PROBABILISTIC MACHINE LEARNING AND BAYESIAN INFERENCE FOR VIBRATION-BASED STRUCTURAL DAMAGE IDENTIFICATION

WANG XIAOYOU

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering

Probabilistic Machine Learning and Bayesian Inference for Vibration-based Structural Damage Identification

WANG Xiaoyou

A thesis submitted in partial fulfilment of the requirements for the degree of **Doctor of Philosophy**

August 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it produces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________(Signed)
WANG Xiaoyou (Name of student)

To My Family

ABSTRACT

Vibration-based structural damage detection involves the acquisition of vibration data, extraction of damage-sensitive features, and identification of novelty. These procedures intrinsically contain uncertainties arising from the measurement noise, varying environmental and operational conditions, and modeling errors. Probabilistic machine learning (ML) techniques have the ability of autonomous feature extraction and model optimization from uncertain data, thus stand out as the natural approach for vibration-based damage detection. This thesis attempts to develop advanced techniques for structural damage identification based on probabilistic ML techniques and Bayesian inference.

The vibration-based damage detection methods have two branches, data-based and inverse model-based. The Bayesian theorem can be introduced to both kinds of methods by incorporating the engineering knowledge or researcher's belief about the unseen model parameters as the prior information for the damage detection.

The Bayesian theorem is first introduced to the data-driven data normalization techniques for automatic model optimization. A linear sparse Bayesian factor analysis (FA) method is developed to discriminate the environmental effects on structural dynamic features from damage. The automatic relevance determination (ARD) prior is defined on the factor loading matrix to determine the number of underlying environmental factors automatically. The method is applied to two laboratory-tested examples (a reinforced concrete slab and a steel frame) under changing environmental conditions for damage detection. Later, an improved method on the basis of the probabilistic kernelized model is developed to remove nonlinear environmental effects. The unknown kernel parameters and the latent variables are estimated automatically in the Bayesian probabilistic framework. The method is finally applied to the benchmark Z24 bridge for damage detection.

Next, the sparse Bayesian learning (SBL) technique is introduced to the finite element (FE) model updating-based damage identification. An SBL technique is developed using the normalized damage sensitive modal parameters. The ARD prior is defined on the damage index to induce sparsity to the results. Due to the nonlinear relationship between the damage index and modal parameter, the evidence involves a complex integral that cannot be calculated directly. The Laplace approximation method and the combination of variational Bayesian inference (VBI) and delayed rejection adaptative Metropolis (DRAM) algorithm are developed to circumvent the intractable evidence. Both methods are applied to the laboratory-tested steel frame for damage detection. The comparison indicates that the analytical Laplace approximation technique is markedly efficient for low-dimensional problems as no sampling is required. The VBI-DRAM algorithm is substantially efficient in dealing with high-dimensional problems.

The last contribution of the thesis is to develop a cutting-edge domain adaptation (DA) technique for structural damage detection that can transfer knowledge from the numerical FE model to the experimental structure and from one structure to the other with different sizes, in which re-collecting the labeled damage data from the new structure is not required. A re-weighting mechanism is introduced to deal with inconsistent label spaces between the labeled source domain and unlabeled target domain, given that structural damage scenarios may be different. The numerical and experimental studies demonstrate the effectiveness of the proposed method. The comparison analysis indicates the superiority of the method, as compared with the models without DA or without the re-weighting mechanism.

PUBLICATIONS

Journal Papers

1. **Wang XY**, Hou RR, Xia Y, Zhou XQ. Laplace approximation in sparse Bayesian learning for structural damage detection. *Mechanical Systems and Signal Processing* 2020; 140: 106701.

2. **Wang XY**, Hou RR, Xia Y, Zhou XQ. Structural damage detection based on variational Bayesian inference and delayed rejection adaptive Metropolis algorithm. *Structural Health Monitoring* 2021; 20(4): 1518–1535.

