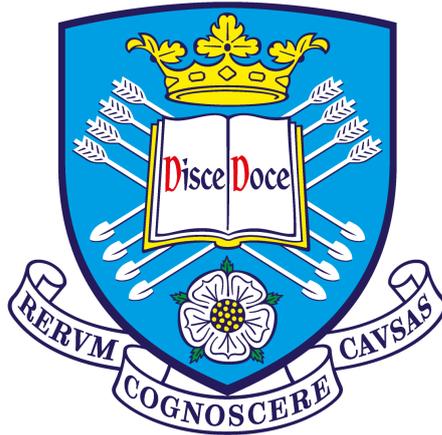


On Risk-Based Decision-Making for Structural Health Monitoring



A Thesis submitted to the University of Sheffield
for the degree of Doctor of Philosophy in the Faculty of Engineering

by

Aidan J. Hughes

Department of Mechanical Engineering

University of Sheffield

September 2022

ABSTRACT

Structural health monitoring (SHM) technologies seek to detect, localise, and characterise damage present within structures and infrastructure. Arguably, the foremost incentive for developing and implementing SHM systems is to improve the quality of operation and maintenance (O&M) strategies for structures, such that safety can be enhanced, or greater economic benefits can be realised. Given this motivation, SHM systems can be considered primarily as decision-support tools. Although much research has been conducted into damage identification and characterisation approaches, there has been relatively little that has explicitly considered the decision-making applications of SHM systems. In light of this fact, the current thesis seeks to consider decision-making for SHM with respect to *risk*. Risk, defined as a product of probability and cost, can be interpreted as an *expected utility*.

The keystone of the current thesis is a general framework for conducting risk-based, SHM generated by combining aspects of *probabilistic risk assessment* (PRA) with the existing statistical pattern recognition paradigm for SHM. The framework, founded on *probabilistic graphical models* (PGMs), utilises Bayesian network representations of fault-trees to facilitate the flow of information between observations of discriminative features to failure states of structures of interest. Using estimations of failure probabilities in conjunction with utility functions that capture the severity of consequences enables risk assessments – these risks can be minimised with respect to candidate maintenance actions to determine optimal strategies. Key elements of the decision framework are examined; in particular, a physics-based methodology for initialising a structural degradation model defining health-state transition probabilities is presented.

The risk-based framework allows aspects of SHM systems to be developed with

explicit consideration for the decision-support applications. In relation to this aim, the current thesis proposes a novel approach to learn statistical classification models within an online SHM system. The approach adopts an active learning framework in which descriptive labels, corresponding to salient health states of a structure, are obtained via structural inspections. To account for the decision processes associated with SHM, structural inspections are mandated according to the *expected value of information* for data-labels. The resulting risk-based active learning algorithm is shown to yield cost-effective improvements in the performance of decision-making agents, in addition to reducing the number of manual inspections made over the course of a monitoring campaign.

Characteristics of the risk-based active learning algorithm are further investigated, with particular focus on the effects of *sampling bias*. Sampling bias is known to degrade decision-making performance over time, thus engineers have a vested interest in mitigating its negative effects. On this theme, two approaches are considered for improving risk-based active learning; semi-supervised learning, and discriminative classification models. Semi-supervised learning yielded mixed results, with performance being highly dependent on base distributions being representative of the underlying data. On the other hand, discriminative classifiers performed strongly across the board. It is shown that by mitigating the negative effects of sampling bias via classifier and algorithm design, decision-support systems can be enhanced, resulting in more cost-effective O&M strategies.

Finally, the future of risk-based decision-making is considered. Particular attention is given to population-based structural health monitoring (PBSHM), and the management of fleets of assets. The hierarchical representation of structures used to develop the risk-based SHM framework is extended to populations of structures. Initial research into PBSHM shows promising results with respect to the transfer of information between individual structures comprising a population. The significance of these results in the context of decision-making is discussed.

To summarise, by framing SHM systems as decision-support tools, risk-informed O&M strategies can be developed for structures and infrastructure such that safety is improved and costs are reduced.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Prof. Keith Worden for being a superb supervisor – a seemingly endless font of knowledge (both technical and otherwise), who has continuously provided support, guidance, and encouragement. I would also like to give profound thanks to my brilliant secondary supervisors; Dr. Rob Barthorpe, Dr. Paul Gardner, and Prof. Nikos Dervilis – each of whom have been generous with their time, expertise, and support. Special thanks also to Dr. Lawrence Bull and Dr. Chuck Farrar, for their help and insight.

I have thoroughly enjoyed the past several years, not least because of the people that I have found myself surrounded by. Thank you to the Dynamics Research Group in Sheffield; you are an amazing group of people, creating a friendly and supportive environment for research. In addition to those already mentioned, I would particularly like to thank: Matt T, Tim Rogers, George, Aris, Tim Rooker, Dan, Lizzy, and David. In addition, I would like to sincerely thank the following people for being an incredibly supportive network of friends: Matt J, Marcus, Max, Chandy, Julian, and Chris.

Outside of work, several other friends have also made life enjoyable – thank you to Sam, Alice, Gina, General Panic, and, of course, the ‘Big Mec Lads’. Cheers to Jack, Andy, Pete, Jon, Dec, Ben, Tom, and Ian for all the catch-ups back home.

My family have been a constant source of encouragement and support. I am deeply grateful to them for the opportunities that they have provided me with, and for the love that they have shown me. Thank you to: my parents, Karen and Andrew; my brother, Calum; my sister, Rachel; and my grandparents, Sonia and Peter. Thanks to Uncle Andrew, for sparking my interest in artificial intelligence and machine learning.

And thanks to Molly, for being the best dog.

Finally, I would like to thank Tina. You have been immensely supportive, caring, and loving. You have helped me in countless ways, and I couldn't have done this without you.

Once again, thank you to those that have helped me – I owe all of you a pint!

Dedicated to Tony

TABLE OF CONTENTS

| | |
|--|------------|
| Abstract | i |
| Acknowledgements | iii |
| 1 Introduction | 1 |
| 1.1 SHM: Aims, Approaches and Challenges | 1 |
| 1.2 SHM for Decision Support | 3 |
| 1.3 Risk-Based Decision-Making | 4 |
| 1.4 Thesis Contribution | 6 |
| 1.5 Chapter Summary | 7 |
| 2 Literature Survey: Risk-based SHM | 9 |
| 2.1 Inspection and Maintenance Planning | 9 |
| 2.2 Value of Information Quantification | 11 |
| 2.3 SHM System Optimisation | 12 |
| 3 Background | 13 |
| 3.1 Structural Health Monitoring | 13 |

| | | |
|----------|--|-----------|
| 3.1.1 | Operational Evaluation | 13 |
| 3.1.2 | Data Acquisition | 14 |
| 3.1.3 | Feature Selection | 14 |
| 3.1.4 | Statistical Modelling for Feature Discrimination | 15 |
| 3.2 | Probabilistic Risk Assessment | 15 |
| 3.2.1 | The History of PRA | 16 |
| 3.2.2 | Initial Information Collection | 18 |
| 3.2.3 | Event-tree Development | 18 |
| 3.2.4 | Fault-tree Development | 18 |
| 3.2.5 | Reliability Modelling | 19 |
| 3.2.6 | Failure Sequence Quantification | 20 |
| 3.2.7 | Consequence Analysis | 20 |
| 3.3 | Summary | 21 |
| 4 | Decision-making Under Uncertainty | 22 |
| 4.1 | Probability Theory | 22 |
| 4.2 | Decision Theory | 24 |
| 4.2.1 | Fundamentals of Decision Theory | 24 |
| 4.2.2 | Limitations | 26 |
| 4.3 | Probabilistic Graphical Models | 27 |
| 4.3.1 | Bayesian Networks | 28 |
| 4.3.2 | Dynamic Bayesian Networks | 29 |
| 4.3.3 | Influence Diagrams | 31 |
| 4.3.4 | Markov Decision Processes | 32 |

| | | |
|----------|---|-----------|
| 4.4 | Summary | 34 |
| 5 | Mapping PRA onto SHM | 35 |
| 5.1 | Definitions | 36 |
| 5.2 | Mapping PRA onto SHM | 38 |
| 5.2.1 | Operational evaluation | 39 |
| 5.2.2 | Failure-mode Modelling | 40 |
| 5.2.3 | Data Acquisition | 41 |
| 5.2.4 | Feature Selection | 42 |
| 5.2.5 | Statistical Modelling | 42 |
| 5.2.6 | Decision Modelling | 43 |
| 5.3 | Discussion | 43 |
| 5.4 | Summary | 44 |
| 6 | Case Study: Four-bay Truss | 45 |
| 6.1 | Four-bay Truss | 45 |
| 6.1.1 | ‘Operational Evaluation’ of Truss | 46 |
| 6.1.2 | Truss Failure Modelling | 48 |
| 6.1.3 | Health-state Transition Modelling | 48 |
| 6.1.4 | Statistical Classifier Development and Training | 50 |
| 6.1.5 | Overall Probabilistic Graphical Model | 53 |
| 6.1.6 | Results | 53 |
| 6.2 | Discussion | 63 |
| 6.3 | Summary | 65 |

| | | |
|----------|--|-----------|
| 7 | Health-state Transition Models | 66 |
| 7.1 | Degradation Modelling: An Overview | 67 |
| 7.2 | Transition Models in Risk-based SHM | 68 |
| 7.3 | Generating Transition Models | 69 |
| 7.3.1 | Operational Evaluation | 69 |
| 7.3.2 | Handling Uncertainty | 71 |
| 7.3.3 | Generating Degradation Models | 71 |
| 7.3.4 | Generating Intervention Models | 72 |
| 7.3.5 | Case Study: Truss | 73 |
| 7.3.6 | Operational Conditions | 74 |
| 7.3.7 | Failure Criteria | 74 |
| 7.3.8 | Transition Modelling | 76 |
| 7.4 | Discussion | 79 |
| 7.4.1 | Importance of Transition Models | 79 |
| 7.4.2 | Challenges | 80 |
| 7.5 | Summary | 82 |
| 8 | Statistical Classifiers for Decision-making | 83 |
| 8.1 | Statistical Classifiers for SHM | 83 |
| 8.2 | Machine Learning Paradigms | 85 |
| 8.2.1 | Supervised and unsupervised learning for SHM | 85 |
| 8.2.2 | Active Learning for SHM | 86 |
| 8.3 | Risk-based Active Learning | 87 |
| 8.3.1 | Classifier Initiation | 87 |

| | | |
|----------|--|------------|
| 8.3.2 | Value of Information | 88 |
| 8.3.3 | Inspection Scheduling | 90 |
| 8.3.4 | Assessing Performance | 90 |
| 8.4 | Numerical Example | 91 |
| 8.4.1 | Decision Process | 92 |
| 8.4.2 | Statistical Classifier | 94 |
| 8.4.3 | Example EVPI Calculation | 98 |
| 8.4.4 | Results | 99 |
| 8.5 | Experimental Case Study | 105 |
| 8.5.1 | Decision Process | 107 |
| 8.5.2 | Statistical Classifier | 110 |
| 8.5.3 | Results | 110 |
| 8.6 | Discussion | 116 |
| 8.6.1 | Sampling Bias | 117 |
| 8.7 | Summary | 118 |
| 9 | Improving Decision-making via Classifier Design | 119 |
| 9.1 | Effects of Sampling Bias | 120 |
| 9.1.1 | Case Study: Visual Example | 121 |
| 9.1.2 | Further Comments | 127 |
| 9.2 | Approaches to Address Sampling Bias | 129 |
| 9.2.1 | Semi-supervised Learning | 129 |
| 9.2.2 | Discriminative Classifiers | 142 |
| 9.3 | Case Study: Z24 Bridge Dataset | 156 |

| | | |
|-----------|---|------------|
| 9.3.1 | Decision Process | 157 |
| 9.3.2 | Results: Gaussian Mixture Model | 157 |
| 9.3.3 | Results: Semi-supervised Learning | 160 |
| 9.3.4 | Results: Discriminative Classifiers | 164 |
| 9.4 | Discussion | 168 |
| 9.5 | Summary | 170 |
| 10 | Towards Risk-based Decision-making for Populations of Structures | 172 |
| 10.1 | Population-based SHM | 173 |
| 10.1.1 | Homogeneous and Heterogeneous Populations | 174 |
| 10.1.2 | Similarity Between Structures | 175 |
| 10.1.3 | Mapping and Transfer | 176 |
| 10.2 | Structures as Hierarchies | 177 |
| 10.3 | Populations of Structures as Hierarchies | 178 |
| 10.4 | Risk-informed PBSHM | 179 |
| 10.4.1 | Operational Evaluation | 180 |
| 10.4.2 | Inferences and Decisions | 181 |
| 10.4.3 | Value of Information Transfer | 183 |
| 10.5 | Conclusions | 183 |
| 11 | Conclusions and Future Work | 184 |
| 11.1 | Summary and Conclusions | 185 |
| 11.2 | Limitations and Challenges | 187 |
| 11.3 | Future Work | 190 |
| 11.4 | Closing Remarks | 192 |

| | |
|--|------------|
| Appendices | 193 |
| A Notes on Probability | 193 |
| A.1 Foundations of Probability Theory | 193 |
| A.2 Basic Rules of Probability | 194 |
| A.3 The Bayesian Interpretation of Probability | 196 |
| A.4 Bayesian Network Independence Structures | 197 |
| B Publications | 198 |
| B.1 Journal Papers | 198 |
| B.2 Book Chapters | 199 |
| B.3 Conference Papers | 199 |
| Bibliography | 200 |

INTRODUCTION

Structural health monitoring (SHM) is a field of engineering concerned with the development and implementation of online damage-identification strategies for mechanical, aerospace, and civil infrastructure [1]. Although many technological advances have been made in structural damage detection, relatively little research has been conducted into one of the primary motivations for SHM – asset management decision support. This thesis presents novel risk-based approaches to decision-making under uncertainty in the context of SHM.

1.1 Structural Health Monitoring: Aims, Approaches and Challenges

In the context of SHM, damage can be defined as material and/or geometric changes that may presently, or at some future time, adversely affect the performance of a structure [1]. SHM systems aim to provide information regarding the damage present in a structure of interest; characteristically, this information is interpreted from measured data via some online process. Once extracted, this information can be used to improve decisions associated with the operation and maintenance (O&M) of high-value and/or safety-critical assets, such that the required level of performance is maintained. The damage information sought by SHM systems, and the engineers tasked with asset management, can be evaluated with respect to levels in a hierarchy, originally presented by Rytter in [2], and later extended in [3]:

1. *Detection*: information related to the presence of damage.
2. *Localisation*: information related to the position of damage.
3. *Classification*: information related to the type of damage.
4. *Quantification*: information related to the extent of damage.
5. *Prognosis*: information related to progression of damage and the remaining useful life of a structure.

Rytter’s hierarchy highlights somewhat of a double-edged sword in SHM. In general, as one progresses further in the hierarchy, the information available becomes increasingly useful for decision-making; however, the effort necessary to extract the information also escalates.

Damage is typically difficult, if not impossible, to measure directly [1]. Therefore, in order to parse damage information from indirect measurements, SHM systems rely on models. There are two fundamental approaches to the development of such models: *physics-based* and *data-based*. Physics-based (also referred to as *white-box*) models are developed using knowledge of the laws of nature that seek to causally describe the observable behaviours exhibited by systems operating at various scales and when subject to various constraints. On the other hand, data-based (or *black-box*) models are inferred using empirical measurements of the behaviours exhibited by systems and generally seek to represent correlations that exist between the obtained observations¹. Finally, there exists a third ‘hybrid’ approach that combines the physics-based and data-based approaches in what are referred to as *grey-box* models.

While the physics-based, data-based and hybrid approaches each have their own merits and drawbacks [5, 6], they ultimately all warrant the use of observed data. Data-based elements naturally require measurements from which correlations can be inferred. Likewise, physics-based models require data for validation to ensure they accurately represent system behaviours. To develop any form of robust model for SHM, it is often necessary to use data from a wide range of damage, operational and environmental conditions. Prior to a monitoring campaign, such data are typically prohibitively expensive to obtain, or otherwise infeasible because of physical

¹Some correlations may, of course, correspond to causal relationships; however, this is famously not guaranteed. Causality can be preserved in data-based models via causal inference algorithms. For further details on this topic, the reader is directed to [4].

limitations and/or safety reasons. This scarcity of informative data poses a significant challenge in the development of SHM systems.

1.2 SHM for Decision Support

As mentioned previously, SHM technologies are highly desirable for asset management applications in a variety of high-value and safety-critical engineering sectors, such as civil, aerospace, manufacturing and nuclear. Nonetheless, until recent years, little attention had been given to decision-making and risk analyses in the context of SHM. This failure to quantify the value of SHM systems has hindered the adoption of SHM systems into industry, despite the significant technological advances [7].

Online monitoring systems are particularly useful for structures that are difficult to inspect *in situ*; for example, structures that are in remote locations, such as offshore wind farms (Figure 1.1), or satellites.



Figure 1.1: The Lillgrund Wind Farm, located off the southern coast of Sweden. Image credit: Hans Blomberg, Vattenfall, CC BY-NC-ND 2.0.

More specifically, damage information provided by monitoring systems can be used to inform, or even automate, various O&M decisions, including:

- when inspections are required;
- when maintenance is required;
- where maintenance is required;
- what maintenance is required;
- when downtime/closures are required.

Historically, such decisions have been alleviated by the adoption of O&M strategies in which systems are inspected at predetermined intervals; however, this approach can result in unnecessary inspections being made. Moreover, because of the nonlinear manner in which many types of damage progress, inspections may fail to recognise imminent structural failures. In contrast, O&M decisions supported by an SHM system can be based upon up-to-date damage information, thereby allowing for bespoke inspection and maintenance strategies. Such strategies yield numerous social and economic benefits including [8]:

- improved structural safety;
- increased availability/capacity/output;
- reduced expenditure on inspections;
- reduced cost of maintenance;
- lifetime extension.

A compelling approach by which bespoke O&M strategies can be generated, and the aforementioned benefits can be realised, is to use the damage information provided by a monitoring system to aid in the quantification of *risk*.

1.3 Risk-Based Decision-Making

The concept of risk is a useful tool for decision-making, capturing both the *likelihood* and *severity* of an adverse event. It follows that, in order to quantify the risk of an

event, one must quantify the *uncertainty* surrounding its occurrence, in addition to the *costs* associated with the consequences.

Uncertainty is pervasive throughout existence; arising as a result of incomplete information (epistemic), or inherent stochasticity (aleatoric), in fields such as physics, economics, medicine, and engineering. Uncertainty arises in the context of SHM for several reasons. As previously stated, structural damage often has to be inferred from indirect measurements; yet, even in scenarios when damage can be measured directly, systematic error and finite measurement precision can cause some degree of uncertainty, thus rendering obtainable damage information imperfect. Furthermore, decision-making in SHM often requires predictions of future damage; as the future is seldom known, the process of forecasting such information introduces further uncertainty. Structural health monitoring systems provide information that improves damage prediction twofold; both by increasing the accuracy of the predictions, and by reducing the prediction uncertainty. There are several approaches for quantifying uncertainty – the current thesis opts to use the well-established notion of *probability*.

Without context, the word ‘consequences’ can sound ominously vague. To elucidate this term, one can return to the definition of structural damage provided in Section 1.1. This definition indicates that the consequence of damage is some adverse impact on performance. It then follows that the severity of a consequence corresponds to the degree to which performance is impacted. Here, it is also worth recognising that a given structure/system may be required to perform well with respect to several (potentially competing) criteria; for example, both economic output and safety.

In general, the severity of consequences can be quantified via costs, or *utilities*. For engineering applications, it is often useful to express such costs in terms of monetary value; this is because it facilitates direct comparison to decidable actions, such as maintenance, that also tend to have monetary costs associated. A challenging aspect of using a monetary basis for the quantification of cost, is that determining exact and consistent values can be difficult, being highly variable with respect to the elicitation process. This issue holds particularly true if one attempts to assign costs directly to the damage states of a structure, which do not intrinsically have costs associated. A key tenet of the current thesis is that, where possible, costs should be assigned to performance criteria, with a mapping established between damage and performance.

As a product of probability and cost, risk can be interpreted as an *expected utility*. When adopting a risk-based approach to decision-making, asset management actions

can be selected such that risk is minimised, or expected utility is maximised.

1.4 Thesis Contribution

This thesis aims to move towards a generalised holistic framework for risk-based decision-making in the context of structural health monitoring. The modelling techniques used to establish the framework are developed to address the primary challenges, mentioned previously, associated with SHM decision support. Specifically, three core contributions to the fields of SHM and physical-asset management are presented:

- A risk-based paradigm for conducting SHM campaigns is presented. The foundation of the paradigm is a framework based upon probabilistic graphical models, which allows uncertainty to be accounted for throughout the decision-making process. The approach utilises technologies from probabilistic risk assessment (PRA), to establish mappings between local damage states of a structure and global failure criteria. Importantly, by considering specific failure modes of a structure, the framework facilitates the assignment of costs within the framework.
- Built upon the decision framework established, an algorithm for developing damage-classification models is presented. The methodology adopts an active-learning approach and is therefore applicable in scenarios where labelled data are scarce, such as SHM. The algorithm builds classification models in an online manner, using limited damage information obtained via structural inspections. As part of the algorithm, inspections are mandated according to *value of information*, thereby considering risk in the context of the SHM decision process.
- By adapting the risk-based active-learning algorithm, it is shown that decision-making performance can be improved via classifier design. In particular, discriminative classifiers and semi-supervised generative classifiers are shown to reduce the number of structural inspections made throughout a monitoring campaign, and result in an overall improvement in decision-making performance. The results are significant as they indicate that the costs of a monitoring campaign can be reduced via classifier design.

The limitations of the techniques developed are discussed. Additionally, potential avenues for future research are highlighted, in the hope that the SHM community generates further research surrounding decision-making techniques.

1.5 Chapter Summary

The outline for the remainder of this thesis is as follows:

Chapter 2 — An up-to-date summary of literature on the topic of decision-making for structural health monitoring is provided.

Chapter 3 — The established paradigm for conducting SHM campaigns is presented. In addition, probabilistic risk assessment is introduced and details of the methodology are provided.

Chapter 4 — Background theory relevant to decision-making under uncertainty is provided. Firstly, probability theory is established as an approach for reasoning under uncertainty. Subsequently, decision theory and the concept of expected utility are introduced as a manner of making decisions under uncertainty. Finally, probabilistic graphical models (PGMs), are presented as a tool for representing structured inference and decision problems.

Chapter 5 — Here, similarities and differences between SHM and PRA are highlighted and discussed. A mapping from PRA onto SHM is developed and an augmented risk-based paradigm for SHM is detailed. The mapping is founded on a hierarchical representation for structures, reminiscent of the fault-tree models utilised in PRA. This representation is used to define key variables used within PGMs for SHM decision processes.

Chapter 6 — The risk-based framework developed in the previous chapter is demonstrated via an experimental case study; specifically, a four-bay truss. An O&M decision-making process is formulated around the dataset, and the necessary models are developed with the use of a finite element model.

Chapter 7 — In this chapter, a closer look is taken at one of the key models utilised in the risk-based decision framework – the transition model. Returning to the four-bay truss case study, a probabilistic degradation model is developed for the truss by Monte Carlo sampling from a finite element model. The approach considers

multiple failure modes of the individual members comprising the structure.

Chapter 8 — Statistical classifiers, another vital component in the risk-based decision framework, are the focus of this chapter. Here, a risk-based approach to active learning is developed that seeks to overcome one of the primary challenges in SHM - data scarcity. The algorithm utilises *expected value of information* to mandate structural inspections; thus providing a methodology for risk-based inspection. Risk-based active learning is demonstrated using a numerical visual example, and using an experimental dataset; namely, the Z24 Bridge benchmark.

Chapter 9 — Further insights into the role of statistical classifiers in decision-making are provided. Here, novel algorithms for risk-based active learning are presented that seek to overcome one of the drawbacks of the original approach – sampling bias. Discriminative classifiers and semi-supervised learning algorithms are incorporated into the risk-based active learning process. These new algorithms are demonstrated again using a visual example and the Z24 Bridge benchmark.

Chapter 10 — This chapter investigates a new perspective for SHM decision-making, based upon *populations* of structures. The hierarchical representation of structures outlined in Chapter 5 is extended such that a hierarchical representation is obtained for populations of structures. Within this hierarchy, inferences and decisions are defined and related to the core technologies underpinning population-based structural health monitoring (PBSHM), such as irreducible element models, and transfer learning.

Chapter 11 — Conclusions: The consonant themes throughout the thesis are highlighted with respect to the novel technologies presented. In addition to the advantages of the proposed methodologies being highlighted, challenges with risk-based approaches to decision-making for SHM are discussed. Finally, areas for future research are offered.

LITERATURE SURVEY: RISK-BASED SHM

Thus far, much of the research in the field of SHM has focussed on the identification, localisation and classification of damage; with numerous advancements having been achieved for these tasks. Decision-theoretic approaches to SHM have received comparatively little attention from the research community, nonetheless, in recent years some progress has been made. The current chapter provides a brief review of the literature that takes a risk-based, or decision-theoretic, approach to SHM. Broadly speaking, the existing research into risk-based SHM can be assigned to one of three categories: (1) inspection and maintenance planning, (2) monitoring system optimisation, and (3) value of information quantification.

2.1 Inspection and Maintenance Planning

Research into risk-based inspection and maintenance planning aims to find methods by which cost-optimal O&M strategies can be determined.

Cost-informed decision-making for miter gates was demonstrated in [9]; this involved using a Bayesian neural network trained on a finite element model to infer damage and forecast performance using a transition matrix. An approach proposed in [10], facilitates cost-efficient reliability-based maintenance. As the latter is a reliability-based approach rather than a risk-based approach, the costs of failure events and

maintenance are not explicitly modelled. Hence, whilst the maintenance strategies developed may be cost-efficient for given safety parameters, they are not necessarily cost-optimal. In [11], expected utility theory is used as an approach for O&M decision-making for civil infrastructure from monitoring observations. In [12, 13], a heuristic risk-based approach to inspection planning is formulated around value of information. In [14], Di Francesco *et al.* use Bayesian multi-level models to partially pool information between several corrosion ‘hot spots’ on a simulated structure. This model is subsequently used to compute expected value of information within a risk-based inspection setting.

Probabilistic graphical models have been applied several times to risk-based inspection and maintenance planning. In particular, dynamic Bayesian networks and influence diagrams for various Markov decision processes have received the most attention. Dynamic Bayesian networks are also employed in [15] for the diagnosis and prognosis of the structural health of an aircraft wing; this includes the probabilistic temporal modelling and prediction of crack growth. In [16], partially-observable Markov decision processes (POMDPs) are used in conjunction with heuristics to find optimal inspection strategies for deteriorating structural systems. In [17], Nielsen details a risk-based approach for the operation and maintenance of off-shore wind turbines, the approach is furthered using probabilistic graphical modelling of POMDPs in [18, 19]. Similarly, Hovgaard and Brincker provide a case study demonstrating a risk-based approach to the monitoring and maintenance of a finite element model of a wind turbine tower experiencing circumferential cracking in [20].

In [21], a continuous-state POMDP was demonstrated on artificial data for maintenance planning via a deteriorating bridge case study. A comparison of grid-based and point-based POMDP solvers with respect to structural inspection and maintenance planning is presented in [22] – favouring the latter for their ability to efficiently solve large decision problems, point-based POMDPs are applied to various structure management case studies in [23–27]. Finally, POMDPs are used in conjunction with deep reinforcement learning in [28]. In doing so, inspection and maintenance planning are accomplished under budgetary constraints.

2.2 Value of Information Quantification

Research into value of information quantification seeks to develop methods that can quantitatively express the value that SHM systems contribute to inspection and maintenance planning, in terms of expected utility.

The value of information provided by monitoring systems is introduced in [29] and [30]. The method presented in the former, based upon decision theory, is demonstrated using a pedestrian footbridge as a case study. Methods for computing the life-cycle value, and the initial investment payback period for monitoring approaches supporting O&M decisions are detailed. The latter provides a general framework for quantifying the value of SHM information derived from Bayesian decision analysis and comprehensively details the influence of structural system characteristics on the value of information.

Memarzadeh and Pozzi present an approach for assessing value of information in sequential decision problems modelled via POMDPs in [31]. They show that value of information at the component level can be used to simplify system-level decision-making, at the expense of global optimality. Similarly, in [32], value of monitoring information is considered for time-dependent decision processes for the purpose of adaptive risk-based inspections.

Several studies have been conducted into the value of monitoring systems for supporting the management of bridge structures. In [33], a comparison is made between the value of information provided by scour monitoring systems and vibration-based SHM systems for mitigating failures of bridges subject to flood conditions. Bayesian methodologies for quantifying the benefit of an SHM system for roadway bridges with seismic hazards are presented in [34]. In [35], a similar Bayesian decision analysis approach is used to quantify the value of monitoring information for a bridge degrading as a result of corrosion and seismic effects.

Giordano *et al.* examine the value of seismic structural health monitoring information for various decision scenarios for a building existing in a seismically active zone in [36].

2.3 SHM System Optimisation

Research into risk-based SHM system optimisation is concerned with finding expected utility maximal or risk minimal arrangements for monitoring systems; this includes aspects such as sensor types, sensor locations, and data acquisition rates.

Flynn and Todd successfully applied a Bayes-risk approach to the decision problem of active-sensor placement for an SHM system on square, gusset and T-shaped plates in [37]. The approach considered the risk of false positives and false negatives of damage identification in discrete regions of the plates. In [38], Pozzi and Der Kiureghian propose value of information as a means to rank competing measurement systems such that those that yield the greatest benefit to the decision-making process can be selected.

Spatial distributions of sensors are optimised with respect to value of information in [39]. This task is accomplished for Gaussian random fields by constructing objective functions representative of the decision process that the monitoring system supports. Sensor quantities and positions are selected such that the value of information acquired from the random field is maximised. In [40], value of information is used to select quantities of interest to be measured by a monitoring system. Finally, Long *et al.* use value of information to optimise several key characteristics of a monitoring system, including the number of sensors, the location of sensors and the noise characteristics of the sensors [41].

BACKGROUND

3.1 Structural Health Monitoring

Structural health monitoring involves implementing damage identification strategies to determine the damage, or health state, of a structure throughout its operational lifetime. Statistical pattern recognition (SPR) offers a natural approach to SHM as it allows the associated uncertainties mentioned in Chapter 1 to be dealt with robustly. It is for this reason that the SPR approach has been the focus of much research over the past few decades. The established SPR paradigm for an SHM campaign is composed of four procedures [1]:

1. Operational Evaluation.
2. Data acquisition.
3. Feature selection.
4. Statistical modelling for feature discrimination.

3.1.1 Operational Evaluation

Operational evaluation seeks to answer several questions concerning the implementation of an SHM system, specifically:

- What is the justification (safety and/or economic) for implementing an SHM system?
- How is damage defined for the system and what are the critical damage states?
- What are the environmental and operational conditions that the monitoring system is required to perform under?
- How does the operational environment limit data acquisition?

For an SHM system to be successfully developed and implemented, a substantial amount of information must be collected during the operational evaluation process. Examples of required information include: monetary cost and reliability of the proposed SHM system; pertinent damage states of the structure and the thresholds at which they can be deemed to have occurred, and the expected temperature and load variations to be experienced by the structure during operation.

3.1.2 Data Acquisition

The data acquisition process is informed by the operational evaluation. The process aims to finalise the types, number and locations of sensors to be used in the SHM system. The data acquisition, storage and transmittal hardware must also be selected. The process is typically constrained by both economic restrictions and the limitations enforced by the expected environmental conditions; for this reason, the data acquisition process is context dependent and relies heavily on the information gathered during the operational evaluation stage.

3.1.3 Feature Selection

Once the data have been acquired, a set of features must be constructed that indicate whether or not there is damage present in the structure. Here, a damage sensitive feature refers to a quantity that can be extracted from measurable data that varies between structural health states of interest¹. This procedure often involves

¹The sensitivity of a given feature can be evaluated with respect to Rytter's hierarchy. For example, the first few natural frequencies of a structure (global properties of a system) can be a good indicator of the presence of damage; however, may provide little information regarding the location of damage.

processing the data acquired from the structure; common practices include domain transformation, dimensionality reduction, and normalisation [1].

3.1.4 Statistical Modelling for Feature Discrimination

Statistical models must be developed to exploit the discrepancy in the features such that differing damage states can be identified. The degree of knowledge regarding the damage state obtained from an SHM system is highly dependent on the statistical model employed and can be evaluated in terms of Rytter's Hierarchy. As mentioned, as SHM systems progress up the hierarchy, the information they yield becomes increasingly useful to agents tasked with deciding upon a course of action for a structure.

3.2 Probabilistic Risk Assessment

Probabilistic risk assessment (PRA) is a method that is widely used for evaluating risks and making decisions associated with the design and management of safety-critical systems and high-value assets. In the context of PRA, risk is characterised by the likelihood of an adverse event occurring, and by the severity of the consequences of the event. The likelihood of occurrence for uncertain adverse events is quantified via probabilistic event-sequences and system modelling. The consequences and expected costs/gains are compared and evaluated by finding an appropriate utility metric - obvious examples include financial cost and loss of human life; however, in many applications these can be overly simplistic [42]. Probabilistic risk assessment is applied in a range of industries including nuclear [43–45], aerospace [46, 47] and chemical process [48]. Whilst the exact methodology used for conducting PRA differs between industries, they generally adhere to the key steps as outlined by the US Nuclear Regulatory Commission (USNRC) and the International Atomic Energy Agency (IAEA) [49]:

1. Initial information collection.
2. Event-tree development.
3. System modelling.

4. Reliability modelling.
5. Failure-sequence quantification.
6. Consequence analysis.

Here, some historical background is provided before each key step in the PRA methodology is detailed further.

3.2.1 The History of PRA

In ancient Mesopotamia, around 3200 BCE, priest-like figures known as *ašipu* were consulted to provide guidance for important decisions such as marriage arrangements and selecting building sites [50]. To aid in making such decisions, the *ašipu* would identify courses of action and use ‘data’ to assess the possible outcomes, such as profit or losses, for each alternative. At the time, the state-of-the-art method for obtaining ‘data’ was to divine signs from the deities. Nonetheless, findings were recorded in a grid, etched into a clay tablet, and presented to the client [51]. It is said that the *ašipu* provide the first recorded version of a primitive form of risk analysis [50].

Today, more modern forms of risk analyses are used in a variety of disciplines including epidemiology, finance, and politics. The probabilistic risk assessment methodology used in modern engineering contexts has its roots in the aerospace sector [42]. In the January of 1967, three astronauts were killed after a fire destroyed the command module during an Apollo 1 launch rehearsal test. In addition to the deaths, the disaster cost the National Aeronautics and Space Administration (NASA) millions of dollars and public support. As a result, an overhaul of NASA’s safety policies was initiated; beginning with quantitative goals specifying probabilities for mission success and injury per mission.

Despite the severe consequences of the 1967 disaster, quantitative risk analyses faced opposition from NASA management, as estimates of failure probabilities were considered to undermine the political viability of the space program. A 1983 risk analysis found the failure probability of solid rocket boosters in the Challenger Space Shuttle to be as high as 1 in 35; this value was rejected by management in favour of a figure based on engineering judgement in the region of 1 in 100,000 [52].

Sadly, in the April of 1986, the Challenger Space Shuttle broke apart shortly after launch as a result of gasket failures in a solid rocket booster. The disaster resulted in seven fatalities. After a congressional report concluded that NASA could not effectively direct resources without quantitative analyses identifying probabilities of failure, risk assessments were broadly adopted to support safety during the design and operations phases of manned space-travel projects. In 2003, the Space Shuttle Columbia disintegrated upon re-entry to the atmosphere, again resulting in seven fatalities. The investigative board found that the flaws in the risk assessment; specifically, that the consequences of the type of damage responsible for the failure were significantly under estimated as a result of organisational deficiencies.

Although the basic risk assessment methodologies originated in the aerospace sector, the first modern PRA was conducted and published by the United States Nuclear Regulatory Commission (USNRC) in 1975 [42]. In addition to the quantification of failure probabilities, the *Reactor Safety Study* also presented analysis of the consequences of failure, quantifying the costs associated with the release of radioactive material into the biosphere [53]. Although the fundamental approach to PRA was considered valid, the study was criticised for mishandling probabilities. Subsequently, the USNRC distanced itself from the report. In 1979, a core-damage accident occurred at the Three Mile Island Nuclear Generating Station (TMI) in Pennsylvania, USA. The accident resulted in the release of radioactive krypton-85 and iodine-131 into the environment. Fortunately, studies have deemed there to have been minimal health implications as a result of the release. Investigations into the event revealed that the *Reactor Safety Study* predicted the loss-of-coolant events, such as the one that occurred at TMI, as being a significant threat to safety. As a result, it was recommended that probabilistic analyses should be utilised more for assessing the risks associated with nuclear power plants. The recommendation spawned a series of publications that sought to rectify the deficiencies of the *Reactor Safety Study* and standardise the risk-assessment methodology. Notably, the 1983 USNRC *Procedures Guide* outlined the key steps involved in PRA, presented above [49].

The years between 1983 and present day were not without nuclear accidents. Most significant, were the accident at the Chernobyl Nuclear Power Plant that occurred in 1986 in the former Soviet Union (now Ukraine), and the accident at the Fukushima Daiichi Nuclear Power Plant that occurred in 2011 in Japan. Subsequent investigations have partially attributed both events to management neglecting information provided in risk assessments [54].

The experiences in the aerospace and nuclear sectors over the past 50 years highlight that probabilistic risk assessment provides a valuable tool for decision support, but also indicates that the results of such analyses must be carefully interpreted as *subjective estimations* of risk, and scrutinised and evaluated with disregard for politics. The following subsections outline the key steps involved in conducting PRAs as presented by the USNRC in [49].

3.2.2 Initial Information Collection

The first step in the PRA methodology relates to the acquisition of contextual information. Specifically, information regarding the design and operation of the structure in question is collated. Details such as component specifications, loading and environmental conditions are considered, such that information gathered at this stage can be used to inform the subsequent steps. Given the large quantity of information required for conducting PRA, an important factor to be considered at this stage is the method by which the necessary information is represented, stored and managed. A common practice is to utilise a database [42]. There is a clear analogy here with the operational evaluation stage for SHM.

3.2.3 Event-tree Development

Event trees outline potential accident sequences - combinations of initiating events and the subsequent system failures or successes that may result in an adverse consequence. The sequences of system failures and successes are known as *top events*. The system failures identified in the event-tree development stage are subsequently modelled as fault trees. An example event tree for a system designed to prevent injury following a jump from a plane is shown in Figure 3.1.

3.2.4 Fault-tree Development

Fault trees are used in PRA to facilitate the quantification of system failure probabilities. The development of fault trees involves expressing the causal relationships between component failures and subsystem failures using Boolean logic gates. The level of detail captured in the fault tree (the level of components which are incorpo-

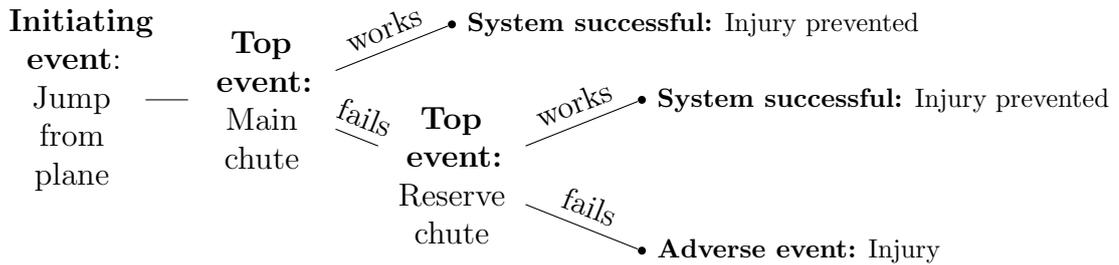


Figure 3.1: An example event tree for a parachute system to prevent fall injuries [55].

rated) is determined by the component level for which meaningful reliability data can be obtained. Failures of components belonging to the most fundamental level incorporated in the fault tree are known as *basic events* and are represented in a fault-tree diagram as circles. Intermediate and top events are defined as combinations of other intermediate, and basic events through Boolean logic gates such as the AND-gate and OR-gate. The fault-tree diagram notation for the AND-gate and OR-gate are shown in Figure 3.2. An example fault tree for the deployment of a reserve parachute is shown in Figure 3.3.

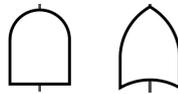


Figure 3.2: Fault-tree diagram representations of the AND-gate (left) and OR-gate (right).

3.2.5 Reliability Modelling

Reliability is defined as the probability that a system, or component, is functioning at a given time. It follows that the complement of reliability provides information regarding the probability that a system or component is failed at a certain time. Determining the reliability of components in a system/structure is vital in PRA such that failure probabilities can be propagated through the fault tree in order to determine the probability of top events. Typically, information regarding the reliability of system components and the frequency of initiating events is gleaned from analysis of the physics of failure or by applying appropriate reliability models to data. Examples of such approaches include the first- and second-order reliability methods, and surrogate-modelling methods [56–59].

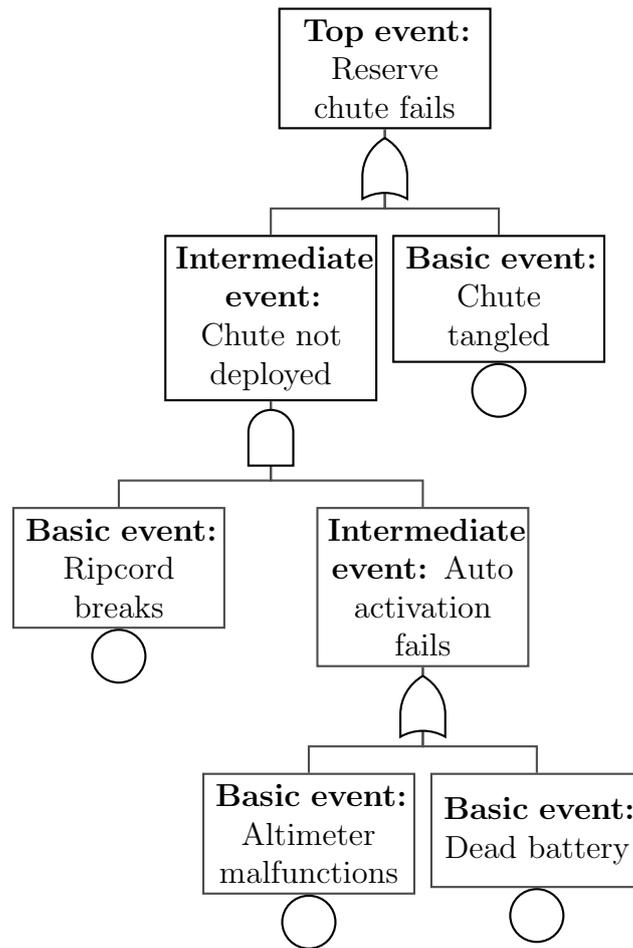


Figure 3.3: An example fault tree for a reserve parachute system [55].

3.2.6 Failure Sequence Quantification

By assigning the components in the fault trees with reliabilities/failure rates, the probability of top events may be computed. Propagating the probabilities of the initiating events and top events through the event tree allows the probability of each possible outcome in the event tree to be calculated.

3.2.7 Consequence Analysis

Consequence analysis involves determining the effects of system failures/adverse events and assigning representative costs such that risk can be assessed. The risk assessment may then be used to inform design decisions, such as increasing safety by

introducing additional redundancies in the system, or optimising cost by removing components that do not cause the risk to fall below an acceptable threshold. The risk assessment may also be used to inform risk-based inspection for a system that is in operation. In some applications, consequences analysis is an extremely involved process. For example, the consequence analysis for a loss of containment accident at a nuclear facility requires modelling of radionuclide transport through the environment and assessing the impact it has on populations.

3.3 Summary

In this chapter, the established paradigms for conducting SHM campaigns and PRAs were presented. In both cases, a multi-step procedure is undertaken. In Chapter 5, similarities and differences between the paradigms are identified and useful aspects of PRA are incorporated into a risk-based framework for SHM. Before a framework is established, the theoretical basis for decision-making in SHM is provided in Chapter 4.

DECISION-MAKING UNDER UNCERTAINTY

It was stated previously that, within SHM, there are various sources of uncertainty; arising because of imperfect damage information and via predictions of future events. In this context, the term ‘uncertainty’ refers to one’s knowledge, or rather lack thereof, regarding the truth conditions of a declarative statement. In order to make decisions for structural-asset management, engineers may wish to reason about declarative statements such as: ‘There is damage in this aeroplane wing’ or ‘The bridge will collapse next week’. There are several approaches for representing uncertainty such as intervals [60], and fuzzy logic [61]. The current thesis opts to use *probability theory*. One advantage of using probability, is that, built upon the theory, there exists a well-established field in mathematics for decision-making known as *decision theory*. The current chapter presents the fundamental concepts from probability theory and decision theory used in risk-based decision-making. In addition, *probabilistic graphical models* (PGMs) are introduced as a powerful tool for reasoning and decision-making under uncertainty.

4.1 Probability Theory

Probability theory is now founded upon Kolmogorov’s axioms [62], the in-depth details of which are not presented here but can be found in [42, 63, 64]. Instead, a

short overview of the foundations of probability theory are presented in Appendix A.1. Conceptually, it is important to understand that probabilities are bounded by the interval $[0, 1]$. The probability that an event X holds true is denoted by $P(X)$, with $P(X) = 1$ implying that X certainly *does* hold true, and $P(X) = 0$ implying it certainly *does not* hold true. It then follows that when $0 < P(X) < 1$, it is implied that there is some degree of uncertainty about whether X holds true.

In many applications, including engineering and SHM, it is useful to consider events as *random variables* taking certain values, or being in certain states, within a given domain. Here, the domain of the random variable X is denoted as \mathcal{X} . If X is a discrete random variable then $\mathcal{X} = \{1, \dots, N\}$ and,

$$P(X = x) \forall x \in \mathcal{X} \quad (4.1)$$

where $P(X = x)$ is the *probability mass function* (pmf) for the variable X . An important property of the pmf, known as the normalisation condition, is as follows,

$$\sum_{x \in \mathcal{X}} P(X = x) = 1. \quad (4.2)$$

If X is a continuous random variable, then $\mathcal{X} \in \mathbb{R}$, and as such, rather than a pmf, it will have an associated *probability density function* (pdf) $p(x)$ such that,

$$P(a < X \leq b) = \int_a^b p(x) dx \quad (4.3)$$

and,

$$\int_{\mathcal{X}} p(x) dx = 1. \quad (4.4)$$

The basic rules required to manipulate probabilities are provided in Appendix A.2. From these rules, one can arrive at the notion of a *conditional* probability $P(X|Y)$, and Bayes' theorem. Bayes' Theorem is given by the expression,

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}. \quad (4.5)$$

The conditional probability $P(X|Y)$ denotes the probability of X occurring given that Y occurs, thereby implying a *dependence* between the variables X and Y . Further insight into Bayes' theorem and the Bayesian interpretation of probability is provided in Appendix A.3. Here, it is worth acknowledging that Bayesian probabilities do not explicitly differentiate between epistemic and aleatoric uncertainty. In the Bayesian setting, it is typically assumed that as a belief is updated using evidence, any reduction in variance can be attributed to epistemic effects. It then follows that, in the limit of infinite data, it can be assumed that any remaining variance captured by a model is associated with aleatoric uncertainties.

4.2 Decision Theory

Whereas probability theory provides a formal approach to reasoning under uncertainty, decision theory provides a formal approach to decision-making under uncertainty. According to decision theory, decision-making can be formulated as an expression of preferences over a set of actions \mathcal{A} .

4.2.1 Fundamentals of Decision Theory

The axioms underpinning decision theory (detailed in [42, 65, 66]) define the notion of a 'rational' decision-maker. Simply put, a decision-maker is said to be rational if they act so as to maximise their *expected utility*. More formally, the von Neumann-Morgenstern theorem states that, for two decidable actions $a, b \in \mathcal{A}$,

$$a \succeq b \iff \text{EU}(a) \geq \text{EU}(b) \quad (4.6)$$

where \succeq denotes a weak preference, indicating that a decision-maker favours a at least as much as b ; and $\text{EU}(\cdot)$ denotes the expected utility associated with doing an action.

In [67], von Neumann and Morgenstern derive the expected utility. Consider a stochastic event X , the mutually-exclusive outcomes of which $x \in \mathcal{X}$ are conditionally dependent on a decision D between actions a and b . The expected utility of action a is computed as follows,

$$\text{EU}(a) = \sum_{x \in \mathcal{X}} P(X = x | D = a) \cdot U(X = x, D = a) \quad (4.7)$$

Here, $P(X|D = a)$ represents the uncertainty in the outcome of X given that action a is executed. U denotes a utility function that provides a mapping, $U : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$. From equation (4.7), one can realise that expected utility has an equivalence to risk. Here, it is worth noting that $U(X, D)$ may be additively separable and thus may be conveniently expressed as the sum of two utility functions $U(X)$ and $U(D)$, that respectively describe the utilities associated with outcomes and actions independently. In such cases, equation (4.7) can be rewritten as,

$$\text{EU}(a) = \left[\sum_{x \in \mathcal{X}} P(X = x | D = a) \cdot U(X = x) \right] + U(D = a) \quad (4.8)$$

For a single decision D over a finite set of actions \mathcal{A} , an *optimal* action a^* can be defined such that the maximum expected utility (MEU) is achieved, where,

$$\text{MEU}(D) = \max_{a \in \mathcal{A}} \text{EU}(a) \quad (4.9)$$

and,

$$a^* = \arg \max_{a \in \mathcal{A}} \text{EU}(a) \quad (4.10)$$

The utility theorem, as originally presented by von Neumann and Morgenstern, assumes that a decision-maker has an objective quantification for both the probabilities and utility functions necessary for assessing expected utility. In [68], Savage extends the von Neumann-Morgenstern theorem such that it applies in scenarios where the decision-maker does not know precisely the required probabilities and utilities, but instead must rely on subjective quantifications. Although the introduction of subjective probabilities and utilities does not change the form of equations (4.6) and (4.7), they must be interpreted in a slightly different way. Savage's theorem states that, in the absence of objective quantifications of probabilities and utilities, a rational decision-maker must act so as to maximise their *perceived* expected utility; as if holding beliefs regarding the probabilities and utilities of outcomes. The theorem

implies that an agent's subjective expected utilities can be inferred by observing preferred actions.

In general, the work presented in the current thesis will assume that utility functions are known but that precise probability distributions are unknown, and subjective quantifications must be obtained via Bayes' Theorem (equation (4.5)).

4.2.2 Limitations

Utility-based decision theory has several well-documented limitations.

It is proven that, for a given decision problem, the preferable strategy remains unchanged for positive affine transformations of the utility function. This characteristic means that one cannot make absolute comparisons between the utilities associated with specific actions and outcomes. As an example, consider two possible outcomes, x_1 with $U(X = x_1) = 10$, and x_2 with $U(X = x_2) = 20$. One cannot say that x_2 is two times more preferable than x_1 since an agent with a utility function $U'(X) = U(X) + 5$ – a positive affine transformation of $U(X)$ – would act indistinguishably from an agent using the original utility function $U(X)$. In a similar vein, as utility functions are only defined up to an affine transformation and, in the case of Savage's theorem, may be subjective, there is no canonical way to make comparisons between the utility functions of multiple different agents. This limitation means that quantities, such as average expected utility across a population of agents, are ill-defined without some prior assumptions.

Here, it is worth acknowledging that there are several well-known examples demonstrating that, in reality, humans do not always behave rationally in the von Neumann-Morgenstern or Savage sense. The St. Petersburg paradox, originally conceived by Nicolas Bernoulli, is presented as a game of chance with infinite expected utility; a fair coin is tossed until a tails occurs, then a prize with utility 2^k is awarded where k is the number of consecutive heads. A von Neumann-Morgenstern rational agent would be willing to pay an arbitrarily large (but still finite) sum to play such a game; however, in practice it is observed that people hold upper limits to the amounts they would pay. The paradox highlights some of the psychological and sociological aspects that influence decision-making; in particular, the notion of *diminishing marginal utility* [69]. The law of diminishing marginal utility is well-established in economics and simply states that each unit of a finite resource accumulated has an ever-decreasing

subjective value when compared to the previous unit. This effect means agents with more limited resource availability are less inclined to risk what resource they possess, even for gambles with favourable expected utilities.

Other related phenomena are demonstrated by the Allais paradox and the Ellsberg paradox, the details of which can be found in [70] and [71], respectively. The Allais paradox provides an example of the *certainty effect* [72], which states that in some scenarios people prefer guaranteed, or zero-variance, outcomes as opposed to more variable outcomes, even when the more variable outcomes hold a higher expected utility. Similarly, the Ellsberg paradox provides evidence for *ambiguity aversion*. The theory of ambiguity aversion suggests that people prefer to take actions where the consequences are known over actions where there is some ambiguity surrounding the them, even when the actions with ambiguous consequences are perceived to yield higher expected utility. These phenomena suggest that certainty has some psychological value that is not captured by the von Neumann-Morgenstern and Savage formulations of expected utility theory.

Despite the limitations, decision theory provides a rigorous and formal methodology for selecting optimal actions under uncertainty and as such it is a valuable tool for physical-asset management and SHM. Further discussions surrounding the limitations, and in particular, the ethical implications that arise when attributing utilities to outcomes in the context of SHM, are left as future work.

4.3 Probabilistic Graphical Models

Many of the core concepts in probability theory and decision theory are elegantly captured by *probabilistic graphical models* (PGMs) [73–75]. Fundamentally, PGMs are graphical representations of joint probability distributions, in which the nodes of the graph represent random variables and the edges connecting nodes imply dependencies between variables. Probabilistic graphical models provide two key benefits over the flat (non-graphical) representation of joint distributions [74]:

- They provide a compact and intuitive representation of probability distributions, which makes complex distributions more easily interpreted, understood and communicated.

- They facilitate efficient computation by exploiting local independence structures within the graphs.

A PGM over a set of N variables X may be specified by a set of M local functions $f(Y_i)$ where $Y_i \subseteq X$, and a graph G comprised of nodes/vertices V and edges E . The joint probability distribution over the variables in X is then given by,

$$P(X_1, \dots, X_N) = K \prod_{i=1}^K f(Y_i) \quad (4.11)$$

where K is a normalisation factor ensuring that the conditions given in equations (4.2) and (4.4) hold.

There are two classes of problem associated with traditional PGMs: *inference* and *learning*. Inference is concerned with obtaining marginal or conditional probabilities for a subset of variables Z , given observed values for another subset of variables Z_{obs} , i.e. $P(Z|Z_{obs})$. Learning is concerned with obtaining the graph structure and parameters given some complete, or incomplete, set of observations over the variables in X , i.e. $G, f(Y_i)|Z_{obs}$. For the graphical models used in the current thesis, it is assumed that the structure of the graphs is known; therefore, the problems presented are centred around inference, or learning parameters that specify local functions.

4.3.1 Bayesian Networks

Bayesian networks (BNs) are a form of PGM. Specifically, they are directed acyclic graphs (DAGs) in which nodes represent random variables and edges connecting nodes represent conditional dependencies between variables. For discrete random variables, the local functions (i.e. the pmfs) that describe the conditional probability distributions between variables are conditional probability tables (CPTs). In the case of continuous random variables, the local functions are specified by conditional pdfs.

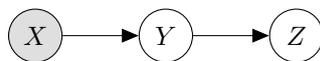


Figure 4.1: An example Bayesian network.

Figure 4.1 shows a simple Bayesian network where X is a *parent* of Y and an *ancestor*

of Z ; Z is said to be the *child* of Y and a *descendant* of X . Node X is independent of other nodes and so is specified by the unconditional distribution $P(X)$. A node that has children but no parents may sometimes be referred to as a *root*. Similarly, a node that has parents but no children may sometimes be referred to as a *leaf*. Observed variables are shaded grey.

Given observations on a subset of nodes in a BN, inference algorithms can be applied to obtain posterior distributions over the unobserved random variables. In some cases, analytical solutions of posterior distributions may be found by using exact inference methods. In general, inference in BNs is *NP-hard* [76]. Fortunately, algorithms have been developed that allow efficient computation, such as Pearl’s belief propagation algorithm for polytree-structured BNs [77]. When finding analytical posterior distributions is an intractable problem, approximate solutions can be found via stochastic simulation methods such as Monte Carlo sampling [78–80]. Efficient inference algorithms for Bayesian networks typically exploit local independence structures – an overview of these structures is provided in Appendix A.4.

Learning the local functions of a BN amounts to learning conditional probability distributions from observed data which comprises an extensive field bridging statistics and computer science known as *probabilistic machine learning*. Comprehensive introductions to the subject are provided in [63, 81]. For the current thesis, specific machine-learning algorithms are introduced when required.

4.3.2 Dynamic Bayesian Networks

Dynamic Bayesian networks (DBNs) are an extension of Bayesian networks that allow for convenient modelling of temporal stochastic processes [82]. As with standard BNs, nodes within DBNs represent the states of system at discrete instances in time, or ‘time-slices’. DBNs differ in that they also incorporate a probabilistic model that defines transitions between states from time-slice to time-slice. Edges between nodes that exist in the same time slice are known as *intra-time-slice* edges, and edges between nodes that exist in different time-slices are known as *inter-time-slice* edges.

In practice, two simplifying assumptions are commonly made about DBNs [75]. Firstly, it is assumed that the Markov property holds. The property arises as a result of the independence that is induced in Bayesian networks when a variable centred within a serial connection is observed. The Markov property asserts that

future states of a stochastic process are conditionally independent of past states, given the present states. Secondly, it is often assumed that a process is stationary, meaning that the structure and parameters of the model do not change over time. The simplest DBN, a Markov chain, follows both of these assumptions. A slight modification to the Bayesian network shown in Figure 4.1 yields the Markov chain shown in Figure 4.2.

