristics of the data set, such as mean, standard derivation and interquartile range. *R-squared* or the coefficient of determination is a standardised version of the RMSE, which is defined as [96],

$$R^2 = 1 - \frac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \overline{y})^2} \tag{4.16}$$

where $\overline{y}$ is the mean of the observed y_n. The relationship between the RMSE and R-squared is,

$$R^2 = 1 - \frac{RMSE^2}{\Sigma^2} \tag{4.17}$$

where Σ is the standard derivation of the observed y_n. It can be seen that when R-squared is equal to one, the regression model fits the data perfectly with the corresponding RMSE equalling zero. For the training data set, R^2 ranges from 0 to 1, but it will be negative when the RMSE is greater than Σ for the test data set.

From Figure 4.1, R-squared can also be defined as,

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} = \frac{\sum_{n=1}^{N}(\hat{y}_n - \overline{y})^2}{\sum_{n=1}^{N}(y_n - \overline{y})^2} \tag{4.18}$$

This equation suggests another intuitive interpretation of R-squared, which is that R-squared represents the ratio of the variance of the observed variable captured by the regression. An R-squared value close to 1 indicates that the corresponding regression model captures most of the variance of the observed variable and can be treated as a good model.

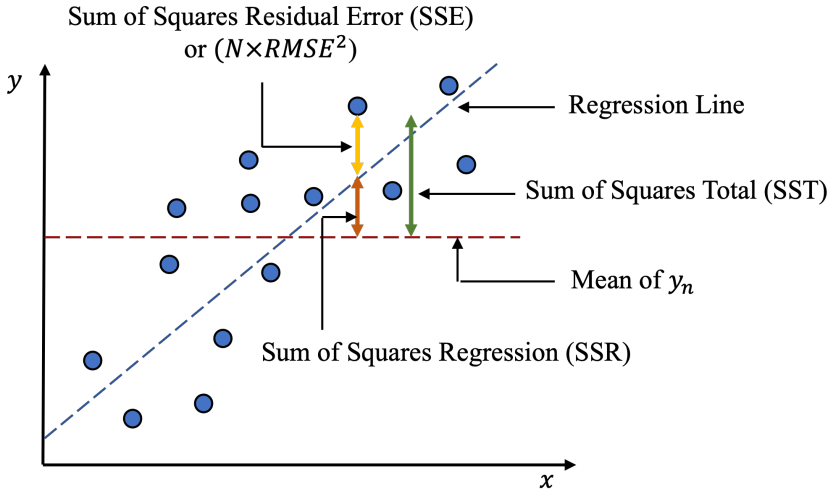


Figure 4.1: Diagrammatic representation of RMSE and R-squared.

4.2.5 Parameters or derivations of a confusion matrix

Another type of supervised-learning algorithm, called classification, is used to predict discrete class labels. To allow the intuitive consideration of the SHM system's performance on classification, the assessment criteria can be composed of parameters or derivations of the confusion matrix [97]. The parameters of the confusion matrix are demonstrated in Table 4.1.

Here are the three most commonly used derived quantities of parameters in the confusion matrix. The variables in the following equations are defined in Table 4.1. The first one is the *true positive rate* (TPR) or probability of detection, which is given by,

$$TPR = \frac{tp}{p} = 1 - \frac{fn}{p} \tag{4.19}$$

Table 4.1: Terminology in a confusion matrix.

		Measured condition	
		Positive (p)	Negative (n)
Predicted condition	Positive (pp)	True positive (tp)	False positive (fp)
	Negative (pn)	False negative (fn)	True negative (tn)

t refers to true; f refers to false; p refers to positive; n refers to negative

the second one is the *false positive rate* (FPR) or probability of false alarm, given by,

$$FPR = \frac{fp}{n} = 1 - \frac{tn}{n} \tag{4.20}$$

and the last one is called *accuracy* (ACC), which can be calculated by,

$$ACC = \frac{tp + tn}{tp + fn + fp + tn} = \frac{tp + tn}{p + n} \tag{4.21}$$

There are some other commonly used quantities capable of assessing classification performance, such as the area under the curve, the precision and the F-score, which can be chosen according to the characteristics of the data. For example, precision, the proportion of positive results correctly classified, should be adopted when studying rare diseases. This is because such studies usually include many more people without disease than with disease. The precision not including the number of true negatives (which should be very large) allows the results to be unaffected by unbalanced data in different categories.

4.3 Summary

This chapter has pointed out that the SPO for SHM in this thesis is framed as a combinatorial optimisation problem, and two basic components of the SPO technique are investigated, including the optimisation algorithm and the optimisation objective.

An overview of the existing algorithms used to search for optimal or sub-optimal results was presented first to clarify their definitions and relationships. Two deterministic optimisation algorithms (exhaustive search and greedy search) and one stochastic optimisation algorithm (GA), which would be applied in the following chapters, were mentioned in this part. An exhaustive search can provide global optima but may require significant computational effort when the number of candidates

is large. A greedy search usually gives locally optimal results but in an efficient way. Two-stage approaches can be applied to improve search results. A GA can improve the possibility of finding the globally optimal point but reduce the computational effort, by involving some random factors in the search process to jump out of the local optima.

Optimisation objectives or evaluation criteria were then listed to indicate which aspects of sensor system performance they were developed for, respectively. The trace-based criterion (EI) and the FIM-based criterion (DFIM) aim to optimise the accuracy of target modal coordinate estimation. Additionally, both criteria can be combined with ADPR to incorporate signal amplitudes into the target. However, EI can only rank sensors within a sensor combination, while DFIM can compare different combinations with the same number of sensors. The last introduced part was about frequently used metrics for assessing the quality of regression or classification models. Using these metrics as optimisation targets can improve the performance of a sensor system in predicting continuous or discrete variables, i.e. improving the damage identification capability of an SHM system.

Chapter 5

Glider-wing experiment

The technologies developed in this thesis will be validated on a real structure, a full wing of a glider aircraft. The detailed information on the experimental design and acquisition of data-sets is given in this chapter. The entire experiment was conducted in the Laboratory for Verification and Validation by the author, with the excellent technical assistance from Robin S. Mills.

A novel aspect of this experiment was that the glider wing was tested in a chamber where the ambient temperature around the wing could be controlled as needed. It means that real data sets, rather than simulated data sets, can be used to develop an SPO framework robust to temperature variation.

5.1 Experiment setups and design

5.1.1 Vibration test system

The components of the adopted vibration test system are shown in Figure 5.1. In this experiment, a Gaussian white-noise excitation was first generated within the Siemens LMS acquisition system, and amplified using the ETS solutions LA500 power amplifier. The generated signal then travelled out to the ETS solutions VT100 electrodynamic shaker to excite vibrations in the tested object.

The object under investigation is a glider wing under a free-free boundary condition

(shown in Figure 5.2). The wing was excited at a point between sensor 22 and sensor 23, as shown in Figure 5.3. The tip area of the wing is more flexible, and it is easier to vibrate the entire structure by setting an excitation point in this area. The vibration data of the wing were measured using PCB resonant piezoelectric accelerometers (Model 353B18) and sampled using a 64-channel acquisition system controlled by LMS software. These accelerometers are uniaxial and measure perpendicular to the surface on which they are mounted. The weight of each sensor is 1.8 gm. Figure 5.3 is a schematic showing the 36 candidate sensors used, with their labelled positions not drawn to scale. These sensors are evenly distributed on the wing structure to provide the candidate sensor-position combinations.

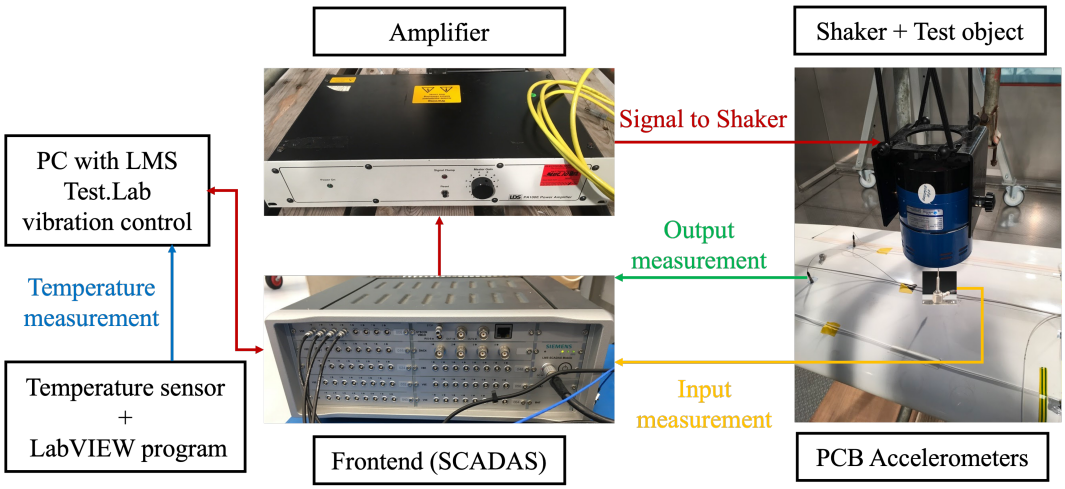


Figure 5.1: Vibration test system applied on the glider wing.

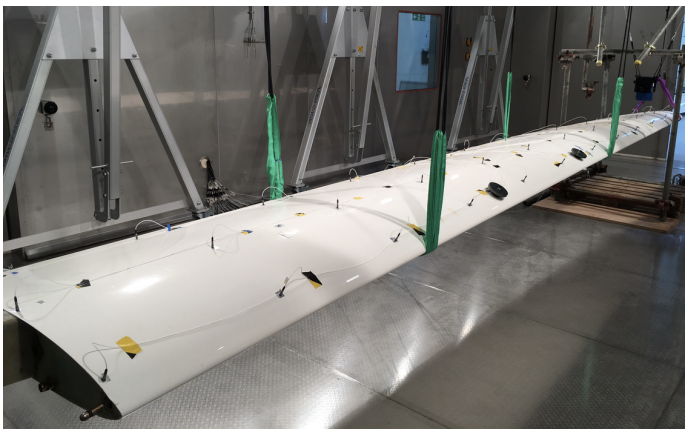


Figure 5.2: Photograph of the experiment setups in the testing chamber.

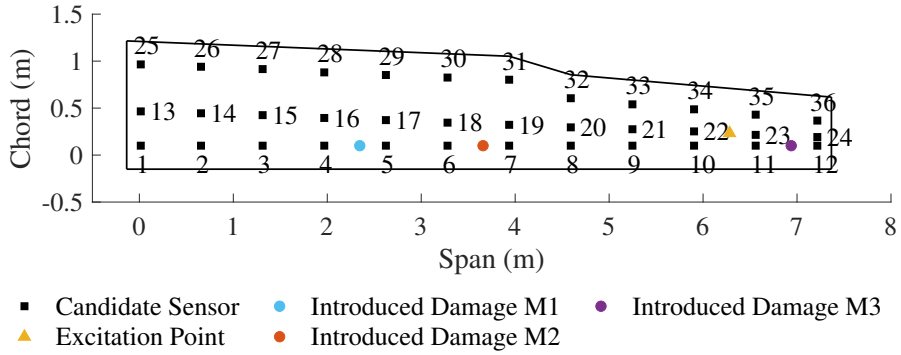


Figure 5.3: Labelled positions of significant points on the glider wing.

5.1.2 Damage scenarios

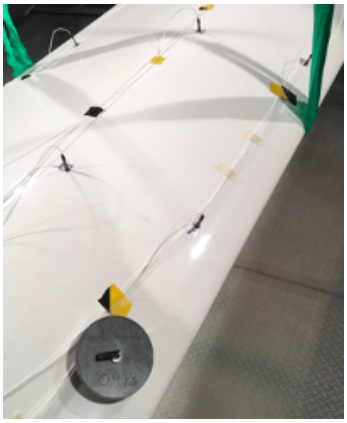
To simulate a reversible damage scenario rather than inflict permanent damage on the wing, the damage was simulated by adding mass blocks at discrete points. Two rounds of experiments were carried out for different damage sizes and conditions.

In the first-round tests, three damage cases were considered: mass addition at locations [M1], [M2] and [M1, M2]. The first mass block (M1) was added between sensors 4 and 5 and had a mass of 860 g. The second mass block (M2) was added between sensors 6 and 7 with a mass of 900 g. The locations for both mass blocks are shown in Figure 5.3. These represent quite large damage in order to produce separated feature clusters. One damage scenario and the two mass blocks that were actually used are shown in Figures 5.4a and 5.4b.

In the second-round tests, the damage of interest was simulated by adding a mass block (60g) onto the structure to provide a more challenging data set. As demonstrated in Figure 5.3, three damage cases were considered: mass addition at locations [M1], [M2], and [M3]. Figures 5.4c and 5.4d demonstrate the last two damage scenarios in practice.

5.1.3 Temperature scenarios

In the first-round tests, seven different temperatures in the chamber were controlled and recorded, ranging from 0 °C to 30 °C at intervals of 5 °C. When the temperature exceeded 30 °C, it was found that the attachment of the shaker and sensors was affected; therefore, setting a higher temperature in the chamber for this experiment



(a) M1 in Test 1

(b) M1 & M2 in Test 1

(c) M2 in Test 2

(d) M3 in Test 2

Figure 5.4: Photographs of damage cases introduced by adding mass blocks on the glider wing.

was abandoned. When the temperature is 0 °C, it is necessary to control the humidity parameters in the room to avoid damage to the experimental equipment because of freezing.

For each temperature, the wing would be immersed in the chamber for two hours after the room temperature reached that expected temperature. Because it is a non-trivial task to measure the temperature of the entire wing, the immersion strategy is a practical solution to ensure that the temperature of the structure is at the required temperature.

In the second-round tests, because of time constraints, six temperatures are considered: 0, 5, 10, 15, 20, 25 °C. During the process of immersing the gliding wing to reach the set temperature, a series of sweep tests were also implemented. An separate temperature measurement system (as shown in Figure 5.1), consisting of a temperature sensor

placed inside the wing and a LabVIEW program, was used to collect real-time temperature labels in this test.

5.2 Data acquisition and feature extraction

The extraction techniques for three commonly-used data types will be presented separately, including time-series data, frequency-domain data and modal data.

5.2.1 Time-series data

The time series of accelerations from 36 accelerometers on the wings were measured and collected for both tests, but the time series of the input force generated by the shaker were collected only in the second-round test.

The time-series data collected during the first-round test are introduced in Table 5.1. For the second-round test, the temperature fell from 25°C to 0°C with a fixed step of 5°C. This process was repeated twice. The data sets collected at each temperature during the first temperature ramp-down were two 80-second time series and 50 four-second time series for the 'healthy' structure at each temperature, as shown in the third column of Table 5.2. The data sets collected at each fixed temperature during the second cooling process are shown in Table 5.2.

While waiting for the structure to reach room temperature, one 80-second time series and corresponding temperature label were collected every three minutes. The data collected during each 5°C down sweep are shown in Table 5.3. Note that it takes 110 seconds to collect an 80-second time series, including a 30-second setup and cooldown.

5.2.2 Frequency response functions

For both tests, the adapted frequency range over which the frequency response functions (FRFs) were taken is $0 - 4096$ Hz, and the frequency resolution (df) was 0.25 Hz. Note that according to the Nyquist-Shannon sampling theorem [98], the sampling rate (f_s)(the number of data points recorded per second), should be at least two times the signal's highest frequency component, which is 4096 Hz in this

Table 5.1: Time series and corresponding labels collected in the first-round test.

Temperature label	Health state			
	Normal	M1	M2	M1 & M2
0°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
5°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
10°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
15°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
20°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
25°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s
30°C	2 × 4s	1 × 4s	1 × 4s	1 × 4s

Table 5.2: The first data set of time series and corresponding labels collected for the second round of tests.

Temperature label	Health state				
	M2	Normal	M1	Normal	M3
0°C	2 × 80s	2 × 80s	2 × 80s	2 × 80s	2 × 80s
	25 × 4s	50 × 4s	25 × 4s	25 × 4s	25 × 4s
5°C	2 × 80s	2 × 80s	2 × 80s	2 × 80s	2 × 80s
	25 × 4s	50 × 4s	25 × 4s	25 × 4s	25 × 4s
10°C	2 × 80s	2 × 80s	2 × 80s	2 × 80s	2 × 80s
	25 × 4s	50 × 4s	25 × 4s	25 × 4s	25 × 4s
15°C	2 × 80s	2 × 80s	2 × 80s	2 × 80s	2 × 80s
	25 × 4s	50 × 4s	25 × 4s	25 × 4s	25 × 4s
20°C	2 × 80s	2 × 80s	2 × 80s	-	-
	25 × 4s	50 × 4s	25 × 4s		
25°C	2 × 80s	2 × 80s	1 × 80s	-	-
	25 × 4s	50 × 4s	25 × 4s		

Table 5.3: The second data set of time series and corresponding labels collected for the second round of tests.

Temperature label	Health state			
	Normal	M1	M2	M3
25°C - 20°C	24 × 80s	-	19 × 80s	-
20°C - 15°C	23 × 80s	18 × 80s	-	-
15°C - 10°C	27 × 80s	-	27 × 80s	-
10°C - 5°C	30 × 80s	29 × 80s	-	-
5°C - 0°C	33 × 80s	-	-	37 × 80s

experiment. Thus, $f_s = 2 \times 4096 = 8192$ Hz, which means that the bit length of the analogue-to-digital converter is 13 [99, Chapter 2, p. 93–96]. Moreover, for a fixed-range converter, the signal should be amplified to cover as much of the converter's range as possible.

In the first test, each FRF is an averaged value of eight measurements, to make the obtained FRFs smoother. Two measurements of the FRF matrix were recorded for the structure under the normal condition at each controlled temperature, with only one measurement for the structure under the damaged condition at each temperature. All the FRFs collected during the first-round test are shown in Table 5.4.

Table 5.4: FRFs and corresponding labels collected in the first-round test.

Temperature label	Health state			
	Normal	M1	M2	M1 & M2
0°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
5°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
10°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
15°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
20°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
25°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.
30°C	2 × 8-avg.	1 × 8-avg.	1 × 8-avg.	1 × 8-avg.

It is worthy clarifying that, since the sampling period T to obtain a one-averaged FRF is 4 seconds, which is equal to $1/df$, the length of the time-series collected during the first-round test should be 32 seconds. However, only the last 4-second time-series data corresponding to the last one-average FRFs were recorded at that time, as shown in Table 5.1.

The reasons for not collecting a large number of observations under the same conditions were that the tests at that time were aimed at preliminarily exploring the effect of temperature on the structure, and the full experimental design had not yet been completed.

In the second test, two measurements of 20-average FRFs and 25 one-average measurements were recorded at each fixed temperature (0 °C to 15 °C, in 5 °C increments), for the normal condition and three damage conditions. During each waiting period, 20-average observations can be collected as the results of the sweep tests: refer to Tables 5.5 and 5.6 for these two data sets.

Moreover, in the second-round test, another two 20-average FRFs and 50 one-average repetitions were recorded for the normal condition at each temperature, as shown in

the third column of Table 5.5.

Table 5.5: The first data set of FRFs and corresponding labels collected for the second round of tests.

Temperature label	Health state				
	M2	Normal	M1	Normal	M3
0°C	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.	25 × 1-avg.	25 × 1-avg.
5°C	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.	25 × 1-avg.	25 × 1-avg.
10°C	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.	25 × 1-avg.	25 × 1-avg.
15°C	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.	25 × 1-avg.	25 × 1-avg.
20°C	2 × 20-avg.	2 × 20-avg.	2 × 20-avg.	-	-
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.		
25°C	2 × 20-avg.	2 × 20-avg.	1 × 20-avg.	-	-
	25 × 1-avg.	50 × 1-avg.	25 × 1-avg.		

Table 5.6: The second data set of FRFs and corresponding labels collected for the second round of tests.

Temperature label	Health state			
	Normal	M1	M2	M3
25°C - 20°C	24 × 20-avg.	-	19 × 20-avg.	-
20°C - 15°C	23 × 20-avg.	18 × 20-avg.	-	-
15°C - 10°C	27 × 20-avg.	-	27 × 20-avg.	-
10°C - 5°C	30 × 20-avg.	29 × 20-avg.	-	-
5°C - 0°C	33 × 20-avg.	-	-	37 × 20-avg.

5.2.3 Mode shape and natural frequency

For experimental modal analysis, the modal parameters, such as mode shape and natural frequency, can be extracted from an FRF. This process is called modal curvefitting or modal parameter estimation [99, Ch. 4, pp. 237–251].

Considering the effect of noise, only the mode shapes and natural frequencies corresponding to the 8-average and 20-average time series were extracted by LMS Test.Lab. See the high-average FRFs in Tables 5.4-5.6 to get an idea of the obtainable modal parameter data sets in this experiment.

The natural frequencies of the normal-state wing at each temperature, corresponding to the 20-average FRFs in the third column of Table 5.5, are given in Table 5.7 as an

example, to give an intuitive idea about the effect of temperature on this structure.

Table 5.7: Natural frequencies and corresponding labels for the wing under the normal condition.

Temperature label	Structural mode					
	1st (Hz)	2nd (Hz)	3rd (Hz)	4th (Hz)	5th (Hz)	6th (Hz)
0°C	8.8360	21.6167	37.9651	40.8269	61.4164	68.3324
5°C	8.8577	21.6000	37.9793	40.7647	61.3753	68.3252
10°C	8.7696	21.4963	37.6829	40.5435	60.7310	68.1490
15°C	8.5974	21.1780	36.9880	40.0681	59.2559	67.2463
20°C	8.5249	20.9834	36.3059	39.5515	58.7107	66.6458
25°C	8.4342	20.7875	35.5758	39.2322	56.9587	65.7209

The first six mode shapes corresponding to the six natural frequencies in the fifth row of Table 5.7 are plotted in Figure 5.5. The modal assurance criterion (MAC) is used here to indicate the level of consistency between two mode shapes, which is given by,

$$MAC(\boldsymbol{\phi}_s^i, \boldsymbol{\phi}_s^j) = \frac{(\boldsymbol{\phi}_s^{i\top} \boldsymbol{\phi}_s^j)^2}{(\boldsymbol{\phi}_s^{i\top} \boldsymbol{\phi}_s^i)(\boldsymbol{\phi}_s^{j\top} \boldsymbol{\phi}_s^j)} \tag{5.1}$$

where $\boldsymbol{\phi}_s^i$ and $\boldsymbol{\phi}_s^j$ are the i^{th} and the j^{th} modal vectors. It takes values between 0 and 1, with larger values corresponding to higher consistency. Possible reasons why the MAC value is close to 0 or 1 can be found in [100].

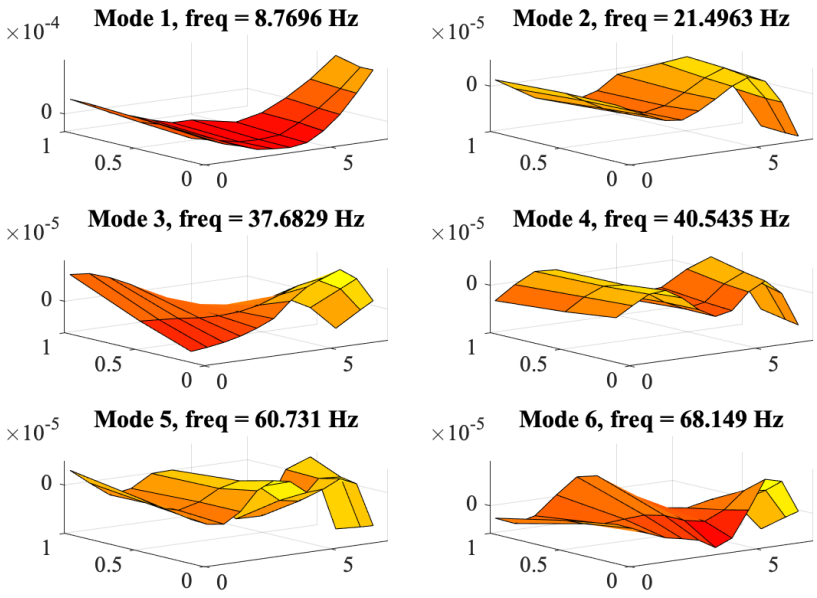
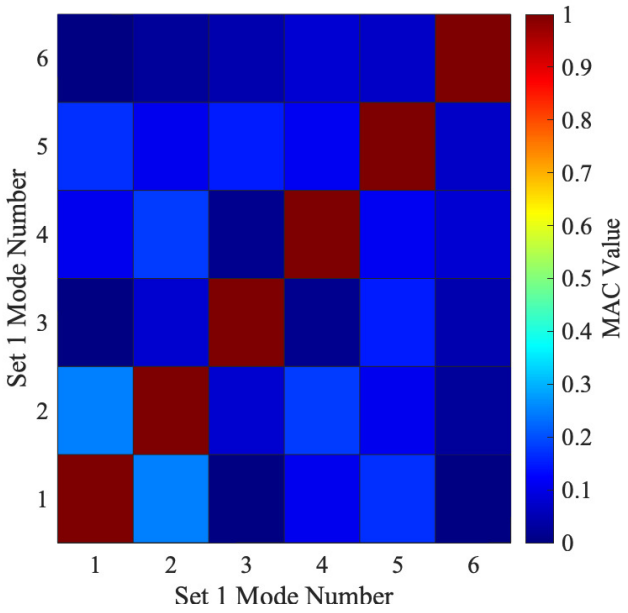
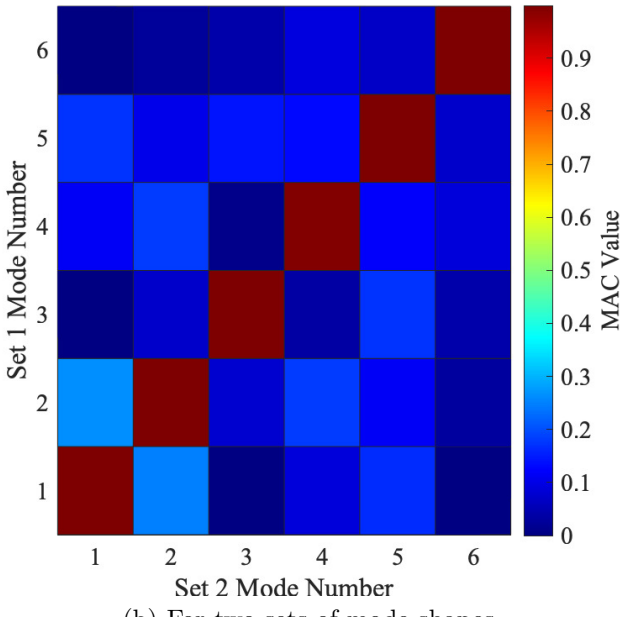


Figure 5.5: The first six mode shapes of the glider wing at 10°C.



(a) For a set of mode shapes.



(b) For two sets of mode shapes.

Figure 5.6: 2D plot of the MAC of the first six mode shapes of the wing under the normal condition.

The MAC matrix of the six mode shapes corresponding to 10°C in Table 5.5 is shown in Table 5.8 and Figure 5.6a. The MAC matrix of two different sets of modes with six mode shapes corresponding to 10°C and 15°C are shown in Table 5.9 and Figure 5.6b. The effect of temperature on the mode shape can be seen by the difference between the values at the same location in the tables 5.8 and 5.9. On a scale from 0 to 1, the maximum absolute value of the difference is 0.0241, corresponding to where Set 1 Mode 3 and Set 2 Mode 5 are located in Figure 5.6b and the same position in Figure 5.6a. By comparing the figures 5.6b and 5.6a, it is difficult to see this largest difference, which implies that the mode shape is not a temperature-sensitive feature.

Table 5.8: The MAC matrix of the six mode shapes of the wing under the normal condition at 10°C.

Natural frequency (Hz)	8.7696	21.4963	37.6829	40.5435	60.7310	68.1490
8.7696	1	0.2491	0.0004	0.1059	0.1693	0.0008
21.4963	0.2491	1	0.0761	0.1834	0.1072	0.0242
37.6829	0.0004	0.0761	1	0.0148	0.1499	0.0452
40.5435	0.1059	0.1834	0.0148	1	0.1108	0.0806
60.7310	0.1693	0.1072	0.1499	0.1108	1	0.0681
68.1490	0.0008	0.0242	0.0452	0.0806	0.0681	1

Table 5.9: The MAC matrix of two sets of six mode shapes of the wing under the normal condition at 10°C and 15°C.

Natural frequency (Hz)	8.5974	21.1780	36.9880	40.0681	59.2559	67.2463
8.7696	0.9970	0.2480	0.0004	0.0894	0.1663	0.0006
21.4963	0.2637	0.9995	0.0766	0.1827	0.1139	0.0278
37.6829	0.0010	0.0733	0.9986	0.0345	0.1740	0.0426
40.5435	0.1148	0.1822	0.0082	0.9923	0.1197	0.0896
60.7310	0.1730	0.1050	0.1435	0.1316	0.9974	0.0704
68.1490	0.0010	0.0236	0.0419	0.0902	0.0665	0.9992

5.3 Conclusions

This chapter clearly illustrated the collected data sets that can be used in this thesis: time-series data, frequency-domain data and modal data. The experimental setup and design were described in as much detail and completeness as possible to provide conditions for replication of this experiment.

To sum up, two rounds of experiments were conducted, both considering three damage scenarios. But the damage size and location were adjusted in the second

round of testing to provide a more challenging and richer data set for the damage identification task. Due to limited time in the laboratory, six temperature ranges were controlled and recorded in the first-round tests (from 0 °C to 30 °C at intervals of 5 °C), while only five were considered in the second round (from 0 °C to 25 °C).

Finally, several suggestions for conducting such experiments are put forward here: (1) Check the linearity of the structural system under test; (2) Adjust the damage size and location according to the research needs; (3) Consider the temperature to avoid damage to the equipment; (4) Automate the data collection and export process as much as possible.

Chapter 6

Canonical-correlation-based optimisation objective

Inspired by the discussion in Section 1.1.3, a metric that can help quickly evaluate the performance of sensor combinations rather than individual sensors will be investigated in this chapter. To be specific, the *sum of squared canonical correlation coefficients* (SSC) is used as an optimisation objective of a sensor system. Two fast calculation methods of this criterion combined with a greedy search are developed and introduced in this chapter [1]. The optimisation criterion for selecting the optimal subset of sensors is developed from the field of feature selection.

Feature selection is the process of selecting the subset of features that are most useful to realise the purpose of a model. When there are reliable data in the design stage of the sensor system, the SPO for SHM, based on statistical pattern recognition can be regarded as a feature selection problem. This chapter only presents case studies of the proposed canonical-correlation-based greedy search from the field of feature selection in general machine learning. Its applications in SPO will be given in Chapters 8 and 9.

[1] Note this work was carried out in collaboration with S. Zhang. Please refer to [101] for the corresponding paper.

6.1 Supervised SSC-based criterion theorem

6.1.1 Correlation coefficients and angles

In Chapter 3, it was shown that the canonical correlation coefficient could be defined as [76],

$$R(\mathbf{X},\mathbf{Y}) = \max_{\mathbf{u},\mathbf{v}} r(\mathbf{X}\mathbf{u},\mathbf{Y}\mathbf{v}) \tag{6.1}$$

where $\mathbf{X} \in \mathbb{R}^{N\times n}$ and $\mathbf{Y} \in \mathbb{R}^{N\times m}$ are two matrices with two co-occurring multivariate random variables with N observations, and the optimal projection directions $\mathbf{u} \in \mathbb{R}^{n\times 1}$ and $\mathbf{v} \in \mathbb{R}^{m\times 1}$ can be obtained by solving the eigenvalue equations given by,

$$\left[(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}} - R_l^2(\mathbf{X},\mathbf{Y})\mathbf{I}\right]\mathbf{u} = \mathbf{0} \tag{6.2a}$$

$$\left[(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}} - R_l^2(\mathbf{X},\mathbf{Y})\mathbf{I}\right]\mathbf{v} = \mathbf{0} \tag{6.2b}$$

where $\mathbf{X}_{\mathrm{M}} = \mathbf{X} - \overline{\mathbf{X}}$ and $\mathbf{Y}_{\mathrm{M}} = \mathbf{Y} - \overline{\mathbf{Y}}$. $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are the means of columns of $\mathbf{X}$ and $\mathbf{Y}$ respectively.

The canonical correlation coefficient can also be considered as the cosine of the angle between the projected $\mathbf{X}$ and the projected $\mathbf{Y}$ [102], which is,

$$R(\mathbf{X},\mathbf{Y}) = \max_{\mathbf{u},\mathbf{v}} \cos\left(\angle(\mathbf{X}\mathbf{u},\mathbf{Y}\mathbf{v})\right) = \cos\left(\min_{\mathbf{u},\mathbf{v}} \angle(\mathbf{X}\mathbf{u},\mathbf{Y}\mathbf{v})\right) \tag{6.3}$$

If considered from the direction of the angle between the vectors, the angle corresponding to the canonical correlation between $\mathbf{X}$ and $\mathbf{Y}$ is defined as [101],

$$\Theta(\mathbf{X},\mathbf{Y}) \triangleq \min_{\mathbf{u},\mathbf{v}} \angle(\mathbf{X}\mathbf{u},\mathbf{Y}\mathbf{v}) \tag{6.4}$$

6.1.2 Canonical-correlation-based greedy search

In order to comprehensively consider the correlation between $\mathbf{X}$ and $\mathbf{Y}$ in all projected directions, the sum of non-zero eigenvalues, i.e. the SSC, is adopted as a criterion for evaluating the effectiveness of a feature set from a sensor set to infer the structural

state, which is given by,

$$\begin{aligned}\sum_{l=1}^{\min(n,m)} R_l^2(\mathbf{X},\mathbf{Y}) &= \mathrm{tr}\left[(\mathbf{X}_{\mathrm{M}}^\top\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^\top\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^\top\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^\top\mathbf{X}_{\mathrm{M}}\right] \\ &= \mathrm{tr}\left[(\mathbf{Y}_{\mathrm{M}}^\top\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^\top\mathbf{X}_{\mathrm{M}}(\mathbf{X}_{\mathrm{M}}^\top\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^\top\mathbf{Y}_{\mathrm{M}}\right]\end{aligned} \tag{6.5}$$

where $\mathrm{tr}[\,\cdot\,]$ represents the trace of a matrix. It is worth mentioning that the letters $\mathbf{u}$ and $\mathbf{v}$ will be redefined and used in the following part of this subsection.