3. **Wang XY**, Li LF, Beck JL, Xia Y. Sparse Bayesian factor analysis for structural damage detection under unknown environmental conditions. *Mechanical Systems and Signal Processing* 2021; 154: 107563.

4. **Wang XY**, Li LF, Tian W, Du Y, Hou RR, Xia Y. Unsupervised one-class classification for condition assessment of bridge cables using Bayesian factor analysis. *Smart Structures and Systems* 2021; Accepted.

5. Hou RR, **Wang XY**, Xia Q, Xia Y. Sparse Bayesian learning for structural damage detection under varying temperature conditions. *Mechanical Systems and Signal Processing* 2020; 145: 106965.

6. Hou RR, **Wang XY**, Xia Y. Sparse damage detection via the elastic net method using modal data. *Structural Health Monitoring* 2021; 14759217211021938.

7. **Wang XY**, Xia Y. Knowledge-transfer for structural damage detection through re-weighted adversarial domain adaptation. *Mechanical Systems and Signal Processing*, Under review.

ACKNOWLEDGEMENTS

The list of my gratitude, to all those who have helped me through many ups and downs during these three years and contributed to my success of getting a PhD, would be as long as the thesis itself. Due to the space limitation, I would like to highlight a few.

I would especially like to express my gratitude to my supervisor Professor Yong Xia for providing me the chance to pursue my PhD, introducing me to the research field of structural health monitoring, sharing his abundant knowledge and experience with me, enlightening me with his in-depth insights into problems, guiding and encouraging me through many ups and downs involved in my research during these three years. I would additionally like to thank him for being patient with me, allowing me to keep my own research pace and providing an excellent study environment. His meticulous attitude towards research, and his kindness and integrity, will continually benefit me in my career in the future.

I would also like to express my sincere thanks to Professor James L. Beck from California Institute of Technology for his helpful suggestions and enlightening ideas in my study on Bayesian inference. I would also like to thank Professor Jianfei Chen from Southern University of Science and Technology for providing me a study environment in November 2019.

My heartfelt thanks also go to my group members of Professor Xia’s group, Drs. Rong rong Hou, Lingfang Li, Qi Xia, Yiding Hu, Funian Li, Yi Zhou, Zhenghao Ding, Ying Chen, Zifeng Huang, as well as Messrs. Wanglin Wu, Wei Tian who is also my undergraduate classmate, Yao Du and Yuxuan Zhang and Miss Yushi Shan for their valuable suggestions and warm assistances both in my research and daily life, making my three years in Hong Kong a wonderful experience. I would also like to thank all technicians in Laboratory in QT005 and Mr. Ryan Wong for their assistance in the experimental work. Thanks also go to my friends Mses. Na Yang, Xinzhe Zhu, Yanlin

Zhang, and Messrs. Bo Xu, Xiong Sha, Qinlin Cai and many others in Hong Kong for all the kindness. I would additionally like to express my sincere gratitude to my dear friends outside Hong Kong. Thank Mses. Lili Liu, Runfang Yang, Xiaoyan Gao, Yuguo Niu, Linlin Gao, Wenjun Ma and Messrs. Jinyang Jiao and Menglei Sang for the encouragement.

Special thanks are due to my family, who shared with me every moment and companied me through many ups and downs in the process of getting the PhD. Without their everlasting love, continuous support, unconditional understanding, and inexhaustible patience, I wound have never started to pursue a PhD. The attitudes of my family towards life, optimism, diligence, and persistence, will benefit me throughout my life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Symbols