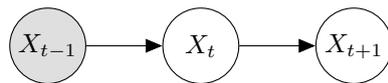


Figure 4.2: A simple Markov chain for the variable X over three time-slices.

In many applications – with SHM as an apt example – one may be unable to directly observe the states of interest for a system, and instead, these states must be inferred from indirect observations of the system. Systems that possess hidden, or *latent* states are said to be *partially observable*. One can extend the Markov chain model such that it accounts for partial observability in order to arrive at the hidden Markov model (HMM). HMM-type models are ubiquitous, allowing probabilistic modelling of time-series, and other sequential datasets, in fields such as natural language processing and economics. A basic HMM is shown in Figure 4.3.

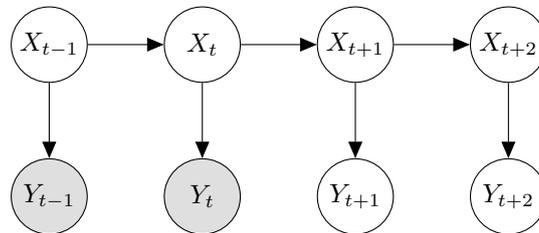


Figure 4.3: A hidden Markov model showing four time-slices.

In a given time-slice, there are two nodes in a basic HMM; a *state* variable and an *observation* variable. In Figure 4.3, it can be seen that the observation Y_t is only dependent on the latent state X_t ; this assumption implies a *generative* model, i.e. the observations are generated according to the state of the system.

Inference problems in DBNs fall into one of four categories [83]:

- *Filtering* – the prediction of the current state based on observations up to the current time, i.e. $P(X_t|Y_{1:t})$.
- *Prediction* – the prediction a future state based on observations up to the current time, i.e. $P(X_{t+n}|Y_{1:t})$.

- *Smoothing* – the prediction of the current state based on past and future observations, i.e. $P(X_t|Y_{1:T})$.
- *Decoding* – the estimation of the most likely sequence of latent states based on past and future observations, i.e. $\arg \max_X P(X_{1:T}|Y_{1:T})$.

Here, the subscript notation $1 : T$ is used to denote all instances of a variable up from time 1 to T , with $1 \leq t \leq T$.

DBNs provide a suitable framework for reasoning under uncertainty, accounting for both the partially observable and the temporal nature of the problems that arise in SHM. To consider the decision theoretic aspect of PGMs, one can turn to *influence diagrams*.

4.3.3 Influence Diagrams

Influence diagrams (IDs) are a form of augmented Bayesian network. IDs draw a distinction between states that are *observable* and actions that are *decidable*. As such, IDs incorporate an ancillary node-type representing decisions, denoted as squares. In addition to decision nodes, influence diagrams incorporate utility nodes, denoted in the graph by rhombi. An example influence diagram is provided in Figure 4.4.

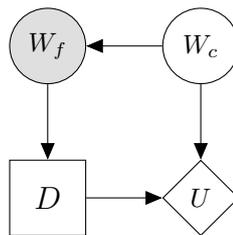


Figure 4.4: An example influence diagram representing the decision D of whether to go outside or stay in under uncertainty in the future weather condition W_c given an observed forecast W_f . Preferences for the various combinations of actions and weather conditions are specified by U .

In influence diagrams, edges between random variable nodes still indicate conditional dependence; however, other edge-types are also possible and carry different meaning. An edge from a decision node to a random variable node indicates that the random variable has some conditional dependence on the the decidable actions of the decision. An edge from a decision or random variable node to another decision node is known

as an *informational link*, and it denotes that the random variable is observed, or the former decision is made, prior to the latter decision being made. Utility nodes can receive incoming edges from both random variable nodes and decision nodes, indicating that the utility function associated with the utility node is dependent on the random variable states and decidable actions, respectively.

Whereas the BN forms of PGMs represent factorisations of joint probability distributions, influence diagrams essentially represent factorisations of expected utilities. Recalling from Section 4.2 that one can define optimal actions with respect to expected utility, one can also find optimal actions for decisions within an influence diagram \mathcal{I} . In general, influence diagrams may contain multiple decision nodes for which optimal actions must be selected. The problem of inference in an influence diagram containing n decision nodes can therefore be summarised as determining an optimal strategy $\mathbf{a}^* = \{a_1^*, \dots, a_n^*\}$ that, when executed, yields $\text{MEU}(\mathcal{I})$. If the i^{th} decision node receives informational links from variables or decisions that are not yet determined, then a_i^* can be interpreted as an optimal *policy* that maps from all possible states for the undetermined variables/decisions to possible actions. Here, it is worth noting that any given $a_i^* \in \mathbf{a}^*$ is not necessarily optimal in isolation, but can be only considered so with respect to the other actions or policies in the optimal strategy \mathbf{a}^* .

Compared to Bayesian networks, inference in influence diagrams is somewhat complicated by the presence of interventions. In scenarios where the analytical computation of all expected utilities across a decision-space is intractable, approximate optimisation algorithms may be employed. Typically, these algorithms involve searching the decision-space for strategies that yield increasingly high expected utilities. Examples include the *single policy updating* (SPU) algorithm [84] and the Monte-Carlo tree search algorithm [85]. Details of inference algorithms for influence diagrams are provided in [73, 75].

4.3.4 Markov Decision Processes

As mentioned earlier in the current section, BNs can be extended to DBNs in order to model temporal processes. In much the same way, influence diagrams can be extended to model temporal decision processes; such models are known as *Markov decision processes* (MDPs). Of particular interest here, are *partially-observable*

Markov decision processes (POMDPs) – the influence diagram counterpart to the HMM. The influence diagram representation of a POMDP is shown in Figure 4.5.

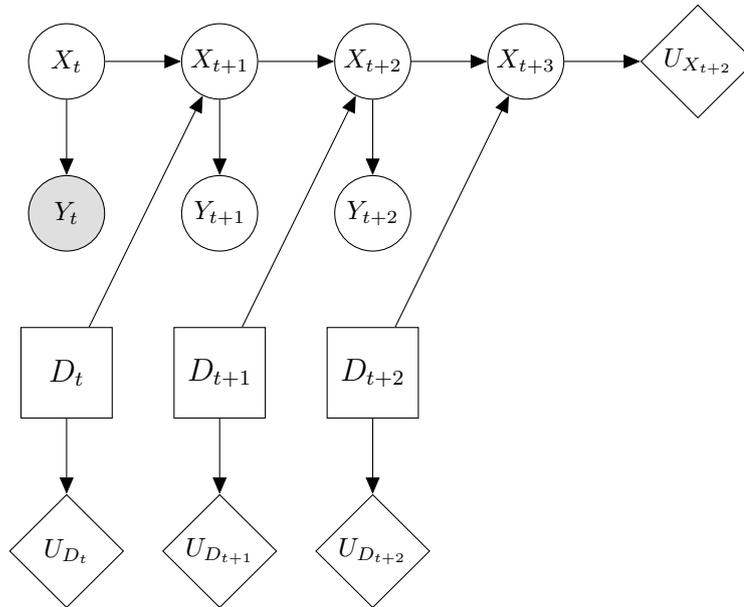


Figure 4.5: A partially observable Markov decision process showing four time-slices.

It can be seen from the ID shown in Figure 4.5, that a latent state X_t , is only dependent on the state and decision in the previous time-slice.

An important distinction between static decision problems and temporal decision problems is that the latter require a *horizon* to be specified. The horizon of a decision problem specifies how far into the future an agent considers when selecting strategies. In general, there are three types of horizon; *finite horizon*, *infinite horizon*, and *indefinite horizon*. For finite horizon problems, the number of time-slices over which the expected utility is to be maximised is known and finite; for infinite horizon problem, the number of time-slices is known to be infinite; and for indefinite horizon problems the number of time-slices is finite but undetermined.

In the field of SHM, structures typically have predefined operational lifetimes and, as such, related decision processes tend to be finite horizon. In some applications, for example indefinite lifetime extension, it may be necessary to consider indefinite horizon decision processes. The current thesis will generally focus on finite horizon decision problems.

4.4 Summary

To summarise, together, probability theory and decision theory provide a rich foundation for representing the O&M decision processes that arise in the context of structural health monitoring. Importantly, the theories allow for the quantification of uncertainty and risk, whilst also providing axioms and criteria by which optimal actions can be selected. Probabilistic graphical models provide a powerful tool for representing probability distributions and expected utilities; they have two key benefits with respect to decision-making for SHM. Firstly, they facilitate efficient computation of posterior risks. Secondly, they allow (often complex) decision processes to be communicated effectively and clearly – a benefit the importance of which cannot be overstated for engineering applications.

MAPPING PRA ONTO SHM

Risk-informed operation and maintenance strategies have the potential to reduce costs and improve safety of structures and infrastructure. To determine optimal strategies for the operation and maintenance of the subject of a monitoring campaign, agents require access to information about the likely current and future health states of the structure of interest. These engineering structures may come in a variety of complexities, ranging from those comprised of just a few parts to those comprised of thousands. Because of the potential complexity of the structures which must be operated and maintained, and of the associated decision problems, a framework for SHM decision-making demands a thorough and systematic approach akin to PRA.

The current chapter aims to address the lack of a generalised framework for conducting risk-based monitoring of structures at the full-system scale by augmenting the current SHM paradigm with practices employed in probabilistic risk assessment; thereby facilitating the decision-making processes that motivate the implementation of SHM systems. In the previous chapters, current paradigms for conducting PRA and SHM were outlined, in addition to background theory regarding the key technologies required for mapping PRA onto SHM; namely, probabilistic graphical models in the form of Bayesian networks and influence diagrams. In the current chapter, a notation is established before an augmented risk-based paradigm for SHM is detailed. Finally, a discussion around the framework is made and further challenges in the SHM decision-process are identified.

5.1 Definitions

To establish a framework for mapping PRA onto SHM, some fundamental concepts will first be defined. In addition, a notation will be established for describing structures that can be expressed as hierarchical graphs. This notation facilitates the generalisation of the approach to any structure that can be represented as a hierarchical graph.

One begins with the full structure of interest \mathcal{S} . It is assumed that \mathcal{S} may be decomposed into a discrete number of constituent elements, which are referred to as *substructures*. Substructures are considered to be entities which may, in principle, be available for independent testing prior to incorporation into the full-scale structure. This decomposition allows a hierarchical framework, where substructures may be further decomposed. At the base of the hierarchy are components, which are substructures that cannot be decomposed further. A joint is the physical mechanism by which substructures are joined together, thus moving up the hierarchy until \mathcal{S} itself is finally obtained. Joints themselves may be physically complex to the extent that they may be deemed substructures.

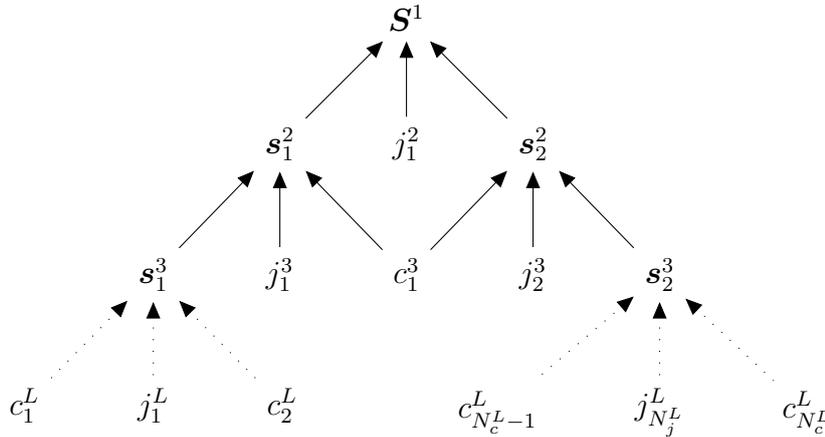


Figure 5.1: A hierarchical graphical representation of a generic structure \mathcal{S} . Superscripts denote the level in the hierarchy and subscript indexes each type of constituent unit in a given level. Dotted edges imply an arbitrary structuring between levels.

Figure 5.1 depicts a graphical representation of a hierarchical structure that may be considered without loss of generality. Nodes represent the global structure and its constituent units and edges represent the dependence of a (sub)structure on its constituent units. At the top, or Level 1, of the hierarchy is the global structure with

the hierarchy level denoted in the superscript. It can be seen that the global structure \mathbf{S}^1 is comprised of two substructures \mathbf{s}_1^2 , \mathbf{s}_2^2 and a joint j_1^2 , i.e. $\mathbf{S}^1 = \{\mathbf{s}_1^2, j_1^2, \mathbf{s}_2^2\}$. These units form the second level of the hierarchy. \mathbf{s}_1^2 and \mathbf{s}_2^2 may in turn be expanded to yield $\mathbf{S}^1 = \{\{\mathbf{s}_1^3, j_1^3, c_1^3\}, j_1^2, \{c_1^3, j_2^3, \mathbf{s}_2^3\}\}$. Progressing down the hierarchy levels, one can continue to expand the substructures into constituent units until the L^{th} level of the hierarchy which is comprised solely of base units (components and joints). By taking the expansion of \mathbf{S}^1 into its constituent base units and discarding the repeated units arising from substructures that share components, one obtains a list of the base units that form a given structure \mathbf{S} . Within each level of the hierarchy, units are numbered via a subscript from 1 to N_u^i , where N_u^i is the number of a constituent unit type u in the i^{th} level of the hierarchy. The notation u_n^i , where i is an integer from 1 to L and n is an integer from 1 to N_u^i , provides a unique identifier for each unit within a structure. To further elucidate this notation; with reference to Figure 5.1, one can consider the joint units ($u = j$) within the third level of the hierarchy ($i = 3$) of which there are in total two ($N_j^3 = 2$). Thus, if one is required to reference the second joint in the third level of the structure's hierarchy, the notation j_2^3 may be used.

It is assumed that there exists a set of features $\boldsymbol{\nu}$, observable from \mathbf{S} , that are produced according to a generative latent state model with latent state $\mathbf{H}^*(t)$, where $\mathbf{H}^*(t)$ is the true health state of \mathbf{S} and may be expressed in terms of the true health states of the constituent components and joints $h^*c_n^i(t)$ and $h^*j_n^i(t)$, respectively, i.e. $\mathbf{H}^*(t) = \{h^*c_1^2(t), \dots, h^*c_{N_c^L}^L(t), h^*j_1^2(t), \dots, h^*j_{N_j^L}^L(t)\}$.

The structure \mathbf{S} also has a predicted time-dependent health state vector $\mathbf{H}(t) = \{hc_1^2(t), \dots, hc_{N_c^L}^L(t), hj_1^2(t), \dots, h^{j_{N_j^L}^L}(t)\}$. Health-state vectors can be constructed from any subset of components and joints.

For the structure/system \mathbf{S} , there exists a set of failure modes of interest $\mathbf{F} = \{F_1, \dots, F_{N_F}\}$ whereby \mathbf{S} ceases to be fit for purpose with respect to critical performance criteria. It is assumed that a given failure mode is dependent on the health states of a subset of components, joints and substructures for which a health-state vector can be constructed. In addition to structural health states, failure modes may also be conditionally dependent on specific operational conditions $\mathbf{o} = \{o_1(t), \dots, o_{n_o}(t)\}$ and environmental conditions $\mathbf{e} = \{e_1(t), \dots, e_{n_e}(t)\}$. Here, it is worth noting that operational conditions may relate to aspects of the environment that can be controlled.

The operational and environmental conditions may also have the potential to alter the distribution of ν , and the time-dependent function $\mathbf{H}(t)$.

For the structure \mathbf{S} , there also exists a set of decisions $\mathbf{d} = \{d_1, \dots, d_{N_d}\}$ which affect $\mathbf{H}^*(t)$, directly, or indirectly by altering $\mathbf{o}(t)$. Finally, there will be a utility function $U(\mathbf{F}, \mathbf{d}, \mathbf{o}, \mathbf{e})$ that reflects the economic gains and costs associated with the possible failure states, operational states, and environmental states, subject to decidable actions.

5.2 Mapping PRA onto SHM

Upon examination, it becomes apparent that there are both differences and similarities between the paradigms for SHM and PRA that can be considered in order to determine which aspects of PRA will be useful for SHM. Whilst it is clear that both SHM and PRA are utilised for the purpose of making decisions in the face of uncertainty, PRA is typically conducted offline for a system experiencing a set of anticipated initiating events. In contrast, the decision processes for which SHM is implemented are online and require continual predictions of the damage state of the structure. It is for this reason that the event-tree development stages and failure-sequence quantification stages in PRA are less applicable to SHM.

Both paradigms begin with collating information regarding the structure and defining the context in which decisions are to be made. In fact, the first three stages of the PRA paradigm involve expressing the structure and context in a logical way which facilitates the quantification of risk and the decision-making process. It is in this formal expression of the structure that the decision-making process in the SHM paradigm stands to benefit. An overview of the risk-based SHM paradigm is as follows:

1. Operational evaluation.
2. Failure-mode modelling.
3. Decision modelling.
4. Data acquisition.
5. Feature selection.

6. Statistical modelling for feature discrimination.

5.2.1 Operational evaluation

With the aim of justifying the use, and defining the context of a risk-based SHM system, the operational evaluation stage seeks to answer many of the same questions as in the standard paradigm. However, some questions require an approach that facilitates the failure-mode modelling and decision-modelling stages.

Foremost, information regarding the components \mathbf{c} , joints \mathbf{j} , substructures \mathbf{s} , and the dependencies between them is required.

When identifying the critical damage states of the structure \mathbf{S} , one should aim to identify the failure modes of interest \mathbf{F} . Critical components, joints and substructures/subsystems that contribute to \mathbf{F} should also be identified at this stage. The anticipated damage states of these components h should be defined. The damage states of the critical substructures/subsystems \mathbf{H} should be defined as a vector in terms of h .

For each failure mode in \mathbf{F} , potential decidable interventions \mathbf{d} should be identified and the ways in which the actions influence the structure, or likelihood of failure modes occurring, should be determined. Utility values \mathbf{U} for all failure-action combinations given by $\mathbf{F} \times \mathbf{d}$ should be quantified. The selection of utility values will determine the behaviour of the decision-making agent, and is analogous to setting a decision threshold in a standard SHM paradigm.

Environmental influences \mathbf{e} should also be identified. It should also be decided whether the SHM system is to evaluate the health of the structure at static, independent instances in time, or predict future health states, thereby requiring a model forecasting the degradation of the structure.

For large, complex structures it may be beneficial to borrow the data management techniques used in PRA, such as databases, to organise the information obtained during the Operational Evaluation stage. This practice will allow for a rigorous and structured approach to the information collection and allow for the identification of aspects of the SHM system that require further specification or more information. Having a formal information structure will also expedite the subsequent failure-mode modelling step which requires detailed knowledge of the physical structure.

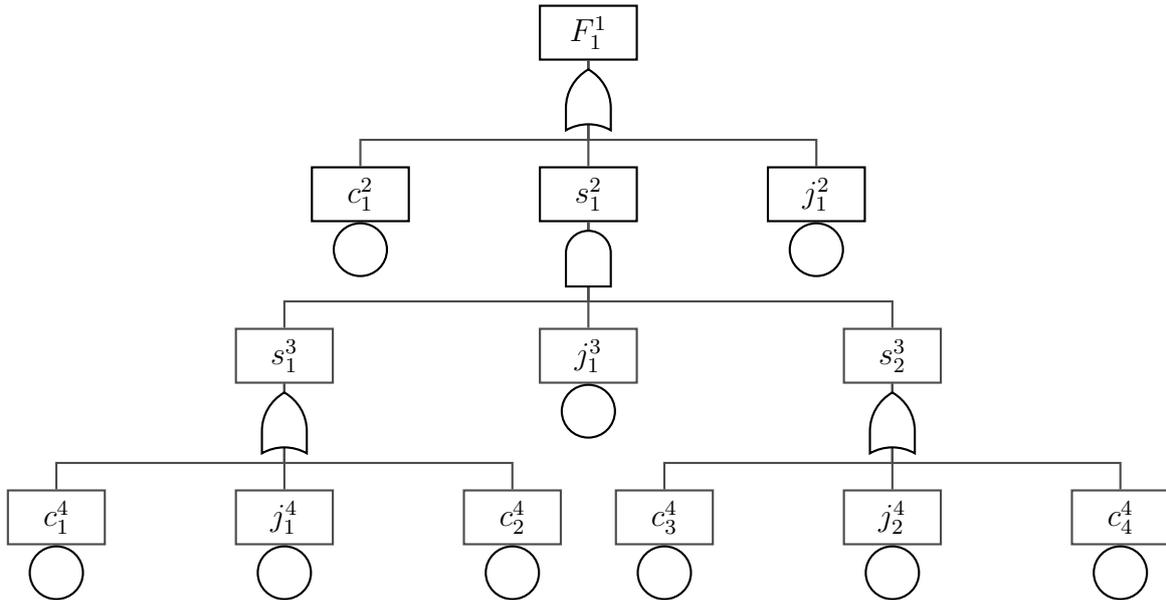


Figure 5.2: A fault tree of a (single) failure mode F_1 , where the superscript denotes the hierarchy level and the subscript is an identifier.

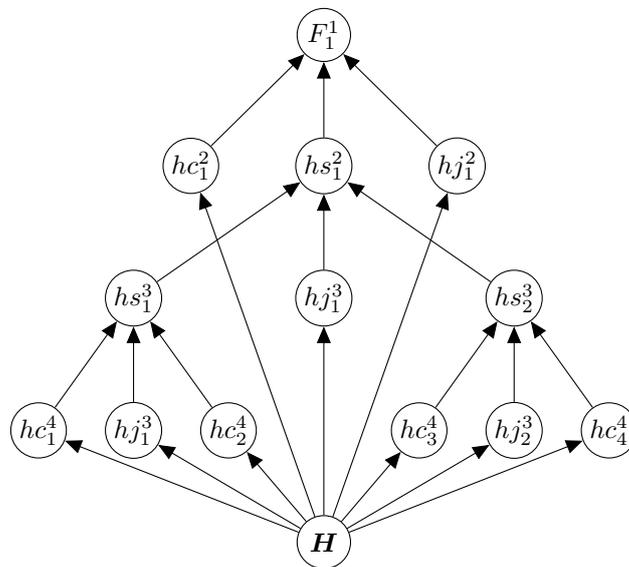


Figure 5.3: A Bayesian network representation of the failure mode F_1 .

5.2.2 Failure-mode Modelling

For each of the failure modes of interest in \mathbf{F} , one should proceed to construct a fault tree, such as that shown in Figure 5.2, based upon the health states of the relevant components, joints and substructures/subsystems. It should be noted here that, in many cases, the exact nature of the failure modes will be unknown and so it

may be necessary to use a best estimate based on engineering judgement.

Fault trees offer a rigorous and consistent structure for expressing the failure modes; however, as statements in Boolean logic they are limited in their flexibility. In the context of SHM, it is desirable to represent the components in a fault tree as having multiple damage states, and it is for this reason that one should map the constructed fault trees into Bayesian networks. Bobbio *et al.* outline a convenient mapping from fault trees into Bayesian networks in [86], whilst also highlighting the additional flexibility that is granted by doing so. Additionally, Bayesian networks are used to represent structural failures in [87].

In the example shown in Figure 5.3, the component health states, substructure health states and failure event are represented as random variables, where the substructure health states are conditioned on the component health states and the failure events is conditioned on the substructure health state. The random variables are defined using a conditional probability distribution (CPD) which may be discrete or continuous.

A node representing the health-state vector of the critical components and joints \mathbf{H} should be included in the fault tree, as this latent state will be predicted during the statistical modelling process. To define the vector \mathbf{H} within the Bayesian network, the conditional dependence between the nodes representing the local health states of the components and joints and \mathbf{H} are expressed as a binary logic table.

The primary function of the failure-mode Bayesian network is to allow the flow of information regarding the health states of a structure to determine a probability of failure. The fault-tree structure of the network facilitates the process of defining the failure events in \mathbf{F} . In addition, the network also allows the computation of marginal distributions for the probability of failure in each component, joint, or substructure allowing damage to be conveniently localised to individual components, or substructures.

5.2.3 Data Acquisition

The data acquisition process should not differ greatly from that in the standard SHM paradigm. However, there is a subtlety that, where possible, the data acquisition system should be designed so as to optimise the decision-making rather than damage identification.

5.2.4 Feature Selection

The feature-selection process, in general, should not differ from that in the standard SHM paradigm. Although, it is worth restating that there is the subtlety that the features should be selected so as to optimise the decision-making and this may lead to novel feature-selection approaches in the future.

5.2.5 Statistical Modelling

For decision-making in SHM, two statistical models are required.

The purpose of the first statistical model is to predict the critical health states \mathbf{H} , given the selected feature set $\boldsymbol{\nu}$. As aforementioned, it is assumed that $\boldsymbol{\nu}$ is produced via a generative latent state model, with latent-state \mathbf{H} , though the classification models used to predict \mathbf{H} do not necessarily need to be generative. Probabilistic classifiers that output a probability distribution over all possible states of \mathbf{H} , such as Gaussian mixture models (GMMs) and relevance vector machines (RVMs), are compatible and are discussed further in Chapter 9. Probabilistic classifiers are instrumental in building robustness to the uncertainty surrounding the true health state of \mathbf{S} into the decision process. Ideally, the chosen statistical model will be capable of consistently identifying the actual health state under all identified operating and environmental conditions \mathbf{e} , or at least appropriately reflect the uncertainty caused by varying conditions in the prediction.

Secondly, if a model describing the degradation of \mathbf{S} (i.e. a transition model for \mathbf{H}) is required. The purpose of the transition model is to forecast health-states forward in time, such that future failure probabilities can be determined via the failure-mode model. The transition model corresponds to the CPD, $P(\mathbf{H}_t | \mathbf{H}_{t-1}, \mathbf{d})$, and is discussed further in Chapter 7.

As with other stages in the current risk-based paradigm for SHM, the statistical models may be developed so as to optimise decision-making. This notion, in the context of statistical classifiers, forms the basis for Chapters 8 and 9.

5.2.6 Decision Modelling

Incorporating decision-theoretic aspects into the framework involves augmenting the Bayesian network developed in the previous stages with nodes for each decision in \mathbf{d} and for the utility function $U(\mathbf{F}, \mathbf{d}, \mathbf{o}, \mathbf{e})$ to produce an influence diagram. The manner in which decisions alter future states should be accounted for in the transition model CPDs accordingly. Multiple utility nodes may be introduced if the utility function is additively separable, as discussed in Chapter 4. Utility nodes are constrained to be leaf nodes and should be dependent on the appropriate failure events, decisions, or operational conditions.

5.3 Discussion

The framework described provides an approach to conducting risk-based SHM that incorporates useful stages of the PRA procedure into the SHM paradigm. Decision-making is facilitated via the inclusion of risk, thereby allowing for the comparison of actions and the selection of one that maximises expected utility.

The PRA paradigm currently practised in industries such as aerospace and nuclear provides a basis for the formalisation of the operational evaluation procedure. Organising the information, specifying the structure, and monitoring system in a database will assist with ensuring all the necessary information required for subsequent stages is acquired and it will also provide a structured method for the retrieval of applicable information at each stage.

The fault tree development process of PRA provides the key novelty of this risk-based approach to SHM. Firstly, it facilitates the definition of key failure modes of interest and provides a structured method for identifying pertinent components whose health states should be targeted by a statistical classifier. The size of decision space for any given structure in the context of SHM is vast and an intimidating problem to begin addressing, with further implications for the tractability of decision problems. By targeting selected failure modes of interest for a structure and modelling them as fault trees, the scope of the decision-maker may be limited, thereby making the problem more approachable; additional failure modes may subsequently be incorporated as an SHM system is further developed/expanded. Mapping the fault trees into Bayesian networks enables the framework to retain information regarding the uncertainties

in the health states thereby allowing robustness in the decision-making. Moreover, intermediate nodes within the Bayesian network representation of a fault tree may be queried, yielding marginal distributions that provide information about the probability of damage within components and/or substructures that will streamline the localisation of damage. This localisation information may be utilised to guide inspection and maintenance engineers to specific locations, potentially saving time and reducing the cost of the actions, particularly for larger structures.

5.4 Summary

To summarise, a risk-based framework for structural health monitoring was presented. Borrowing practices frequently used in probabilistic risk assessment – such as the use of fault trees to model system failures – the framework facilitates robust decision-making under uncertainty. In the next chapter, the risk-based decision framework will be demonstrated via an experimental case study.

CASE STUDY: FOUR-BAY TRUSS

In the current chapter, the probabilistic risk-based paradigm for SHM, introduced in the previous chapter, is applied to a practical case study. Specifically, by following the steps outlined, an influence diagram representing a maintenance decision process for a four-bay truss is developed, and subsequently analysed with respect to experimental data.

6.1 Four-bay Truss

To demonstrate the risk-based approach to decision-making for SHM, the framework was applied to a four-bay truss structure, identical to that used in [88], and visualised in Figure 6.1. For clarity, the example will be limited to consider a single failure mode and a single binary decision. The failure of joints will also be ignored.

The truss itself was composed of 20 members – eight horizontal, four vertical and eight cross – each with a cross-sectional area of 177 mm^2 . The overall length of the structure was 1 m and the height was 0.25 m, The members were pinned together with steel bolts in lubricated holes.

To generate an experimental dataset, the truss was subjected to a preload of 5 kg at point P and additional consecutive loads of 10, 20 and 30 kg at each of the eight points L in turn. For each of the 24 load cases, microstrains were measured at the midpoints of the 12 horizontal and vertical members. This process was performed

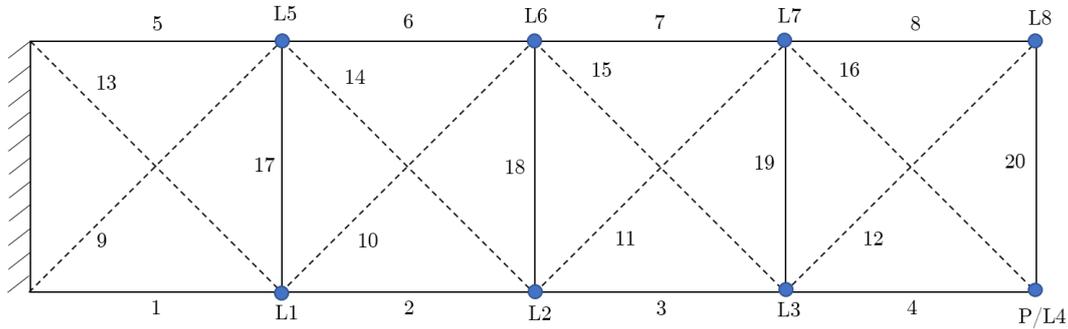


Figure 6.1: A two-dimensional four-bay truss comprised of 20 members, eight of which are removable and denoted by a dashed line. Loads are applied at points L, and a preload is applied at point P. Load positions are shown as blue dots. The bays are numbered left to right from 1 to 4.

eight times in total, once with each cross-member removed.

In addition to the experimental data, a finite element simulation of the truss was developed whereby removal of a cross-member was achieved by assigning a highly-reduced Young's modulus of 1 MPa. As well as the 10, 20 and 30 kg loads used in the experiment, the truss was simulated with loads of 5, 15 and 25 kg. Furthermore, strains were obtained for the truss under each load case in its undamaged condition, i.e. with all cross-members intact.

6.1.1 'Operational Evaluation' of Truss

In order to construct a risk-based decision framework for the truss, one must first define it formally. As it was elected to ignore joints, the global truss structure \mathbf{T} can be defined as four substructures, one for each bay i.e. $\mathbf{T} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$. As only the failures of cross-members were considered, each bay can in turn be defined as two components, for example, $\mathbf{b}_1 = \{m_9, m_{13}\}$. Consequently, the critical health states of the global structure may be summarised as $\mathbf{T} = \{m_9, m_{10}, m_{11}, m_{12}, m_{13}, m_{14}, m_{15}, m_{16}\}$.

A single failure mode $F_{\mathbf{T}}$ of the truss was considered; the full or partial collapse of the structure. This failure mode corresponds to the event where the truss is no longer able to support the load/preload, hence, $F_{\mathbf{T}}$ occurs when both cross-members in a single bay fail.

In an attempt to minimise the occurrence of the failure mode of interest $F_{\mathbf{T}}$, a

single binary decision d was identified; a choice between ‘do nothing’ and ‘perform maintenance’. In addition, utilities were assigned to the failure event and the decidable actions somewhat arbitrarily, though in a manner which intends to reflect the relative costs associated with failure and maintenance in real-world engineering applications. There is a small positive utility associated with the structure being in its operational state, and a large cost associated with the structure being in its failed state. The ‘do nothing’ action is given zero utility and the ‘perform maintenance’ action is given a cost greater than the utility gained from the structure being operational in a given time-step, although less than the cost of failure. The utilities assigned to the failure and decision are shown in Tables 6.1 and 6.2, respectively. It is worth noting that, in many practical applications, the specification of utility function is non-trivial and is an active topic of research outside the scope of the current chapter. For further reading related to the specification of utility functions, the reader is directed to [89]. Hence, for this case study, relative utility values are selected to be only somewhat representative of the SHM context. Further discussion of the challenges associated with assigning utilities is provided in Section 6.2.

Table 6.1: A table showing the entries of the utility function $U(F_{\mathbf{T}})$ where $F_{\mathbf{T}} = 0$ and $F_{\mathbf{T}} = 1$ denote the truss being operational and failed, respectively.

| $F_{\mathbf{T}}$ | $U(F_{\mathbf{T}})$ |
|------------------|---------------------|
| 0 | 15 |
| 1 | -285 |

Table 6.2: A table showing the entries of the utility function $U(d)$ where $d = 0$ and $d = 1$ denote the ‘do nothing’ and ‘perform maintenance’ actions, respectively.

| d | $U(d)$ |
|-----|--------|
| 0 | 0 |
| 1 | -100 |

For the purposes of demonstration, it is assumed that the load on the truss will be uncertain, varying in discrete time within the interval $[0, w_{max}]$ where w_{max} is defined in subsection 6.1.3. Furthermore, it is assumed that the location of the load also varies in discrete time and that, in the limit of infinite time-slices, each of the eight locations is visited an equal number of times.

6.1.2 Truss Failure Modelling

The failure mode F_T of the truss can be represented by the fault tree shown in Figure 6.2, where the failure of a bay \mathbf{b} is defined as the AND-gate of two cross-member m failures, and the failure of the truss F_T is defined as the or-gate of the bay failures.

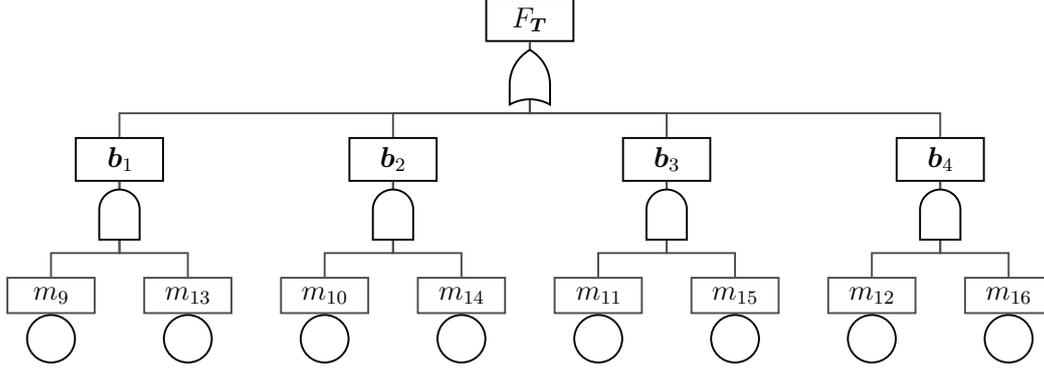


Figure 6.2: A fault tree of failure mode F_T for a four-bay truss. The failure mode F_T occurs if at least one bay \mathbf{b} fails. A bay \mathbf{b} will fail if both cross-members m fail.

To map the fault tree for the failure event F_T into a probabilistic graphical model, hb and hm will be used to denote the random variables that represent the local binary health states of the bays and cross-members, respectively, where 0 corresponds to intact and 1 corresponds to failed. Additionally, \mathbf{H} will be used to denote the random variable vector for the health state of the global structure where $\mathbf{H} = \{hm_9, hm_{10}, hm_{11}, hm_{12}, hm_{13}, hm_{14}, hm_{15}, hm_{16}\}$. For conciseness, the vector \mathbf{H} will, on occasion, be summarised as $\mathbf{H} = H$, where H is the decimal representation of the 8-bit binary number specified by the vector. Finally, the F_T notation will be retained to represent the random variable corresponding to the failure event. The Bayesian network corresponding to the fault tree shown in Figure 6.2 is shown in Figure 6.3. The conditional probability distributions specifying $P(F_T|hb_1, hb_2, hb_3, hb_4)$ (or $P(F_T|hb)$ for brevity) and $P(hb_1|hm_9, hm_{13})$ (or $P(hb_1|hm_{b_1})$ for brevity) are shown in Tables 6.3 and 6.4, respectively.

6.1.3 Health-state Transition Modelling

The purpose of health-state transition modelling is to develop the conditional probability distribution $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d)$ that predicts the future health state of the truss forward in time, given the current health state and the decided action.

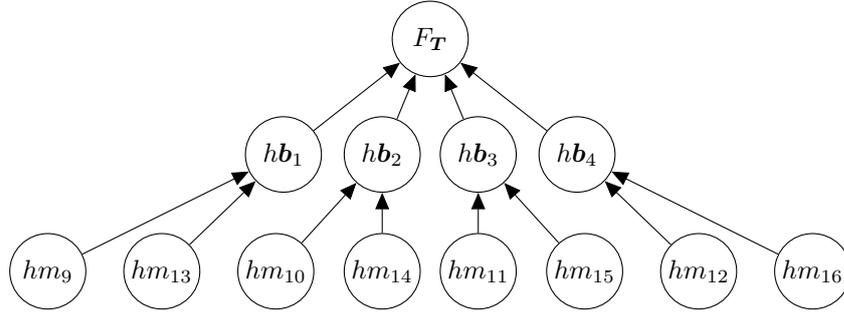


Figure 6.3: A Bayesian network representation of failure mode F_T for a four-bay truss.

Table 6.3: A table showing the entries of the conditional probability distribution $P(F_T|hb_1, hb_2, hb_3, hb_4)$ where $hb = 0$ and $hb = 1$ denote a bay being intact and failed, respectively, and $F_T = 0$ and $F_T = 1$ denote the truss being operational and failed, respectively.

| hb_1 | hb_2 | hb_3 | hb_4 | $P(F_T = 1 hb)$ | $P(F_T = 0 hb)$ |
|----------|----------|----------|----------|-----------------|-----------------|
| 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| 0 | 1 | 1 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 |

For the purpose of this demonstration, it was decided that the ‘perform maintenance’ action simply returns the structure to its undamaged state, i.e. with no cross-members failed, with probability 1, independent of \mathbf{H}_t .

With regard to the ‘do nothing’ action, it was first assumed that the truss would not spontaneously transition from a more advanced damaged state to a lesser one, that is to say, cross-members would not self-repair in the absence of intervening maintenance, or, without maintenance the health-state of the structure monotonically degrades as a function of time.

The assumed loading range $[0, w_{max}]$ was discretised into 100 evenly-spaced increments; combined with the eight possible load locations, this resulted in 800 unique considered load cases L_c .

For a given load case L_c , a transition in health state was defined as $\mathbf{H}_{t+1} = \mathbf{H}_t +$

Table 6.4: A table showing the entries of the conditional probability distribution $P(h\mathbf{b}_1|hm_9, hm_{13})$ where $hm = 0$ and $hm = 1$ denote a member being intact and failed, respectively, and $hb = 0$ and $hb = 1$ denote a bay being intact and failed, respectively.

| hm_9 | hm_{13} | $P(h\mathbf{b}_1 = 1 h\mathbf{m}_{b_1})$ | $P(h\mathbf{b}_1 = 0 h\mathbf{m}_{b_1})$ |
|--------|-----------|--|--|
| 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 0 |

$\delta\mathbf{H}_{t \rightarrow t+1}$. $\delta\mathbf{H}_{t \rightarrow t+1}$ is an 8-bit binary vector with i^{th} entry equal to 1 if the yield stress of aluminium (300 MPa) is exceeded in member m_{i+8} when the truss is simulated in health state H_t subject to load case L_c , and equal to 0 otherwise. The conditional probability of transitioning from $\mathbf{H}_t = H_t$ to $\mathbf{H}_{t+1} = H_{t+1}$, $P(\mathbf{H}_{t+1} = H_{t+1}|\mathbf{H}_t = H_t, L_c)$, was assigned unity if $\delta\mathbf{H}_{t \rightarrow t+1} = \mathbf{H}_{t+1} - \mathbf{H}_t$ for $\mathbf{H}_t = H_t$ and $\mathbf{H}_{t+1} = H_{t+1}$ for load case L_c , and assigned zero otherwise. The full transition matrix $P(\mathbf{H}_{t+1}|\mathbf{H}_t)$ was then populated, where the entry $P(\mathbf{H}_{t+1} = H_{t+1}|\mathbf{H}_t = H_t)$ is given by,

$$P(\mathbf{H}_{t+1} = H_{t+1}|\mathbf{H}_t = H_t) = \frac{\sum_{L_c=1}^{N_{L_c}} P(\mathbf{H}_{t+1} = H_{t+1}|\mathbf{H}_t = H_t, L_c)}{N_{L_c}} \quad (6.1)$$

where N_{L_c} is the total number of load cases considered and $N_{L_c} = 800$.

For illustrative purposes, the maximum load w_{max} was determined by asserting $P(\mathbf{H}_{t+1} \neq 0|\mathbf{H}_t = 0) = 0.005$ and the value of w_{max} that satisfied the condition was found to be approximately 6900 kg. This is obviously somewhat arbitrary, and in practice the maximum load for a structure may be estimated during the operational evaluation stage.

The transition model developed provides a means of forecasting future health states of the truss, a heatmap representation of the transition matrix is shown in Figure 7.1.

6.1.4 Statistical Classifier Development and Training

The purpose of the statistical classifier is to obtain a probability distribution over the current health state, given a set of observed features, i.e. $P(\mathbf{H}_t|\boldsymbol{\nu})$. In the current case

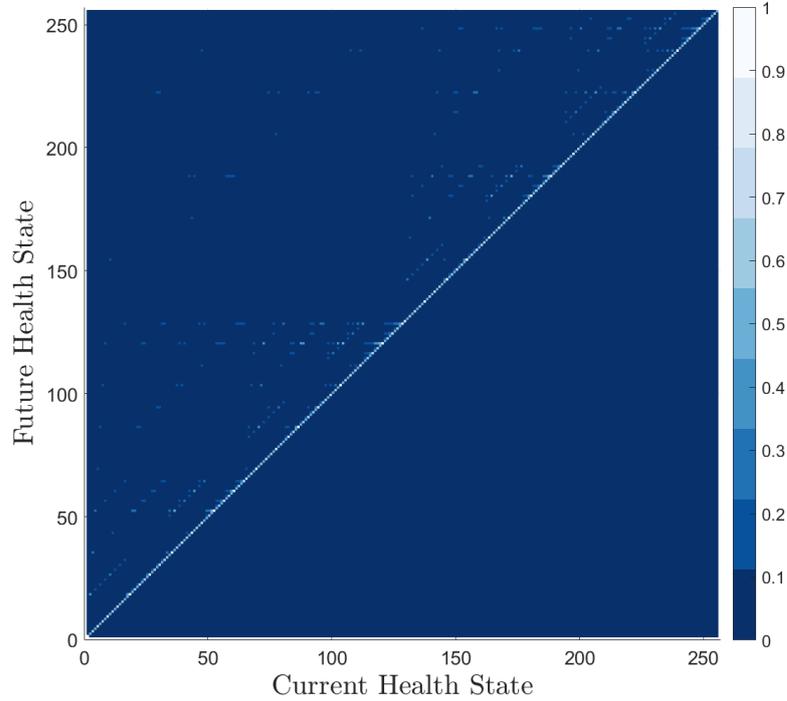


Figure 6.4: A heatmap showing the conditional probability distribution transition matrix $P(\mathbf{H}_{t+1}|\mathbf{H}_t)$.

study, the damage indicative features are the strains measured from the horizontal and vertical members of the truss. The classifier selected for the current case study was comprised of two components; a detector and a localiser, corresponding to the first two stages of Rytter’s hierarchy. Whilst generative models may better reflect the causality of the problem at hand, where it is assumed that the features are generated as a result of the latent state and $P(\mathbf{H}_t|\boldsymbol{\nu})$ may be computed via Bayes’ Theorem, discriminative classifiers that directly learn a mapping from the feature space to the label space are also applicable in the risk-based decision framework.

A Gaussian novelty detector was implemented to determine the probability that the structure is currently in its undamaged state $P(\mathbf{H}_t = 0|\boldsymbol{\nu})$, and, as its complement, the probability that the structure is currently in a damaged state $P(\mathbf{H}_t \neq 0|\boldsymbol{\nu})$. The first two principal components of the simulated strain data for the undamaged structure are compared to those from the damaged structure in Figure 6.5. Inspection of Figure 6.5 reveals that it is possible to discriminate between the undamaged and damaged finite element simulation data using only the first principal component of the strains, hence the novelty detector uses the first principal component as the

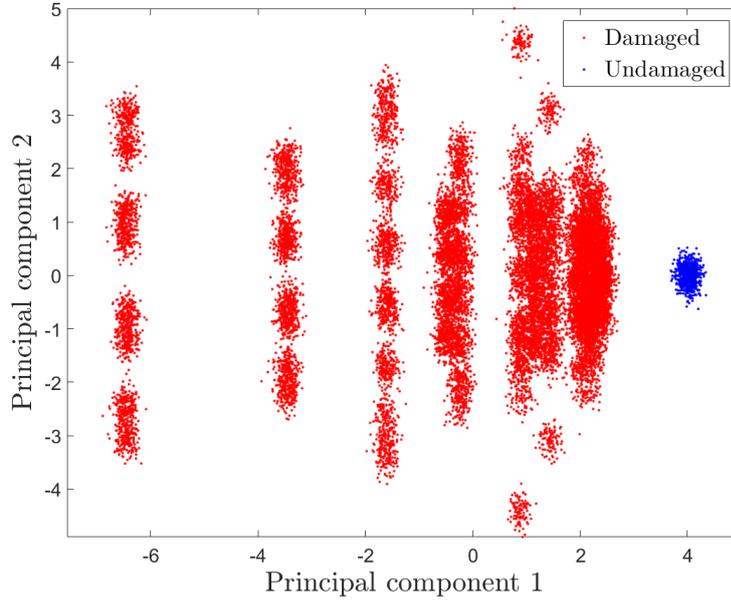


Figure 6.5: A comparison of the distributions of the first two principal components of the strain data obtained from the finite element model of the undamaged and damaged truss.

discriminative feature. As such, this principal component projection was learned from the training data and the detector was formed by computing the mean μ and standard deviation σ of the univariate distribution of the first principal component. If the first principal component of an incoming set of strains were to lie within the range $\mu \pm 3\sigma$, it was asserted that the observed strains came from the structure in its undamaged condition with confidence 0.997, i.e. $P(\mathbf{H}_t = 0|\boldsymbol{\nu}) = 0.997$. If the first principal component lay outside the $\pm 3\sigma$ confidence interval, $P(\mathbf{H}_t = 0|\boldsymbol{\nu})$ was given by the probability mass in the tail of the Gaussian probability density function parametrised by μ and σ .

The function of the localiser component of the statistical classifier is to distribute $P(\mathbf{H}_t \neq 0|\boldsymbol{\nu})$ over the remaining 255 health states corresponding to the various combinations of cross-member failures. For consistency with [88], the classifier selected for this purpose was an artificial neural network (ANN) with an input node for each of the twelve strain measurements, an output node for each cross-member, and three hidden layers with twelve, twelve and eight nodes, respectively. The activation function used was the hyperbolic tangent function. As an identical classifier to that used in [88] was implemented, only health states $\mathbf{H}_t = \{1, 2, 4, 8, 16, 32, 64, 128\}$ are considered by the localiser in this case study.

In accordance with [88], a training dataset was constructed from the finite element data for the loads of 10, 20 and 30 kg and a validation dataset from the finite element data for the loads of 5, 15 and 25 kg. In both instances, 100 repetitions of the datasets were produced and superimposed with a noise pattern of 1 microstrain RMS. The optimal weights were computed using the scaled conjugate gradient (SCG) back-propagation algorithm [90] and evaluated and selected based on the classification performance of the network on the validation dataset. Although not inherently probabilistic, a pseudo-probabilistic interpretation for the activations of the output nodes was acquired via the use of a softmax function [91].

6.1.5 Overall Probabilistic Graphical Model

The failure model, transition models, classifier, decisions and utilities can be combined to form a partially-observable Markov decision process represented by the limited memory influence diagram (LIMID) shown in Figure 6.6. Figure 6.6 shows the decision process for two decisions over three time-slices. The informational link connecting ν_{t+0} to d_{t+0} implies that the features are observed prior to the decision being made. Similarly, the edge connecting d_{t+0} to d_{t+1} implies d_{t+0} is decided before d_{t+1} .

It should be noted that the model shown in Figure 6.6 assumes a generative model for the features ν ; for discriminative classifiers the direction of the conditioning edge connecting nodes H_{t+0} and ν_{t+0} would be reversed.

6.1.6 Results

Novelty Detector

When applied to the experimental strain data from the damaged truss, the univariate Gaussian novelty detector was able to correctly identify 175 out of the 192 observations as novel with respect to the simulated undamaged data, thereby yielding an overall accuracy of 91.1% and shown in the confusion matrix in Figure 6.7.

The 8.9% misclassification can be elucidated by examining the distribution of the experimental data as projected through the principal component mapping learned from the finite element simulation data, shown in Figure 6.8. Separability between

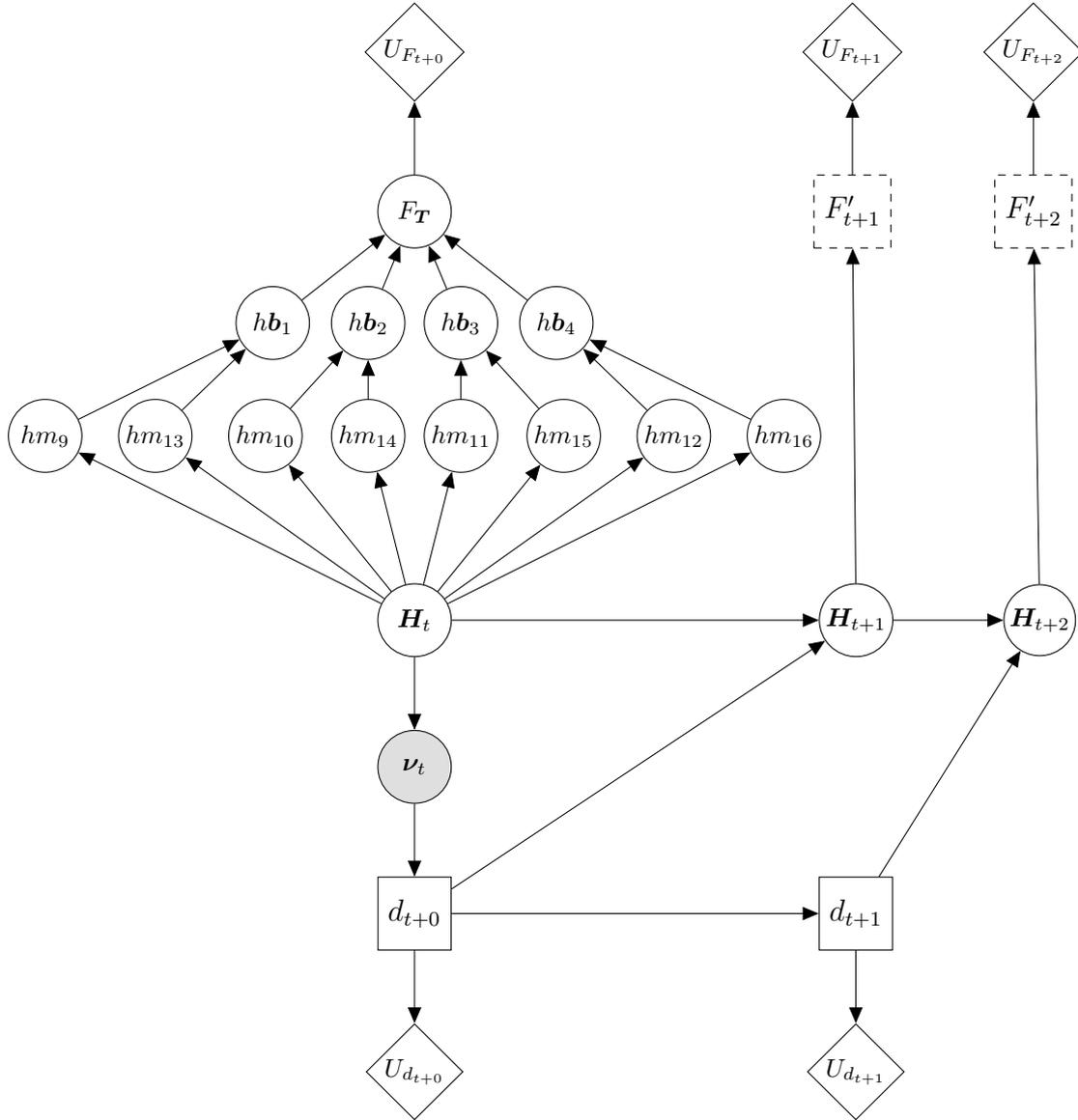


Figure 6.6: An influence diagram representing the partially observable Markov decision process for determining the utility-optimal maintenance strategy for the cross-members of a four-bay truss given observations of strains made from the horizontal and vertical members. Observed variables are shaded grey. The fault-tree failure models for the latter time steps have been represented as the nodes F'_t for compactness.

the projected damaged and undamaged datasets is absent where values of the first principal component are approximately 3.9. Finally, it is worth acknowledging that the NaN¹ arises in Figure 6.7 as there are no examples of the undamaged structure (Class 1) in the experimental dataset.

¹Here, NaN denotes ‘not a number’ – the typical result of a division by zero in many programming languages.

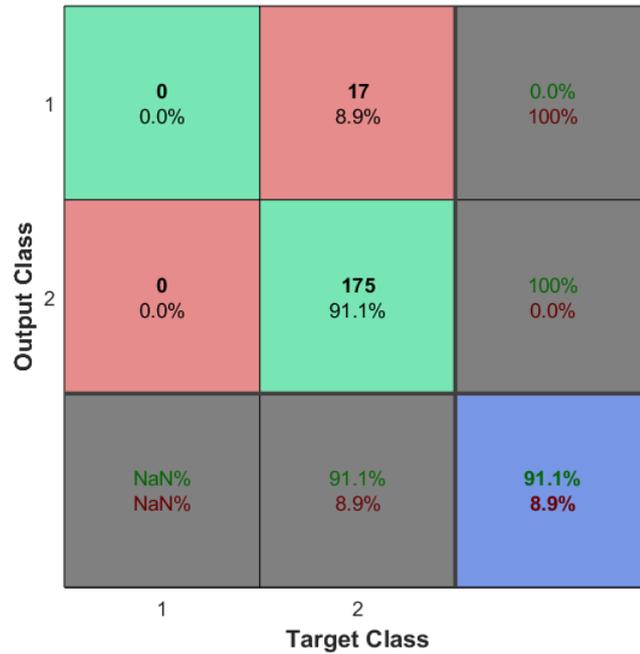


Figure 6.7: A confusion matrix detailing the classification performance of the novelty detector on the experimental data.

Localiser

When applied to the experimental strain data, the neural network localiser has an overall classification accuracy of 59.9%. The full confusion matrix is shown in Figure 6.9. Whilst 60% classification accuracy may be considered low, for a eight-class problem the neural network provides a significant improvement over simply guessing, which would yield an accuracy of 12.5%. The imperfect classifier has been deliberately chosen here as a possible source of uncertainty to highlight that improvements in decision performance can be realised even with imperfect information.

The misclassification error of 40.1% can be explained by considering the physics of the problem at hand. The selected damage sensitive features were the strains in the horizontal and vertical members of the truss; however, significant changes in the strains are only expected of members in the load path between the end fixture and applied mass. Therefore, the features are largely insensitive to damage when the mass is closer to the fixture than the damaged cross-member and any differences can be attributed to the strains induced by the preload. In 72 of the 192 (37.5%) cases, the damaged cross-member is not in the load path.

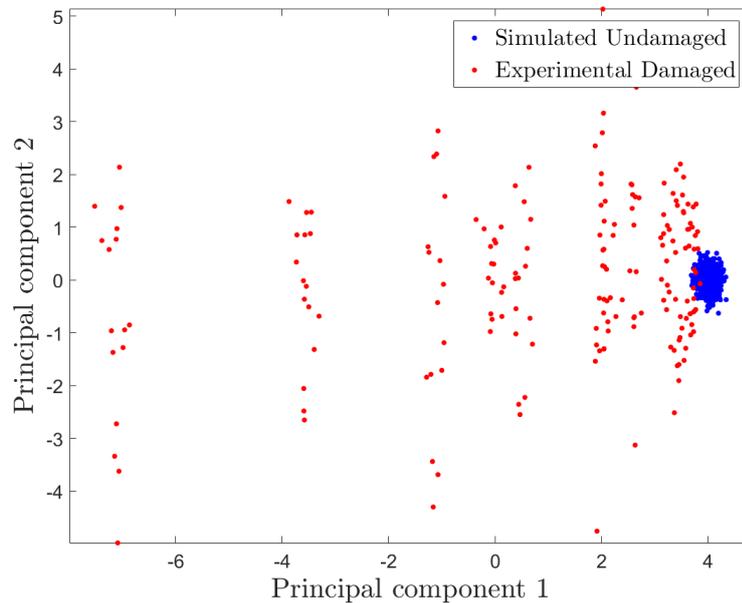


Figure 6.8: A comparison of the distributions of the first two principal components of the strain data obtained from the finite-element model of the undamaged truss and the strain data from the experiment performed on the damaged truss mapped through the projection learned from the simulated training data.