According to [101], for the dataset with the number of observations smaller than the number of features, a method called *h-correlation* can be used to speed up the calculation of this criterion by the orthogonalisation of the feature matrix $\mathbf{X}_{\mathrm{M}}$ and the label matrix $\mathbf{Y}_{\mathrm{M}}$. Here, $\mathbf{W} \in \mathbb{R}^{N\times n}$ and $\mathbf{V} \in \mathbb{R}^{N\times m}$ are orthogonal bases for the column space of $\mathbf{X}_{\mathrm{M}}$ and the column space of $\mathbf{Y}_{\mathrm{M}}$ respectively, which can be obtained by the singular value decomposition. The SSC can then be calculated by,

$$\sum_{l=1}^{\min(n,m)} R_l^2(\mathbf{X},\mathbf{Y}) = \sum_{l=1}^{\min(n,m)} R_l^2(\mathbf{W},\mathbf{V}) = \sum_{i=1}^{n}\sum_{j=1}^{m} h_{i,j} \tag{6.6}$$

where,

$$h_{i,j} = r^2(\mathbf{w}_i,\mathbf{v}_j),\ \mathbf{W} = (\mathbf{w}_1,\ldots,\mathbf{w}_n),\ \mathbf{V} = (\mathbf{v}_1,\ldots,\mathbf{v}_m)$$

Combined with a greedy search algorithm, the SSC computed by the h-correlation method at each iteration is given by,

$$\sum_{l=1}^{\min((p+q),m)} R_l^2((\mathbf{X}_{\mathrm{s}},\mathbf{X}_{\mathrm{a}}),\mathbf{Y}) = \sum_{l=1}^{\min(p,m)} R_l^2(\mathbf{W}_{\mathrm{s}},\mathbf{V}) + \sum_{l=1}^{\min(q,m)} R_l^2(\mathbf{W}_{\mathrm{a}},\mathbf{V}) \tag{6.7}$$

where $\mathbf{X}_{\mathrm{s}}$ is composed of the p selected features in the previous steps and $\mathbf{X}_{\mathrm{a}}$ represents the q features available for selection at the current step. $\mathbf{W}_{\mathrm{s}} \in \mathbb{R}^{N\times p}$ and $(\mathbf{W}_{\mathrm{s}}, \mathbf{W}_{\mathrm{a}}) \in \mathbb{R}^{N\times(p+q)}$ are orthogonal bases for the column space of $\mathbf{X}_{\mathrm{s}}$ and the column space of $\mathbf{X}_{\mathrm{a}}$ respectively; $\mathbf{W}_{\mathrm{a}}$ is orthogonal to $\mathbf{W}_{\mathrm{s}}$.

For a dataset with the number of observations larger than the number of features, a method named the *θ-angle* can be adopted [101]. However, unlike the h-correlation method, the coordinate matrices of $\mathbf{X}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{M}}$ with respect to the same matrix $\mathbf{U} \in \mathbb{R}^{N\times(n+m)}$ need to be obtained first, which are represented as $[\mathbf{X}_{\mathrm{M}}]_{\mathbf{U}} \in \mathbb{R}^{(n+m)\times n}$ and $[\mathbf{Y}_{\mathrm{M}}]_{\mathbf{U}} \in \mathbb{R}^{(n+m)\times m}$. The columns of $\mathbf{U}$ must form an *orthonormal basis*, which

means $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$. Then,

$$R(\mathbf{X}, \mathbf{Y}) = \cos\left(\Theta([\mathbf{X}_{\mathrm{M}}]_{\mathbf{U}}, [\mathbf{Y}_{\mathrm{M}}]_{\mathbf{U}})\right) \tag{6.8}$$

This is the key to transitioning from the orthogonal feature-based superposition to the angle-based superposition.

The orthogonal vector bases $\mathbf{W} \in \mathbb{R}^{(n+m)\times n}$ and $\mathbf{V} \in \mathbb{R}^{(n+m)\times m}$ for the coordinate matrices $[\mathbf{X}_{\mathrm{M}}]_{\mathbf{U}}$ and $[\mathbf{Y}_{\mathrm{M}}]_{\mathbf{U}}$ can then be found, and the SSC can be calculated by using another superposition property, which is given by,

$$\sum_{l=1}^{\min(n,m)} R_l^2(\mathbf{X}, \mathbf{Y}) = \sum_{l=1}^{\min(n,m)} \cos^2\left(\Theta_l(\mathbf{W}, \mathbf{V})\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \theta_{i,j} \tag{6.9}$$

where,

$$\theta_{i,j} = \cos^2\left(\angle(\mathbf{w}_i, \mathbf{v}_j)\right),\ \mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n),\ \mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_m) \tag{6.10}$$

Similarly, combining the θ-angle method with a greedy search algorithm, in one selection step, the SSC is calculated as,

$$\sum_{l=1}^{\min((p+q),m)} R_l^2\big((\mathbf{X}_{\mathrm{s}}, \mathbf{X}_{\mathrm{a}}), \mathbf{Y}\big) = \sum_{l=1}^{\min(p,m)} \cos^2\left(\Theta_l(\mathbf{W}_{\mathrm{s}}, \mathbf{V})\right) + \sum_{l=1}^{\min(q,m)} \cos^2\left(\Theta_l(\mathbf{W}_{\mathrm{a}}, \mathbf{V})\right) \tag{6.11}$$

where $\mathbf{W}_{\mathrm{s}} \in \mathbb{R}^{(n+m)\times p}$ and $(\mathbf{W}_{\mathrm{s}}, \mathbf{W}_{\mathrm{a}}) \in \mathbb{R}^{(n+m)\times(p+q)}$ are orthogonal bases for the column space of $[\mathbf{X}_{\mathrm{s}}]_{\mathbf{U}}$ and the column space of $[\mathbf{X}_{\mathrm{a}}]_{\mathbf{U}}$ respectively. $\mathbf{W}_{\mathrm{a}}$ is orthogonal to $\mathbf{W}_{\mathrm{s}}$.

In this thesis, one feature is added at each iteration of a greedy search. The SSC of each candidate feature set can be calculated by,

$$\sum_{l=1}^{\min((p+1),m)} R_l^2((\mathbf{X}_{\mathrm{s}}, \mathbf{x}_{\mathrm{a}i}), \mathbf{Y}) = \sum_{l=1}^{\min(p,m)} R_l^2(\mathbf{W}_{\mathrm{s}}, \mathbf{V}) + R^2(\mathbf{w}_{\mathrm{a}i}, \mathbf{V}) \tag{6.12}$$

or,

$$\sum_{l=1}^{\min((p+1),m)} R_l^2\big((\mathbf{X}_{\mathrm{s}}, \mathbf{x}_{\mathrm{a}i}), \mathbf{Y}\big) = \sum_{l=1}^{\min(p,m)} \cos^2\left(\Theta_l(\mathbf{W}_{\mathrm{s}}, \mathbf{V})\right) + \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}i}, \mathbf{V})\right) \tag{6.13}$$

where $i = 1, \ldots, q$. It can be seen that, in each iteration, the feature that maximises

$R^2(\mathbf{w}_{ai}, \mathbf{V})$ or $\cos^2\left(\Theta(\mathbf{w}_{ai}, \mathbf{V})\right)$ will be selected until the desired number of features is reached. The pseudocode of this canonical-correlation-based greedy search algorithm is given in Algorithm 1. In addition, an example is given in Appendix C to illustrate the specific process of implementing this algorithm to facilitate algorithm reproduction.

Note that the labels can usually be represented by a vector, i.e. $m = 1$. In this situation, the applied criterion reduces to the squared multiple correlation coefficient $R^2(\mathbf{X}, \mathbf{y})$. However, $m > 1$ may happen, for example, when the responses are multi-class categorical labels, which do not have an inherent order. This is the situation in this chapter. These labels can be dummy encoded to a matrix $\mathbf{Y}$, in which each category is transformed into a set of binary variables to use m variables to represent $(m + 1)$ categories. One advantage of the dummy encoding is that it can make the label matrix $\mathbf{Y}$ retain full column rank after subtracting its column mean.

6.1.3 Comparison of elapsed time

In this part, a synthetic dataset is generated to compare the runtime of defined h-correlation- and θ-angle-based feature-selection methods. The feature and response matrices $\mathbf{X}_{\mathrm{a}}$ and $\mathbf{Y}$ are generated from random numbers uniformly distributed between 0 to 1. There are no pre-selected features, which means $\mathbf{X}_{\mathrm{s}}$ is an empty matrix. The number of samples is 5000, and the number of variables in $\mathbf{X}_{\mathrm{a}}$ and $\mathbf{Y}$ are 700 and 50 respectively, i.e. $N = 5000$, $q = 700$, $m = 50$. The built-in function `canoncorr` of MATLAB is adopted as the baseline method to compute this criterion as well. These three methods are used to greedily select features according to the SSC. The results are shown in Figure 6.1.

It can be seen that the baseline method is much slower than the other two methods, even at the first iteration. To illustrate the reasons more simply, the first iteration is taken as an example to compare the asymptotic upper bound notation $\mathcal{O}$ of these three greedy algorithms. When selecting the first feature, the baseline algorithm requires solving equation (6.5), which has a computational complexity of $\mathcal{O}(m^2N)$ for each candidate feature. There are q candidate features, so the complexity of this iteration is $\mathcal{O}(m^2Nq)$. For the h-correlation-based algorithm, before the first iteration, $\mathbf{Y}_{\mathrm{M}}$ needs to be orthogonalised to obtain $\mathbf{V}$, and its rate of growth is $\mathcal{O}(m^2N)$. Then, the computation of the SSC values for q candidate features requires

Algorithm 1: Pseudocode for feature selection based on h-correlation and the θ-angle.

Input : Initial selected feature matrix $\mathbf{X}_{\mathrm{s}} \in \mathbb{R}^{N\times p}$
Initial candidate feature matrix $\mathbf{X}_{\mathrm{a}} \in \mathbb{R}^{N\times q}$
Response matrix $\mathbf{Y} \in \mathbb{R}^{N\times m}$
Total number of features to be selected t

Output : Final selected feature matrix $\mathbf{X}_{\mathrm{s}} \in \mathbb{R}^{N\times t}$
Vector of indices of additionally selected features, $\mathbf{i}_{\mathrm{s}} \in \mathbb{R}^{(t-p)\times 1}$, preserving the initial order in $\mathbf{X}_{\mathrm{a}}$

$n \leftarrow$ the number of variables in $(\mathbf{X}_{\mathrm{s}}, \mathbf{X}_{\mathrm{a}})$;
$m \leftarrow$ the number of variables in $\mathbf{Y}$;
$N \leftarrow$ the number of observations;
$p \leftarrow$ the number variables in $\mathbf{X}_{\mathrm{s}}$;
Centre $\mathbf{X}_{\mathrm{s}}$, $\mathbf{X}_{\mathrm{a}}$ and $\mathbf{Y}$ to $\mathbf{X}_{\mathrm{M_s}}$, $\mathbf{X}_{\mathrm{M_a}}$ and $\mathbf{Y}_{\mathrm{M}}$;
if $N > n + m$ **then**
  let $\mathbf{F}_{\mathbf{X}_{\mathrm{s}}} = [\mathbf{X}_{\mathrm{M_s}}]_{\mathbf{U}}$, $\mathbf{F}_{\mathbf{X}_{\mathrm{a}}} = [\mathbf{X}_{\mathrm{M_a}}]_{\mathbf{U}}$ and $\mathbf{F}_{\mathbf{Y}} = [\mathbf{Y}_{\mathrm{M}}]_{\mathbf{U}}$, with $(\mathbf{F}_{\mathbf{X}_{\mathrm{s}}}, \mathbf{F}_{\mathbf{X}_{\mathrm{a}}}) \in \mathbb{R}^{(n+m)\times n}$, $\mathbf{F}_{\mathbf{Y}} \in \mathbb{R}^{(n+m)\times m}$ and $\mathbf{U} \in \mathbb{R}^{N\times(n+m)}$ as orthogonal basis of $(\mathbf{X}_{\mathrm{M_s}}, \mathbf{X}_{\mathrm{M_a}}, \mathbf{Y}_{\mathrm{M}})$;
else
  $\mathbf{F}_{\mathbf{X}_{\mathrm{s}}} = \mathbf{X}_{\mathrm{M_s}}$, $\mathbf{F}_{\mathbf{X}_{\mathrm{a}}} = \mathbf{X}_{\mathrm{M_a}}$, $\mathbf{F}_{\mathbf{Y}} = \mathbf{Y}_{\mathrm{M}}$;
end
Orthogonalise $\mathbf{F}_{\mathbf{Y}}$ to itself to form $\mathbf{V}$, which is composed of $(\mathbf{v}_1, \cdots, \mathbf{v}_m)$;
if $p \neq 0$ **then**
  Orthogonalise $\mathbf{F}_{\mathbf{X}_{\mathrm{s}}}$ to itself to form $\mathbf{W}_{\mathrm{s}}$, which is composed of $(\mathbf{w}_{\mathrm{s}1}, \cdots, \mathbf{w}_{\mathrm{s}p})$;
  Orthogonalise $\mathbf{F}_{\mathbf{X}_{\mathrm{a}}}$ to $\mathbf{W}_{\mathrm{s}}$ to form $\mathbf{W}_{\mathrm{a}}$, which is composed of $(\mathbf{w}_{\mathrm{a}1}, \cdots, \mathbf{w}_{\mathrm{a}q})$;
else
  $\mathbf{W}_{\mathrm{a}} \leftarrow \mathbf{F}_{\mathbf{X}_{\mathrm{a}}}$, which is composed of $(\mathbf{w}_{\mathrm{a}1}, \cdots, \mathbf{w}_{\mathrm{a}q})$;
end
while $p < t$ **do**
  Compute $R^2(\mathbf{w}_{\mathrm{a}i}, \mathbf{V}) = \sum_{j=1}^{m} \dfrac{\mathbf{w}_{\mathrm{a}i}^{\top}\mathbf{v}_j}{\|\mathbf{w}_{\mathrm{a}i}\|\|\mathbf{v}_j\|}$;
  Find feature index $i_{\max}$, such that $R^2(\mathbf{w}_{\mathrm{a}i}, \mathbf{V})$ is the maximum with $i = 1, \ldots, q$;
  Remove the feature $\mathbf{x}_{i_{\max}}$ from $\mathbf{X}_{\mathrm{a}}$, and append it to $\mathbf{X}_{\mathrm{s}}$;
  Save the index of $\mathbf{x}_{i_{\max}}$ in the initial matrix $\mathbf{X}_{\mathrm{a}} \in \mathbb{R}^{N\times m}$ into $\mathbf{i}_{\mathrm{s}}$;
  $p \leftarrow p + 1$;
  $q \leftarrow q - 1$;
end

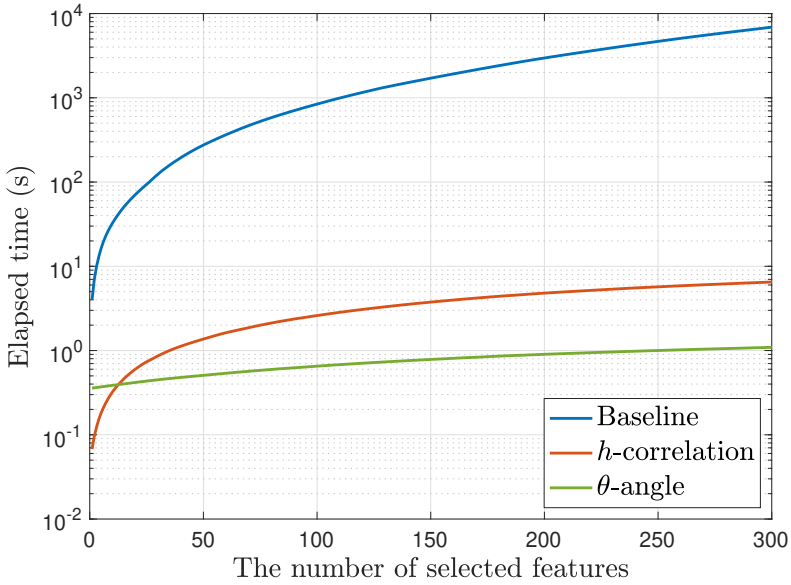


Figure 6.1: The elapsed time of the definition, h-correlation- and the θ-angle-based feature selection methods.

the inner product of each centred feature value with each column of $\mathbf{V}$. Therefore, the complexity of finding the first optimal feature in this iteration is $\mathcal{O}(mNq)$. Similarly, for the θ-angle-based algorithm, the computational complexity of the first iteration is $\mathcal{O}(m(q+m)q)$.

It is also found that the algorithm based on the θ-angle is slower than the algorithm based on h-correlation until the selection of the 13^{th} feature. This result is because, for the θ-angle-based algorithm, before the first iteration, the orthonormal basis $\mathbf{U}$ of $(\mathbf{X}_{\mathrm{M_a}}, \mathbf{Y}_{\mathrm{M}})$ needs to be found by singular-value decomposition. The asymptotic running time is $\mathcal{O}((m+q)^2 N)$, which dominates the computational complexity of selecting the first few features. However, since the singular-value decomposition only needs to be done once, it can be seen from Figure 6.1 that the algorithm based on the θ-angle is faster in general.

6.2 Application to the real data-sets

To show the effectiveness of the proposed criterion, the SSC and seven mutual information (MI) based criteria [25]: minimal-Redundancy-Maximal-Relevance (mRMR), Mutual Information Maximisation (MIM), Joint Mutual Information (JMI), Conditional Mutual Information Maximisation (CMIM), Conditional Infomax Feature

Extraction (CIFE), Interaction Capping (ICAP), and Double Input Symmetrical Relevance (DISR), are used to select features for eight real datasets for classification and regression prediction.

6.2.1 Classification tasks

For the classification tasks, three UCI datasets [103] and one MINST dataset [104] are chosen to compare the SSC with the seven MI-based criteria. These four datasets are summarised in Table 6.1.

Table 6.1: An example of the greedy search for classification based on the SSC.

	Lymph	CNAE	Mfeat	MNIST
Data Type	Categorical & Discrete	Discrete	Continuous	Discrete
No. of Instances	142	1080	2000	700000
No. of Features	18	856	649	784
No. of Classes	2	9	10	10
Classifier	SVM	LDA	SVM	LDA

The *Lymph* dataset is to classify lymphography information into two conditions, namely metastatic and malignant lymph, based on 18 medical diagnostic attributes. The categorical features are assigned a value by ordinal encoding. For the National Classification of Economic Activities (CNAE) dataset [105], 1080 business description documents of Brazilian companies are classified into nine economic activities according to the frequency of occurrence of 856 words in the documents. The multi-feature digit (Mfeat) dataset [106] is to classify ten handwritten numerals (i.e. 0 to 9) from a collection of Dutch utility maps by using 649 features extracted from raw images. The continuous features in this dataset are standardised to z-scores. Furthermore, when using MI-based criteria, the continuous features are discretised into five equal-width bins. For the MNIST dataset [104], ten handwritten digits (i.e. 0 to 9) are recognised using 784 pixels with different grey levels as features. When using the SSC-based criterion for all four datasets, the label is encoded by the dummy encoding.

A linear SVM or LDA model is adopted for classification, given the selected features. Ten-fold cross-validation was used to calculate the average of ten ACC scores on the validation data for comparing the performance of eight feature-selection criteria. The corresponding results are shown in Figure 6.2. It can be seen that the SSC-based

criterion is competitive in selecting features for classification tasks. When the ACC results are averaged over the number of selected features, the proposed SSC achieves the best results in three of the four datasets, as shown in Figure 6.2e.

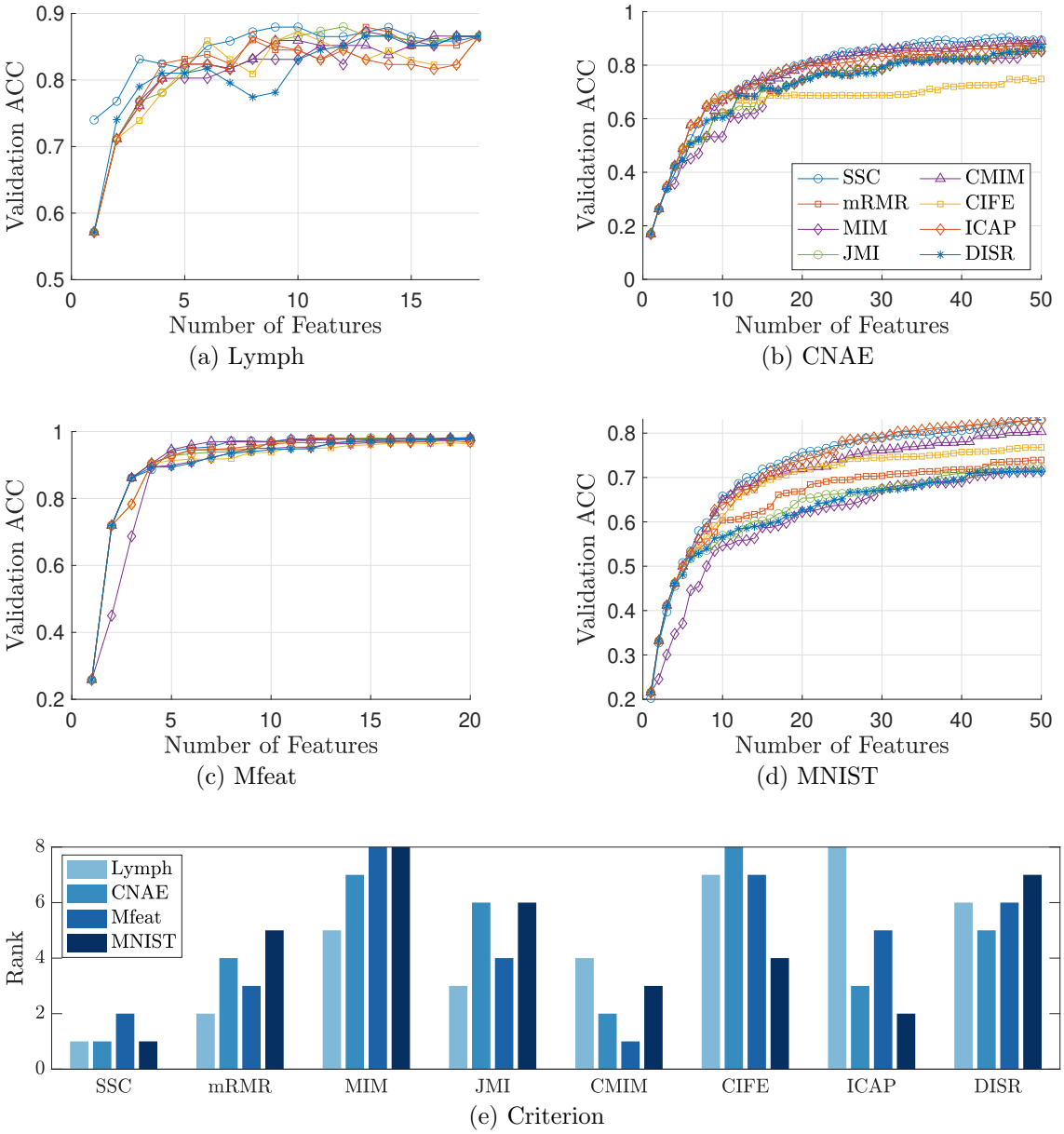


Figure 6.2: The comparison of the averaged ACC results between the different feature ranking criteria.

6.2.2 Regression tasks

Four UCI datasets are used for the regression tasks to compare the SSC with the seven MI-based criteria. Table 6.2 gives a summary of these datasets.

Table 6.2: An example of the greedy search for regression based on the SSC.

	Student	Parkinson	Conductor	Energy
Data Type	Categorical & Discrete	Continuous	Continuous	Continuous
No. of Instances	649	1040	21263	19735
No. of Features	30	26	81	2924

The *Student* dataset [107] is to predict the final grade of the Portuguese class for students in secondary education of two Portuguese schools. 30 candidate features, including the demographic, social and school-related attributes, are collected from questionnaires, and the categorical features are assigned an ordinal value. For the *Parkinson* dataset [108], the Unified Parkinson Disease Rating Scale (UPDRS) scores are predicted using 26 linear and time-frequency-based features extracted from voice samples of different subjects, such as sustained vowels, numbers, words and short sentences. The *conductor* dataset [109] is to predict the critical temperature of a superconductor by 81 features extracted from the chemical formula of superconductors. The *Energy* dataset [110] is used to predict the energy usage of appliances during the operation phase of a building. To account for the nonlinear relationship between features and the response value, the original 24 features are expanded to 2924 using polynomials of degree one to three. Features in all four datasets are all standardised to z-scores. When using MI-based criteria, the expanded features of the Energy dataset are discretised into ten equal-width bins, while the others are into five equal-width bins.

Given the selected features, a linear regression (LR) model is applied for regression tasks. Similarly to classification tasks, ten-fold cross-validation was used to calculate the average value of RMSE on the validation datasets, and the corresponding results can be seen in Figure 6.3. Clearly, the SSC-based criterion outperforms these MI-based criteria. The reason is that because all four regression tasks have a single output, the SSC degenerates to R-squared, which is monotonically decreasing with the RMSE of the LR model. See Chapter 4 for more details on this relationship.

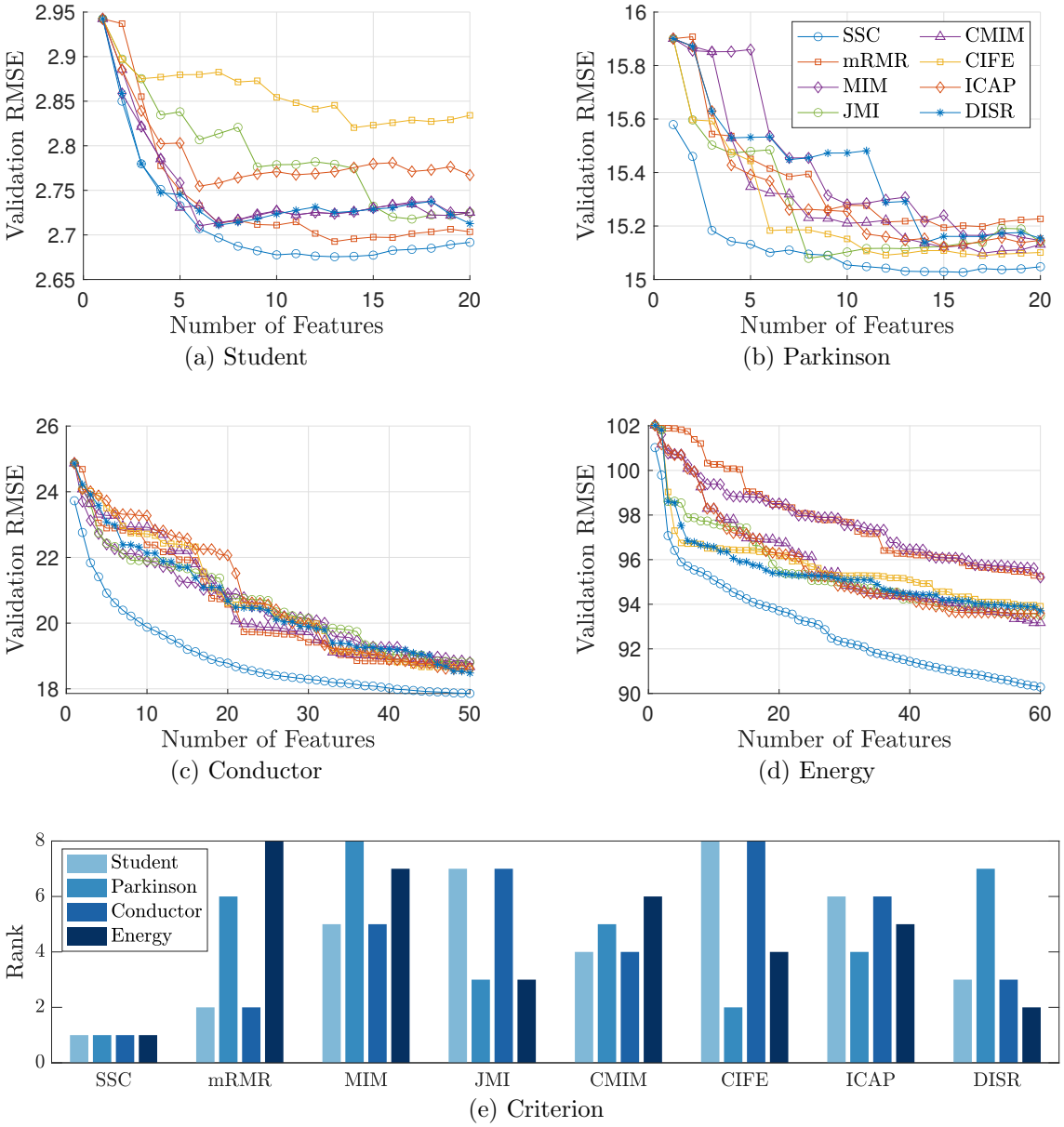


Figure 6.3: The comparison of the averaged RMSE results between the different feature ranking criteria.

6.3 SSC-based unsupervised feature selection method

As said in Section 1.1.3, unsupervised feature-selection strategies also require a target, which is not class labels or response variables, but general characteristics of the collected data. This understanding drives the transition from the SSC-based supervised feature selection in the previous section to the unsupervised one.

In this section, an unsupervised feature selection method combing the SSC-based criterion and PCA is developed, which can simultaneously select multiple features for regression or classification tasks. In short, compared with the supervised SSC-based criterion theorem, the SSC-based unsupervised feature selection method requires one more step: PCA is performed on the feature matrix $\mathbf{X}$ to obtain the derived matrix $\mathbf{Y}$ as the label matrix. Considering the scalability of this method for large datasets, the fast calculation strategies based on greedy search are also adopted. A case study of the MNIST dataset is used to compare the characteristics of the proposed unsupervised feature selection strategy and the supervised strategy introduced in the previous section.

6.3.1 Performance comparison based on MNIST dataset

In the design phase of a sensor system, data can be obtained from simulated models or laboratory tests. Data availability is largely project-dependent, which can be divided into four levels. At the first level, only normal condition data are available. At the second level, data for the normal condition and damage states are accessible. For level three, both data and labels for normal and damage states can be obtained. At the fourth level, data and labels for different health states and the labels for different normal conditions can be used. It can be seen that the higher the level, the more information, but it also means the more difficult it is to obtain.

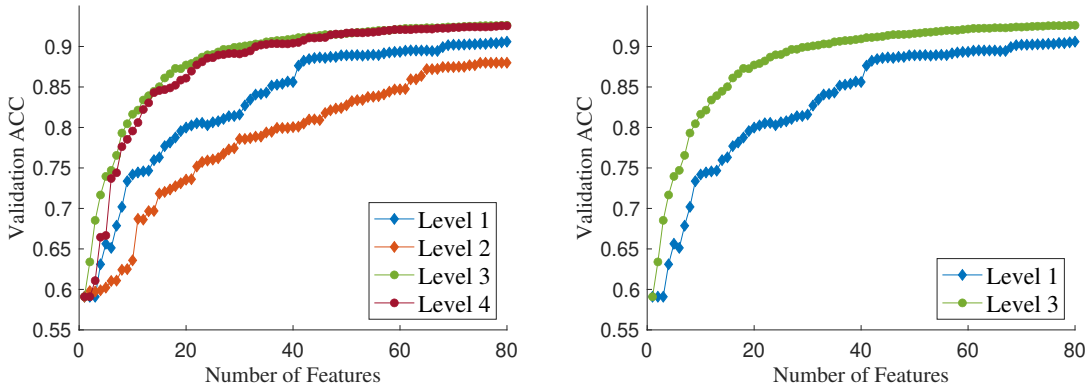
Depending on the availability of datasets, two feature-selection strategies proposed in this thesis can be employed. That is, for the first and second levels, the unsupervised feature selection can be used, while for the third and fourth levels, the supervised feature-selection method can be applied.

The MNIST dataset is used to simulate these four possible scenarios. The first four digits, from zero to three, are considered to be the normal pattern data. Data of another three digits, from four to six, are used as the data for three anticipated novel patterns. Data of digits seven to nine are used as the data for three unexpected novel patterns. Therefore, the available data for Level one is zero to three, and for Level two is zero to six. For the third level, data for zero to six plus a label for zero to three and three labels for four to six are usable. The fourth level has access to data and labels from zero to six. Note that the main reason for using the MNIST dataset is that it can intuitively show the relationship between selected pixels/features and

labels, which helps one further analyse and understand the results.

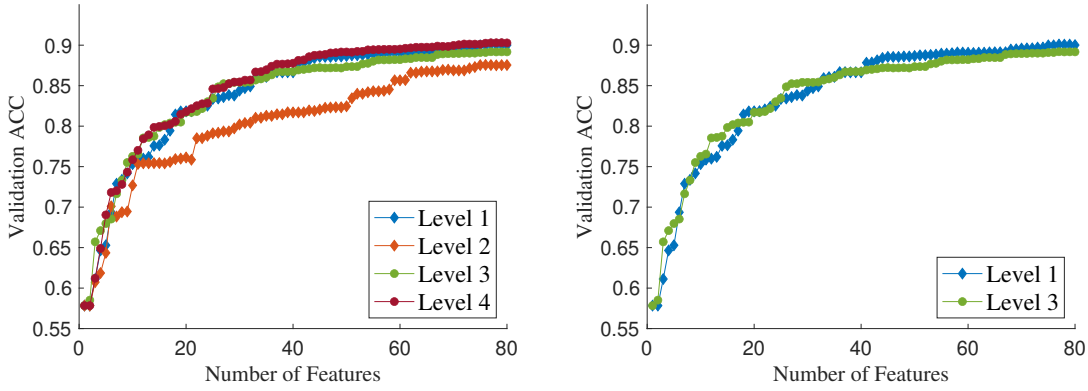
6.3.2 Results and discussion

Given the selected features, the LDA model is adopted for classification. Ten-fold cross-validation was used to calculate the average of ten ACC scores on the validation data to evaluate the diagnostic performance of the selected feature combinations with pixels from 1 to 80. The first p components of the derived label matrix $\mathbf{Y}$ capturing at least 95% of the total variance are selected to speed up the calculation again.



(a) For four levels. (b) For two levels.