$\mathbf{D}_n$	n^{th} observation data
$\boldsymbol{\mathcal{W}}$, $\widehat{\boldsymbol{\mathcal{W}}}$	Factor loading matrix and the maximum a posteriori (MAP) estimate of $\boldsymbol{\mathcal{W}}$
$\mathbf{w}_i$, $\boldsymbol{w}_i$	i^{th} column of $\boldsymbol{\mathcal{W}}$, i^{th} row of $\boldsymbol{\mathcal{W}}$
$\mathbf{z}_n$	n^{th} latent variable vector
$\boldsymbol{\varepsilon}_n$, $\boldsymbol{\varepsilon}_r$, $\boldsymbol{e}_r$	n^{th} or r^{th} error vector between the model prediction and measurement data
μ, $\boldsymbol{\mu}$, $\hat{\boldsymbol{\mu}}$	Mean value, mean value vector and the MAP estimate of $\boldsymbol{\mu}$
N_m	Number of measured modes
M	Dimension of each $\mathbf{z}_n$
$\mathbf{I}$	Unity matrix
$\boldsymbol{\Psi}$, $\widehat{\boldsymbol{\Psi}}$	Covariance matrix associated with $\boldsymbol{\varepsilon}_n$ and the MAP estimate
c	Normalizing constant in the Bayesian equation
$\mathbb{E}$	Expectation
α_i, $\hat{\alpha}_i$	Hyper-parameter reflecting the precision of each variable in the prior PDF and its most probable value
$\mathbf{H}$, $\mathbf{P}$, $\mathbf{A}$, $\mathbf{G}$	Hessian matrix
$\boldsymbol{f}$	Mapping function
β, γ	Parameter associated with precision
L, J	Objective function

$\mathbb{k}$, $\mathbb{K}$	Kernel vector, kernel matrix
$\boldsymbol{\Sigma}$	Covariance matrix in the Gaussian process
$\mathbf{K}$, $\overline{\mathbf{K}}$	Stiffness matrix in the undamaged and damaged states
$\boldsymbol{M}$	Structural mass matrix
s_i, $\bar{s}_i$	Element stiffness parameter in the undamaged and damaged states
$\boldsymbol{\theta}$, $\overline{\boldsymbol{\theta}}$	Stiffness reduction parameter and its most probable value
$\phi_{j,r}$, $\boldsymbol{\phi}_r$, $\boldsymbol{\phi}$	r^{th} modal shape at jth point, r^{th} modal shape vector, modal shape matrix
λ_r, $\boldsymbol{\lambda}$, $\hat{\lambda}_r$	r^{th} eigenvalue, eigenvalue vector, MAP estimate of λ_r
$\boldsymbol{\mathcal{D}}$	Measured modal parameter
$\mathcal{M}$	Structural model class
N_p	Number of measured points
$\boldsymbol{Q}$, S	Proposed PDF to approximate the intractable PDF
$\boldsymbol{\mathcal{L}}$	Lower bound in the variational Bayesian inference
$q(\quad)$	PDF proportional to the target PDF to be approximated
C	Covariance matrix of S
N_t	Bound for adjusting C
X_s, X_t	Source data, target data
Y_s, Y_t	Source label, target label
K	Number of classes
G	Feature generator
C_1, C_2	Two classifiers with varied parameters
θ_f, θ_{c_1}, θ_{c_2}	Parameters of the feature generator, classifier C_1 and C_2
w_k,$\mathbf{W}$	k^{th} weight parameter, weight matrix in the domain adaptation

Abbreviations

AANN	Auto-Associative Neural Network
AM	Adaptative Metropolis
ANN	Artificial Neural Network
ARD	Automatic Relevance Determination
ARX	Auto-Regressive and eXogeneous
BN	Batch Normalization
CADA	Common Adversarial DA
CNN	Convolutional Neural Network
DA	Domain Adaptation
DB	Dense Block
DBN	Deep Belief Network
DenseNet	Densely Connected Convolutional Network
DL	Deep Learning
DP	Dropout
DR	Delayed Rejection
DRAM	Delayed Rejection Adaptative Metropolis
DRBM	Deep Restricted Boltzmann Machine
DS	Damage Scenario
EM	Expectation-Maximization
EOV	Environmental and Operational Variation
FA	Factor Analysis
FE	Finite Element
GAP	Global Average Pooling