Decision Process Results

The decision algorithm was tested on a dataset comprised of the experimental strains for the truss in its damaged conditions and, due to the lack of experimental data, finite element simulation data for the truss in its undamaged condition. Equal proportions of the undamaged and damaged data were used; 192 sets of strains from each.

The decision algorithm used for testing was similar to that shown in Figure 6.6, except using the discriminative pseudo-probabilistic ANN classifier rather than a generative model. The graphical model was implemented in MATLAB using the Bayes Net Toolbox [92] and solved using the junction-tree algorithm for influence diagrams described in [93]. Utilities were as shown in Tables 6.1 and 6.2. Three failure events in consecutive time-steps were considered in conjunction with two ‘do nothing’/‘perform maintenance’ decisions in the first two time-steps with a single observation made during the first time-step.

The performance of the decision process was evaluated with a metric similar to that

| | | | | | | | | | |
|---|--------------|------------|------------|------------|------------|-------------|------------|------------|-------|
| 1 | 23 12.0% | 0 0.0% | 2 1.0% | 4 2.1% | 0 0.0% | 3 1.6% | 4 2.1% | 16 8.3% | 44.2% |
| 2 | 0 0.0% | 16 8.3% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% |
| 3 | 0 0.0% | 0 0.0% | 11 5.7% | 0 0.0% | 5 2.6% | 0 0.0% | 0 0.0% | 0 0.0% | 68.8% |
| 4 | 0 0.0% | 0 0.0% | 3 1.6% | 15 7.8% | 2 1.0% | 0 0.0% | 3 1.6% | 3 1.6% | 57.7% |
| 5 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 15 7.8% | 0 0.0% | 0 0.0% | 0 0.0% | 100% |
| 6 | 0 0.0% | 2 1.0% | 7 3.6% | 4 2.1% | 0 0.0% | 21 10.9% | 3 1.6% | 3 1.6% | 52.5% |
| 7 | 1 0.5% | 4 2.1% | 1 0.5% | 1 0.5% | 2 1.0% | 0 0.0% | 14 7.3% | 2 1.0% | 56.0% |
| 8 | 0 0.0% | 2 1.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% |
| | 95.8% | 66.7% | 45.8% | 62.5% | 62.5% | 87.5% | 58.3% | 0.0% | 59.9% |
| | 4.2% | 33.3% | 54.2% | 37.5% | 37.5% | 12.5% | 41.7% | 100% | 40.1% |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| | Target Class | | | | | | | | |

Figure 6.9: A confusion matrix detailing the classification performance of the artificial neural network localiser on the experimental data.

| | | | |
|---|--------------|--------------|-------|
| 0 | 584 76.0% | 12 1.6% | 98.0% |
| 1 | 40 5.2% | 132 17.2% | 76.7% |
| | 93.6% | 91.7% | 93.2% |
| | 6.4% | 8.3% | 6.8% |
| | 0 | 1 | |
| | Target Class | | |

Figure 6.10: A confusion matrix detailing the ‘accuracy’ of the decision algorithm for both decisions in the three time-slice problem for $U(F_{\mathcal{T}} = 1) = -285$.

of a classifier's overall accuracy. Whereas, in the simplest sense, a classification accuracy is a comparison between the predicted outputs and the target outputs, 'decision accuracy' is defined as a comparison between the actions selected by an agent using the statistical classifier being evaluated, and the optimal actions selected by an agent given perfect information, i.e. an agent in possession of the true target outputs of the classifier. A decision is considered 'correct' if the agent utilising the classifier selects the same action as an agent with perfect information up to the current time. It follows that a decision can be considered incorrect if the decided action differs from that selected by an agent operating with perfect information. A quantitative value for 'decision accuracy' may be calculated by taking the ratio:

$$\text{decision accuracy} = \frac{\text{number of correct decisions}}{\text{total number of decisions}} \quad (6.2)$$

The target and output classes '0' and '1' for Figures 6.10, 6.11 and 6.12 correspond to the 'do nothing' and 'perform maintenance' actions, respectively.

Figure 6.10 shows the performance of the decision algorithm across all 786 decisions associated with the test dataset. It can be seen that an overall 'accuracy' of 93.2% was achieved, meaning that the optimal decision given perfect information of the health state was selected in 716 of the cases when the statistical classifier was used to infer the health state. In 40 of the cases, the 'perform maintenance' action was selected unnecessarily, this is a result of the uncertainty in the health state triggering a more, conservative action to be taken; this form of error is akin to a 'false positive' or type I error. In 12 of the cases the 'do nothing' action was deemed to be optimal whereas, had perfect information of the structures health state been available, the optimal decision would, in fact, have been 'perform maintenance'. This form of error is akin to a 'false negative' or type II error.

The severity/significance of type I and type II errors is dependent on the context of the SHM system. For example, for an offshore wind structure, erroneously sending inspection/maintenance engineers has a higher cost relative to the failure event, whereas for a bridge the cost of inspection/maintenance is relatively lower with respect to the cost of failure. In this risk-based decision-framework, the costs are explicitly modelled and may be used to inform a preferential selection of classifier with regard to type I and type II errors.

Figures 6.11 and 6.12 provide a breakdown of the 'decision accuracy' shown in Figure

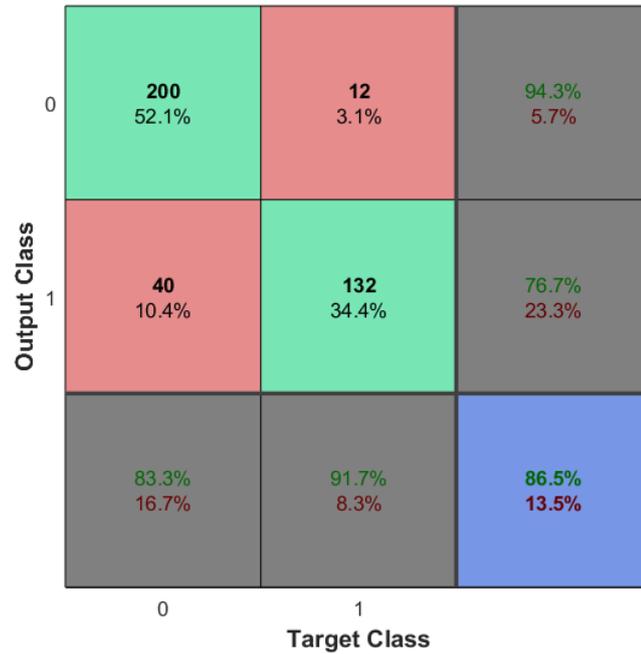


Figure 6.11: A confusion matrix detailing the ‘accuracy’ of the decision algorithm for the first decision in the three time-slice problem for $U(F_{\mathcal{T}} = 1) = -285$.



Figure 6.12: A confusion matrix detailing the ‘accuracy’ of the decision algorithm for the second decision in the three time-slice problem for $U(F_{\mathcal{T}} = 1) = -285$.

6.10 for the first and second decisions, respectively. Figure 6.11 shows that, for the first decision, the algorithm was able to select the correct action in 332 of the 384 cases, yielding an overall ‘decision accuracy’ of 86.5%. Additionally, it can be seen that in 40 cases the ‘perform maintenance’ action was selected incorrectly, and in 12 cases the ‘do nothing’ action was selected incorrectly. Comparing Figures 6.10 and 6.11 reveals that all type I and type II errors occur during the first decision. Logically, Figure 6.12 shows that the algorithm has a ‘decision accuracy’ of 100% with respect to the second decision. Moreover, Figure 6.12 shows that all optimal decisions during the second time slice are ‘do nothing’. This result can be explained by considering two possibilities. During the first time step, if the algorithm has decided that maintenance is warranted, then, under the assumed transition model $P(\mathbf{H}_{t+1}|\mathbf{H}_{t+0}, d = 1)$, the structure is guaranteed to be in its undamaged health state in the second time step in which case further maintenance is unwarranted given $P(\mathbf{H}_{t+2}|\mathbf{H}_{t+1}, d = 0)$ and $U(F_T = 1) = -285$. Alternatively, if, during the first time step, the algorithm deems ‘do nothing’ to be the optimal decision, then the degradation forecast over two time steps is not sufficient to warrant a ‘perform maintenance’ action in the second time step; however, this decision, of course, is dependent on the utilities specified within the model.

Figure 6.13 shows the number of state transitions that an initial undamaged state may go through before a ‘perform maintenance’ action is decided when the cost of the failure event and cost of maintenance are varied and all other utilities remain as specified in Tables 6.1 and 6.2. It can be seen that the number of transitions decays exponentially as the cost of failure increases. This result is in accordance with the intuitive understanding that, as the cost of failure tends to infinity, the number of transitions before maintenance should be decided asymptotically approaches zero. It can also be seen from Figure 6.13 that for a given failure event cost the time until maintenance decreases with cost of maintenance. It should also be noted that, logically, if the cost of maintenance exceeds the cost of failure, the structure will be allowed to operate until failure.

To investigate the influence of the cost of failure upon the overall ‘accuracy’ of the decision algorithm, the decision process used to produce Figure 6.10 was repeated for varying $U(F_T = 1)$. It should be noted that the utility of ‘perform maintenance’ action was fixed at $U(d = 1) = -100$. In addition to varying $U(F_T = 1)$, the decision algorithm was executed assuming a uniform distribution over the health states targeted by the classifier rather than the distribution as predicted by the

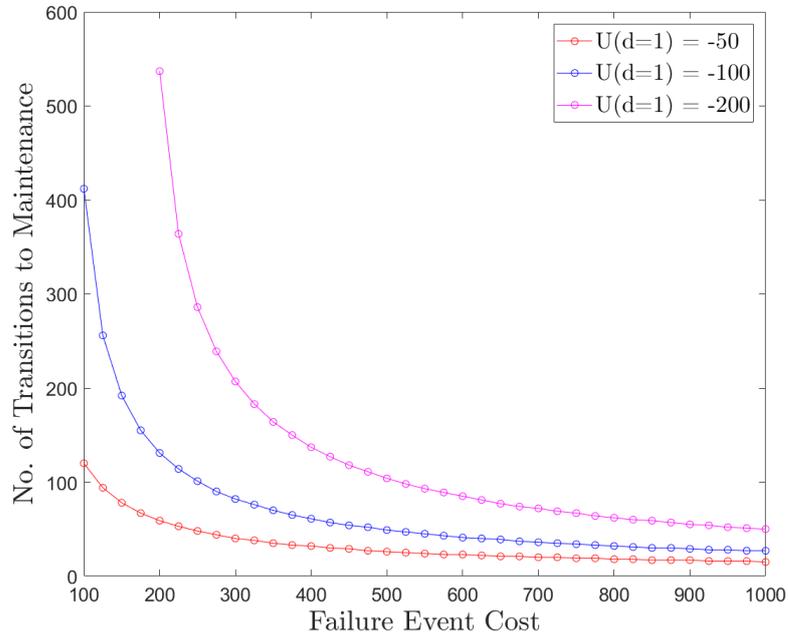


Figure 6.13: The variation in the number of state transitions the undamaged structure will go through before maintenance is decided, as a function of the failure event cost and maintenance cost. Failure event cost is defined as $-U(F_{\mathcal{T}} = 1)$.

classifier. Figure 6.14 shows how the ‘decision accuracy’ of each algorithm varies with the cost of the failure event.

Figure 6.14 shows that the agents using the classifier and the uniform distribution assumption all have perfect ‘accuracy’ when $U(F_{\mathcal{T}} = 1) = -100$. This result can be attributed to the fact that when the cost of maintenance is less than or equal to the cost of failure, the optimal decision is always ‘perform maintenance’; therefore, the algorithms all select this action independent of the information available regarding the health state of the structure.

It can be seen from Figure 6.14 that, with the exception of the somewhat trivial case of $U(F_{\mathcal{T}} = 1) = -100$, the agent utilising the classifier to incorporate probabilistic information about the health state of the structure into the decision, consistently performs better than the agent that assumes a uniform distribution over health states. This result is to be expected, as the use of the classifier allows for the identification of the undamaged condition and differing damage states. On the other hand, as assuming a uniform distribution is entirely ignorant of the health state, the selected decision will be invariant for a given failure-event cost. Therefore, even when the structure is undamaged, assuming a uniform distribution will result in a maintenance

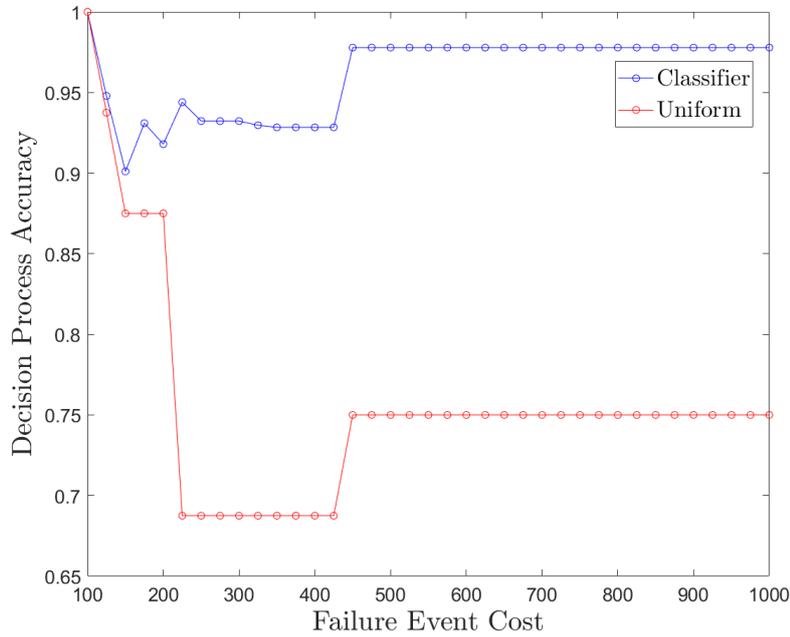


Figure 6.14: A comparison of the ‘accuracy’ of the decision process as a function of the cost of the failure event when the health state is inferred using the statistical classifier and a uniform distribution over health states is assumed. Failure event cost is defined as $-U(F_{\mathcal{T}} = 1)$.

action to be selected whilst the failure event cost is sufficiently high.

The ‘accuracies’ achieved by agents utilising the classifier and the uniform assumption follow similar trends for the range of failure costs shown in Figure 6.14. For $U(F_{\mathcal{T}} = 1) \leq -450$ the accuracy is constant. This result likely arises because of the high cost of failure – given the perfect knowledge of the health state, the optimal decision for all damage cases other than undamaged is ‘perform maintenance’. For the agent utilising the classifier, misclassification of damage location does not influence the decided action in this cost range. As previously mentioned, the uniform assumption is ignorant of the health state and the decided actions are invariant in the range $U(F_{\mathcal{T}} = 1) \leq -450$.

In Figure 6.14, a lower ‘decision accuracy’ is seen in the failure cost range $-450 < U(F_{\mathcal{T}} = 1) \leq -225$. This observation can likely be explained by the fact that, given perfect information, some locations of damage warrant the ‘do nothing’ action. It follows that, the uncertainty in health state reflected by the classifier and uniform distribution assumption cause the ‘perform maintenance’ action to be deemed appropriate instead.

In the failure cost range $U(F_{\mathcal{T}} = 1) \geq -200$, an increase in accuracy is seen for the algorithm assuming a uniform distribution over the health states. This observation is a result of the health-state invariant optimal decision becoming ‘do nothing’; in this failure cost range this assumption is able to correctly decide actions for the undamaged cases and the less severe damage locations. The ‘decision accuracy’ of the algorithm employing the classifier fluctuates in this range, this may be because the decided actions are sensitive to the distribution of uncertainties over the health states.

6.2 Discussion

The framework described and demonstrated in the current chapter provides an approach to risk-based decision-making in the context of SHM. Decision-making is facilitated via the inclusion of aspects of PRA such as fault-tree modelling and risk, thereby allowing for the comparison of actions and the identification of a strategy that maximises expected utility.

Although a relatively simple case study - specifically a truss - is used for illustrative purposes, the probabilistic risk-based framework may be extended straightforwardly to many of the other structural types that exist in mechanical or civil infrastructure. This approach is conditional on there being sufficient information available in the SHM operational evaluation stage that health states of interest and key failure modes may be identified and precisely defined with respect to the hierarchical representation of the structure. Furthermore, the framework is not specific to the sensor and data type used in the case study; in fact, for certain SHM applications, where information of the dynamics is required, different sensors and data types will be a necessity. The framework may be applied in such cases via the actions taken during the data acquisition, feature extraction and statistical modelling stages of the risk-based SHM paradigm.

Whilst the framework demonstrated addresses some of the problems surrounding the SHM decision process, there remain a number of challenges. One challenge, that has been widely acknowledged in the SHM community, is that data from the damage states of interest for a structure are seldom available prior to the implementation of an SHM system. This issue hinders the development of the probabilistic classifiers on which the decision process is highly dependent, and a

choice must be made regarding the approach to the statistical modelling (discussed further in Chapters 8 and 9). One option is to take a model-driven approach [94], that utilises outputs from physics-based models of the structure in its damage states of interest to learn a classifier in a supervised manner, pre-implementation of the SHM system. Subsequently, the classifier can be continuously updated and validated with data obtained during the monitoring campaign. Alternatively, a semi-supervised approach can be taken in which a clustering algorithm is applied to the data acquired throughout the monitoring campaign. Clusters are attributed damage-state labels via the incorporation of labelled data into the clustering algorithm with damage-state labels for data points that may be obtained from inspection of the structure [95, 96].

The results presented here show that the performance of a probabilistic risk-based decision algorithm is dependent on the available information regarding the health state of the structure. As demonstrated, gains with regards to utility can be made in the absence of high-accuracy classifiers, provided uncertainty is accounted for. This result provides motivation for moving towards the use of true probabilistic classifiers in the context of SHM, be they discriminative (such as relevance vector machines (RVMs) [97]) or generative (such as GMMs [98]).

In addition to being dependent on the statistical classifier used, the optimality of decisions is highly contingent on the appropriateness of the transition model used; if the degradation of the structure is not accurately modelled, erroneous actions may be taken. Facing a similar issue to the statistical modelling process, oftentimes, data describing the transitions between the health states of interest are not held *a priori*. Again, one is faced with the choice of taking a model-driven approach involving the simulation of the degradation, or a data-driven approach that utilises data obtained during the monitoring campaign. The development of data-driven transition models, or validation of model-driven transition models is an awkward problem. Due to the fact that one is performing interventions on the structure during operation, information on transitions between health states is regularly censored, meaning the quantities of data spanning all state transitions of interest required for developing/validating transitions models may never be acquired, and particularly troublesome for one-off/bespoke structures. This issue is revisited in Chapter 7.

To ensure the desired performance of the decision algorithm, a vital stage in the risk-based framework is to assign utilities/costs to failure events and actions. Currently, within the literature there is no general formalised approach to how these values should be elicited, nor is there a consensus on how the risk preferences of an SHM

decision-maker should be specified; should an agent be risk averse, risk neutral, or risk seeking? The issue at hand is one of both a technical and ethical nature, and whilst it will not be discussed in further detail in the current chapter, it is highlighted to stimulate the conversations required for progress in the area of risk-informed decision-making for SHM.

6.3 Summary

In summary, a probabilistic risk-based framework for structural health monitoring was presented. Borrowing practices frequently used in probabilistic risk assessment – such as the use of fault trees to model system failures – the framework facilitates robust decision-making under uncertainty and provides advancements in the utility-optimal operation and maintenance of structures. In Chapters 7 and 8, the role played by two of the key sub-models underpinning the decision process will be discussed in further detail; specifically, the health-state transition model and the statistical classifier/observation model.

HEALTH-STATE TRANSITION MODELS

As discussed previously, an agent tasked with making decisions regarding the operation and management of a structure may utilise health-state information inferred via an SHM system to make better informed and more optimal decisions. However, given solely information regarding the structural health state at the current instance in time, the agent may only make well-informed decisions *ad hoc*. In order to make well-informed decisions on policies that include *preventative* interventions, an agent requires information about the future health states of the structure. This information can be gained by developing transition models that forecast future health states given the current health state, decidable actions, and operational and environmental conditions.

The current chapter presents a methodology for determining a health-state transition matrix for use in a probabilistic risk-based decision paradigm for the operation and maintenance of structures as developed in [99], with particular focus on the modelling of *degradation*. The methodology relies on utilising physics-based models in conjunction with knowledge of anticipated operational conditions that may be obtained during the operational evaluation stage of an SHM campaign. The methodology is demonstrated using the four-bay truss study introduced in the previous chapter. Finally, the importance of health-state transition models within the risk-based decision framework is discussed, and the notable challenges associated with their development will be highlighted.

7.1 Degradation Modelling: An Overview

Transition models for which the conditional decided action is ‘do nothing’ essentially forecast the unchecked degradation of a structure.

Degradation models of differing complexities have been used within the field of engineering for reliability assessment, maintenance planning and prognosis [100]. In general, the models can be categorised in terms of a combination of the following criteria; physics-based or data-based, deterministic or probabilistic, and continuous state or discrete state. A commonly used degradation model is Paris’ law for fatigue-crack growth, given by the following equation,

$$\frac{da}{dN} = C(\Delta K)^m \quad (7.1)$$

where a is the crack length, N is the load cycle, ΔK is the stress intensity range and C and m are constants. After a little thought, one can reason that equation (7.1) is a deterministic, physics-based model of a continuous state. Different categories of degradation model are applicable in different scenarios, depending on the context. For example, in a situation where little is known of the underlying physics governing the degradation, but data are readily available, one may opt for a data-based model. Conversely, if the physics are known but data availability is low a physics-based model may be more suitable. Whether continuous or discrete states are modelled also depends on the nature of the application; considerations for this include the required model fidelity and the computational cost/time. Without delving too far into metaphysics, it is reasonable to assert that, in general, the future is inherently uncertain. For this reason, with regard to the use of deterministic versus probabilistic models, the latter have a distinct advantage, as they are capable of representing uncertainty. Fortunately, many deterministic degradation models can be used to obtain probabilistic outputs via methods such as sequential Monte Carlo sampling [101].

In the context of SHM and decision-making, a variety of health-state transition models have been employed. In [20], a probabilistic interpretation of Paris’ law is used to develop a degradation model in a maintenance decision process for a simulated wind turbine tower. In [21], a continuous health-state variable is given nonlinear Gaussian transition models in a partially-observable Markov decision

process (POMDP) based on a normalised unscented Kalman filter; this approach has the property that there is a non-zero probability that the health-state transitions to a less-damaged state, meaning that the structural degradation is not strictly monotonic. In [9], qualitative data obtained from the inspection of miter gate components is used to derive a health-state transition matrix for a Markovian decision process for optimal maintenance decisions.

7.2 Transition Models in Risk-based SHM

The approach proposed in Chapter 5 and demonstrated in Chapter 6 facilitates decision-making in the context of SHM by incorporating aspects of probabilistic risk assessment into a probabilistic graphical model framework. An underlying assumption of the decision framework presented previously, that facilitates the modelling process, is that structures can be represented as a hierarchical combination of discrete substructures/regions [99]. A consequence of this assumption is that the health-states of interest are all represented as discrete random variables, hence, the transition models required are matrices. For a given decided action a , and assuming a finite number N of possible discrete global health states, the conditional probability table $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d_t = a)$ is given by an $N \times N$ square matrix. In such a matrix, the i, j^{th} entry is the probability of transitioning from the i^{th} to the j^{th} health state and $i, j \in \mathbb{Z} : 1 \leq i, j \leq N$. Additionally, it is currently assumed that the Markov decision process is time-homogeneous, i.e. $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d_t = a)$ is invariant with respect to t . This assumption implies that, given no intervention is made ($d_t = \text{'do nothing'} \forall t$), the future global structural health-state is forecast as,

$$P(\mathbf{H}_{t+n}) = P(\mathbf{H}_t) \cdot P(\mathbf{H}_{t+1}|\mathbf{H}_t, d_t = \text{'do nothing'})^n \quad (7.2)$$

where n is the number of discrete time-slices forecast over, and $P(\mathbf{H}_t)$ and $P(\mathbf{H}_{t+n})$ are $1 \times N$ multinomial probability distributions over the global health-states at times t and $t + n$, respectively. For conciseness and clarity, the conditional dependence of the transition process on environmental and operational conditions is omitted.

7.3 Generating Transition Models

As with the established paradigm for conducting an SHM campaign (detailed in [1]), the risk-based approach is formed of several distinct stages. The risk-based approach consists of: operational evaluation, failure-mode modelling, decision modelling, data acquisition, feature selection and statistical modelling. Most crucial to the development of transition models is the operational evaluation stage. The current section outlines the information that must be obtained for the development of transition models, provides discussion around the quantification of the uncertainty in operational conditions and offers an explanation of how the quantified uncertainty may be used in conjunction with a physics-based model to develop transition models.

7.3.1 Operational Evaluation

As mentioned previously, the operational evaluation stage, for both the traditional and probabilistic risk-based structural health monitoring paradigms, seeks to assess the context in which a structural health monitoring campaign is to be conducted. It is during this stage that the operational and environmental conditions for the structure of interest are considered. Furthermore, failure modes of interest are determined and key health-states of the structure identified.

For the development of transition models in the probabilistic risk-based approach, during the operation evaluation stage, it is necessary to identify factors that will influence the way in which the structure will degrade. Many of these factors may be specific to the type of structure on which SHM is being conducted. Information regarding the operational conditions that must be obtained includes the anticipated forcing amplitudes, locations and temporal variations. These operational conditions will influence the fatigue life of the structure. Environmental conditions are also important to consider. Examples of important environmental factors include operating temperatures and the presence/absence of water. The anticipated operational temperature ranges are important to consider as these potentially introduce thermally-induced stresses in addition to other temperature effects on material properties such as fracture toughness. Furthermore, whether the structure will be in the presence of water is a key factor as this may introduce structural degradation mechanisms such as corrosion and erosion. An important consideration to make

when considering operational and environmental conditions is that degradation mechanisms may interact with one another. A notable example of this effect occurring is within the core of light-water nuclear reactors, where stainless steel structural components experience accelerated brittle fracturing as a result of interplay between multi-physical phenomena in a process known as irradiation-assisted stress corrosion cracking (IASCC) [102].

With the operational and environmental conditions of the structure considered and potential degradation mechanisms determined, the failure modes of interest for the structure and critical substructures, components and joints can be identified. Subsequently, it is important to define damage for each critical substructure, component and joint, i.e. the possible local health-states. As per the ‘classification’ stage of Rytter’s hierarchy, different components/joints may be susceptible to different types of damage, depending on factors such as materials and local operational and environmental conditions. For example, composite components may experience delamination whereas metallic components may experience fatigue cracking. For each component, criteria for each of the relevant failure mechanisms should be specified.

Irrespective of the type of damage associated with each component/joint, it is reasonable to assert that the discrete random variables corresponding to the local health states will have a cardinality of at least two. In the most simple case, each local health-state variable could possess states corresponding to ‘undamaged’ and ‘failed’, where the ‘failed’ state represents the component being unfit-for-purpose. In some scenarios, it may also be desirable to consider extents of damage (as indicated by Rytter’s hierarchy), and the functionality of the component/joint at varying damage extents. Some components/joints may possess health states associated with the presence of damage whilst continuing to function at their full, or partial capacity. Although these states are not necessarily associated with any immediate risk with regard to the failure of the global structure, they may still be important to consider as they may increase the propensity for transitioning to other more advanced damage states that do have high risk associated. An example of a component that may require this consideration is a load-bearing structural member in which partial thickness cracks may form.

7.3.2 Handling Uncertainty

For most applications of structural health monitoring, perfect knowledge of the operational and environmental conditions will not be available prior to the implementation of the system. It is for this reason that uncertainties should be considered and quantified where possible. While there exists a number of methodologies for the quantification of uncertainty, as discussed in Chapter 4, here, it is elected to continue using probability theory for consistency with the probabilistic risk-based decision framework.

For each of the key environmental and operation conditions, statistical distributions quantifying the ranges, likely values and/or variance in the conditions should be elicited from an expert judgement, and where possible, observed data. In a Bayesian setting, these distributions may be updated as measurements are collected, and the transition models re-estimated.

7.3.3 Generating Degradation Models

To generate the degradation transition models, physics-based models are used in the approach proposed in the current chapter. For SHM applications, the advantage of a physics-based model is that it can be used without prior data. A drawback of the approach is that the physics-based model will be in an unvalidated state. Nonetheless, in many ways, this is a natural approach as, prior to the implementation of an SHM system, the physics-based model can simply be considered as an informative prior. Updating and validation of transition models are discussed later in the current chapter.

The purpose of the model is to simulate the structure and specifically its critical components in each of the global health-states and under specified operational and environmental conditions. The simulated structure can then be evaluated with respect to the failure criteria identified in the operational evaluation stage, to determine whether state transitions occur.

With respect to modelling the degradation of a structure, the purpose of the physics-based model is to determine a distribution over the quantities of interest in which the failure criteria are specified, conditioned on the uncertain operational and environmental conditions. In the case that the physics-based model employed is inherently

stochastic (such as a probabilistic fracture mechanics model), this conditional distribution may be determined analytically. In the case that the physics-based model employed is deterministic (such as a finite-element model), this distribution may be determined by applying sampling methods to the probability distributions for the operational and environmental conditions, and querying the physics-based model accordingly.

Once a distribution over the quantities of interest has been determined, a distribution over local failure events can be produced by considering the logical operations defining the failure criteria. Again, this distribution is conditioned on the operational and environmental conditions. This conditional distribution over local failure events can then be mapped into transitions in the global health state by utilising the definition of \mathbf{H} as a vector containing the local health-states of the critical components, joints, and substructures.

At this stage, it is necessary to marginalise out the variable operational and environmental conditions to obtain the distribution $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d = 0)$. Additionally, to ensure a valid probability distribution is produced, normalisation should be conducted.

7.3.4 Generating Intervention Models

In addition to degradation models, transition models must include probability distributions specifying how structural health-states transition given each possible intervention in the decision space, i.e. $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d \neq 0)$. Naturally, developing transition models for specific actions (such as repairs) is a problem that is highly dependent on the context.

In many respects, the simplest intervention one can make to a structure is replacement of a component or components. For such interventions, it can be assumed that the replaced components return to an undamaged condition with high probability.

More nuanced interventions, such as repairs, are more challenging to model for reasons discussed in Section 7.4.2. In general, it can be assumed that maintenance interventions improve the health-state of a structure, though this is not necessarily always the case as it is possible an intervention may inadvertently compromise structural integrity. As with the degradation model, it is important that the intervention

models within the transition model reflect the uncertainties surrounding the outcomes. When modelling the effects of interventions on structural health, it is important to consider causality. Generally, it is desirable to know that a given intervention *causes* a change in the structural health, or operational, states, rather than simply being correlated with the change. If only a correlation is known, then it is feasible that an intervention may in fact be ineffective, with the effect being explained by confounding variables¹. For this reason, it may be necessary to learn probabilistic health-state transition models via causal inference techniques to precisely determine the effects of interventions [103, 104].

7.3.5 Case Study: Truss

To demonstrate how probability distributions quantifying uncertainty in operational conditions may be used in conjunction with a physics-based model to generate a transition model for a risk-based SHM decision process, the methodology was applied to the four-bay truss case study introduced in Chapter 6, and shown in Figure 6.1. For illustrative purposes, fictitious operational conditions were assumed.

To avoid obfuscating the development of the transition model, it was elected to ignore the failure of joints and the horizontal and vertical members and again focus on the failures of the cross-members. Denoting the local health states of the eight cross-members as hm_9 to hm_{16} , the global health state of the structure can be expressed as the vector $\mathbf{H} = \{hm_9, \dots, hm_{16}\}$. Additionally, for the purposes of demonstration, binary health-states for each of the eight cross-members were considered resulting in 256 possible global health states. Once again, a convenient referencing scheme for the global health-states is adopted where \mathbf{H} is given a superscript of the decimal number represented by the 8-bit binary number (with ascending powers of two from left to right) specified by the vector \mathbf{H} , i.e. the undamaged health state $\mathbf{H} = \{0, 0, 0, 0, 0, 0, 0, 0\}$ is denoted as \mathbf{H}^0 , and the health-state corresponding to the failure of the cross-members in the first bay $\mathbf{H} = \{1, 0, 0, 0, 1, 0, 0, 0\}$ is denoted as \mathbf{H}^{17} .

Finally, a binary decision d was considered for the structure, with possible courses of action ‘do nothing’ and ‘perform maintenance’; for conciseness, these actions will be denoted with $d = 0$ and $d = 1$, respectively. In this case study, it is assumed that the

¹Here, confounding variables refers to a variable that influences both an independent and dependent variable.

‘perform maintenance’ action is equivalent to the replacement of all cross-members with the structure consequently returned to its undamaged state.

7.3.6 Operational Conditions

Operational conditions were assumed for the structure such that the stress experienced in cross-members has a degree of stochasticity. Specifically, it was assumed that there would be uncertainty in both the load and the location that the load is applied to the structure at each time step. In addition to the variable load, a constant preload of 5 kg was applied to the structure at point P.

The magnitude of the load w was assumed to vary in accordance with the discrete uniform distribution,

$$w \sim \mathcal{DU}(0, w_{max}; n) \quad (7.3)$$

where w_{max} was determined such that $P(\mathbf{H}_{t+1}^0 | \mathbf{H}_t^0, d_t = 0) = 0.8$ and each load magnitude had probability of $P(w) = \frac{1}{n}$ with $n = 100$.

The position of the load was also assumed to vary according to a discrete uniform distribution over eight candidate locations labelled L1 to L8 in Figure 6.1. This distribution may be formalised as:

$$L \sim \mathcal{DU}(1, 8) \quad (7.4)$$

Hence, the operational conditions can be summarised as a vector $\mathbf{c}_o = \{w, L\}$. In total, 800 possible operational conditions were considered.

7.3.7 Failure Criteria

For each cross-member, three modes of failure were considered; yielding under tension, buckling under compression, and supercritical crack growth.

A cross-member was considered to have failed by yielding, if the tensile stress in the member exceeded the ultimate tensile stress of aluminium, where $\sigma_{UTS} = 300$ MPa.

The event of a cross-member m_i failing via yielding is denoted as Y_i .

A cross-member was considered to have failed by buckling when the compressive stress within a member exceeded the buckling stress σ_b . The critical buckling stress for a slender beam is given by the following equation [105],

$$\sigma_b = \frac{\pi^2 EI}{A(KL)^2} \quad (7.5)$$

where E is the Young's modulus, I is the cross-sectional second moment of area, A is the cross-sectional area, K is the effective length factor and is dependent on the boundary conditions, and L is the length of the member. As the truss was constructed in a way that allows in-plane rotation at the ends of each member, a pinned-pinned boundary condition was assumed, resulting in an effective length factor of $K = 1$. Taking the Young's modulus of aluminium to be $E = 70$ GPa, the critical buckling stress was found to be a compressive stress of $\sigma_b = 270$ MPa. The event of a cross-member m_i failing via buckling is denoted as B_i .

The final failure method considered for the cross-members was supercritical crack growth. For this failure mechanism, it was assumed that each member possessed a crack in the centre across the entire width of the member and at the midpoint along the length with probability 0.1. The size of the crack in meters was assumed to be continuous uniformly distributed according to,

$$2a \sim \mathcal{U}(0, b) \quad (7.6)$$

where $2a$ is the crack size and $b = 0.0125$ is the half width of the cross-members.

Assuming the cross-members can be modelled as a finite plate and with plane strain conditions, the mode I stress intensity factor K_I for a cracked member can be given by the following equation [106],

$$K_I = G\sigma\sqrt{\pi a} \quad (7.7)$$

where σ is the applied stress, and G is a geometric factor given by,

$$G = \frac{1 - \frac{a}{2b} + 0.326(\frac{a}{b})^2}{\sqrt{1 - \frac{a}{b}}} \quad (7.8)$$

A cracked cross-member was considered to have failed when the stress intensity factor exceeded the critical stress intensity factor K_c . For the aluminium members, it was taken that $K_c = 24 \text{ MPa} \cdot \text{m}^{\frac{1}{2}}$. The event of a cross-member m_i failing via supercritical cracking is denoted as C_i .

The initial variable structural conditions can be summarised in a vector $\mathbf{c}_s = \{2a_9, \dots, 2a_{16}\}$, where $2a_i$ is the crack length present in cross-member m_i . Here, it should be noted that the \mathbf{c}_s is considered independently of \mathbf{H} .

7.3.8 Transition Modelling

As previously mentioned, the truss finite element model was used. The finite element model was validated with a set of strain measurements taken from the physical truss in its undamaged condition.

A wrapper function was produced to iterate over the global health states \mathbf{H}_t . Additionally, the function was used to generate random samples \mathbf{c}^* from the probability distributions specifying the uncertain operational and structural conditions $\mathbf{c} = \{\mathbf{c}_o, \mathbf{c}_s\}$. Afterwards, the function queried the finite element model to obtain the stresses in the cross-members for the given global health-state and a random sample of operational and structural conditions.

Asserting $d = 0$, for an initial global health-state \mathbf{H}_t and a randomly-sampled set of conditions \mathbf{c}^* , a health-state transition was defined as $\mathbf{H}_{t+1} = \mathbf{H}_t + \delta\mathbf{H}$ where $\delta\mathbf{H} = \{\delta hm_9, \dots, \delta hm_{16}\}$ is an 8-bit binary vector and,

$$\delta hm_i = \mathbb{1}[(Y_i \vee B_i \vee C_i) | \mathbf{H}_t, d = 0, \mathbf{c}^*] \quad (7.9)$$

where $\mathbb{1}$ denotes the indicator function and \vee denotes the inclusive-or logical operator. Here, equation (7.9) corresponds to evaluating cross-member failures with respect to the previously-discussed criteria for yielding, buckling and cracking. Subsequently, the conditional probability of transitioning from \mathbf{H}_t^i to \mathbf{H}_{t+1}^j given \mathbf{c}^* was specified such that,

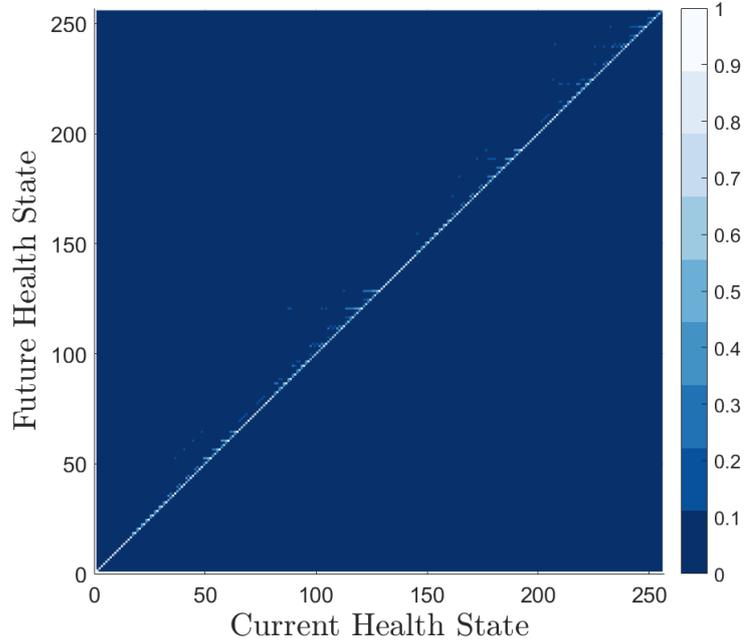


Figure 7.1: A heatmap showing the transition matrix $P(\mathbf{H}_{t+1} | \mathbf{H}_t, d = 0)$.

$$P(\mathbf{H}_{t+1}^j | \mathbf{H}_t^i, d = 0, \mathbf{c}^*) = \begin{cases} 1 & \text{if } \delta \mathbf{H} = \mathbf{H}_{t+1}^j - \mathbf{H}_t^i \\ 0 & \text{otherwise.} \end{cases} \quad (7.10)$$

To populate the transition matrix $P(\mathbf{H}_{t+1} | \mathbf{H}_t, d = 0)$, the variability in the conditions \mathbf{c} must be marginalised out and the distribution normalised; this was achieved by calculating the i, j^{th} entry of the transition matrix as,

$$P(\mathbf{H}_{t+1}^j | \mathbf{H}_t^i, d = 0) = \frac{\sum_1^{N_s} P(\mathbf{H}_{t+1}^j | \mathbf{H}_t^i, d = 0, \mathbf{c}^*)}{N_s} \quad (7.11)$$

where N_s is the number of queries of the finite-element model per \mathbf{H}_t .

The transition model for the action corresponding to ‘do nothing’ was estimated with the described procedure using $N_s = 10^4$. The heatmap of the resulting transition matrix $P(\mathbf{H}_{t+1} | \mathbf{H}_t, d = 0)$ is shown in Figure 7.1. A dominant lighter colour line can be seen along the diagonal in Figure 7.1; this indicates that the structure has a tendency to remain in the same health-state over a single time-step. Furthermore, it can be seen that the elements in the lower-right triangle of the graph (which

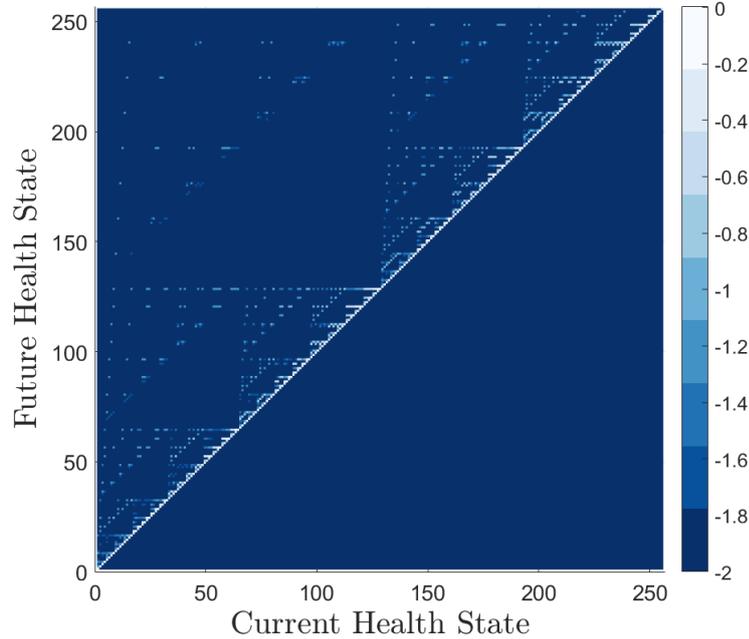


Figure 7.2: A heatmap showing the log probability of the transition matrix with an offset, $\log_{10}(P(\mathbf{H}_{t+1}|\mathbf{H}_t, d = 0) + 0.01)$.

corresponds to the lower-left triangle of the transition matrix) consists entirely of zero elements; a result of the implicit constraint imposed through Equations (7.9) and (7.10) that the structure monotonically degrades. Taking the \log_{10} of the conditional probability distribution (with an offset of $+0.01$ so that zero elements may be plotted with finite values) reveals further structure in the transition matrix as lower probability transitions are made more visible, as can be seen in Figure 7.2. Figure 7.2 shows that the transition matrix has fractal pattern akin to the Sierpiński triangle. Because the global health-state is represented as an 8-bit binary vector, the set of all allowable transitions assuming only monotonic degradation (i.e. once a bit is ‘turned on’, it cannot be ‘turned off’), form a Sierpiński triangle [107]. The possible transitions shown in Figure 7.2 are a subset of the Sierpiński triangle with some elements missing because of physical effects disallowing some transitions; for example, if the truss were to collapse because of the failure of the first bay, then the members in the other bays would no longer be able to fail as the structure would cease to support the load.

For completeness, the transition matrix for the ‘perform maintenance’ action $P(\mathbf{H}_{t+1}|\mathbf{H}_t, d = 1)$, was specified by making the assumption that the replacement of

all cross-members returns the structure to the undamaged health states, as shown by the following function:

$$P(\mathbf{H}_{t+1}^j | \mathbf{H}_t, d = 1) = \begin{cases} 1 & \text{for } j = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (7.12)$$

The current section has demonstrated a methodology for developing a health-state transition model for a structure by means of a case study. The next steps would be to evaluate and test the transition model, though this is omitted here as it is outside the scope of the current chapter.

7.4 Discussion

The current sections aims to highlight and discuss the importance of health-state transition models in the context of risk-based decision-making for SHM and for the specific problem of prognosis. Additionally, discussion will be made around the challenges associated with the development of the transition models.

7.4.1 Importance of Transition Models

In general, when it comes to decision-making, possessing information or beliefs regarding future events/states is crucial. This statement becomes most apparent when taking this notion *ad extremum*. At one extreme, if one possesses no information or belief regarding future events/states, then there is no reason for one to expect that any single course of action is better than any other. At the other end, if one somehow becomes clairvoyant and possesses perfect information regarding future events/states, then it follows that one would be able to make perfect decisions such that maximum rewards may be reaped.

As it happens, almost all decision problems, including those pertaining to SHM, fall somewhere between these two extremes, where belief and partial knowledge regarding future events/states is possessed. Nonetheless, in the context of SHM, increased expected utility gain provides a strong argument for striving towards improved knowledge regarding future health states by the development of transition models.

In addition to allowing closer-to-optimal decisions to be made within the risk-based framework, a good transition model allows for a pseudo-prognosis for the structure to be made by utilising equation (7.2). By propagating the belief in the current health-state forward in time according to equation (7.2), and by evaluating the risk of failure associated with the predicted distribution over future health-states, at each time step until the risk exceeds the cost of one of the candidate courses of action, one can obtain an estimate for the anticipated number of time-steps until an action should be taken. Whilst this result is not as powerful as a true-prognosis that yields remaining useful life, this information is still beneficial, as it provides the expected time available to execute a course of action.

7.4.2 Challenges

There are numerous challenges associated with the development of transition models.

A primary challenge pertains to the validation of transition models. For many applications of SHM, the monitoring campaign will be for a newly-built structure from which data are yet to be acquired at the time that the transition model must be developed. Without any observed state transitions to validate the model, one must rely solely on prior knowledge of the underlying physics that govern the degradation. One possible option is to independently validate the physics-based models used to develop the transition model via hybrid testing [108, 109], or performing experiments on individual components or substructures. Alternatively, in situations where an SHM system is being retrofitted to an existing structure, there may be data from historical inspections detailing health-state transitions that may be used to validate the degradation model. Finally, it is worth reiterating here that the transition model should be updated using data obtained throughout the monitoring campaign; a Bayesian approach for achieving this is demonstrated in [110].

The issue of validation is further complicated if the structure of interest is unique. For such a structure, even in a scenario where one is able to update the transition model with observed state transitions, it is possible, and in many cases likely, that only a small subset of the total possible state transitions will be observed throughout the operational lifetime; thereby leaving potentially large portions of the transition model without validation. In the context of population-based SHM [111–113], a single transition model may be applied to all members of a fleet of homogeneous

structures and also updated with state transitions observed from each instance of the structure. This idea is discussed further in Chapter 10.

Another challenge is the cost, both in terms of money and time, associated with the development of transition models. The development cost of a transition model will depend highly on the complexity of the structure for which a model is being developed, and the range of operational and environmental conditions that must be considered. For complex structures, the high-fidelity models capable of the multi-physics that may be required to simulate all the necessary failure mechanisms to develop a transition model are expensive and time-consuming to develop, often requiring teams of highly-skilled engineers. The financial argument for the development of such models should be constructed and evaluated during the operational evaluation stage of the SHM process, taking into account whether the structure is of high-value, or safety-critical.

The computational cost of the development and implementation of the transitional model should also be considered. During the development of the transition model, it is possible that a physics-based model is queried numerous times. For complex structures, and high-fidelity models, these simulations require large computing times. As the number of influential operating and environmental conditions increases, the number of samples required to adequately cover the input space will also increase. Taking this factor into account, with the possibility that high-fidelity physics-based simulations may need to be queried many many times, the calculation of the transition models may have prohibitively-long computation times. A possible solution to this issue would be to use a surrogate model, where an interpolation function that is relatively cheap to query is trained on a subset of the outputs of the comparatively-expensive physics-based model [114].

Finally, a challenge pertaining to maintenance action transition models is left as an open topic for research and discussion. In a few (limited) cases, such as when repair corresponds to replacement of all failed components (as is assumed for the case study in the current chapter), it may be reasonable to assert that the structure returns to its original undamaged case. However, in general, for less extreme and more realistic approaches to structural repair, this does not hold. In fact, it is possible that the state to which the structure transitions was not considered during the original development of the transition model; in other words, an undamaged original structure may behave in substantially different manner to an undamaged repaired structure [115]. Here, the challenge lies with determining reasonable assumptions that allow

one to avoid redeveloping the transition model after every intervention, or to conceive of methods for adapting the health states considered within the risk-based decision framework.

7.5 Summary

The aim of the current chapter has been to present a general methodology for developing structural health-state transition models for use in a probabilistic risk-based decision framework for SHM. Using a four-bay truss for a case study, a degradation model in the form of a probabilistic transition matrix was developed by considering uncertain operational conditions in conjunction with a physics-based model. Finally, discussions were made focussing on the challenges with developing health-state transition models, but also on the importance of the models for both the risk-based decision framework, and their application to the problem of prognosis in SHM.

STATISTICAL CLASSIFIERS FOR DECISION-MAKING

As discussed in Chapters 5 and 6, one can approach decision-making in SHM by adopting a probabilistic risk-based framework [99], in which failure events and decidable actions - like maintenance - are assigned costs and utilities, respectively. Decisions are made so as to maximise expected utility gain or minimise expected utility loss. In accordance with [1], the framework presented in Chapter 5 utilises a statistical pattern recognition (SPR) approach to damage detection and localisation in the inference of structural health states. The current chapter examines the role of statistical classifiers within SHM decision processes and proposes a risk-based active-learning algorithm for the development of classifiers with explicit consideration for decision-making performance.

8.1 Statistical Classifiers for SHM

A critical challenge associated with the learning of statistical classifiers when adopting an SPR approach to SHM, is a lack of labelled data corresponding to health-states of interest. This challenge arises in SHM, and high-value engineering asset management in general, as obtaining such a comprehensive dataset would require expensive testing programmes that would involve deliberately damaging the asset. This lack of labelled

data severely limits the use of supervised learning in SHM applications¹. While unsupervised methods may be used to find statistical patterns within data, such patterns are of limited use in decision-support, because of the lack of contextual information that would be provided by data labels. Several methods have been investigated as a means to overcome this challenge, including the use of physics-based models [116] and transfer learning [113, 117]. An alternative approach, that also enables the online development of classifiers, is *active learning*. An active-learning framework for SHM has been developed in [95], in which probabilistic classifiers direct the acquisition of new labelled data according to an uncertainty measure, given the current model. Within the active-learning framework for SHM, newly-acquired labelled data correspond to diagnostic information provided by an engineer, following an inspection of a structure. While this formulation provides a principled methodology for allocating inspection resources in a manner that optimises classification accuracy, in some scenarios it may be desirable/utility-optimal to consider the active learning of a classifier with respect to the context in which the classifier is being applied; supporting O&M decision-making.

Active learning has seen limited use in health and performance monitoring applications. A study applying active-learning methods to a generative model for predicting machining tool-wear is presented in [118]. Artificial neural networks with active sampling have been utilised for an image classification task to detect defects in civil structures [119]. In [120], a Bayesian convolutional neural network incorporating entropy-based active sampling is proposed as an approach for monitoring tools. Additionally, a particle filter-based damage progression model is aided by actively selected data in [121]. Previously, applications of active learning to SHM have all adopted an information-theoretic perspective. The novelty of the work presented in the current chapter lies with the formulation of the active-learning process from a decision-theoretic perspective. This novelty is achieved by applying active learning in the context of probabilistic risk-based SHM and considering the *expected value of perfect information* (EVPI) with respect to a maintenance decision process.

¹In some cases, when it can be assumed that a structure is undamaged initially, novelty detectors can be constructed in a supervised manner.

8.2 Machine Learning Paradigms

8.2.1 Supervised and unsupervised learning for SHM

In taking a data-based statistical pattern recognition approach to SHM, one employs machine-learning tools to learn patterns in data acquired from structures in order to infer information about structural health states such as the presence, location, and type of damage. For classification in general, the i^{th} measured data point $\mathbf{x}_i \in \mathcal{X}$ can be categorised according to a descriptive label $y_i \in \mathcal{Y}$ where y_i corresponds to the ground truth of the classification problem. For SHM, observations \mathbf{x}_i correspond to features extracted from the raw data acquired from a structure via signal processing, and the descriptive labels y_i relate to structural health-state information.

As aforementioned, probabilistic classifiers are desirable in SHM. For probabilistic classifiers, the features \mathbf{x}_i are defined as random vectors existing in a D -dimensional feature space $\mathcal{X} \in \mathbb{R}^D$. Additionally, the descriptive labels y_i are defined by a discrete random variable such that $y_i \in \mathcal{Y} = \{1, \dots, K\}$, where \mathcal{Y} is the label space and K is the number of classes required to uniquely identify the structural health states of interest.

Traditionally in SHM, classifiers are learned using one of two frameworks; *supervised* or *unsupervised* learning [81].

For a supervised classifier f , a mapping between the feature space and the label space is learned, i.e. $f : \mathcal{X} \rightarrow \mathcal{Y}$. Supervised learning requires a fully-labelled training-set \mathcal{D}_l such that [122],

$$\mathcal{D}_l = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathcal{X}, y_i \in \mathcal{Y}\}_{i=1}^n \quad (8.1)$$

for n collected data points. In the context of SHM, a fully-labelled training-set is often prohibitively expensive to obtain, or otherwise unavailable.

Conversely, unsupervised-learning techniques (e.g. k -means clustering [123]) may be applied when only unlabelled data are available and the training-set \mathcal{D}_u is of the form,

$$\mathcal{D}_u = \{\mathbf{x}_i | \mathbf{x}_i \in \mathcal{X}\}_{i=1}^m \quad (8.2)$$

for m collected data points. The issue with unsupervised techniques is that, without label information corresponding to the structural health conditions, the models learned are of limited usefulness for decision-making. This drawback arises as there is no context associated with the model and thus a related decision process cannot be specified. For a more in-depth discussion of the use of supervised and unsupervised learning in SHM, the reader is directed to [95].

8.2.2 Active Learning for SHM

Active learning is a form of *partially-supervised learning* [122]. Partially-supervised learning algorithms are characterised by their use of both labelled and unlabelled data, such that the dataset is,

$$\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u \quad (8.3)$$

Active-learning algorithms automatically query unlabelled data in \mathcal{D}_u to obtain labels allowing the labelled dataset \mathcal{D}_l to be extended. A generalised active learning heuristic is presented in Figure 8.1.

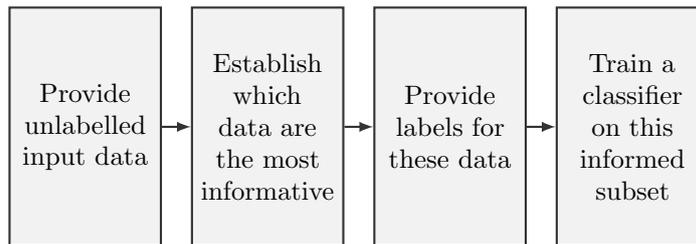


Figure 8.1: The general active-learning heuristic from [95].

The probabilistic active-learning framework for SHM developed in [95] details an approach built around a supervised probabilistic mixture model, trained and retrained on \mathcal{D}_l as it is extended via the active querying process. The approach presented in [95] uses measures of uncertainty to guide querying of incipient data points; specifically, preferentially obtaining labels for data points that have high entropy (information) [124], or low likelihood, given the current model.

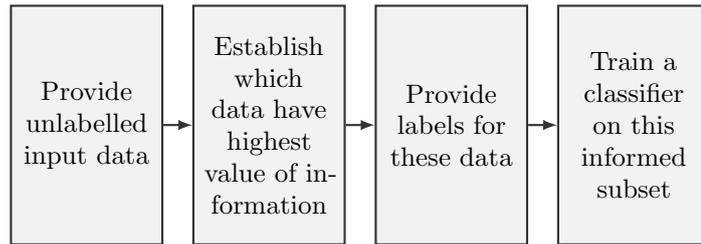


Figure 8.2: The general risk-based active learning heuristic.

After some thought, one can realise that the active-learning approach overcomes several of the challenges associated with supervised and unsupervised learning in SHM, as decision-making may be facilitated by the acquisition of class labels, whilst limiting the expenditure necessary to obtain them.

8.3 Risk-based Active Learning

The current chapter proposes a variation on the active-learning framework presented in [95] where, instead of using uncertainty measures to guide querying of data points according to their information or likelihood, incipient data are queried according to the expected value of perfect information with respect to the decision process, modelled using the risk-based SHM framework, in which the classifier is being applied. A generalised framework for risk-based active learning is presented in Figure 8.2.

8.3.1 Classifier Initiation

To begin the risk-based active learning process, one must initiate the classifier to be learned. Typically, this initialisation is done in a supervised manner with the available labelled data \mathcal{D}_l . In the case that there are no available labelled data, i.e. $\mathcal{D}_l = \emptyset$, and one has opted for a Bayesian learning approach, the classifier may be initiated using only the prior distribution.

Here, it is worth noting that the risk-based approach to active learning assumes that the number of classes (health states of interest) to be targeted by the model is known *a priori*. Within an uncertainty-based approach to active learning, one can infer the number of classes from data [95]. In contrast, for the risk-based approach to active learning, it is required that the classes targeted by the classifier correspond

to those represented in the decision process. The prescription of the target classes limits the flexibility of the classifier, while also facilitating the computation of value of information.