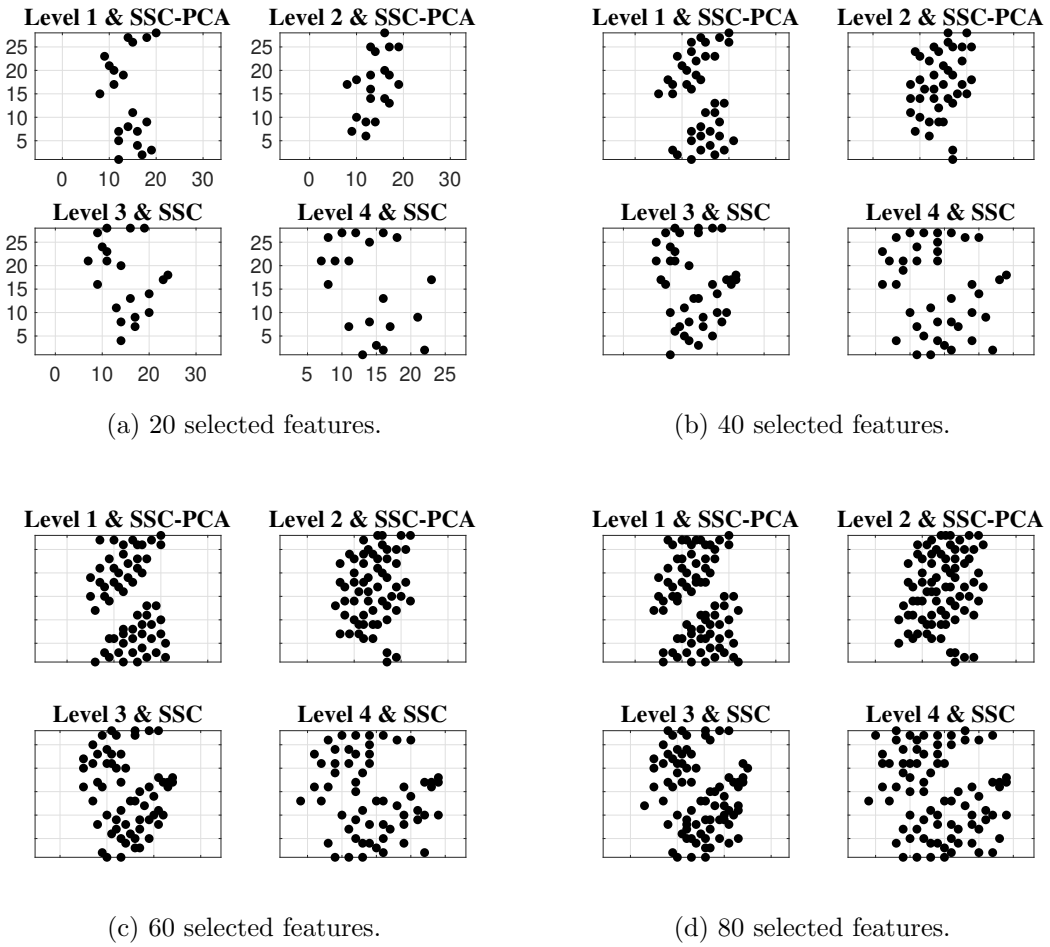
Figure 6.4: The performance comparison of datasets with different levels of data availability in identifying the normal pattern and expected novel patterns.



(a) For four levels. (b) For two levels.

Figure 6.5: The performance comparison of datasets with different levels of data availability in identifying the normal pattern and unexpected novel patterns.

The performance results of the two feature-selection methods applied for four different datasets are shown in Figures 6.4a and 6.5a. The performance curves corresponding to Level 1 and Level 3 are specially placed in Figures 6.4b and 6.5b for comparison, because they are often encountered. The distributions of 20, 40, 60 and 80 selected features are demonstrated in Figure 6.6. It can be seen that the selected feature distribution corresponding to the fourth level is the most dispersed, followed by Level three.



(a) 20 selected features.

(b) 40 selected features.

(c) 60 selected features.

(d) 80 selected features.

Figure 6.6: The distribution of selected features obtained by two feature selection methods for four different datasets.

From Figures 6.4a and 6.5a, it can be seen that diagnostic performance for the third and fourth levels is similar. One possible reason is that, for level 4, the labels of normal patterns are not used to handle the effects of their inconsistencies; this means that there may be no significant benefit if the additional information is not handled in an appropriate manner. It can also be seen that the second-level performance is

the worst. This observation may be because of the fact that using data for different patterns in an unsupervised manner makes the obtained system insensitive to their differences.

Figures 6.4b and 6.5b indicate that Level three outperforms Level one in detecting expected patterns, especially when the number of features is small. However, for detecting unexpected patterns, the performance of these two levels is very similar. When the features increase, the performance of Level one becomes better than that of Level three because Level three is more dedicated to detecting the expected patterns.

6.4 Conclusions

In this chapter, supervised and unsupervised feature selection methods based on canonical correlation coefficients and greedy search have been proposed to assess a feature combination. Both of them need a target and assess the correlation of features with the target. However, the difference between them is that the former requires collected target variables, such as labels or responses, while the latter can 'extract' targets from features, such as possible clusters of data or projected low-dimensional data. These two methods can be chosen depending on the level of data availability for different projects.

Furthermore, two approaches have been developed to speed up the above feature selection methods, which are named h-correlation and the θ-angle. The computational speed advantage of these two methods was demonstrated by a case study of a synthetic data set. When the number of features is equal to or greater than the instances, the h-correlation approach should be used, otherwise the θ-angle should be applied. The two methods are not unrelated. From the pseudocode in Algorithm 1, it can be found that the θ-angle only requires one more singular value decomposition step than h-correlation.

Finally, it is worth noting that both supervised and unsupervised methods are good at selecting features for regression or classification tasks. It can be seen from ACC or RMSE results corresponding to different feature selection strategies, on eight real data sets. However, they cannot intuitively reflect the final performance of the monitoring system. It seems that further steps are needed to fill this gap, which will be discussed in the next chapter.

Chapter 7

Designed-system performance assessment criteria

Based on the previous chapter, it can be known that some criteria can be calculated fast to evaluate a sensor combination but cannot give an intuitive idea about the performance of this sensor system. Furthermore, as discussed in Section 1.2.2, the study of system-assessment criteria for the design process has received little attention in the reviewed literature of SPO.

This chapter endeavours to address this gap by introducing a hierarchical evaluation system with two stages to improve the optimisation design of the SHM sensor network. It will be shown that this method can support decision-making with respect to the number, type and location of sensors.

7.1 A systematic assessment framework

7.1.1 Stage 1 - optimal sensor combination

The task at hand is achieving optimal spatial deployment of sensors for binary or multiclass classification. In this chapter, the *multiple correlation coefficient* (MCC) is investigated as a possible early-stage assessment criterion, which is designed to measure the maximal linear association between a set of independent variables and a

dependent variable [111]. Here, the N observations with n independent variables are represented by a matrix $\mathbf{X}$ of dimension $N \times n$. An $N \times 1$ vector of labels associated with each observation is represented by $\mathbf{y}$. The essence of the MCC analysis of $\mathbf{X}$ and $\mathbf{y}$ is to find a projection direction of $\mathbf{X}$ that can maximise the linear correlation between $\mathbf{y}$ and the projected $\mathbf{X}$, which can be realised by,

$$R(\mathbf{X}, \mathbf{y}) = r(\hat{\mathbf{y}}, \mathbf{y}) = \frac{\hat{\mathbf{y}}_{\mathrm{M}}^{\top} \mathbf{y}_{\mathrm{M}}}{\sqrt{\hat{\mathbf{y}}_{\mathrm{M}}^{\top} \hat{\mathbf{y}}_{\mathrm{M}}} \sqrt{\mathbf{y}_{\mathrm{M}}^{\top} \mathbf{y}_{\mathrm{M}}}} \tag{7.1}$$

where,

$$\hat{\mathbf{y}} = \mathbf{X}^{*} \beta, \quad \text{s.t. } \mathbf{X}^{*} = [\mathbf{1},\ \mathbf{X}] \tag{7.2}$$

$\hat{\mathbf{y}}$ refers to a vector containing the predicted values of the label. $\top$ indicates the matrix transpose. Subscript M means the variables are centralised. $\mathbf{X}^{*}$ is an augmented matrix of dimension $N \times (n+1)$. β can be obtained from the normal equations for a linear multiple regression with a least-squares cost function, which can be written as,

$$\beta = ({\mathbf{X}^{*}}^{\top} \mathbf{X}^{*})^{-1} {\mathbf{X}^{*}}^{\top} \mathbf{y} \tag{7.3}$$

The MCC is a scalar, taking values from 0 to 1. A sensor combination corresponding to a higher MCC value indicates that the features from these sensors may perform better when predicting the labels. The cost function defined by the MCC is an efficient means of searching for the optimal spatial deployment of sensors with different numbers.

Note that the labels can usually be represented by a vector, but it may happen that the labels are represented by a matrix; for example, when the labels can be one-hot encoded to a matrix to avoid adding order information. In this case, the SSC introduced in Chapter 6 can be adopted to replace the MCC. However, there is no research to theoretically prove which encoding method can make the selected features better predict the labels in the model training step. Furthermore, considering computational efficiency, the MCC is selected as an example in this chapter.

It is suggested that the criterion applied at this stage can be any criterion used in filter methods for feature selection, which measures the relevance of features from sensors with the dependent variables, without actually training a model on features.

7.1.2 Stage 2 - optimal sensor number and type

Based on the characteristics of the feature set distribution, the purpose of classification and the consequences of misclassification, an appropriate criterion can be selected for the second stage of a system design. The parameters of the confusion matrix can be used, which are demonstrated in Table 4.1.

The results based on these criteria are related to the selected type of classifier. Therefore, at this stage, it is necessary to determine the possible classifiers which will be used for classifying the data once the structure is put into use. Since the number of optimal sensor combinations with different sensor numbers and sensor types is expected to be relatively small at this stage, thanks to the optimisation conducted in Stage 1, it is feasible to train the associated classifiers and compare them with each other.

For binary classification, the receiver operating characteristic (ROC) curve is a commonly-used criterion to indicate the diagnostic ability of a classification system. The ROC curve consists of the true positive rate and the false positive rate. The area under the curve (AUC) makes it easy to compare one ROC curve to another. Thus, AUC is a criterion considered in the sensor-system evaluation step for the cases of binary classification.

For multiclass classification, a derivation from the confusion matrix, ACC, is taken as an example to assess system performance. Higher accuracy means a better decision can be made by the established SHM system.

Note that the criteria used in the second stage should be indicators that can intuitively reflect the performance of the final system. This means that the criteria for the second stage are highly dependent on the project requirements. In addition, it can be realised that the criteria used in the first stage should preferably be more related to the machine learning algorithms used in the second stage. For example, if a linear classifier is used in the second stage, a linear correlation-based criterion can be employed in the first stage. If a nonlinear classifier is applied in the second stage, a mutual information-based criterion can be applied in the first stage.

7.2 A numerical case study

A state-space model of a plate structure with 24 degrees of freedom (DoFs) is used for demonstration in this chapter with the simulation carried out in MATLAB version R2018b. The parameters of the 24-DoF model are given in Table 7.1. Each lumped-mass node provides a candidate location for the sensor, as demonstrated in Figure 7.1a. The simulated boundary conditions are fixed-free on the short side and free-free on the long side. Figure 7.1b shows the first four mode shapes of this structure.

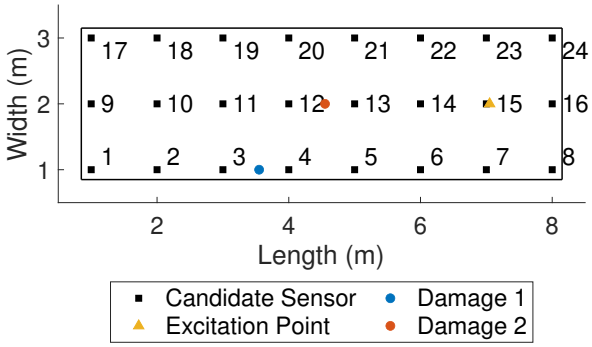
Table 7.1: The parameters of the plate structure.

Parameter	Mean (μ)	SD (σ)	Distribution
m, kg	1	0	N/A
k, N/m	1000	30	Normal (independent)
c, N/(m/s)	1	0.03	Normal (dependent on k)
Damage 1, N/m	200	0	N/A
Damage 2, N/m	200	0	N/A

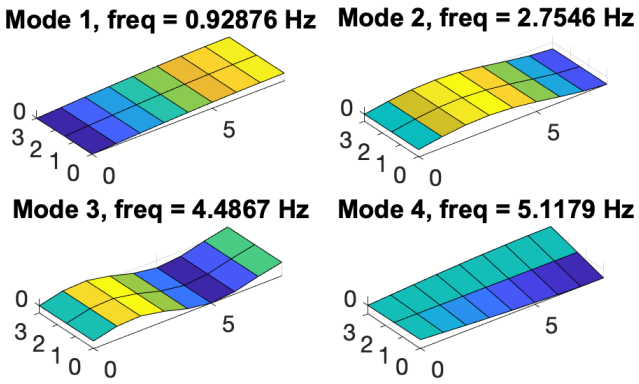
m: Mass; k: Stiffness; c: Damping

Observations of this structure are simulated by using the state-space model to acquire time-series data and then via the Fourier transform to obtain the FRFs. A white-noise Gaussian excitation is employed as the input force. Noise with a prescribed signal-to-noise ratio (SNR) is introduced into the simulated time-series data to represent the noise effects corresponding to different sensor types. Note that the type of sensor is characterised by the SNR. There is a rule of thumb that the higher the sensor cost, the higher the SNR of data. Four SNRs ranging from 45 dB to 30 dB (with an interval of 5 dB) are adopted here to simulate four sensor types; they are high-precision sensors, medium-precision sensors, low-precision sensors, and extremely low-precision sensors respectively.

This approach was used to provide predicted normal condition data and damaged-state data for the design stage of an SHM system. In a practical setting, the damage cases of interest would be determined by failure mode and effect analysis [112]. In this case study, damage is simulated by reducing the local stiffness. The position and extent of the simulated damage can be changed according to the research needs. Two damage cases are considered in this case study by reducing the stiffness value at two positions by 200 N/m, which are included in Table 7.1. The positions of the damage are shown in Figure 7.1a.



(a) Labels for simulated conditions.



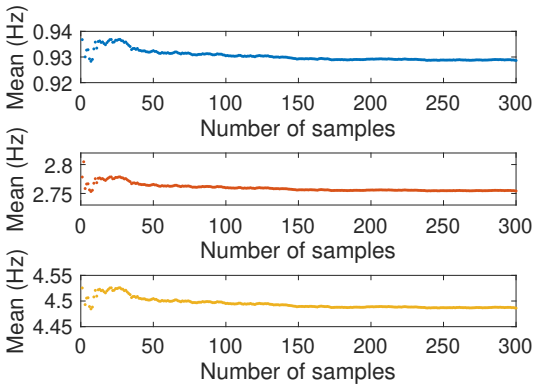
(b) First 4 mode shapes.

Figure 7.1: Schematic of the simulated 2-D plate and the corresponding mode shapes.

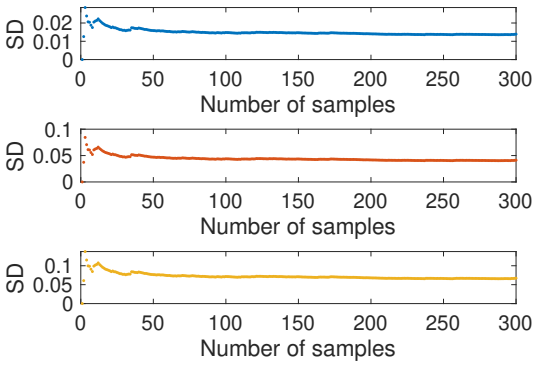
7.2.1 Uncertainty analysis

The focus of this chapter is not on how to provide realistic data sets, so here the introduction of uncertainty in the stiffness parameters is a simple example of how to consider the discordance between the simulated model and the real structure. Because proportional damping dependent on stiffness was used in this simulation, an uncertain damping effect was also involved. Other parameters related to material properties or geometry were held constant. The mean values and standard deviations of the uncertain parameters of the 24-DoF model are given in Table 7.1.

Latin hypercube sampling was adopted to efficiently generate the random samples of the uncertain parameters. After partitioning a cumulative curve into N intervals with equal cumulative probability, Latin hypercube sampling proceeds by randomly selecting one sample from each interval. The convergence of mean value and standard deviation of the first three natural frequencies for different sample sizes was used to



(a) Mean Value



(b) Standard deviation

Figure 7.2: Mean value and standard deviation for the first three natural frequencies of model samples. Blue, orange and yellow correspond to the first, second and third natural frequencies.

determine the required sample size, as shown in Figure 7.2. The percentage changes of these six quantities relative to their previous values were used as indicators of convergence. It was found that these indicators started below 0.02% and 1.5% after 150 samples. These two values were assumed to be the two thresholds here for the convergence of the mean and standard deviation respectively. Thus, 150 was selected as the number of samples. After obtaining 150 samples of this plate structure in the healthy state, 150 samples for two damage cases were obtained respectively by reducing the stiffness at the position of Damage 1 or Damage 2.

In addition to the parametric uncertainty, the uncertainty caused by the influence of noise on the time-series data was also taken into account. For each model sample, 20 time-series observations were simulated under each specified signal-to-noise ratio to consider the effect of noise. Therefore, 3000 measurements are available for each of

the normal state, damaged state 1 and damaged state 2 in the case study.

7.2.2 Components for the sensor placement optimisation

Before conducting SPO, effective features should be extracted from the output signals of all candidate sensors. One option for generating robust features using FRFs is the multivariate outlier analysis approach introduced in [113]. Here, a frequency range (FR) with a specific resolution can be selected to generate features by using the MSD technique to compute the discordance between an observation and a training set (refer to equation (3.43)).

This kind of feature can naturally account for uncertainty effects in the case that the normal condition data includes measurements under different conditions, designed to be filtered out [13]. Therefore, the normal condition data for 150 random samples of the plate structure were used to calculate the mean value and the covariance required in equation (3.43).

By comparing the averaged FRFs of the 150 samples for the normal condition and two damaged states separately, the third peak and its vicinity of the FRFs (frequency range between 4.3 Hz and 4.7 Hz), were selected to generate MSD-based features, in which range the normal state data and the damaged state data are more distinguishable.

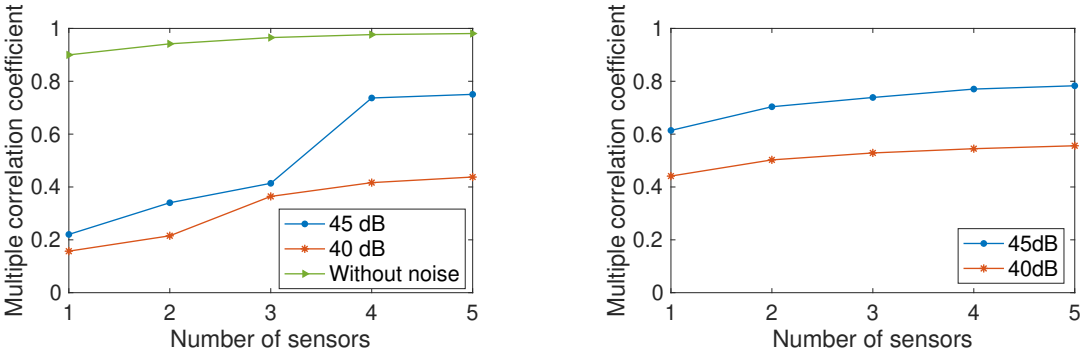
In order to provide an upper bound on performance, an exhaustive search was utilised to search out the optimal sensor combinations with different numbers of sensors. The cost function (or fitness function), was constructed from the MCC. The preferred classification algorithm here to facilitate the final system evaluation is a linear SVM, since it is simple and makes no assumption about the data distribution. The hyperparameters of this algorithm are optimised by minimising five-fold cross-validation loss. When the type of sensor changes, the same optimisation process was repeated.

7.3 Results and discussion

7.3.1 The effect of noise on optimisation

To evaluate whether it is necessary to consider the noise effect in the optimisation process, a set of comparisons based on Damage Case 2 are made here. The effectiveness of a sensor combination is assessed via the MCC.

At first, MSD-based features of 150 sampling models were used to decide a group of optimal sensor combinations with different numbers of sensors. These features only included the influence of uncertain model parameters. The optimal MCC values calculated by using these 300 features (150 normal condition features and 150 damaged state features), are plotted in Figure 7.3a, marked in green. The corresponding optimal sensor combinations determined at the previous step were then applied to calculate the MCC values for features extracted from sensor networks with two different SNRs (45 dB and 40 dB) (See Figure 7.3a for the results).



(a) Data set involving uncertainty.

(b) Data set involving uncertainty and noise.

Figure 7.3: Results of SPO based on different data sets. The optimal MCC values calculated using data set only involving uncertainty are marked in green.

In contrast, the other two groups of optimal sensor combinations with different numbers of sensors were obtained by directly using features involving both parametric uncertainties and noise with different levels (including 45 dB and 40 dB). The corresponding results are shown in Figure 7.3b.

By comparing the results in the above two figures, it can be seen that the MCC values in Figure 7.3a for two conditions with different SNRs are always smaller than the MCC values in Figure 7.3b for the same two conditions. Thus, the SPO using

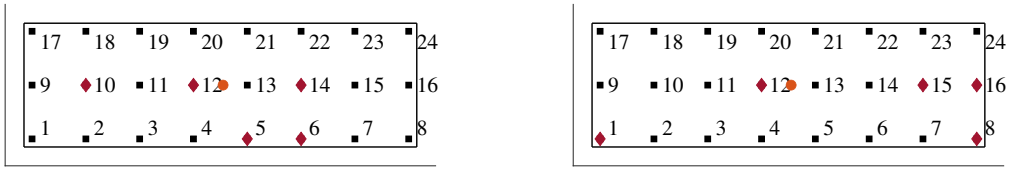
features only involving parametric uncertainties can not provide optimal results when the background noise is considered in the simulation of sensor outputs. This result indicates that, as expected, the noise has a significant impact on the result of optimisation. Therefore, at the design stage of a sensor system, the influence of noise should be incorporated into the optimisation process.

7.3.2 Sensor placement optimisation for binary classification

The normal condition and second damage cases were taken as an example to carry out the SPO for a binary classification. For four situations with different SNRs, an exhaustive search was used to find the optimal sensor combinations with the number of sensors ranging from 1 to 5. The optimal results are presented in Table 7.2, and the corresponding MCC values can be referred to in Figure 7.5a. The optimal sensor distributions corresponding to the highest and lowest SNR are shown in Figure 7.4. It can be seen that as the SNR decreases, the distribution of sensors moves from the nodes centred on the Damage 2 location to the nodes providing high-amplitude signals.

Table 7.2: Optimal sensor combinations for the detection of Damage 2.

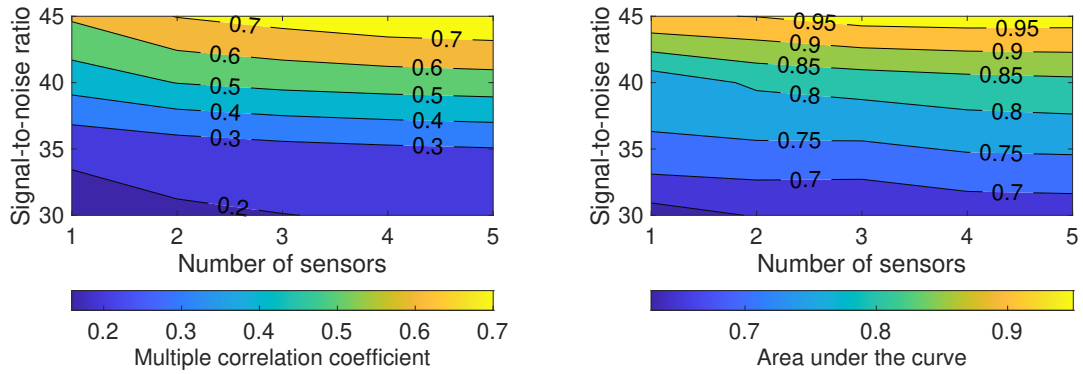
No. of sensors	Signal-to-noise ratio			
	45 dB	40 dB	35 dB	30 dB
1	12	11	9	1
2	12,14	9,11	9,16	1,16
3	9,12,14	9,11,17	9,15,16	1,15,16
4	5,10,12,14	1,9,11,17	9,15,16,20	1,8,15,16
5	5,6,10,12,14	1,9,11,17,18	5,9,15,16,22	1,8,12,15,16



(a) 45 dB. (b) 30 dB.

Figure 7.4: Two examples of optimal sensor distributions for the detection of Damage 2. The selected sensor locations are marked in red. The Damage 2 location is marked in orange.

After determining the optimal sensor combinations with different numbers of sensors, the ROC curves for these optimal sensor networks can be calculated to finally decide



(a) Contour plot of MCC.

(b) Contour plot of AUC.

Figure 7.5: Two types of contour plots for binary classification.

the type and number of sensors to be installed. To compare these curves conveniently, the values of AUC are calculated and plotted in Figure 7.5b.

Figures 7.5a and 7.5b indicate that those sensor combinations with higher MCC values have a higher AUC. Therefore, it is reasonable to use the MCC as a criterion to effectively determine the spatial combination of sensors. However, MCC values can not reflect the performance of an established sensor system intuitively. Therefore, it is necessary to calculate the ROC curve and AUC value via the selected classification algorithm to facilitate the final decision on whether or not to implement the system.

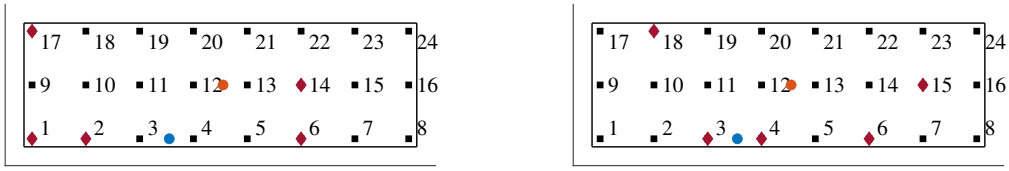
7.3.3 Sensor placement optimisation for multiclass classification

The optimal spatial distributions of sensors for detecting Damage Cases 1 and 2 are presented in Table 7.3 and Figure 7.6. The optimal objective function values (MCC) are shown in Figure 7.7a. For this multiclass classification task, the selected sensor distribution was found to exhibit the following phenomenon. When the introduced SNR is high, the selected sensors will be respectively distributed around the two areas centred on the two damaged locations. As the noise level increases (SNR decreases), the distance between these two areas becomes smaller. The five-fold cross-validation classification accuracy of the SHM systems composed of these optimal sensor networks is calculated and plotted in Figure 7.7b.

As shown in Figures 7.7a, 7.7b, the contour plots of MCC and accuracy achieved by

Table 7.3: Optimal sensor combinations for the detection of Damage 1 and 2.

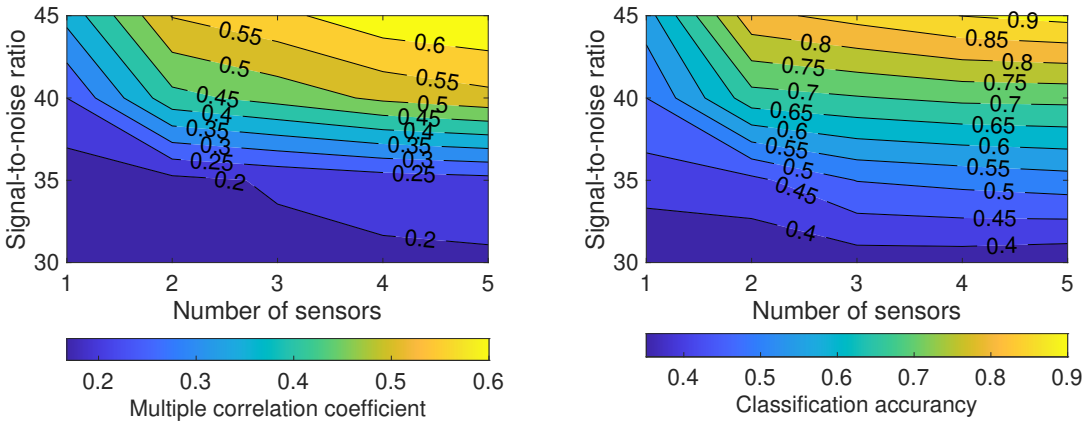
No. of sensors	Signal-to-noise ratio			
	45 dB	40 dB	35 dB	30 dB
1	2	1	3	3
2	2,18	1,13	3,15	3,15
3	1,2,18	1,2,13	1,3,18	3,4,15
4	1,2,14,17	1,2,5,13	1,3,15,18	3,4,12,15
5	1,2,6,14,17	1,2,5,12,13	1,3,5,15,18	3,4,6,15,18



(a) 45 dB. (b) 30 dB.

Figure 7.6: Two examples of optimal sensor distributions for the detection of Damage 1 and 2. The selected sensor locations are marked in red. The Damage 1 location is marked in blue. The Damage 2 location is marked in orange.

the fully-trained classifier exhibit similar trends across the number of sensors and sensor accuracy. Therefore, it is deemed practical to use MCC as a low-cost criterion for the multiclass classification to decide a spatial sensor deployment.



(a) Contour plot of MCC. (b) Contour plot of accuracy.

Figure 7.7: Two types of contour plots for multiclass classification.

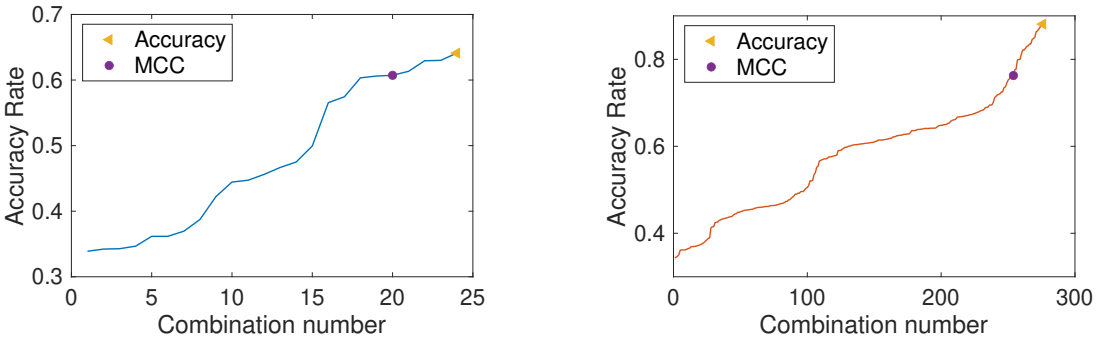
7.3.4 Performance comparison of assessment criteria for two stages

In order to quantitatively measure the damage detection performance of the sensors selected by MCC and the efficiency of the MCC as the selection criteria, a set of comparisons is made here. First, accuracy is directly used as the criterion to search out the optimal sensor combinations. In order to reduce the hyperparameter optimisation time to make it possible to calculate the accuracy of all possible sensor combinations, each model sample under one noise level uses only four measurements to construct the data set used here. Therefore, the number of measurements for the normal state, Damage State 1 and Damage State 2 is 600 respectively. One- and two-sensor location(s) are selected from 24 candidates. The considered SNR is 45 dB. At this noise level, the accuracy of all possible sensor combinations is shown in Figure 7.8, and arranged in ascending order. The optimal results are presented in Table 7.4.

On the contrary, for the same data set, the accuracy values of the best sensor combinations determined based on the MCC are calculated, and the optimal results are given in Table 7.4 and Figure 7.8.

Table 7.4: Optimal results and calculation time for the two assessment criteria.

Criterion	No. of sensors					
	1	Accuracy rate	Time (s)	2	Accuracy rate	Time (s)
Accuracy	9	64%	5.82×10^{3}	2,12	88%	2.00×10^{5}
MCC	2	61%	0.37	2,19	76%	1.11



(a) One sensor (b) Two sensors

Figure 7.8: Sorting accuracy and optimal results of two assessment criteria. Blue and orange correspond to the selection of one or two sensors.

When one sensor is chosen, the accuracy of the optimal sensor combination obtained by using MCC is 3% less than the best result received by directly using accuracy as the objective function. When two sensors are chosen, the difference between them increases to 12%. Therefore, more sensors would be required if the proposed assessment framework is used to design a sensor system with expected performance. Furthermore, a technique that helps to select a better criterion in Stage 1 can be developed to reduce the performance difference between the two stages.

When it comes to the calculation time, it can be seen that compared with the accuracy-based method, the MCC-based optimisation reduces the required computation time by several orders of magnitude. The difference in the calculation time also increases with the number of selected sensors. Therefore, in practical applications with a large number of candidates, the MCC-based criterion can be applied to reduce the number of combinations, and accuracy can further refine the selection of sensor combinations with different sensor numbers and sensor types.

7.4 Conclusions

In this chapter, a systematic framework for sensor system design combined with a machine learning algorithm has been established. Briefly, the main contribution is to divide the SPO into two stages, including selecting the best sensor combination and determining the number of sensors and sensor types, and at the same time, suggesting appropriate evaluation criteria for each stage.

There are key distinctions between the cost functions used for two different stages of a sensor-system design. When the task is to select an optimal sensor combination from a large number of candidates, the cost functions need to be quick to evaluate to enable efficient optimisation, and it is acceptable for them to be abstracted from the main aim of the system to some extent. For the final system evaluation, to decide the number of sensors and the sensor type, the priority is to make the most accurate evaluation possible of how the system will perform in practice.

A case study of a plate structure was carried out. The parametric uncertainty and background noise were involved in the optimisation process to consider the discrepancy between numerical simulation and the real situation. The results indicate that the proposed assessment system reduces the required computation time for the optimal

design by several orders of magnitude while considering the detection performance of the designed sensor system.

Chapter 8

On SPO robust to environmental variations

One challenge in establishing an effective SHM system is the impact of environmental variability on damage identification, which is discussed in Section 1.1.2. It is therefore, advantageous to consider any environmental effects when conducting SPO.

One approach to this problem is to check the robustness of SPO technique to environmental variations and consider whether it is necessary to take account of these environmental factors as part of the optimisation process, which will be demonstrated in Section 8.1. Then, in Section 8.2, practical strategies will be applied to improve the robustness of the SPO results, which are sensitive to environmental variations.