GMM	Gaussian Mixture Model
GP	Gaussian Process
GPLVM	Gaussian Process Latent Variable Model
HBM	Hierarchical Bayesian Model
ICA	Independent Component Analysis
KDE	Kernel Density Estimation
KL	Kullback–Leibler
MAC	Modal Assurance Criterion
MAP	Maximum A Posteriori
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
ML	Machine Learning
MLP	Multilayer Perceptron
MMD	Maximum Mean Discrepancy
MPV	Most Probable Value
PCA	Principal Component Analysis
PDF	Probability Density Function
RADA	Re-weighted Adversarial Domain Adaptation
RBF	Radial Basis Function
RC	Reinforced Concrete
ReLU	Rectified Linear Units
RNN	Recurrent Neural Network
RVM	Relevance Vector Machine
SBL	Sparse Bayesian Learning
SHM	Structural Health Monitoring
SPE	Squared Prediction Error

SRF	Stiffness Reduction Factor
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TL	Transition Layer
t-SNE	t-Distributed Stochastic Neighbor Embedding
VBI	Variational Bayesian Inference

CHAPTER 1

INTRODUCTION

1.1 Damage Detection Methods

Civil structures are subjected to extreme natural or man-made hazards during the service life, such as earthquakes, typhoons, fires, and collisions. Small damages can accumulate incrementally with the material fatigue or corrosion, deteriorate rapidly in the case of critical fracture, and jeopardize the entire structure. Therefore, the damage should be detected as early as possible to prevent the catastrophic collapse of the structures that cause life losses and economic and social impacts.

Vibration-based structural damage detection methods are developed on the premise that structural vibration properties (such as frequencies and mode shapes) are functions of structural physical properties (e.g. mass, stiffness, damping and boundary condition). Therefore, monitoring the variations of the vibration quantities makes it possible to detect the changes in the physical properties, that is, the damage. The process typically involves the long-term monitoring of structures for data acquisition, feature extraction, and novelty detection. Some damage identification methods are based on the structural finite element (FE) model updating, which is a process to minimize the discrepancy between the model predictions and measured vibration data in either time-domain or frequency-domain. This kind of approach is referred to as the model-based approach, which is a typical inverse problem. By contrast, the data-based approach gets rid of physics-based modeling. The relationship between the uncovered features and damage state of the structure is learned by the data analysis models. The model parameters, outputs or the model prediction errors then become the damage index.

Both kinds of approaches have their pros and cons. The data-driven approaches are efficient whereas require a large amount of data to cover the range of variations. The model-based approaches could attain the damage localization and quantification in the element level, while heavily depend on an accurate FE model, which inevitably involves modeling errors (e.g. boundary conditions, joints, and materials). In general, almost all vibration-based damage detection approaches intrinsically involve uncertainties, which arise from the measurement noise, varying environmental and operational factors, and modeling errors.

1.2 Probabilistic Machine Learning

Machine learning (ML) refers to the data analysis methods that imitate the way that human learns. The goal of ML is to develop models or algorithms to learn from data for underlying feature extraction, automatic pattern recognition and then use the uncovered patterns for classification, prediction or other tasks of decision making (Jordan and Mitchell, 2015). Unlike traditional data analysis methods that need to pre-select models and model parameters, ML automates analytical model building and optimization. For example, the nonlinear problem can be estimated by the kernel function in the reproducing kernel Hilbert space without defining a specific mapping function explicitly.

ML has been a mainstay of data-driven damage detection methods in past decades through establishing non-parametric (e.g. clustering, k-nearest neighbor, and support vector machine) or parametric models (e.g. perceptron and neural network) for data analysis (Farrar and Worden, 2012). Most ML methods can be divided into two types according to whether the labels of damage data are available or not, namely, supervised or unsupervised learning. In the supervised learning, when the label is real-valued, the regression model is established and optimized to fit the relationship between the input data and label; and when the label is categorical, the data are classified for damage pattern recognition (Murphy, 2012). In the unsupervised learning, the training data are unlabeled. The goal is to recognize and classify the data according to the pre-defined metric (e.g. similarity of the underlying features or reconstruction error of the measurement data). A small number of studies fall between supervised and unsupervised methods, which are semi-supervised with a small portion of labeled data only. It is noteworthy that the scope of using ML for damage detection is not limited to the data-driven methods, but also can be physics-based (Farrar and Worden, 2012). Particularly, if the FE model has been updated and validated, the data-driven method can also be improved by exploiting information from the physical models, such as the physics-guided ML methods. Specifically, the updated FE model could simulate labeled damage data to guide the training of deep learning (DL) models for damage identification.