8.3.2 Value of Information

The expected value of perfect information (EVPI) is often understood as the price that a decision-maker should be willing to pay in order to gain access to perfect information regarding an otherwise uncertain or unknown state. More formally, EVPI can be defined as [73],

$$\text{EVPI}(d|X) := \text{MEU}(\mathcal{I}_{X \rightarrow d}) - \text{MEU}(\mathcal{I}) \quad (8.4)$$

where $\text{EVPI}(d|X)$ is the expected value of observing (with perfect information) a variable X before making a decision d , \mathcal{I} corresponds to an original influence diagram for a decision process involving a decision node d and a random variable node X , and $\mathcal{I}_{X \rightarrow d}$ corresponds to a modified influence diagram incorporating an additional edge from X to d . The MEU for the decision process modelled by \mathcal{I} is as follows,

$$\text{MEU}(\mathcal{I}) = \max_d \sum_{y \in \mathcal{Y}} P(y|d) \cdot U(y, d) \quad (8.5)$$

where Y is the subset of random variables in \mathcal{I} with utility functions specified by U .

The EVPI has the important characteristic that it is strictly non-negative (when disregarding the cost of making the additional observation), i.e. $\text{EVPI}(d|X) \geq 0$. This proposition can be realised by considering the conditional probability distributions over which the expected utility is optimised; the conditional probability distributions defined by the influence diagram \mathcal{I} are a subset of those defined by $\mathcal{I}_{X \rightarrow d}$. Furthermore, $\text{EVPI}(d|X) = 0$ if, and only if, the optimal policy for d in \mathcal{I} remains optimal for d in $\mathcal{I}_{X \rightarrow d}$; put simply, information has non-zero value when its possession results in a policy change.

In the context of an SHM decision process, the expected value of inspection can be represented as $\text{EVPI}(d_t|\mathbf{H}_t)$, where the influence diagram \mathcal{I} corresponds to the decision process where the health state of the structure is inferred only via the

observation of discriminative features with use of the statistical classifier. The modified influence diagram includes an additional edge from \mathbf{H}_t to d_t indicating the inspection of the structure within the discrete time-slice t . The modified influence diagram $\mathcal{I}_{\mathbf{H}_t \rightarrow d_t}$ is shown in Figure 8.3. In Figure 8.3, the edge highlighted in red denotes inspection, thus is not present in the original influence diagram \mathcal{I} . For the computation of the EVPI, it is assumed that inspection of the structure returns the ground-truth health state at the current time.

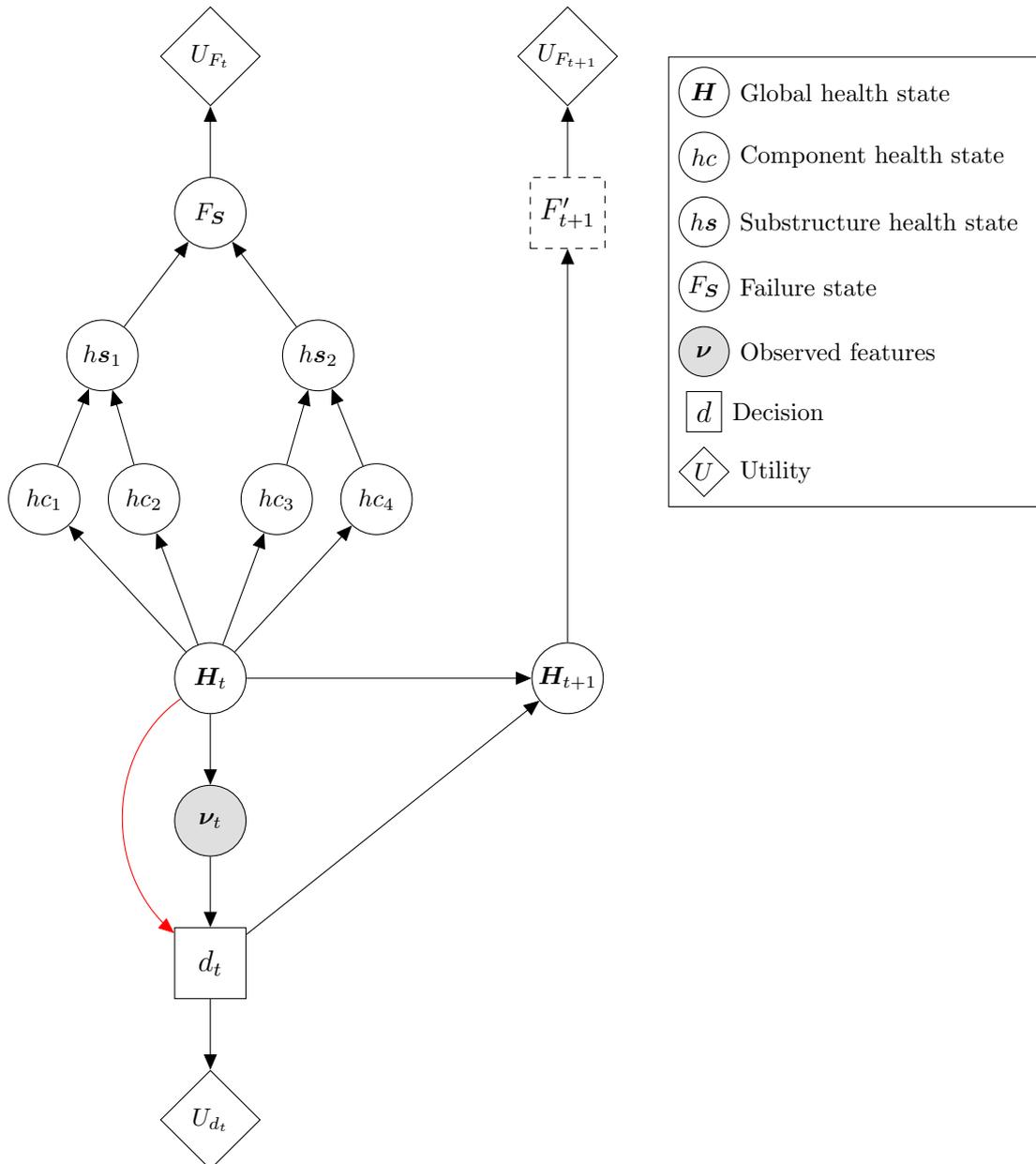


Figure 8.3: A modified influence diagram $\mathcal{I}_{\mathbf{H}_t \rightarrow d_t}$. The additional edge, not present in \mathcal{I} , is shown in red.

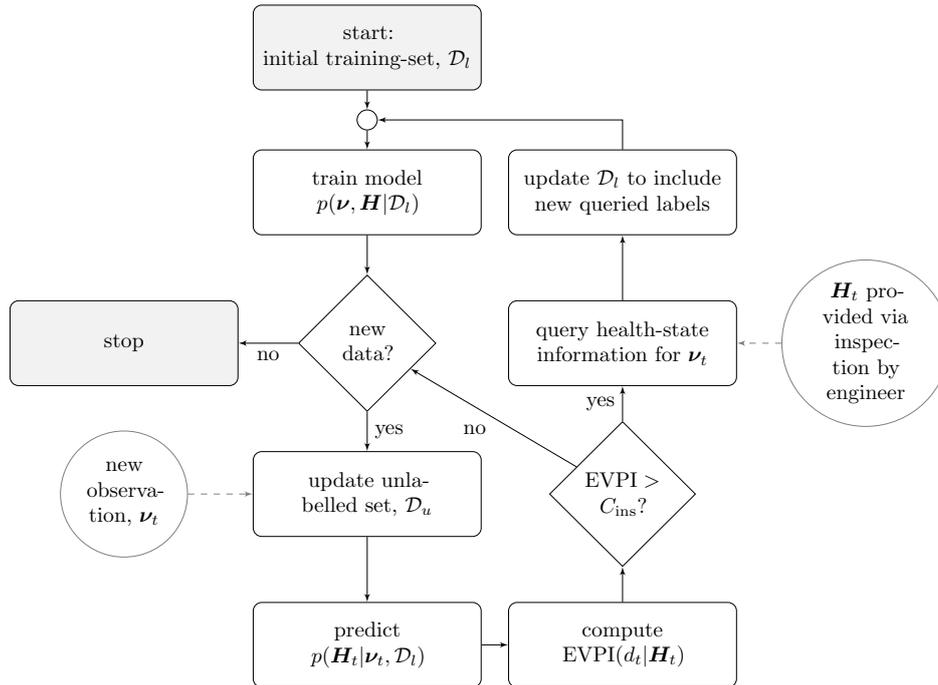


Figure 8.4: Flow chart to illustrate the risk-based active learning process for classifier development and inspection scheduling.

An example calculation of EVPI is provided in Section 8.4.3.

8.3.3 Inspection Scheduling

The EVPI of an unlabelled data point provides a convenient measure for determining whether a structure warrants inspection; if the EVPI for a data point ν_t exceeds the cost of inspection C_{ins} , the structure should be inspected prior to d_t and the corresponding health-state label for \mathbf{H}_t obtained. Subsequently, (ν_t, \mathbf{H}_t) can be incorporated into \mathcal{D}_l and the classifier retrained. The risk-based active-learning process for inspection scheduling and the development of statistical classifiers for risk-based SHM is shown in Figure 8.4.

8.3.4 Assessing Performance

Typically, classifier performance is evaluated using measures of classification accuracy – a popular choice being the f_1 -score. As the focus of this chapter is the development of classifiers in the context of decision-making, classification accuracy is of secondary

concern. Rather here, the ‘decision accuracy’ metric introduced in Chapter 6 is adopted for evaluating the more salient measure of decision-making performance.

By virtue of its similarities with classification accuracy, decision accuracy provides a simple and intuitive metric to assess decision-making performance and is therefore deemed appropriate for the current thesis. However, it is worth acknowledging here a limitation of ‘decision accuracy’ as a measure of performance. In the same way that classification accuracy weights false-positives (type-I errors) and false-negatives (type-II errors) equally, decision accuracy considers the decision-making equivalents to be of equal concern. Naturally, in many SHM applications, one may prefer unnecessary action over neglectful inaction so as to avoid catastrophic structural failures. To account for such nuances, one may opt to use a utility-based metric [125], or a weighted receiver operating characteristic (ROC) [126]. Additionally, these types of performance metric can also be extended to problems with non-binary decision domains and can reflect the relative preferability of candidate actions within such domains.

8.4 Numerical Example

To demonstrate risk-based active learning for SHM in a visual manner, the framework was applied to a representative numerical case study. Consider a structure S with four distinct health states of interest $H \in \{1, 2, 3, 4\}$:

- State 1 corresponds to the structure being undamaged and fully functional.
- State 2 corresponds to the structure possessing minor damage whilst being fully functional.
- State 3 corresponds to the structure possessing significant damage whilst operating at a reduced operational capacity.
- State 4 corresponds to the structure possessing critical damage and being non-operational, i.e. failed.

8.4.1 Decision Process

One can consider a decision process for the structure S , in which an agent at time t is tasked with making a decision d_t , such that some degree of operational capacity is maintained and the structure is not in the failed state at time $t + 1$. The influence diagram for such a decision process is shown in Figure 8.5.

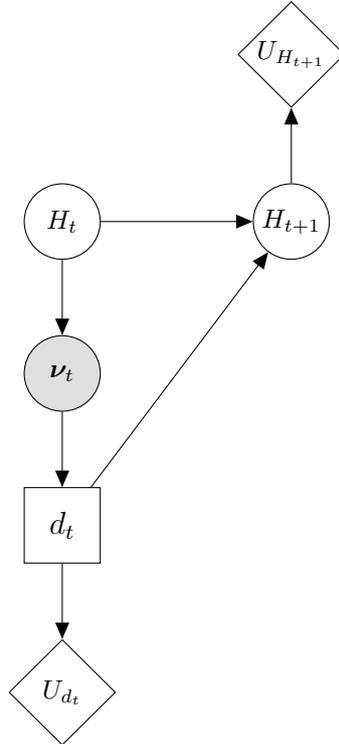


Figure 8.5: An influence diagram representation of the decision process associated with structure S .

Here, the decision d_t is a binary choice where, $d_t = 0$ corresponds to ‘do nothing’ and $d_t = 1$ corresponds to ‘perform maintenance’. The utilities associated with d_t are specified by the utility function $U(d_t)$ and are shown in Table 8.1. It is assumed that the ‘do nothing’ action has no utility associated with it, whereas the ‘perform maintenance’ action has negative utility relating to the expenditure associated with material and labour costs for structural maintenance. In practice, the specification of utility functions is non-trivial; however, may be achieved via expert elicitation during the operational evaluation stages of an SHM campaign. The elicitation of utility functions is an active area of research [89] but is beyond the scope of the current thesis. Hence, for the purposes of the current case study, the relative values of utilities are selected to be somewhat representative of the SHM context.

Table 8.1: The utility function $U(d_t)$ where $d = 0$ and $d = 1$ denote the ‘do nothing’ and ‘perform maintenance’ actions, respectively.

| | | |
|----------|---|-----|
| d_t | 0 | 1 |
| $U(d_t)$ | 0 | -30 |

For the ‘do nothing’ action $d_t = 0$, it is assumed that the structure monotonically degrades, with a propensity to remain in its current health state. This assumption is reflected in the CPD $P(H_{t+1}|H_t, d_t = 0)$ shown in Table 8.2.

Table 8.2: The conditional probability table $P(H_{t+1}|H_t, d_t)$ for $d_t = 0$.

| | | | | | |
|-------|---|-----------|------|-------|-------|
| | | H_{t+1} | | | |
| | | 1 | 2 | 3 | 4 |
| H_t | 1 | 0.8 | 0.18 | 0.015 | 0.005 |
| | 2 | 0 | 0.8 | 0.15 | 0.05 |
| | 3 | 0 | 0 | 0.8 | 0.2 |
| | 4 | 0 | 0 | 0 | 1 |

For the ‘perform maintenance’ action $d_t = 1$, it is assumed that the structure is returned to its undamaged health state with probability 0.99 and remains in its current state with probability 0.01. The conditional probability distribution $P(H_{t+1}|H_t, d_t = 1)$ is shown in Table 8.3.

Table 8.3: The conditional probability table $P(H_{t+1}|H_t, d_t)$ for $d_t = 1$.

| | | | | | |
|-------|---|-----------|------|------|------|
| | | H_{t+1} | | | |
| | | 1 | 2 | 3 | 4 |
| H_t | 1 | 1 | 0 | 0 | 0 |
| | 2 | 0.99 | 0.01 | 0 | 0 |
| | 3 | 0.99 | 0 | 0.01 | 0 |
| | 4 | 0.99 | 0 | 0 | 0.01 |

For brevity, within the current decision process it is assumed that utilities may be attributed directly to the future health state H_{t+1} of the structure without the need for modelling a specific failure mode. The utility function used for the current case study was specified so as to reflect the relative utility values that may be expected in a typical SHM application. The utility function $U(H_{t+1})$ is given in Table 8.4. Here, States 1 and 2, in which the structure S is fully functional, are assigned some positive utility; State 3 in which the structure is functional but with reduced capacity is assigned a lesser positive utility; and State 4, for which the structure is

non-operational, is assigned a relatively-large negative utility to reflect the loss of functionality and some additional severe consequence associated with the failure, e.g. risk to human life.

Table 8.4: The utility function $U(H_{t+1})$.

| H_{t+1} | 1 | 2 | 3 | 4 |
|--------------|----|----|---|-----|
| $U(H_{t+1})$ | 10 | 10 | 5 | -75 |

Finally, it is also assumed that the health state H_t may be inferred via the use of a statistical classifier by observing a set of discriminative features $\boldsymbol{\nu}_t = \{\nu_t^1, \nu_t^2\}$. The ground-truth health state at time t may be obtained via inspection at the cost of $C_{\text{ins}} = 7$.

8.4.2 Statistical Classifier

While the risk-based active learning algorithm is not restricted to any particular type of classifier, the statistical model employed for the current case study is one similar to that used in [95] - a mixture of four multivariate Gaussian distributions learned in a supervised manner from the initial labelled dataset \mathcal{D}_l . Each Gaussian component defines a generative model for the discriminative features $\boldsymbol{\nu}_t$, given each of four possible health states of interest in the domain of H_t ,

$$p(\boldsymbol{\nu}_t | H_t = k) = \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k) \quad (8.6)$$

where $\boldsymbol{\mu}_k$ and Σ_k are parameters of the multivariate Gaussian distribution corresponding to the mean and covariance, respectively. In addition to the mean and covariance parameters of the Gaussian components, the mixture model requires specification of $p(H_t)$; it is assumed that,

$$H_t \sim \text{Cat}(\boldsymbol{\lambda}) \quad (8.7)$$

where the categorical distribution is parametrised by a set of *mixing proportions* $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ such that,

$$P(H_t = k) = \lambda_k \quad (8.8)$$

and,

$$\sum_{k=1}^4 P(H_t = k) = \sum_{k=1}^4 \lambda_k = 1 \quad (8.9)$$

The parameters of the Gaussian mixture model that describe the generative statistical distribution $p(H_t, \boldsymbol{\nu}_t)$ can be summarised as,

$$\Theta = \{(\boldsymbol{\mu}_1, \Sigma_1, \lambda_1), \dots, (\boldsymbol{\mu}_4, \Sigma_4, \lambda_4)\} \quad (8.10)$$

In order to learn the parameters Θ from \mathcal{D}_l in a manner that avoids over-fitting, a Bayesian approach was adopted here. In this approach, the parameters Θ were treated as random variables with a prior placed over them. For conjugacy with the multivariate Gaussian distribution, a Normal-inverse-Wishart (NIW) prior was chosen such that,

$$\boldsymbol{\mu}_k, \Sigma_k \sim \text{NIW}(\mathbf{m}_0, \kappa_0, v_0, S_0) \quad (8.11)$$

where \mathbf{m}_0 , κ_0 , v_0 and S_0 are hyperparameters of the probabilistic mixture model. These hyperparameters can be interpreted in the following way [81]: \mathbf{m}_0 is the prior mean for each class mean $\boldsymbol{\mu}_k$, and κ_0 specifies the strength of the prior; S_0 is proportional to the prior mean for each class covariance Σ_k , and v_0 specifies the strength of that prior. The hyperparameters were specified such that each class H_t was initially represented as a zero-mean and unit-variance Gaussian distribution.

As a conjugate to the categorical distribution, a Dirichlet prior was placed over the mixing proportions $\boldsymbol{\lambda}$,

$$\boldsymbol{\lambda} \sim \text{Dir}(\boldsymbol{\alpha}) \quad (8.12)$$

and

$$p(\boldsymbol{\lambda}) \propto \prod_{k=1}^4 \lambda_k^{\alpha_k - 1} \quad (8.13)$$

where $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_4\}$ are hyperparameters of the mixture model. The hyperparameters were specified such that the prior probability of each class in the mixture model was $\frac{1}{4}$, i.e. each state is equally weighted.

Posterior estimates of the parameters may be calculated using the labelled dataset \mathcal{D}_l . As conjugate priors were used, updates of the parameters may be computed analytically to obtain the posterior NIW distribution given by [64],

$$\boldsymbol{\mu}_k, \Sigma_k | H_t = k, \mathcal{D}_l \sim \text{NIW}(\mathbf{m}_n, \kappa_n, v_n, S_n) \quad (8.14)$$

where $\mathbf{m}_n, \kappa_n, v_n, S_n$ are the updated parameters and are computed as follows,

$$\mathbf{m}_n = \frac{\kappa_0}{\kappa_0 + n_k} \mathbf{m}_0 + \frac{n_k}{\kappa_0 + n_k} \bar{\boldsymbol{\nu}}_k \quad (8.15)$$

$$\kappa_n = \kappa_0 + n_k \quad (8.16)$$

$$v_n = v_0 + n_k \quad (8.17)$$

$$S_n = S_0 + S + \kappa_0 \mathbf{m}_0 \mathbf{m}_0^\top - \kappa_n \mathbf{m}_n \mathbf{m}_n^\top \quad (8.18)$$

where n_k is the number of observations in \mathcal{D}_l with label k , $\bar{\boldsymbol{\nu}}_k$ is the sample mean of observations with label k , and S is the empirical scatter matrix given by the uncentered sum-of-squares for observations in class k , $S = \sum_{i \in \mathbb{I}_k} \boldsymbol{\nu}_i \boldsymbol{\nu}_i^\top$ where \mathbb{I}_k is the set of indices for observations with label k .

The posterior distribution of the mixing proportions $\boldsymbol{\lambda}$ remains Dirichlet, and is given by [64],

$$p(\boldsymbol{\lambda}|\mathcal{D}_l) \propto \prod_{k=1}^4 \lambda_k^{n_k + \alpha_k - 1} \quad (8.19)$$

To make class predictions for unlabelled data in \mathcal{D}_u , the posterior predictive distributions over the labels and observations can be obtained by marginalising out the parameters of the model. The posterior predictive distribution for unlabelled observations is obtained via the following marginalisation,

$$p(\boldsymbol{\nu}_t|H_t = k, \mathcal{D}_l) = \int \int p(\boldsymbol{\nu}_t|\boldsymbol{\mu}_k, \Sigma_k) p(\boldsymbol{\mu}_k, \Sigma_k|H_t = k, \mathcal{D}_l) d\boldsymbol{\mu}_k d\Sigma_k \quad (8.20)$$

resulting in the Student- t distribution [81],

$$\boldsymbol{\nu}_t|H_t = k, \mathcal{D}_l \sim \mathcal{T}(\mathbf{m}_n, \frac{\kappa_n + 1}{\kappa_n(v_n - D + 1)} S_n, v_n - D + 1) \quad (8.21)$$

where \mathbf{m}_n , κ_n , v_n , S_n are the updated hyperparameters and D is the dimensionality of the feature space. Here, the first two parameters of the Student- t distribution correspond to the mean and scale, respectively. The third parameter specifies the *degrees of freedom*. The full functional form of the Student- t distribution can be found in [81].

Similarly, the posterior predictive distribution over the labels is obtained via the following marginalisation,

$$p(H_t|\mathcal{D}_l) = \int p(H_t|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\mathcal{D}_l) d\boldsymbol{\lambda} \quad (8.22)$$

resulting in,

$$p(H_t = k|\mathcal{D}_l) = \frac{n_k + \alpha_k}{n + \alpha_0} \quad (8.23)$$

where $n = \sum_{k=1}^4 n_k$ and $\alpha_0 = \sum_{k=1}^4 \alpha_k$.

Finally, the predictive distribution for the class labels, given a new unlabelled observation $\boldsymbol{\nu}_t$, can be obtained using the posterior predictive distribution and applying Bayes' rule [95],

$$p(H_t = k | \boldsymbol{\nu}_t, \mathcal{D}_l) = \frac{p(\boldsymbol{\nu}_t | H_t = k, \mathcal{D}_l) p(H_t = k | \mathcal{D}_l)}{p(\boldsymbol{\nu}_t | \mathcal{D}_l)} \quad (8.24)$$

In summary, a Gaussian mixture model was trained in a supervised Bayesian manner on \mathcal{D}_l and subsequently retrained as \mathcal{D}_l was extended via the risk-based active querying process. The classifier developed allows a probability distribution over possible health states to be obtained following an observation of the discriminative features.

8.4.3 Example EVPI Calculation

To further elucidate the risk-based active-learning approach, an illustrative calculation of the EVPI is provided. This calculation is based upon the current case study as presented in Section 8.4.

Consider Figure 8.5 as the influence diagram of the original decision process \mathcal{I} . The corresponding modified influence diagram $\mathcal{I}_{H_t \rightarrow d_t}$ is shown below in Figure 8.6.

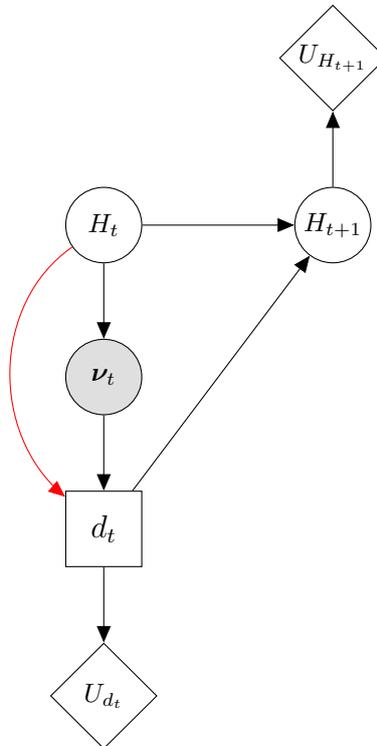


Figure 8.6: An influence diagram representation of the modified decision process associated with structure S .

With the EVPI given by equation (8.4), one must first calculate the MEU for the decision process represented by \mathcal{I} . The MEU for \mathcal{I} is given by,

$$\text{MEU}(\mathcal{I}) = \max_{d_t} \left[\sum_{H_t} \sum_{H_{t+1}} P(H_t|\boldsymbol{\nu}_t)P(H_{t+1}|H_t, d_t)U(H_{t+1}) + U(d_t) \right] \quad (8.25)$$

For the purposes of the demonstration, it will be assumed that the classifier presented in Section 8.4.2 has predicted $P(H_t|\boldsymbol{\nu}_t) = \{0.4, 0.3, 0.2, 0.1\}$, otherwise, the remaining conditional probability distributions and utility functions are as specified in Section 8.4. Employing equation (8.25), one can compute that $\text{MEU}(\mathcal{I}) = -4.4$.

Next, one must calculate the MEU for the decision process represented by the influence diagram $\mathcal{I}_{H_t \rightarrow d_t}$. This is given by,

$$\text{MEU}(\mathcal{I}_{H_t \rightarrow d_t}) = \sum_{H_t} \max_{d_t} \left[\sum_{H_{t+1}} P(H_t|\boldsymbol{\nu}_t)P(H_{t+1}|H_t, d_t)U(H_{t+1}) + U(d_t) \right] \quad (8.26)$$

Applying equation (8.26) and again taking $P(H_t|\boldsymbol{\nu}_t) = \{0.4, 0.3, 0.2, 0.1\}$, one determines $\text{MEU}(\mathcal{I}_{H_t \rightarrow d_t}) = 1.015$.

Finally, the EVPI of observing H_t prior to making the decision d_t , or the expected value of inspecting the structure, can be trivially calculated with equation (8.4) to be,

$$\text{EVPI}(d_t|H_t) = 1.015 - -4.4 = 5.415$$

As per the risk-based active-learning procedure outlined in Figure 8.4, calculated values of EVPI, such as the one presented above, can be compared to C_{ins} to determine whether an inspection is necessary.

8.4.4 Results

The complete dataset associated with the structure S in its various health states of interest $H_t \in \{1, 2, 3, 4\}$ was comprised of 1997 data points and is shown in Figure 8.7. The data were generated to be representative of typical changes in SHM feature

spaces as a result of progressive damage. One half of the data were randomly selected and set aside to form an independent test set, the remaining data were used to form the dataset \mathcal{D} .

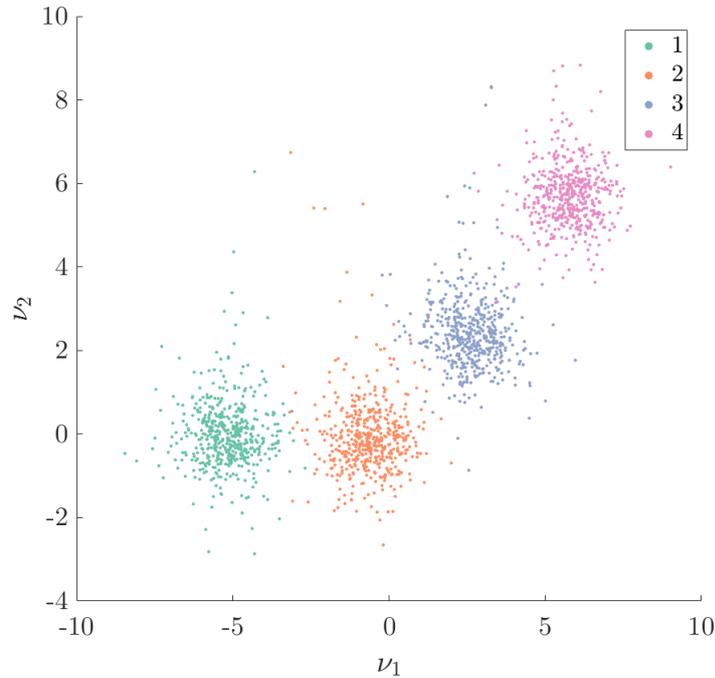


Figure 8.7: The 2-dimensional feature space of the complete dataset.

A small (1.5%) random subset of \mathcal{D} was annotated with the corresponding labels to initialise \mathcal{D}_l and the statistical classifier; it was ensured that at least one data point from each of the four classes was included in \mathcal{D}_l upon initialisation. The remaining data \mathcal{D}_u were left unlabelled to be sequentially presented to the decision model in the risk-based active-learning process. An example of an initial model learned from \mathcal{D}_l is shown in Figure 8.8 with data points in \mathcal{D}_l circled.

It can be seen from Figure 8.8 that, in general, the initial model fits the data poorly as a result of insufficient data. The best fitted Gaussian component of the mixture model is for $H_t = 2$; this is to be expected, since this cluster lies close to the zero-mean of the prior distribution. The EVPI across the feature space given the initial model is shown in Figure 8.9.

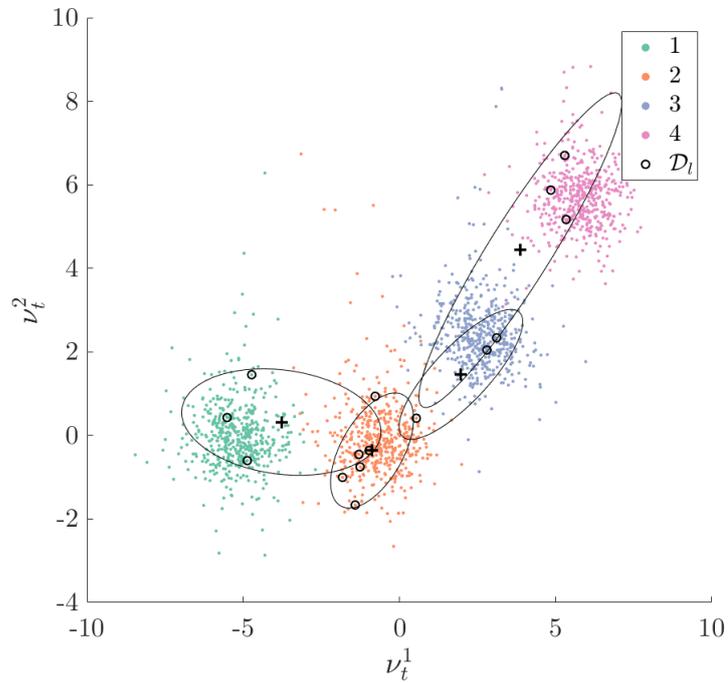


Figure 8.8: An initial statistical classifier $p(\nu_t, H_t, \Theta)$ learned from the original labelled dataset \mathcal{D}_t ; *maximum a posteriori* (MAP) estimate of the mean (+) and covariance (ellipses represent 2σ).

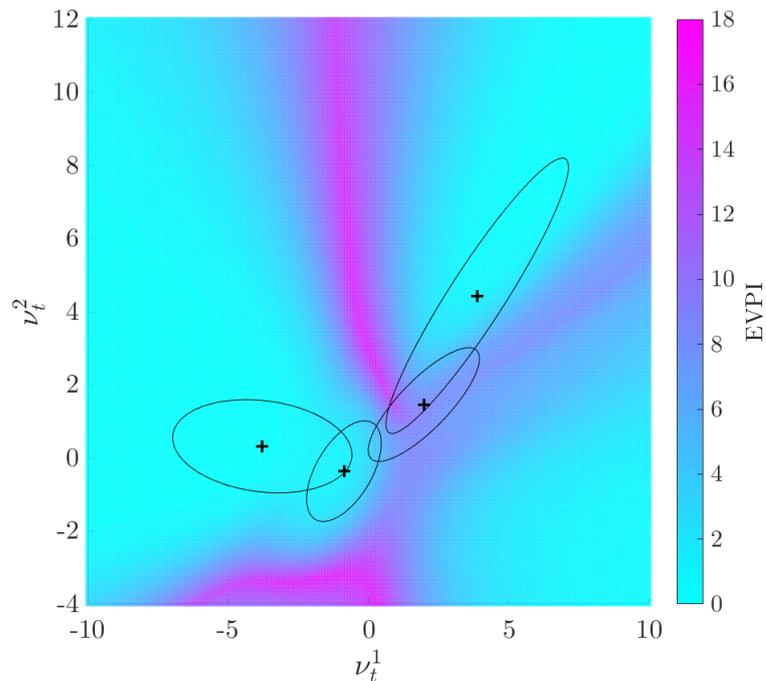


Figure 8.9: The EVPI over the two-dimensional feature space, given the initial model shown in Figure 8.8.

Figure 8.9 shows the regions of high and low value of information. At this stage, it is worth reminding oneself that the expected value of information arises when obtaining information results in a change in policy; areas of high EVPI correspond to regions in the feature space where there is classification uncertainty in the vicinity of decision boundaries given the current model. Bearing this in mind, one can justify the observation that the feature space around the learned Gaussian component for Class 1 (corresponding to the structure being undamaged) has low EVPI. Additionally, the region of the feature space in the overlap between the learned Gaussian distributions for Class 1 and Class 2 (minor damage), where there is uncertainty between the two classes, has *low* EVPI, as the optimal policy is unchanged regardless of whether the structure is in State 1 or State 2. On the other hand, areas with high EVPI correspond to regions of the feature space where there is classification uncertainty between a benign state and a more worrisome state; an example of this is the dominant vertical band of high EVPI feature space that corresponds to the set of points equidistant (Mahalanobis distance) from the learned clusters for State 1 and State 4.

The labelled dataset \mathcal{D}_l was extended according to the risk-based active learning process presented in Section 8.3. Data points ν_t in \mathcal{D}_u were considered in random order one-at-a-time and had their EVPI computed given the current model. A data point was queried, annotated with a label, and incorporated into \mathcal{D}_l if the data point met the condition $\text{EVPI} > C_{\text{ins}}$. After each query, the statistical classifier was retrained on the newly-extended \mathcal{D}_l . After being presented with each data point in \mathcal{D}_u one at a time, a final model trained on the fully-extended dataset in \mathcal{D}_l is shown in Figure 8.10.

It can be seen from Figure 8.10 that data in Classes 2 and 3 have been preferentially queried, particularly in the overlap between clusters; this is to be expected because of the associated classification uncertainty between states that warrant different maintenance policies. It can be seen that very few samples have been made for data points belonging to States 1 and 4, resulting in poorly-fitting learned distributions. This result can be explained by again considering the EVPI of the feature space shown in Figure 8.8, which shows low value of information in the regions around those data.

Figure 8.11 shows the EVPI over the feature space after the statistical classifier has been trained on the fully-extended version of \mathcal{D}_l . Similarities can be seen between Figure 8.9 and Figure 8.11 such as areas of low EVPI and the bands of high EVPI

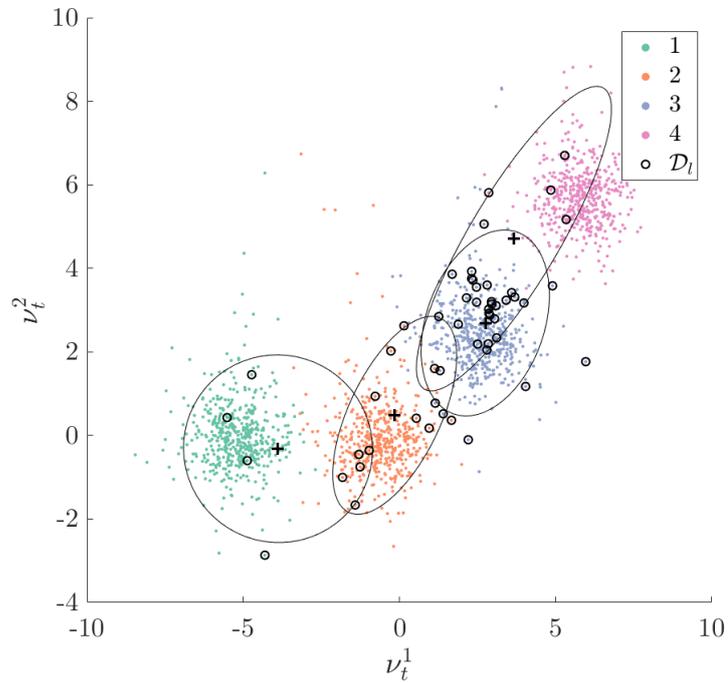


Figure 8.10: A final statistical classifier $p(\boldsymbol{\nu}_t, H_t, \Theta)$ learned from the extended labelled dataset \mathcal{D}_t ; *maximum a posteriori* (MAP) estimate of the mean (+) and covariance (ellipses represent 2σ).

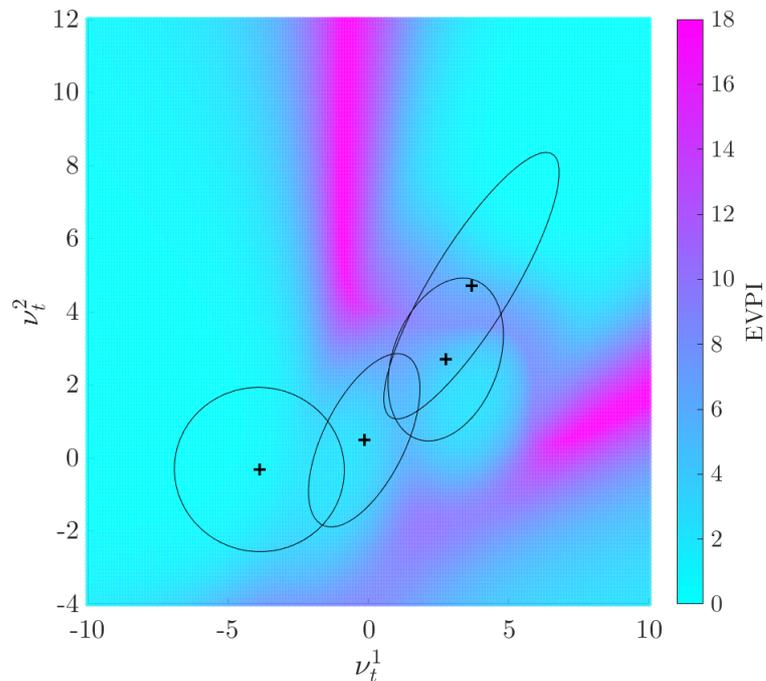


Figure 8.11: The EVPI over the two-dimensional feature space given the final model shown in Figure 8.10.

in regions equidistant from clusters corresponding to states that warrant differing maintenance actions. Figure 8.11 also shows a region of high EVPI that lies between the estimated means for the distributions learned for States 3 and 4; this is where the boundary between the structure being operational, and the structure being non-operational and requiring repair, lies. This result may indicate that the risk-based active learning approach preferentially incorporates data that strengthen decision boundaries.

To evaluate the change in classifier decision performance throughout the active learning process, the decision accuracy achieved on the independent test set was computed after each retraining of the classifier. The active learning process was repeated 1000 times using different random number generator seeds, so that the data selected to be in \mathcal{D} and the initial subset \mathcal{D}_l were randomly varied. The mean and standard deviation of the decision accuracy as a function of the number of queries is shown in Figure 8.12. The mean and standard deviation of the decision accuracy for a classifier trained on \mathcal{D}_l when extended with randomly-queried data points is also shown for comparison.

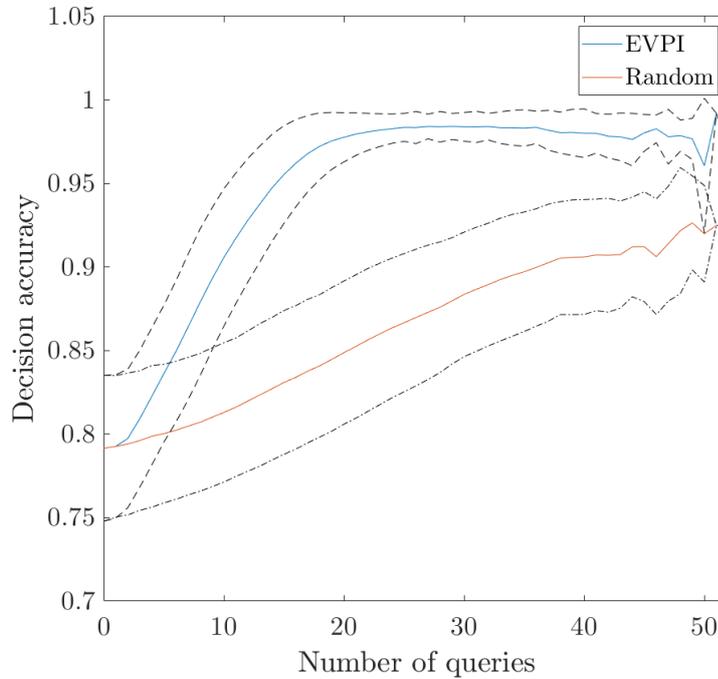


Figure 8.12: The variation in decision accuracy with number of label queries for an agent utilising a statistical classifier trained on \mathcal{D}_l extended via (i) risk-based active querying (EVPI) and (ii) random sampling (Random). The dashed lines show $\pm 1\sigma$.

The decision accuracy can be seen to increase more rapidly when guided by querying according to EVPI, rather than random sampling, which suggests that incorporating data labels with high expected value into a classifier will improve an agent's decision-making. After approximately 20 queries, the improvement in decision accuracy gained per query is greatly reduced. Upon close examination, it can be seen that after 28 queries, the decision accuracy begins to decrease slightly. This may be a result of sampling bias introduced via the active learning process and will be further discussed in Section 8.6. The variability in the decision accuracy can be seen to collapse for number of queries greater than 50; this result is simply due to the fact only one of the 1000 trials queried more than 50 times from the unlabelled data.

In summary, a risk-based approach to active learning was demonstrated for a representative numerical case study. A simple maintenance decision problem was formed for a structure with four key health states of interest; EVPI was used to trigger 'inspections' to obtain class-label information corresponding to structural health states.

8.5 Experimental Case Study

The risk-based active-learning framework was applied to a dataset obtained from the Z24 Bridge [127]. The Z24 bridge was a concrete highway bridge near Solothurn in Switzerland, between the municipalities of Koppigen and Utzenstorf. Prior to demolition, the bridge was the focus of a cross-institutional research project (SIMCES), the goal of which was to provide a benchmark and prove the feasibility of vibration-based SHM [128, 129]. The Z24 Bridge benchmark has been widely used for SHM and modal analysis applications. The monitoring campaign on the bridge spanned a period of 12 months, during which time the bridge was instrumented with sensors to capture both the dynamic response of the structure and environmental conditions including air temperature, deck temperature, humidity and wind speed [130].

From the dynamic response measurements acquired, the first four natural frequencies of the structure were obtained. The variations in these natural frequencies throughout the monitoring campaign are shown in Figure 8.13. The dataset contains 3932 observations of the natural frequencies in total. Towards the end of the campaign, incremental damage was introduced to the structure artificially, beginning at ob-

ervation 3476. Additionally, throughout the campaign, the bridge exhibited cold temperature effects, particularly prominent between observations 1200 and 1500. These effects are believed to be a result of the stiffening of the asphalt layer in the bridge deck induced by very low ambient temperatures.

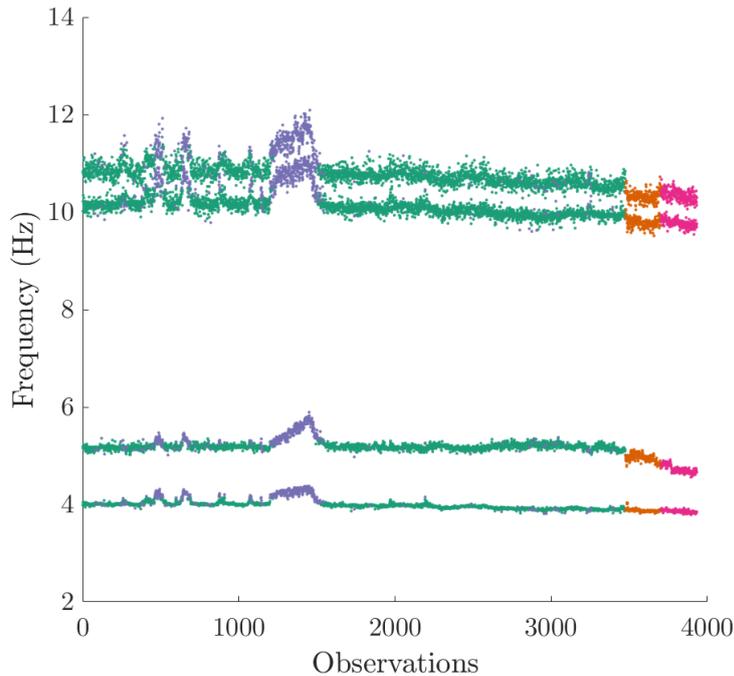


Figure 8.13: Time history of the first four natural frequencies for the Z24 bridge.

To define a classification problem on which to apply risk-based active learning, the first four natural frequencies of the bridge were selected as the discriminative features such that $\nu_t \in \mathbb{R}^4$. Furthermore, there are assumed to be four distinct classes of interest $H_t \in \{1, 2, 3, 4\}$:

- Class 1: normal undamaged condition (green)
- Class 2: cold temperature undamaged condition (blue)
- Class 3: incipient damaged condition (orange)
- Class 4: advanced damaged condition (pink).

Here, it has been assumed that the damaged data may be separated into two halves; the earlier half corresponding to incipient damage and the later half corresponding

to advanced damage. It is believed that this is a reasonable assumption, given the incremental addition of damage during the monitoring campaign [127]. Furthermore, it is supported by the unsupervised clustering achieved using the Dirichlet process in [96]. The ‘ground truth’ labels for the normal undamaged data are separated from the cold temperature effects using outlier analysis via the Minimum Covariance Determinant algorithm [131, 132].

8.5.1 Decision Process

Again, to undertake risk-based active learning, one must consider the decision process for the structure. As the bridge is long-since demolished, here, a similar decision process to that considered for the visual example presented in Section 8.4 will be considered. This decision process focusses on a binary decision d_t where $d_t = 0$ corresponds to ‘do nothing’ and $d_t = 1$ corresponds to ‘perform maintenance’. Because of the similar nature of the problems, the influence diagram shown in Figure 8.5 can be used to represent the decision process considered for the Z24 bridge. However, the utility functions and conditional probability distributions defined by the influence diagram must be altered to reflect the operational context of the bridge.

The utilities associated with the decision d_t , specified by the utility function $U(d_t)$, are presented in Table 8.5. It is assumed that the ‘do nothing’ action has zero utility, whereas the ‘perform maintenance’ action has negative utility. Once again, the utilities for the current case study have been selected such that the relative values are appropriate for demonstrating the risk-based active learning approach. As discussed earlier, more representative or exact utilities may be obtained with the aid of expert judgement.

Table 8.5: The utility function $U(d_t)$ for the Z24 bridge where $d = 0$ and $d = 1$ denote the ‘do nothing’ and ‘perform maintenance’ actions, respectively.

| | | |
|----------|---|------|
| d_t | 0 | 1 |
| $U(d_t)$ | 0 | -100 |

For the decision action ‘do nothing’, the state transitions are specified such that the structure monotonically degrades, with a propensity to remain in its current state. For the current case study, however, there is the subtlety that States 1 and 2 both correspond to the undamaged health state but under differing environmental

conditions. Here, the earlier assumption that there is a propensity to remain in the current state is exploited; simply, asserting that if the temperature is cold for any given measurement, it is more likely than not that the subsequent measurement will also be made at a cold temperature. The same reasoning is also applied to normal temperature conditions (if only weather forecasting really were so simple!). These assumptions are reflected in the conditional probability distribution $P(H_{t+1}|H_t, d_t = 0)$, shown in Table 8.6. Here, the assumed distributions are sufficiently representative to demonstrate the risk-based active-learning process. However, in practice, one may use degradation modelling, climate modelling and expert elicitation to develop these transition models. Moreover, it is worth acknowledging that, for some applications, it may be desirable to remove the effects of environmental and operational variables. This removal process has been demonstrated in [133].

Table 8.6: The conditional probability table $P(H_{t+1}|H_t, d_t)$ for $d_t = 0$.

| | | H_{t+1} | | | |
|-------|---|-----------|------|-------|-------|
| | | 1 | 2 | 3 | 4 |
| H_t | 1 | 0.7 | 0.28 | 0.015 | 0.005 |
| | 2 | 0.43 | 0.55 | 0.15 | 0.05 |
| | 3 | 0 | 0 | 0.8 | 0.2 |
| | 4 | 0 | 0 | 0 | 1 |

The counterpart to the conditional probability distribution shown in Table 8.6, $P(H_{t+1}|H_t, d_t = 1)$, specifying the state transition probabilities given $d_t = 1$ is shown in Table 8.7. This conditional probability distribution is based on the assumption that the ‘perform maintenance’ action returns the structure to one of the two undamaged states with high probability. The probabilities constituting this distribution, given that the structure is in one of the two undamaged states initially, are specified such that the future undamaged state is independent of the action being undertaken. This assumption is, of course, a natural one to make, if one reasons that (chaos theory aside) the act of repairing the bridge does not influence the weather. The remaining probabilities, conditional on the structure being in one of the two damaged states initially, are specified in a similar manner. Firstly, it is assumed that there is a small probability that the maintenance has no effect and the structure remains in its damaged state. Secondly, it is assumed that the remaining probability is attributed to each of the undamaged states in accordance with the stationary distribution obtained when considering the probability of being in normal or cold temperatures in the distant future. This assumption is made as the classes corresponding to damaged states do not distinguish between temperatures and therefore provide no information

regarding which of the undamaged states the structure is likely to be returned to. More formally, it is asserted that,

$$P(H_{t+1} = 1 \vee H_{t+1} = 2 | H_t = 3 \vee H_t = 4, d_t = 1) = P(H_{t+1} = 1 \vee H_{t+1} = 2 | H_t = 1 \vee H_t = 2, d_t = 1)^n \quad (8.27)$$

as $n \rightarrow \infty$, and where \vee is the OR logical operator. As it is implicitly assumed that the states H_t are mutually exclusive this is also equivalent to the XOR logical operator.

Table 8.7: The conditional probability table $P(H_{t+1} | H_t, d_t)$ for $d_t = 1$.

| | | H_{t+1} | | | |
|-------|---|-----------|--------|------|------|
| | | 1 | 2 | 3 | 4 |
| H_t | 1 | 0.7143 | 0.2857 | 0 | 0 |
| | 2 | 0.4388 | 0.5612 | 0 | 0 |
| | 3 | 0.5996 | 0.3904 | 0.01 | 0 |
| | 4 | 0.5996 | 0.3904 | 0 | 0.01 |

For simplicity, it is once again assumed that utilities may be attributed directly to the states of interest for this decision problem. The utility function $U(H_{t+1})$ is shown in Table 8.8. Here, the undamaged states associated with the bridge are assigned small positive utilities as reward for the bridge being functional with minimal risk of failure. The incipient damage state is assigned moderately-sized negative utility to reflect possible reduced functionality or low-to-moderate risk of failure. The advanced damage state is assigned a very large negative utility to reflect the high risk associated with the failure of the bridge.

Table 8.8: The utility function $U(H_{t+1})$.

| H_{t+1} | 1 | 2 | 3 | 4 |
|--------------|----|----|-----|-------|
| $U(H_{t+1})$ | 10 | 10 | -50 | -1000 |

Finally, the cost of inspection is specified to be $C_{\text{ins}} = 30$. This moderate cost reflects the time required to inspect a large-scale structure such as a bridge.

8.5.2 Statistical Classifier

Once again, a probabilistic Gaussian mixture model was used as the statistical classifier; trained repeatedly in a supervised manner as \mathcal{D}_l was extended via a risk-based querying process. As previously mentioned, the discriminative features used were the first four natural frequencies of the bridge, normalised with respect to the mean and standard deviations. The targets of the classifier were the four states of interest corresponding to two undamaged and two damaged states.

8.5.3 Results

As previously mentioned, the Z24 dataset comprises 3932 observations. Once again, these data were divided in half to form a training dataset and a test dataset. A small (1%) random subset of the training dataset \mathcal{D} was assigned to the initial labelled dataset \mathcal{D}_l , and the remaining data assigned to the unlabelled dataset \mathcal{D}_u to be sequentially presented to the decision model in the risk-based active-learning process. To facilitate visualisation, the four-dimensional feature space was projected down onto two-dimensions using *principal component analysis* (PCA) [134]. The two-dimensional projection of the Z24 dataset is shown in Figure 8.14.

An example of an initial model learned from the subset \mathcal{D}_l is shown in Figure 8.15 with data in \mathcal{D}_l circled. It can be seen from Figure 8.15, that it appears that the best-learned Gaussian distribution corresponds to the class $H_t = 1$. The reasons for this are twofold, the data for this class are positioned closest to zero-mean of the prior, and more datapoints belonging to this class were, by chance, included in the initial labelled dataset \mathcal{D}_l . The other three distributions appear to have been learned poorly, being heavily influenced by the prior.

Figure 8.16 shows the EVPI contours over the projected feature space. It can be seen that there are regions of high expected value around the edges of the normal undamaged cluster (Class 1); this is intuitively understood by considering the adjacent regions of low expected value. The low-value region bounded by the high-value region occurs as a result of the classifier being confident that the structure is in its undamaged condition and therefore the decision-maker is confident that the optimal decision is ‘do nothing’. Conversely, the low-value region enclosing the high-value region occurs as a result of the tolerable risk of damage/failure having

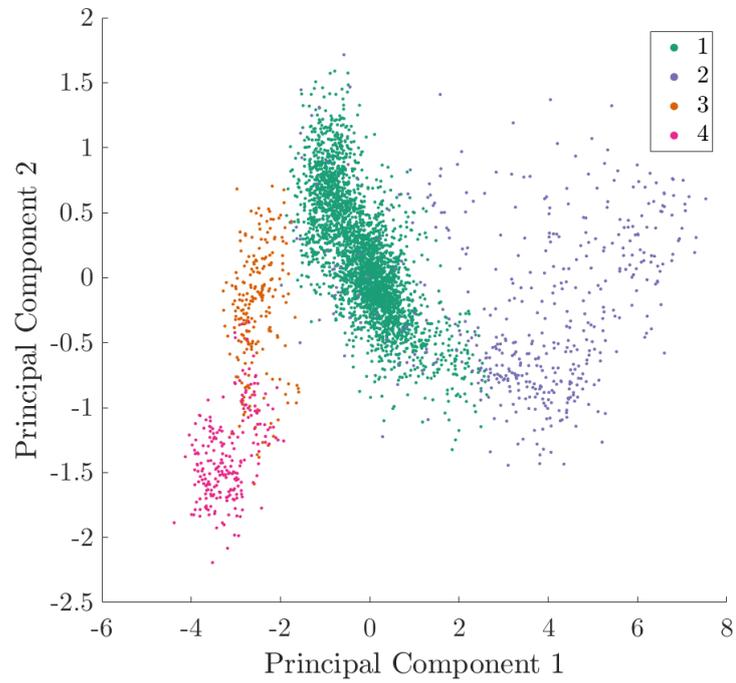


Figure 8.14: A two-dimensional PCA projection of the feature space for the complete Z24 dataset.

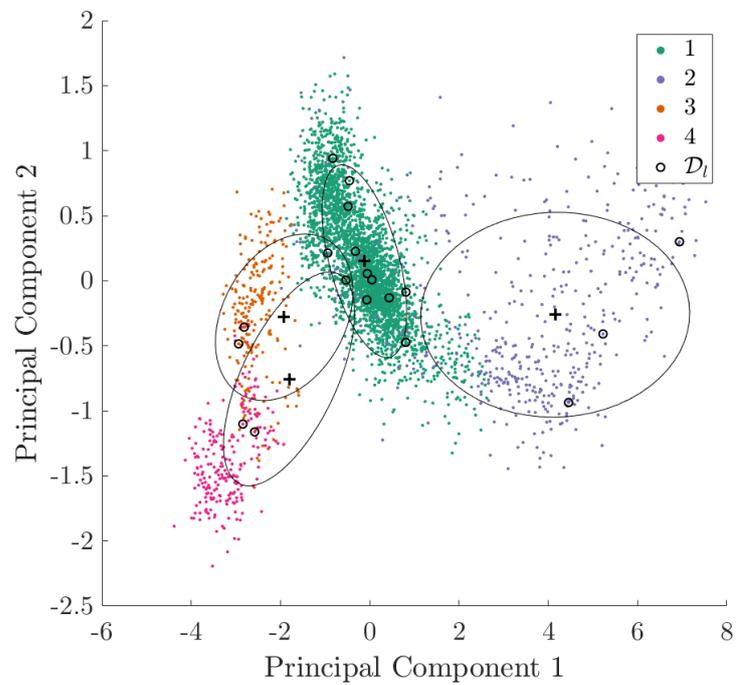


Figure 8.15: A two-dimensional projection of an initial statistical classifier $p(\nu_t, H_t, \Theta)$ learned from the initial labelled dataset \mathcal{D}_t ; *maximum a posteriori* (MAP) estimate of the mean (+) and covariance (ellipses represent 2-sigma).

been exceeded; as such, the decision-maker is confident that the optimal decision for data in this region is ‘perform maintenance’. Again, Figure 8.16 shows that there are larger swathes of high-value, corresponding to regions of high uncertainty between undamaged and damaged states.