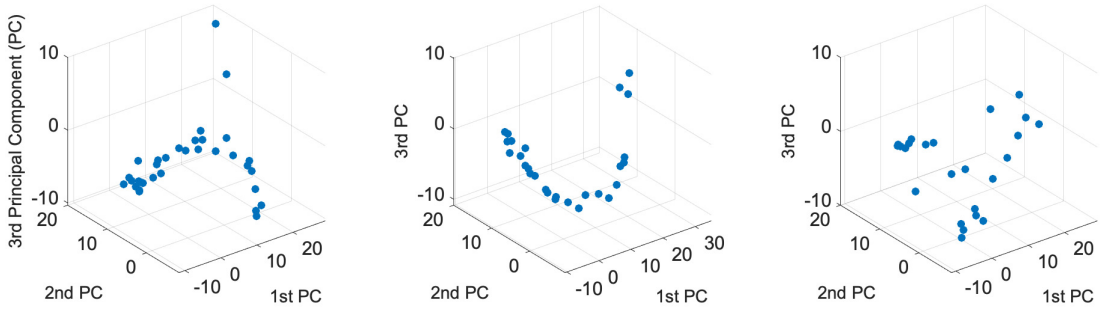
8.1 On robustness of optimal sensor placement to environmental variations

This section studies the robustness of an SPO method to variations of temperature around the given structure. The dataset from the second-round tests performed on a glider wing is suitable for this research. Damage is simulated by adding a mass block (60g) onto the structure, and two damage cases are considered: mass addition at locations M1 and M2. Features extracted from time-series data, frequency-domain

data and modal parameters are used for SPO respectively.

8.1.1 Feature sets extracted from three types of data

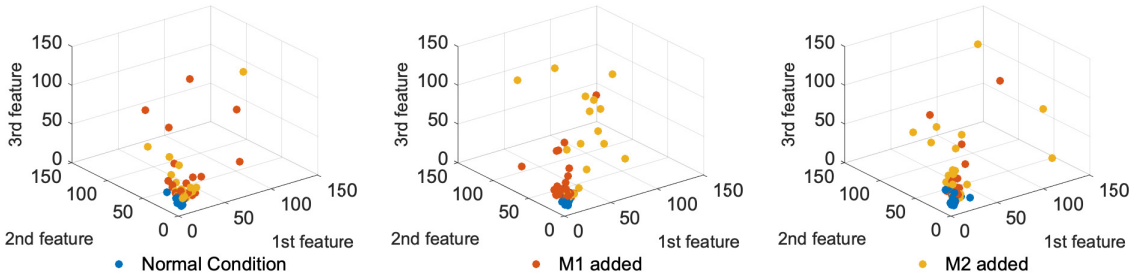
For the time-series data set, 80-second time series collected during each 5°C sweep-down process (see Chapter 5) were used to obtain the coefficients of an autoregressive (AR) model as features. The actual time series adopted were from 10 seconds to 70 seconds to avoid the effects of device startup and termination; They were also preprocessed using a simple moving average (SMA) to eliminate noise from the data. Each time series for each temperature was fitted with an AR model with 20 lags, and the coefficients for all lags were used as features. The first three principal components of features from all sensors for three temperatures are shown in Figure 8.1. More details about extracting features from time-series data using an SMA and an AR model are in Appendices A.1 and A.2.



(a) 5 °C (b) 15 °C (c) 25 °C

Figure 8.1: The first three principal components of the AR model-based features extracted from all sensors.

For the FRF data set consisting of 25 single-average FRFs for each structural state at each temperature, the MSD was calculated for a selected frequency range to form a feature matrix, which can naturally eliminate the noise effect when all data have a similar noise level [114]. The adopted frequency range was 10 Hz to 50 Hz, covering the three natural frequencies corresponding to the three selected mode shapes. A feature-bagging method is also used to avoid the singular covariance used in the MSD calculation by sampling spectral lines with repetition among a relatively-broad frequency range and averaging the features calculated from M sample data sets of spectral lines [114]. Features from three sensors 5, 14 and 31, collected at three temperatures, are given as examples, as plotted in Figure 8.2.



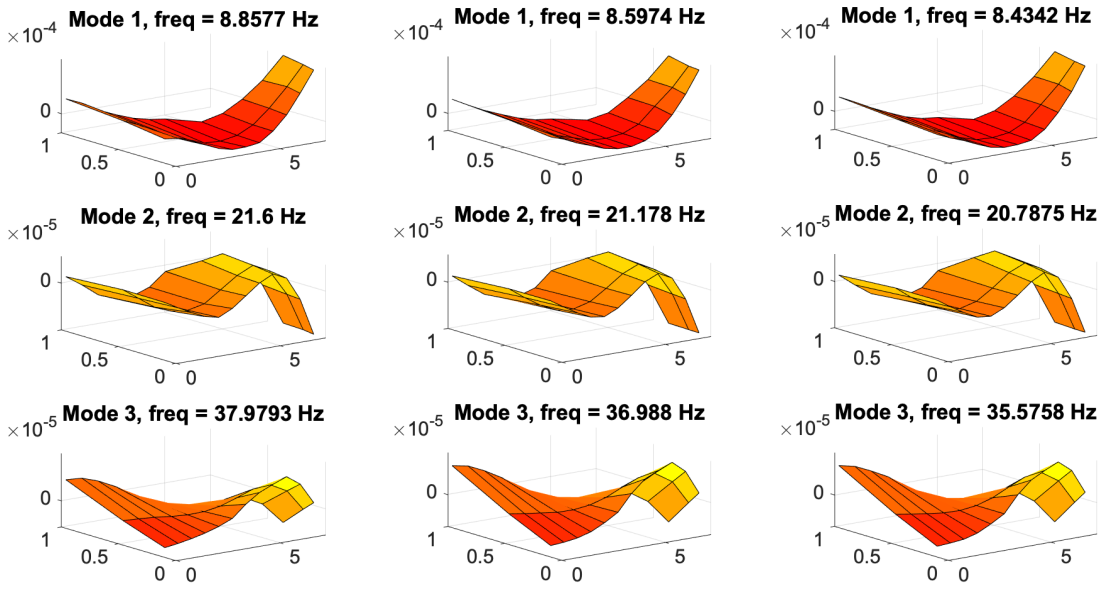
(a) 5 °C (b) 15 °C (c) 25 °C

Figure 8.2: The MSD-based features extracted from three sensors 5, 14 and 31.

For the mode shape data set, the first three mode shapes were chosen and extracted from the 20-average FRFs to form the target mode matrix for each temperature. These mode shapes and corresponding natural frequencies at three different temperatures are shown in Figure 8.3. It can be seen that as the ambient temperature increases from 5 °C to 25 °C, the natural frequencies of the three modes decrease by 4.78%, 3.76% and 6.33% separately. Furthermore, the difference in mode shapes is hardly observed in Figure 8.3 because the modal shape is a feature that can reflect the global change of the structure, and the change caused by the temperature variation is not significant enough to be regarded as a substantial global change.

8.1.2 Optimisation techniques

For the SPO using features from an AR model, a criterion based on the canonical correlation coefficient and the PCA is applied, which was introduced in Chapter 6, Section 6.3. Only the normal condition time series are needed to provide features. For the SPO using the MSD-based features, another criterion based on the canonical correlation coefficient is applied, which was introduced in Chapter 6, Section 6.1. The FRFs and corresponding labels for the normal condition and three damage states are required to provide information for the SPO stage. A fast calculation approach for these two criteria, combined with a greedy search is used to find the potentially sub-optimal solution for different temperatures. For the SPO using mode shapes as features, a trace-based criterion (introduced in Chapter 4, Section 4.2), combined with an iterative technique is adopted to find a sub-optimal result under different temperatures. Only the mode shapes for the normal condition are needed.



(a) 5 °C (b) 15 °C (c) 25 °C

Figure 8.3: The first three mode shapes of the glider wing.

These three SPO technologies may sacrifice the performance of the designed sensor system to a certain extent, but they have higher computational efficiency; This means they are still practical when the size of the search space is large, which is common in practice. In the following part, the word 'sub-optimal' is replaced with 'optimal', keeping in mind the greedy nature of the algorithms. Here, three to ten sensors are selected in all cases.

8.1.3 Robustness of SPO based on coefficients of an AR model

Two strategies are used here to do the SPO. The same optimisation technique combing SSC and PCA is implemented in both cases. The first one is to select the optimal features from all sensors directly. The components capturing at least 95% of the total variance for all controlled temperatures are used to form a label matrix $\mathbf{Y}$. The second strategy is to select optimal features from 20 features for each sensor first, then select the optimal features among the selected features from all sensors. In the first step, the features with a *squared canonical correlation coefficient* larger than 0.9 at all controlled temperatures are selected to form feature matrices for the

second step. Seven of the 20 features from each sensor are selected. The components of the derived label matrix $\mathbf{Y}$ capturing at least 95% of the total variance for all controlled temperatures are used in both steps. Both strategies take into account the interaction between features from different sensors. However, a major difference between these SPO strategies is that, the interaction of features from one sensor is first considered for the second strategy.

The optimal results of these two SPO techniques based on the AR model coefficients are shown in Figures 8.4 and 8.5. When considering all the optimal sensor distributions with the number of sensors from three to ten, there are 25 and 29 sensors contained in these optimal distributions designed for the two strategies respectively; this means that the optimisation results are sensitive to temperature changes. Furthermore, as the number of selected sensors increases, the number of selected sensors obtained by the one-step approach changes from smaller to larger than that obtained by the two-step approach. This phenomenon means that as the number of selected sensors increases, the relationship between features from different sensors is more susceptible to temperature changes than the relationship between features from the same sensor.

8.1.4 Robustness of SPO based on the MSD

The responses in this case study are multi-class categorical labels, which do not have an inherent order. Thus, these labels can be dummy encoded to a matrix $\mathbf{Y}$, in which each category is transformed into a set of binary variables to use m variables to represent $(m + 1)$ categories.

The optimal results of the MSD-based SPO designed for M1 detection are shown in Figure 8.6. The results optimised for the detection of M1 and M2 are shown in Figure 8.7. When considering all the optimal sensor distributions with the number of sensors from three to ten, there are 32 and 31 sensors contained in these optimal distributions designed for the two mentioned tasks respectively. Furthermore, no sensor is always selected at all temperatures for any cases. Therefore, for the SPO with the MSD-based features, the temperature effect should obviously be considered in the optimisation process.

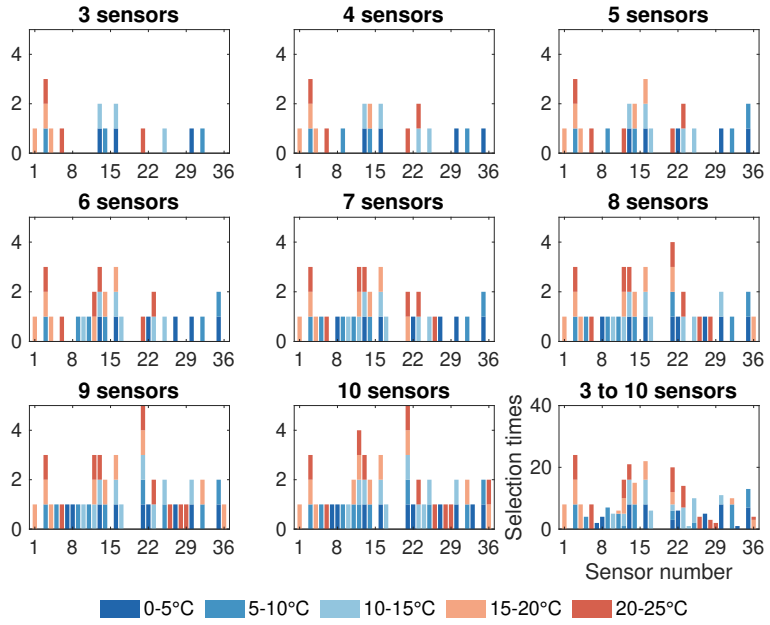


Figure 8.4: Histogram of the SSC-PCA-based optimal sensor distributions obtained by one step.

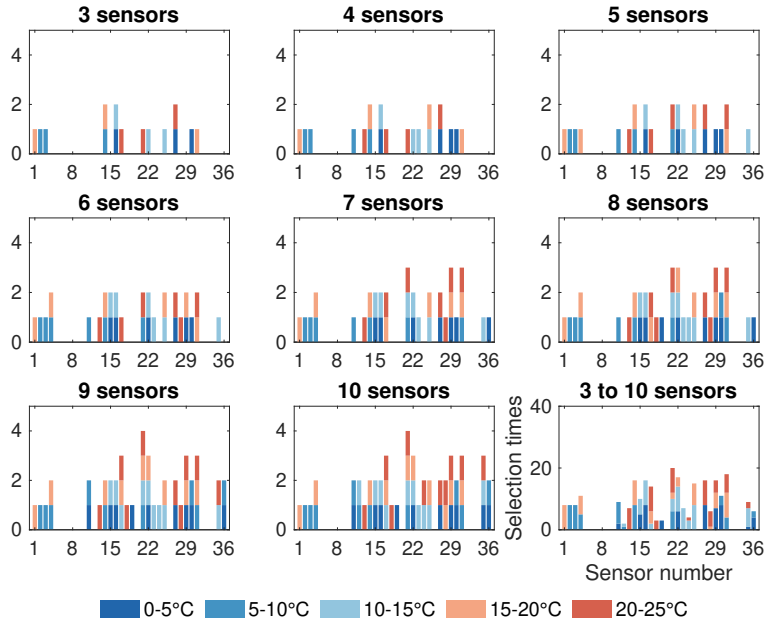


Figure 8.5: Histogram of the SSC-PCA-based optimal sensor distributions obtained by two steps.

8.1.5 Robustness of SPO based on mode shapes

The optimal results for the SPO based on mode shapes are shown in Figure 8.8. It can be seen that the obtained sensor distributions with a fixed number of sensors always contain the same sensor at all controlled temperatures. When considering all the optimal sensor distributions, two sensors, including sensors 10 and 36, will always be selected at all temperatures. The number of sensors contained in these optimal results is obviously smaller than that obtained by the SPO based on the coefficients of an AR model or the MSD-based features. This observation indicates that, since all the OFs are based on the relationship among features, the modal features from the candidate sensors have a more consistent relationship at different temperatures compared to the coefficients of an AR model or the MSD-based features.

The signal strength is then taken into account in the optimisation process by using the modified distribution vector $\mathbf{E_d^{EI-DPR}}$. The optimal results are visualised in Figure 8.9. It was found that there is a significant increase in the number of the same sensors contained in the optimal sensor distributions corresponding to different temperatures. When considering all the optimal sensor distributions, although there are still two sensors (12 and 36) selected at all temperatures, the total number of selected sensors decreases from 16 to 13.

Compared with the EI optimisation results, the results corresponding to EI-ADPR are more robust to temperature variations. This phenomenon suggests that, compared with mode shapes, the ordering of signal amplitudes is more insensitive to temperature changes because the multiplication of the signal amplitude by the EI-based OF makes the optimisation results more consistent across different temperatures. In summary, it is acceptable to ignore the temperature factor in the SPO based on the mode shapes, especially for the case with the signal strength considered.

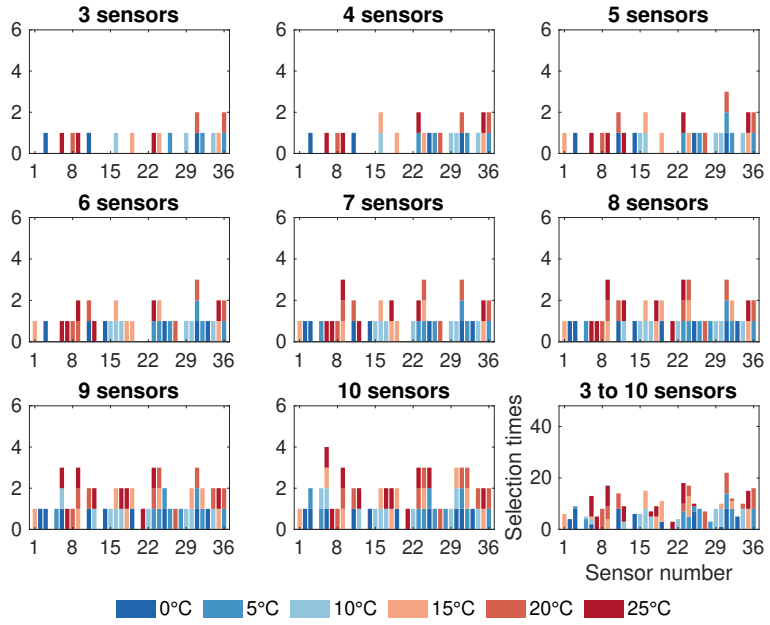


Figure 8.6: Histogram of the SSC-based optimal sensor distributions for M1 detection.

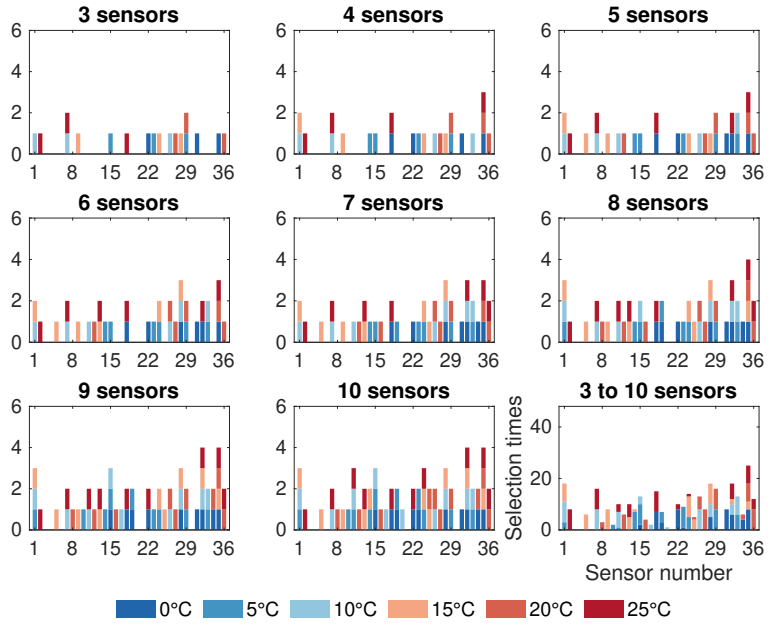


Figure 8.7: Histogram of the SSC-based optimal sensor distributions for M1M2 detection.

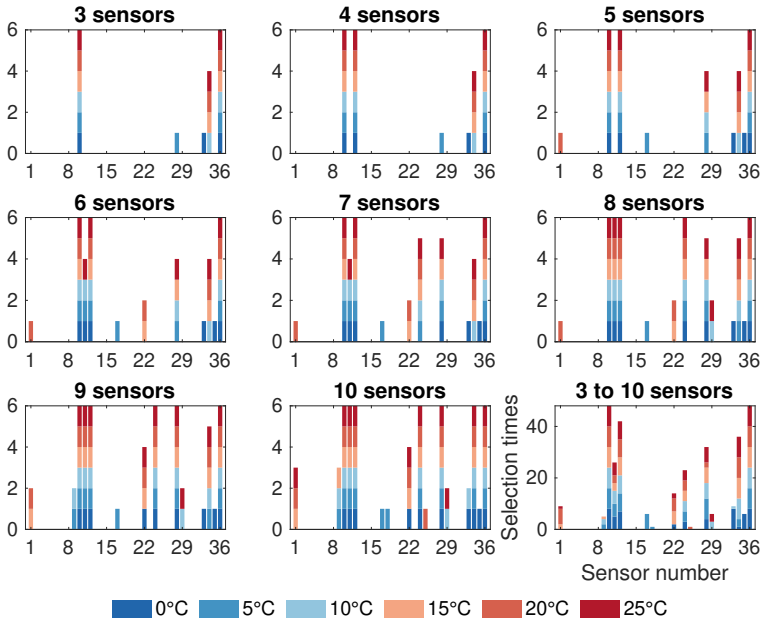


Figure 8.8: Histogram of the EI-based optimal sensor distributions.

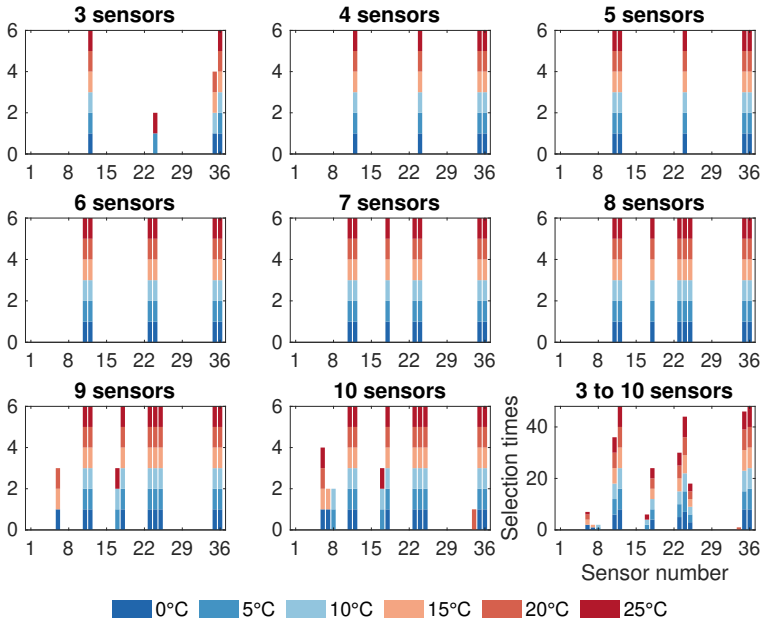


Figure 8.9: Histogram of the EI-ADPR-based optimal sensor distributions.

8.2 A general framework of SPO considering environmental variations

For situations where environmental factors need to be considered, two strategies, considering environmental variations in the optimum design of the sensor deployment in an SHM system, are developed and evaluated in this section. The two strategies both involve constructing an objective function (OF) based upon an established damage classification technique and an optimisation of sensor locations using a GA.

The key difference between the two strategies explored here is whether any sources of benign variation are deemed to be observable or not. To demonstrate the whole process more concretely, a set of data and the corresponding temperature labels gathered from the first test on a glider wing are taken as an example to establish a general framework to consider a certain environmental effect in SPO.

8.2.1 Feature derivation

For laboratory-based studies, if the temperature of a structure can be controlled and recorded, then both the normal condition data and damage data will be labelled with the corresponding temperature. Thus, features for the normal condition data and damage data at each temperature can be calculated respectively. This means that these temperature labels are used to clarify the effect of temperature on features, which can help identify the effect of damage at a specific temperature. On this basis, a distance calculation approach can be taken to calculate features. The noise effect on the high-averaged data can be ignored, so it is assumed acceptable to use only the mean value of the high-averaged data set to represent the data set itself. As there is thus no uncertainty associated with the feature values, the feature comparisons are between crisp numbers at equal temperatures and it is sufficient to use the Euclidean squared-distance (ESD) as the comparison metric between feature vectors. Following the methodology given in equation (8.1), the ESD statistic D_{E} is calculated as,

$$D_{\mathrm{E}}^2 = \sum_{i=1}^{n} (x_i - \mu_i)^2 \tag{8.1}$$

where x_i is the amplitude value of the i^{th} spectral line, μ_i is the averaged amplitude value of a spectral line set corresponding to a frequency from an observation set and n is the dimension of a sample, i.e. the number of spectral lines in a sample.

Note that the ESD-based features here are feature clusters with corresponding temperature labels, representing an idealised baseline for comparisons. This option is possible here because of the precise temperature control and recording in the experiments. In general, high levels of noise or limited measurements for averaging will result in uncertainties on the features which should be taken into account. If sufficient measurements are available to estimate the covariance, the MSD can be used to quantify the discordancy. In practice, for the monitoring of *in situ* structures, the ambient temperature will be uncontrollable and may not be recorded.

In situations where temperature measurements are not available for feature vectors, the influence of temperature can still be removed by the linear technique introduced in Chapter 3, Section 3.2. To make possible the existence of a linear subspace that can account for the confounding effects and avoid the singular covariance used in the MSD calculation, a feature-bagging method can be used.

In this thesis, a feature-bagging method is used to objectively derive effective features within a specific frequency range (FR) for FRFs from each sensor. In general terms, a feature-bagging process is an average of all basic features derived from sampled data sets [115]. Here, it is applied by averaging the distance features calculated from M-sample data sets of spectral lines from a relatively-large FR. After obtaining a committee model with M feature values, the average value D_{A}^2 of these features for an observation from one sensor can be calculated as,

$$D_{\mathrm{A}}^2 = \frac{1}{M} \sum_{m=1}^{M} D_m^2 \tag{8.2}$$

where D_m^2 is the feature (squared distance) value corresponding to the m^{th} sampled set of spectral lines.

To generate M sample data sets of spectral lines, bootstrap sampling is applied in the selected FR to sample spectral lines (independent variables in a feature vector),

with repetition. The sampling size n should be set based on the specific data set. The number of M can be set based on the total number of spectral lines in the initial frequency range, divided by the sampling size. Feature vectors provided by bootstrap sampling over a large FR are high-dimensional enough to make possible the existence of a linear subspace that can account for the confounding effects. Further detail, including pseudo-codes for the MSD-based feature derivation heuristic, may be found in [114]. Such a linear approach, based on the MSD is explored in this part. Note that the MSD-based features in this research are used to represent temperature-insensitive features obtained following the application of the linear projection approach in [13].

8.2.2 Case study data sets

Data used in this section are from the first experiment on the glider wing. These represent quite large damage, in order to produce separated feature clusters, which will be explained in detail later. The feature-bagging method was conducted to generate features using measurements with or without temperature labels in this case study. The sampling size n should be less than the number of observations for the normal condition to avoid a singular covariance matrix in the MSD calculation. Because there are two observations for each normal condition state at each controlled temperature, and data for seven temperatures were recorded, the total number of observations for the normal condition is 14. Therefore, the sampling size n was set to 10 and the covariance determinants were checked to avoid the singular matrix. The number of samples M, was set to be equal to, or slightly larger, than d/n, where d is the dimension of the original FR.

By plotting FRFs for the structure in different health states, an FR from 50 Hz to 250 Hz is selected to generate the sample data sets of spectral lines. Here, FRFs for the structure with the temperature at 15 °C are used as an example. As can be seen in Figure 8.10, this frequency range is effective because it is sensitive to all damage cases, i.e. it is easy to distinguish the FRFs for these three damage cases from the FRF of the normal condition case. The resolution of the FRFs is 0.25 Hz, so the dimension of the original FR space is 801. Thus, the number of samples can be set to 90 here.

The same sampling size and the number of samples are taken for the derivation of the ESD- and MSD-based features, which is beneficial for comparing the optimal

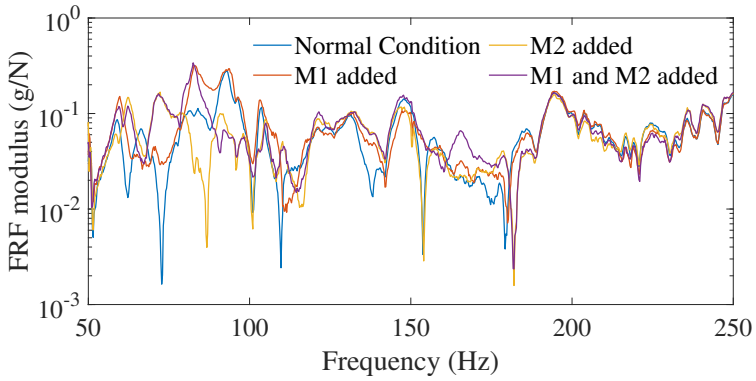


Figure 8.10: An example of the adopted frequency range: FRFs between 50 Hz and 250 Hz collected at 15 °C from sensor 17 for four condition cases.

results when different features are extracted from the same data set. In addition, thanks to the feature bagging method, only one discordance measure is calculated for one observation from each sensor. The selection of a single feature for each sensor based on a relatively-large frequency range is attractive as it leaves aside the feature selection task. Relevant examples of the semi-objective method and the objective method based on the neural network for feature selection can refer to [113] and [116].

8.2.3 Optimisation objective based on an SVM model

To construct an OF having a direct relationship with the damage detection ability of an SHM system, a relationship between an OF and a classifier distinguishing healthy-state data and damaged-state data needs to be established. For the purposes of simplicity and computational efficiency, the research here is limited to a linear classifier. The supervised-learning algorithm used here to build a linear classifier is the SVM technique [117], introduced in Chapter 3. The initial reason for the selection of this classification algorithm is that an SVM makes no assumption about the prior distribution of data, which is difficult to know exactly in practice.

In order to illustrate the concepts of sensor placement optimisation in this chapter, the detection problem itself is deliberately simplified by considering large 'damage' cases. Furthermore, averaged features are used in order to minimise the effects of measurement noise. In this situation, the health states of the structure are strictly separable in the feature space. This is a useful property as it allows a minimal version of the SVM tailored to separable classes and amenable to linear decision boundaries. This choice does not represent a restriction on the sensor-optimisation problem, as

the cases of smaller damage or curved decision boundaries are both addressable with appropriately adapted versions of the SVM [117].

For different input data sets, the SVM will provide different best hyperplanes that match different maximum margin widths. The misclassification rate is also calculated to check the separability. It is obvious that if there is a sensor combination providing a non-zero misclassification rate, it should be abandoned directly. Because there are no hyperparameters to be optimised in this process, the OF can be evaluated quickly. It is beneficial to search out the best sensor combination from a large number of sensor candidates.

In this research, the combination of sensor locations which provides feature vectors that make the healthy-state data and the damaged-state data most separated will be selected as the optimum; i.e. the sensor deployment with the largest max-margin obtained by an SVM. To find the optimal deployment of sensors, an integer GA is adopted here, which can search out the globally-optimal result with variables that are integer-valued. Linear constraints are used to make sure that non-repetitive sensors exist in an optimal sensor deployment. These constraints refer to the fact that the difference between any two selected sensor indices should not be less than 1, which can be expressed as follows,

$$IDX_i - IDX_{i+1} \geq 1, \quad i = 1, \ldots, I-1 \tag{8.3}$$

where IDX_i is the index of a selected sensor and I is the number of selected sensors.

Although three damage cases are considered with mass additions at locations [M1], [M2] and [M1, M2], only binary classification is considered here to distinguish the normal condition case and the three damage cases collectively. In this way, the relationship between the damage position and the sensor deployment can be studied. If more than two structural states need to be considered in the optimisation process, a multiple-classification algorithm can be used to classify the data and an appropriate criterion should be adopted as the optimisation objective, according to the requirements of the project. The details of the specific learning algorithm which delivers the OF values do not affect the general framework proposed.

8.2.4 Feature bagging results and analysis

In order to improve the efficiency of the calculation, it is necessary to normalise the feature vectors before training a classifier by mapping the min and max values of all given dimensions to 0 and 1. The feature-bagging results of the ESD-based features and the MSD-based features for the normal condition case and three damage cases from 36 sensors at seven different temperatures are shown in Figures 8.11 to 8.14.

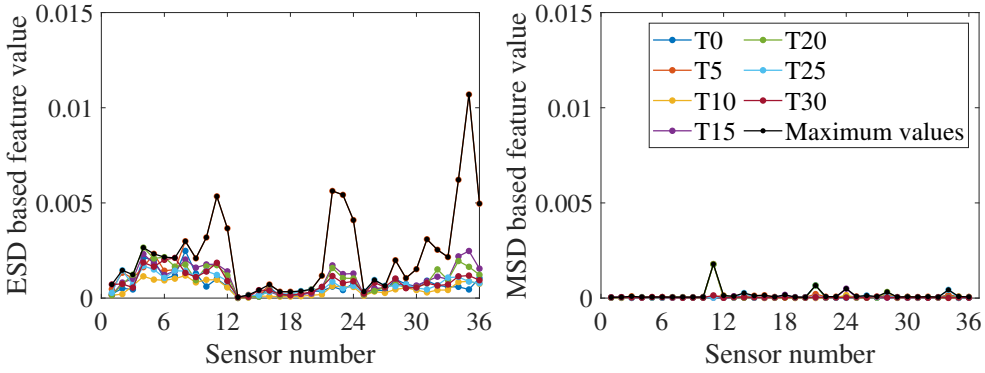


Figure 8.11: Feature-bagging results based on the ESD (left) and the MSD (right) for the normal condition. Seven controlled temperatures T0–T30 and the maximum value among the features for seven different temperatures are shown by colour markers. The ESD-based features refer to features obtained by the normalised approach with temperature labels. The MSD-based features refer to features obtained by the linear approach without temperature labels.

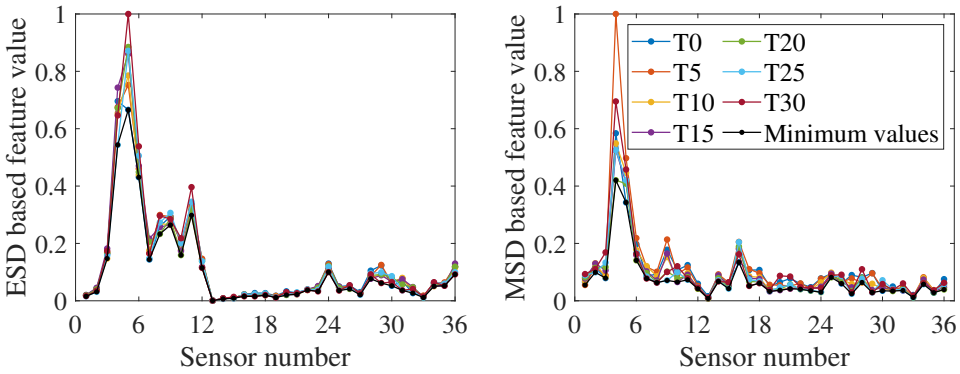


Figure 8.12: Feature-bagging results based on the ESD (left) and the MSD (right) for the damage case with M1 added. Seven controlled temperatures T0–T30 and the minimum value among the features for seven different temperatures are shown by colour markers.

By analysing the positions of sensors providing comparatively large discordance values in different damage cases, it is apparent that data collected from sensors close to the added mass blocks are sensitive to this damage, as may have been expected.

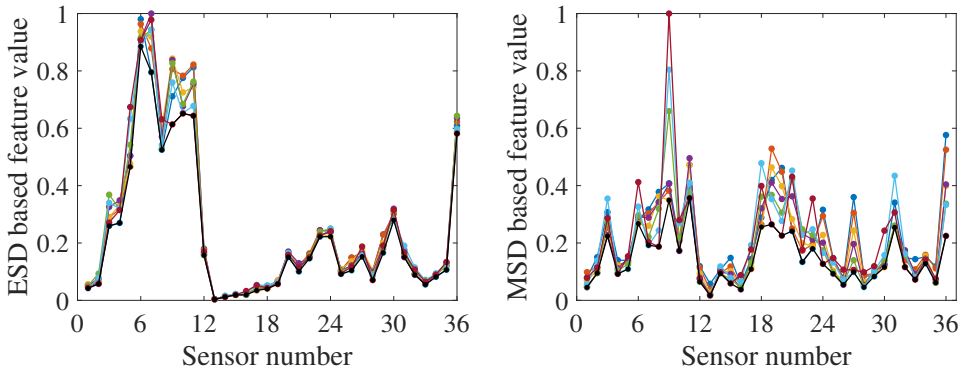


Figure 8.13: Feature-bagging results based on the ESD (left) and the MSD (right) for the damage case with M2 added.