The Bayesian theorem that interprets the uncertainties quantitively can be introduced to the ML models for data mining. The ML model incorporating the probability theory is referred to as probabilistic ML, rather than Bayesian ML, considering that the Bayesian theorem strictly corresponds to the MAP estimate while some probabilistic methods, such as the maximum likelihood estimate, do not strictly contain the prior information (Murphy, 2012). The incorporation of the Bayesian theorem could add prior knowledge to the data probability distribution, to encode researcher’s belief about the unseen model parameters. For example, the Bayesian theorem can be introduced to regression and classification tasks for desired model optimization. One representative is the development from support vector machine (SVM) to relevance vector machine (RVM) (Tipping, 2001). The DL can also be interpreted in a probabilistic manner, such as the variational autoencoder and Bayesian neural network that consider all possible values of parameters rather than a point estimate (Kendall and Gal, 2017).

As vibration-based damage detection intrinsically contains uncertainties, the probabilistic ML approaches stand out as the natural approach. The probabilistic theory can be introduced to the existing non-probabilistic ML-based damage detection methods. There are several aspects worthy of attention and development:

(1) Some ML techniques are not originally proposed for the engineering problem, thus need to be adjusted or improved by incorporating the engineering knowledge as the prior information for the damage detection. For example, for most engineering systems, there are only a few relevant factors that account for the variations of a collection of observations. Besides, in the preliminary failure stage, the damage occurs in limited locations (elements). The sparse Bayesian learning (SBL) can be introduced for automatic model optimization.

(2) Most unsupervised methods are limited to the first level of damage detection, that is, the presence of damage. The supervised damage detection methods could attain damage localization and quantification while face the challenge of lacking labeled damage data from practical structures. The structural FE model can be built to generate labeled damage data under all possible damage scenarios (DSs). However, the FE model cannot perfectly predict the structural behavior even has been updated. Instead of expecting to build an error-free FE model, a more feasible idea is to explore the similarity between the labeled data from the FE model and unlabeled data from the real structure, and then extract the invariant damage-sensitive features shared by the two datasets.

(3) Civil structures always have different materials, geometries, and topologies and are under varying environmental conditions. The data-driven DL techniques need to be re-trained by re-collecting labeled data when the environmental conditions or structural properties change. The methods that can facilitate the knowledge transfer between different but similar structures are worth studying.

1.3 Research Objectives

This study aims to develop advanced techniques for structural damage detection based on probabilistic ML techniques and Bayesian inference. The research objectives are:

1) To develop probabilistic ML-based data normalization techniques for damage detection by incorporating the Bayesian framework:

   - Develop a sparse Bayesian latent variable method to detect the underlying

  environmental factors automatically;
  - Develop an improved probabilistic kernelized method to remove nonlinear environmental effects with model parameters optimized automatically;

2) To develop an SBL method for structural damage identification based on the model updating:
   - Develop an analytical Laplace approximation-based method to circumvent the intractable integration due to the nonlinear relationship between the damage index and modal parameters;
   - Develop a method applicable to all kinds of probability distributions (standard and nonstandard) by combing variational Bayesian inference (VBI) and delayed rejection adaptative Metropolis (DRAM) sampling together;
3) To develop an unsupervised data-driven DL technique for structural damage identification that could transfer the knowledge between structures with discrepancies:
   - Attain knowledge transfer from the FE model to an experimental structure allowing the existence of modeling errors;
   - Attain knowledge transfer from one structure to the other with different sizes.