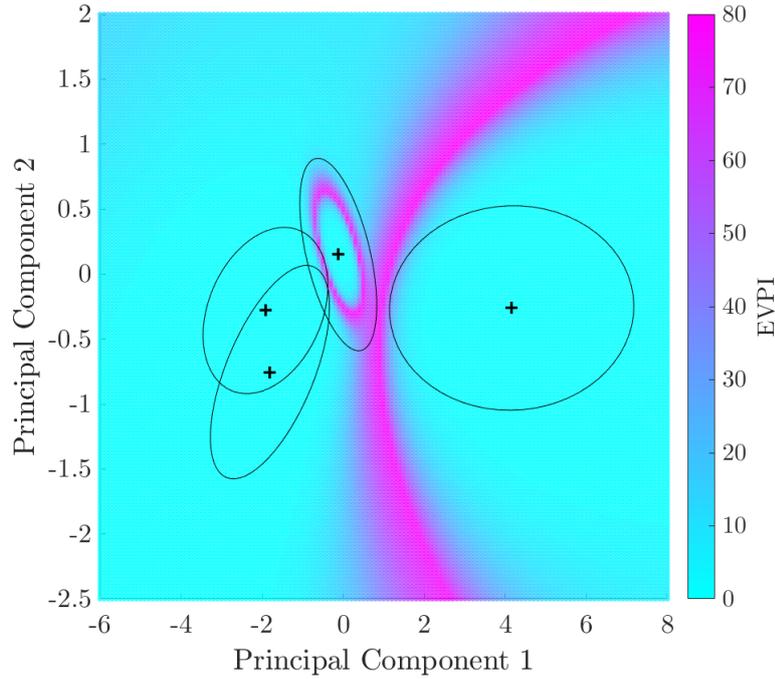


Figure 8.16: The EVPI over the two-dimensional PCA projection of the feature space given the two-dimensional projection of the initial model shown in Figure 8.15.

The labelled dataset \mathcal{D}_l was extended via the risk-based active learning process as data were presented to the decision process in sequential order, one-at-a-time. Again, labels for datapoints were queried and the classifier retrained on the extended dataset when the criterion $\text{EVPI} > C_{\text{ins}}$ was satisfied. The final model, corresponding to the updated version of the initial model shown in Figure 8.15 following risk-based active learning, is shown in Figure 8.17.

It can be seen in Figure 8.17, that the active learner has preferentially queried datapoints belonging to Class 1 and, to a lesser extent, Class 2. Data on the boundary between Classes 1 and 3 appear to have been particularly heavily sampled. This observation is, again, to be expected when considering the distribution of EVPI shown in Figure 8.16, which shows low-value regions in the vicinity of advanced-damage data and high-value regions on the boundary of Class 1.

Figure 8.18 shows the EVPI over the projected feature space given a projection of the

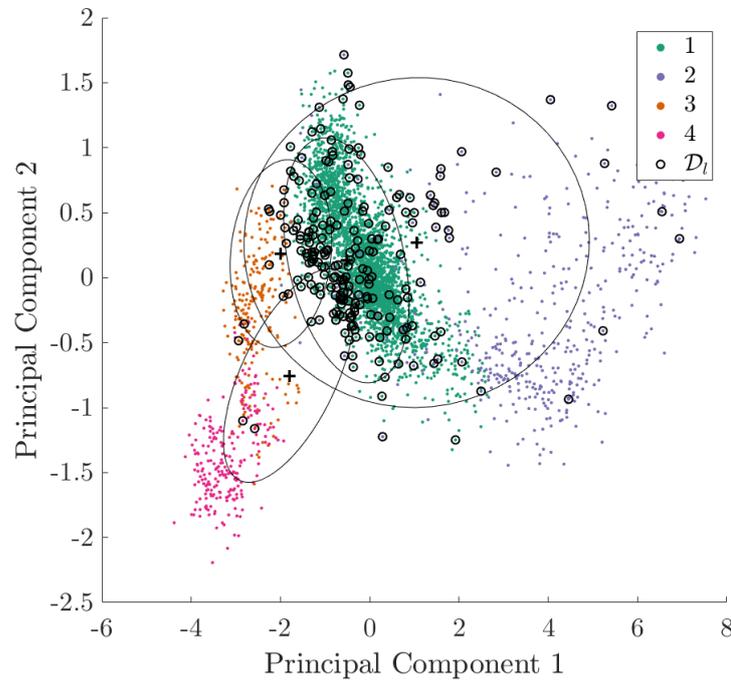


Figure 8.17: A two-dimensional PCA projection of a final model statistical classifier $p(\boldsymbol{\nu}_t, H_t, \Theta)$ learned from the extended labelled dataset \mathcal{D}_i ; *maximum a posteriori* (MAP) estimate of the mean (+) and covariance (ellipses represent 2-sigma).

final model. Figure 8.18 shows very well-defined ‘rings’ of high expected value. Once again, the areas inside these rings can be considered regions on the feature space where the classifier is sufficiently confident that the structure is currently, and will be in the subsequent time-step, in an undamaged state such that the decision-maker can be confident that ‘do nothing’ is the optimal action without the need for inspection of the structure. Likewise, areas outside of the rings correspond to regions of the feature space where the decision-maker may be confident that ‘perform maintenance’ is the optimal action.

By comparison of Figures 8.16 and 8.18, it can be seen that the swathes of high-value feature space are replaced with a ring of high-value, associated with the cold temperature undamaged class. Additionally, the ring associated with the normal undamaged class becomes tighter. These phenomena indicate that, with hindsight, one can deem an agent utilising the initial classifier to be over-confident in its decision-making. For the current case study, it is to be expected that the areas in which a decision-maker can be confident in the ‘do nothing’ action are small because of the large negative utility associated with the advanced damage class and the large

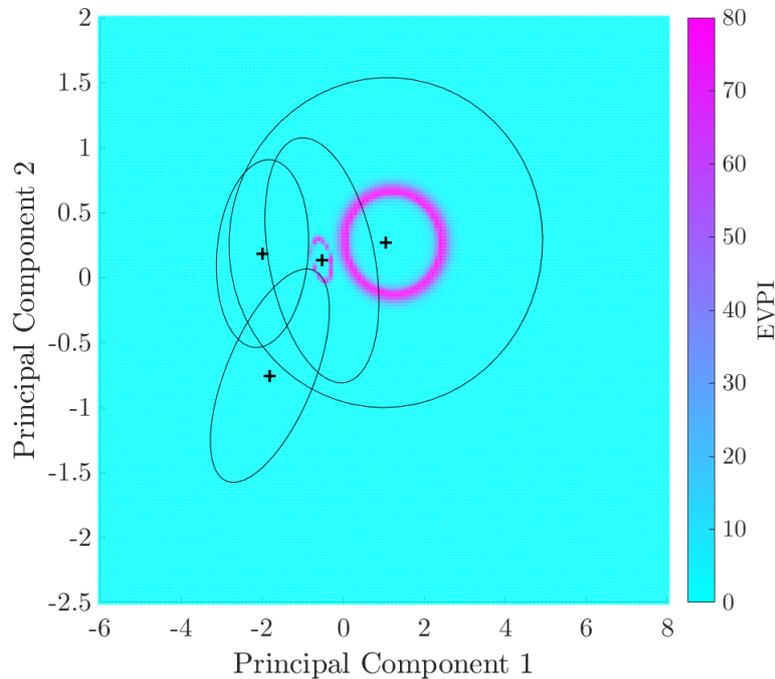


Figure 8.18: The EVPI over the two-dimensional PCA projection of the feature space given a two-dimensional projection of the final model shown in Figure 8.17.

resulting risk. In fact, if one reduces the cost associated with the bridge being in its advanced damage state, then one can expect the rings to grow larger as the region of tolerable risk expands. This can be observed in the merging of the rings shown in Figure 8.19. The final model shown in Figure 8.19 was learned with the cost associated with the advanced damage state reduced from 1000 to 500.

Throughout the active-learning process, the decision-making performance of an agent was evaluated by estimating the decision accuracy for data in the independent test set. The active-learning process was repeated 1000 times with differing random number generator seeds, such that the data in \mathcal{D} and the initial subset \mathcal{D}_i were randomly varied. The mean and standard deviation of the decision accuracy as a function of the number of queries is shown in Figure 8.20. For comparison, the decision accuracy for a classifier training on a labelled dataset extended via random sampling is also shown.

The decision accuracy can be seen to improve almost monotonically when using risk-based active learning, as opposed to random sampling which initially results in a degradation in performance before improving. For risk-based active learning, in addition to the mean value of decision accuracy reaching close to unity with fewer

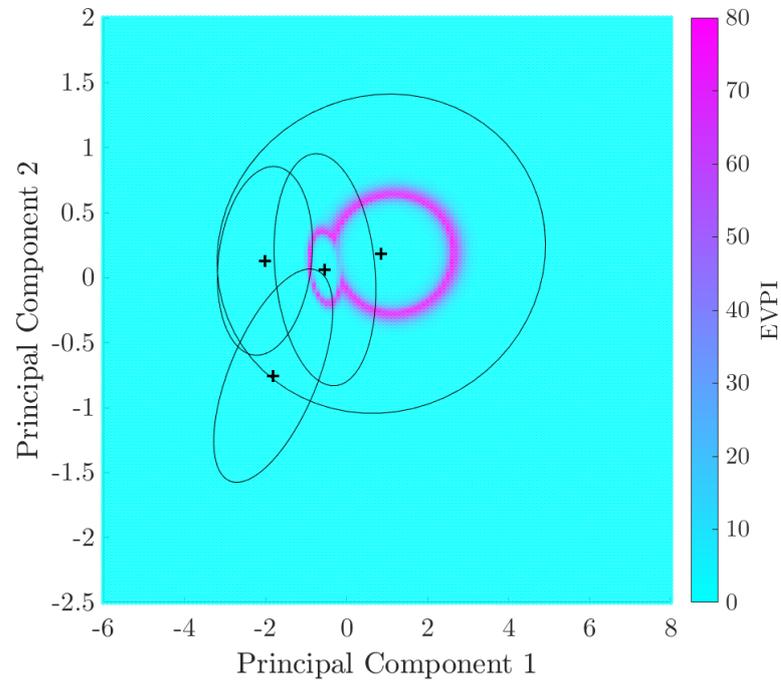


Figure 8.19: The EVPI over the two-dimensional PCA projection of the feature space given a 2-dimensional projection of the final model learned when $U(H_t = 4) = -500$.

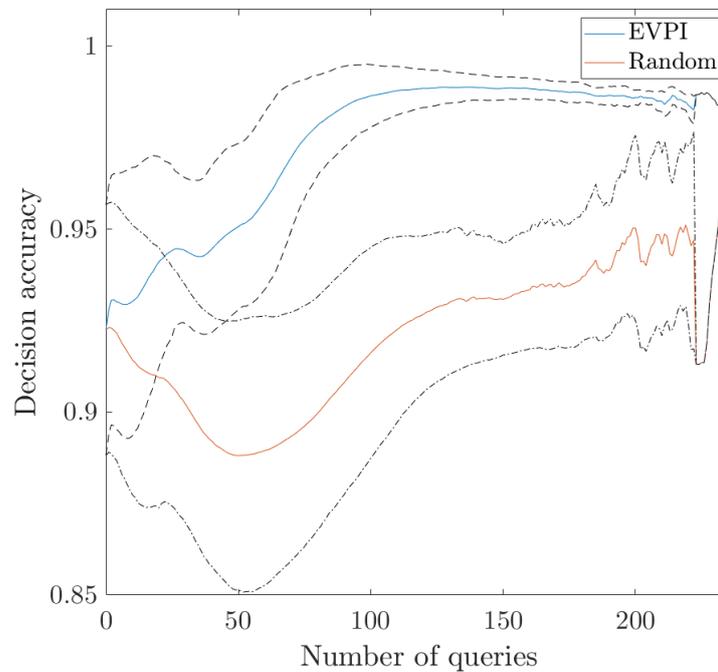


Figure 8.20: The variation in decision accuracy with number of label queries for an agent utilising a statistical classifier trained on \mathcal{D}_t extended via (i) risk-based active querying (EVPI) and (ii) random sampling (Random). The dashed lines show $\pm 1\sigma$.

queries than random sampling, the variance of the decision accuracy also converges more rapidly. This result indicates that using risk-based active learning may result in more consistent decision-making performance. Upon closer examination, it can be seen that, for higher numbers of queries, the decision accuracy slightly declines. This observation and the initial decrease in accuracy for random sampling can be explained by *sampling bias*, which is discussed further in Section 8.6. Once again, the variability in decision accuracy can be seen to collapse for high numbers of queries and, again, this is explained by the fact that only one of the 1000 trials made more than 220 queries. In the case of random querying, a sharp drop-off in decision accuracy is observed at the point that the variability collapses – such an observation is expected as it is logical that a model that is performing poorly at a given time will proceed to make further queries in order to improve.

To summarise, risk-based active learning was applied to a ‘real-world’ case study, specifically the Z24 Bridge benchmark for SHM. A simple maintenance decision problem was constructed for the bridge. A probabilistic Gaussian mixture model was employed to distinguish between four salient states of interest using natural frequencies as a discriminative feature. The EVPI of incipient data points with respect to the decision process was used to guide the querying of health-state labels where querying would correspond to the inspection of the bridge.

8.6 Discussion

The results presented in Sections 8.4 and 8.5 indicate that making inspections of structures according to a risk-based active-learning heuristic may provide a cost-effective method of developing statistical classifiers for use in decision processes in situations when no, or limited, labelled data are available *a priori*. Moreover, it is interesting to note from Figures 8.10 and 8.17 that classifiers do not necessarily need to accurately model the entire feature space to be useful in decision processes, and in fact, well-fitting models can be foregone in favour of decision performance. That being said, a statistical model that is able to accurately predict health-state labels has the potential to provide additional information such as damage locations and types that may be useful in directing more specific types of maintenance and coordinating repair teams - albeit at a potentially higher cost. Indeed, as the dimensionality of a decision space increases, the need for high classification accuracy becomes more

crucial. Nevertheless, it can be said with confidence that the appropriate machine learning paradigm to be used for statistical classifier development in SHM is highly dependent on the context in which the monitoring system is being employed; taking into account factors such as: data availability, knowledge of relevant physics, and the decision-support application of the monitoring system itself.

The risk-based approach to active learning presented in the current chapter is not without its limitations. Notably, in order to evaluate EVPI as presented, it is necessary to assume that the number of classes (i.e. health states of interest) are known prior to the implementation of a monitoring system. This limits the flexibility of the classifier and may result in the mistreatment of unforeseen health states and associated failure events within the decision framework. Additionally, it is assumed that perfect information of the health state can be acquired by means of inspection. During the inspection procedure, human error may be introduced, and in some scenarios it may be vital to account for this uncertainty in the active learning framework by relaxing the perfect information assumption when possible. For interesting research on how imperfect information from inspections can be modelled, the reader is directed to [16, 135, 136].

8.6.1 Sampling Bias

A noteworthy observation from Sections 8.4 and 9.3 that bears further discussion is that, whilst decision accuracy may be increased via risk-based active learning, after a certain number of queries have been made, any incipient data points whose true labels may have high expected value with respect to a structural maintenance decision process, may, in fact, be detrimental to the performance of a decision-maker when incorporated into the statistical classifier on which it relies.

A likely explanation for this observation is the phenomenon known as *sampling bias*; a known issue associated with active learning that has been documented to impact upon classification performance [95, 137]. Sampling bias occurs when specific regions of feature-space are over/under-sampled resulting in a training dataset that is not representative of the underlying distributions; this can clearly be observed in Figure 8.10. In the numerical example presented in Section 8.4, sampling bias manifests as unrepresentative mixing proportions λ , which may result in overconfident misclassifications that subsequently cause erroneous actions to be decided.

To overcome the dangers of sampling bias, there are several potential solutions which could be considered. One option would be to establish a heuristic-based methodology for switching between value of information-based and uncertainty-based measures for guided sampling. This approach would have the effect of establishing decision boundaries by querying regions of the feature space with high value of information, while also exploring low likelihood and high information (including low-value) regions of the feature space. An alternative approach would be to incorporate semi-supervised learning techniques [138, 139], such that the label predictions for the unlabelled data \mathcal{D}_u may be utilised to retain representative estimates for the mixing proportions. Alternatively, discriminative classifiers, that relax the assumptions required for generative models, could be employed. Semi-supervised learning and discriminative classifiers are examined further in the next chapter.

8.7 Summary

The aim of the current chapter has been to present a risk-based active learning approach for SHM. The approach utilises the expected value of perfect information of incipient data points to instigate inspections, such that structural health information may be obtained and incorporated into probabilistic classifiers. The methodology was demonstrated on a numerical case study to aid in visualisation and understanding of the risk-based active-learning process. Additionally, the approach was demonstrated on an experimental dataset obtained from a previously-existing bridge, thereby highlighting its potential applicability to ‘real-world’ engineering problems. The results of the case studies indicated that the risk-based approach to active learning has the potential to provide a cost-effective solution to the development of decision-supporting SHM systems. This finding is valuable as, ordinarily, the comprehensively labelled datasets necessary for the fully-supervised learning of statistical classifiers are seldom available at the inception of a monitoring system. The next chapter investigates how statistical classifiers and learning algorithms may be designed to improve decision-making performance.

IMPROVING DECISION-MAKING VIA CLASSIFIER DESIGN

As discussed previously, statistical pattern recognition (SPR) has been established as the state of the art for making data-informed predictions in the context of physical-asset management technologies such as structural health monitoring (SHM) systems [1] and digital twins [140, 141]. Statistical classifiers are a fundamental component of the SPR approach to these decision-supporting technologies – enabling the categorisation of acquired data into groups, or classes. For example, in the context of SHM, by associating the target classes of a classifier with salient health-states (e.g. undamaged, nascent damage, severe damage), inferences can be made regarding the condition of a structure of interest.

It was mentioned in the previous chapter that the traditional *supervised* and *unsupervised* machine-learning paradigms are of limited applicability. This predicament arises because, oftentimes, there is a scarcity of labelled data corresponding to salient damage, operational, and environmental states of a structure. As demonstrated in Chapter 8, risk-based active learning provides a methodology capable of overcoming the challenges associated with data scarcity when a statistical classifier is being used within a predefined decision process.

While being apt for dealing with challenges associated with online decision-support systems, active-learning algorithms are not without their own pitfalls. Notably, generative models (such as those used in [95, 142]), learned via active learning, are

often susceptible to sampling bias because the preferential nature of the querying process causes inequity in the amount of data observed for each class. This problem is pertinent, as sampling bias has been found to degrade the performance of classifiers, and decision-making agents alike [95, 142, 143].

The current chapter proposes and examines two novel methods for addressing sampling bias in active-learning algorithms for improving the performance of decision-makers utilising the learned classifiers in the context of SHM decision-support.

The first approach involves adapting the algorithm presented in [142], via the introduction of semi-supervised learning. Here, two formulations of semi-supervised learning are considered; expectation-maximisation with respect to the generative model learned in a supervised manner, and latent-state smoothing with respect to the hidden Markov model that underpins the asset-management decision processes. The second novel approach replaces the generative Gaussian mixture model, used in previous literature [95, 142], with discriminative classifiers. Specifically, two formulations of multiclass relevance vector machines (mRVMs) [144] are considered. In total, these approaches result in four new formulations of risk-based active learning. Finally, further novelty is provided in the form of valuable discussions on the role of statistical classifiers in asset management technology, and how decision-support can be improved via classifier design.

9.1 Effects of Sampling Bias

Although active learning has been found to be an effective way of constructing highly-informative and valuable datasets when labels are costly to obtain, such algorithms are somewhat of a double-edged sword because of a phenomenon known as *sampling bias* [137, 143]. Sampling bias occurs because of an active learning scheme preferring data labels obtained in specific regions of a feature space, as guided by the specified query measures and heuristics. Consequently, unrepresentative training datasets are formed in which the data diverges from the underlying generative distribution. In some applications, sampling bias can degrade the performance of classifiers [137, 143]. Risk-based approaches to active learning are more susceptible to the effects of sampling bias than traditional information-based approaches because data with high-value of information are often a subset of data with high-information content [95, 142]. This characteristic can be understood if one realises that not all

data are equally informative from a classification perspective, and further, that not all information is equally valuable from a decision-making perspective.

Thus far, active learning algorithms for SHM have centred around generative classifiers [95, 142]; learning joint probability distributions across the input and label space. Specifically, mixtures of Gaussian distributions, or Gaussian mixture models (GMMs), were utilised in [95, 142]. To highlight the effects of sampling bias, a case study on a representative synthetic dataset is presented. In particular, the case study focusses on the performance of a decision-making agent utilising a risk-based active-learning algorithm for the development of a GMM.

9.1.1 Case Study: Visual Example

To draw attention to the effects of sampling bias in risk-based active-learning algorithms, a modified version of the synthetic dataset used in [142] and the previous chapter is adopted. Once again, the dataset consists of a two-dimensional input space $\mathbf{x}_t = \{x_t^1, x_t^2\}$ and a four-class label space $y_t \in \{1, 2, 3, 4\}$, where the Classes 1 through 4 correspond to increasing severities of damage. Here it is worth noting that, for convenience and for consistency with the relevant literature, a minor change of notation has been adopted for the current chapter; specifically, inputs are now denoted as \mathbf{x} (previously $\boldsymbol{\nu}$) and health-state labels are denoted with y (previously \mathbf{H}_t).

While the original dataset comprised 1997 data points ordered according to health state – from Class 1 through to Class 4; for the current case study, the dataset was extended by drawing additional independent samples from a generative distribution learned from the original dataset. This procedure was conducted a total of five times resulting in 11997 data points that repeatedly progress from Class 1 through to Class 4, thereby emulating the process of structural deterioration and subsequent repair – a pattern that could conceivably be experienced by the fictitious structure of interest S . A visualisation of the extended dataset is presented in Figure 9.1.

In order to employ risk-based active learning for the development of a statistical classifier, a decision process for the structure S must first be specified. Here, an identical binary maintenance decision process to that used in the previous chapter is selected for the current study. As such, the relevant utility functions and conditional probability distributions can be found in Tables 8.1 to 8.4.

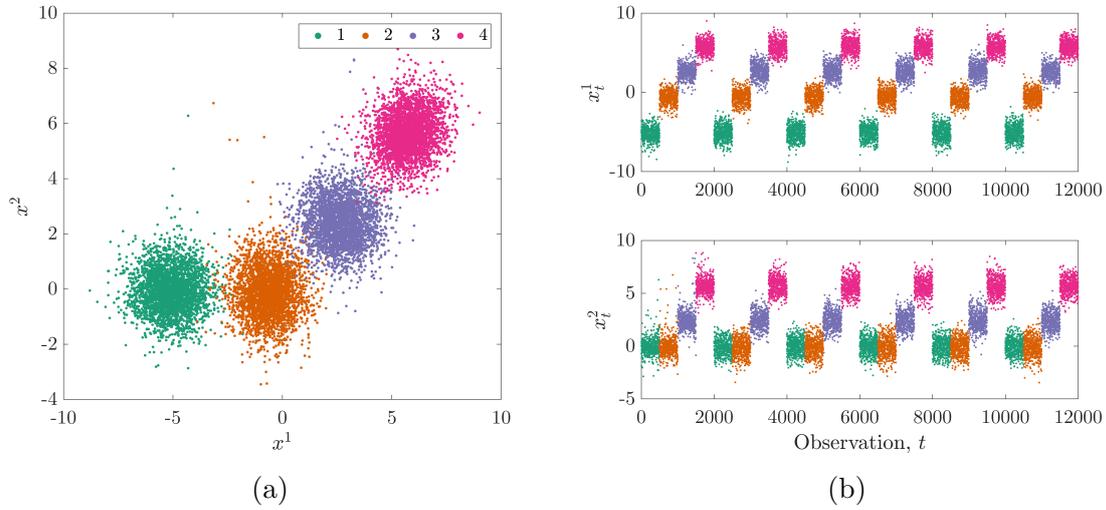


Figure 9.1: Visualisation of the extended synthetic dataset in: (a) the feature space and (b) discrete time t .

It is again assumed that the ground-truth health state at discrete-time instance t can be obtained via inspection at the cost of $C_{\text{ins}} = 7$.

As previously discussed, generative classifiers, specifically GMMs, have garnered the vast proportion of attention with regard to active learning for SHM [95, 142]. As such, the current case study adopts a classifier specified by a mixture of four multivariate Gaussian distributions learned in a supervised manner from a labelled dataset \mathcal{D}_l .

Concisely, a Gaussian mixture model was trained in a supervised Bayesian manner on \mathcal{D}_l . Within the risk-based active learning algorithm, the learning procedure presented in Chapter 8 is reapplied each time the labelled dataset \mathcal{D}_l is extended following the inspection of structure S that is mandated when $\text{EVPI}(d_t|y_t) > C_{\text{ins}}$.

Results

In order to assess the effects of sampling bias, the risk-based active-learning approach was used to learn a GMM within the decision process outlined previously. This process was repeated 1000 times. For each repetition, the dataset was randomly halved into a test set and a training set \mathcal{D} . From the training set, a small ($\sim 0.2\%$) random subset retain their corresponding ground-truth labels. These data form the initialised labelled dataset \mathcal{D}_l . The remaining majority of data from \mathcal{D} have their

ground-truth labels hidden, forming the unlabelled dataset \mathcal{D}_u .

Figures 9.2 and 9.3 show a GMM from one of the 1000 repetitions, before and after the risk-based active-learning process, respectively.

It can be seen from Figure 9.2a that, initially, the model fits poorly; the means and covariances for each class being heavily influenced by the zero-mean and unit-variance priors because of the lack of data. Figure 9.2b shows the EVPI over the feature space induced by considering the initial model within the context of the decision process. The near-symmetric ‘sharp’ regions of high EVPI (pink) can be attributed to the cluster for Class 4; in particular, the major axis of the covariance ellipse (i.e. the dominant eigenpair of the covariance matrix). Intuitively, the regions of high value of information occur between the clusters for classes with severe and milder consequences.

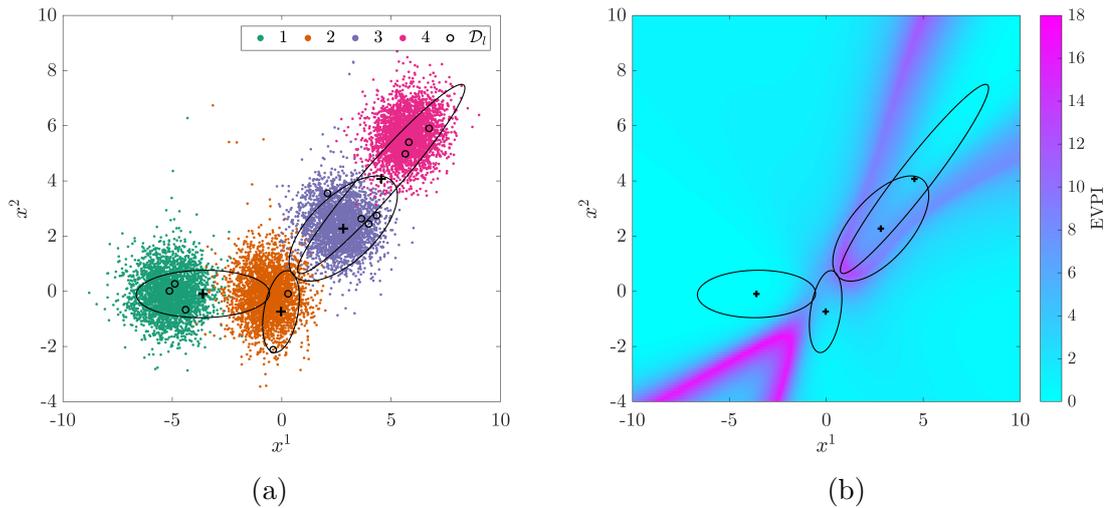


Figure 9.2: A statistical classifier $p(y_t, \mathbf{x}_t, \Theta)$ prior to risk-based active learning; *maximum a posteriori* (MAP) estimates of the mean (+) and covariance (ellipses represent 2-sigma) are shown. (a) shows the initial model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

From Figure 9.3a, one can see that during the active-learning process, querying is concentrated in highly-localised regions of the feature space. Specifically, regions between the clusters for Class 3 (moderate damage) and Class 4 (severe damage), have been queried preferentially. It is clear from Figure 9.3a that the subset of data \mathcal{D}_l is not representative of the underlying distribution, indicating that sampling bias is present. Nonetheless, the queried data have been somewhat successful in learning a decision boundary - as can be deduced by observing the region of high EVPI between

the means of clusters for Class 3 and Class 4 in Figure 9.3b.

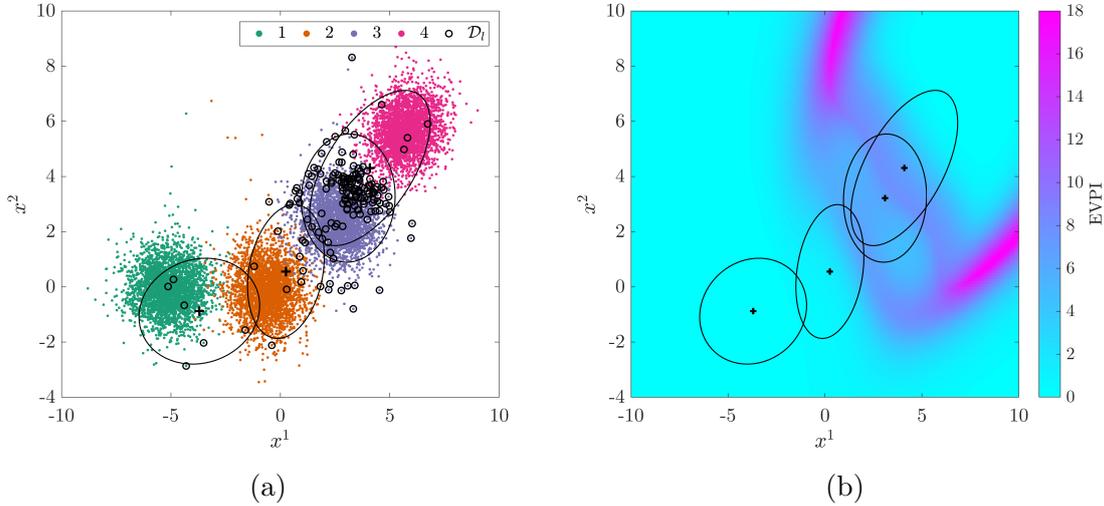


Figure 9.3: A statistical classifier $p(y_t, \mathbf{x}_t, \Theta)$ following risk-based active learning; *maximum a posteriori* (MAP) estimates of the mean (+) and covariance (ellipses represent 2-sigma) are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

Figures 9.4 and 9.5 provide further evidence of sampling bias in the risk-based active learning approach.

Figure 9.4 shows the means and standard deviations, calculated from the 1000 repetitions, of the relative representation of each class within \mathcal{D}_l . Figure 9.4a shows the class proportions for \mathcal{D}_l subject to risk-based active learning whereas Figure 9.4b shows the class proportions for \mathcal{D}_l subject to an equivalent number of random queries. As one would expect, the random querying is initially biased, as the learning algorithm is presented with ordered data. However, as more random queries are made, the proportion for each class converges to approximately 25%. In stark contrast, under querying guided according to EVPI, the class proportions diverge early in the querying process. Figure 9.4a shows that, on average, Class 3 garners a majority representation after fewer than 50 queries. This disparity in class representation is maintained throughout the risk-based active-learning process with Class 3 reaching a peak representation of approximately 80%.

Figure 9.5 compares the total number of queries for data point indices in D_u over the 1000 repetitions for EVPI-based and random querying. Although individual data points within the training data D are not fixed within D_u because of the random selection of D_l , only a small proportion of data are in D_l and therefore the index

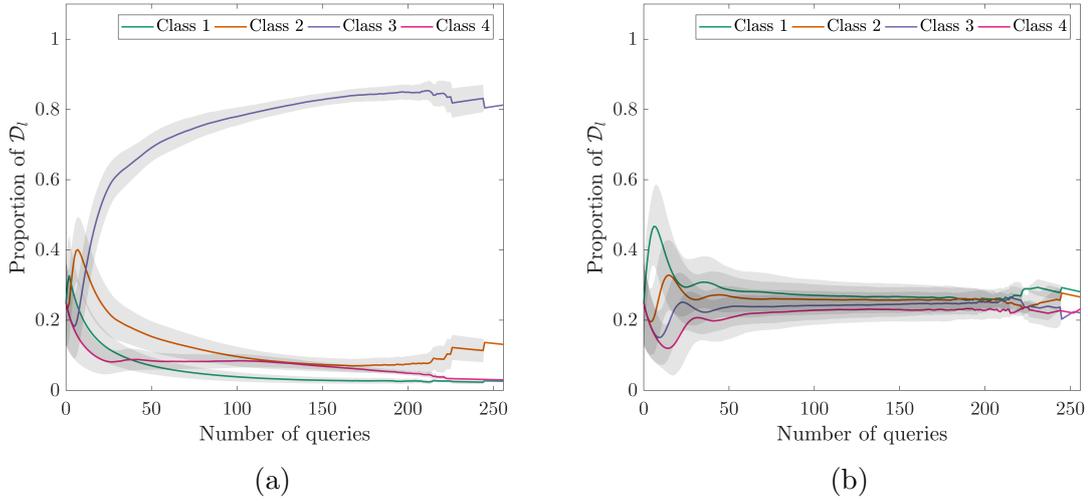


Figure 9.4: Variation in class proportions within \mathcal{D}_l with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended via (a) risk-based active querying and (b) random sampling. Shaded area shows $\pm 1\sigma$.

of a datapoint will differ by no more than 12 between D and D_u . Furthermore, the histogram ‘bins’ represent groups containing 25 indices each, thereby minimising the amount of leakage from differing indices. As one would expect, random sampling results in each data point being queried approximately by an equal amount, as is indicated by the class proportions shown in Figure 9.4. Figure 9.5 shows risk-based active learning results in data corresponding to Class 3 being queried preferentially which, of course, results in a biased labelled dataset on which the classifier is trained.

Both the positive and negative consequences of sampling bias can be realised by considering Figure 9.6. Figure 9.6a shows the median and interquartile range (calculated from 1000 repetitions) of the decision accuracy for an agent utilising the learned GMMs, evaluated throughout the risk-based active-learning process. For comparison, also shown in Figure 9.6a, is the same performance measure for an agent utilising GMMs learned from a labelled dataset comprised of an equal number of randomly-queried data points. Here, ‘decision accuracy’ is the measure of decision-making performance as used in the previous chapters [99].

It can be seen from Figure 9.6a, that a high level of decision-making performance can be achieved with fewer queries by adopting a risk-based approach to active learning, when compared to random sampling. This result is achieved because of the preferential querying process that induces sampling bias in the labelled dataset. That being said, the performance of the agent utilising risk-based active learning

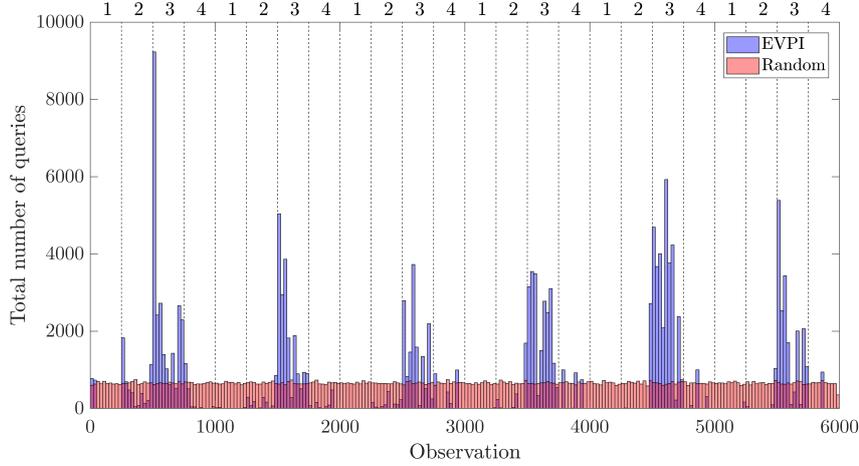


Figure 9.5: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u over 1000 runs when adopting: (a) risk-based active learning (EVPI) and (b) random sampling (Random) in order to learn a GMM. The average locations of classes within \mathcal{D}_u are numbered on the upper horizontal axis and transitions between classes are denoted as a dashed line.

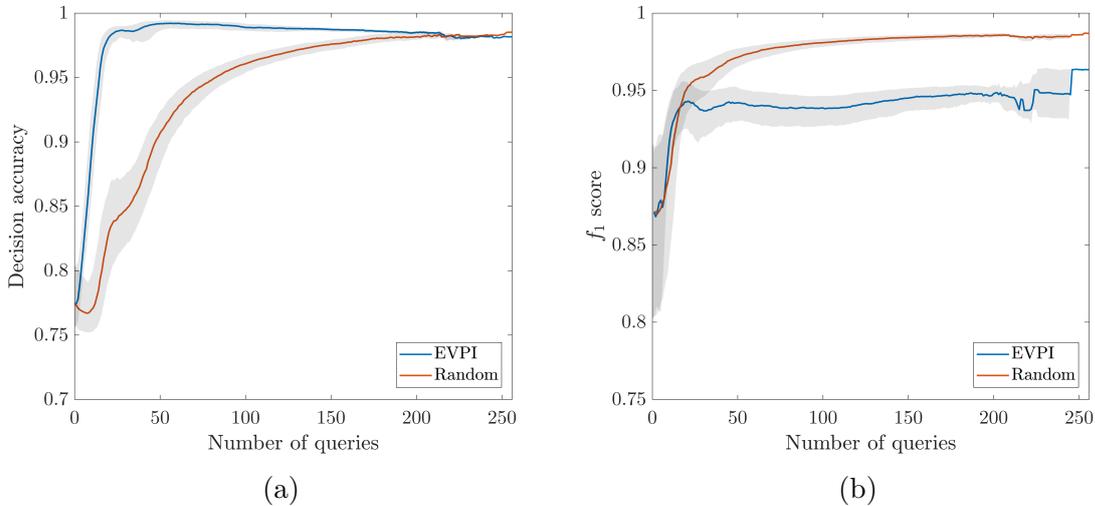


Figure 9.6: Variation in median: (a) decision accuracy and (b) f_1 -score with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended via: (i) risk-based active querying (EVPI) and (ii) random sampling (Random). Shaded area shows the interquartile range.

gradually degrades as more queries are made. While randomly obtaining queries improves decision accuracy at a lesser rate, a decline in performance is not observed and, in fact, random sampling eventually surpasses guided querying in terms of decision-making performance. These results indicate that the decline in decision accuracy is a result of the sampling bias that is exacerbated by the latter queries in

the active learning process. This proposition is further supported by Figure 9.6b.

Figure 9.6b shows the average classification performances (f_1 -score [145]) of agents utilising EVPI-guided and random approaches to querying. The f_1 -score is defined as,

$$f_1 = \frac{\text{tp}}{\text{tp} + 0.5(\text{fp} + \text{fn})} \quad (9.1)$$

where tp is the number of true-positives, fp is the number of false-positives and fn is the number of false negatives.

It can be seen from Figure 9.6b, that active querying and random querying increase classification performance at similar rates initially. However, the classification performance for risk-based active learning plateaus at a lower value than that of random sampling. This observation provides supplementary evidence that classifiers developed via risk-based active learning represent the underlying distribution poorly¹. Once again, this observation is to be expected when considering the disparity in class representation within \mathcal{D}_l for risk-based active learning. The drastic difference between the decision accuracy and classification accuracy demonstrates that risk-based active learning algorithms prioritise decision-making performance over classification performance.

9.1.2 Further Comments

As an archetypal example of a probabilistic generative model, one can rationalise the effects of sampling bias on decision-making agents utilising such models by examining the posterior parameter estimates for the GMM, given in equations (8.21) and (8.23). Here, a reminder is provided that, as a result of the Bayesian inference on distribution parameters, components of the posterior GMM are Student- t distributed, rather than Gaussian as the name would imply.

One way in which bias is introduced to a posterior model is via the estimation of the posterior means and covariances (equations (8.14) and (8.18), respectively). These updates depend on weighted averages of the prior mean and sample mean. The

¹In the limit of infinite queries, random sampling will achieve classification performance equivalent to the ‘gold standard’ of fully-supervised learning.

sample mean $\bar{\mathbf{x}}_k$ becomes biased because, for a given label, regions of the feature space that are shared with or close to another label are sampled preferentially. Whilst these parameters do introduce bias, this bias facilitates the rapid improvement in decision-making performance achieved via risk-based active learning. These biased estimates cause the distributions corresponding to classes of differing optimal strategies to drift closer together, thereby narrowing the decision boundary region inferred via the model.

The bias introduced via the posterior mixing proportions (equation (8.23)) can likely be ascribed to the deterioration in decision performance. This posterior update is simply a summation of the prior weight for a given class and the number of samples corresponding to that class in \mathcal{D}_l . As is demonstrated in Figure 9.4, the risk-based active-learning process causes the class proportions in \mathcal{D}_l to greatly deviate from those in the underlying distribution. This effect undoubtedly results in misclassifications, and therefore erroneous decisions, particularly for data points close to decision boundaries.

Finally, an additional effect of sampling bias, not immediately obvious from the case study presented above, arises via the posterior estimates for the Student- t distribution's degrees of freedom parameter. As previously mentioned, this parameter governs the weightedness of the distribution tails; distributions with fewer degrees of freedom possess tails that are weighted more heavily and thus hold a greater probability mass away from the mean. In contrast, distributions with a greater number of degrees of freedom have more probability mass concentrated around the mean. From equations (8.17) and (8.21), it can be seen that degrees-of-freedom is a monotonically-increasing function of the number of samples for a given class. This characteristic has implications for how an agent utilising an actively-learned model responds when presented with outlying observations. As some classes are seldom queried during risk-based active learning, a classifier trained accordingly will make overconfident predictions on outlying data in favour of classes that are under-represented in the training data. This result has particularly concerning consequences for structural-maintenance decision processes if benign classes are under-represented, as extreme structural damage states may go ignored. Further discussion around the treatment of outlying data for classifiers used in an SHM decision-support context is provided in Section 9.5.

In essence, risk-based active learning provides a useful methodology for developing statistical classifiers for decision-support applications because of its ability to consider

the value of information with respect to the cost of obtaining said information and thereby improve decision-making performance in a cost-effective manner. Nevertheless, in some scenarios, because of inherent issues with sampling bias, the use of generative models alone may not serve as the most appropriate foundation for risk-based active-learning algorithms. Although the decline in decision performance observed in the previous subsection is small, ultimately, it corresponds to a loss in resource/utility and therefore it is desirable to rectify this issue. The following sections outline and demonstrate approaches to curtail the detrimental effects of sampling bias in such algorithms.

9.2 Approaches to Address Sampling Bias

As discussed previously, sampling bias can occur in generative models learned by risk-based active learning. The problem is from posterior estimates of distribution parameters and a result of imbalanced class representation and non-uniform coverage of the feature space within \mathcal{D}_l .

The current section aims to address these issues by considering two main approaches; semi-supervised learning, and discriminative classifiers. Two methods for semi-supervised learning are considered; expectation-maximisation, and latent-state smoothing. The multiclass relevance vector machine (mRVM) is considered as the discriminative classifier. Two formulations of the mRVM are considered; mRVM₁ and mRVM₂ [144, 146].

9.2.1 Semi-supervised Learning

Alongside active learning, semi-supervised learning is a form of partially-supervised learning – utilising both \mathcal{D}_l and \mathcal{D}_u to inform the classification mapping. The fundamental principle of semi-supervised learning that distinguishes it from active learning is as follows; data in the unlabelled dataset \mathcal{D}_u can be given *pseudo-labels* that are informed by the ground-truth labels available in \mathcal{D}_l . By incorporating pseudo-labels for unqueried data into the risk-based active learning algorithm, class imbalance and inadequate coverage of the feature space can be rectified.

There are several approaches to semi-supervised learning. The simplest approach,

self-labelling, involves training a classifier using \mathcal{D}_l in a supervised manner [138]. Pseudo-labels can then be assigned to the unlabelled data according to the predictions of this classifier. The model can subsequently be retrained, incorporating both ground-truth labels and pseudo-labels. The effectiveness of self-labelling is strongly conditioned on the implementation and the underlying supervised-learning algorithm. More advanced approaches that utilise *low-density-separation* [138], and *graph-based* learners [147], are available.

Semi-supervised learning has been applied to pattern recognition problems for SHM [148]. This method of learning brings several benefits, such as an increased utilisation of information obtained via costly structural inspections.

Expectation-Maximisation

The first of the approaches to semi-supervised learning considered here, aims to exploit the generative mixture model form of the statistical classifier presented in Section 8.4.2. Generative models can conveniently account for labelled and unlabelled data; this is achieved by modifying the *expectation-maximisation* (EM) algorithm [149], typically used for unsupervised density estimation, such that the log-likelihood of the model is maximised over both unlabelled and labelled data.

The aim of semi-supervised learning via the EM algorithm is to infer updated distribution parameters Θ from $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$. The *maximum a posteriori* (MAP) estimate of these updated parameters is specified as follows,

$$\Theta|\mathcal{D} = \operatorname{argmax}_{\Theta} \left\{ \frac{p(\mathcal{D}|\Theta)p(\Theta)}{p(\mathcal{D})} \right\} = \operatorname{argmax}_{\Theta} \left\{ \frac{p(\mathcal{D}_u|\Theta)p(\mathcal{D}_l|\Theta)p(\Theta)}{p(\mathcal{D}_u, \mathcal{D}_l)} \right\} \quad (9.2)$$

Implicit in the factorisation of $p(\mathcal{D}|\Theta)$ in equation (9.2), is the assumption that \mathcal{D}_l and \mathcal{D}_u are conditionally independent. This assumption holds for random querying, as samples selected in this manner are independent and identically distributed (iid). Unfortunately, active learning violates this assumption, as data in \mathcal{D}_l are not iid because of the preferential querying process and iterative model updating [137, 143]. Convenient assumptions such as this are frequently relied upon in statistical and engineering analyses (particularly in active-learning contexts [95]). As such, the assumption is embraced for the current case study in order to demonstrate the

decision-making performance that one may achieve despite the violation.

To circumvent numerical instabilities, the MAP estimate in equation (9.2) is formulated as a maximisation of the expected joint log-likelihood across \mathcal{D} [138],

$$\begin{aligned} \mathcal{L}(\Theta|\mathcal{D}_l, \mathcal{D}_u) \propto & \sum_{t=1}^m \log \sum_{k=1}^K p(\tilde{\mathbf{x}}_t|y_t = k, \Theta)p(y_t = k|\Theta) \\ & + \sum_{t=1}^n \log[p(\mathbf{x}|y_t = k)p(y_t = k|\Theta)] + \log p(\Theta) \end{aligned} \quad (9.3)$$

where the first term corresponds to the log-likelihood of the model over \mathcal{D}_u , the second term corresponds to the log-likelihood of the model over \mathcal{D}_l and the final term is the log-prior-likelihood of the model parameters. It should be noted that the first term contains a summation over the label space; this marginalises out y_t , which is considered a latent variable for data in \mathcal{D}_u .

During each E-step of the EM algorithm, the unlabelled observations are classified using the current estimate of the model parameters. During the M-step, updates for parameters Θ are found using the predicted labels determined via the E-step, in addition to the acquired labelled data.

The E-step and M-step are more formally defined as follows [149].

E-step: A *responsibility matrix* R is computed for the unlabelled data, corresponding to the posterior label prediction,

$$R[t, k] = r_{tk} = p(\tilde{y}_t = k|\tilde{\mathbf{x}}_t, \Theta), \quad \forall \tilde{\mathbf{x}}_t \in \mathcal{D}_u \quad (9.4)$$

The posterior label predictions for observations in \mathcal{D}_l are given by the known labels y_t , and can be represented using discrete delta functions [63],

$$p(y_t = k|\mathbf{x}_t) = \delta_{k,y_t}, \quad \forall (\mathbf{x}_t, y_t) \in \mathcal{D}_l \quad (9.5)$$

where δ_{k,y_t} is the Kronecker delta function – equal to 1 when $y_t = k$, and 0 otherwise.

Whereas, for \mathcal{D}_l , the number of observations corresponding to each class n_k are

known, these values are indeterminate for data in \mathcal{D}_u . As such, it is convenient to define for each class their *effective counts* r_k in D_u ; these may be calculated from the responsibility matrix as follows [81],

$$r_k = \sum_{t=1}^m r_{tk} \quad (9.6)$$

The total (effective) counts per class over \mathcal{D} can be summarised as,

$$N_k = n_k + r_k \quad (9.7)$$

M-step: Updates for Θ are computed via modified versions of equations (8.14) to (8.18) and equation (8.23). The mean and covariance parameters are updated as follows,

$$\mathbf{m}_n = \frac{\kappa_0}{\kappa_0 + N_k} \mathbf{m}_0 + \frac{N_k}{\kappa_0 + N_k} \bar{\mathbf{x}}_k \quad (9.8)$$

$$\bar{\mathbf{x}}_k \triangleq \frac{\sum_{t=1}^n \delta_{k,y_t} \mathbf{x}_t + \sum_{t=1}^m r_{tk} \tilde{\mathbf{x}}_t}{N_k} \quad (9.9)$$

$$\kappa_n = \kappa_0 + N_k \quad (9.10)$$

$$v_n = v_0 + N_k \quad (9.11)$$

$$S_n = S_0 + S_k + \kappa_0 \mathbf{m}_0 \mathbf{m}_0^\top - \kappa_n \mathbf{m}_n \mathbf{m}_n^\top \quad (9.12)$$

$$S_k \triangleq \sum_{t=1}^n \delta_{k,y_t} \mathbf{x}_t \mathbf{x}_t^\top + \sum_{t=1}^m r_{tk} \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t^\top \quad (9.13)$$

which lead to MAP estimates given by,

$$\hat{\boldsymbol{\mu}}_k = \mathbf{m}_n \quad (9.14)$$

$$\hat{\Sigma}_k = \frac{S_n}{v_n + D + 2} \quad (9.15)$$

The mixing proportions $\boldsymbol{\lambda}$ are updated as follows,

$$\frac{\alpha_k + N_k - 1}{\alpha_0 + N - K} \quad (9.16)$$

where $N = |\mathcal{D}| = n + m$.

The EM algorithm iterates between E-steps and M-steps, resulting in a hill-climbing search that terminates when the log-likelihood (equation (9.3)) converges to a local maximum in the parameter space. The EM algorithm is sensitive to initial conditions, and in many applications requires multiple random initialisations. For the current application, however, the model is initialised using the labelled dataset; this additional information mitigates the need to re-initialise. Examination of the parameter updates provided in equations (9.8) to (9.16) reveals that the information introduced by the incorporation of unlabelled data via EM learning has the ability to reduce the bias in parameter estimates by considering a broader span of the feature space.

Within the risk-based active learning algorithm, it is possible to apply EM every time a new unlabelled observation is acquired; however, this would be computationally expensive. To limit the computational cost of the modified active-learning algorithm, the EM update was only applied following the retraining of the model subsequent to the acquisition of a new ground-truth label obtained via inspection. A flow chart detailing the risk-based active-learning algorithm modified to incorporate EM is shown in Figure 9.7.

Latent Health-State Smoothing

The second approach to semi-supervised learning aims to exploit the Markovian model underpinning the decision processes outlined in Chapter 5. The Markov property asserts that future states of a stochastic process are conditionally independent of past states, given the present states, i.e. $p(y_{T+1}|y_{1:T}) = p(y_{T+1}|y_T)$. Here, the notation

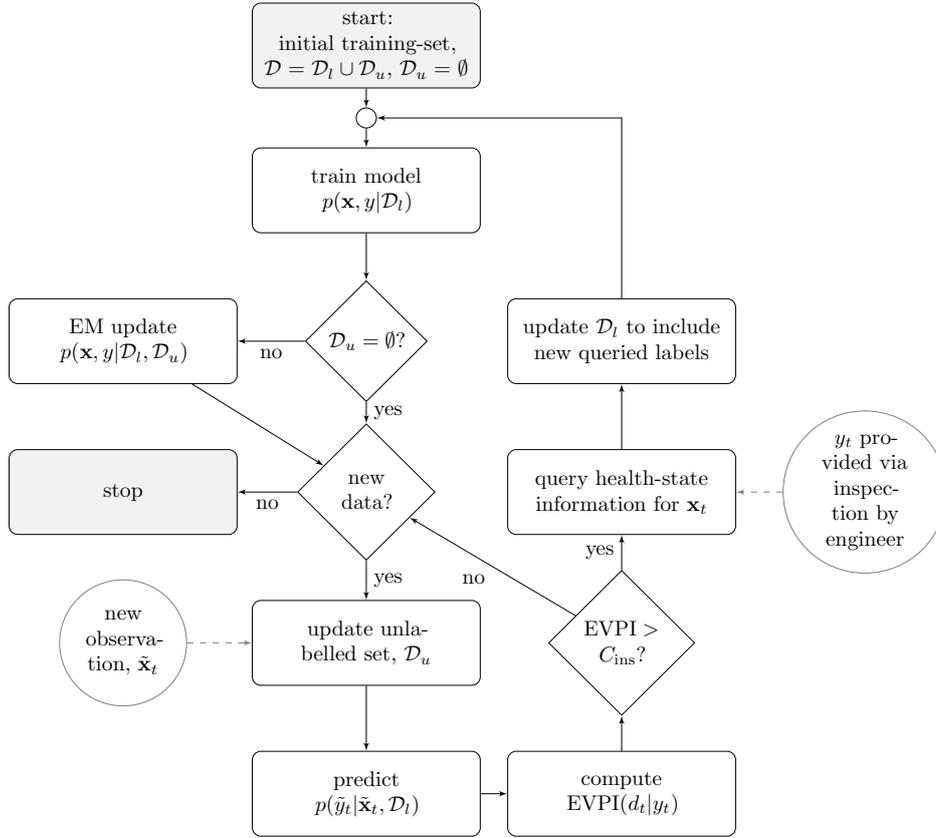


Figure 9.7: Flow chart to illustrate the risk-based active learning process incorporating expectation-maximisation.

$1:T$ is employed to refer to all instances from 1 to T , inclusive.

Hidden Markov models (HMMs) are a statistical state-space representation of stochastic processes for which the Markov property holds. Within HMMs, it is typically assumed that latent states are partially-observable and must be inferred via indirect observations. For the current application, one can form a HMM for latent states between inspections at times $t = a$ and $t = b$. As they have occurred in the past, modifications can be made to the HMM such that decisions between a and b are treated as observed exogenous inputs that influence the subsequent latent state. A Bayesian network of this process can be seen in Figure 9.8.

Given that a modified HMM can be formed for states between two inspections, posterior marginal distributions conditioned on both inspections and intermediate observations can be obtained for the latent states via a smoothing algorithm. For this application, the *forward-backward* algorithm was employed to infer the smoothed distributions [150]. The forward-backward algorithm is a message-passing algorithm

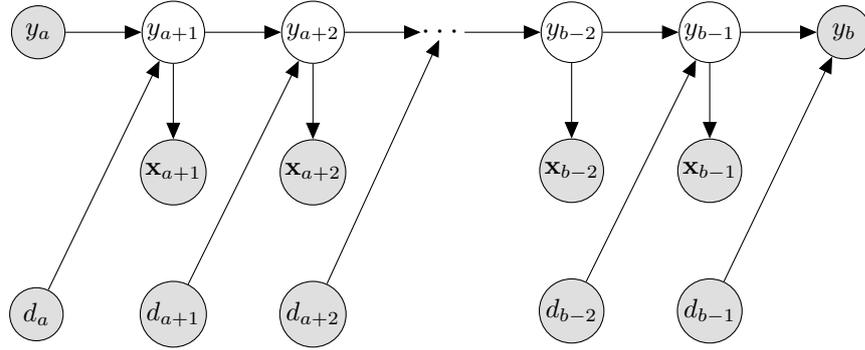


Figure 9.8: A Bayesian network representation of a modified HMM with latent states conditionally dependent on known historical actions.

and involves three key steps: (1) computing forward messages, (2) computing backward messages, and (3) computing smoothed distributions.

The forward messages ϕ_t relate to the joint probability distribution $p(\tilde{\mathbf{x}}_{a+1:t}, \tilde{y}_t | y_a)$ $\forall t \in (a, b)$, and can be computed recursively by invoking the Markov property,

$$\phi_t = p(\tilde{\mathbf{x}}_t | y_t) \sum_{y_{t-1}} \phi_{t-1} \cdot p(y_t | y_{t-1}, d_{t-1}) \quad (9.17)$$

where the message ϕ_a is initialised as δ_{k, y_a} . Here, $p(y_t | y_{t-1}, d_{t-1})$ is equivalent to the transition matrix $p(y_{t+1} | y_t, d_t)$ and $p(\tilde{\mathbf{x}}_t | y_t)$ is the posterior predictive distribution of the statistical classifier.

The backward messages ψ_t relate to the conditional distribution $p(\tilde{\mathbf{x}}_{t:b-1} | \tilde{y}_{t:b-1}, y_b)$ $\forall t \in (a, b)$, and can again be computed recursively as follows,

$$\psi_t = \sum_{y_{t+1}} \psi_{t+1} \cdot p(\tilde{\mathbf{x}}_{t+1} | y_{t+1}) \cdot p(y_t | y_{t-1}, d_{t-1})^\top \quad (9.18)$$

where the message ψ_b is initialised as δ_{k, y_b} . Here, it is worth recalling that $p(y_{t+1} | y_t, d_t)$ can be considered to be a square $K \times K$ matrix for a given d_t .

To obtain the posterior smoothed distributions $p(\tilde{y}_t | y_a, y_b, \tilde{\mathbf{x}}_{a+1:b-1}, d_{a:b-1})$, one can simply multiply together the forward and backward messages and normalise to unity [151],

$$p(\tilde{y}_t | y_a, y_b, \tilde{\mathbf{x}}_{a+1:b-1}, d_{a:b-1}) \propto \phi_t \psi_t \quad (9.19)$$

By taking the MAP of the smoothed distribution, pseudo-labels \hat{y}_t may be assigned to $\tilde{\mathbf{x}}$ such that,

$$\mathcal{D}_u = \{(\tilde{\mathbf{x}}_t, \hat{y}_t) | \tilde{\mathbf{x}}_t \in X, \hat{y}_t \in Y\}_{t=1}^m \quad (9.20)$$

and where,

$$\hat{y}_t = \operatorname{argmax}_{\tilde{y}_t} p(\tilde{y}_t | y_a, y_b, \tilde{\mathbf{x}}_{a+1:b-1}, d_{a:b-1}) \quad (9.21)$$

Using the pseudo-labels provided for \mathcal{D}_u , in addition to \mathcal{D}_l , updated estimates for Θ can be computed in a supervised manner via equations (8.14) to (8.18) and (8.23).