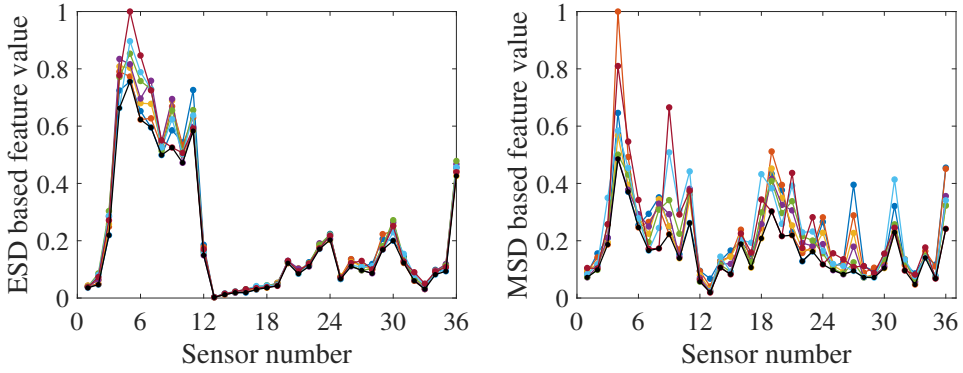


Figure 8.14: Feature-bagging results based on the ESD (left) and the MSD (right) for the damage case with M1 and M2 added.

For example, two of the largest discordance values calculated via the ESD or the MSD for damage cases 1 come from sensors 4 and 5, adjacent to the mass block M1. Furthermore, by comparison of Figure 8.11 and Figures 8.12 to 8.14, it can be seen that the influence of temperature on the two types of features extracted from a subset of sensors is much smaller than the influence of the damage. This observation suggests that both feature extraction techniques can help to provide effective features robust to temperature variations but sensitive to the damage.

Furthermore, from Figures 8.12 to 8.14, it can be seen that a sensor at a fixed location on the gilder wing has different sensitivity to the occurrence of the same damage at different temperatures. A reasonable explanation is that at different temperatures, the physical parameters of various structural components (such as stiffness and cross-sectional area) will change. This variation results in the same part of the structure responding differently to the same damage at different temperatures. Thus, the effect of one type of damage on different locations of the structure at a certain temperature

cannot be applied after the temperature of the structure changes. In addition, because the minimum feature values consist of features for different temperatures, it is impossible to set one temperature as the most unfavourable temperature for damage detection. Therefore, it is necessary to involve the temperature when analysing the structural response and training a classifier.

8.2.5 Optimal results and analysis

The GA is used to optimise sensor sets containing between one and five sensors for detecting three different damage cases separately by using the ESD- or MSD-based features. The resulting optimal combinations are listed in Tables 8.1-8.3. It can be seen that sensor locations selected by the SVM optimisation technique using the ESD-based features are mainly distributed on the leading edge of the glider wing, while the results for the MSD-based features show that optimal sensors are much more scattered on the structure. This phenomenon suggests that the ESD-based features are more sensitive to damage locations than the MSD-based features. More analysis will be given in the following paragraphs.

Table 8.1: Optimal sensor combination designed for the M1 case detection and corresponding max-margin widths.

No. of sensors	ESD	Max-margin width	MSD	Max-margin width.
1	5	0.6634	4	0.4202
2	4, 5	0.9597	4, 5	0.5859
3	4,5,6	1.0511	4,5,16	0.6156
4	4,5,6,11	1.1030	4,5,16,25	0.6210
5	4,5,6,9,11	1.1411	4,5,6,16,25	0.6367

Table 8.2: Optimal sensor combination designed for the M2 case detection and corresponding max-margin widths.

No. of sensors	ESD	Max-margin width	MSD	Max-margin width
1	6	0.8819	11	0.3549
2	6, 7	1.2582	9, 19	0.5775
3	6,7,11	1.4548	9,11,20	0.7082
4	6,7,9,10	1.6012	9,11,19,36	0.8467
5	6,7,9,10,11	1.7229	9,11,19,20,36	0.9283

By comparing the selected sensor locations in Tables 8.1 to 8.3 with the minimum values of features shown in Figures 8.12 to 8.14, an expected phenomenon is observed:

Table 8.3: Optimal sensor combination designed for the M1M2 case detection and corresponding max-margin widths.

No. of sensors	ESD	Max-margin width	MSD	Max-margin width
1	5	0.7534	4	0.4859
2	4, 5	1.0436	4, 19	0.6395
3	4,5,11	1.2666	4,11,19	0.7438
4	4,5,6,11	1.4115	4,5,11,19	0.8344
5	4,5,6,7,11	1.5431	4,5,11,19,36	0.9070

generally, locations corresponding to the larger minimum feature values are selected as the optimal sensor locations. However, it can also be seen that the distribution of feature vectors from one sensor for different temperatures can also affect the results of the SVM optimisation technique.

To demonstrate the effect of feature distribution visually, data from two sensors (9 and 19) selected by the SVM optimisation technique with the MSD-based features adopted to detect Damage Case 2 in Table 8.2 and two sensors (9 and 11) providing the two largest minimum-feature values for the same damage case in Figure 8.13 (left) are used to build two classifiers and calculate the maximum margin widths. The results are shown in Figures 8.15 and 8.16. The max-margin width calculated by the data from sensors 9 and 19 is 0.58, which is larger than the margin width (0.56) corresponding to sensors 9 and 11.

To explore the relationship between the max-margin width corresponding to the optimal sensor combination and the number of selected sensors, the results for the M1 case are taken as an example and demonstrated in Figure 8.17. Here the relative position of the line for SPO with temperature labels is higher than those without temperature labels. A similar phenomenon also occurs for the M2 and M1M2 cases; this indicates that the linear method without using the temperature labels employed to eliminate the influence of temperature will sacrifice sensitivity to damage to a certain extent. This observation is also in line with the expectation that more information can provide better results.

However, one obvious advantage of the linear method is that it is not necessary to strictly control the temperature of a structure to a specific degree, which can be a tricky and/or expensive process even in the limited number of cases where it is feasible. Additionally, the number of measurements can be greatly reduced if the normal condition data for all temperatures are used to provide a non-singular

covariance used in distance calculation. This condition helps to reduce the cost and time of data collection. Therefore, there is a trade-off between the cost and feasibility associated with temperature control and label collection, and the lack of sensitivity associated with the linear method to address the confounding effects conveniently.

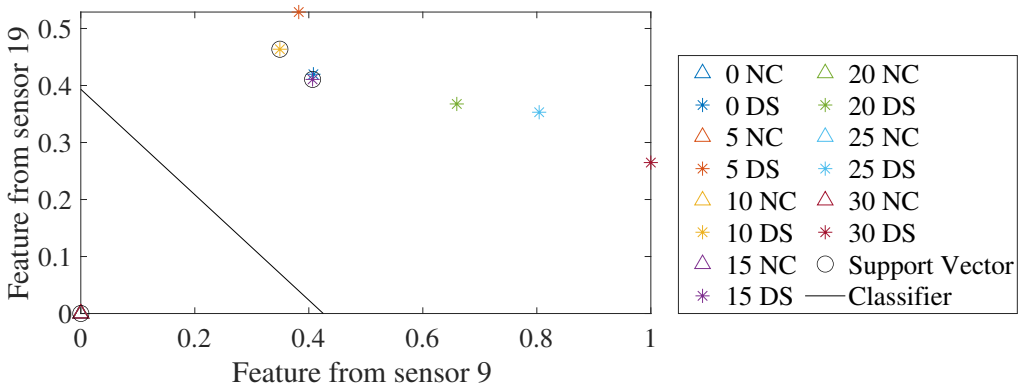


Figure 8.15: Distribution of the MSD-based features from two optimal sensors selected by the SVM optimisation technique. △ refers to the normal condition and ∗ refers to the damaged state.

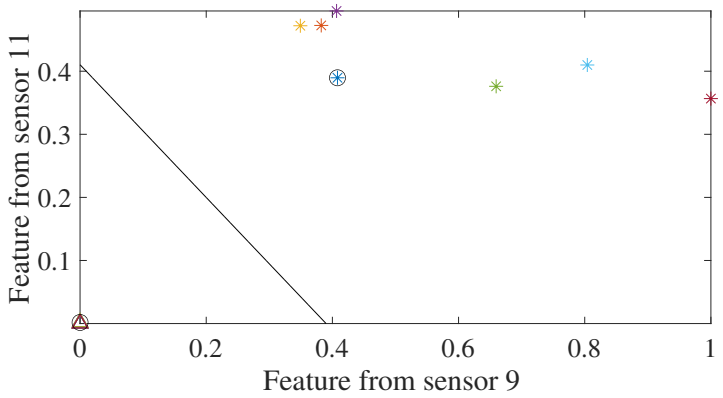


Figure 8.16: Distribution of the MSD-based features from two sensors providing the two largest minimum-feature values.

A method that may warrant investigation, is to partition a large temperature range into several segments, with the linear approach applied within each segment and the normalised approach used to combine the feature values for each temperature segment with the SVM optimisation technique to process SPO.

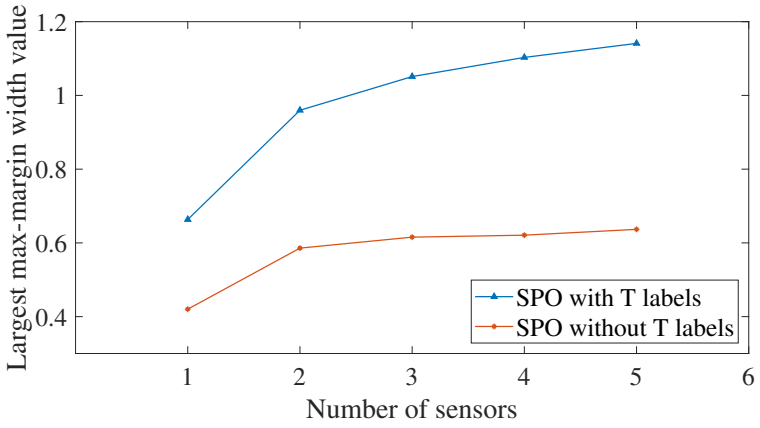


Figure 8.17: Optimal max-margin widths and corresponding number of selected sensors for the M1 case detection.

8.3 Conclusions

This chapter's first main content analysed the robustness of SPO techniques to environmental variations, to illustrate the effect of an environmental parameter in the optimisation process. Three commonly-used features were considered, i.e. the AR model coefficients, the MSD-based features and mode shapes; the entire process was demonstrated in detail with temperature as an example.

Data of a glider wing collected at different temperatures were used for a case study. Mode shapes were combined with EI and EI-ADPR technologies to optimise the sensor combinations for modal identification. In addition, the MSD-based features or the AR model coefficients were combined with a greedy search to optimise the sensor deployments for damage identification. A fast calculation of the optimisation objective, SSC, was adopted to speed up the calculation.

The results mainly show that the modal features from the candidate sensors have a more consistent relationship at different temperatures, especially when the signal amplitude is considered. Hence, the participation of temperature is avoidable for the SPO based on mode shapes. In contrast, the relationship of the MSD-based features or the coefficients of an AR model for different structural states is greatly affected by temperature. Therefore, SPO using the MSD-based features or the AR model coefficients inevitably must take this environmental factor into consideration.

The second topic of this chapter illustrated two SPO techniques designed for damage detection while taking into account temperature effects; the key contributions are:

(1) investigation of two approaches for extracting damage-sensitive features in the presence of either recorded or unrecorded temperature variations and, (2) investigation of appropriate optimisation functions for evaluating the resulting sensor combinations.

A case study of a glider wing shows that, compared to the normalised method using the temperature label, the linear method that does not require temperature labels provides features that are less sensitive to damage. However, it is cheaper and more convenient to extract temperature-robust features in practical engineering.

Meanwhile, the proposed optimisation criterion – maximum margin width – is an effective criterion considering the damage detection ability of the designed SHM system, provided that the health-state classes are separable in the feature space. If this condition is not satisfied, the criterion should be extended to soft-margin SVM classifiers. In addition, the criterion shown in Chapter 6 can also be considered for non-separable situations.

Chapter 9

On SPO robust to sensor failure

In this chapter, the mode shape data of a glider wing are used to study 'fail-safe' sensor distributions, i.e. those whose sub-distributions still have good performance. The performance of the fail-safe sensor distribution with multiple sensors placed in the same position will also be studied. Two types of optimisation objectives are investigated, one for modal identification and the other for damage identification.

The scope of the current chapter belongs to the classical optimisation criteria based on model parameters (see Section 1.2.1.). However, except for the traditional objective functions based on mode shapes, a predefined criterion based on canonical correlations is creatively used to evaluate a sensor layout, which can transfer the functionality of linear discriminant analysis into the parameter estimation issue.

Here, *sensor failure* refers to a situation where the sensor fails to provide any usable information, regardless of whether it is damaged or unable to connect to the network. A wired sensor network is adopted, and the failures of each sensor are independent.

Under this circumstance, in this chapter, the improved fail-safe optimisation method will be proposed to comprehensively consider the fitness of the mother distribution and worst sub-distribution for modal identification and fault detection, which was ignored in [66]. Moreover, wireless sensor networks must handle the effects of sensor failures by placing redundant sensors on critical spots [67]. Therefore, this chapter will discuss whether it is worth designing a wired sensor network with redundant sensors.

9.1 Proposed optimisation objectives

The optimisation objectives used for SPO, depend on the purpose of the system, and include the determinant of the Fisher information matrix (DFIM), the determinant of the FIM weighted by the average-driving-point residue (DFIM-ADPR), the sum of squared canonical correlation coefficients (SSC), and the sum of squared canonical correlation coefficients weighted by the average-driving-point residue (SSC-ADPR). The first two objectives, developed by [42] and [94], are widely used in SPO tasks, but the last two objectives are used for SPO for the first time in this thesis. These objectives are selected or proposed from three perspectives, including, the linear independence of mode shapes, the vibration energy and the information on the damage of interest.

The criterion based on canonical correlations was used in SPO for the first time in Chapter 6. However, because of the large content of that chapter, in order to avoid unclear focus, this section reviews the related studies and explains in detail the reasons for adopting the SSC criterion in SPO.

Firstly, the SSC metric is usually more computationally efficient than criteria depending on model training. In existing research, correlation-based criteria have been widely applied in feature selection [101, 118, 119], and some efforts have been made to extend its application to the SPO field. Chen *et al.* [120] defined a coefficient to quantify the geometric correlation of probability distributions of data collected by pairs of sensors. Sensor pairs with larger correlation indexes were selected to provide the data for predicting the shapes of all probability density functions. However, the redundancy and interaction of information from different sensors were not considered in this study. Lu *et al.* [121] narrowed the range of sensor selection by maximising the coefficient based on a correlation matrix and then minimising the correlation of the data provided by these candidate sensors to determine the final optimisation result, but still ignored the information interaction. In contrast, the criteria based on canonical correlations can inherently take account of the information redundancy and interaction [101, 118], which is another reason for using this kind of criterion in this chapter.

For a detailed introduction to the first two criteria, please refer to Chapter 4. See Chapter 6 for a fast calculation of SCC. The standard calculation process of SSC and SSC-ADPR is described as follows.

9.1.1 Sum of squared canonical correlation coefficients

For the damage identification task, to make the sensor system have good performance in identifying certain damage states of concern, information about these high-potential failures should be included in the sensor system optimisation process. These damage states, such as damage type, location and severity, can be determined by conducting failure mode analysis and referring to historical data about damage states collected on the existing structures.

The SSC is used here as a criterion to measure the maximal linear association between two co-occurring multivariate random variables with N observations, which are collected as matrices $\mathbf{X} \in \mathbb{R}^{N\times n}$ and $\mathbf{Y} \in \mathbb{R}^{N\times m}$ [76]. Here, a matrix $\mathbf{X}$ represents the N samples of n features. A matrix $\mathbf{Y}$ refers to labels represented by N variables associated with each sample. This criterion can help to directly measure the effectiveness of data for predicting structural states to a certain extent, without the need for model training. Furthermore, since a set of features in $\mathbf{X}$ is evaluated together, rather than evaluating each feature individually, the interaction of features can be considered, avoiding redundant features.

The acquisition of the SSC can be realised by finding pairs of projection directions $\mathbf{U} \in \mathbb{R}^{n\times L}$ and $\mathbf{V} \in \mathbb{R}^{m\times L}$, so that the Pearson's correlation coefficient between $\mathbf{X}_{\mathrm{M}}\mathbf{u}_l$ and $\mathbf{Y}_{\mathrm{M}}\mathbf{v}_l$ for the l^{th} pair of projection directions is maximised [77], i.e.

$$\begin{aligned} R_l(\mathbf{X},\mathbf{Y}) &\triangleq \max_{\mathbf{u}_l,\mathbf{v}_l} r(\mathbf{X}\mathbf{u}_l,\mathbf{Y}\mathbf{v}_l) \\ &\triangleq \max_{\mathbf{u}_l,\mathbf{v}_l} \frac{\mathbf{u}_l^\top \mathbf{X}_{\mathrm{M}}^\top \mathbf{Y}_{\mathrm{M}}\mathbf{v}_l}{\sqrt{\mathbf{u}_l^\top \mathbf{X}_{\mathrm{M}}^\top \mathbf{X}_{\mathrm{M}}\mathbf{u}_l}\sqrt{\mathbf{v}_l^\top \mathbf{Y}_{\mathrm{M}}^\top \mathbf{Y}_{\mathrm{M}}\mathbf{v}_l}} \end{aligned} \tag{9.1}$$

where $\mathbf{X}_{\mathrm{M}} = \mathbf{X} - \overline{\mathbf{X}}$ and $\mathbf{Y}_{\mathrm{M}} = \mathbf{Y} - \overline{\mathbf{Y}}$. $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are the means of columns of $\mathbf{X}$ and $\mathbf{Y}$ respectively. If the $\mathbf{X}\mathbf{u}_l$ and $\mathbf{Y}\mathbf{v}_l$ are standardised to have unit variance, equation (9.1) can be expressed as,

$$\max_{\mathbf{u}_l,\mathbf{v}_l} \mathbf{u}_l^\top \mathbf{X}_{\mathrm{M}}^\top \mathbf{Y}_{\mathrm{M}}\mathbf{v}_l, \quad \text{s.t. } \mathbf{u}_l^\top \mathbf{X}_{\mathrm{M}}^\top \mathbf{X}_{\mathrm{M}}\mathbf{u}_l = 1,\ \mathbf{v}_l^\top \mathbf{Y}_{\mathrm{M}}^\top \mathbf{Y}_{\mathrm{M}}\mathbf{v}_l = 1 \tag{9.2}$$

According to Chapter 4, the corresponding projection directions $\mathbf{u}_l \in \mathbb{R}^{n\times 1}$ and $\mathbf{v}_l \in \mathbb{R}^{m\times 1}$ can be obtained by solving the eigenvalue equations given by [122],

$$\left[(\mathbf{X}_{\mathrm{M}}\mathbf{X}_{\mathrm{M}}^{\top})^{-1}\mathbf{X}_{\mathrm{M}}\mathbf{Y}_{\mathrm{M}}^{\top}(\mathbf{Y}_{\mathrm{M}}\mathbf{Y}_{\mathrm{M}}^{\top})^{-1}\mathbf{Y}_{\mathrm{M}}\mathbf{X}_{\mathrm{M}}^{\top} - \lambda\mathbf{I}\right]\mathbf{u}_l = \mathbf{0} \tag{9.3a}$$

$$\left[(\mathbf{Y}_{\mathrm{M}}\mathbf{Y}_{\mathrm{M}}^{\top})^{-1}\mathbf{Y}_{\mathrm{M}}\mathbf{X}_{\mathrm{M}}^{\top}(\mathbf{X}_{\mathrm{M}}\mathbf{X}_{\mathrm{M}}^{\top})^{-1}\mathbf{X}_{\mathrm{M}}\mathbf{Y}_{\mathrm{M}}^{\top} - \lambda\mathbf{I}\right]\mathbf{v}_l = \mathbf{0} \tag{9.3b}$$

where the eigenvalue λ equals the $R_l^2(\mathbf{X}, \mathbf{Y})$, and the number of the non-zero eigenvalues is not greater than $\min(n,\, m)$, that is, L is not greater than $\min(n,\, m)$. Since $R_l(\mathbf{X}, \mathbf{Y})$ ranges between 0 and 1, the eigenvalue $R_l^2(\mathbf{X}, \mathbf{Y})$ takes values between 0 and 1.

On this basis, the sum of non-zero eigenvalues, i.e. SSC, is adopted as a criterion for evaluating the effectiveness of a feature set from a sensor set to infer the structural state, which is given by,

$$\begin{aligned}\sum_{l=1}^{L} R_l^2(\mathbf{X}, \mathbf{Y}) &= \mathrm{tr}\left[(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}\right] \\ &= \mathrm{tr}\left[(\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}})^{-1}\mathbf{Y}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}(\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}})^{-1}\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{Y}_{\mathrm{M}}\right]\end{aligned} \tag{9.4}$$

where $\mathrm{tr}[\,\cdot\,]$ represents the trace of a matrix. The computational complexity of equation (9.4) is $\mathcal{O}(n^2N)$ when $m \leq n$ or $\mathcal{O}(m^2N)$ when $m > n$. It can be seen that the optimal sensor deployment with the largest SSC among sensor combinations with the same number of sensors is conducive to the most accurate estimation of the structural labels because the features are more correlated with labels.

9.1.2 SSC average driving-point residue

If mode shapes are used as features for SPO, with the aim of damage identification, the selected n mode shapes collected under different structural states can form a feature matrix $\mathbf{X}_s \in \mathbb{R}^{K\times(I\times J)}$ corresponding to the s^{th} combination of I sensors. Rows refer to K observations for different health states, and columns refer to the relative displacement of I selected sensor locations on a structure in J modes. What is worth noting is that if only one mode shape is going to be considered, i.e. $J = 1$, the feature corresponding to i^{th} sensor location will be a vector $\mathbf{x}_s^i \in \mathbb{R}^{K\times 1}$; if more than one mode shape is considered, i.e. $J > 1$, the feature corresponding to i^{th} sensor location will be a matrix $\mathbf{X}_s^i \in \mathbb{R}^{K\times J}$.

Similarly, the ADPR coefficient can be combined with SSC to improve the stability

of the optimal sensor layout to noise, which is given by,

$$\sum_{l=1}^{L} R_l^2(\mathbf{X}_s, \mathbf{Y}_s)_{\mathrm{ADPR}} = \sum_{l=1}^{L} R_l^2(\mathbf{X}_s, \mathbf{Y}_s) \sum_{k=1}^{K} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(x_s^{kij})^2}{\omega_j^k} \tag{9.5}$$

where x_s^{kij} is the entry corresponding to the i^{th} degree of freedom of the j^{th} target mode from the k^{th} observation of the s^{th} target modal matrix and ω_j^k is the natural frequency corresponding to the j^{th} target mode in the k^{th} observation. When the number of sensors is fixed, the sensor deployment with the largest SSC-ADPR will be considered the optimal sensor placement.

9.1.3 Discussion and handling of singular matrices

From the equations (4.12), (4.14), (9.4) and (9.5), it can be seen that these four optimisation objectives are not applicable for the SPO with singular matrices, including the mode shape matrix ($\mathbf{\Phi}$), the feature matrix ($\mathbf{X}$) and the label matrix ($\mathbf{Y}$), which are not invertible. This means that it is necessary to check whether these matrices are singular before optimisation using these four objectives.

Since the mode shapes are independent when the number of degrees of freedom is equal to or greater than the number of identified modes and the selected degrees of freedom are not centrally distributed, the singularity problem for a mode shape matrix ($\mathbf{\Phi}$) can be solved by increasing the candidate sensor locations and changing the distribution of the candidate sensor locations. For the feature matrix ($\mathbf{X}$), the sensor combination corresponding to a singular feature matrix can be discarded directly. Usually, the label matrix ($\mathbf{Y}$) is of full rank. If not, the redundant label information should be removed first.

9.2 Optimisation considering sensor failure and sensor redundancy

In order to make the performance of the sensor system less sensitive to sensor failures, two novel optimisation ideas are proposed by combining the above optimisation objectives with the effects of, or countermeasures to, sensor failures, to obtain the fitness of sensor distribution.

The first approach, called the *fail-safe sensor distribution*, will take into account the effects of critical sensor failures when calculating the fitness of the sensor distribution. The second approach is called the *fail-safe sensor distribution with redundancy*, where the fitness results of the sensor distribution will be affected by the critical failure handling scheme. The motivation for the second optimisation idea is to explore, with a fixed number of sensors, a better strategy to place all sensors in different locations, to collect more information, or use a subset of sensors as backup sensors for key sensors. The evaluation criterion is based on the performance of the sensor system before and after the failure of the key sensor.

To make the formulation of these two optimisation strategies more explicit in the following sections, SPO is framed here as a combinatorial optimisation problem. This thesis does not address the case of describing sensor locations in continuous space. As described in Chapter 4, the number of combinations S for selecting I sensors from Z candidate sensors in different locations is given by,

$$S = {}_Z C_I = \frac{Z!}{I!(Z-I)!} \tag{9.6}$$

9.2.1 Fail-safe sensor optimisation

The process of SPO with sensor failures under consideration can be divided into two stages. At the first stage, the 'fail-safe' fitness of an algorithm is defined and obtained by taking the worst fitness value of the $(I - I_{\mathrm{F}})$-sensor distributions generated from an I-sensor distribution by removing I_{F} sensors. I_{F} refers to the number of sensors that are allowed to fail without requiring any remedial action. Note that the I-sensor distribution refers to the I sensors being put in I different locations. Because $(I - I_{\mathrm{F}})$ sensor locations will be selected from the mother distribution of I sensor locations, the number of combinations S_{FS} for the fail-safe fitness is,

$$S_{\mathrm{FS}} = S \cdot {}_I C_{(I-I_{\mathrm{F}})} = S \cdot {}_I C_{I_{\mathrm{F}}} = \frac{Z!}{I_{\mathrm{F}}!(Z-I)!(I-I_{\mathrm{F}})!} \tag{9.7}$$

Note that ${}_I C_{(I-I_{\mathrm{F}})}$ equals ${}_I C_{I_{\mathrm{F}}}$, because the number of all child distributions containing functioning sensors is equal to the number of all combinations with failed sensors. For the sake of brevity, the more concise expression, ${}_I C_{I_{\mathrm{F}}}$, will be used below. Then, the maximum/minimum fail-safe fitness values are going to specify the optimal fail-safe sensor distribution.

The first stage here provides a more general description of the fail-safe sensor distribution idea introduced in [66]. However, there may be multiple optimal combinations where sub-distributions corresponding to the worst fitness are the same. Nonetheless, the search algorithm will provide only one of these optimal I-sensor combinations of different fitness, but the same fail-safe fitness.

Therefore, after obtaining an optimal fail-safe sensor distribution given by a search algorithm, at the second stage, the following steps should be taken to further consider the performance of the I-sensor distribution:
(1) Replace the sensors for which failures lead to the worst performance, with other candidate sensors from the $Z-I$ unselected sensors. The number of updated I-sensor distributions is,

$$S_{\mathrm{UPD}} = {}_{(Z-I)}C_{I_{\mathrm{F}}} = \frac{(Z-I)!}{I_{\mathrm{F}}!(Z-I-I_{\mathrm{F}})!} \tag{9.8}$$

(2) Calculate the performance of the $(I-I_{\mathrm{F}})$-sensor distributions generated by the updated I-sensor distribution. The number of the child distributions for these updated mother distributions is given by,

$$S_{\mathrm{UPDFS}} = S_{\mathrm{UPD}} \cdot {}_{I}C_{I_{\mathrm{F}}} = \frac{I!(Z-I)!}{(I_{\mathrm{F}}!)^2(Z-I-I_{\mathrm{F}})!(I-I_{\mathrm{F}})!} \tag{9.9}$$

Compare the worst fitness value of these $(I-I_{\mathrm{F}})$-sensor distributions with the corresponding fail-safe fitness result of the optimal sensor layout found at the first stage. Retain the sensor distribution if both fitness values are the same.
(3) Calculate the performance of all such retained I-sensor distributions and the optimal sensor distribution obtained at the first stage. Select the sensor distribution with the largest fitness value as the final optimal fail-safe sensor distribution, which is named the optimal improved-fail-safe sensor distribution.

An example is given in Section 9.4.1 to show all the steps of the fail-safe optimisation strategy. Note that the three steps in the second stage are not the most efficient way to find the optimal fail-safe sensor distribution with the largest fitness value for an exhaustive search (ES). Multiple optimal fail-safe combinations can be directly found by filtering out the combinations with the best fail-safe fitness. However, the three steps described above are applicable when not all candidate solutions are available.

9.2.2 Fail-safe sensor optimisation with redundancy

To further improve the performance of the sensor system after any sensor fails, compared with putting I sensors in I different locations, I_{F} of the I sensors can be used as redundant sensors and be added to I_{F} critical locations of the $I - I_{\mathrm{F}}$ sensor locations. This idea can help to avoid the worst performance of the $[(I - I_{\mathrm{F}}) - I_{\mathrm{F}}]$-sensor distributions generated by an $(I - I_{\mathrm{F}})$-sensor distribution without increasing the number of sensors used. Based on this, the fail-safe SPO with sensor backup under consideration can also be divided into two stages.

At the first stage, $I - I_{\mathrm{F}}$ locations will be selected first from the Z candidate locations. The maximum/minimum fail-safe fitness value is going to be reached to find the optimal fail-safe sensor distribution with $I - I_{\mathrm{F}}$ sensors. Backup sensors are then added to the locations of the I_{F} key sensors, to avoid the occurrence of the $[(I - I_{\mathrm{F}}) - I_{\mathrm{F}}]$-sensor distributions with the worst performance. The 'fail-safe with redundancy' fitness is then defined and obtained by taking the worst fitness value of the $(I - I_{\mathrm{F}})$-sensor distributions with redundancy, generated from an I-sensor distribution with redundancy by omitting I_{F} 'non-critical' sensors (sensors without a backup sensor) in turn. Note that the I-sensor distribution with redundancy means that I sensors are placed in $I - I_{\mathrm{F}}$ different locations, with I_{F} pairs of sensors in the same location. The number of combinations S_{FSR}, for the fail-safe-with-redundancy fitness is,

$$S_{\mathrm{FSR}} = {}_{Z}C_{(I-I_{\mathrm{F}})} \cdot {}_{(I-I_{\mathrm{F}})}C_{I_{\mathrm{F}}} \cdot {}_{[(I-I_{\mathrm{F}})-I_{\mathrm{F}}]}C_{I_{\mathrm{F}}} = \frac{Z!}{(I_{\mathrm{F}}!)^2(Z - I + I_{\mathrm{F}})!(I - 3I_{\mathrm{F}})!} \quad (9.10)$$

The maximum/minimum fail-safe with redundancy fitness value is then going to specify the optimal I-sensor distribution with redundancy. Furthermore, it is necessary to clarify that the intended function of redundant sensors (or backup sensors) is to provide the same information when some of the original sensors cease to provide valid information. Therefore, the existence of the redundant sensors is to avoid the worst-case scenario, after the failure of the key sensors. Data from redundant sensors are not used when computing the fail-safe with redundancy fitness. On this basis, optimal results for the fail-safe SPO with redundancy can be obtained.

Similarly, more than one optimal fail-safe sensor distribution with redundancy may be included for the case with the same number of selected sensors. Considering the performance of these sensor distributions before any sensor fails, the sensor distribu-

tion with the greatest fitness should be chosen as the final optimal fail-safe sensor distribution with redundancy. Therefore, after obtaining the optimal fail-safe sensor distributions with redundancy at the first stage, the results should be evaluated and updated via the following steps at the second stage:
(1) Replace the I_{F} sensors whose failure corresponds to the optimal fail-safe with redundancy fitness with other candidate sensors from the $Z-(I-I_{\mathrm{F}})$ unselected sensors.
(2) Compare the worst fitness value of the $(I-I_{\mathrm{F}})$-sensor distributions with redundancy, generated by omitting I_{F} 'non-critical' sensors in turn, with the corresponding fail-safe with redundancy fitness result of the optimal sensor deployment found at the first stage. And retain the I-sensor distribution with redundancy if the two fitness values are the same. The number of the child distributions with redundant sensors for the updated mother distributions with redundant sensors is,

$$S_{\mathrm{UPDFSR}} = {}_{[Z-(I-I_{\mathrm{F}})]}C_{I_{\mathrm{F}}} \cdot {}_{[(I-I_{\mathrm{F}})-I_{\mathrm{F}}]}C_{I_{\mathrm{F}}} = \frac{(I-2I_{\mathrm{F}})!(Z-I+I_{\mathrm{F}})!}{(I_{\mathrm{F}}!)^2(Z-I)!(I-3I_{\mathrm{F}})!} \tag{9.11}$$

(3) Calculate the performance of all retained I-sensor distributions with redundancy and the optimal I-sensor distribution with redundancy obtained at the first stage. Select the sensor distribution with the largest fitness value as the final optimal fail-safe sensor distribution with redundancy, which is called the optimal improved-fail-safe sensor distribution with redundancy.

Section 9.4.2 gives an example of fail-safe sensor optimisation with redundancy, that includes all the steps above. Again, the three steps introduced here are not the most efficient way to find the best results for an ES, but they are more versatile given the different search algorithms.

9.2.3 Discussion and suggestions on the application of fail-safe strategies

The initial motivation for considering sensor failures in the sensor optimisation process is that when one or two sensors fail, the performance of the sensor system remains within acceptable limits and no maintenance is required. Such sensor systems can be of value in situations where maintenance is difficult or expensive. When many sensors are not providing data properly, this is a very suspicious situation that needs

to be detected by the maintenance team. So it doesn't make much engineering sense to design for the failure of many sensors. Moreover, it must be clarified that the fail-safe strategies here can only consider the worst-case scenario after sensor failure. However, since the worst case corresponds to the failure of several critical sensors, the loss of a smaller number of critical or non-critical sensors should correspond to better system performance than the worst case.