1.4 Thesis Organization

The thesis is organized into eight chapters, as illustrated in Figure 1.1. Chapter 1 introduces the background and objectives of the thesis in general. Chapter 2 conducts a literature review of vibration-based damage detection methods. Moreover, the limitations and challenges of existing studies are issued. The rest five chapters introduce the newly developed methods.

In Chapter 3, a sparse Bayesian factor analysis (FA) technique is developed to assess the structural state under unknown and changing environmental conditions. The automatic relevance determination (ARD) prior is defined on the factor loading matrix to determine the number of environmental factors automatically.

The environmental influence may be nonlinear due to the non-uniform sensitivities of different materials to the environmental variations. Therefore, an improved method based on the probabilistic kernelized model is developed In Chapter 4, to eliminate the nonlinear environmental effects. The Gaussian kernel is employed to deal with the nonlinearity.

An SBL method is developed in Chapter 5 for structural damage identification based on the Laplace approximation technique. The item to be integrated into the evidence is assumed to be a Gaussian distribution. The posterior PDFs of the variable and hyper-parameters associated with the uncertainties are then derived in a closed-form.

Chapter 6 further improves the Bayesian method for damage detection by integrating the VBI and DRAM algorithms. A full Bayesian analysis of the posterior PDFs is conducted. The merit of the VBI-DRAM method is that it is not limited to the Gaussian distribution but is applicable to all kinds of distributions.

In Chapter 7, a re-weighted adversarial domain adaptation (RADA) method is developed to transfer knowledge between structures for damage detection using time series responses. The novelty of the method is that it considers the label space inconsistency. Consequently, it can transfer the knowledge learned from an FE model to an experimental structure and from one structure to the other with different sizes.

Chapter 8 concludes the thesis and discusses the future work.