As with the decision process underpinning risk-based active learning, latent-state smoothing relies on the specified observation and transition models. If poor models are chosen, erroneous pseudo-labels may be provided for unlabelled data.

Similar to the EM approach, the smoothing approach to semi-supervised learning addresses sampling bias by reducing the class imbalance and generating a training dataset that is more representative of the underlying distribution. Unlike the EM approach, latent-state smoothing utilises information encoded within the decision process transition model, which can yield some powerful inferences. One can imagine a scenario where it is determined via inspection that the structure is undamaged at time a , and time b , and furthermore, it is known that no interventions were performed on the structure inside the interval (a, b) . Under these conditions, and given that the relevant transition model allows only monotonic degradation, it can be inferred that the structure is undamaged with unit probability for all instances in (a, b) .

Latent-state smoothing fits naturally within the risk-based active-learning framework and can be applied after each inspection. When applied immediately following the first inspection, it is assumed that at the onset of the active-learning algorithm, the structure was in its undamaged state. For the current chapter, latent-states are smoothed only once, using the version of the classifier available up to the most recent inspection. An alternative approach in which all historical latent-states are

smoothed after every inspection using the most up-to-date classifier available may yield better performance, however, would be more computationally expensive. A flow chart detailing the risk-based active-learning algorithm modified to incorporate smoothing is shown in Figure 9.9.

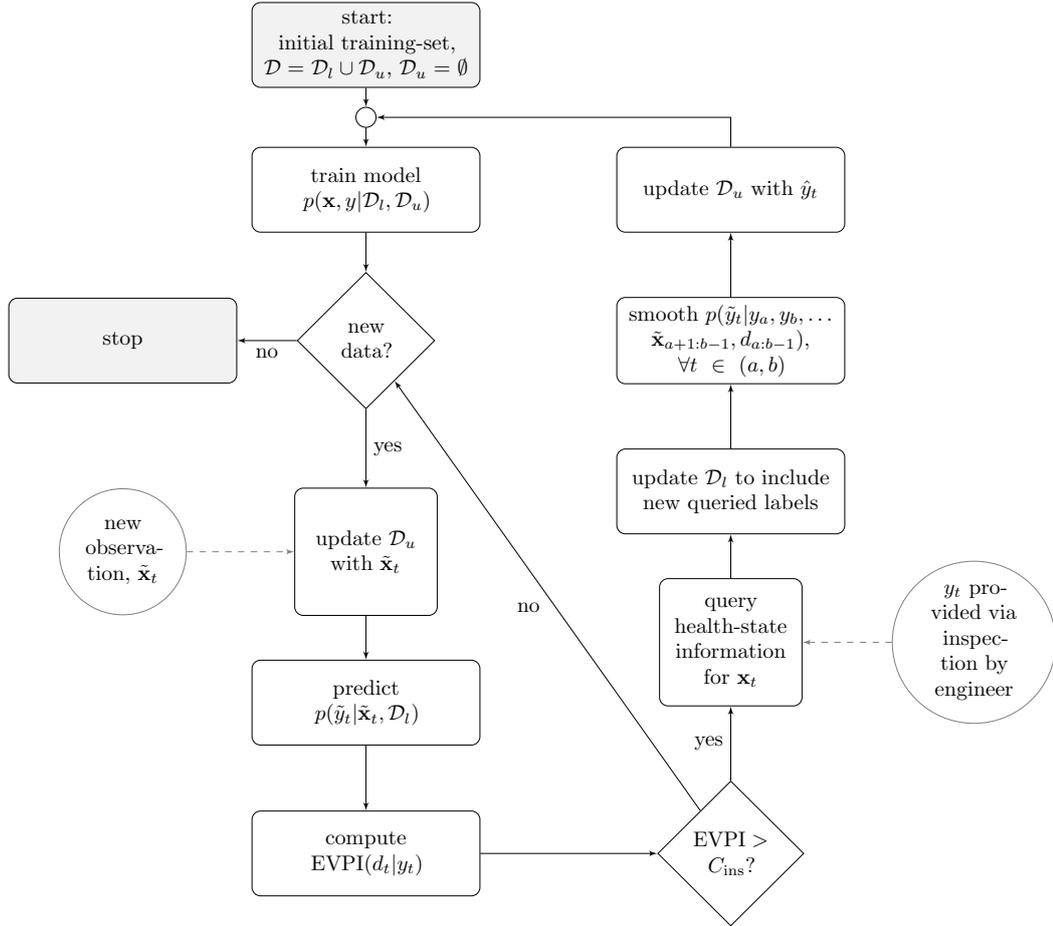


Figure 9.9: Flow chart to illustrate the risk-based active learning process incorporating latent-state smoothing.

Results

The EM and smoothing approaches to semi-supervised learning were each incorporated into the risk-based active-learning process and applied to the case study presented in Section 9.1.1. Once again, 1000 repetitions were conducted, each with randomly-selected training and test datasets and with \mathcal{D}_l randomly initialised as a small subset of the training data. For the smoothing approach, decisions were specified according to the hidden labels associated with \mathcal{D}_u , such that the state

transitions present in the dataset were consistent with the transition models specified in Tables 8.2 and 8.3.

Figures 9.10 and 9.11 show a GMM for one of the 1000 runs after the risk-based active-learning process incorporating EM and smoothing, respectively. The initial model, prior to risk-based active learning, is unchanged from that shown in Figure 9.2.

It can be seen from Figures 9.10a and 9.11a that, similar to the GMM without semi-supervised learning, risk-based active learning with semi-supervised learning results in labels being obtained for localised regions of the feature space, with Class 3 (moderate damage) being preferentially queried. Nonetheless, these figures show that the clusters learned in a semi-supervised manner fit the data very well. Furthermore, examination of Figures 9.10b and 9.11b, reveals that the resulting EVPI distributions over the feature spaces distinctly differ from that in Figure 9.3b. For both the EM and smoothing approaches, the introduction of semi-supervised learning into the risk-based active-learning process has enabled the inference of a well-defined decision-boundary indicated by the band of high EVPI separating the 2-sigma clusters for Class 3 and Class 4.

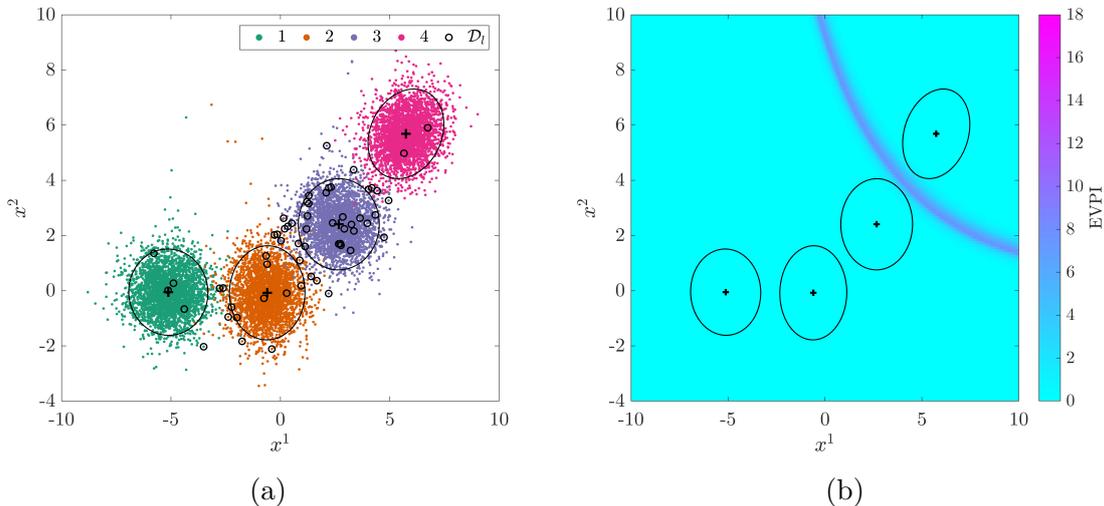


Figure 9.10: A statistical classifier $p(y_t, \mathbf{x}_t, \Theta)$ following risk-based active learning with semi-supervised learning via EM; *maximum a posteriori* (MAP) estimates of the mean (+) and covariance (ellipses represent 2-sigma) are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

Figure 9.12 shows the effective class proportions in the training data \mathcal{D} averaged over

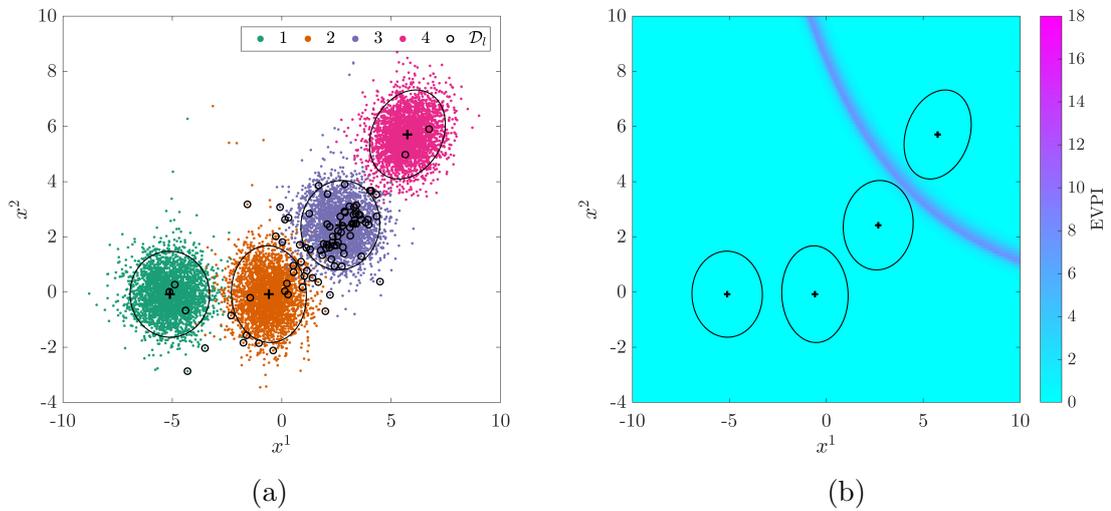


Figure 9.11: A statistical classifier $p(y_t, \mathbf{x}_t, \Theta)$ following risk-based active learning with semi-supervised learning via smoothing; *maximum a posteriori* (MAP) estimates of the mean (+) and covariance (ellipses represent 2-sigma) are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

the 1000 runs for both the EM and smoothing approaches to active learning. These figures, when considered in the context of the equations specifying the relevant model parameter updates, provide some explanation as to why incorporating semi-supervised learning yields improved model fits over standard risk-based active learning. It can be seen from Figures 9.12a and 9.12b that the effective class proportions in the latter stages of the querying process are much more representative of the underlying data distribution when semi-supervised learning is employed - approximately 25% of the dataset represented by each of the four classes. As postulated in Section 9.2.1, having a more representative training dataset results in better-fitting generative distributions, and consequently, a better-defined decision boundary.

Figure 9.13 shows how the number of queries varies between each approach to risk-based active learning. It is immediately obvious from Figure 9.13, that incorporating semi-supervised learning, via either EM or smoothing, substantially reduces the number of queries made. The significance of this result becomes most apparent if one recalls that, in the proposed SHM decision context, the number of queries can be mapped directly onto inspection expenditure. This result is to be expected as, when employing semi-supervised learning, supplementary information from the unlabelled dataset is utilised to specify the model each time a query is made. This characteristic allows clusters to become well-defined more quickly, reducing the area of high-EVPI

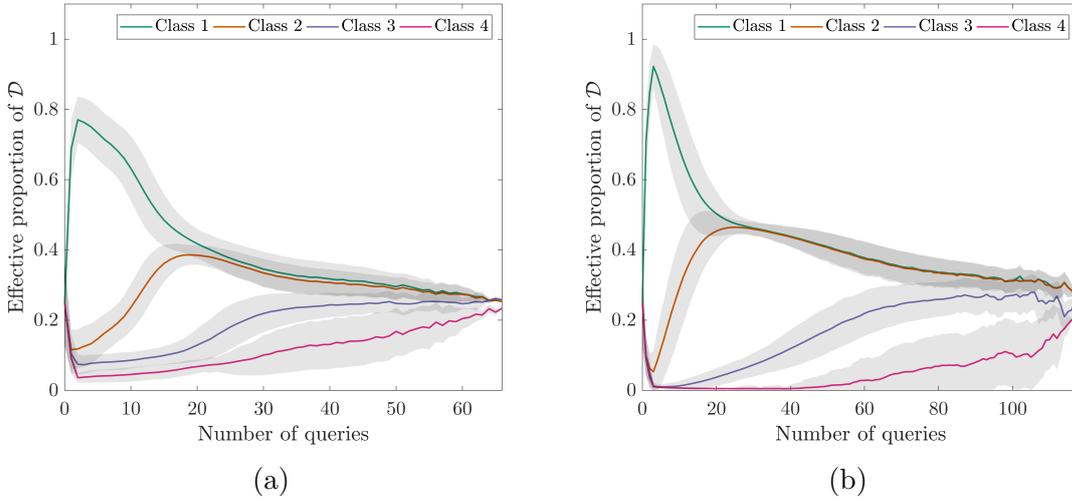


Figure 9.12: Variation in effective class proportions within \mathcal{D}_l with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended by risk-based active learning with semi-supervised updating via: (a) expectation-maximisation and (b) latent-state smoothing. Shaded area shows $\pm 1\sigma$.

regions (as is visible in Figures 9.10 and 9.11) meaning fewer queries are made later in the dataset. This result is evident from Figure 9.14.

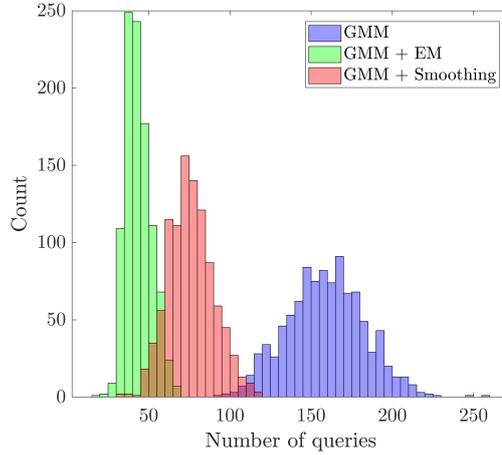


Figure 9.13: Histograms showing the distribution of the number of queries from 1000 runs of the risk-based active learning of: (i) a GMM (blue) (ii) a GMM semi-supervised via expectation-maximisation (green) and (iii) a GMM semi-supervised via latent-state smoothing (red).

Figure 9.14 compares the total number of queries for each index in \mathcal{D}_u over the 1000 repetitions of risk-based active learning conducted with a GMM, a GMM with EM and a GMM with smoothing. It can be seen from Figure 9.14, that the incorporation

of semi-supervised learning into the risk-based active approach results in relatively more queries being obtained during the first occurrence of each class, with relatively fewer queries being made at later occurrences. It can be seen that Class 3 is heavily investigated at each occurrence when semi-supervised methods are not employed. This phenomenon is to be expected when one considers the differences between the high-EVPI regions shown in Figure 9.3b and Figures 9.10b and 9.11b; as previously discussed semi-supervised learning results in a well-defined decision boundary thereby reducing the likelihood that data will have high value of information.

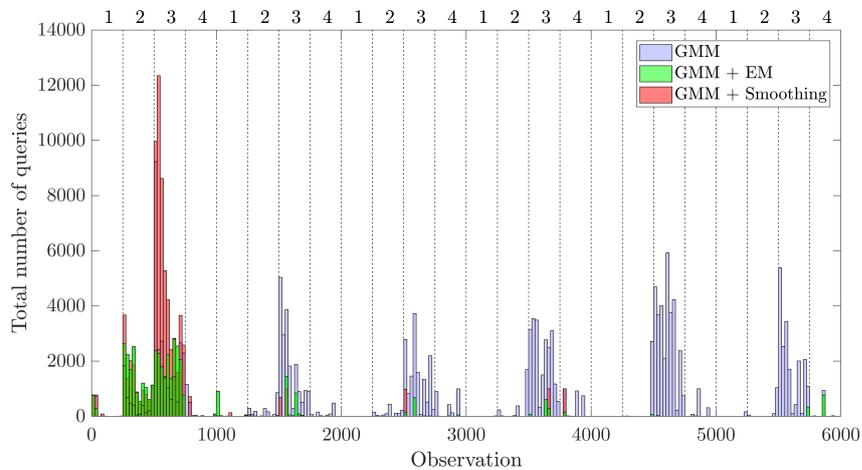


Figure 9.14: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u from 1000 runs of risk-based active learning for: (i) a GMM (blue) (ii) a GMM with EM (green) and (iii) a GMM with smoothing (red). The average locations of classes within \mathcal{D}_u are numbered on the upper horizontal axis and transitions are denoted as a dashed line.

Figure 9.15 shows median decision accuracies and f_1 -scores throughout the query process, and compares those measures for risk-based active learning with, and without, semi-supervised learning. From Figure 9.15a, one may be inclined to deduce that the introduction of semi-supervised learning has been detrimental to the performance of risk-based active learning, as both EM and smoothing result in a delayed increase in decision accuracy. However, when considered alongside Figure 9.14, one can realise that, because semi-supervised learning results in increased querying early in the dataset, decision accuracy over the whole dataset is improved. Furthermore, a decline in decision performance is not observed for the algorithms incorporating semi-supervised learning; this result is because of the reduction in sampling bias obtained via the inclusion of unlabelled data. Figure 9.15b shows the f_1 -score classification performance; these results provide further indication that the models

learned via semi-supervised risk-based active learning better represent the underlying distribution of data and that the detrimental effects of sampling bias have been reduced.

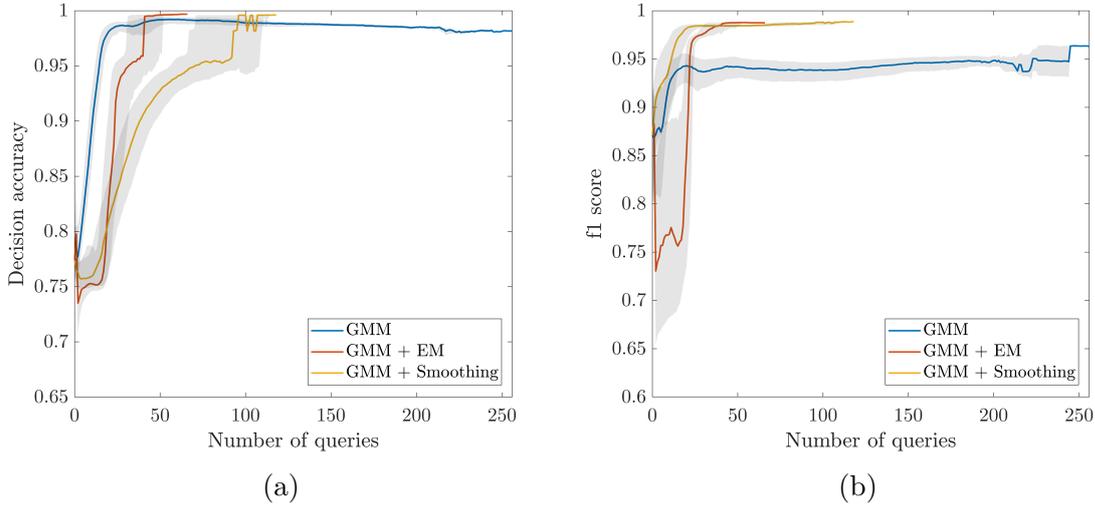


Figure 9.15: Variation in median: (a) decision accuracy and (b) f_1 -score with number of label queries for an agent utilising a GMM learning via risk-based active learning: (i) without semi-supervised updates, (ii) with semi-supervised updates via EM and (iii) with semi-supervised updated via smoothing. Shaded area shows the interquartile range.

In summary, the case study has shown that, for some applications, semi-supervised learning provides a suitable approach to reducing the effects of sampling bias. The generative distributions obtained via these semi-supervised risk-based active learning approaches better fit the underlying data distributions, whilst also establishing well-defined decision boundaries in a cost-effective manner. The semi-supervised learning approach to reducing the negative impact of sampling bias on decision-making performance is re-examined with respect to an experimental dataset in Section 9.3.3.

9.2.2 Discriminative Classifiers

Probabilistic discriminative models (sometimes referred to as conditional models) provide an alternative approach to generative models for developing statistical classifiers. Whereas generative models seek to first represent the underlying joint probability distribution $p(\mathbf{x}, y)$, discriminative models seek to learn the predictive conditional distribution $p(y|\mathbf{x})$ directly. For discriminative classifiers, the mapping $f : X \rightarrow Y$ is typically specified via boundaries that partition the feature space X

according to Y . In probabilistic discriminative models, these classification boundaries are ‘soft’, allowing for nondeterministic classification. For example, data lying on a binary classification boundary will be attributed equal probabilities of belonging to each class. Unlike with generative models, discriminative models do not require assumptions to be made regarding the underlying distribution of data (e.g. it was assumed earlier that data were Gaussian distributed), as only separating boundaries are learned. Thus, discriminative models do not necessarily require training datasets to be representative of the base distribution – rather, they rely on data being representative of the classification boundaries. Because of this characteristic, and considering the proximity of queries to decision-boundaries, it is believed that discriminative classifiers may provide improved robustness to the type of sampling bias prevalent in risk-based active learning.

There are numerous approaches to probabilistic discriminative modelling. Arguably, *logistic regression* is the simplest example of such a model. For logistic regression, a separating plane (or hyperplane) between two classes is learned. A sigmoid function then maps the distance of a data point from the plane to the probability of class membership. Support vector machines (SVMs) are also a popular form of discriminative classifier that have found application in many domains including engineering [152]. SVMs seek to find (hyper)planes with maximal margins between classes such that classification errors in the training dataset are minimised; a principal in statistical learning theory known as *empirical risk minimisation*². Furthermore, SVMs are capable of determining nonlinear classification boundaries by utilising nonlinear kernels to project data into higher dimensions in which the data are more separable. Fundamentally, SVMs are deterministic classifiers; however, the outputs can be modified to have a probabilistic interpretation via post-processing in the form of Platt scaling [91].

For its robust uncertainty quantification, a Bayesian extension of the SVM, known as the *relevance vector machine* (RVM) is selected as the probabilistic discriminative classifier for the current chapter. Additionally, as RVMs are sparse models utilising a small subset of the training data, they potentially have enhanced robustness to sampling bias as problematic data can be ignored. Details of RVMs and their multiclass extension (mRVMs) are provided in the following subsection.

²Here, ‘risk’ is distinct from the definition used elsewhere in the current thesis. A mapping between the two could be formulated by considering the costs associated with misclassifications, however, this is outside the scope of the current thesis.

Relevance Vector Machines

Originally introduced by Tipping in [97], the RVM is a computationally-efficient Bayesian model capable of achieving high accuracies for both regression and classification tasks, via the use of basis functions specified via a kernel function and a sparse subset of the training data. In [146, 153], the RVM is extended from binary classification to multiclass classification via the multinomial probit link and multinomial probit likelihood. As previously mentioned, RVMs are able to achieve high computational efficiency; this is achieved by using a subset of n^* ‘relevant’ samples from \mathcal{D}_l such that $n^* \ll n$. This reduced subset is denoted \mathcal{A} . Two approaches for achieving this sparsity are presented in [144], and termed mRVM₁ and mRVM₂.

For the multiclass classification of an unlabelled data point $\tilde{\mathbf{x}}_t$, mRVMs employ a fundamentally (generalised) linear model-form as a foundation,

$$\mathbf{f}_t = \mathbf{W}^\top \mathbf{k}(\mathbf{x}_l, \tilde{\mathbf{x}}_t) \quad (9.22)$$

where $\mathbf{f}_t = \{f_1, \dots, f_K\}^\top$ is a vector of K auxiliary variables that provide a ranking system by which the class membership of an unlabelled data point may be assessed. $\mathbf{k}(\mathbf{x}_l, \tilde{\mathbf{x}}_t)$ is an $n \times 1$ vector for which the i^{th} element is specified by the *kernel function* $k(\mathbf{x}_i, \tilde{\mathbf{x}}_t)$. The kernel function specifies a set of basis functions that reflect the similarity between $\tilde{\mathbf{x}}_t$ and training inputs $\mathbf{x}_l = \{\mathbf{x}_i | (\mathbf{x}_i, y_i) \in \mathcal{D}_l\}_{i=1}^n$; nonlinearity can be introduced into the RVM by selecting a nonlinear kernel function. For compactness, the vector $\mathbf{k}(\mathbf{x}_l, \tilde{\mathbf{x}}_t)$ is denoted as \mathbf{k}_t herein. $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_K\}$ is an $n \times K$ matrix of tunable parameters referred to as *weights*, and where $\mathbf{w}_k = \{w_{1,k}, \dots, w_{n,k}\}^\top$. These weights act as a voting system that indicates which data in \mathcal{D}_l are important, or ‘relevant’, for discriminating between classes.

In accordance with [144], the auxiliary variables in \mathbf{f}_t are assumed to adhere to a standardised noise model:

$$f_k | \mathbf{w}_k, \mathbf{k}_t \sim \mathcal{N}(\mathbf{w}_k^\top \mathbf{k}_t, 1) \quad (9.23)$$

Predicted class labels \hat{y}_t are assigned to otherwise unlabelled data via the auxiliary variables \mathbf{f}_t by using a criterion specified by the multinomial probit link,

$$\hat{y}_t = k \iff f_k > f_j \forall j \neq k \quad (9.24)$$

Moreover, a probabilistic representation of class membership can be obtained via the following marginalisation,

$$p(\tilde{y}_t = k | \mathbf{W}, \mathbf{k}_t) = \int p(\tilde{y}_t = k | \mathbf{f}_t) p(\mathbf{f}_t | \mathbf{W}, \mathbf{k}_t) d\mathbf{f}_t \quad (9.25)$$

where $p(\tilde{y}_t = k | \mathbf{f}_t) = \delta_{k, \hat{y}_t}$. This marginalisation yields the multinomial probit likelihood [146],

$$p(\tilde{y}_t = k | \mathbf{W}, \mathbf{k}_t) = \mathbb{E}_{p(u)} \left[\prod_{j \neq k} \Phi(u + (\mathbf{w}_k - \mathbf{w}_j)^\top \mathbf{k}_t) \right] \quad (9.26)$$

where $u \sim \mathcal{N}(0, 1)$ and Φ denotes the Gaussian cumulative distribution function. Here, it is worth noting that, in practice, equation (9.26) cannot be computed analytically and is instead approximated via Gauss-Hermite quadrature [144].

Having specified the predictive model, the relevant supervised-learning problem for mRVMS can be expressed as the Bayesian inference of weights \mathbf{W} from \mathcal{D}_l . In order to conduct Bayesian inference, a prior must first be placed upon the model parameters \mathbf{W} ; in this case,

$$w_{i,k} \sim \mathcal{N}(0, \alpha_{i,k}^{-1}) \quad (9.27)$$

where the scale parameters $\alpha_{i,k}$ can be summarised in an $n \times K$ matrix \mathbf{A} and are assigned the following hyperprior,

$$\alpha_{i,k} \sim \Gamma(\tau, \nu) \quad (9.28)$$

where τ and ν are hyperparameters related to the shape and scale, respectively. Given small values for the hyperparameters τ and ν , the prior and hyperprior presented above result in a Student- t distribution with zero-mean and small variance over the weights in \mathbf{W} . This restrictive distribution causes few weights to be non-zero, thereby inducing sparsity in the model [97, 144].

The training of the mRVM model is accomplished by updating the model parameters via an EM algorithm [149]. For a detailed exposition of the learning procedure, the reader is directed to [153]. Here, the EM steps are provided.

The model weight parameters are updated as the MAP estimate of the posterior distribution over the weights; given by $\hat{\mathbf{W}} = \arg \max_{\mathbf{W}} p(\mathbf{W} | \mathbf{F}_l, \mathbf{K}_l, \mathbf{A}, \mathcal{D}_l)$, where $\mathbf{F}_l = \{\mathbf{f}_1, \dots, \mathbf{f}_K\}^\top$ are auxiliary variables for data in \mathcal{D}_l , and \mathbf{K}_l denotes $\mathbf{K}(\mathbf{x}_l, \mathbf{x}_l)$. It follows that, for a given class, the update for the weights across data in \mathcal{D}_l is given by,

$$\hat{\mathbf{w}}_k = (\mathbf{K}_l \mathbf{K}_l^\top + \mathbf{A}_k)^{-1} \mathbf{K}_l \mathbf{f}_k^\top \quad (9.29)$$

where \mathbf{A}_k is a diagonal matrix formed from the k^{th} column of \mathbf{A} , i.e. $\mathbf{A}_k = \text{diag}(\alpha_{1,k}, \dots, \alpha_{n,k})$.

Following the formulation in [144, 153], the posterior distribution over the auxiliary variables is derived to be a product of truncated Gaussian distributions³. For a given class j , the auxiliary variables can be updated $\forall k \neq j$ as follows,

$$\tilde{f}_{k,i} \leftarrow \hat{\mathbf{w}}_k^\top \mathbf{k}_i - \frac{\mathbb{E}_{p(u)} [\mathcal{N}(\hat{\mathbf{w}}_k^\top \mathbf{k}_i - \hat{\mathbf{w}}_j^\top \mathbf{k}_i, 1) \prod_{\kappa \neq j,k} \Phi(u + \hat{\mathbf{w}}_j^\top \mathbf{k}_i - \hat{\mathbf{w}}_\kappa^\top \mathbf{k}_i)]}{\mathbb{E}_{p(u)} [\Phi(u + \hat{\mathbf{w}}_j^\top \mathbf{k}_i - \hat{\mathbf{w}}_k^\top \mathbf{k}_i) \prod_{\kappa \neq j,k} \Phi(u + \hat{\mathbf{w}}_j^\top \mathbf{k}_i - \hat{\mathbf{w}}_\kappa^\top \mathbf{k}_i)]} \quad (9.30)$$

and for class j ,

$$\tilde{f}_{j,i} \leftarrow \hat{\mathbf{w}}_j^\top \mathbf{k}_i - \sum_{\kappa \neq j} (\tilde{f}_{\kappa,i} - \hat{\mathbf{w}}_\kappa^\top \mathbf{k}_i) \quad (9.31)$$

where $\tilde{f}_{k,i}$ denotes the mean of the distribution over the latent auxiliary variable corresponding to class k for data point i in \mathcal{D}_l .

By continuing to follow [153], one can derive the update for the hyperpriors $\alpha_{i,k}$ as the mean of the posterior distribution $p(\mathbf{A} | \mathbf{W}, \mathcal{D}_l)$, given by,

³Derivations and visualisations for conically-truncated Gaussian distributions are shown in [154].

$$\tilde{\alpha}_{i,k} = \frac{2\tau + 1}{w_{i,k}^2 + 2\nu} \quad (9.32)$$

The iterative learning process requires the repeated application of the updates given in equations (9.32), (9.29), (9.30) and (9.31), until a convergence criterion is met. Suitable convergence criteria are provided in [144].

As discussed in Section 9.2.2, RVMS utilise a subset of ‘relevant’ samples to construct basis functions, with mRVM₁ adopting a bottom-up approach and mRVM₂ adopting a top-down approach.

For mRVM₁, \mathcal{A} is initialised as an empty set with samples subsequently added or removed from \mathcal{A} based upon their contribution to the objective function given by the marginal log-likelihood $\mathcal{L}(\mathbf{A}) = \log p(\mathbf{F}_l | \mathbf{K}_l, \mathbf{A}) = \log \int p(\mathbf{F}_l | \mathbf{K}_l, \mathbf{W}) p(\mathbf{W} | \mathbf{A}) d\mathbf{W}$. In accordance with [144], to ensure $\mathcal{L}(\mathbf{A})$ is differentiable such that the fast type-II Maximum-Likelihood approach [155], can be employed, it is assumed that the scale parameter for each sample is common across classes, i.e. $\forall k \in \{1, \dots, K\}$, $\alpha_{i,k} = \alpha_i$. By following the procedure detailed in [155], one arrives at the following marginal likelihood decomposition,

$$\mathcal{L}(\mathbf{A}) = \mathcal{L}(\mathbf{A}_{-i}) + \ell(\alpha_i) \quad (9.33)$$

where,

$$\mathcal{L}(\mathbf{A}_{-i}) = \sum_{k=1}^K -\frac{1}{2} \left[N \log 2\pi + \log |\mathcal{C}_{-i}| + \mathbf{f}_k^\top \mathcal{C}_{-i}^{-1} \mathbf{f}_k \right] \quad (9.34)$$

and,

$$\ell(\alpha_i) = \sum_{k=1}^K \frac{1}{2} \left[\log \alpha_i - \log(\alpha_i + s_i) + \frac{q_{k,i}^2}{\alpha_i + s_i} \right] \quad (9.35)$$

with $\mathcal{C}_{-i} = \mathbf{I}_{n-1} + \mathbf{K}_{l,-i}^\top \mathbf{A}_{-i}^{-1} \mathbf{K}_{l,-i}$, where \mathbf{I}_{n-1} denotes the identity matrix of size $n-1$, and where the subscript $-i$ is used to indicate matrices with entries corresponding to the i^{th} data point in \mathcal{D}_l removed. The quantities s_i and $q_{k,i}$, introduced in equation (9.35) are given by [155],

$$s_i = \mathbf{k}_i^\top \mathbf{C}_{-i}^{-1} \mathbf{k}_i \quad \text{and} \quad q_{k,i} = \mathbf{k}_i^\top \mathbf{C}_{-i}^{-1} \mathbf{f}_k \quad (9.36)$$

and can be interpreted as a ‘sparsity factor’ and ‘quality factor’, respectively. The sparsity factor indicates how much the descriptive information provided by the i^{th} data point is already provided by the existing samples. The quality factor provides a measure of the i^{th} sample’s ability to describe class k . Maximising $\mathcal{L}(\mathbf{A})$ with respect to α_i by following the procedure presented in [144, 155], one can quantify the contribution of data point i to the objective function as,

$$\theta_i = \sum_{k=1}^K q_{k,i}^2 - K s_i \quad (9.37)$$

For each iteration in the learning procedure, the contribution θ_i is then used to construct \mathcal{A} with samples satisfying $\theta_i > 0$ included, and other samples excluded. The model parameter update in equation (9.29) can be expressed utilising the sparse subset of data \mathcal{A} as follows:

$$\hat{\mathbf{W}}_* = (\mathbf{K}_* \mathbf{K}_*^\top + \mathbf{A}_*)^{-1} \mathbf{K}_* \tilde{\mathbf{F}}^\top \quad (9.38)$$

where \mathbf{K}_* denotes $\mathbf{K}(\mathbf{x}_l, \mathcal{A})$ and is $n^* \times n$, and \mathbf{A}_* is $n^* \times n^*$. Finally, the update given in equation (9.32) becomes,

$$\alpha_i = \frac{K s_i^2}{\theta_i} \quad (9.39)$$

Whereas mRVM₁ provides a constructive approach to the formation of \mathcal{A} of relevance vectors, the approach termed mRVM₂ sculpts \mathcal{A} from \mathcal{D}_l . This is achieved by excluding samples with scales $\alpha_{i,k}$ sufficiently large that $w_{i,k}$ is negligible. The i^{th} data point can be considered insignificant and can be removed from \mathcal{A} when $\alpha_{i,k} > 10^5 \forall k \in \{1, \dots, K\}$. Here, it is worth noting that, within the mRVM₂ approach, once a sample has been removed from \mathcal{A} it cannot be reintroduced to the model. This characteristic is in contrast to mRVM₁, which allows a previously pruned sample to be reintroduced if, during a later iteration, the sample is deemed to have positive contribution. In the context of risk-based active learning, this constraint of mRVM₂ should prove to be insignificant as, within the active-learning process, the

model is repeatedly retrained, with all data in \mathcal{D}_t considered.

By using the sparse subset \mathcal{A} to form the basis functions for the mRVM classification model, improvements in computational efficiency are achieved. Furthermore, in the context of risk-based active learning, it is hypothesised that probabilistic discriminative classifiers will show robustness to sampling bias over the generative classifiers – in part because detrimental or superfluous data are excluded from the model by virtue of its sparsity, and because discriminative classifiers in general, do not rely upon assumptions regarding the underlying distribution of the data.

Results

Both mRVM₁ and mRVM₂ were incorporated into the risk-based active-learning process and applied to the case study presented in Section 9.1.1. As it is well-studied and flexible, both models were formed using a Gaussian kernel. The Gaussian kernel function has the following form,

$$k(\mathbf{x}_t, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x}_t - \mathbf{x}_i\|^2) \quad (9.40)$$

where γ is introduced as a hyperparameter and is fixed at $\frac{1}{D}$ in accordance with [144, 156].

As with the previous statistical classifiers, the risk-based active-learning process was applied to each model 1000 times with randomly-selected initial labelled datasets \mathcal{D}_t .

Figures 9.16 and 9.17 show, from one of the 1000 repetitions, an mRVM₁ classifier before and after the risk-based active-learning process, respectively.

Figure 9.16 shows that the mRVM₁ model, subject to the convergence criteria presented in [144], was unable to effectively discriminate between classes when learned from the initial limited subset of training data. In Figure 9.16a, the absence of contours corresponding to $p(y_t = k|\mathbf{x}_t) = 0.5$ indicates that, if using the initial model shown, one would be unable to classify data with a high degree of confidence. This result is corroborated by the approximately uniform EVPI over the feature space, presented in Figure 9.16b.

Figure 9.17 indicates that, following the risk-based active learning procedure, the

updated $mRVM_1$ classifier has established classification boundaries – the contours corresponding to $p(y_t = k|\mathbf{x}_t) = 0.5$ are now visible and, in general, are a good fit to the data. In Figure 9.17a, similar to the GMM, the risk-based active-learning algorithm preferentially queries data close to the boundary between Class 3 (significant damage) and Class 4 (critical damage). From Figure 9.17b, it can be seen that a band of high EVPI has been established close to the classification boundary between classes 3 and 4. Again, this band corresponds to a decision boundary and indicates the region of the feature space where a decision-making agent should inspect the structure; below this band the agent can be confident that the optimal decision is ‘do nothing’, and above this band the agent can be confident that the optimal decision is ‘perform maintenance’.

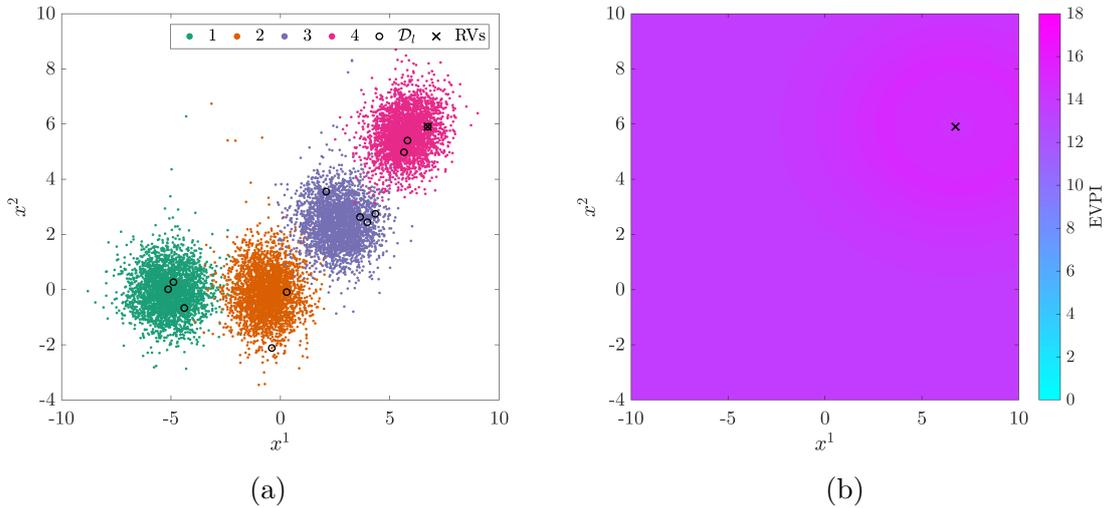


Figure 9.16: An $mRVM_1$ statistical classifier $p(y_t|\mathbf{x}_t; \mathbf{W})$ prior to risk-based active learning; relevance vectors are shown (\times). (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

Similar to Figures 9.16 and 9.17, Figures 9.18 and 9.19 show, from one of the 1000 repetitions, an $mRVM_2$ classifier before and after the risk-based active-learning process.

Unlike $mRVM_1$, the $mRVM_2$ approach to relevance-vector selection resulted in an initial model capable of discriminating between the four classes – as demonstrated by the contours corresponding to $p(y_t = k|\mathbf{x}_t) = 0.5$ visible in Figure 9.18. In general, the discriminative boundaries fit the data well, except for that corresponding to Class 2 (minor damage). The resulting EVPI distribution over the feature space, shown in Figure 9.18b, indicates that a decision boundary (albeit somewhat nebulous) has

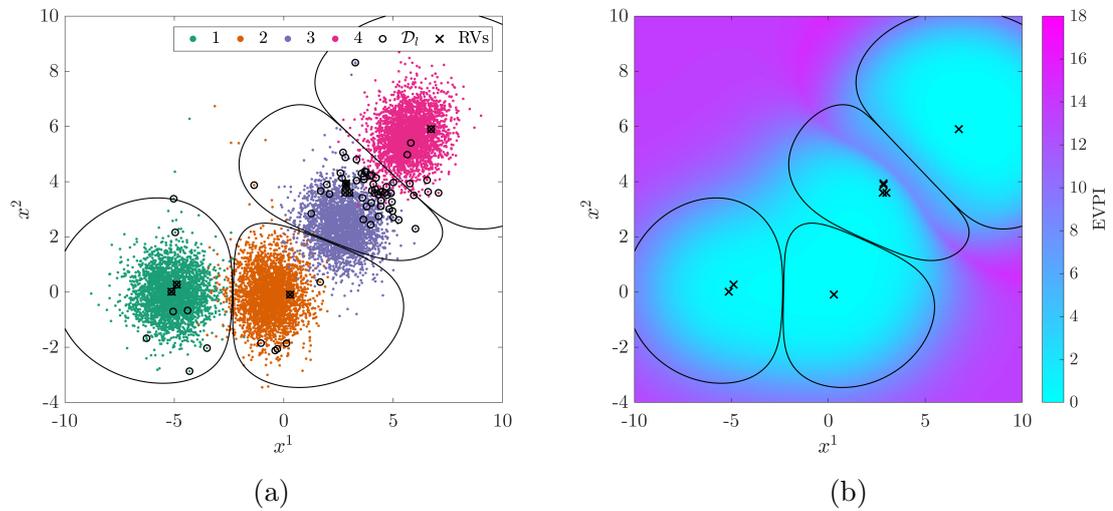


Figure 9.17: An mRVM_1 statistical classifier $p(y_t|\mathbf{x}_t; \mathbf{W})$ after risk-based active learning; relevance vectors (\times) and contours (lines) denoting $p(y_t = k|\mathbf{x}_t) = 0.5$ are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

been established, even from the very limited labelled data.

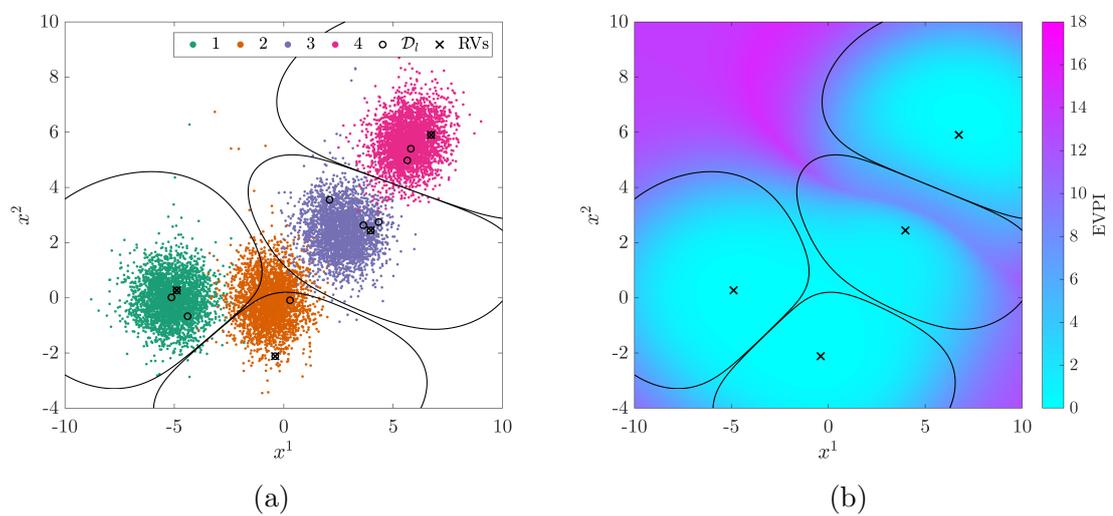


Figure 9.18: An mRVM_2 statistical classifier $p(y_t|\mathbf{x}_t; \mathbf{W})$ prior to risk-based active learning; relevance vectors (\times) and contours (lines) denoting $p(y_t = k|\mathbf{x}_t) = 0.5$ are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

It can be seen from Figure 9.19a that, once again, the active-learning algorithm results in data being queried preferentially close to the boundary between Classes 3 and 4. It can be seen from the probability contours that the updated model fits the

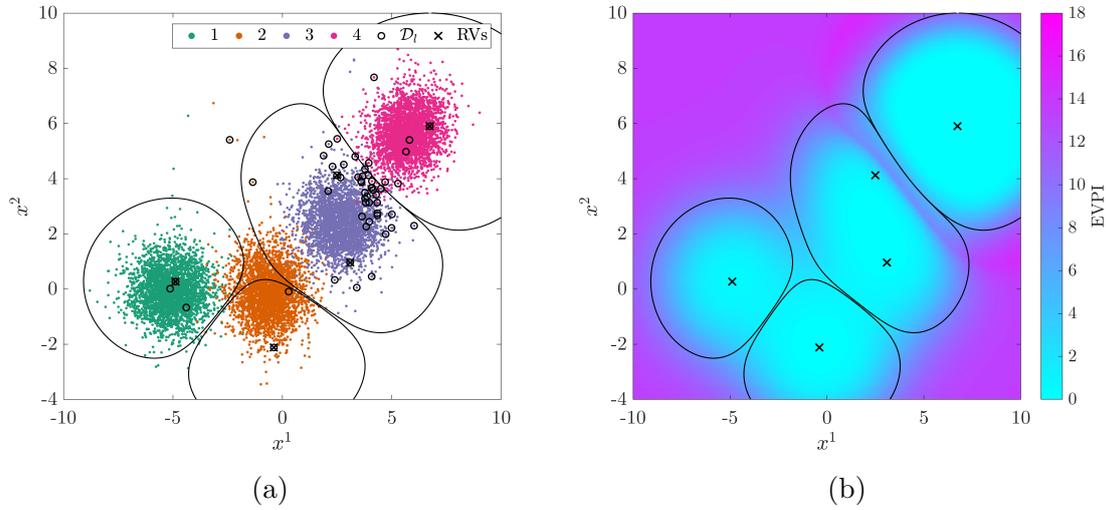


Figure 9.19: An mRVM₂ statistical classifier $p(y_t|\mathbf{x}_t; \mathbf{W})$ after risk-based active learning; relevance vectors (\times) and contours (lines) denoting $p(y_t = k|\mathbf{x}_t) = 0.5$ are shown. (a) shows the final model overlaid onto the data with labelled data \mathcal{D}_l encircled and (b) shows the resulting EVPI over the feature space.

data marginally better than the initial model. From Figure 9.19b, it can be seen that the inferred decision boundary is now more well-defined, represented by the narrow band of high EVPI close to the classification boundary between Classes 3 and 4.

From the EVPI surfaces shown in Figures 9.16b, 9.17b, 9.18b, and 9.19b, it is apparent that, when using an mRVM model, data far from those already observed are classified with high uncertainty, meaning that high EVPI would be associated with any outlying observations. This result is in contrast to EVPI surfaces obtained when utilising generative models, where, as discussed in Section 9.1.2, sampling bias resulted in over-confident predictions for outlying data. The decision support implications for the differing ‘attitudes’ towards outliers, induced by the choice of a generative versus discriminative classifier, are discussed further in Section 9.4.

Figures 9.20a and 9.20b show the class proportions in the labelled dataset \mathcal{D}_l , averaged over the 1000 runs of risk-based active learning applied to mRVM₁ and mRVM₂, respectively. Both figures show similar trends in class proportions as the original GMM, with Class 3 disproportionately represented ($\sim 60\%$) in the final iteration of \mathcal{D}_l , above Classes 1, 2, and 4. Here, it is worth recognising from Figures 9.17 and 9.19, that within the sparse subset \mathcal{A} , the classes are more equally represented. For mRVM₁, Class 1 corresponds to two of the eight relevance vectors, Classes 2 and 4 each correspond to one, and Class 3 corresponds to four of the eight relevance vectors.

Similarly, for mRVM_2 , Classes 1, 2 and 4 are each represented by one relevance vector, with Class 3 represented by two out of the five relevance vectors. This result suggests that utilising a sparse subset of data somewhat mitigates the prevalence of sampling bias in the training data.

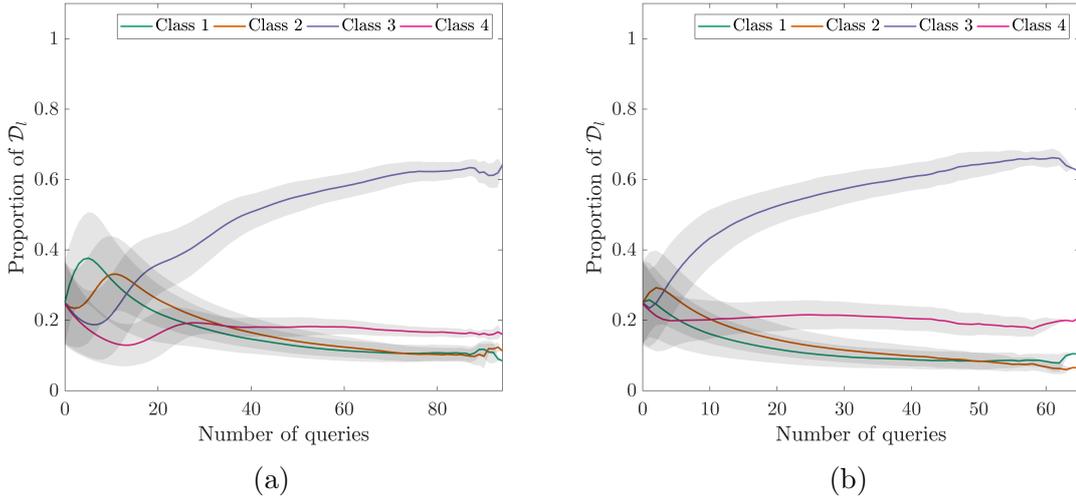


Figure 9.20: Variation in class proportions within \mathcal{D}_l with number of label queries for an agent utilising an: (a) mRVM_1 and (b) mRVM_2 statistical classifier learned from \mathcal{D}_l extended via risk-based active learning. Shaded area shows $\pm 1\sigma$.

The histograms presented in Figure 9.21 show the distributions of the number of queries made by the risk-based active-learning algorithm as applied to mRVM_1 and mRVM_2 for each of the 1000 runs. Both mRVM_1 and mRVM_2 result in fewer queries than the GMM. Again, this result is significant, as it indicates that expenditure on queries has been reduced. Furthermore, the variance of the distributions corresponding to the mRVM algorithms is reduced compared to that for the GMM, indicating that relatively few queries are made, consistently.

Figure 9.22 compares the total number of queries for each index in \mathcal{D}_u over the 1000 repetitions of risk-based active learning conducted using a GMM, an mRVM_1 classifier, and an mRVM_2 classifier. It can be seen from Figure 9.22, that data points queried by the GMM formulation also tend to be queried by the mRVM formulations, though with a greatly-reduced total. In contrast to the semi-supervised approaches presented in Section 9.2.1, the risk-based active learning algorithms built upon mRVM s query data more heavily from the later occurrences of Classes 3 and 4. This observation can, in part, be attributed to the fact that the high-EVPI regions associated with the decision boundary inferred via the mRVM models are slightly

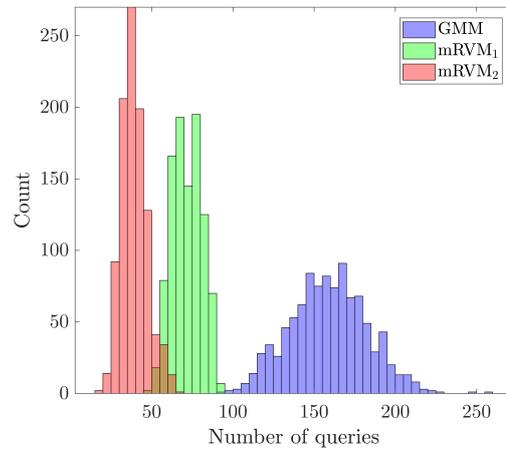


Figure 9.21: Histograms showing the distribution of the number of queries from 1000 runs of the risk-based active learning of: (i) a GMM (blue) (ii) an mRVM₁ (green) and (iii) an mRVM₂ (red) statistical classifier.

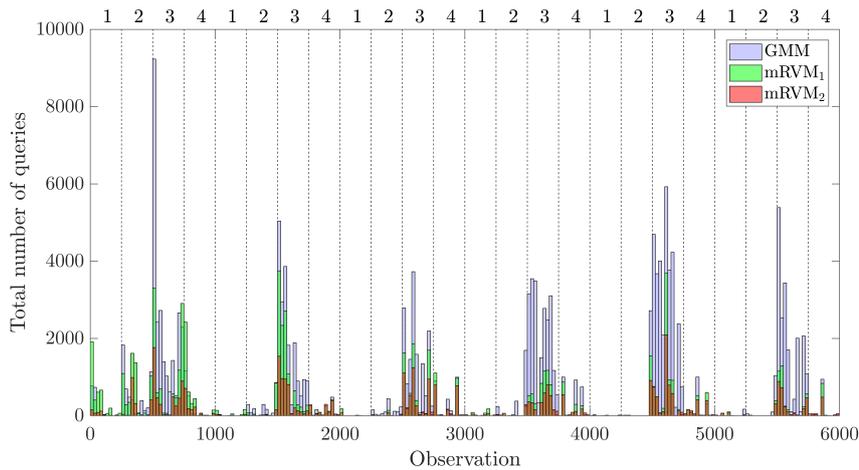


Figure 9.22: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u from 1000 runs of risk-based active learning for: (i) a GMM (blue) (ii) an mRVM₁ (green) and (iii) an mRVM₂ (red) statistical classifier. The average location of classes within \mathcal{D}_u are numbered on the upper horizontal axis and transitions are denoted as a dashed line.

more nebulous than those obtained via semi-supervised learning. In addition, when using the mRVM classifiers, one would expect outlying data to be queried at any time during the risk-based active-learning process because of the uncertainty in the class-label prediction.

Figure 9.23 provides a comparison of the median decision accuracies and f_1 -scores throughout the querying processes aggregated from the 1000 runs of the risk-based

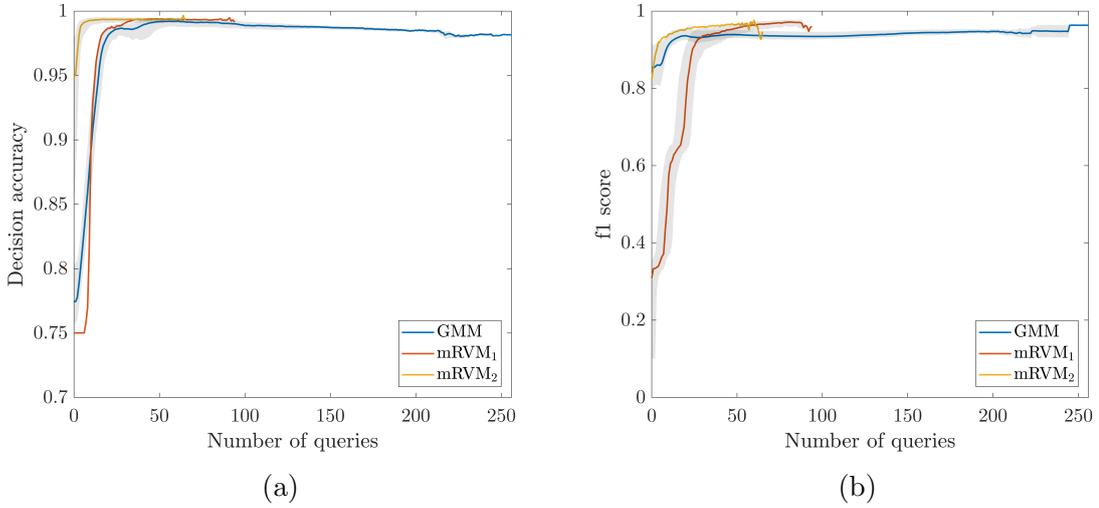


Figure 9.23: Variation in median: (a) decision accuracy and (b) f_1 -score with number of label queries for an agent utilising: (i) a GMM (ii) an mRVM₁ and (iii) an mRVM₂ statistical classifier learned via risk-based active learning. Shaded area shows the interquartile range.

active-learning process, applied to the mRVM₁ and mRVM₂ classifiers. It can be seen from Figure 9.23a, that both mRVM₁ and mRVM₂ result in superior decision accuracy performance over the GMM, plateauing at a slightly higher decision accuracy and reaching that plateau earlier. Although mRVM₁ begins with a lower decision accuracy than the GMM, the performance of the decision-maker rapidly improves with queries made. On the other hand, mRVM₂ begins with significantly higher decision accuracy than both the GMM and the mRVM₁ classifiers, yet still improves rapidly with the early queries. From Figure 9.23a, an important observation regarding the use of the discriminative models is that, for later queries, there is no decline in decision-making performance as additional labelled data are obtained, indicating that such models, as hypothesised, possess improved robustness to sampling bias. Figure 9.23b shows the f_1 -score classification performance of the models on the test dataset. Both models achieve improved classification performance over the GMM.