Furthermore, the idea of failure safety is unnecessary when the budget for sensors and the corresponding data storage and processing is large enough to design a densely-distributed sensor system, which means no sensor in this system is critical or irreplaceable. That is, in that sensor system, each sensor has at least one 'backup' sensor to provide similar information. However, the fail-safe strategies in this chapter aim to consider the impact of the failure of certain critical sensors in the optimisation process of a sensor system and whether it is worthwhile to pre-place a backup sensor for critical sensors. Therefore, it should only be applied when designing a sensor system which includes critical sensors.

When all the degrees of freedom of the finite element model are used as candidate sensor positions, engineering experience can be used to reduce the density of candidate sensor distribution preliminarily. For example, sensors are evenly distributed to cover the area to be detected, and points with larger signal amplitudes are selected. In addition, a more objective criterion can be used to avoid arranging the sensors at adjacent locations where substantially the same information will be collected. For example, the linear correlation of features can also be used to determine whether a feature contains some uncorrelated information from the selected ones.

9.3 Illustrative examples

A data set from the second-round tests on a glider wing in an environmental chamber (shown in Fig. 5.2) is used for this research.

9.3.1 Data set for the SPO on modal identification

For SPO with the aim of modal identification, three mode shapes of the wing under the normal condition at 15 °C were chosen to form the s^{th} target mode matrix

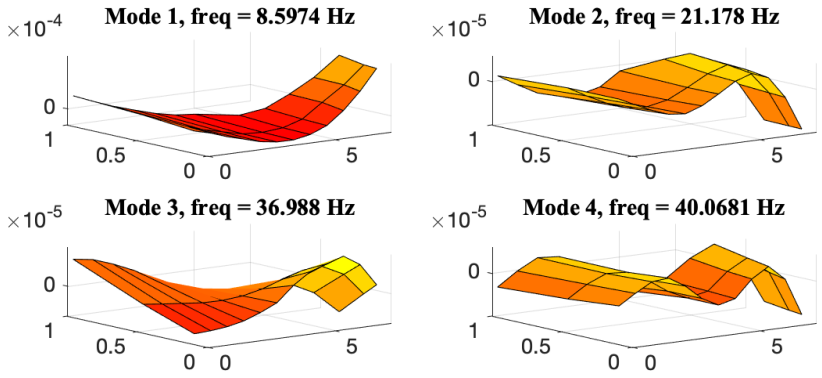


Figure 9.1: The first four mode shapes of the glider wing at 15 °C.

$\mathbf{\Phi}_s \in \mathbb{R}^{I\times 3}$ for the s^{th} combination of the I sensor locations, including the first, second and fourth mode shapes. The first four mode shapes and corresponding natural frequencies, extracted from the high-averaged FRFs collected at 15 °C, are shown in Figure 9.1. They are denoted as Mode 1 to 4.

The reason to adopt these three mode shapes is that by comparing the optimisation results obtained by the two proposed fail-safe methods using the first three mode shapes and the selected three mode shapes separately, optimisation results corresponding to the selected three mode shapes include more possible scenarios. Moreover, according to equation (4.3), at least three sensors must be selected to provide a target modal matrix with full column rank, which means that the mode shapes are linearly independent [42].

9.3.2 Data set for the SPO on damage identification

For SPO with the aim of damage identification, mode shapes of the glider wing under four different health states with corresponding labels were used to form the feature matrix and the label matrix. For each mode shape, only one measurement is taken for the structure in each health state at each controlled temperature, and only three controlled temperatures are considered: 5, 10, 15 °C. Thus, the available number of the observations K is the number of health states times the number of controlled temperatures, which is equal to 4×3. This means that the number of columns $(I \times J)$ in the feature matrix should not exceed $K = 12$ to form a full-column rank feature matrix. Note that mode shapes for three temperatures are adopted here because mode shapes extracted from FRFs (if there are more than one), measured

at one temperature may constitute a rank-deficient feature matrix $\mathbf{X}_{\mathrm{M}}$. $\mathbf{X}_{\mathrm{M}}^{\top}\mathbf{X}_{\mathrm{M}}$ will then be a singular matrix, which cannot be used for the SSC calculation.

Therefore, to allow more sensors to be selected to show the possible situation of optimisation results as comprehensively as possible, each of the three mode shapes used in the SPO for modal identification was adapted to form the feature matrix $\mathbf{X}_s \in \mathbb{R}^{12\times I}$, corresponding to the s^{th} combination of I sensors separately. Meanwhile, because the labels here are nominal data without an inherent order, four categorical variables are encoded into a matrix by dummy coding, which uses N features to represent $N+1$ labels/categories. Therefore, the size of the label matrix is 12×3.

For each mode shape, the SSC between the feature matrix and the label matrix of the structural state is calculated for sensor distributions with different numbers of sensors. An ES is used to find the optimal sensor combinations with one to six sensors for three mode shapes separately. The combinations having the largest SSC will be selected and the results are in Appendix B.1. The comparison of optimal fitness values for three selected mode shapes is presented in Figure 9.2. Because the SSC corresponding to the second mode shape (Mode 2) is always greater than that of others, this mode shape will be used to provide features to illustrate the proposed SPO ideas for damage identification.

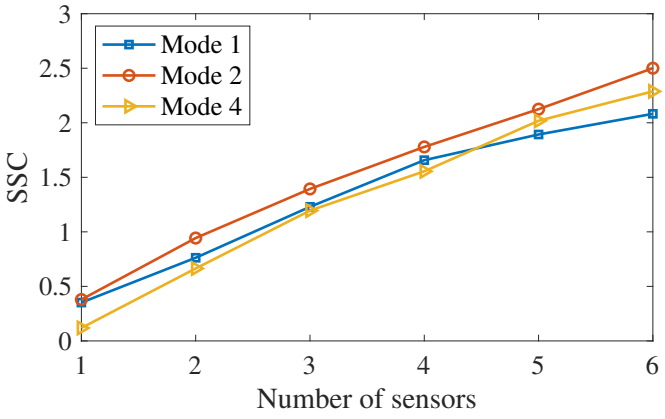


Figure 9.2: Comparison of ES results based on the SSC for three selected mode shapes.

The ADPR values calculated from the above two data sets were scaled so that their range is within the interval [0,1], reducing the software's precision requirements for numeric calculations of objective functions involving signal strength.

9.3.3 Discussion on optimisation algorithms

Two optimisation algorithms are chosen here to show the whole optimisation process specifically: a deterministic algorithm – an ES – and a stochastic algorithm – a genetic algorithm (GA). Because of limitations in computer performance, a GA is mainly used to realise the search tasks. Results from the ES will be used for validation. It is necessary to say that the number of mother combinations presented in Section 9.2 is based on an ES. Finding the optimal result with a GA usually requires evaluating far fewer combinations of the mother distribution. In the search process of the two algorithms, the combination process of obtaining child distributions of a mother distribution is the same.

The parameters used for the GA are as follows: population size is 50, probability of crossover is 0.8, and probability of mutation is 0.01. The number of elite surviving to the next generation equals 0.05 $\times$ population size (round toward positive infinity to the nearest integer). Linear fitness scaling is used. The algorithm stops when the average relative change in the fitness function is less than or equal to a tolerance. Linear constraints defined in equation (8.2.3) are also used here. Considering the randomness of the initial population for the GA, it was run 10 times for each case, and the best among the 10 results was chosen as the final optimisation result. For more information on GAs and how to choose or optimise the parameters for a GA, please refer to [123].

An ES technique and a GA are first used to search for the optimal sensor distribution without considering sensor failure. The comparison of these two results can help to check the effectiveness of the GA. Furthermore, these two sets of results can be compared with the following results that consider sensor failure aspects, which are designed to be searched out via a GA. Part of the results corresponding to a small number of sensors has also been searched exhaustively to verify the effectiveness of the GA application.

Note that considering the scalability of the algorithm, an evolutionary algorithm is suggested, because it will still be applicable when the number of candidate sensor combinations is enormous. However, the reliability of optimal results will be an issue, since evolutionary algorithms, such as the GA, may provide poor suboptimal results. For this concern, a compromise solution is given here. A greedy algorithm can be used to search for suboptimal results, which can often be obtained in much less time

than an ES. The GA results are then compared to the greedy search results to ensure that the GA results are at least as good as the greedy-search results.

9.4 Results and discussion

To more intuitively demonstrate the two ideas introduced in Section 9.2, take $I_{\mathrm{F}} = 1$ as an example, which means that the number of sensors assumed to have failed in the design process is one. The number of backup sensors then corresponding to the key sensors is also one. It is then easy to find that the sub-distribution with the worst fail-safe with redundancy fitness should typically be obtained by performing the combination ${}_{[(I-I_{\mathrm{F}})-I_{\mathrm{F}}]}C_{I_{\mathrm{F}}}$ in equation (9.10). However, for the case of $I_{\mathrm{F}} = 1$, these sub-distributions of sensors are the child distributions corresponding to the second-worst fail-safe fitness in the last step, which helps simplify the code.

At the risk of duplication, it should be stated that the fitness based on any of the four optimisation objectives mentioned in Section 9.1 needs to be maximised to obtain the optimal sensor distribution.

9.4.1 Optimisation results of improved fail-safe SPO

Firstly, the SPO based on DFIM is used to demonstrate the two-stage process of obtaining the optimal improved fail-safe sensor distribution. At the first stage, optimal fail-safe sensor distributions with four and five sensors are found by an ES and listed in Table 9.1. The optimisation results for this type, with four to eight sensors corresponding to a GA, are given in Table 9.2. The reasons for choosing the number of sensors to be four to eight are as follows. The minimum number of selected sensors depends on the number required to obtain valid fitness values for all four optimisation objectives. When it comes to selecting the maximum number of sensors, it needs to be said that, considering the reliability of the conclusions obtained, the findings of this chapter will be based on the analysis of the results of the deterministic search algorithm. Therefore, the computational cost limits the maximum number of selected sensors by an ES to six. The optimal sensor distributions with seven and eight sensors obtained by the GA were used to verify the applicability of the findings obtained from the optimal sensor distributions of four to six sensors.

The design of up to 8 sensors was selected from the 36 sensors, indicating that this illustrative example considers the case where the sensors are not densely arranged on the structure, which enables each sensor to collect mutually non-redundant information to a certain extent. In practice, the sensors are generally not arranged in large numbers on a simple structure or a part of a complex structure considering data transmission, storage, and processing. However, this approach can be used to optimise systems with densely-distributed sensors to account for the failure of some key sensors that collect critical information.

Table 9.1: Optimal fail-safe sensor distributions obtained by an ES combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness
4	4,12[a],33,35	6.169e-08
5	4,12,31,34,36	3.182e-07
6	12,16,24,28,34,35	1.147e-06

[a] The failure of a sensor at the position marked in blue will result in the worst fail-safe fitness.

Table 9.2: Optimal fail-safe sensor distributions obtained by an GA combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness
4	4,12,33,35	6.169e-08
5	5,12,31,34,36	3.182e-07
6	12,16,24,28,34,35	1.147e-06
7	12,16,22,24,28,31,34	1.980e-06
8	12,16,22,24,28,31,34,36	3.334e-06

From Tables 9.1 and 9.2, it can be seen that the sensor distributions with five sensors obtained by the GA and the ES are a little different. However, by checking the fail-safe fitness values of these two distributions, the result turns to be the same. This result means that when the number of selected sensors is the same, multiple optimal fail-safe sensor distributions with the same fail-safe fitness value can be obtained. In this case, the fitness values of the mother distributions of these optimal fail-safe sensor distributions can be used to evaluate their performance and rank them.

At the second stage, considering the performance of these optimal fail-safe sensor distributions before any sensor fails, the sensor distribution with the largest DFIM should be selected as the final optimal fail-safe sensor distribution. Therefore, after obtaining the optimal fail-safe sensor distributions given by an ES or a GA at the first stage, these results should be evaluated and updated following the steps in Section 9.2.1. The situation of all optimal fail-safe distributions with four to eight sensors

generated by the GA results is given in Table 9.3, and the optimal improved-fail-safe distributions with four to eight sensors are listed in Table 9.4. It can be seen that the improved fail-safe approach can bring some benefits without increasing the cost of the sensing equipment. By a similar two-stage optimisation process, the optimal improved-fail-safe sensor distributions corresponding to an ES are shown in Table 9.5.

Table 9.3: Multiple optimal fail-safe sensor distributions corresponding to the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
4	4,12,33,35	6.169e-08	4.923e-07
	4,24,33,35	6.169e-08	4.851e-07
	4,33,35,36	6.169e-08	4.840e-07
5	4,12,31,34,36	3.182e-07	1.408e-06
	5,12,31,34,36	3.182e-07	1.284e-06
	12,13,31,34,36	3.182e-07	1.073e-06
	12,16,31,34,36	3.182e-07	1.479e-06
	12,17,31,34,36	3.182e-07	1.234e-06
	12,25,31,34,36	3.182e-07	1.107e-06
	12,28,31,34,36	3.182e-07	1.621e-06
	12,29,31,34,36	3.182e-07	1.151e-06
6	12,16,24,28,34,35	1.147e-06	2.783e-06
7	12,16,22,28,31,34,24	1.980e-06	3.940e-06
	12,16,22,28,31,34,36	1.980e-06	4.140e-06
8	12,16,22,28,31,34,36	3.334e-06	6.128e-06

Table 9.4: Optimal improved-fail-safe sensor distributions obtained by a GA combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
4	4,12[b],33,35	6.169e-08	4.923e-07
5	12,28,31,34,36	3.182e-07	1.621e-06
6	12,16,24,28,34[a],35	1.147e-06	2.783e-06
7	12,16,22,28,31,34,36	1.980e-06	4.140e-06
8	12,16,22,24,28,31,34,36	3.334e-06	6.128e-06

[a] The failure of a sensor at the position marked in blue will result in the worst fail-safe fitness.
[b] From a fail-safe perspective, the replaceable sensor location is marked in magenta.

Table 9.5: Optimal improved-fail-safe sensor distributions obtained by an ES combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
4	4,12,33,35	6.169e-08	4.923e-07
5	12,28,31,34,36	3.182e-07	1.621e-06
6	12,16,24,28,34,35	1.147e-06	2.783e-06

The same process, dependent on an ES and a GA, is repeated to search for the optimal improved fail-safe sensor distributions corresponding to the other three optimisation objectives – DFIM-ADPR, SSC and SSC-ADPR, and the results are listed in Tables B.4 to B.9 in Appendix B.2.

9.4.2 Optimisation results of improved fail-safe with redundancy SPO

The two-stage process of the improved-fail-safe SPO with redundancy is illustrated using an objective function of a GA based on DFIM. At the first stage, the optimisation results of the fail-safe sensor distributions with redundancy obtained by a GA are shown in Table 9.6. As before, more than one optimal result is included for the case with the same number of selected sensor locations, which is shown in Table 9.7. Considering the performance of these sensor distributions before any sensor fails, the sensor distribution with the largest DFIM should be selected as the optimal improved-fail-safe sensor distribution with redundancy. Therefore, after obtaining the optimal fail-safe sensor distributions with redundancy given by a GA, these results should be evaluated and updated following the steps in Section 9.2.2 and the improved optimisation results are shown in Table 9.8. These improved optimal distributions with five and six sensors found by an ES are given in Table 9.9.

For other optimisation objectives, including the DFIM-ADPR, the SSC and the SSC-ADPR, the results of the optimal improved-fail-safe sensor distributions with redundancy obtained by a GA and an ES are shown in Tables B.10 to B.15 in Appendix B.3.

Table 9.6: Optimal fail-safe sensor distributions with redundancy obtained by a GA combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness
5	13[d],31,34,36 (36)[c]	1.631e-07
6	12,16,28,34 (34),36	7.242e-07
7	4,12,16,34 (34),35,36	1.306e-06
8	4,12,16,24,28,34 (34),36	2.147e-06

[c] Sensors in parentheses are redundant sensors placed to avoid the worst fail-safe fitness.

[d] The failure of a sensor at the position marked in green will result in the worst fail-safe with redundancy fitness.

Similar to Section 9.4.1, a small number of sensors were selected in this part. However,

Table 9.7: Multiple optimal fail-safe sensor distributions with redundancy corresponding to the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
5	4,31,34,36 (36)	1.631e-07	7.274e-07
	5,31,34,36 (36)	1.631e-07	6.619e-07
	13,31,34,36 (36)	1.631e-07	5.535e-07
	16,31,34,36 (36)	1.631e-07	7.623e-07
	17,31,34,36 (36)	1.631e-07	6.351e-07
	28,31,34,36 (36)	1.631e-07	8.333e-07
	29,31,34,36 (36)	1.631e-07	5.908e-07
6	12,16,28,34 (34),36	7.242e-07	1.587e-06
7	4,12,16,34 (34),35,36	1.306e-06	2.614e-06
	4,12,28,34 (34),35,36	1.306e-06	2.822e-06
8	4,12,28,34 (34),35,36	2.147e-06	3.432e-06

Table 9.8: Optimal improved-fail-safe sensor distributions with redundancy obtained by a GA combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
5	28[b],31,34,36 (36)[c]	1.631e-07	8.333e-07
6	12,16,28[d],34 (34),36	7.242e-07	1.587e-06
7	4,12,28,34 (34),35,36	1.306e-06	2.822e-06
8	4,12,16,24,28,34 (34),36	2.147e-06	3.432e-06

[b] From a fail-safe perspective, the replaceable sensor location is marked in magenta.
[c] Sensors in parentheses are redundant sensors placed to avoid the worst fail-safe fitness.
[d] The failure of a sensor at the position marked in green will result in the worst fail-safe with redundancy fitness.

Table 9.9: Optimal improved-fail-safe sensor distributions with redundancy obtained by an ES combined with the DFIM.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
5	28,31,34,36 (36)	1.631e-07	8.333e-07
6	12,16,28,34 (34),36	7.242e-07	1.587e-06

this approach can also be used in situations where a large number of sensors need to be placed; this may cause some sensors to be brought close together to collect almost the same information. So the question becomes one of, how to find the optimal redundant sensor positions when the available sensor positions are limited and the number of sensors is greater than the sensor positions. In this case, the fail-safe SPO with redundancy in this chapter can be used as an alternative method to optimise redundant sensor placements.

9.4.3 Performance evaluation

In this part, three aspects will be discussed: the effectiveness of the GA algorithm, the performance of the sensor distributions obtained by the fail-safe SPO and those obtained by the fail-safe with redundancy SPO.

To better accomplish these tasks, except for the optimal sensor distributions considering sensor failure, the optimal sensor distributions only taking account of the fitness of mother distributions are found via ES and GA optimisation strategies. This kind of SPO strategy is called 'classical' SPO here. Four optimisation objectives – DFIM, DFIM-ADPR, SSC and SSC-ADPR are all involved. The corresponding results are shown in Tables B.16 and B.23 in Appendix B.4. Moreover, in the case of a certain number of sensors, to compare the results of the three SPO strategies more intuitively, the results of the classical SPO, the fail-safe SPO and the fail-safe with redundancy SPO corresponding to four optimisation objectives are plotted in blue, orange and yellow in Figures 9.3 to 9.6 separately.

Before comparing the optimal results obtained by the two applied search algorithms, it should be noted that: (1) when comparing the optimal results corresponding to an ES and a GA obtained by classical SPO (marked with a blue dash-dotted line and a blue asterisk, respectively), the fitness value before the sensor failure should be used, as shown in the left sub-figures and marked in blue. This choice is because, for classical SPO, the sensor deployment is optimised according to the fitness of the mother distribution. (2) For the optimal results obtained by the other two optimisation strategies, it is necessary to compare the optimal fitness values corresponding to two search algorithms after the failure of the marked sensor, as shown in the right sub-figures and marked with orange or yellow dash-dotted lines and asterisks, because, for these two SPO, the sensor deployment is optimised on the worst child fitness.

It can be seen that the optimal sensor deployments with three to five sensors obtained by the ES and the GA are the same. Furthermore, the fitness, fail-safe fitness or fail-safe with redundancy fitness of the optimal sensor combinations with six sensors obtained by both search algorithms are close to each other. Thus the GA adopted here is effective for searching out optimisation results.

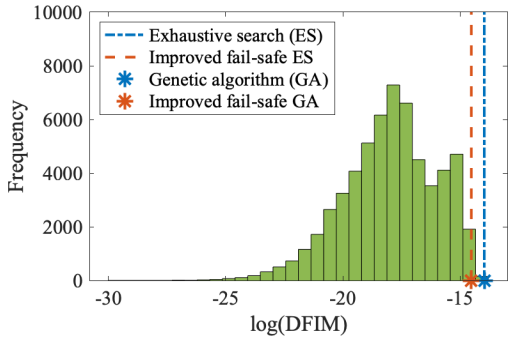
To check the performance of the sensor distributions obtained by the fail-safe SPO and the fail-safe with redundancy SPO, different histograms of the fitness values

corresponding to all sensor distributions with the same number of sensors from an ES are also plotted in Figures 9.3 to 9.6 in green.

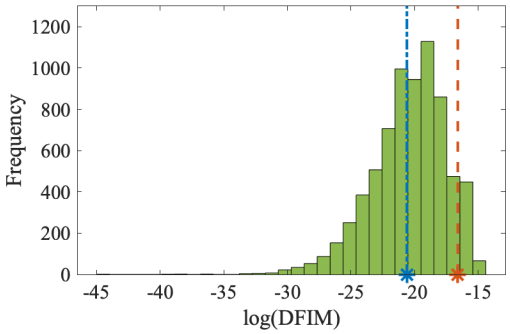
For the fail-safe SPO strategy, the expected result is that the performance of the sensor system before any sensor failure is very similar between the classical and fail-safe methods, while the worst-case performance of the sensor system obtained by the fail-safe method is better than that of the sensor system corresponding to the classical method after a critical sensor failure. This phenomenon can be observed particularly in four pairs of figures: 9.3a & 9.3b, 9.4c & 9.4d, 9.4e & 9.4f, and 9.6e & 9.6f.

Overall, it can be found that the obtained improved-fail-safe sensor distribution provides good performance when all sensors are working normally. Besides, it also ensures that acceptable performance can be maintained when any sensor fails. For the case when the optimal improved-fail-safe sensor distribution (in orange), is the same as the optimal sensor distribution with the largest fitness (in blue), the obtained fail-safe sensor distribution does not compromise the system performance before any sensor fails, as demonstrated in Figures 9.3c & 9.3d and 9.6a & 9.6b.

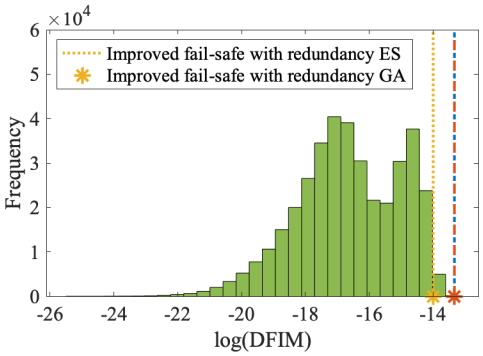
When it comes to the fail-safe with redundancy SPO, results indicate that when the number of sensors is greater than five, the performance corresponding to its optimal result changes the least before and after the sensor failure; this means that the redundant sensor is useful to increase the sensor system's ability to withstand sensor failures. However, for this kind of SPO, a pair of sensors is placed at the same position, which will collect the same information. Therefore, although the sensor failure has the least effect on the performance of the sensor distribution, the optimal sensor distribution found via this optimisation strategy almost always has the worst performance before and after a sensor fails compared with the other optimisation approaches.



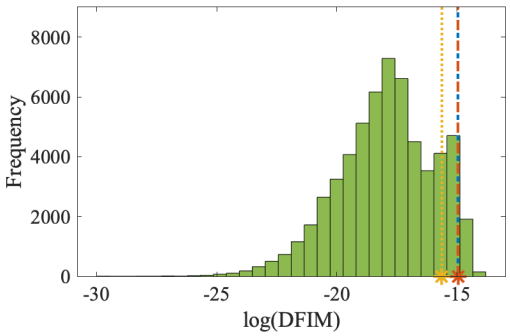
(a) Before a sensor fails – four-sensor distributions.



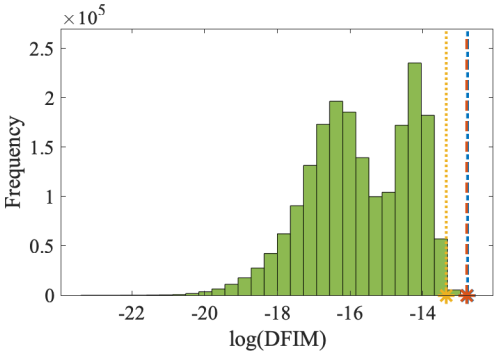
(b) After a sensor fails – three-sensor distributions.



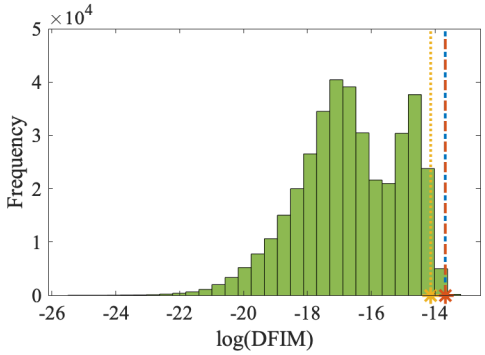
(c) Before a sensor fails – five-sensor distributions.



(d) After a sensor fails – four-sensor distributions.

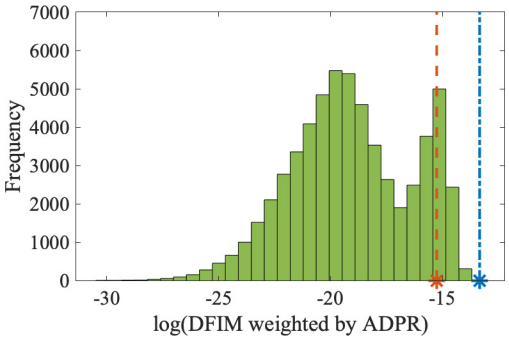


(e) Before a sensor fails – six-sensor distributions.

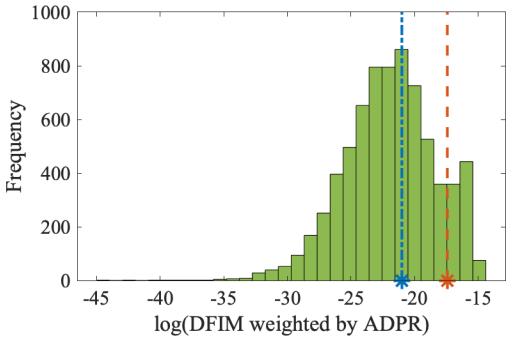


(f) After a sensor fails – five-sensor distributions.

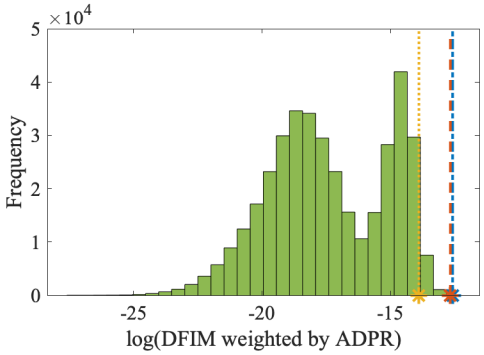
Figure 9.3: Comparison of the DFIM results of a GA, an improved-fail-safe GA and an improved-fail-safe with redundancy GA, with the ES results.



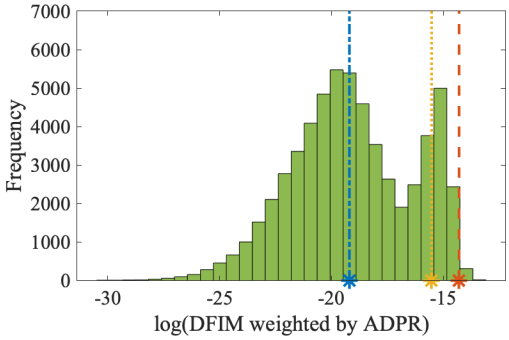
(a) Before a sensor fails – four-sensor distributions.



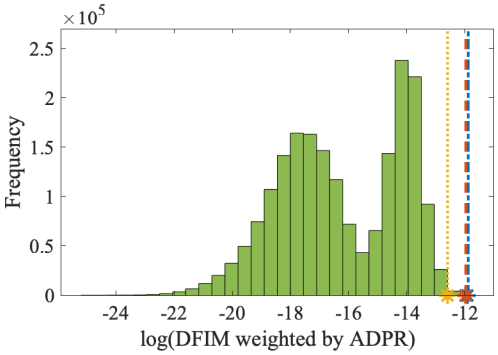
(b) After a sensor fails – three-sensor distributions.



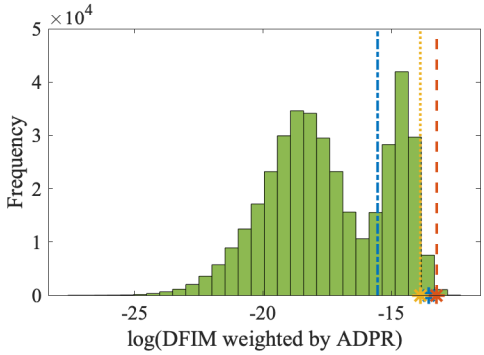
(c) Before a sensor fails – five-sensor distributions.



(d) After a sensor fails – four-sensor distributions.

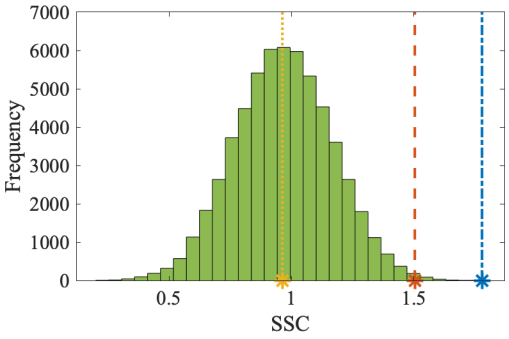


(e) Before a sensor fails – six-sensor distributions.

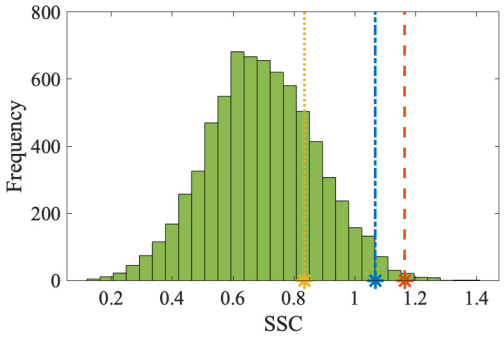


(f) After a sensor fails – five-sensor distributions.

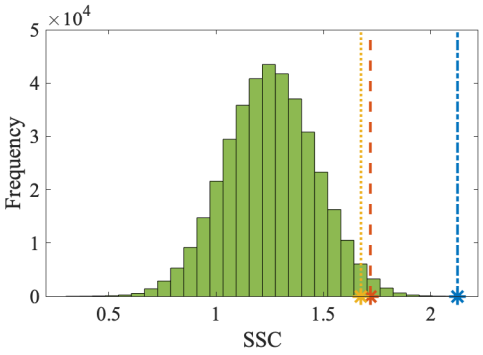
Figure 9.4: Comparison of the DFIM-ADPR results of a GA, an improved-fail-safe GA and an improved-fail-safe with redundancy GA, with the ES results. See Figures 9.3a and 9.3a for legends.



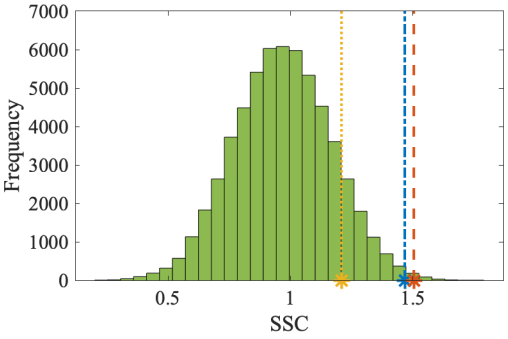
(a) Before a sensor fails – four-sensor distributions.



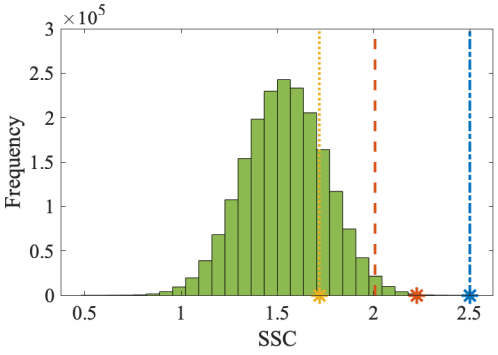
(b) After a sensor fails – three-sensor distributions.



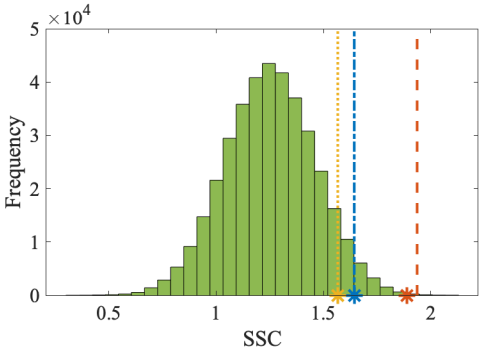
(c) Before a sensor fails – five-sensor distributions.



(d) After a sensor fails – four-sensor distributions.

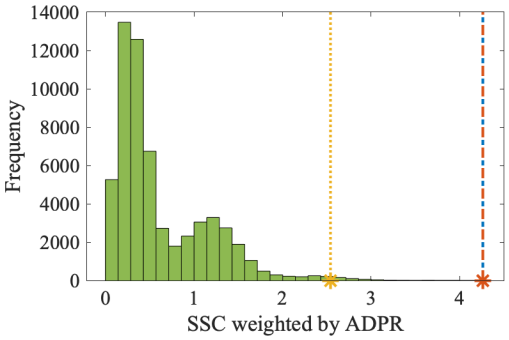


(e) Before a sensor fails – six-sensor distributions.

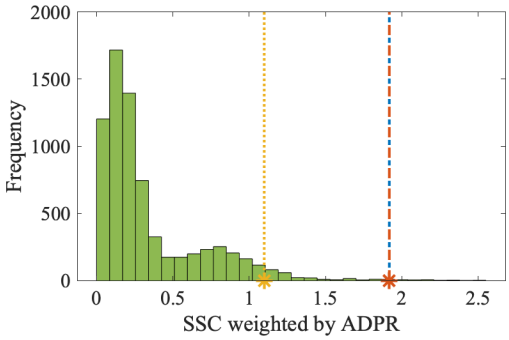


(f) After a sensor fails – five-sensor distributions.