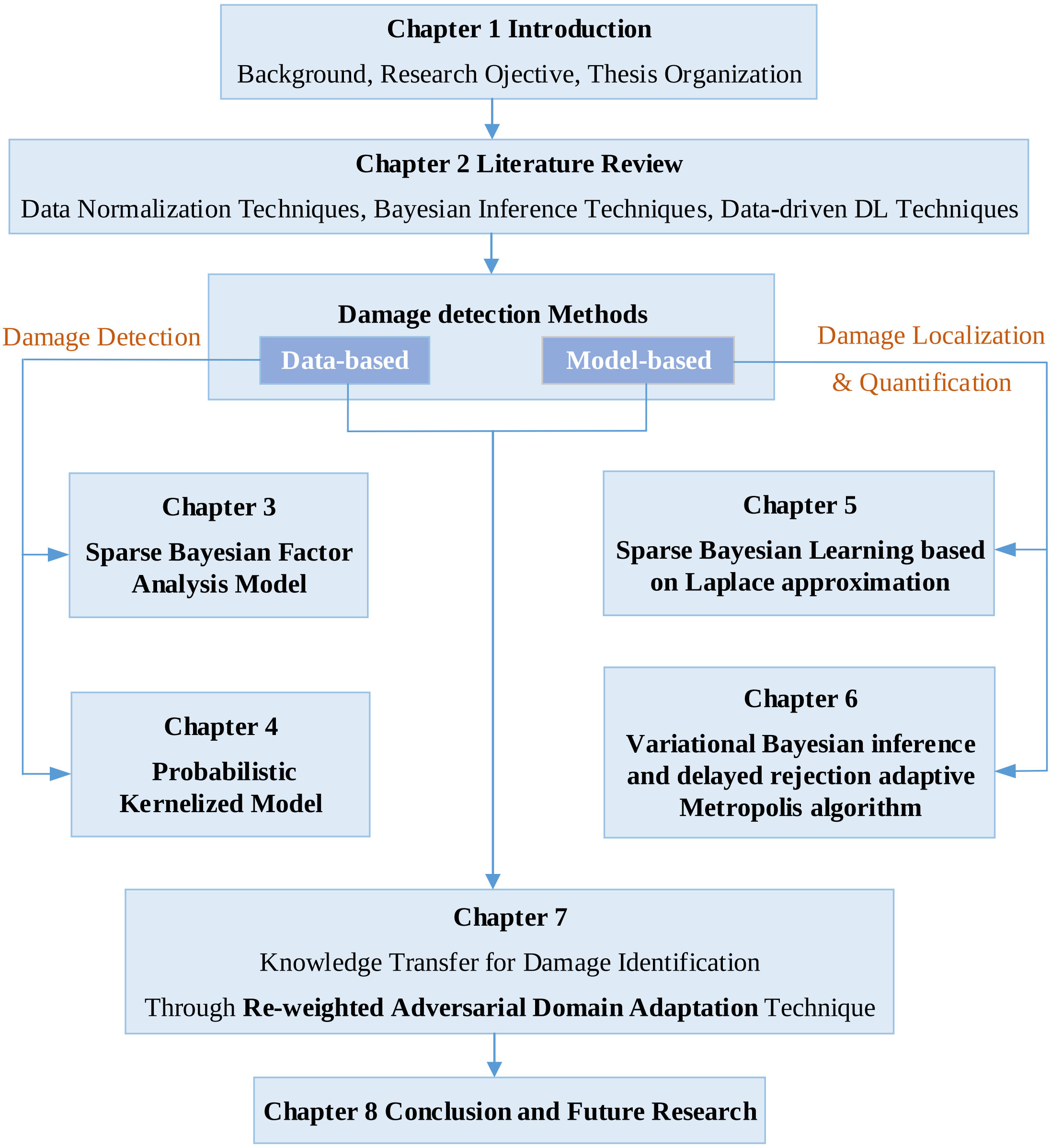


Figure 1.1 Organization of Thesis

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The vibration-based structural damage detection methods can be classified into data-based (or non-model-based) and model-based methods (Doebling et al., 1998; Farrar and Worden, 2012). The former relies on the data processing and comparison of dynamic features in the intact state and the counterpart to be predicted without establishing structural physical models. The latter requires to build an FE model of the structure, and then update and predict model parameters by minimizing the discrepancy between the model-predicted and monitored structural vibration properties.

The prerequisite of the data-based methods is that damage induces variations in structural vibration properties; thus anomaly can be detected by comparing the vibration properties over time (Doebling et al., 1996, 1998). However, normal environmental and operational variations (EOVs) may also produce changes in the structural dynamic features. If the environmental and operational effects are not considered appropriately, the damage may be masked, leading to a false structural condition assessment result. Therefore, removing environmental and operational effects on the extracted features before they are applied to structural damage detection is of high necessity. The process to remove the normal environmental and operational effects from the damage-sensitive dynamic features is referred to as data normalization (Farrar and Worden, 2012).

In the model-based methods, an objective function is generally established to evaluate the discrepancy between the model-predicted and measured properties (Sohn et al., 2003). The structural model parameters are identified by minimizing the objective function. However, measurement data are generally contaminated with noise, which introduces uncertainties in future extraction and damage detection processes. Most of the traditional damage detection methods treat the problem as deterministic and neglect the uncertainties. During the recent decades, researchers have developed probabilistic model updating or damage identification techniques, such as perturbation techniques (Xia and Hao, 2003), Monte Carlo simulations (Hjelmstad and Shin, 1997), and Bayesian inference methods (Beck and Katafygiotis, 1998; Katafygiotis and Beck, 1998).

Additionally, the huge amount of data generated in the long-term monitoring stage also imposes significant challenges on the efficient, effective and accurate damage detection. The era of big data has witnessed the development and progress of artificial intelligence technology, paving the way for the implementation of structural health monitoring (SHM) utilizing DL techniques, which have powerful ability and high efficiency in processing a large amount of data. The DL technique prevails recently with the development of a few powerful algorithms such as deep autoencoder, convolutional neural network (CNN), recurrent neural network (RNN) and so on (Sun et al., 2020). Conducting structural damage detection by adopting state-of-the-art deep learning techniques is a research hotspot.