In summary, for the current case study, substituting the generative GMM classifier out of the risk-based decision process, in favour of RVM-based discriminative models, has mitigated the adverse effects on decision-making performance caused by the inherent bias introduced via the risk-based active-learning process. Furthermore, well-defined decision boundaries were inferred from a reduced number of queries as compared to the generative model. These results were achieved as discriminative models do not rely on prior assumptions regarding the distribution of data.

9.3 Case Study: Z24 Bridge Dataset

In the previous section, the effects of sampling bias on decision-making performance, and approaches for counteracting these effects, were demonstrated on a somewhat idealised case study, in which data were Gaussian distributed and classes were fairly separable. In the current section, the effectiveness of semi-supervised learning, and discriminative classifiers, at mitigating the effects of sampling bias, are assessed using the experimental dataset obtained from the Z24 Bridge introduced in the previous chapter [127]. For reference, visualisations of the Z24 Bridge dataset are provided in 9.24.

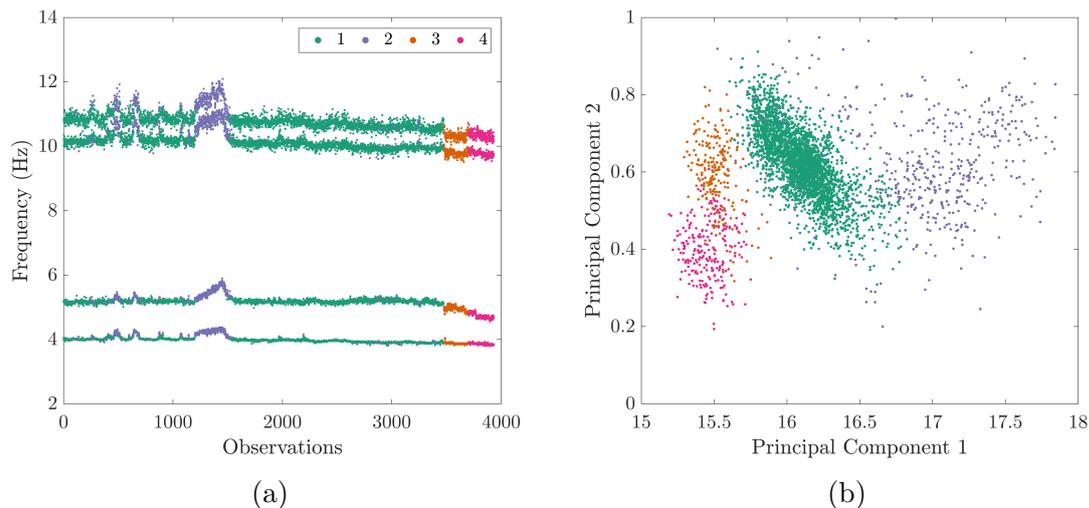


Figure 9.24: Visualisation of the Z24 Bridge dataset in: (a) the discrete time and (b) the projected feature space.

Upon examination of Figure 9.24b, it becomes apparent that there is significant overlap between Class 1 and Class 2. There is also some degree of overlap between the data corresponding to Classes 1 and 3, Classes 2 and 3, and Classes 3 and 4. Moreover, the distribution of Z24 Bridge data deviates somewhat from Gaussianity. Finally, it is obvious that class imbalance is present in the dataset, with most observations corresponding to Class 1. Combined, these characteristics of the Z24 Bridge dataset should represent a greater challenge to the risk-based active-learning approaches demonstrated on a more idealised case study in Section 9.2.

9.3.1 Decision Process

As with the previous case study, an O&M decision process must be specified for the Z24 Bridge dataset in order to apply a risk-based active learning approach to classifier development. A decision process identical to that used for the Z24 Bridge case study in the previous chapter is adopted here. As such, the graphical representation of the process provided by the influence diagram in Figure 8.3, applies also to the current case study. The parameters populating the conditional probability distributions and utility functions must be specified to reflect the operational context of the Z24 Bridge and can be found in Tables 8.5 to 8.8.

Again, the cost of inspection is specified to be $C_{\text{ins}} = 30$. The moderate cost is intended to reflect resources required to inspect a large-scale structure such as the Z24 Bridge.

9.3.2 Results: Gaussian Mixture Model

To facilitate comparison, risk-based active learning was applied to a Gaussian mixture model learned via the Bayesian approach outlined in Section 8.4.2. For the Z24 Bridge case study, the features used to discriminate between the four health states of the structure were the first four natural frequencies shown in Figure 9.24a, such that $\mathbf{x}_t \in \mathbb{R}^4$.

The dataset was separated into two halves. One half formed the training dataset \mathcal{D} , with the other half forming an independent test set. A small (0.3%) random subset of \mathcal{D} was used to initialise the labelled dataset \mathcal{D}_l . The remaining data were assigned to \mathcal{D}_u to be sequentially evaluated with respect to the decision process in the risk-based active-learning process. Once again, 1000 repetitions of the active-learning process were conducted, each with randomly-initialised \mathcal{D}_l .

Figure 9.25 shows the median and interquartile range for the decision accuracy and f_1 -score as a function of number of queries. Additionally, the performance measures are provided for an agent utilising a GMM trained using an equivalent number of samples obtained via unguided (random) querying. From Figure 9.25a, it can be seen that risk-based active learning results in decision-making performance improving at an increased rate compared to random sampling. However, yet again, a gradual decline in decision accuracy can be observed over later queries when said queries

are guided by EVPI. Figure 9.25b shows that the classification performances of the predictive model are fairly similar for EVPI-based and random querying. Each gradually improves with number of queries, albeit non-monotonically. The large interquartile ranges for both approaches to querying indicate that the classification performance is highly sensitive to which data are available in the supervised learning process – a phenomenon reported in [95].

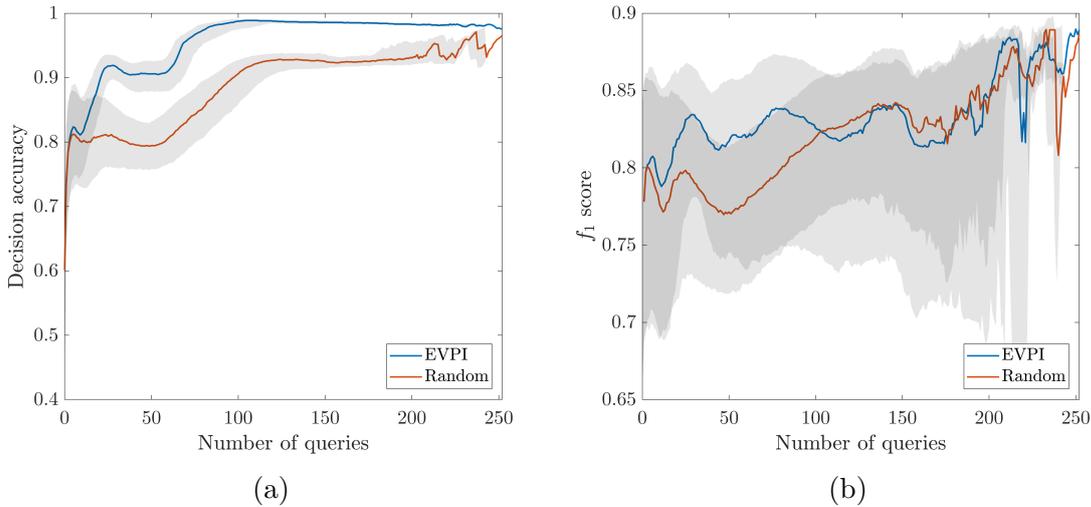


Figure 9.25: Variation in median: (a) decision accuracy and (b) f_1 score with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended via: (i) risk-based active querying (EVPI) and (ii) random sampling (Random). Shaded area shows the interquartile range.

Figure 9.26 shows the mean proportions that each class contributes to the labelled dataset \mathcal{D}_l throughout the learning process. Figure 9.26a provides the class proportions for EVPI-based querying, whereas Figure 9.26b provides the proportions for random querying. Logically, both datasets show an increase in the proportion for Class 1, as this is the first class to present itself from the perspective of the decision-maker. One difference between datasets is that the risk-based active-learning approach results in Class 2 (cold temperature) gaining an elevated level of representation in \mathcal{D}_l ; this is perhaps understandable if one considers the overlap present in the dataset, especially that between Class 2 and Class 3. Throughout later queries, one can see that Class 1 gains increased representation in \mathcal{D}_l . This increase correlates with the slight decline in decision accuracy observed, indicating that data acquired during this time cause the cluster corresponding to Class 1 to be altered such that it is suboptimal for decision-making.

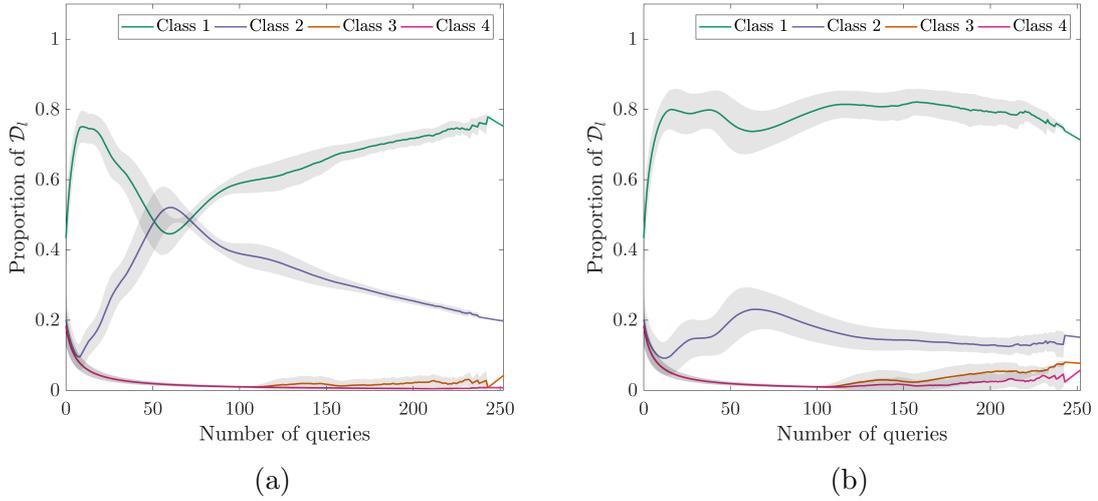


Figure 9.26: Variation in class proportions within \mathcal{D}_l with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended via: (a) risk-based active querying and (b) random sampling. Shaded area shows $\pm 1\sigma$.

Figure 9.27 shows the total number of times each observation in \mathcal{D}_u was queried throughout the 1000 runs, with the observations in \mathcal{D}_u presented for reference. As one would expect, random sampling results in uniform querying across the dataset. On the other hand, the risk-based active learning approach exhibits definite preferences for specific observations. One tends to see an increase in observation data corresponding to previously-unseen classes; however, the exception to this is Class 4 (advanced damage) – indicating that the data in the vicinity of the cluster corresponding to Class 4 have low value of information. This result is, of course, because a decision-maker observing data in this region of the feature space can be certain that the optimal policy is ‘perform maintenance’ because of the extremely high cost associated with this class. The most heavily-queried data points are in the vicinity of observation number 1600. These data belong to Classes 1 or 2 – both undamaged classes. The fact that they are so heavily queried, however, implies that they consistently lie close to the decision boundary between the candidate actions ‘do nothing’ and ‘perform maintenance’.

To summarise, when utilising a GMM as the statistical classifier within a decision process for the Z24 Bridge dataset, one can obtain improved decision-making performance; however, as with the previous case study, a decline in performance still occurs with queries that are made later in the learning process.

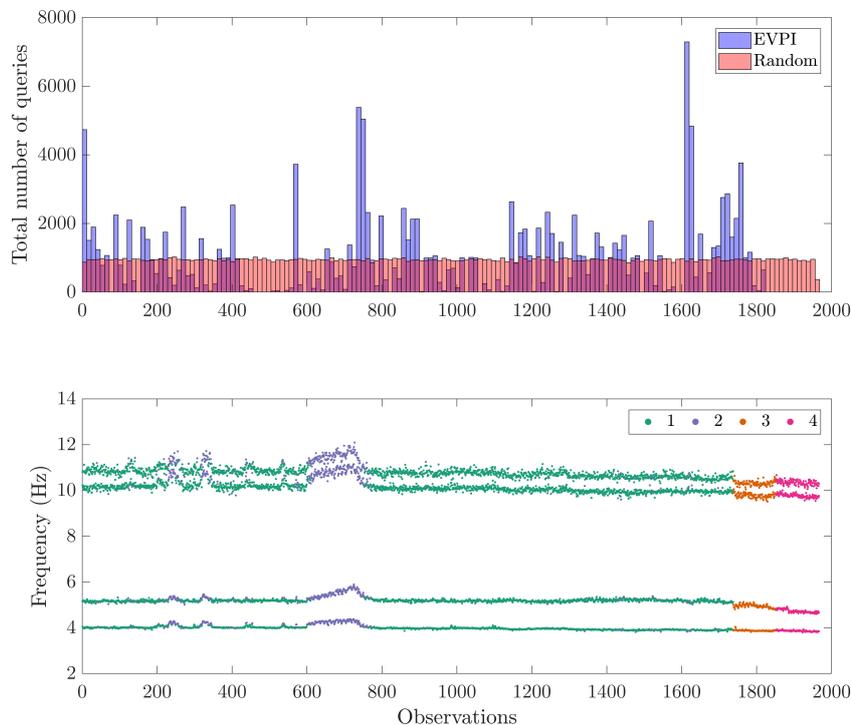


Figure 9.27: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u over 1000 runs when adopting: (a) risk-based active learning (EVPI) and (b) random sampling (Random) in order to learn a GMM. For reference, the unlabelled dataset is also provided.

9.3.3 Results: Semi-supervised Learning

The modified approaches to risk-based active learning, incorporating semi-supervised learning via EM and latent-state smoothing, were each applied to the Z24 Bridge case study.

Figure 9.28 shows, for each algorithm, the median decision and classification performances as a function of the number of queries. Additionally, the performances for the risk-based active-learned GMM are provided for comparison.

It is immediately apparent from both Figure 9.28a and Figure 9.28b, that the semi-supervised approaches have failed in improving robustness to sampling bias; in fact, the semi-supervised learning approaches have been detrimental to decision-making performance. Neither expectation-maximisation updates to the generative model parameters, nor pseudo-labelling of unlabelled data via a smoothing algorithm,

result in a classifier that surpasses the standard GMM in terms of decision accuracy. Similarly, both approaches fail to surpass the standard GMM in terms of classification performance.

For the current case study, the poor performance of the semi-supervised methods can be attributed to the fact that the algorithms further increase dependence on the key assumption required for learning generative models; namely, that the density estimations selected are representative of the underlying distribution. This increased dependence arises as the generative models, learned from an unrepresentative dataset, are being used to inform the pseudo-labels that are subsequently used to update the model. As the Z24 Bridge data is non-Gaussian and as Gaussian density estimation was used, the semi-supervised steps in the modified risk-based learning algorithm yield biased pseudo-labels and therefore exacerbate the effects of sampling bias. These results correspond with the results of the active-learning EM approach in [157], in which introducing EM to the active-learning process degrades the classification performance of a GMM on the Z24 Bridge dataset.

Figure 9.28 shows that the smoothing approach to semi-supervised learning outperforms EM in terms of decision accuracy and f_1 -score. This observation is likely because the smoothing algorithm utilises temporal information from the transition model. This additional information source somewhat tempers the dependence on the (biased) classification model that one solely relies on in the EM approach.

Figure 9.29 shows the effective class proportions for data in the training set \mathcal{D} , throughout the risk-based active learning of semi-supervised GMMs. From these figures one can see that the semi-supervised algorithms initially induce a heavy bias in favour of Class 1 within the training dataset. These class proportions are likely to inflate both the component mixture weights and the covariance for the cluster belonging to Class 1. This initial bias explains the dramatic decrease in decision accuracy and f_1 -score that can be observed over the first few queries in Figure 9.28.

In addition to decreasing the decision-making performance, the introduction of semi-supervised learning to risk-based active learning also has a detrimental effect on the number of queries made. Figure 9.30 shows the distribution of the number of queries made over 1000 randomly-initialised repetitions. It is clear from Figure 9.30, that both EM and latent-state smoothing result in a higher mean for the number of queries made, when compared to the standard GMM. This result implies that using these modified approaches to risk-based active learning result in increased expenditure on

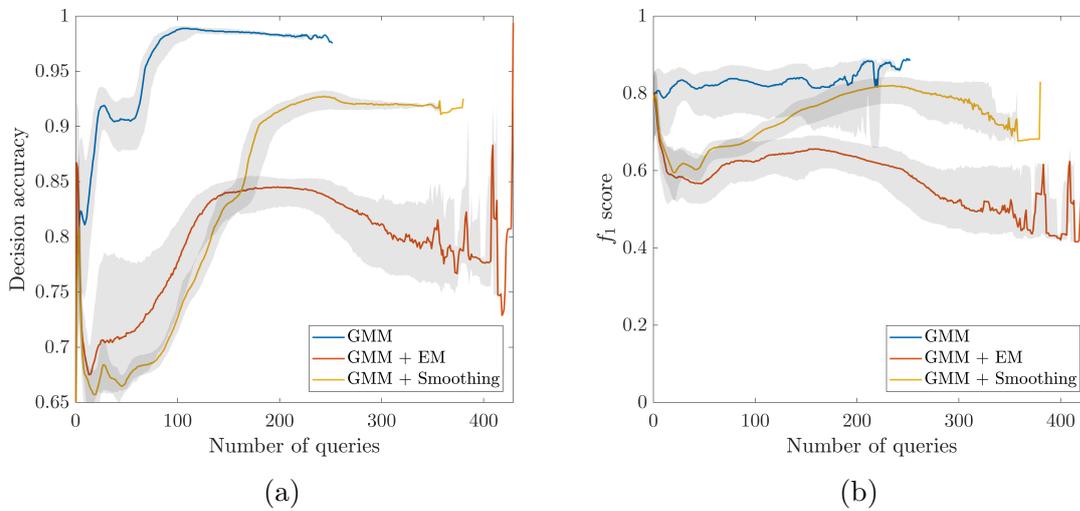


Figure 9.28: Variation in median: (a) decision accuracy and (b) f_1 score with number of label queries for an agent utilising a GMM learning via risk-based active learning: (i) without semi-supervised updates, (ii) with semi-supervised updates via EM and (iii) with semi-supervised updated via smoothing. Shaded area shows the interquartile range.

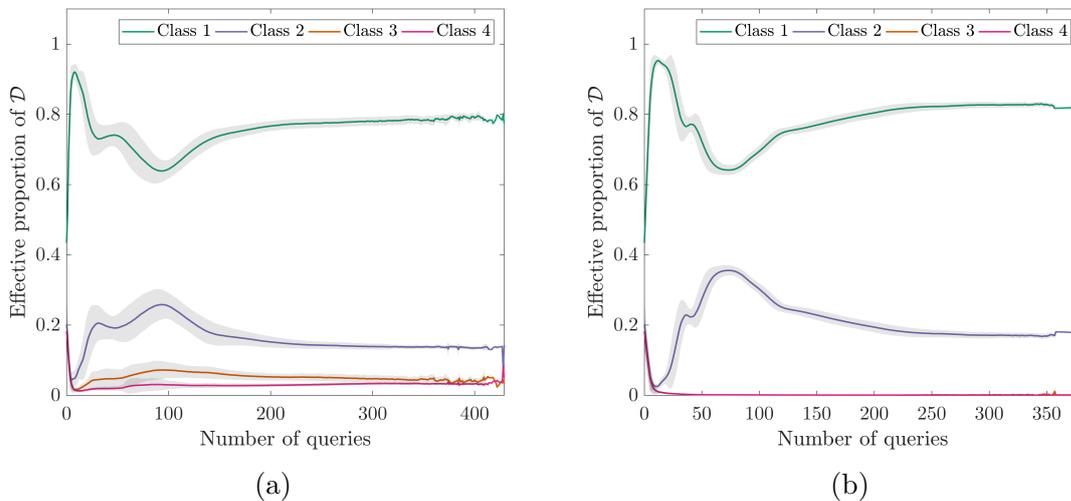


Figure 9.29: Variation in class proportions within \mathcal{D}_l with number of label queries for an agent utilising a GMM learned from \mathcal{D}_l extended risk-based active learning with semi-supervised updating via: (a) expectation-maximisation and (b) latent-state smoothing. Shaded area shows $\pm 1\sigma$.

structural inspections. Interestingly, the distribution of the number of queries made using the risk-based active learning algorithm that incorporates EM, is bimodal. The dominant mode corresponds to worsened performance in comparison to the standard GMM; however, the secondary mode corresponds to an improved performance. This

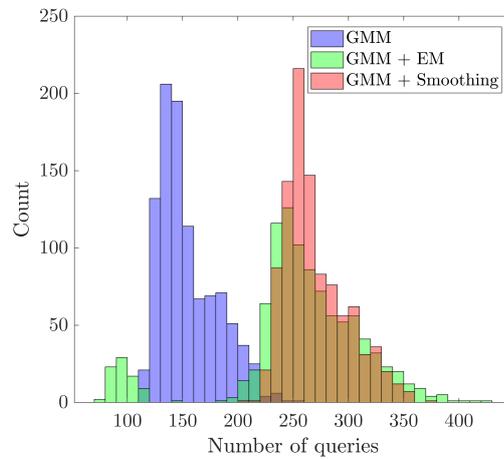


Figure 9.30: Histograms showing the distribution of the number of queries from 1000 runs of the risk-based active learning of: (i) a GMM (blue) (ii) a GMM semi-supervised via expectation-maximisation (green) and (iii) a GMM semi-supervised via latent-state smoothing (red).

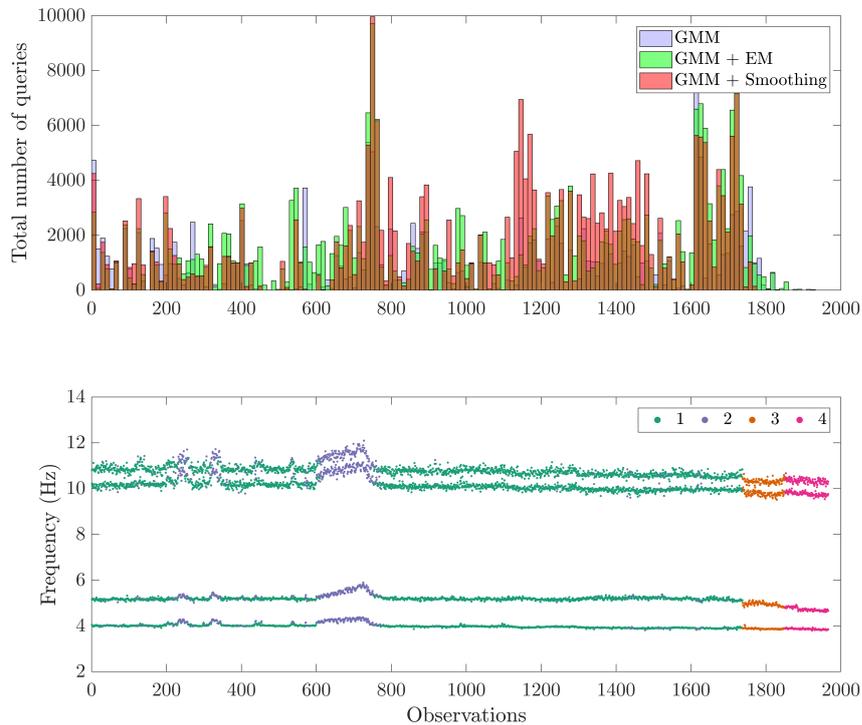


Figure 9.31: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u from 1000 runs of risk-based active learning for: (i) a GMM (blue) (ii) a GMM with EM (green) and (iii) a GMM with smoothing (red).

result suggests that there exists a modality of the learning algorithm in which the EM algorithm can converge to a classifier that results in a well-defined decision boundary. As the variability in results arises because of the randomisation of the initial labelled dataset \mathcal{D}_l , it can be inferred that the occurrence of this secondary modality is conditional on which labelled data are available for learning the fully-supervised model, prior to the EM updates. It can be seen from Figure 9.30 that more than 400 queries were made in very few trials; this explains the collapse in the variability in decision and classification performance observed in Figure 9.28.

Figure 9.31 shows the total number of queries for each observation index in \mathcal{D}_u , over the 1000 repetitions conducted for each of the two variants of the risk-based active learning process. Additionally, the queries made by the agent utilising the standard GMM are provided for comparison. As one would expect, given the histograms in Figure 9.30, the queries made by the semi-supervised algorithms all but obscure those made when using the standard GMM. These semi-supervised algorithms result in data being heavily queried across many observations in \mathcal{D}_u , indicating that a well-defined decision boundary is not obtained - an effect of the exacerbated sampling bias. Here, it is worth noting that, despite numerous queries elsewhere in the dataset, Class 4 remains more-or-less unqueried. Again, this result indicates that, because of the extreme cost associated with the advanced damage state, a decision-maker will confidently select the action ‘perform maintenance’ when in the vicinity of the cluster corresponding to Class 4.

In summary, the introduction of semi-supervised learning via EM, and latent-state smoothing, resulted in a degradation in the performance of an agent tasked with selecting actions within a maintenance decision process defined around the Z24 Bridge dataset. This degradation occurred because of the interaction between important characteristics of the dataset (class imbalance, non-Gaussianity, and non-separability) and the strengthened assumptions (iid training data, and representative distribution forms) invoked via the semi-supervised learning algorithms. These interactions resulted in the amplification of sampling bias, leading to worsened performance.

9.3.4 Results: Discriminative Classifiers

Discriminative classifiers, in the form of $mRVM_1$ and $mRVM_2$, were substituted into decision processes in place of the GMM. Risk-based active learning was employed to

develop these classification models for the Z24 Bridge dataset.

The decision accuracy and f_1 -score as functions of the number of queries made are provided for the risk-based active learning of mRVM_1 and mRVM_2 in Figure 9.32.

From Figure 9.32a, it can be seen that both mRVM_1 and mRVM_2 surpass the GMM in terms of decision-making performance. mRVM_1 begins with very low performance, once again indicating that the constructive approach to selecting relevance vectors was unable to converge, because of the limiting size of \mathcal{D}_l . Over the first few queries, however, the performance rapidly increases as \mathcal{D}_l expands such that convergence of the relevance-vector selection can be achieved. As with the previous case study, mRVM_2 achieves good performances on even the limited initial \mathcal{D}_l . Furthermore, improvements in decision accuracy are gained at a similar, if not a slightly greater, rate when compared to the GMM. This result means that superior decision-making performance is obtained throughout the querying process when adopting an mRVM_2 classifier. It is again worth noting that neither formulations of RVM suffer from a degradation in decision-making performance over later queries - indicating robustness to sampling bias in \mathcal{D}_l . Again, this result can be attributed to the discriminative nature of the models, in addition to the fact that extraneous data are excluded from \mathcal{A} .

From Figure 9.32b, it can be seen that the RVMs have inferior f_1 -scores in comparison to the GMM. Distinct improvements in classification performance arise in correlation with the appearance of new classes in \mathcal{D}_u (see Figure 9.35). The performances plateau at approximately 0.6 as data corresponding to Class 4 are not queried, resulting in the classifiers struggling to discriminate between Classes 3 and 4. Nonetheless, as these classes share an optimal policy with respect to the decision process, decision-making performance is not impacted by these misclassifications – as demonstrated in Figure 9.32a.

Figure 9.33 shows the class proportions in \mathcal{D}_l throughout the risk-based active learning of the mRVM classifiers. If one compares Figures 9.33a and 9.33b with Figure 9.26b, it becomes apparent that the class proportions following the active learning of mRVM_1 and mRVM_2 are not representative of the overall dataset. In fact, the class imbalance present in the full dataset is lessened in \mathcal{D}_l , with Classes 1 and 2 garnering approximately 40% representation each, and Class 3 possessing the remaining 20%. This deviation in class representation can be attributed to the RVM's sparsity – as only a few prototypical examples are required to sufficiently represent

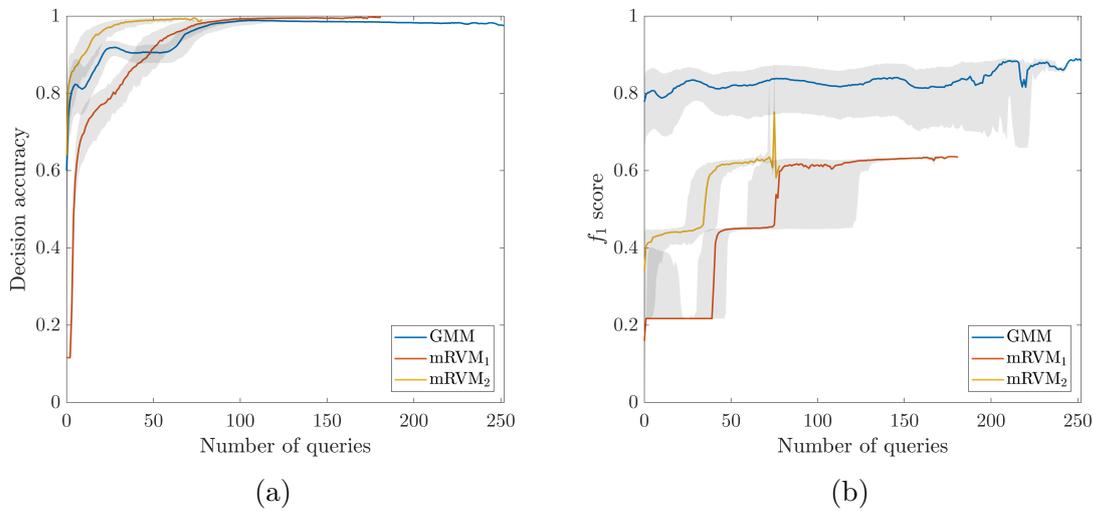


Figure 9.32: Variation in median: (a) decision accuracy and (b) f_1 score with number of label queries for an agent utilising: (i) a GMM (ii) an mRVM₁ and (iii) an mRVM₂ statistical classifier learned via risk-based active learning. Shaded area shows the interquartile range.

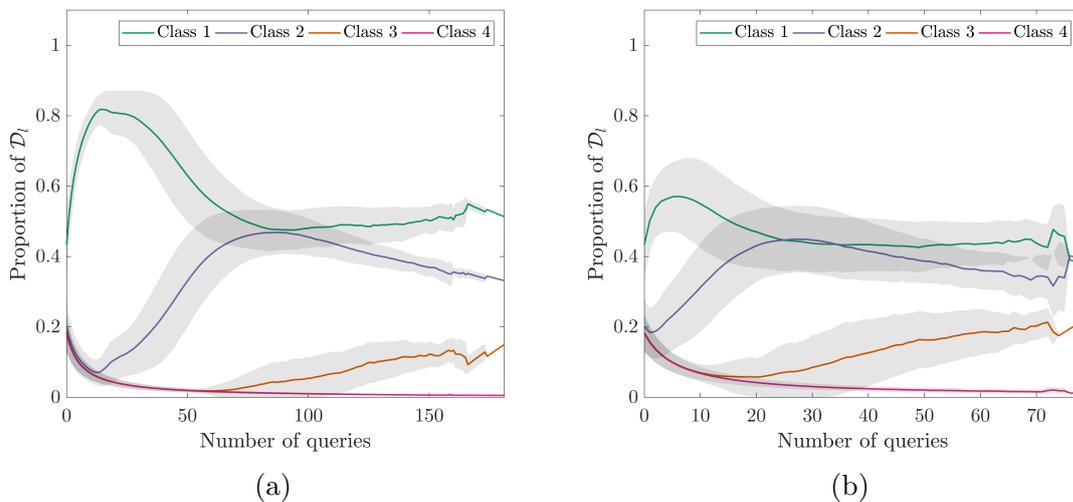


Figure 9.33: Variation in class proportions within \mathcal{D}_l with number of label queries for an agent utilising an: (a) mRVM₁ and (b) mRVM₂ statistical classifier learned from \mathcal{D}_l extended via risk-based active learning. Shaded area shows $\pm 1\sigma$.

each class; once a class is established following a few queries, data associated with the class need not be queried further. This result is further supported by the step-like improvements in f_1 -score.

The distributions for the number of queries made by each active learner are shown in Figure 9.34. Similar to the previous case study, mRVM₁ and mRVM₂ both query

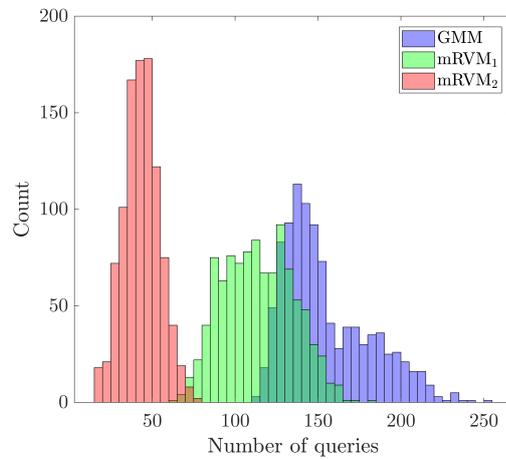


Figure 9.34: Histograms showing the distribution of the number of queries from 1000 runs of the risk-based active learning of: (i) a GMM (blue) (ii) an mRVM₁ (green) and (iii) an mRVM₂ (red) statistical classifier.

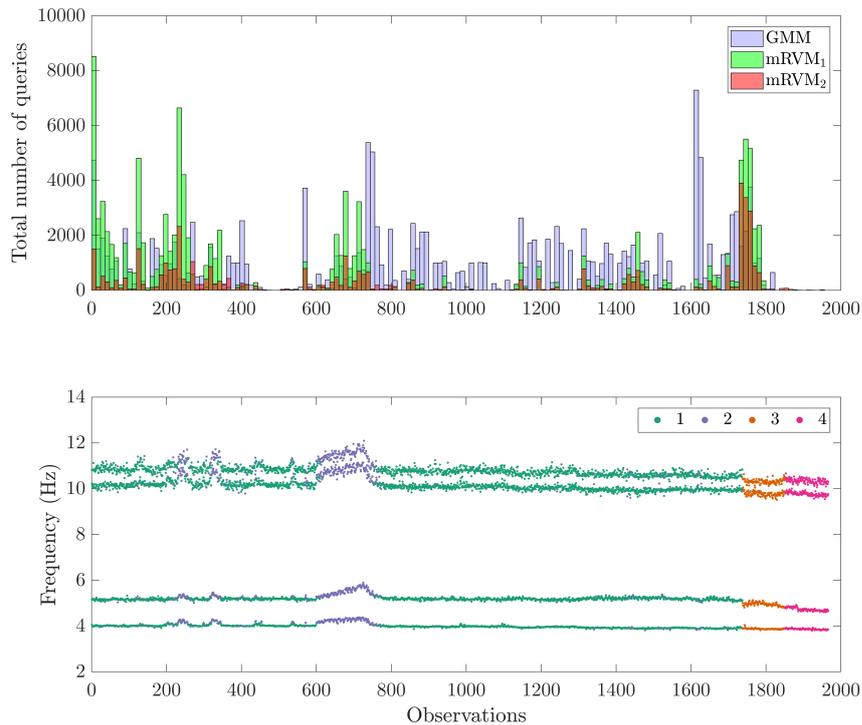


Figure 9.35: Histograms showing the distribution of the number of queries for each observation in \mathcal{D}_u from 1000 runs of risk-based active learning for: (i) a GMM (blue) (ii) an mRVM₁ (green) and (iii) an mRVM₂ (red) statistical classifier. A visualisation of \mathcal{D}_u is provided for reference.

fewer times on average when compared to the GMM. This result can also be explained by the RVM's ability to represent classes of data using only a few prototypical samples. Between mRVM_1 and mRVM_2 , mRVM_2 yields superior performance in terms of the number of queries made, with both lower mean and variance. As previously mentioned, fewer queries corresponds to a reduced expenditure on inspections, with lower variance indicating more consistent performance.

Finally, Figure 9.35 shows the total number of queries for each observation index in \mathcal{D}_u . As with Figure 9.34, it is apparent from Figure 9.35 that, overall, the use of RVMs as a statistical classifier reduces the total number of queries made. It can be seen that mRVM_1 results in increased querying early in the dataset; this is to be expected as it was previously observed that this formulation struggles to construct a model capable of discriminating between classes, resulting in high EVPI for all regions of the feature space. As with the GMM, both RVM approaches result in an increase in the number of queries made as new patterns in the data present – specifically around observation indices 200, 700, and 1800 in \mathcal{D}_u . Again, this result is in agreement with the trends in the f_1 -score shown in Figure 9.32b. The relative reduction in the number of queries made between observations 800 to 1600, for the RVMs compared to the GMM, can be attributed to the sparse representation achieved by the RVMs; as data for these classes had already been obtained, the EVPI associated with the classes was sufficiently low that inspection was not necessary, with the exception of outlying data.

In summary, substituting mRVM s in place of the GMM, results in improved decision-making performance and robustness to sampling bias. Notably, the number of queries made throughout the learning process is also decreased by the introduction of discriminative classifiers into the decision process. These benefits are obtained by virtue of the RVM's characteristic sparsity, in addition to the reduced reliance on assumptions regarding the distribution of data. Whilst mRVM_1 did yield noteworthy improvements in decision-making performance, mRVM_2 proved to be the stand-out model formulation.

9.4 Discussion

The results presented in Sections 9.2 and 9.3 hold several implications with respect to the development of decision-supporting SHM systems and thus bear further

discussion.

From the case studies presented, it is abundantly clear that variability in performance arises according to several factors. Specifically, variation was observed between the various types of classifier employed within a decision process, and the approach by which the models are learned.

In general, the discriminative classifiers consistently yielded better performance than the GMM (standard and semi-supervised), both in terms of decision accuracy, robustness to sampling bias, and the number of queries made. In contrast, the decision-making performance of the semi-supervised generative models were highly dependent on the assumed distributions used for density estimation and the underlying distribution of data. Based on these results, faced with the conundrum of selecting a classifier to be used in a decision-supporting SHM system, an engineer may opt to use a discriminative classifier without second thought. However, the selection may require more nuanced deliberations. For example, digital twins are a form of asset management technology highly applicable to structures and infrastructure [140, 158, 159]. In this context, SHM systems are a necessary component of digital twins. Furthermore, recent work has identified generative models as being a core component of digital twin technology [160]. It follows that, if the context in which SHM system is being applied necessitates a generative model, such as within a digital twin, a discriminative model can be deemed an unsuitable choice of classifier. In this scenario, one may have to accept the effects of sampling bias, or find a method of accounting for them within a generative model. If *a priori* knowledge of data distributions is available (such as via expert elicitation, or transfer learning [113]), then density estimates can be carefully selected such that EM, latent-state smoothing, or an alternative approach to semi-supervision can be incorporated into the risk-based active learning algorithm.

Some degree of performance variability was also observed between mRVM_1 and mRVM_2 . While fundamentally, these two classifiers share identical model-forms, they differ in the approach used to select relevance vectors. It was observed that mRVM_2 outperformed mRVM_1 . This result occurred as mRVM_2 has the advantage of beginning the training process by considering the relevance of all data in \mathcal{D}_l simultaneously in the initial set \mathcal{A} . In contrast, mRVM_1 is constrained to consider the contribution of data in the context of the current active set A , which is initialised as empty. Although mRVM_2 proved the better of the two formulations for case studies presented in the current chapter, for applications where \mathcal{D}_l grows very large,

mRVM₂ becomes computationally inefficient compared to mRVM₁, as discussed in [144]. It follows that, the quantity of data one expects to obtain throughout a decision-supporting SHM campaign must also factor into the selection of classification models learned via a risk-based active approach.

Another consideration that must be made in the selection of statistical classifiers used within SHM decision processes, risk-based active learning, and asset management technologies in general, is the uncertainty quantification with respect to outlying data, or out-of-distribution (OOD) data. As previously mentioned, the generative models presented in the current chapter make overconfident predictions for the class labels of outlying data – a result of data scarcity. This observation has concerning implications for the decision-support systems, as an agent utilising a model that overconfidently classifies OOD data will be especially prone to making suboptimal decisions. While these decisions could yield benign consequences, such as an unwarranted down-time, they could also lead to missed inspections and ultimately structural failures with potentially severe consequences. Fortunately, the discriminative classifiers considered for the current chapter proved to have excellent uncertainty representation for outlying data, yielding high EVPI and therefore reliably triggering inspections where necessary.

Finally, a notable observation is that the number of queries made during the risk-based active learning can be reduced via the choice of statistical classifier within a decision process. This result is highly significant in the context of SHM decision-support because *queries correspond to structural inspections*. This correspondence indicates quite directly, that monetary savings can be achieved by selecting a classifier that results in minimal queries, thereby reducing the total operational costs incurred over the lifetime of structure. The aforementioned result highlights the great value to be gained via SHM systems, and provides further motivation for their development and implementation.

9.5 Summary

Risk-based active learning provides a methodology for the online development of statistical classifiers, specifically those being used in decision-supporting systems. In the context of SHM, this process is accomplished by querying labels for observed data via structural inspection, according to the expected value of the label information with

respect to an O&M decision process. Although risk-based active learning has been demonstrated to yield improvements to decision-making performance, the generative classification models considered prior to the current chapter suffer ill-effects related to the sampling bias introduced by the guided querying. This bias typically manifests as deterioration in decision-making performance later in the querying process.

The current chapter proposed two approaches to address the issue of sampling bias; semi-supervised learning and discriminative classification models. Each approach was applied to a visual example, and an experimental case study. From these case studies, it was found that semi-supervised learning yielded variable performance, dependent on the correspondence between the underlying data distributions and the density estimates selected for the generative models. On the other hand, the discriminative models were found to mitigate the deterioration in decision-making performance. In addition, the case studies provided results with high significance to SHM decision-support in general. Specifically, attention was drawn to the fact that utilising differing classifiers in a decision process model can result in the decision-making agent possessing drastically different attitudes towards outlying data. Finally, it was also shown that the choice of classifier used in the process model can greatly affect the number of queries – and therefore inspections – made throughout an SHM campaign. This finding was particularly significant as it implies that, with the careful selection of the classifier within an SHM system, resource expenditure can be greatly reduced.

TOWARDS RISK-BASED DECISION-MAKING FOR POPULATIONS OF STRUCTURES

As previously discussed, SHM technologies aim to detect damage within mechanical, civil and aerospace structures and infrastructure [1]. By inferring information about the health of a structure from discriminative features extracted from data acquired throughout a monitoring campaign, these systems can facilitate informed predictions relating to one or more of the problems regarding the health of a structure, as summarised in Rytter’s hierarchy [2]; namely, detection, localisation, classification, quantification, and prognosis. By informing predictions with data from a monitoring system, one can also inform decision-making regarding the operation and maintenance of structures, and this can yield benefits such as improved safety, reduced operation costs and operational-lifetime extension.

The approach to decision-making for SHM presented in previous chapters adopts a probabilistic risk-based perspective. Several submodels have been identified as elements that are required to sufficiently define SHM decision processes; these submodels include statistical classifiers for inferring health-states and health-state transition models for forecasting damage. In order to achieve robust decision-making, these submodels require labelled data for learning and/or validation.

As has been highlighted throughout the current thesis, a critical challenge associated

with the development of SHM systems is the scarcity of the data necessary for the learning and validation of models. Population-based structural health monitoring (PBSHM) provides a holistic framework for alleviating data scarcity in the development of predictive models for SHM [111–113, 161]. The core principal of PBSHM is that predictions about individual structures can be improved with the use of information transferred from other similar structures.

The current chapter aims to examine PBSHM in the context of the decision-making processes necessary for the operation of both individual structures and populations of structures, and introduce some key concepts and definitions necessary for furthering risk-based decision-making for populations of structures. This aim is achieved by extending the hierarchical representation of structures, introduced in Chapter 5, to hierarchical representations of *populations of structures*. Throughout the current chapter, an inventory of offshore wind farms are referenced as a motivating example.

10.1 Population-based SHM

The foundations of PBSHM have been presented in a series of journal papers, each detailing the fundamental concepts of the approach; homogeneous populations [111], heterogeneous populations [112], mapping and transfer [113], and the geometric spaces in which structures exist [161]. By adopting a population-based approach to SHM, such that knowledge and information can be transferred between similar structures, there is the potential for improved diagnostic and prognostic capabilities [162].

In the most general sense, a population can be considered to simply be a set of structures. Given the broad nature of this definition, in order to achieve useful transfer of knowledge and information between structures, it is discerning to consider specific classes of populations based upon the similarity of the constitutive structures. Thus, the notions of homogeneous and heterogeneous populations are introduced in [111–113].

10.1.1 Homogeneous and Heterogeneous Populations

Within a population, structures may share common characteristics such as geometries, topologies, materials, and boundary conditions. Consider a population of wind turbines in an offshore wind farm and suppose these turbines are of the same model; developed to the same ISO standards and possessing common components, materials, aerodynamic design and so on. Qualitatively, these structures can be regarded as nominally identical. Populations comprised exclusively of nominally-identical structures are termed *homogeneous populations*. Specific instances of structures in a homogeneous population can be considered to be perturbations of a population *form* [111]. For further discussions on population forms, the reader is directed to [111]. Other examples of homogeneous populations include a fleet of Airbus A380s, an array of small modular nuclear reactors, and the Global Positioning System (GPS) satellite constellation.

Variation between structures in homogeneous populations may arise because of factors such as environmental conditions and manufacturing defects. Returning to the example of an offshore wind farm, one could imagine that two turbines at differing locations in the farm may experience different geotechnical conditions – perhaps as a result of varying geological composition in the seabed. Variability in such conditions could affect the boundary conditions of the monopile turbine towers and therefore modify the behaviours and data exhibited by these otherwise nominally-identical structures.

In essence, *heterogeneous* populations form the complement of the set of homogeneous populations [162]; that is, heterogeneous populations are not exclusively comprised of structures that are nominally identical. Heterogeneous populations represent more general sets of structures and allow for differing designs, large variability in boundary conditions, and even multiple types of structure. While there may be stark differences between individual structures in a heterogeneous population, there may nonetheless be similarities that can be exploited to achieve useful knowledge and information transfer.

Consider again the offshore wind farm example and suppose that the population is comprised of wind turbines each with three blades. Suppose also that the operating company manage an additional wind farm in a distinct location, comprised of four-blade turbines. Useful inferences could be achieved by considering these wind farms

as two homogenous populations, however, further insights could also be gained by considering them as a single heterogeneous population. For example, similarities may be present in the tower design between both types of wind turbine; hence, by considering a larger population from which to make observations, improved predictive models can be developed for this specific substructure. Other types of heterogeneous populations that may be useful to consider include inventories of aircraft comprised of a variety of models, and multiple suspension bridges with differing designs (e.g. single-span, multi-span).

Thus far, similarities between structures have been described somewhat qualitatively, however, to better indicate where information transfer may work, it is useful to quantify this similarity.

10.1.2 Similarity Between Structures

Graph theory provides a rigorous and rich framework for representing and comparing discrete structured objects and has proved to be an invaluable modelling tool in fields such as chemistry and proteomics.

In [112], the notion of the irreducible element (IE) model for structures is introduced as a representation of structures with relatively low-dimension when compared to alternatives such as finite element, or CAD models. The IE representation involves abstracting a structure into discrete elements having geometrical and material information (e.g. beams, plates, shells), and relationships (e.g. joints), so as to sufficiently capture the nature of a structure. Here, the ‘nature’ one wishes to capture pertains to health monitoring problems associated with a structure.

Once an IE representation of a structure has been obtained, the information can be encoded into an attributed graph (AG). Whereas the purpose of the IE model is to present key characteristics of a structure in a human-readable format, the purpose of the AG is to embed a structure into a metric space, so as to facilitate the efficient pair-wise comparison of structures. With structures embedded into such a space via AGs, one can utilise graph-matching algorithms to find common subgraphs between sets of structures. These subgraphs indicate substructures that are common within sets of structures and can be used to inform where transfer may be applicable. Furthermore, measures of closeness within the space of AGs (or common subgraphs) can be used to quantify similarity; in [112] the Jaccard index is used and in [163] a

variety of graph kernels are demonstrated.

In summary, structures can be mapped into a graphical domain to facilitate comparison, identify common substructures and quantify similarity. By conducting this similarity assessment for structures within a population, one can determine where it is likely that information and knowledge can be successfully transferred between individual structures.

10.1.3 Mapping and Transfer

As mentioned previously, the primary benefit in taking a population-based approach to SHM is in gaining the ability to transfer knowledge and information between sufficiently-similar individual structures; thereby overcoming issues associated with data scarcity.

The sharing of knowledge and information between individual structures can be achieved via a number of methodologies. One manner in which this can be achieved is by having a statistical representation of the aforementioned population form, as demonstrated in [111]. Another approach, presented in [164], shares datasets in joint hierarchical statistical models of a population. Methodologies founded upon *transfer learning* have also been successfully demonstrated [113]. The principal of transfer learning is closely aligned with the goals of PBSHM; specifically, a branch of transfer learning known as *domain adaptation*. In domain adaptation, datasets are adapted in a manner that allows a model constructed for a *source* domain to generalise to a *target* domain.

For knowledge/information transfer to be successful, it is imperative that these source and target domains are comparable. This constraint can be adhered to by employing the similarity assessment outlined in the previous section.

Thus far, PBSHM has been considered with respect to predictions and inferences. Before incorporating decisions into the PBSHM framework, background on the risk-based approach to decision-making for traditional SHM is provided.

10.2 Structures as Hierarchies

A key assumption implicit in the development of the fault-tree failure models within the risk-based SHM decision framework, is that structures can be represented as a hierarchy, or, in other terms, as a system of systems of systems. As discussed in Chapter 5, to obtain a hierarchical representation for S , one must first decompose S into a discrete number of constituent elements, which are referred to as *substructures*. A simplified diagram illustrating the hierarchical representation of a structure is shown in Figure 10.1. The levels in the hierarchy that specifies the system of systems of systems shown are denoted as \mathcal{S}^1 , \mathcal{S}^2 , and \mathcal{S}^3 – corresponding to the component, substructure and substructure levels, respectively. Within each level of the hierarchy, elements can be listed.

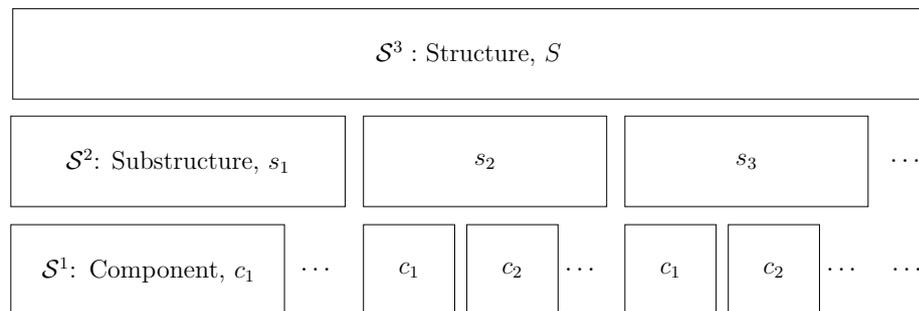


Figure 10.1: A structure as systems of systems.

Returning to the example of a wind farm, it would be perfectly reasonable to consider a single turbine as an individual structure, representing the \mathcal{S}^3 level in a hierarchy. In the \mathcal{S}^2 level one may consider substructures such as the drive train, blades, or tower. Finally, in the \mathcal{S}^1 level one may have components such as the gearbox or bearings comprising the drive train, or the web and shells comprising the blades.

The hierarchical representation of structures facilitates the specification of the decision processes that motivate the development and implementation of SHM technologies. This facilitation is achieved by decomposing structures into constituent substructures and components which can then be used to define failure modes of the structure. Given a finite set of failure modes of interest, one can then specify critical components, and therefore health states, to be targeted by a monitoring system.

10.3 Populations of Structures as Hierarchies

A natural method for incorporating decision-making into PBSHM, is to extend the hierarchical representation of structures to hierarchical representations of populations. The number of levels required in a hierarchy is of course dependent on context. However, it is deemed that an additional three levels provide sufficient generality for most PBSHM applications, and indeed the discussions in the current chapter.

The additional levels necessary to extend the hierarchical representation to populations of structures can be summarised as follows:

- \mathcal{S}^4 – Type/Model Inventory: This level of the hierarchy corresponds to the lowest population level and represents an organisational grouping in which all individual structures in the population are of the same type/model and can be considered to be nominally identical. Thus, populations at this level in the hierarchy are homogeneous.
- \mathcal{S}^5 – Group Inventory: This next population level corresponds to a set of \mathcal{S}^4 inventories for which it is necessary or convenient to consider it as a group for operational reasons such as asset management. As a group inventory may be formed of disparate type/model inventories, in general, group inventories are heterogeneous populations.
- \mathcal{S}^6 – Global Inventory: This level of the hierarchy corresponds to the total set of structural assets operated or owned by an organisation or company. Again, this level will generally represent a heterogeneous population.

Figure 10.2 depicts the continuation of the hierarchical representation from \mathcal{S}^3 to \mathcal{S}^6 . In Figure 10.2, an inventory I is considered as a system of systems of systems of systems. Once again, a list can be formed of the constituent elements for each level in the hierarchy.

To further elucidate this extension of the hierarchy, once again, consider the example of an organisation operating offshore wind farms. As previously indicated, a wind farm comprised exclusively of turbines of a single type or model can form a homogeneous population; this corresponds to \mathcal{S}^4 in the hierarchy. In the case that the organisation is responsible for multiple wind farms, or a single farm with a mixture of turbine types,

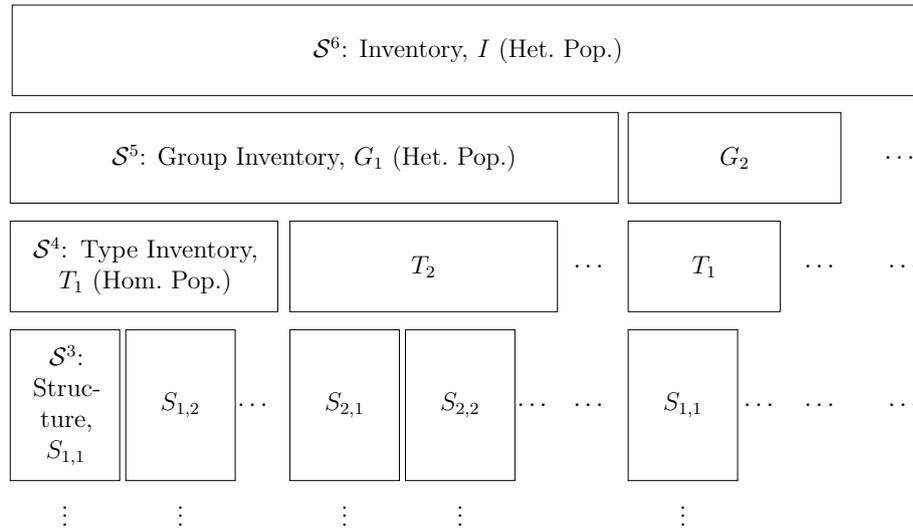


Figure 10.2: An inventory as a system of systems of systems of systems.

one may wish to organise these type/model inventories into group inventories. For example, these group inventories may be formed from type inventories according to the geographical jurisdiction of sub-divisions within the organisation, or even formed from a collection of type inventories that are overseen by a single maintenance crew. Should these populations each be comprised of a different model of wind turbine, the group inventories formed would be heterogeneous populations and correspond to \mathcal{S}^5 in the hierarchy. Alternatively, if all the wind farms consist of a single type of turbine, \mathcal{S}^4 and \mathcal{S}^5 can be merged and the group inventories are instead homogeneous populations. Finally, the group inventories owned by the wind farm organisation can be aggregated as an inventory in the \mathcal{S}^6 level of the hierarchy. This level would represent the organisation’s total structural assets and could amount to, for example, multiple wind farms spread across the globe, maritime vessels, and aircraft that may be used for inspection, maintenance or other operational activities.

As is the case for traditional SHM, the hierarchical representation of structures and populations of structures can help facilitate decision-making for PBSHM in several ways. These decision processes are discussed further in the following section.

10.4 Risk-informed PBSHM

Numerous decisions must be made throughout the life cycle of a PBSHM system. Most obvious are the operation and maintenance decisions an organisation may have

to make, following the installation of a monitoring system, such as inspections and repairs. Equally important, however, are the decisions that must be made prior to implementation, such as those made in the operational evaluation stage of PBSHM.

10.4.1 Operational Evaluation

One significant way in which adopting a hierarchical risk-based approach to PBSHM facilitates decision-making occurs very early on, in the operational evaluation stage. By considering specific failure modes and constructing fault trees for individual structures, one can decide the key elements of a structure which should be modelled in IEs and AGs. In other words, the specification of failure modes as combinations of component and substructure failures can be used to inform the granularity at which IEs and AGs are constructed. A further benefit of the population-based approach is that, as structures are considered nominally identical, large proportions of the fault trees may be mapped across a homogeneous population, with the exception of perhaps environment-specific failure modes.