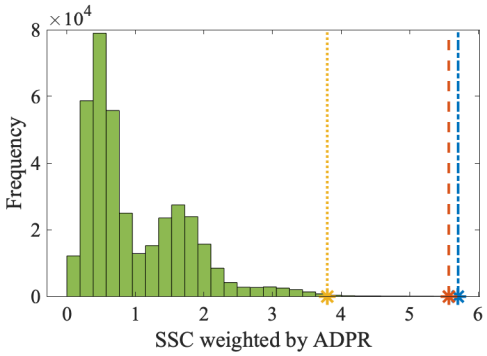
Figure 9.5: Comparison of the SSC results of a GA, an improved-fail-safe GA and an improved-fail-safe with redundancy GA, with the ES results. See Figures 9.3a and 9.3a for legends.



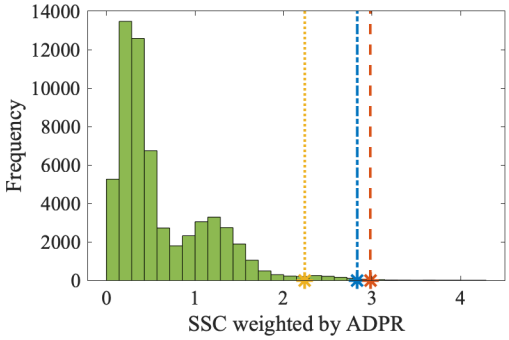
(a) Before a sensor fails – four-sensor distributions.



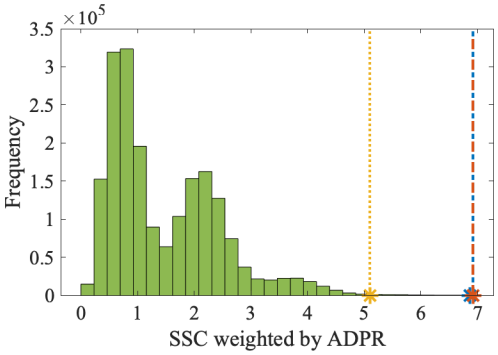
(b) After a sensor fails – three-sensor distributions.



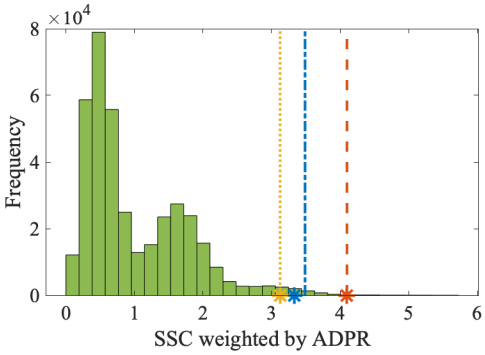
(c) Before a sensor fails – five-sensor distributions.



(d) After a sensor fails – four-sensor distributions.

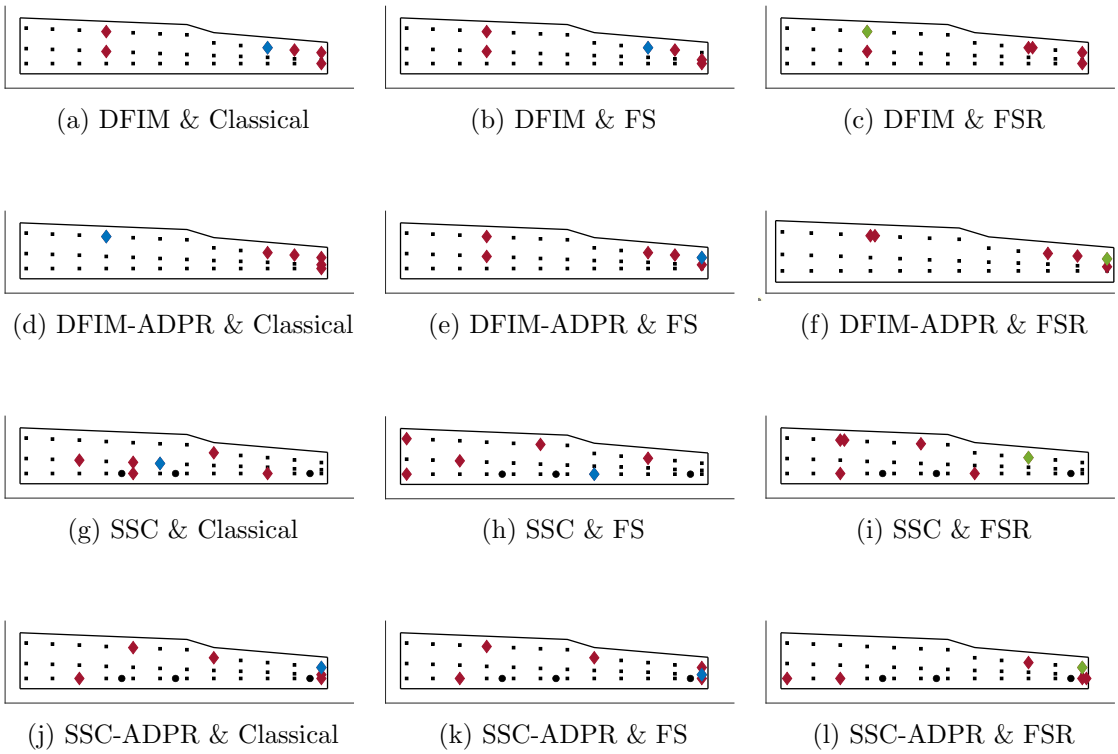


(e) Before a sensor fails – six-sensor distributions.



(f) After a sensor fails – five-sensor distributions.

Figure 9.6: Comparison of the SSC-ADPR results of a GA, an improved-fail-safe GA and an improved-fail-safe with redundancy GA, with the ES results. See Figures 9.3a and 9.3a for legends.



(a) DFIM & Classical (b) DFIM & FS (c) DFIM & FSR

(d) DFIM-ADPR & Classical (e) DFIM-ADPR & FS (f) DFIM-ADPR & FSR

(g) SSC & Classical (h) SSC & FS (i) SSC & FSR

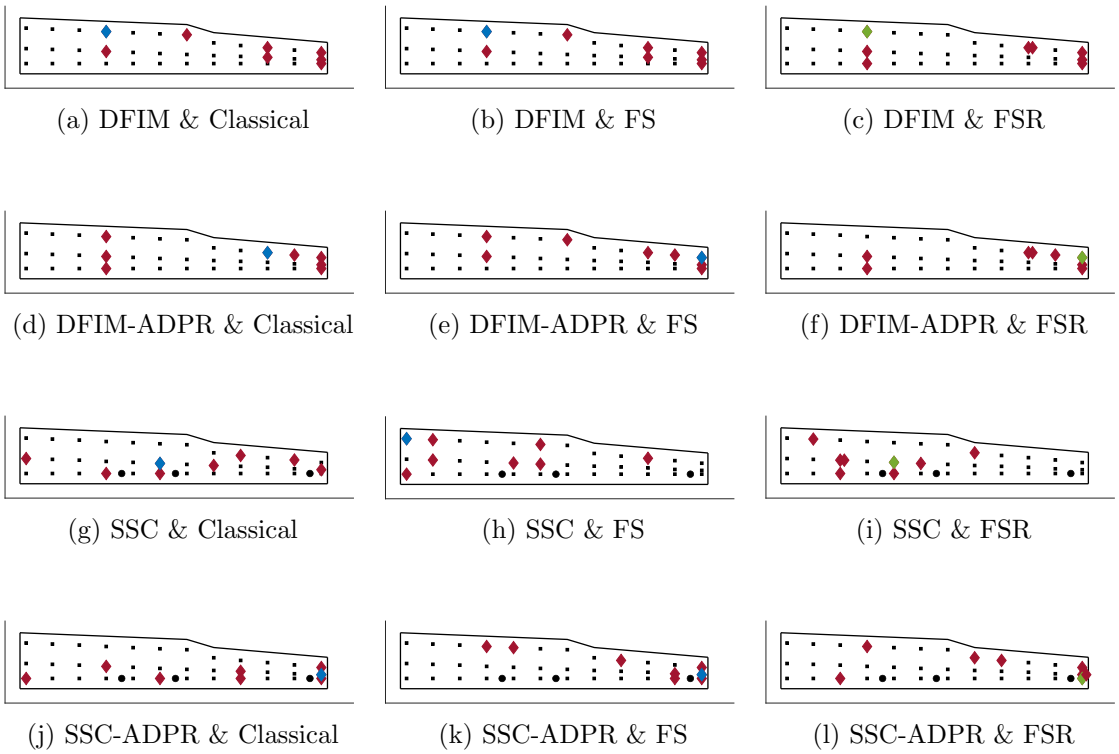
(j) SSC-ADPR & Classical (k) SSC-ADPR & FS (l) SSC-ADPR & FSR

Figure 9.7: Comparison of optimal six-sensor distributions obtained by three ES-based optimisation strategies combined with four optimisation objectives. The selected sensor locations are marked in red. The sensors whose failure will lead to the worst and the second-worst child fitness are marked in blue and green. FS refers to fail-safe. FSR refers to fail-safe with redundancy. The three black dots represent the three considered damage scenarios in the second-round tests.

Note that the two improved fail-safe strategies cannot directly trade off the fitness of the mother distribution and the fail-safe fitness of the child distribution. Although the GA results in Figures 9.5e & 9.5f show the trade-off possibility, a possible reason for the phenomenon shown in Figures 9.5e & 9.5f is that the GA fails to find the optimal fail-safe sensor distribution, but instead finds a sub-optimal fail-safe sensor distribution, which sacrifices less performance of the sensor system without a failed sensor to consider the effect of sensor failure.

9.4.4 Optimal sensor distribution comparisons

In this part, the optimal sensor distributions with six and eight sensors found by an ES and a GA respectively are plotted in Figures 9.7 and 9.8, to analyse the distribution characteristics of sensors obtained using different optimisation objectives.



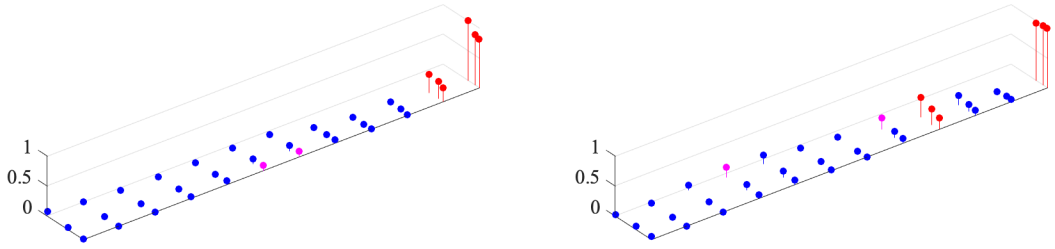
(a) DFIM & Classical (b) DFIM & FS (c) DFIM & FSR

(d) DFIM-ADPR & Classical (e) DFIM-ADPR & FS (f) DFIM-ADPR & FSR

(g) SSC & Classical (h) SSC & FS (i) SSC & FSR

(j) SSC-ADPR & Classical (k) SSC-ADPR & FS (l) SSC-ADPR & FSR

Figure 9.8: Comparison of optimal eight-sensor distributions obtained by three GA-based optimisation strategies combined with four optimisation objectives. The selected sensor locations are marked in red. The sensors whose failure will lead to the worst and the second-worst child fitness are marked in blue and green.



(a) ADPR used in mode identification task. (b) ADPR used in damage identification task.

Figure 9.9: Rescaled ADPR corresponding to 36 candidate sensor locations. The first six maximum values are marked in red, and the next two are marked in magenta.

For each optimisation objective, the sensor distributions obtained by the two search algorithms are similar. This phenomenon also reflects the reliability of GA results. So far, it can be said that a GA can be used to search for a reliable optimisation result when optimising a sensor system that is too complex for ES.

The specific features of optimal distributions included are as follows: (1) by comparing the optimal sensor distributions considering ADPR with the distribution of locations with larger ADPR values shown in Figure 9.9, it can be seen that the addition of signal strength in the fitness will have a certain impact on the result, but it is limited. (2) Considering the signal strength in the objective function has a greater impact on the optimisation results based on SSC than those based on DFIM. By comparing Figures 9.3 and 9.5, it can be seen that a possible reason for this phenomenon is that the difference in DFIM corresponding to different sensor combinations is larger than the difference in SSC. Therefore, the addition of ADPR into the optimisation objective has a more limited impact on the ranking of DFIM-based fitness. (3) As expected, the optimal sensor distribution corresponding to SSC fitness is closer to the damage points of interest. (4) Fail-safe SPO makes the selected sensors more evenly dispersed on the structure to be monitored. This kind of deployment explains to a certain extent why the performance of the fail-safe sensor distribution is good, both before and after the sensor failure.

9.5 Conclusions

This chapter has designed an optimisation technique considering the effect of sensor failure for a wired sensor system. Two SPO strategies, including the improved-fail-safe SPO and the improved-fail-safe SPO with redundancy, were proposed for modal identification and damage identification issues. From three perspectives (the accuracy of the target modal-coordinate estimate, the vibration energy and the information on the damage of interest), four objective functions, including the DFIM, the DFIM-ADPR, the SSC and the SSC-ADPR, were adopted and combined with two search algorithms, including a deterministic algorithm – an ES and a stochastic algorithm – a GA to obtain the optimal sensor layouts.

A case study of a glider wing was demonstrated. By comparing with the classical SPO, it has been found that the improved-fail-safe sensor distribution can provide good performance when all sensors are working normally and guarantee good performance when any sensor fails. Furthermore, the redundant sensor can help enhance the sensor system's ability to withstand sensor failures, but the overall performance is not attractive; this shows that it is not advisable to place redundant sensors in a wired sensor system to avoid the worst system performance because of damage to a

key sensor. In the case of a certain number of sensors, the improved fail-safe SPO can make the most effective use of sensor resources by placing the sensors in different locations.

Chapter 10

Conclusions and future work

The final chapter will summarise the novel contributions contained in the thesis and their benefits and limitations. In addition, possible applications will be discussed. Plans for future work are presented at the end of this chapter.

10.1 Conclusions and benefits and limitations

The focus of this thesis is how to design a sensor system robust to environmental variations and sensor failures for SHM systems based on statistical approaches. The scope of this study belongs to the classical optimisation design, and there is an assumption that some reliable data are available during the design phase, obtained through model verification, validation and uncertainty quantification.

A comprehensive review was provided at the beginning of the thesis about the SHM and SPO, which are based on statistical models. This review suggested the need for some research on the design of sensor systems, considering the versatility of design strategies and the performance requirements of the SHM system. There are three directions mainly explored in this thesis, which are:
(1) Senor system design strategies to consider both the computational feasibility of the methods and final system performance.
(2) Sensor system design approaches to dealing with the confounding effects of changing environmental conditions.
(3) Sensor system design methods involving the consequences of sensor failures and

the countermeasures to sensor failures.

The background theories of machine learning algorithms for statistical pattern recognition were provided in Chapter 3. They can help derive a statistical model from training data to infer the health of a structure as new data emerges. Two supervised algorithms, LAD and SVM, were then introduced. The equivalence between LDA and CCA was also demonstrated, which set the stage for the study of the SSC criteria in later chapters. In addition, two existing linear approaches, PCA and MSD, were introduced and compared in this chapter as the second main content. Both of them can be used to create features sensitive to damage but insensitive to changing environmental or operational conditions, when two conditions are satisfied: (1) The training data set should be collected under an appropriate range of environmental or operational conditions that may be encountered. (2) A linear subspace in the original data matrix exists to account for changing environmental or operational conditions. Although there are some restrictions on using them, they are both computationally convenient and efficient.

The topic was then moved to the basics of SPO for SHM in Chapter 4. This chapter clarified that this thesis only studied the SPO that can be treated as a combinatorial problem. Continuous cases, such as sensing area optimisation or sensor size optimisation, were not considered. Some commonly-used optimisation algorithms and optimisation objectives were systematically presented. The search algorithm can be selected according to the available computational capability. The optimisation objective can be chosen according to the characteristics of the data or the expected performance of the monitoring system.

The primary data sets used for case studies in this thesis were from the experiment on a glider wing, which was introduced in great detail in Chapter 5. The experiment was conducted in an environmental chamber and the wing was tested in different health states across a series of controlled temperatures. Three types of data were collected, including time-series data, frequency-domain data and modal data. Two rounds of experiments were carried out for different temperature conditions and damage cases. The first round aims to explore the temperature effect on the structure preliminarily. More comprehensive and challenging data were obtained in the second round of testing, which corresponded to smaller damage sizes and more damage locations.

Chapter 6 began to investigate part of the first direction mentioned above, namely, sensor system design strategies considering the computational feasibility of the methods. The SSC was recommended as a criterion for assessing sensor combinations, which was inherently immune to information redundancy and interaction issues. This criterion can help to directly measure the effectiveness of data for predicting structural states to a certain extent, without actually training a model on data. Two fast calculation approaches of this criterion, h-correlation and the θ-angle, were provided to speed up its calculation. It can be seen that the advantage of combining this criterion with two fast computational strategies is the ability to evaluate sensor combinations in a very efficient manner. However, these two acceleration strategies can only be realised when combined with a greedy search. It means these selection approaches can obtain only a suboptimal sensor combination. Furthermore, the SSC can only evaluate linear relationships between two data matrices. To account for nonlinear dependencies, some further processing steps will be required.

Considering data availability at the sensor system's design stage, in addition to providing a supervised method to evaluate the sensor combination, an unsupervised way to achieve a similar function was also proposed in Chapter 6. Compared to the supervised method, the unsupervised method needs one more step to acquire the 'labels' from the feature set, which may be possible data clusters or projected low-dimensional data. Note that there are four levels of data availability in this thesis:
(1) Level one contains normal condition data only.
(2) Level two contains data for the normal condition and damage states.
(3) Level three contains both data and labels for normal and damage states.
(4) Level four contains data and labels for different health states and the labels for different normal conditions.
For the first and second levels, the unsupervised strategy can be applied. For the other two levels, the supervised strategy can be used. It also needs to be clarified that the methods developed in Chapter 6 belong to the field of feature selection. One synthetic data set, and eight real-world data sets from open data platforms were used as case studies. An attempt was made here to summarise this chapter in the context of SPO to make it easier for readers to fit these methodologies into the goal of this article, sensor system design.

Chapter 7 more completely addressed the first direction mentioned above by establishing a tiered evaluation framework with two stages for sensor system design. At

the first stage, a superior (optimal or sub-optimal) sensor combination for a fixed number of sensors was determined. To make this framework more practical, the case of having a large number of candidate sensors was also considered, which implied a large number of candidate sensor combinations. Therefore, the criteria (objectives) used at this stage should be computationally efficient, or some measures can be taken to improve the calculation speed of the criteria. The criteria used in the filter methods for feature selection are recommended at this stage, which measure the relevance of features from sensors with the expected output variables, without the need for model training. The MCC in this chapter and SSC and MI in Chapter 6 was given as three examples in this thesis to consider three different cases: the case with a response vector or a label vector having a ranked ordering, the case with a response matrix or a label matrix without ordering information, and the case in which the relationship between features and output variables is non-linear. At the second stage, the sensor type and the number of sensors were decided to finalise the configuration of the superior sensor combination, by using criteria that can visually indicate the final system performance. These criteria can be selected according to the characteristics of the data and specific project requirements. Note that it is strongly suggested to ensure that there is a connection between the criteria used at the two stages. For example, if the objective for stage two is the purpose of designing this system, then the objective for stage one can be abstracted from the purpose of the system. A case study of a numerical model with uncertainty analysis was used to concretely demonstrate the whole idea mentioned above.

Chapter 8 focussed on the second direction – SPO robust to environmental variations. Before developing any new techniques working in this direction, first, the robustness of the SPO technique to changing environments was examined for commonly-used features extracted from three types of data, including the time-domain data, frequency-domain data and modal parameters. These features are AR model coefficients, MSD-based features and mode shapes. The case study used glider wing data sets collected at different ambient temperatures. The results showed that the SPO results for mode shapes combined with EI and EI-DPR were robust to the temperature changes. However, the SPO results for MSD-based and AR-model-coefficient features combined with SSC were sensitive to temperature variations. One possible reason behind these phenomena is that the robustness of the results is highly dependent on whether the relationship between features is affected by temperature.

Therefore, the second half of Chapter 8 was about SPO considering temperature

effects. Two SHM SPO strategies that offered robustness to environmental effects by using ESD-based features or MSD-based features were introduced and evaluated for the cases with data availability at levels three and four. The difference between these two strategies is whether the labels for temperature variation are observable or not. The similarity is that both involve the temperature effect in optimisation by taking temperature into account when deriving features for SPO. Note that both strategies can still be used if the damage state data are unavailable, but the collected dependent variable should be replaced with the extracted dependent variable, which is similar to the idea of the unsupervised feature selection method in Chapter 6. Then, although the results in the first half of this chapter showed that the SPO with features derived from time series or FRFs needed to consider the environmental factors, the case study of this part was dedicated only to features derived from the collected FRFs. This situation was because the feature vector of AR model coefficients was not large enough to provide a linear subspace to account for the temperature effect. Other approaches to obtaining effective features from time series while considering the temperature factor should be explored. The results of the case study indicated that the robust SPO strategy using ESD-based features provided a sensor system with higher sensitivity to damage detection. However, the strategy using MSD-based features had a lower requirement for the data availability level.

The last research direction in this thesis, the SPO technique considering the sensor-failure effect, was studied in Chapter 9. The improved-fail-safe SPO and improved-fail-safe SPO with redundancy were proposed for two types of optimisation objectives – modal identification and damage identification. The difference between these two SPO strategies is that the improved-fail-safe SPO attempts to place the available sensors in different locations while considering the system performance before and after the failure of some critical sensors; however, the improved fail-safe SPO with redundancy also retains some sensors as backup sensors for critical sensors, given the same number of sensors. They are similar in that both strategies can only be applied to design sensor systems that include critical sensors. The case study used the mode shape data sets from the second round of tests. Four optimisation objectives were selected, including the DFIM and the DFIM-ADPR, which considered the modal identification issue, and SSC and SSC-ADPR, which considered the damage identification issue. The results showed that, for both identification problems, the sensor distribution obtained by the improved-fail-safe method provided a good performance when the whole system worked normally. When any sensor failed, the performance was still acceptable. Compared with the results of the classical

method and the improved fail-safe method with redundancy, the distribution of sensors selected by the improved-fail-safe method was most dispersed. The sensor distribution obtained by the fail-safe SPO with redundancy increased the ability of the monitoring system to withstand sensor failures. However, the overall performance was not attractive as some sensors were used as backup sensors to collect the same information. This strategy can be applied to find optimal redundant sensor locations when the number of sensors exceeds the available sensor locations.

Last but not least, it can be seen that the strategies developed in Chapters 8 and 9 can be attributed to the first stage of the design framework introduced in Chapter 7, since the optimisation objectives used in these two chapters are not directly related to the final performance of the SHM system based on machine learning algorithms. As stated at the beginning of this section, the availability of reliable data needs to be assumed for sensor system design using the method developed in this thesis. If this assumption cannot be met, these methods can also be used to select data from a large number of measurements to build a compact and robust model. It can help reduce the computational power required to run a real-time SHM system and improve the monitoring performance of the system.

10.2 Future work and outlook

The fast feature selection strategies presented in Chapter 6 can only evaluate the linear relationship between features and output variables. Therefore, it is necessary to modify these strategies or develop other techniques to consider the non-linear relationship, which is crucial for building an SHM system to handle more complex real-world problems. A possible direction is to utilise the kernel trick to evaluate their relations in a kernel-defined feature space [124] or the locally linear embedding technique to introduce a non-linear relationship [125].

In addition, due to the greedy nature of the algorithm used in the fast calculation, the results provided by these calculation methods are sub-optimal. Other more advanced techniques will be tried to help find better solutions. The method of interest is called Shapley value, which assigns each feature its contribution to the output [126]. However, the computational feasibility of obtaining these contribution values is still an issue, and how to speed up the computation to obtain the results could be explored.

By reviewing classical vibration-based monitoring methods and recently developed new sensing techniques [7, 127], measurable quantities in the context of structural health monitoring can be digital signals, such as acceleration and velocity, or visualised images, such as radar satellite data and camera-based images. These fast feature selection strategies could be used to automatically extract features from visualised images to further explore the possible application scenarios of these methods.

The evaluation framework proposed in Chapter 7 only considers the performance of the SHM system to varying degrees. In actual engineering projects, it is inevitable to refer to cost-related criteria. Therefore, cost-related evaluation studies should also be conducted in the future to determine the final sensor system configuration. The cost-related decision-making framework presented in [128] could be used for reference.

The strategies developed in Chapters 8 and 9, including two SPO strategies that are robust to environmental variations and two SPO strategies that are robust to sensor failures, are mainly demonstrated in a supervised way. These strategies can also be applied in an unsupervised manner. In this case, the collected output variables should be replaced with the extracted output variables, as previously mentioned. This work could help to fully demonstrate the benefits of these strategies.

Furthermore, in Chapter 9, in addition to mode shapes, other data types, including time- and frequency-domain data, will be covered to further explore the feasibility and characteristics of the improved fail-safe SPO techniques. If the data are suitable, the three optimisation methods (classical, improved fail-safe, improved fail-safe with redundancy) will be compared in the presence of uncertainty in the optimisation process, and a cost-related optimisation objective or multi-objective optimisation will be considered.

Finally, two additional directions related to sensor system design are suggested here:
(1) Extend the current research area to Bayesian optimisation design. A good example can be seen in [47].
(2) Investigate methods to incorporate possible parametric and non-parametric uncertainties of structural systems to provide valuable data at the design stage, including intrusive and non-intrusive formulations. Improving the availability and reliability of the data during the design phase can further advance the field of optimal design.

Appendix A

A.1 A simple moving average

A simple moving average (SMA) of the n^{th} data point is the unweighted mean of the previous $(m-1)$ data point and itself, which is given by [129],

$$SMA_n = \frac{1}{m} \sum_{i=n-m+1}^{n} p_i \tag{A.1}$$

where p_i is the value of the i^{th} original data point, and m is the number of entries included in the average.

To remove noise in the time series data so that it is suitable for fitting an AR model, the appropriate number of entries m needs to be determined. The time series from 36 sensors at five controlled temperatures (from 5°C to 25°C at intervals of 5°C) were used to achieve this goal. For each m, a newly generated time-series data set was obtained, and an AR model with 20 lags was trained for each time series. The RMSE normalised by the standard derivation of the observed time series was used as the criterion for accessing the suitability of each m. For m ranging from 1 to 60, the averaged results of the normalised RMSE of 36 sensors at five temperatures are shown in Figure A.1. It can be seen that the number of entries for the SMA calculation should be 15, which corresponds to the AR model with the smallest normalised RMSE for all five temperatures.

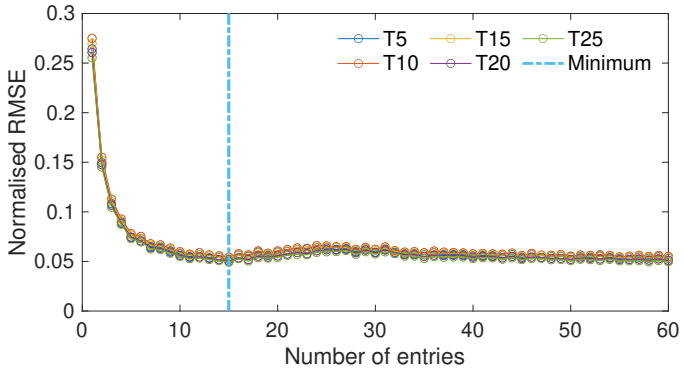


Figure A.1: The relationship between the number of SMA lags and the normalised RMSE.

A.2 An autoregressive model

An AR is one of the most commonly used discrete-time models to describe a system whose inherent behaviour can be characterised by a linear approximation [130]. It can be formulated as,

$$y(n) + a_1 y(n-1) + \cdots + a_m y(n-m) = \xi(n) \tag{A.2}$$

where $y(n)$ is the n^{th} output signal point, m is the model order, $y(n-m)$ is the m^{th} past series value (lag), and $\{\xi(n)\}$ is the noise sequence which is assumed to be independently and identically distributed. The model parameters $[a_1, \ldots, a_k]$ can be estimated by a least-squares algorithm.

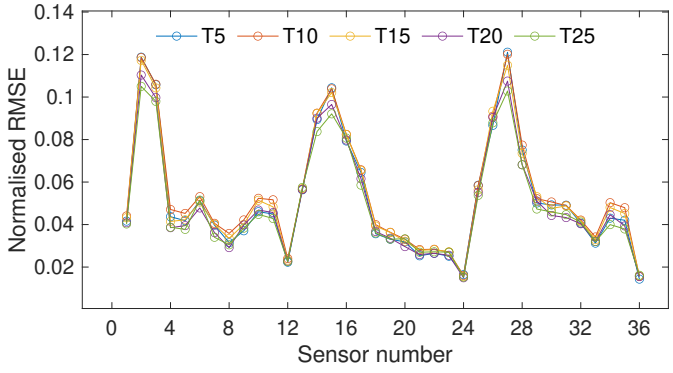


Figure A.2: The normalised RMSE of AR models with 20 lags corresponding to 36 sensors at five different temperatures.

To find the suitable number or lags, m from one to 80 was tried, and again the RMSE normalised by the standard derivation of the observed time series was used

as the criterion to access each m. $m = 20$ was finally selected as the order of the AR model. The normalised RMSE values of AR models with 20 lags corresponding to 36 sensors at five different temperatures are shown in Figure A.2.

Appendix B

B.1 Optimal sensor distributions for three selected modes

Table B.1: Optimal sensor distributions obtained by an ES combined with the SSC for Mode 1.

No. of sensors	Sensor Number
1	15
2	2,17
3	2,15,32
4	1,12,17,36
5	4,28,32,35,36
6	3,5,7,15,20,34

Table B.2: Optimal sensor distributions obtained by an ES combined with the SSC for Mode 2.

No. of sensors	Sensor Number
1	34
2	8,34
3	8,27,34
4	1,8,27,34
5	3,15,17,18,32
6	5,10,15,17,18,32

Table B.3: Optimal sensor distributions obtained by an ES combined with the SSC for Mode 4.

No. of sensors	Sensor Number
1	11
2	16,31
3	16,34,35
4	3,11,16,34
5	7,8,13,29,35
6	3,10,16,25,27,31

B.2 Improved fail-safe SPO results

Table B.4: Optimal improved-fail-safe sensor distributions obtained by a GA combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
4	5,12,28[b],36	2.707e-08	2.384e-07
5	24,28,31,34,36	6.195e-07	3.170e-06
6	4,24,28,34,35,36[a]	1.781e-06	6.260e-06
7	4,12,24,28,34,35,36	5.047e-06	1.266e-05
8	12,16,24,28,31,34,35,36	8.679e-06	1.924e-05

[a] The failure of a sensor at the position marked in blue will result in the worst fitness.
[b] From a fail-safe perspective, the replaceable sensor location is marked in magenta.

Table B.5: Optimal improved-fail-safe sensor distributions obtained by an ES combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
4	5,12,28,36	2.707e-08	2.384e-07
5	24,28,31,34,36	6.195e-07	3.170e-06
6	16,24,28,34,35,36	1.794e-06	6.324e-06

Table B.6: Optimal improved-fail-safe sensor distributions obtained by an GA combined with the SSC.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
2	3,34	0.363	0.781
3	8,27,34	0.756	1.393
4	3,8,30,34	1.164	1.504
5	3,8,27,30,34	1.504	1.719
6	1,15,18,25,30,34	1.890	2.225
7	1,3,17,18,25,30,34	2.083	2.327
8	1,4,17,18,25,26,30,34	2.359	2.513

Table B.7: Optimal fail-safe sensor distributions obtained by a ES combined with the SSC.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
2	3,34	0.363	0.781
3	8,27,34	0.756	1.393
4	3,8,30,34	1.164	1.504
5	3,8,27,30,34	1.504	1.719
6	1,8,15,25,30,34	1.937	2.008

Table B.8: Optimal improved-fail-safe sensor distributions obtained by an GA combined with the SSC weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
2	24,36	0.144	1.098
3	12,24,36	0.776	2.542
4	3,12,24,36	1.918	4.262
5	3,12,16,24,36	2.980	5.571
6	3,12,24,28,32,36	4.100	6.917
7	3,12,21,24,28,32,36	4.995	7.634
8	11,12,23,24,28,29,33,36	5.903	9.018

Table B.9: Optimal improved-fail-safe sensor distributions obtained by a ES combined with the SSC weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
2	24,36	0.144	1.098
3	12,24,36	0.776	2.542
4	3,12,24,36	1.918	4.262
5	3,12,16,24,36	2.980	5.571
6	3,12,24,28,32,36	4.100	6.917

B.3 Improved fail-safe with redundancy SPO results

Table B.10: Optimal improved-fail-safe sensor distributions with redundancy obtained by a GA combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
5	28,31,34,36 (36)[c]	1.812e-07	9.313e-07
6	24,28 (28),34,35,36[d]	9.469e-07	3.382e-06
7	12,16,24,28,34 (34),36	2.575e-06	6.4660e-06
8	4,12,16,24,34 (34),35,36	5.277e-06	1.181e-05

[c] Sensors in parentheses are redundant sensors placed to avoid the worst fitness.

[d] The failure of a sensor at the position marked in green will result in the second worst fitness.

Table B.11: Optimal improved-fail-safe sensor distributions with redundancy obtained by an ES combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
5	28,31,34,36 (36)	1.812e-07	9.313e-07
6	24,28 (28),34,35,36	9.469e-07	3.382e-06

Table B.12: Optimal improved-fail-safe sensor distributions with redundancy obtained by a GA combined with the SSC.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
3	8,34 (34)	0.379	0.943
4	8 (8),21,34	0.834	0.963
5	3,8,27,34 (34)	1.209	1.675
6	3,8,27 (27),30,34	1.567	1.719
7	3,15 (15),17,18,26,32	1.911	2.235
8	3,5,15 (15),17,18,26,32	2.217	2.391

Table B.13: Optimal improved-fail-safe sensor distributions with redundancy obtained by an ES combined with the SSC.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
3	8,34 (34)	0.379	0.943
4	8 (8),21,34	0.834	0.963
5	3,8,27,34 (34)	1.209	1.675
6	3,8,27 (27),30,34	1.567	1.719

Table B.14: Optimal improved-fail-safe sensor distributions with redundancy obtained by a GA combined with the SSC weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
3	12,24 (24)	0.268	1.296
4	12,24 (24),36	1.098	2.542
5	12,16,24 (24),36	2.240	3.798
6	1,3,12 (12),24,36	3.125	5.098
7	3,12,24 (24),28,32,36	4.272	6.917
8	3,12,24 (24),28,32,33,36	5.207	7.870

Table B.15: Optimal improved-fail-safe sensor distributions with redundancy obtained by an ES combined with the SSC weighted by the ADPR.

No. of sensors	Sensor number	Fail-safe fitness	Fitness
3	12,24 (24)	0.268	1.296
4	12,24 (24),36	1.098	2.542
5	12,16,24 (24),36	2.240	3.798
6	1,3,12 (12),34,36	3.125	5.098

B.4 SPO results based on the fitness of mother distributions

Table B.16: Optimal sensor distributions obtained by an ES combined with the DFIM.

No. of sensors	Sensor Number	Fitness	Worst child fitness
3	28,34,36	4.399e-07	N/A
4	12,28[a],34,36	8.566e-07	1.124e-09
5	12,28,31,34,36	1.621e-06	3.182e-07
6	12,16,28,34,35,36	2.855e-06	1.142e-06

[a] The failure of a sensor at the position marked in blue will result in the worst fitness.

Table B.17: Optimal sensor distributions obtained by a GA combined with the DFIM.

No. of sensors	Sensor Number	Fitness	Worst child fitness
3	28,34,36	4.399e-07	N/A
4	12,28,34,36	8.566e-07	1.124e-09
5	12,28,31,34,36	1.621e-06	3.182e-07
6	12,16,28,34,35,36	2.855e-06	1.142e-06
7	12,16,22,28,31,34,36	4.140e-06	1.980e-06
8	12,16,22,24,28,31,34,36	6.128e-06	3.334e-06

Table B.18: Optimal sensor distributions obtained by an ES combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor Number	Fitness	Worst child fitness
3	28,34,36	4.658e-07	N/A
4	24,28,34,36	1.623e-06	7.576e-10
5	12,24,28,34,36	3.457e-06	4.598e-09
6	12,24,28,34,35,36	6.887e-06	1.793e-07

Table B.19: Optimal sensor distributions obtained by a GA combined with the DFIM weighted by the ADPR.

No. of sensors	Sensor Number	Fitness	Worst child fitness
3	28,34,36	4.658e-07	N/A
4	24,28,34,36	1.623e-06	7.576e-10
5	12,24,28,34,36	3.457e-06	4.598e-09
6	12,24,28,31,34,36	6.682e-06	1.309e-06
7	4,12,16,24,34,35,36	1.181e-05	4.772e-06
8	4,12,16,24,28,34,35,36	1.873e-05	7.594e-06

Table B.20: Optimal sensor distributions obtained by an ES combined with the SSC.

No. of sensors	Sensor Number	Fitness	Worst child fitness
1	34	0.379	N/A
2	8,34	0.943	0.289
3	8,27,34	1.393	0.756
4	1,8,27,34	1.779	1.067
5	3,15,17,18,32	2.125	1.467
6	5,10,15,17,18,32	2.500	1.644

Table B.21: Optimal sensor distributions obtained by a GA combined with the SSC.

No. of sensors	Sensor Number	Fitness	Worst child fitness
1	34	0.379	N/A
2	8,34	0.943	0.289
3	8,27,34	1.393	0.756
4	1,8,27,34	1.779	1.067
5	3,15,17,18,32	2.125	1.467
6	5,10,15,17,18,32	2.500	1.644
7	5,15,18,24,26,27,28	2.596	1.752
8	4,6,13,18,20,24,33,35	2.653	1.934

Table B.22: Optimal sensor distributions obtained by an ES combined with the SSC weighted by the ADPR.

No. of sensors	Sensor Number	Fitness	Worst child fitness
1	24	0.268	N/A
2	12,24	1.296	0.131
3	12,24,36	2.542	0.776
4	3,12,24,26	4.262	1.918
5	3,12,24,32,36	5.708	2.833
6	3,12,24,29,32,36	6.918	3.489

Table B.23: Optimal sensor distributions obtained by a GA combined with the SSC weighted by the ADPR.

No. of sensors	Sensor Number	Fitness	Worst child fitness
1	24	0.268	N/A
2	12,24	1.296	0.131
3	12,24,36	2.542	0.776
4	3,12,24,26	4.262	1.918
5	3,12,24,32,36	5.708	2.833
6	12,21,22,24,34,36	6.860	3.331
7	4,11,12,16,24,33,36	8.279	4.951
8	1,6,9,12,16,21,24,36	9.287	4.631

Appendix C

C.1 An illustration of the canonical-correlation-based fast feature selection

The seven instances from the Fisher's iris dataset [72] are given in Table C.1. There are four features used to classify three species of Iris flowers. The goal is to select three of the four features for this classification task.

Sepal Length	Sepal Width	Petal Length	Petal Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3	1.4	0.2	setosa
7	3.2	4.7	1.4	versicolor
6.4	3.2	4.5	1.5	versicolor
6.3	3.3	6	2.5	virginica
5.8	2.7	5.1	1.9	virginica
7.1	3	5.9	2.1	virginica

Table C.1: Fisher's Iris Dataset Sample.

The feature matrix is given by,

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = \begin{pmatrix} 5.1 & 4.9 & 7 & 6.4 & 6.3 & 5.8 & 7.1 \\ 3.5 & 3 & 3.2 & 3.2 & 3.3 & 2.7 & 3 \\ 1.4 & 1.4 & 4.7 & 4.5 & 6 & 5.1 & 5.9 \\ 0.2 & 0.2 & 1.4 & 1.5 & 2.5 & 1.9 & 2.1 \end{pmatrix}^{\top},$$

and the dummy encoded label matrix is,

$$\mathbf{Y} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}^{\top},$$

where $(1, 0)$ represents setosa, $(0, 1)$ represents versicolor, and $(0, 0)$ represents virginica. Therefore, let $N = 7$, $n = 4$, and $m = 2$. The process of the fast canonical correlation-based feature selection method is shown below.

First, centre $\mathbf{X}$ into $\mathbf{X}_{\mathrm{M}}$, which is given by,

$$\mathbf{X}_{\mathrm{M}} = \begin{pmatrix} -0.9857 & -1.1857 & 0.9143 & 0.3143 & 0.2143 & -0.2857 & 1.0143 \\ 0.3714 & -0.1286 & 0.0714 & 0.0714 & 0.1714 & -0.4286 & -0.1286 \\ -2.7429 & -2.7429 & 0.5571 & 0.3571 & 1.8571 & 0.9571 & 1.7571 \\ -1.2000 & -1.2000 & 0.0000 & 0.1000 & 1.1000 & 0.5000 & 0.7000 \end{pmatrix}^{\top}.$$

Then, centre $\mathbf{Y}$ into $\mathbf{Y}_{\mathrm{M}}$, which is given by,

$$\mathbf{Y}_{\mathrm{M}} = \begin{pmatrix} 0.7143 & 0.7143 & -0.2857 & -0.2857 & -0.2857 & -0.2857 & -0.2857 \\ -0.2857 & -0.2857 & 0.7143 & 0.7143 & -0.2857 & -0.2857 & -0.2857 \end{pmatrix}^{\top}.$$

As $N > n + m$, let $\mathbf{F}_{\mathrm{X}} = [\mathbf{X}_{\mathrm{M}}]_{\mathbf{U}}$ and $\mathbf{F}_{\mathrm{Y}} = [\mathbf{Y}_{\mathrm{M}}]_{\mathbf{U}}$, where,

$$\mathbf{F}_{\mathrm{X}} = \begin{pmatrix} 1.8286 & -0.9433 & 0.4666 & -0.1315 & -0.0102 & 0.0041 \\ -0.1782 & -0.2150 & 0.2037 & 0.5233 & 0.0306 & -0.0016 \\ 4.7824 & 0.2504 & -0.0530 & -0.0040 & 0.0116 & -0.0086 \\ 2.1445 & 0.4445 & -0.0823 & 0.1837 & -0.0543 & 0.0111 \end{pmatrix}^{\top},$$

$$\mathbf{F}_{\mathrm{Y}} = \begin{pmatrix} -1.1535 & 0.1772 & 0.2389 & 0.0557 & -0.0799 & -0.0090 \\ 0.2424 & -1.0723 & -0.4584 & 0.0952 & -0.0301 & -0.0022 \end{pmatrix}^{\top}.$$

Here, the orthonormal basis $\mathbf{U}$ is obtained by singular value decomposition of the

matrix $(\mathbf{X}_{\mathrm{M}}, \mathbf{Y}_{\mathrm{M}})$, and $\mathbf{U}$ is given by,

$$\mathbf{U} = \begin{pmatrix} -0.5724 & -0.5810 & 0.1497 & 0.0927 & 0.3680 & 0.1695 & 0.3736 \\ 0.0263 & 0.1509 & -0.6540 & -0.4183 & 0.4083 & 0.4540 & 0.0327 \\ 0.3015 & -0.0626 & 0.0298 & -0.4881 & 0.0159 & -0.4662 & 0.6697 \\ 0.3769 & -0.3212 & -0.0979 & 0.1931 & 0.6321 & -0.4206 & -0.3625 \\ 0.5397 & -0.6196 & 0.0554 & -0.0855 & -0.3155 & 0.4621 & -0.0366 \\ 0.0668 & -0.0759 & -0.6273 & 0.6251 & -0.2359 & -0.1207 & 0.3679 \end{pmatrix}^{\top}.$$

Then, the modified Gram-Schmidt process is applied to $\mathbf{F}_{\mathrm{Y}}$ to obtain $\mathbf{V}$, that is,

$$\mathbf{V} = \begin{pmatrix} -1.1535 & 0.1772 & 0.2389 & 0.0557 & -0.0799 & -0.0090 \\ -0.2190 & -1.0014 & -0.3628 & 0.1175 & -0.0621 & -0.0058 \end{pmatrix}^{\top}.$$

Now, $\mathbf{X}_{\mathrm{s}}$ is empty, and $\mathbf{X}_{\mathrm{a}}$ is the same as $\mathbf{X}$. Correspondingly, $\mathbf{F}_{\mathrm{X}_s}$ is empty and $\mathbf{F}_{\mathrm{X}_a}$ is the same as $\mathbf{F}_{\mathrm{X}}$. Since $\mathbf{W}_{\mathrm{s}}$ is empty, $\mathbf{W}_{\mathrm{a}} = \mathbf{F}_{\mathrm{X}_a}$.

The angle between $\mathbf{w}_{\mathrm{a}i}$ and $\mathbf{V}$ are evaluated by,

$$\begin{aligned} \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}1}, \mathbf{V})\right) &= \theta_{1,1} + \theta_{1,2} = 0.7386 + 0.0242 = 0.7628 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}2}, \mathbf{V})\right) &= \theta_{2,1} + \theta_{2,2} = 0.1047 + 0.1217 = 0.2264 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}3}, \mathbf{V})\right) &= \theta_{3,1} + \theta_{3,2} = 0.9184 + 0.0595 = 0.9779 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}4}, \mathbf{V})\right) &= \theta_{4,1} + \theta_{4,2} = 0.8331 + 0.1273 = 0.9604. \end{aligned} \tag{C.1}$$

Because the petal length ($\mathbf{x}_3$) has the maximum cosine of the angle, it is put into $\mathbf{X}_{\mathrm{s}}$. The rest of the features contained in $\mathbf{X}_{\mathrm{a}}$ are sepal length, sepal width, and petal width.

According to the current $\mathbf{X}_{\mathrm{s}}$ and $\mathbf{X}_{\mathrm{a}}$, update $\mathbf{F}_{\mathrm{X}_s}$ and $\mathbf{F}_{\mathrm{X}_a}$, that is,

$$\mathbf{W}_{\mathrm{s}} = \mathbf{F}_{\mathrm{X}_s} = \begin{pmatrix} 4.7824 & 0.2504 & -0.0530 & -0.0040 & 0.0116 & -0.0086 \end{pmatrix}^{\top}. \tag{C.2}$$

By the modified Gram-Schmidt process, $\mathbf{F}_{\mathrm{X}_a}$ can be orthogonalised to $\mathbf{W}_{\mathrm{s}}$, which is

given by,

$$\mathbf{W}_{\mathrm{a}} = \begin{pmatrix} 0.0596 & -1.0359 & 0.4862 & -0.1300 & -0.0145 & 0.0073 \\ 0.0133 & -0.2049 & 0.2016 & 0.5232 & 0.0310 & -0.0019 \\ -0.0177 & 0.3313 & -0.0584 & 0.1855 & -0.0595 & 0.0150 \end{pmatrix}^{\top}. \quad \text{(C.3)}$$

The angle between $\mathbf{w}_{\mathrm{a}i}$ and $\mathbf{V}$ are evaluated by,

$$\begin{aligned} \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}1}, \mathbf{V})\right) &= \theta_{1,1} + \theta_{1,2} = 0.0107 + 0.4352 = 0.4458 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}2}, \mathbf{V})\right) &= \theta_{2,1} + \theta_{2,2} = 0.0011 + 0.0830 = 0.0841 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}3}, \mathbf{V})\right) &= \theta_{3,1} + \theta_{3,2} = 0.0296 + 0.4348 = 0.4644. \end{aligned} \quad \text{(C.4)}$$

Because the petal width ($\mathbf{x}_4$) has the maximum cosine of the angle, the features now contained in $\mathbf{X}_{\mathrm{s}}$ are petal length and petal width, and the features in $\mathbf{X}_{\mathrm{a}}$ are sepal length and sepal width.

According to the current $\mathbf{X}_{\mathrm{s}}$ and $\mathbf{X}_{\mathrm{a}}$, update ($\mathbf{F}_{\mathrm{X}_s}$ and $\mathbf{F}_{\mathrm{X}_a}$) again. The matrix $\mathbf{W}_{\mathrm{s}}$ is obtained by appending the third column of (C.3) to (C.2), which is given by,

$$\mathbf{W}_{\mathrm{s}} = \begin{pmatrix} 4.7824 & 0.2504 & -0.0530 & -0.0040 & 0.0116 & -0.0086 \\ -0.0177 & 0.3313 & -0.0584 & 0.1855 & -0.0595 & 0.0150 \end{pmatrix}^{\top}.$$

By the modified Gram-Schmidt process, each column of $\mathbf{F}_{\mathrm{X}_a}$ is orthogonalised to $\mathbf{W}_{\mathrm{s}}$, respectively, to obtain $\mathbf{W}_{\mathrm{a}}$, that is,

$$\mathbf{W}_{\mathrm{a}} = \begin{pmatrix} 0.0135 & -0.1714 & 0.3339 & 0.3541 & -0.1698 & 0.0463 \\ 0.0151 & -0.2383 & 0.2075 & 0.5045 & 0.0370 & -0.0034 \end{pmatrix}^{\top}.$$

The angle between $\mathbf{w}_{\mathrm{a}i}$ and $\mathbf{V}$ are evaluated by,

$$\begin{aligned} \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}1}, \mathbf{V})\right) &= \theta_{1,1} + \theta_{1,2} = 0.0105 + 0.0277 = 0.0382 \\ \cos^2\left(\Theta(\mathbf{w}_{\mathrm{a}2}, \mathbf{V})\right) &= \theta_{2,1} + \theta_{2,2} = 0.0004 + 0.1103 = 0.1108. \end{aligned} \quad \text{(C.5)}$$

The sepal width ($\mathbf{x}_2$), which has the maximum cosine of the angle, is put into $\mathbf{X}_{\mathrm{s}}$. The feature selection process ends here.

To verify the equality between the definition-based method and the θ-angle method,

the left-hand side of (6.9) is given by,

$$\begin{aligned} R_1^2((\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_2), \mathbf{Y}) + R_2^2((\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_2), \mathbf{Y}) &= 0.9905 + 0.5626 \\ &= 1.5531, \end{aligned}$$

and the right-hand side of (6.9) is given by the sum of the maxima in (C.1), (C.4) and (C.5), that is $0.9779 + 0.4644 + 0.1108 = 1.5531$.

Bibliography

[1] C. R. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, Ltd., Chichester, United Kingdom, 2012.

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

[3] E. Cross. *On structural health monitoring in changing environmental and operational conditions*. PhD thesis, University of Sheffield, 2012.

[4] B. Spencer Jr, M. E. Ruiz-Sandoval, and N. Kurata. Smart sensing technology: Opportunities and challenges. *Structural Control and Health Monitoring*, 11 (4):349–368, 2004.

[5] J. P. Lynch and K. J. Loh. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest*, 38(2): 91–130, 2006.

[6] M. Mitra and S. Gopalakrishnan. Guided wave based structural health monitoring: A review. *Smart Materials and Structures*, 25(5), 2016. doi: 10.1088/0964-1726/25/5/053001.

[7] S. Sony, S. Laventure, and A. Sadhu. A literature review of next-generation smart sensing technology in structural health monitoring. *Structural Control and Health Monitoring*, 26(3), 2019. doi: 10.1002/stc.2321.

[8] C. R. Farrar, H. Sohn, and K. Worden. Data normalization: A key for structural

health monitoring. Technical report, Los Alamos National Lab., New Mexico, United States, 2001.

[9] K. Worden, H. Sohn, and C. R. Farrar. Novelty detection in a changing environment: Regression and interpolation approaches. *Journal of Sound and Vibration*, 258(4):741–761, 2002.

[10] Y. Ni, H. Zhou, and J. Ko. Generalization capability of neural network models for temperature-frequency correlation using monitoring data. *Journal of Structural Engineering*, 135(10):1290–1300, 2009.

[11] K. Worden, E. J. Cross, and J. M. Brownjohn. Switching response surface models for structural health monitoring of bridges. In *Surrogate-Based Modeling and Optimization*, pages 337–358. Springer, 2013.

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

[13] A. Deraemaeker and K. Worden. A comparison of linear approaches to filter out environmental effects in structural health monitoring. *Mechanical Systems and Signal Processing*, 105:1–15, 2018.

[14] E. J. Cross, K. Worden, and Q. Chen. Cointegration: A novel approach for the removal of environmental trends in structural health monitoring data. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2133):2712–2732, 2011.

[15] J. Kullaa. Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring. *Mechanical Systems and Signal Processing*, 25(8):2976–2989, 2011.

[16] A. Liu, L. Wang, L. Bornn, and C. Farrar. Robust structural health monitoring under environmental and operational uncertainty with switching state-space autoregressive models. *Structural Health Monitoring*, 18(2):435–453, 2019.

[17] R. Marks, A. Clarke, C. Featherston, L. K. C. Paget, and R. Pullin. Using genetic algorithms to optimize an active sensor network on a stiffened aerospace panel with 3D scanning laser vibrometry data. In *Proceedings of the 11th International Conference on Damage Assessment of Structures*, volume 628. IOP Publishing, 2015. doi: 10.1088/1742-6596/628/1/012116.

[18] R. Marks, A. Clarke, C. A. Featherston, and R. Pullin. Optimization of acousto-ultrasonic sensor networks using genetic algorithms based on experimental and numerical data sets. *International Journal of Distributed Sensor Networks*, 13 (11), 2017. doi: 10.1177/1550147717743702.

[19] D.-S. Li, H.-N. Li, and C.-P. Fritzen. Load dependent sensor placement method: Theory and experimental validation. *Mechanical Systems and Signal Processing*, 31:217–227, 2012.

[20] M. Kuhn, K. Johnson, *et al. Applied Predictive Modeling*, volume 26. Springer, New York, United States, 2013.

[21] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[22] Y. Li, T. Li, and H. Liu. Recent advances in feature selection and its applications. *Knowledge and Information Systems*, 53(3):551–577, 2017.

[23] M. A. Hall and L. A. Smith. Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper. In *Proceedings of the 12th International FLAIRS Conference*, page 235–239. AAAI Press, 1999.

[24] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[25] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján. Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13(1):27–66, 2012.

[26] H. Kaya, F. Eyben, A. A. Salah, and B. Schuller. CCA based feature selection with application to continuous depression recognition from acoustic speech features. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3729–3733. IEEE, 2014.

[27] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad. A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2):907–948, 2020.

[28] R. Varshavsky, A. Gottlieb, M. Linial, and D. Horn. Novel unsupervised feature filtering of biological data. *Bioinformatics*, 22(14):e507–e513, 2006.

[29] V. M. Rao and V. Sastry. Unsupervised feature ranking based on representation entropy. In *First International Conference on Recent Advances in Information Technology*, pages 421–425. IEEE, 2012.

[30] M. Banerjee and N. R. Pal. Feature selection with SVD entropy: Some modification and extension. *Information Sciences*, 264:118–134, 2014.

[31] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006.

[32] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In *Advances in Neural Information Processing Systems*, volume 18, pages 507–514, 2005.

[33] Z. Zhao and H. Liu. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1151–1157, 2007.

[34] I. T. Jolliffe. *Principal Component Analysis*, chapter 6, pp. 137-145. Springer-Verlag, New York, United States, second edition, 2002.

[35] M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. In *IEEE International Conference on Data Mining*, pages 115–122. IEEE, 2002.

[36] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian. Feature selection using principal feature analysis. In *Proceedings of the 15th ACM International Conference on Multimedia*, pages 301–304, 2007.

[37] A. J. Ferreira and M. A. Figueiredo. An unsupervised approach to feature discretization and selection. *Pattern Recognition*, 45(9):3048–3060, 2012.

[38] S. Wang, W. Pedrycz, Q. Zhu, and W. Zhu. Unsupervised feature selection via maximum projection and minimum redundancy. *Knowledge-Based Systems*, 75:19–29, 2015.

[39] J. F. Cadima and I. T. Jolliffe. Variable selection and the interpretation of principal subspaces. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(1):62–79, 2001.

[40] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. *Feature Extraction: Foundations and Applications*. Springer, Heidelberg, Germany, 2008.

[41] D. C. Kammer. Sensor placement for on-orbit modal identification and correlation of large space structures. *Journal of Guidance, Control, and Dynamics*, 14(2):251–259, 1991.

[42] L. Yao, W. A. Sethares, and D. C. Kammer. Sensor placement for on-orbit modal identification via a genetic algorithm. *AIAA Journal*, 31(10):1922–1928, 1993.

[43] D. C. Kammer. Effect of model error on sensor placement for on-orbit modal identification of large space structures. *Journal of Guidance, Control, and Dynamics*, 15(2):334–341, 1992.

[44] D. C. Kammer. Effects of noise on sensor placement for on-orbit modal identification of large space structures. *Journal of Dynamic Systems Measurement and Control, Transactions of the ASME*, 114(3):436–443, 1992.

[45] M. M. Jaya, R. Ceravolo, L. Z. Fragonara, and E. Matta. An optimal sensor placement strategy for reliable expansion of mode shapes under measurement noise and modelling error. *Journal of Sound and Vibration*, 487, 2020. doi: 10.1016/j.jsv.2020.115511.

[46] C. Papadimitriou. Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration*, 278(4-5): 923–947, 2004.

[47] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(8):235–284, 2008.

[48] K. Worden and A. P. Burrows. Optimal sensor placement for fault detection. *Engineering Structures*, 23(8):885–901, 2001.

[49] A. T. Eshghi, S. Lee, H. Jung, and P. Wang. Design of structural monitoring sensor network using surrogate modeling of stochastic sensor signal. *Mechanical Systems and Signal Processing*, 133, 2019. doi: 10.1016/j.ymssp.2019.106280.

[50] E. B. Flynn and M. D. Todd. A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing. *Mechanical Systems and Signal Processing*, 24(4):891–903, 2010.

[51] W. Ostachowicz, R. Soman, and P. Malinowski. Optimization of sensor placement for structural health monitoring: A review. *Structural Health Monitoring*, 18(3):963–988, 2019.

[52] Y. Tan and L. Zhang. Computational methodologies for optimal sensor placement in structural health monitoring: A review. *Structural Health Monitoring*, 19(4):1287–1308, 2020.

[53] R. J. Barthorpe and K. Worden. Emerging trends in optimal structural health monitoring system design: From sensor placement to system evaluation. *Journal of Sensor and Actuator Networks*, 9(3):31–51, 2020.

[54] S.-K. Choi, R. Grandhi, and R. A. Canfield. *Reliability-based Structural Design*. Springer Science & Business Media, London, United Kingdom, 2006.

[55] R. C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Siam, Philadelphia, United States, 2013.

[56] T. Imamura, W. C. Meecham, and A. Siegel. Symbolic calculus of the wiener process and wiener-hermite functionals. *Journal of Mathematical Physics*, 6 (5):695–706, 1965.

[57] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63 (3):425–464, 2001.

[58] S. Haykin. *Neural Networks, A Comprehensive Foundation*. Prentice Hall, New Jersey, United States, second edition, 2004.

[59] R. F. Guratzsch and S. Mahadevan. Structural health monitoring sensor placement optimization under uncertainty. *AIAA Journal*, 48(7):1281–1289, 2010.

[60] R. Castro-Triguero, S. Murugan, R. Gallego, and M. I. Friswell. Robustness of optimal sensor placement under parametric uncertainty. *Mechanical Systems and Signal Processing*, 41(1-2):268–287, 2013.

[61] S. Riedmaier, B. Danquah, B. Schick, and F. Diermeyer. Unified framework and survey for model verification, validation and uncertainty quantification. *Archives of Computational Methods in Engineering*, 28(4):2655–2688, 2021.

[62] H. W. Coleman and W. G. Steele. *Experimentation, Validation, and Uncertainty Analysis for Engineers*. John Wiley & Sons, Hoboken, United States, fourth edition, 2018.

[63] N. R. Council. *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. The National Academies Press, Washington, United States, 2012.

[64] P. C. Chang, A. Flatau, and S.-C. Liu. Health monitoring of civil infrastructure. *Structural Health Monitoring*, 2(3):257–267, 2003.

[65] J. M. Brownjohn. Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):589–622, 2007.

[66] S. Side, W. Staszewski, R. Wardle, and K. Worden. Fail-safe sensor distributions for damage detection. In *Proceedings of the DAMAS*, volume 97, pages 135–146, 1997.

[67] M. Z. A. Bhuiyan, G. Wang, J. Cao, and J. Wu. Deploying wireless sensor networks with fault-tolerance for structural health monitoring. *IEEE Transactions on Computers*, 64(2):382–395, 2013.

[68] M. Iqbal, M. Naeem, A. Anpalagan, A. Ahmed, and M. Azam. Wireless sensor network optimization: Multi-objective paradigm. *Sensors*, 15(7):17572–17620, 2015.

[69] S. Abdollahzadeh and N. J. Navimipour. Deployment strategies in the wireless sensor network: A comprehensive review. *Computer Communications*, 91-92: 1–16, 2016.

[70] J. Kullaa. Detection, identification, and quantification of sensor fault in a sensor network. *Mechanical Systems and Signal Processing*, 40(1):208–221, 2013.

[71] T. Sun and S. Chen. Class label versus sample label-based CCA. *Applied Mathematics and Computation*, 185(1):272–283, 2007.

[72] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[73] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, Cambridge, United Kingdom, 2012.

[74] R. A. Fisher. The statistical utilization of multiple measurements. *Annals of Eugenics*, 8(4):376–386, 1938.

[75] F. Binet and G. Watson. Algebraic theory of the computing routine for tests of significance on the dimensionality of normal multivariate systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 18(1):70–78, 1956.

[76] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4): 321–377, 1936.

[77] S. Zhang and Z.-Q. Lang. Orthogonal least squares based fast feature selection for linear classification. *Pattern Recognition*, 123, 2022. doi: 10.1016/j.patcog.2021.108419.

[78] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664, 2004.

[79] T. F. Cox. *An Introduction to Multivariate Data Analysis*, chapter 22, pp. 141-143. Hodder Education, London, United Kingdom, 2005.

[80] M. Barker and W. Rayens. Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(3):166–173, 2003.

[81] L. Zhang, K. Li, E.-W. Bai, and G. W. Irwin. Two-stage orthogonal least squares methods for neural network construction. *IEEE Transactions on Neural Networks and Learning Systems*, 26(8):1608–1621, 2015.

[82] E. Balkanski, B. Mirzasoleiman, A. Krause, and Y. Singer. Learning sparse combinatorial representations via two-stage submodular maximization. In *International Conference on Machine Learning*, volume 48, pages 2207–2216. PMLR, 2016.

[83] X. Xu, M. Wang, Y. Wang, and D. Ma. Two-stage routing with optimized guided search and greedy algorithm on proximity graph. *Knowledge-Based Systems*, 229, 2021. doi: 10.1016/j.knosys.2021.107305.

[84] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[85] K. Sörensen. Metaheuristics—the metaphor exposed. *International Transactions in Operational Research*, 22(1):3–18, 2015.

[86] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[87] M. Dorigo, V. Maniezzo, and A. Colorni. Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 26(1):29–41, 1996.

[88] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr. A survey on metaheuristics for stochastic combinatorial optimization. *Natural Computing*, 8(2):239–287, 2009.

[89] S. Luke. *Essentials of Metaheuristics*. Lulu, second edition, 2013.

[90] C. R. Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447, 1975.

[91] W. H. Greene. *Econometric Analysis*. Pearson Education, Inc., New Jersey, United States, fifth edition, 2003.

[92] J. V. Beck and K. J. Arnold. *Parameter Estimation in Engineering and Science*. John Wiley & Sons, New York, United States, 1977.

[93] D. Middleton. *An Introduction to Statistical Communication Theory*. Wiley-IEEE Press, New York, United States, 1996.

[94] N. Imamovic and D. Ewins. Optimization of excitation dof selection for modal tests. In *Proceedings of the 15th International Modal Analysis Conference*, volume 3089, pages 1945–1951, 1997.

[95] N. Imamovic. *Validation of Large Structural Dynamics Models Using Modal Test Data*. PhD thesis, Imperial College London, 1998.

[96] N. R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley & Sons, Hoboken, United States, third edition, 1998.

[97] G. A. Miller and P. E. Nicely. An analysis of perceptual confusions among some english consonants. *The Journal of the Acoustical Society of America*, 27 (2):338–352, 1955.

[98] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[99] N. M. M. Maia and J. M. M. e. Silva. *Theoretical and Experimental Modal Analysis*. Research Studies Press LTD., Baldock, United Kingdom, 1997.

[100] M. Pastor, M. Binda, and T. Harčarik. Modal assurance criterion. *Procedia Engineering*, 48:543–548, 2012.

[101] S. Zhang, T. Wang, K. Worden, and E. J. Cross. Canonical-correlation-based fast feature selection. *arXiv preprint arXiv:2106.08247*, 2021.

[102] W. W. Rozeboom. Linear correlations between sets of variables. *Psychometrika*, 30(1):57–71, 1965.

[103] D. Dua and C. Graff. UCI machine learning repository, 2019. URL `http://archive.ics.uci.edu/ml`.

[104] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[105] P. M. Ciarelli and E. Oliveira. Agglomeration and elimination of terms for dimensionality reduction. In *Ninth International Conference on Intelligent Systems Design and Applications*, pages 547–552. IEEE, 2009.

[106] M. van Breukelen, R. P. W. Duin, D. M. J. Tax, and J. E. den Hartog. Handwritten digit recognition by combined classifiers. *Kybernetika*, 34(4): 381–386, 1998.

[107] P. Cortez and A. M. G. Silva. Using data mining to predict secondary school student performance. In *Proceedings of the Fifth Future Business Technology Conference*, pages 5–12. EUROSIS-ETI, 2008.

[108] B. E. Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgen, S. Delil, H. Apaydin, and O. Kursun. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, 2013.

[109] K. Hamidieh. A data-driven statistical model for predicting the critical temperature of a superconductor. *Computational Materials Science*, 154:346–354, 2018.

[110] L. M. Candanedo, V. Feldheim, and D. Deramaix. Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140:81–97, 2017.

[111] P. D. Allison. *Multiple Regression: A Primer*. Pine Forge Press Inc., California, United States, 1999.

[112] J. B. Bowles and C. E. Peláez. Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis. *Reliability Engineering & System Safety*, 50(2):203–213, 1995.

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

[114] L. Bull, K. Worden, R. Fuentes, G. Manson, E. Cross, and N. Dervilis. Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data. *Journal of Sound and Vibration*, 453: 126–150, 2019.

[115] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, London, United Kingdom, 2012.

[116] K. Worden, G. Manson, G. Hilson, and S. Pierce. Genetic optimisation of a neural damage locator. *Journal of Sound and Vibration*, 309(3-5):529–544, 2008.

[117] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Science & Business Media, New York, United States, 2013.

[118] M. A. Hall. *Correlation-based Feature Selection for Machine Learning*. PhD thesis, The University of Waikato, 1999.

[119] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th International Conference on Machine Learning*, pages 856–863, 2003.

[120] Z. Chen, Y. Bao, Z. Tang, J. Chen, and H. Li. Clarifying and quantifying the geometric correlation for probability distributions of inter-sensor monitoring data: A functional data analytic methodology. *Mechanical Systems and Signal Processing*, 138, 2020. doi: 10.1016/j.ymssp.2019.106540.

[121] W. Lu, R. Wen, J. Teng, X. Li, and C. Li. Data correlation analysis for optimal sensor placement using a bond energy algorithm. *Measurement*, 91:509–518, 2016.

[122] W. Dillon and M. Goldstein. *Multivariate Analysis: Methods and Applications*. John Wiley & Sons, New York, United States, 1984.

[123] A. E. Eiben, J. E. Smith, *et al*. *Introduction to evolutionary computing*. Springer-Verlag, Berlin/Heidelberg, Germany, 2015.

[124] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, United States, 2004.

[125] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[126] L. S. Shapley. Notes on the n-person game—II: The value of an n-person game. *The RAND Corporation*, 1951.

[127] D. Goyal and B. Pabla. The vibration monitoring methods and signal processing techniques for structural health monitoring: A review. *Archives of Computational Methods in Engineering*, 23:585–594, 2016.

[128] A. J. Hughes, L. A. Bull, P. Gardner, R. J. Barthorpe, N. Dervilis, and K. Worden. On risk-based active learning for structural health monitoring. *Mechanical Systems and Signal Processing*, 167, 2022. doi: 10.1016/j.ymssp.2021.108569.

[129] C. A. Ellis and S. A. Parbery. Is smarter better? a comparison of adaptive, and simple moving average trading strategies. *Research in International Business and Finance*, 19(3):399–411, 2005.

[130] S. A. Billings. *Nonlinear system identification: NARMAX Methods in the Time, Frequency, and Spatio-temporal Domains*. John Wiley & Sons, West Sussex, United Kingdom, 2013.