The extension of the hierarchy to represent populations of structures via the inclusion of levels \mathcal{S}^4 to \mathcal{S}^6 prompts one to consider how failures may be defined at the population level. One possible way to approach the failure of a population would be to consider the critical missions for the operating organisation. Depending on the nature of the organisation – whether they are non-commercial or commercial – these missions may be related to performance measures such as availability and/or profitability. Consider the wind farm example. Suppose that the operating organisation are required to supply energy from the wind farm to an electrical grid while maintaining a total population availability of 99%. This population can then be considered to have failed if the population structural availability falls below 99%. This population failure may be specified then by extending the fault tree; defining the population failures as a combination of individual failures. In addition, the organisation may wish to specify a failure condition based upon profitability, perhaps based upon a performance criterion related to a moving-average of the total power output. Again, this failure could be represented as a combination of individual structure failures and environmental conditions. This distinct failure mode is likely to be highly correlated with the availability failure mode; fortunately, the probabilistic graphical models employed in the risk-based approach can account for these ‘common-cause’ failures. This approach to defining population failures can be applied at any of the population

levels within the hierarchical representation by considering combinations of failures in the levels below.

Defining failures at the population level within the hierarchy allows one to assign costs during the operational evaluation stage. Following on from this, population-scale actions can be also be defined.

10.4.2 Inferences and Decisions

A fundamental process of decision-making for PBSHM is reasoning under uncertainty. This is typically achieved via inferences. Within the hierarchical framework for PBSHM, different types of inferences can be defined:

- I-inference: This type of inference corresponds to those usually made in traditional SHM, and occur within the individual structure levels \mathcal{S}^3 to \mathcal{S}^1 . An example of an I-inference is the process of determining a probability distribution over the health states of an individual structure using data acquired from that structure.
- L-inference: This type of inference occurs between levels in the hierarchical representation of structures. These may also be types of I-inference, for example determining the probability of failure for a (sub)structure given local component health states. Other L-inferences may include those relating to the validation and verification of predictive models (V&V). For example, one may be able to validate a predictive model for a structure at the \mathcal{S}^3 level with data measured from substructures or components at the \mathcal{S}^2 and \mathcal{S}^1 levels, respectively.
- P-inference: This type of inference occurs across populations. If the inference is across a type inventory in \mathcal{S}^4 , i.e. a homogeneous population, they can be denoted as HomP-inferences. These inferences across populations may utilise technologies such as forms [111]. An example of a HomP-inference is inferring the health state of a member in a population using data aggregated across all members in the population. On the other hand, if a P-inference is between populations containing different types of structure, such as within a group inventory in \mathcal{S}^5 , then the inferences can be referred to as HetP-inferences. HetP-inferences may involve using transfer learning techniques such as domain adaptation [113]. An example of a HetP-inference is transferring

the degradation (transition) model for a blade from a population of four-blade wind turbines to a population of three-blade wind turbines.

These inferences within the hierarchical representation of populations, facilitate reasoning under uncertainty using PBSHM systems; this can naturally be extended to decision-making under uncertainty, by considering the following types of decision:

- I-decision: This type of decision is made at the individual structure levels in the hierarchy, \mathcal{S}^1 to \mathcal{S}^3 . Again, this type of decision corresponds to decisions one may make with a traditional SHM system. An example of an I-decision is selecting a maintenance strategy for an individual structure, substructure, or component for repair. Unlike in traditional SHM, in the risk-informed PBSHM approach, I-decisions can be informed by I-, L- and P-inferences alike.
- L-decision: The actions selected via this type of decision operate between levels of the hierarchical representation. As with L-inferences, these decisions may pertain to the V&V of predictive models. For example, deciding whether one can proceed with using a structural model validated on substructures. Another example of this type of decision relates to resource allocation. Suppose one has a limited budget to carry out some structural testing to acquire data for mode updating. Under these circumstances, one should aim to decide on a set of tests, and the levels at which these tests are carried out, such that the largest improvement in model performance is obtained for the given budget.
- P-decision: This type of decision is made at the population levels in the hierarchy, \mathcal{S}^4 to \mathcal{S}^6 . These actions may pertain to resource management. For example, one may decide to send a team of engineers to perform inspections on a type inventory based on the probability of failure for a population rather than the probability of failure of an individual structure. Scheduling inspections in this manner could save both time and expenditure. Again, these decisions may be informed via I-, L- and P-inferences.

To summarise, the hierarchical representation of populations of structures facilitates both making inferences and making decisions for PBSHM, by allowing for the definition of specific types of inferences and decisions.

10.4.3 Value of Information Transfer

Introduced in Chapter 8, value of information is a concept in decision theory defined as the amount of money/resource a decision-maker should be willing to pay to gain access to information prior to making a decision. VoI has seen some application to SHM in recent works [142, 165, 166]. Extending the risk-based approach to decision-making from traditional SHM to PBSHM opens up the possibility of value of information transfer, i.e. the price a decision-maker should be willing to pay in order to gain information via transfer, prior to making a decision. This value arises as a result of change in maximum expected utility that can be achieved should a change in optimal policy occur as a result of the additional information made available via transfer. This notion of value of information-transfer yields the thought-provoking implication that, in some contexts, it may be an optimal decision to allow a (sub)structure to fail, since the data obtained throughout the failure process may improve the management of the other individuals in a population.

10.5 Conclusions

To conclude, PBSHM provides a general framework for overcoming issues of data scarcity associated with developing predictive models for detecting and forecasting damage within structures. This advantage is achieved via technologies that allow for the transfer of information between individual structures within a population. Adopting a probabilistic risk-based approach to SHM, allows inferences made about the health-states of individual structures to inform operation and maintenance decisions via the use of hierarchical representations of structures and fault trees. The current chapter extends this hierarchical representation of structures to representations of populations, such that decision processes can be defined over populations. Other advantages can be gained by adopting a risk-based approach to PBSHM; for example, the identification of critical components and substructures can be used to inform the development of irreducible element models and the associated attributed graphs.

Risk-based PBSHM provides an exciting avenue for future research, promising improved decision-support capabilities beyond asset management for individual structures. The final chapter of this thesis suggests further topics for future work, and summarises the main contributions, conclusions, and limitations of the work.

CONCLUSIONS AND FUTURE WORK

Enhanced decision support for asset management provides one of the primary motivations for the development and implementation of SHM technologies; however, actual uptake of SHM has been limited by the (lack of) availability of cost-benefit analyses. Framing SHM systems in the context of decision-support allows the value of the technology to be quantified, thereby giving impetus to the broader adoption of SHM system into industry. The work presented in the current thesis focusses on topics related to risk-based approaches to decision-making in the context of SHM, in which operation and maintenance strategies are selected so as to maximise *expected utility*.

To recap, the primary technological contributions of this thesis were as follows. Firstly, a risk-based paradigm for conducting SHM campaigns was proposed. By adopting hierarchical representations of structures, and augmenting the traditional SHM paradigm with elements of probabilistic risk analysis, the framework developed facilitates constructing SHM decision processes as probabilistic graphical models. These models are capable of representing the inherent uncertainty associated with structural damage detection and forecasting. Furthermore, the failure-mode models borrowed from probabilistic risk assessment aid in defining the scope of an SHM campaign and facilitate the specification of costs and utilities – a vital component required for risk-based decision-making.

Next, risk-based active learning was developed as an approach for developing statistical classification models with explicit consideration for decision-support applications.

Within the risk-based active-learning algorithm, data are queried according to their expected value of information with respect to an O&M decision process. By comparing the value of information for incipient data to the cost of inspection, the algorithm provides a natural way for mandating structural inspections. Moreover, the online algorithm can be applied when comprehensive labelled datasets are unavailable – a primary challenge with learning statistical classification models in SHM.

Finally, modified risk-based active-learning algorithms were designed so as to improve decision-making performance. In particular, discriminative classifiers and semi-supervised learning techniques were used to mitigate the detrimental effects of sampling bias. Significantly, this research demonstrated that the choice of classification model used within a decision-support system can drastically impact the decision-making performance of an agent.

In addition to several technological contributions, discursive contributions were provided on the topic of transition models, and the promising prospect of risk-based O&M of populations of structures. The following subsection will provide a summary of the work presented in the current thesis and highlight the significant conclusions with respect to risk-based decision-making for SHM.

11.1 Summary and Conclusions

Chapter 1 introduced SHM systems as a decision-support tool and outlined the main approaches for conducting SHM; namely, data-based, physics-based, and hybrid. The key challenges that arise within SHM were also introduced in the context of decision-making – most notably, the scarcity of comprehensive datasets representative of all pertinent damage conditions, prior to the implementation of monitoring systems.

Chapter 2 provided a literature review surrounding risk-based approaches to decision-making in the context of SHM. Chapters 3 and 4 provided background information of probabilistic risk assessment and SHM, and probability theory and decision theory, respectively.

In order to make reasoned O&M decisions for structures of interest, an agent requires some belief regarding the risks (or expected utilities) associated with structural failures, and how differing strategies influence those risks. The keystone of the current thesis was established in Chapter 5 – a risk-based paradigm for conducting

SHM, using salient principals and technologies from probabilistic risk analysis to augment the established paradigm for SHM. The generalised risk-based framework, based upon probabilistic graphical models, sought to provide a means for relating monitoring data to the crucial information required for assessing risks and making informed decisions regarding the operation and maintenance of physical assets. The decision framework was demonstrated via an experimental case study in Chapter 6. The case study demonstrated that SHM information can be used to improve O&M decision-making – even, in certain scenarios, with imperfect classification models.

Chapter 7 focussed on transition models – a fundamental component of Markov decision processes. Transition models are used in the context of SHM to forecast predictions for future structural health states, such that preventative actions can be taken. In Chapter 7, a Monte Carlo sampling was used to generate a degradation model using prior knowledge of physical failure mechanisms and operational conditions. While being a principled approach that overcomes the challenge associated with lack of observed state transitions prior to the implementation of monitoring systems, the purely physics-based methodology lacks validation, because of the absence of data. Validating transition models, particularly those used in maintenance decision processes remains a significant challenge in SHM and is discussed further in Section 11.2.

Statistical classifiers were the topic of Chapter 8. Within partially-observable decision processes, classification models are used to provide probability distributions over discrete latent states. For SHM, classifiers are used to predict current structural health states via features extracted from monitoring data. In Chapter 8, risk-based active learning was proposed as an online approach to developing statistical classifiers with explicit consideration for decision-support applications. It was shown that learning a GMM, using data labelled via structural inspections guided by expected value of information, increased the rate at which decision-making performance improves throughout a monitoring campaign. This result is significant in the context of decision-making for SHM, as it demonstrates that O&M costs can be reduced via risk-informed machine learning.

Statistical classifiers were also the subject of Chapter 9. The performance of active-learning algorithms, such as that presented in Chapter 8, can degrade over time because of the preferential nature of the querying process – a phenomenon known as sampling bias. Chapter 9 presented an investigation into approaches for mitigating the effects of sampling bias. Two approaches for mitigating sampling bias were

investigated; semi-supervised learning, and discriminative classifiers. Both approaches found some degree of success in mitigating the detrimental effects of sampling bias. However, on the case studies presented, the multiclass relevance vector machine - the chosen form of discriminative model - was shown to be dominantly superior in terms of maintaining strong decision-making performance. In addition to improvements in ‘decision accuracy’ over the course of a monitoring campaign, it was found that the design of classification algorithms have a significant impact on the number of queries made during the active-learning process. In particular, this result is notable in the context of SHM decision-support as it indicates that the number of inspections made throughout a monitoring campaign is dependent on the classification models used. It follows that the overall O&M costs over the lifetime of a structure can be reduced by classifier design.

Chapter 10 presented ideas on the topic of decision-making for populations of structures. The hierarchical representation of structures used to specify structural failure modes in Chapter 5 was extended to accommodate levels of homogeneous and heterogeneous populations. With reference to the hierarchical representations of populations, aspects of risk-based decision-making that could be improved via PBSHM were highlighted. Particularly exciting, is the prospect of gaining additional data for learning and validating models within decision processes via knowledge transfer technologies such as transfer learning.

11.2 Limitations and Challenges

There are several limitations of the works presented in the current thesis. In addition, there remain several unaddressed challenges associated with decision-making for SHM that must be overcome in order to achieve robust cost-optimal O&M of structural assets.

One limitation associated with the probabilistic risk-based decision framework detailed in Chapter 5, arises as a result of specifying the scope of an SHM system. In predetermining which failure modes, health states, operational conditions, and environmental conditions are targeted by a monitoring system, and which interventions are considered in the decision process, one risks misinterpreting data generated from conditions outside the scope of the system as data generated by conditions from within the scope. This effect could potentially result in severe ramifications, such as

catastrophic structural failures, should out-of-distribution (OOD) data originating from malignant states be misclassified as benign. As discussed in Chapter 9, using models that possess good uncertainty quantification for OOD data can provide some degree of protection from such misclassifications. Nonetheless, where there are unknown or unaccounted-for confounding influences, there remains a heightened risk of misclassification. In addition, misclassifications may arise in the form of examples produced via an adversarial attack. Here, an important consideration is that adversarial examples can become increasingly common for high-dimensional feature spaces and label spaces, which may arise in SHM when trying to discriminate between numerous combinations of health, environmental, and operational conditions [167].

Two limitations of the work presented in the current thesis relate to the temporal nature of SHM decision-making. Work presented in the current thesis largely considered decision processes comprised of only a few time-slices. By considering only small definite-horizon problems, the decision-making approach is akin to a greedy algorithm – for each time-slice, locally-optimal actions are selected. This approach does not guarantee globally-optimal strategies over the operational lifetime of a structure, and only approximates globally-optimal strategies. Where there are discrepancies between the globally-optimal strategy and the greedy approximation, there will be unrealised potential gains, or avoidable losses, associated with the asset management. While the approaches presented previously are general and can be extended to larger horizons, doing so will incur greater computational costs and so a trade-off must be made between the time-horizon used to model decision processes and the effort expended in solving them. Another limitation that must be considered when applying the risk-based framework to online decision-making relates to the fixed time-step size used to model the decision process. Three factors play a role in the selection of the time-step used; the time-scales at which the degradation mechanisms of the structure operate, the time required to process data and solve the influence diagram, and the time required to execute any necessary actions. Ideally, the time-step selected would allow sufficient time for both the computation of the optimal strategy and the implementation of any strategy selected. However, the time-scales at which these processes may be achieved do not necessarily align with the time-scales at which structural degradation occurs – some degradation mechanisms occur over very short time-scales (e.g. supercritical crack growth). In its current fixed time-step form, an operator using the risk-based framework for online decision-support would have to be reconciled with the impacts associated with any potential

discrepancies between these time-scales and the time-step used to model the decision process; either by accepting that some rapid failures may go undetected, or that some detected failure sequences may be unpreventable.

Another limitation with the risk-based decision framework and risk-based approach to active learning presented is that, in their current formulations, there is no consideration of financial budgets and resource availability. Currently, the approaches select maintenance strategies that yield locally-maximal expected utility and inspections with positive net expected value. These decisions are made irrespective of whether funds are available to meet the up-front costs of undertaking the decided actions. This property means that potential expenditure on inspections and maintenance over the course of a monitoring campaign is bounded only by the sum of the maximum possible expenditure in each time-slice over all time-slices in a monitoring campaign. This is an idealistic upper bound as, in practice, operators do not have unlimited resources available for O&M, and in many applications the budgetary constraint will supersede the maximal upper bound.

In addition to specific limitations with the approaches to decision-making presented in the current thesis, there remain more general challenges that must still be overcome in order to achieve broadly-applicable risk-based decision-support technologies that incorporate information from SHM systems.

Arguably, the most significant challenges lie in the validation of transition models. In order to develop trustworthy transition models, data that correspond to observed state transitions are required; ideally, to comprehensively validate a transition model, each possible transition would be observed multiple times. In the context of SHM there are two compounding factors that limit the availability of such data; the size of the transition space and the cost of inspections. For an SHM system targeting n global health states of interest, c operational and environmental conditions, and a potential actions, the transition space is, in general, $\mathcal{O}(n^2ca)$, where n also grows exponentially with the number of local health states considered. Over the course of an operational lifetime, any single structure may only go through a small fraction of the possible transitions, thereby limiting the subregions of the transition model that can be validated online. Furthermore, given the magnitude of the transition space and cost of inspections, obtaining sufficient data to achieve comprehensive validation data would be prohibitively expensive in many applications.

A further challenge, minimally addressed by the work presented in the current

thesis, pertains to the specification of the utility/cost functions used within the decision-process models. Although considering a limited number of specific structural failure modes via fault trees can facilitate the identification of adverse events, for SHM applications there is no formal methodology for quantifying the severity of the consequences in terms of cost/utility. The issue of cost quantification poses a significant (and potentially contentious) challenge – strategies are generally sensitive to transformations in utility functions, and performance criteria such as safety and profitability are sensitive to O&M strategies. In many scenarios, critical performance criteria such as these will have an inverse relationships; and so effort must be made to specify utility functions that reflect the preferences of users, operators, and owners. For certain safety-critical applications, determining such utility functions may require grappling with uncomfortable ethical questions such as ‘what is the cost of a human life?’. In addition to ethical considerations, there are technical and logistical challenges with the elicitation of utility functions. In much the same way that there is no standardised method for eliciting informative priors for Bayesian models in engineering applications, there is no standard for eliciting utility functions.

11.3 Future Work

Considering the work presented in the current thesis, and the associated limitations and challenges, new directions for research to be conducted in the future can be identified. Here, a few directions for future research are highlighted.

- The risk-based approach to structural health monitoring presented in Chapters 5 and 6 recommends, and facilitates, new approaches to many of the tasks that must be completed in order to fully specify an SHM system. In particular, the framework opens up the possibility for decision-theoretic, or risk-based, approaches to data acquisition, data processing, and feature selection. Traditionally, these tasks are accomplished using information-theoretic approaches; however, by considering the decision-support application of an SHM system, one can instead optimise these processes with respect to expected utility and value of information. For data acquisition, one potentially-viable approach would be to formulate a Bayesian experimental design problem for sensor types and placements; however, rather than optimising with respect to an information measure (e.g. Shannon information), one optimises with respect to value of

information. For data processing and feature selection, one route to follow would be to construct a risk-informed dimensionality reduction for classifier inputs, such that separability for data around decision-boundaries is maximised. Accomplishing this would facilitate a cost-effective trade-off between the computation times for predictions and decision-making performance.

- Considering longer or indefinite time-horizons is one potential avenue for further work that may provide improved O&M of structures. Dynamic programming methods and Bellman's equations [168] provide a suitable approach and can be naturally applied to the O&M Markov decision process presented in the current thesis. A key aspect of Bellman's equations is that future utilities are reduced by a multiplier $0 < \gamma < 1$ such that current rewards are prioritised¹. Importantly, this discounting of future utilities guarantees, under certain other conditions, stationary strategies for infinite-horizon decision problems meaning that by determining the current state of a structure, one can know the optimal course of action at any time. A challenging extension to this problem, for which a more involved future would be required, arises when considering non-stationary utilities and costs. Such utility functions come into play as a result of fluctuations in the value of resources; for example, the cost of labour, or the price of certain materials.
- Another avenue of research worthy of further research pertains to the allocation of finite resources; in the context of SHM, this resource may be monetary or otherwise, e.g. labour. Giving decision-makers budgetary constraints will more accurately reflect the manner in which most organisations operate and manage structural assets in practice. Game theory and multi-agent reinforcement learning provide suitable machinery for enacting such constraints. Multi-agent reinforcement learning seeks to train multiple agents operating in a shared environment. In the context of SHM, this shared environment can comprise of both the states of a structure and the total budget allocated for the O&M of that structure [28]. One novel approach may be to assign each agent with the task of mitigating a specific failure mode of the structure, each with differing risks. Within the shared environment, optimal strategies will be found such that the overall risk of failure all while the budgetary constraints are adhered to.
- Arguably, the most broad area worthy of future research is surrounding risk-

¹A bird in the hand is worth $\frac{1}{\gamma}$ in the bush.

based SHM for populations/fleets, as discussed in Chapter 10. One avenue for future research in this area would be to extend the previous bulletpoint to exact budgetary constraints at population level; again, multi-agent reinforcement learning may be of use in achieving this. Perhaps the most impactful topic of research within risk-based PBSHM, however, is the transfer of information between structures for the development and validation of degradation transition models. As discussed previously, transition models are particularly challenging to develop, as the space of possible state transitions for a given structure may be vast while the observations within the space will typically be sparse. Overcoming this challenge is vitally important for the widespread adoption of SHM technologies as decision-support systems, as without valid forecasting models, operators and regulators cannot place confidence in the recommendations of such systems. Further research into population-based SHM is required to specify methodologies of aggregating data across similar structures to develop transition models that can be shared across a population.

11.4 Closing Remarks

The primary tenet of this thesis is that structural health monitoring technologies are fundamentally decision-support tools; and therefore, such systems should be designed, developed, and implemented, accordingly. The novel research presented in the current thesis adopts a risk-based approach to formulate structural health monitoring with explicit consideration for the associated decision-support applications. Although further work is necessary to expedite the adoption of SHM systems into industry, the work in this thesis has demonstrated that by considering decision-theoretic quantities – such as expected utilities and value of information – a risk-based approach to the development of SHM systems can yield improved decision-making capabilities and cost-effective O&M strategies that can enhance the safety of our structures and infrastructure.

NOTES ON PROBABILITY

A.1 Foundations of Probability Theory

For completeness, the measure-theoretic formulation of probability theory is provided in the current appendix.

The rigorous foundations of probability were established by Kolmogorov in 1933 [62]. Using aspects of set theory, Kolmogorov introduced the concept of a probability measure space as a triple (Ω, \mathcal{F}, P) :

- Often referred to as the *sample space*, or sometimes as the *universe of discourse*, Ω is the set containing all possible outcomes from a random experiment [125]. To avoid a degenerate case, it is asserted that $\Omega \neq \emptyset$.
- Typically known as the *event space*, \mathcal{F} is a subset of the power set of Ω , i.e. $\mathcal{F} \subset \mathcal{P}(\Omega)$. The power set $\mathcal{P}(\Omega)$ refers to the set of all possible subsets of Ω , including the empty set \emptyset and Ω itself. More strictly, \mathcal{F} is a σ -algebra (alternatively known as σ -field) on the set Ω . As a σ -algebra, the set \mathcal{F} is closed when considering three types of operation acting on subsets of Ω ; specifically, the complement, and (countable) unions and intersections. This property carries with it the following implications [125]:

- i. $A \in \mathcal{F} \implies A' \in \mathcal{F}$, where A' denotes the complement of A .

ii. $A_\alpha \in \mathcal{F}, \forall \alpha \implies \bigcup_\alpha A_\alpha \in \mathcal{F}$, where α indexes over a countably infinite set A and \bigcup_α denotes the repeated union over the set indexed by α .

- P is a probability measure on the measurable space (Ω, \mathcal{F}) . Measures are often thought of as a generalisation of familiar concepts such as length, area, and volume. In general, it can be said that a measure provides a mapping from a measurable space into \mathbb{R} – in the case of Kolmogorov probability, P acts as a positive normalised measure, i.e. $P : \mathcal{F} \rightarrow [0, 1]$.

From this definition of a probability measure space, Kolmogorov's probability axioms can be summarised as follows [125]:

1. $P(A) \in [0, 1], \forall A \in \mathcal{F}$.
2. $P(\Omega) = 1$ (and $P(\emptyset) = 0$).
3. P is σ -additive; that is, any countable sequence of disjoint sets A_α satisfies $P(\bigcup_\alpha A_\alpha) = \sum_\alpha P(A_\alpha)$.

A random variable X can be thought of as a particular event or state of the world. Formally they are defined as functions on the sample space, i.e. $X(\omega), \omega \in \Omega$. To elucidate this statement, and to relate the random variable notation used in the current thesis, the reader is encouraged to consider X to be the outcome of a coin toss. For this example, $\{\text{heads}, \text{tails}\} \in \Omega$ and one can define the random variable function $X(\omega) = 1$ if $\omega = \text{heads}$ and equal to zero otherwise. For conciseness, and to improve readability, the current thesis suppresses the notation indicating the underlying sample space. In addition, rather than assigning numeric values to functions, states may be referred to by context-related words, i.e. $X(\omega) = 1$ becomes $X = \text{heads}$.

A.2 Basic Rules of Probability

In order to effectively use probabilities, some fundamental rules are required. A comprehensive overview of these rules can be found in [63, 64, 81].

When considering two events, X and Y , the probability of a union, i.e. the probability of X or Y , is given by,

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y) \quad (\text{A.1})$$

where $P(X \cap Y)$ is the probability of an intersection, i.e. the probability of X and Y . If X and Y are *mutually exclusive*, then $P(X \cap Y) = 0$ and Equation A.1 reduces to,

$$P(X \cup Y) = P(X) + P(Y). \quad (\text{A.2})$$

Often, $P(X \cap Y)$ is referred to as the *joint* probability and denoted $P(X, Y)$. The joint probability can be factorised using the *product rule*,

$$P(X, Y) = P(X|Y)P(Y) = P(Y|X)P(X). \quad (\text{A.3})$$

Here, $P(X|Y)$ denotes the *conditional* probability of X given Y . Similarly, $P(Y|X)$ denotes the probability of Y given X . Conveniently, the *chain rule* states that a joint probability distribution over multiple variables can be factorised by successive applications of the product rule.

Given $P(X, Y)$, one can *marginalise* out variables. For discrete random variables, marginalisation amounts to a summation of the possible values for a given variable,

$$P(Y) = \sum_{x \in \mathcal{X}} P(X, Y) = \sum_{x \in \mathcal{X}} P(Y|X = x)P(X = x). \quad (\text{A.4})$$

For continuous random variables, marginalisation amounts to integration over the domain,

$$P(Y) = \int_{\mathcal{X}} P(X, Y) dX = \int_{\mathcal{X}} P(Y|X = x)P(X = x) dX. \quad (\text{A.5})$$

In both cases, X , x , and \mathcal{X} are interchangeable for Y , y , and \mathcal{Y} .

Finally, independence relations between variables can be expressed. Marginal independence is defined as follows,

$$X \perp Y \iff P(X, Y) = P(X)P(Y) \quad (\text{A.6})$$

and conditional independence,

$$X \perp Y|Z \iff P(X, Y|Z) = P(X|Z)P(Y|Z) \quad (\text{A.7})$$

A.3 The Bayesian Interpretation of Probability

There are three main interpretations of probability; namely, classical, frequentist, and subjective. An introduction to each interpretation is provided in [169]. As it is the interpretation commonly used for risk analysis, the current thesis adopts the *subjective*, or *Bayesian* interpretation. In adopting a Bayesian perspective, one interprets $P(X)$ as a ‘degree of belief’ regarding the state of X .

A trivial rearrangement of equation (A.3) yields Bayes’ Theorem (equation (4.5)); restated here for convenience,

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Bayes’ Theorem is a powerful tool as it expresses how beliefs can be adjusted in the light of new evidence. From equation (4.5), the goal is to obtain $P(X|Y)$ – a *posterior* belief about the state of X , given that an observation has been made on Y . $P(X)$ denotes the *prior* belief. The prior encodes the belief about X before any evidence is considered. $P(Y|X)$ and $P(Y)$ denote the *likelihood* and *evidence*, respectively. The likelihood captures the degree to which Y is supported by X . On the other hand, the evidence captures the total support for Y over the domain \mathcal{X} ; in practice, $P(Y)$ is often computed via equation (A.4) or (A.5). It follows that the quotient $\frac{P(Y|X)}{P(Y)}$ captures the support that Y provides to X .

In SHM applications, the significance of Bayes’ Theorem is most apparent – posterior beliefs regarding structural damage, and failures, can be obtained by considering the evidence provided by monitoring systems. Additionally, the subjective interpretation of probability is a natural one. As mentioned earlier, in order to make decisions, engineers must express their ‘degree of belief’ for hypotheses such as ‘The bridge will collapse next week’. Fortunately, the Bayesian approach is amenable to decision theory [68, 170].

A.4 Bayesian Network Independence Structures

When considering the flow of information within BNs, there are three fundamental structures of interest; *serial connections*, *diverging connections* and *converging connections*. Serial connections, also known as causal chains, are of the form $X \rightarrow Y \rightarrow Z$. For a serial connection where the state of Y is unknown, X and Z are dependent; information about the state of X will influence one's belief about the state of Z and vice versa. For a serial connection where the state of Y is known, information regarding the state of either X or Z cannot influence one's belief about the state of Y , thereby preventing the flow of information between X and Z ; X and Z are independent given Y . In causal chains, this independence structure is commonly referred to as the *Markov property*. Diverging connections are of the form $X \leftarrow Y \rightarrow Z$ and represent two possible effects X and Z of a common cause Y . As with serial connections, for diverging connections X and Z are independent given Y and are dependent otherwise. Converging connections are of the form $X \rightarrow Y \leftarrow Z$ and represent a common effect Y of two possible causes X and Z . For converging connections where the state of Y is unknown, X and Z are independent. However, if the state of Y is known, information regarding the state of one of the possible influences one's belief about the state of the other cause; X and Z are dependent given Y . This dependence property of converging connections is commonly referred to as *explaining away* or *inter-causal inference*. To gain an intuitive understanding of the flow of information within Bayesian networks, including explaining away and the Markov property, the reader is directed to the Holmes and Watson example in [75].

PUBLICATIONS

B.1 Journal Papers

- **A.J. Hughes**, L.A. Bull, P. Gardner, N. Dervilis & K. Worden (2022). On robust risk-based active-learning algorithms for enhanced decision support. *Mechanical Systems and Signal Processing*, 181, 109502.
- **A.J. Hughes**, L.A. Bull, P. Gardner, R.J. Barthorpe, N. Dervilis & K. Worden (2022). On risk-based active learning for structural health monitoring. *Mechanical Systems and Signal Processing*, 167, 108569.
- **A.J. Hughes**, R.J. Barthorpe, N. Dervilis, C.R. Farrar & K. Worden (2021). A probabilistic risk-based decision framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 150, 107339.
- P. Gardner, M. Dal Borgo, V. Ruffini, **A.J. Hughes**, Y. Zhu & D.J. Wagg (2020). Towards the development of an operational digital twin. *Vibration*, 3(3), 235-265.

B.2 Book Chapters

- L.A. Bull, **A.J. Hughes**, T.J. Rogers, P. Gardner, K. Worden & N. Dervilis (2020). Partially Supervised Learning for Data-Driven Structural Health Monitoring. In *Structural Health Monitoring Based on Data Science Techniques* (pp. 389-411). Springer, Cham.

B.3 Conference Papers

- **A.J. Hughes**, R.J. Barthorpe & K. Worden (2022). On health-state transition models for risk-based structural health monitoring. In *Dynamics of Civil Structures, Volume 2* (pp. 49-60). Springer, Cham.
- **A.J. Hughes**, L.A. Bull, P. Gardner, R.J. Barthorpe, N. Dervilis & K. Worden (2021). A risk-based active learning approach to inspection scheduling. In *Proceedings of the 10th International Conference on Structural Health Monitoring of Intelligent Infrastructure*. International Society for Structural Health Monitoring of Intelligent Infrastructure (ISHMII).
- **A.J. Hughes**, R.J. Barthorpe, P. Gardner, D.J. Wagg, T.J. Rogers, E.J. Cross & K. Worden (2021). On decision-making for adaptive models combining physics and data. In *Proceedings of ISMA 2020-International Conference on Noise and Vibration Engineering and USD-2020-International Conference on Uncertainty in Structural Dynamics* (pp. 3623-3637). KU Leuven.
- P. Gardner, M. Dal Borgo, V. Ruffini, Y. Zhu & **A.J. Hughes** (2020). Towards the Development of a Digital Twin for Structural Dynamics Applications. In *Model Validation and Uncertainty Quantification, Volume 3* (pp. 165-179). Springer, Cham.
- **A.J. Hughes**, R.J. Barthorpe, C.R. Farrar & K. Worden (2021). An Augmented Risk-Based Paradigm for Structural Health Monitoring. In *Dynamics of Civil Structures, Volume 2* (pp. 201-212). Springer, Cham.
- **A.J. Hughes**, K. Worden, & R.J. Barthorpe (2019). On an application of probabilistic risk-assessment to structural health monitoring. *Structural Health Monitoring 2019*.

BIBLIOGRAPHY

- [1] C.R. Farrar and K. Worden. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, Ltd, 2013.
- [2] A. Rytter. *Vibration Based Inspection of Civil Engineering Structures*. PhD thesis, Aalborg University, 1993.
- [3] K. Worden and J.M. Dulieu-Barton. An overview of intelligent fault detection in systems and structures. *Structural Health Monitoring*, 3(1):85–98, 2004.
- [4] P. Spirtes. Introduction to causal inference. *Journal of Machine Learning Research*, 11(54):1643–1662, 2010.
- [5] P.A. Gardner. *On Novel Approaches to Model-Based Structural Health Monitoring*. PhD thesis, University of Sheffield, 2018.
- [6] E.J. Cross, S.J. Gibson, M.R. Jones, D.J. Pitchforth, S. Zhang, and T.J. Rogers. Physics-informed machine learning for structural health monitoring. In *Structural Health Monitoring Based on Data Science Techniques*, pages 347–367. Springer, 2022.
- [7] D. Straub, E. Chatzi, E. Bismut, W.M.G. Courage, M. Döhler, M.H. Faber, J. Köhler, G. Lombaert, P. Omenzetter, M. Pozzi, S. Thöns, D. Val, H. Wenzel, and D. Zonta. Value of information: A roadmap to quantifying the benefit of structural health monitoring. In *ICOSSAR - 12th International Conference on Structural Safety and Reliability*, pages 3018–3029, 2017.
- [8] E.J. Cross. *On Structural Health Monitoring in Changing Environmental and Operational Conditions*. PhD thesis, University of Sheffield, 2012.

- [9] M.A. Vega and M.D. Todd. A variational Bayesian neural network for structural health monitoring and cost-informed decision-making in miter gates. *Structural Health Monitoring*, 21(1):4–18, 2020.
- [10] M. Gobbato, J.B. Kosmatka, and C.J. P. A recursive Bayesian approach for fatigue damage prognosis : An experimental validation at the reliability component level. *Mechanical Systems and Signal Processing*, 45(2):448–467, 2014.
- [11] C. Cappello, D. Zonta, and B. Glišić. Expected utility theory for monitoring-based decision-making. *Proceedings of the IEEE*, 104(8):1647–1661, 2016.
- [12] D. Straub and M.H. Faber. Risk based inspection planning for structural systems. *Structural Safety*, 27(4):335–355, 2005.
- [13] D. Straub and M.H. Faber. Computational aspects of risk-based inspection planning. *Computer-Aided Civil and Infrastructure Engineering*, 21(3):179–192, 2006.
- [14] D. Di Francesco, M. Chryssanthopoulos, M.H. Faber, and U. Bharadwaj. Decision-theoretic inspection planning using imperfect and incomplete data. *Data-Centric Engineering*, 2:e18, 2021.
- [15] C. Li, S. Mahadevan, Y. Ling, S. Choze, and L. Wang. Dynamic Bayesian network for aircraft wing health monitoring digital twin. *AIAA Journal*, 55(3):930–941, 2017.
- [16] J. Luque and D. Straub. Risk-based optimal inspection strategies for structural systems using dynamic Bayesian networks. *Structural Safety*, 76:68–80, 2019.
- [17] J.J. Nielsen and J.D. Sørensen. On risk-based operation and maintenance of offshore wind turbine components. *Reliability Engineering & System Safety*, 96(1):218–229, 2011.
- [18] J.S. Nielsen. *Risk-Based Operation and Maintenance of Offshore Wind Turbines*. PhD thesis, Aalborg University, 2013.
- [19] J.S. Nielsen and J.D. Sørensen. Methods for risk-based planning of O&M of wind turbines. *Energies*, 7(10):6645–6664, 2014.

- [20] M.K. Hovgaard and R. Brincker. Limited memory influence diagrams for structural damage detection decision-making. *Journal of Civil Structural Health Monitoring*, 6(2):205–215, 2016.
- [21] R. Schöbi and E.N. Chatzi. Maintenance planning using continuous-state partially observable Markov decision processes and non-linear action models. *Structure and Infrastructure Engineering*, 12(8):977–994, 2016.
- [22] K.G. Papakonstantinou and M. Shinozuka. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part I: Theory. *Reliability Engineering & System Safety*, 130:202–213, 2014.
- [23] K.G. Papakonstantinou and M. Shinozuka. Planning structural inspection and maintenance policies via dynamic programming and Markov processes. Part II: POMDP implementation. *Reliability Engineering & System Safety*, 130:214–224, 2014.
- [24] K.G. Papakonstantinou and M. Shinozuka. Optimum inspection and maintenance policies for corroded structures using partially observable Markov decision processes and stochastic, physically based models. *Probabilistic Engineering Mechanics*, 37:93–108, 2014.
- [25] K.G. Papakonstantinou, C. Andriotis, and M. Shinozuka. Point-based POMDP solvers for life-cycle cost minimization of deteriorating structures. In *Life-Cycle of Engineering Systems*, pages 427–434. 2016.
- [26] K.G. Papakonstantinou, C.P. Andriotis, and M. Shinozuka. POMDP and MOMDP solutions for structural life-cycle cost minimization under partial and mixed observability. *Structure and Infrastructure Engineering*, 14(7):869–882, 2018.
- [27] P.G. Morato, K.G. Papakonstantinou, C.P. Andriotis, J.S. Nielsen, and P. Rigo. Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Structural Safety*, 94:102140, 2022.
- [28] C.P. Andriotis and K.G. Papakonstantinou. Deep reinforcement learning driven inspection and maintenance planning under incomplete information and constraints. *Reliability Engineering & System Safety*, 212:107551, 2021.

- [29] D. Zonta, B. Glišić, and S. Adriaenssens. Value of information: Impact of monitoring on decision-making. *Structural Control and Health Monitoring*, 21(7):1043–1056, 2014.
- [30] S. Thöns. On the value of monitoring information for the structural integrity and risk management. *Computer-Aided Civil and Infrastructure Engineering*, 33(1):79–94, 2018.
- [31] M. Memarzadeh and M. Pozzi. Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling. *Reliability Engineering & System Safety*, 154:137–151, 2016.
- [32] W.H. Zhang, J. Qin, D.G. Lu, M. Lui, and M.H. Faber. VoI analysis of temporally continuous SHM information in the context of adaptive risk-based inspection planning. *Structural Safety*, 99:102258, 2022.
- [33] P.F. Giordano, L.J. Prendergast, and M.P. Limongelli. Quantifying the value of SHM information for bridges under flood-induced scour. *Structure and Infrastructure Engineering*, pages 1–17, 2022.
- [34] P.F. Giordano and M.P. Limongelli. The value of structural health monitoring in seismic emergency management of bridges. *Structure and Infrastructure Engineering*, 18(4):537–553, 2022.
- [35] L. Iannacone, P.F. Giordano, P. Gardoni, and M.P. Limongelli. Quantifying the value of information from inspecting and monitoring engineering systems subject to gradual and shock deterioration. *Structural Health Monitoring*, 21(1):72–89, 2022.
- [36] P.F. Giordano, S. Quqa, and M.P. Limongelli. Impact of Decision Scenarios on the Value of Seismic Structural Health Monitoring. In *European Workshop on Structural Health Monitoring*, pages 24–33, 2023.
- [37] 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.
- [38] M. Pozzi and A. Der Kiureghian. Assessing the value of information for long-term structural health monitoring. In *Health Monitoring of Structural and Biological Systems 2011*, volume 7984, page 79842W. SPIE, 2011.

- [39] C. Malings and M. Pozzi. Value of information for spatially distributed systems: Application to sensor placement. *Reliability Engineering & System Safety*, 154: 219–233, 2016.
- [40] W.H. Zhang, J. Qin, D.G. Lu, S. Thöns, and M.H. Faber. VoI-informed decision-making for SHM system arrangement. *Structural Health Monitoring*, 21(1):37–58, 2022.
- [41] L. Long, M. Döhler, and S. Thöns. Determination of structural and damage detection system influencing parameters on the value of information. *Structural Health Monitoring*, 21(1):19–36, 2022.
- [42] T. Bedford and R. Cooke. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge University Press, 2001.
- [43] R.R. Fullwood and R.E. Hall. *Probabilistic risk assessment in the nuclear power industry*. Pergamon Books Inc., 1988.
- [44] B.J. Garrick and R.F. Christie. Probabilistic risk assessment practices in the USA for nuclear power plants. *Safety Science*, 40(1-4):177–201, 2002.
- [45] S. Tolo, E. Patelli, and M. Beer. Risk assessment of spent nuclear fuel facilities considering climate change. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(2), 2017.
- [46] J.R. Fragola and G. Maggio. Space shuttle operational risk assessment. *AIP Conference Proceedings*, 361(May):719–720, 1996.
- [47] E. Paté-Cornell and R. Dillon. Probabilistic risk analysis for the NASA space shuttle: a brief history and current work. *Reliability Engineering & System Safety*, 74(3):345–352, 2001.
- [48] K. Solomon, J. Giesy, and P. Jones. Probabilistic risk assessment of agrochemicals in the environment. *Crop Protection*, 19(8-10):649–655, 2000.
- [49] US Nuclear Regulatory Commission. PRA Procedures Guide. *NUREG/CR-2300*, 1–2, 1983.
- [50] V.T. Covello and J. Mumpower. Risk analysis and risk management: An historical perspective. *Risk Analysis*, 5(2):103–120, 1985.

- [51] A.L. Oppenheim. *Ancient Mesopotamia: Portrait of a Dead Civilization*. The University of Chicago Press, 1977.
- [52] E.W. Colglazier and R.K. Weatherwas. Failure estimates for the space shuttle. In *Abstracts for Society for Risk Analysis Annual Meeting 1986*, page 80, Boston, MA, November 1986.
- [53] US Nuclear Regulatory Commission. Reactor Safety Study: an Assessment of Accident Risks in US Commercial Nuclear Power Plants. *WASH-1400, NUREG-75/014*, 1975.
- [54] C. Synolakis and U. Kânoğlu. The Fukushima accident was preventable. *Philosophical Transactions of the Royal Society A*, 373(20140379):20140379, 2015.
- [55] US Nuclear Regulatory Commission. Probabilistic Risk Assessment (PRA). <https://www.nrc.gov/about-nrc/regulatory/risk-informed/pr.html>. Accessed: 03-11-2018.
- [56] A.M. Hasofer and C.L. Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division*, 100(1):111–121, 1974.
- [57] K. Breitung. Asymptotic approximations for multi-normal integrals. *Journal of the Engineering Mechanics Division*, 110(3):357–366, 1984.
- [58] A. Der Kiureghian, H.Z. Lin, and S.J. Hwang. Second-order reliability approximations. *Journal of the Engineering Mechanics Division*, 113(8):1208–1225, 1987.
- [59] B.J. Bichon, J.M. McFarland, and S. Mahadevan. Efficient surrogate models for reliability analysis of systems with multiple failure modes. *Reliability Engineering and System Safety*, 96(10):1386–1395, 2011.
- [60] G. Alefeld and G. Mayer. Interval analysis: Theory and applications. *Journal of Computational and Applied Mathematics*, 121(1–2):421–464, 2000.
- [61] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.
- [62] A.N. Kolmogorov. *Foundations of the Theory of Probability (1933)*. Chelsea Publishing Co., 1950.

- [63] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.
- [64] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013.
- [65] J.W. Pratt, H. Raiffa, and R. Schlaifer. *An Introduction to Statistical Decision Theory*. MIT Press, 1995.
- [66] K. Steele and H.O. Stefánsson. Decision Theory. In *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2020.
- [67] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, 1944.
- [68] L.J. Savage. *The Foundations of Statistics*. Dover Publications, 1954.
- [69] A. Marshall. *Principles of Economics*. Macmillan and Co, 1895.
- [70] M. Allais. Rational man's behavior in the presence of risk: critique of the postulates and axioms of the American school. *Econometrica*, 21:503–546, 1953.
- [71] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75(4):643–669, 1961.
- [72] A. Tversky and D. Kahneman. Rational choice and the framing of decisions. *The Journal of Business*, 59(4):S251–S278, 1986.
- [73] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- [74] L.E. Sucar. *Probabilistic Graphical Models: Principles and Applications*. London, 2015.
- [75] U.B. Kjaerulff and A.L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. New York, 2008.
- [76] G.F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.
- [77] J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, 1986.

- [78] 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, 1953.
- [79] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [80] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.
- [81] K.P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.
- [82] K.P. Murphy. *Dynamic Bayesian networks: Representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.
- [83] S. Sarkka. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [84] S.L. Lauritzen and D. Nilsson. Representing and Solving Decision Problems with Limited Information. *Management Science*, 47(9):1235–1251, 2001.
- [85] S. Gelly and D. Silver. Monte-Carlo tree search and rapid action value estimation in computer Go. *Artificial Intelligence*, 175(11):1856–1875, 2011.
- [86] A. Bobbio, L. Portinale, M. Minichino, and E. Ciancamerla. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering and System Safety*, 71(3):249–260, 2001.
- [87] S. Mahadevan, R. Zhang, and N. Smith. Bayesian networks for system reliability reassessment. *Structural Safety*, 23(3):231–251, 2001.
- [88] K. Worden, A.D. Ball, and G.R. Tomlinson. Fault location in a framework structure using neural networks. *Smart Materials and Structures*, 2(3):189–200, 1993.
- [89] A. Valkonen and B. Glišić. Evaluation tool for assessing the influence of structural health monitoring on decision-maker risk preferences. *Structural Health Monitoring*, page 1475921721992016, 2021.
- [90] M.F. Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993.

-
- [91] J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 10(3):61–74, 1999.
- [92] K. Murphy. *The Bayes Net Toolbox for Matlab*. 2001.
- [93] F. Jensen, F.V. Jensen, and S.L. Dittmer. From influence diagrams to junction trees. In *Uncertainty Proceedings 1994*, pages 367–373. Morgan Kaufmann, 1994.
- [94] R.J. Barthorpe. *On Model- and Data-based Approaches to Structural Health Monitoring*. PhD thesis, The University of Sheffield, 2011.
- [95] L.A. Bull, T.J. Rogers, C. Wickramarachchi, E.J. Cross, K. Worden, and N. Dervilis. Probabilistic active learning: An online framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 134:106294, 2019.
- [96] T.J. Rogers, K. Worden, R. Fuentes, N. Dervilis, U.T. Tygesen, and E.J. Cross. A Bayesian non-parametric clustering approach for semi-supervised Structural Health Monitoring. *Mechanical Systems and Signal Processing*, 119:100–119, 2019.
- [97] M.E. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.
- [98] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer Science + Business, LLC, 2006.
- [99] A.J. Hughes, R.J. Barthorpe, N. Dervilis, C.R. Farrar, and K. Worden. A probabilistic risk-based decision framework for structural health monitoring. *Mechanical Systems and Signal Processing*, 150:107339, 2021.
- [100] A.F. Shahraki, O.P. Yadav, and H. Liao. A review on degradation modelling and its engineering applications. *International Journal of Performability Engineering*, 13(3):299–314, 2017.
- [101] M. Corbetta, C. Sbarufatti, A. Manes, and M. Giglio. Real-time prognosis of crack growth evolution using sequential Monte Carlo methods and statistical model parameters. *IEEE Transactions on Reliability*, 64(2):736–753, 2015.

- [102] S.M. Bruemmer, E.P. Simonen, P.M. Scott, P.L. Andresen, G.S. Was, and J.L. Nelson. Radiation-induced material changes and susceptibility to intergranular failure of light-water-reactor core internals. *Journal of Nuclear Materials*, 274 (3):299–314, 1999.
- [103] J. Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96 – 146, 2009.
- [104] B. Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 765–804. 2022.
- [105] C.H. Yoo and S.C. Lee. *Stability of Structures: Principles and Applications*. Elsevier, 2011.
- [106] T.L. Anderson. *Fracture Mechanics: Fundamentals and Applications*. Taylor and Francis, 2005.
- [107] N.J.A. Sloane. The On-Line Encyclopedia of Integer Sequences: A001317. <https://oeis.org/A001317>. Accessed: 08-12-2020.
- [108] J.E. Carrion and B.F. Spencer. *Model-based strategies for real-time hybrid testing*. University of Illinois at Urbana-Champaign, 2007.
- [109] D.P. McCrum and M.S. Williams. An overview of seismic hybrid testing of engineering structures. *Engineering Structures*, 118:240–261, 2016.
- [110] M.A. Vega, Z. Hu, and M.D. Todd. Optimal maintenance decisions for deteriorating quoin blocks in miter gates subject to uncertainty in the condition rating protocol. *Reliability Engineering and System Safety*, 204:107147, 2020.
- [111] L.A. Bull, P. Gardner, J. Gosliga, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maguire, C. Campos, and K. Worden. Foundations of population-based SHM, Part I: Homogeneous populations and forms. *Mechanical Systems and Signal Processing*, 148:107141, 2021.
- [112] J. Gosliga, P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities. *Mechanical Systems and Signal Processing*, 148, 2021.
- [113] P. Gardner, L.A. Bull, J. Gosliga, N. Dervilis, and K. Worden. Foundations of population-based SHM, Part III: Heterogeneous populations – mapping and transfer. *Mechanical Systems and Signal Processing*, 148:107142, 2021.

- [114] B. Sudret, S. Marelli, and J. Wiart. Surrogate models for uncertainty quantification: An overview. In *11th European Conference on Antennas and Propagation (EUCAP)*, pages 793–797, 2017.
- [115] P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. Overcoming the problem of repair in structural health monitoring: Metric-informed transfer learning. *Journal of Sound and Vibration*, 510(116245), 2021.
- [116] R.J. Barthorpe, A.J. Hughes, and P. Gardner. A forward model driven structural health monitoring paradigm: Damage detection. In *Proceedings of IMAC XXXIX – the 39th International Modal Analysis Conference*, 2021.
- [117] G. Michau and O. Fink. Domain adaptation for one-class classification: Monitoring the health of critical systems under limited information. *arXiv:1907.09204 [stat.ML]*, 2019.
- [118] L.A. Bull, K. Worden, T.J. Rogers, C. Wickramarachchi, E.J. Cross, T. McLeay, W. Leahy, and N. Dervilis. A probabilistic framework for online structural health monitoring: active learning from machining data streams. In *Journal of Physics: Conference Series*, volume 1264, page 012028. IOP Publishing, 2019.
- [119] C. Feng, M.Y. Liu, C.C. Kao, and T.Y. Lee. *Deep Active Learning for Civil Infrastructure Defect Detection and Classification*, pages 298–306. 2017.
- [120] G. Martínez-Arellano and S. Ratchev. Towards an active learning approach to tool condition monitoring with bayesian deep learning. In *ECMS*, 2019.
- [121] D. Chakraborty, N. Kovvali, A. Papandreou-Suppappola, and A. Chattopadhyay. An adaptive learning damage estimation method for structural health monitoring. *Journal of Intelligent Material Systems and Structures*, 26(2): 125–143, 2015.
- [122] F. Schwenker and E. Trentin. Pattern classification and clustering: A review of partially supervised learning approaches. *Pattern Recognition Letters*, 37: 4–14, 2014.
- [123] A. Likas, N. Vlassis, and V.J. J. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451–461, 2003.
- [124] D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

- [125] C.A. Friedman and S. Sandow. Model performance measures for expected utility maximising investors. *International Journal of Theoretical and Applied Finance*, 6(4):355–401, 2003.
- [126] C.D. Brown and H.T. Davis. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80:355–401, 2006.
- [127] J. Maeck and G. De Roeck. Description of Z24 benchmark. *Mechanical Systems and Signal Processing*, 17(1):127–131, 2003.
- [128] G. De Roeck. The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience. *Structural Control Health Monitoring*, 10(2):127–134, 2003.
- [129] J. Maeck, B. Peeters, and G. De Roeck. Damage identification on the Z24-bridge using vibration monitoring. *Smart Materials and Structures*, 10(3):512–517, 2001.
- [130] B. Peeters and G. De Roeck. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering Structural Dynamics*, 30(2):149–171, 2001.
- [131] N. Dervilis, E.J. Cross, R.J. Barthorpe, and K. Worden. Robust methods of inclusive outlier analysis for structural health monitoring. *Journal of Sound and Vibration*, 333(20):5181–5195, 2014.
- [132] P.J. Rousseeuw and K.V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223, 1999.
- [133] 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.
- [134] I.T. Jolliffe and J. Cadima. Maximum entropy sampling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- [135] D.P. Johnson. Inspection uncertainty: The key element in nondestructive inspection. *Nuclear Engineering and Design*, 43(1):219–226, 1977.

- [136] J. Bennetts, G. Webb, S. Denton, P.J. Vardanega, and N. Loudon. Quantifying uncertainty in visual inspection data. In *Maintenance, Safety, Risk, Management and Life-Cycle Performance of Bridges*, pages 2252–2259. CRC Press, 2018.
- [137] S. Dasgupta and D. Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th International Conference on Machine Learning*, pages 208–215. Association for Computing Machinery, 2008.
- [138] O. Chapelle, B. Scholkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.
- [139] L.A. Bull, K. Worden, G. Manson, and N. Dervilis. Active learning for semi-supervised structural health monitoring. *Journal of Sound and Vibration*, 437: 373–388, 2018.
- [140] M. Grieves and J. Vickers. Digital twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In *Transdisciplinary Perspectives on Complex Systems*, pages 85–113, 2017.
- [141] S.A. Niederer, M.S. Sacks, M. Girolami, and K. Willcox. Scaling digital twins from the artisanal to the industrial. *Nature Computational Science*, 1(5): 313–320, 2021.
- [142] 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*, 167:108569, 2022.
- [143] S. Dasgupta. Two faces of active learning. *Theoretical Computer Science*, 412 (19):1767–1781, 2011.
- [144] I. Psorakis, T. Damoulas, and M. Girolami. Multiclass relevance vector machines: Sparsity and accuracy. *IEEE Transactions on Neural Networks*, 21(10): 1588–1598, 2010.
- [145] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27 (8):861–874, 2006.
- [146] T. Damoulas and M. Girolami. Probabilistic multi-class multi-kernel learning: On protein fold recognition and remote homology detection. *Bioinformatics*, 24(10):1264–1270, 2008.

- [147] S. Chen, F. Cerda, P. Rizzo, J. Bielak, J. Garrett, and J. Kovačević. Semi-supervised multiresolution classification using adaptive graph filtering with application to indirect bridge structural health monitoring. *IEEE Transactions on Signal Processing*, 62(11):2879–2893, 2014.
- [148] L.A. Bull, K. Worden, and N. Dervilis. Towards semi-supervised and probabilistic classification in structural health monitoring. *Mechanical Systems and Signal Processing*, 140:106653, 2021.
- [149] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.
- [150] G. Einicke. *Smoothing, Filtering and Prediction: Estimating the Past, Present and Future*. InTech, 2012.
- [151] J. Binder, K. Murphy, and S. Russell. Space-efficient inference in dynamic probabilistic networks. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence - Volume 2*, pages 1292–1296. Morgan Kaufmann Publishers Inc., 1997.
- [152] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [153] T. Damoulas and M. Girolami. Combining feature spaces for classification. *Pattern Recognition*, 42(11):2671–2683, 2009.
- [154] S. Rogers and M. Girolami. Multi-class Semi-supervised Learning with the ε -truncated Multinomial Probit Gaussian Process. In *Gaussian Processes in Practice*, pages 17–32. PMLR, 2007.
- [155] M.E. Tipping and A.C. Faul. Fast marginal likelihood maximisation for sparse Bayesian models. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, volume R4, pages 276–283. Proceedings of Machine Learning Research, 2003.
- [156] S. Manocha and M. Girolami. An empirical analysis of the probabilistic k-nearest neighbour classifier. *Pattern Recognition Letters*, 28(13):1818–1824, 2007.

- [157] L.A. Bull. *Towards Probabilistic and Partially-Supervised Structural Health Monitoring*. PhD thesis, University of Sheffield, 2020.
- [158] K. Worden, E.J. Cross, R.J. Barthorpe, D.J. Wagg, and P. Gardner. On digital twins, mirrors, and virtualizations: Frameworks for model verification and validation. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(3), 2020.
- [159] P. Gardner, M. Dal Borgo, V. Ruffini, A.J. Hughes, Y. Zhu, and D.J. Wagg. Towards the development of an operational digital twin. *Vibration*, 3(3): 235–265, 2020.
- [160] G. Tsialiamanis, D.J. Wagg, N. Dervilis, and K. Worden. On generative models as the basis for digital twins. *Data-Centric Engineering*, 2:e11, 2021.
- [161] G. Tsialiamanis, C. Mylonas, E. Chatzi, N. Dervilis, D.J. Wagg, and K. Worden. Foundations of population-based SHM, Part IV: The geometry of spaces of structures and their feature spaces. *Mechanical Systems and Signal Processing*, 157:107692, 2021.
- [162] K. Worden, L.A. Bull, P. Gardner, J. Gosliga, T.J. Rogers, E.J. Cross, E. Papatheou, W. Lin, and N. Dervilis. A brief introduction to recent developments in population-based structural health monitoring. *Frontiers in Built Environment*, 6:146, 2020.
- [163] C.T. Wickramarachchi, J. Gosliga, E.J. Cross, and K. Worden. On the use of graph kernels for assessing similarity of structures in population-based structural health monitoring. In *Proceedings of the Thirteenth International Workshop on Structural Health Monitoring*, 2022.
- [164] M. Dhada, M. Girolami, and A.K. Parlikad. Anomaly detection in a fleet of industrial assets with hierarchical statistical modeling. *Data-Centric Engineering*, 1:e21, 2020.
- [165] A. Kamariotis, E. Chatzi, and D. Straub. Value of information from vibration-based structural health monitoring extracted via bayesian model updating. *Mechanical Systems and Signal Processing*, 166:108465, 2022.
- [166] A.J. Hughes, L.A. Bull, P. Gardner, N. Dervilis, and K. Worden. On robust risk-based active-learning algorithms for enhanced decision support. *Mechanical Systems and Signal Processing*, 181:109502, 2022.

-
- [167] M.D. Champneys, A. Green, J. Morales, M. Sliva, and D. Mascarenas. On the vulnerability of data-driven structural health monitoring models to adversarial attack. *Structural Health Monitoring*, 20(4):1476–1493, 2022.
- [168] R. Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954.
- [169] T.L. Fine. *Theories of Probability*. Academic Press, 1973.
- [170] M.H. DeGroot. *Probability and Statistics*. Pearson, 1975.