Given the fast development of SHM in the past decades, there have been several comprehensive literature reviews in the area (Doebling et al., 1998; Salawu, 1997; Sohn et al., 2003; Sun et al., 2020). In this chapter, SHM techniques in three aspects are focused on and reviewed, including the data normalization, Bayesian inference, and data-driven DL techniques, without exhausting all aspects in this broad topic.

2.2 Data Normalization Techniques

In the following, the environmental and operational factors influencing structural dynamic features are firstly summarized, followed by the data normalization methods, which can be classified into regression methods with environmental data and ML-based methods without environmental data.

2.2.1 Environmental and Operational factors

2.2.1.1 Environmental Factors

Temperature is a major factor that may affect structural dynamic characteristics. The temperature's effects arise from several aspects. First, the Young's modulus of structural materials varies with temperature. Second, structural boundary conditions may be affected when the temperature changes. Third, the non-uniformly distributed temperature over a large-scale structure may cause an asynchronous change in structural physical parameters. Roberts and Pearson (1998) monitored a 9-span, 840 m long bridge and found that temperature accounted for a 3%~4% frequency change in a year. Farrar et al. (1996) reported that the first modal frequency of the Alamosa Canyon bridge varied by nearly 5% during a day caused by temperature variations. Peeters and De Roeck (2001) monitored the Z24 bridge for around ten months before it was artificially damaged. The results indicated that the first four frequencies varied by 14%~18% during the monitored period. They also found that the structural stiffness increased significantly below 0°C, which was due to the frozen asphalt layer of the bridge deck. The one-year monitoring of the Ting Kau Bridge in Hong Kong (Ni et al. 2005) revealed that the bridge's natural frequencies varied from 0.2% to 1.52% when the temperature changed from 2.8 to 53.5 °C. Xia et al. (2006) periodically monitored the vibration properties of a reinforced concrete (RC) slab for around two years. The first two modal frequencies were observed to decrease by 0.13% ~ 0.23% when the temperature increases one degree, whereas no obvious relationship between mode shapes and the temperature was found. Moser and Moaveni (2011) showed that the first six natural frequencies of the Dowling Hall Footbridge on the campus of Tufts University varied by 4% ~ 8% as the corresponding temperature ranged from −14 to

39 °C.

Humidity is another factor that may lead to variations in the structural vibration properties. Several studies reported that the concrete structures absorbed moisture in wet weather. As a result, the structural mass increased, and the frequencies were altered. Bendat and Piersol (1980) found that the bridge's mass increased by 3%~6% as a result of the absorption of moisture. Xia et al. (2006) found that the natural frequencies of an RC slab was reduced by 0.03% when the environmental humidity increased by 1%. Sohn (2007) reported that the mass of the Alamosa Canyon bridge increased by 1.6% due to moisture. However, Peeters and De Roeck (2001) reported that no obvious relation between humidity and the first four natural frequencies of the Z24 bridge was observed.

Additionally, the wind-induced variability in the vibration properties of long-span bridges was investigated. In Fujino et al. (2000), natural frequencies of the cable-stayed and suspension bridges were observed to decrease with the increase of the wind speed. The structural damping may also be altered when wind velocity exceeded a certain level (Fujino et al. 2000). Mahmoud et al. (2001) monitored the Hakucho Suspension Bridge and found that the amplitude of the structural vertical response was nearly a quadratic function of the wind speed. Besides, the wind speed mainly influenced the low natural frequencies of the bridge. Nagayama et al. (2005) studied the relationship between wind speed and structural vibration properties of a full-scale suspension bridge. They observed that the frequencies decreased with the increase of root mean square accelerations. Siringoringo and Fujino (2008) reported that the natural frequencies of a suspension bridge decreased as the wind velocities increased while the damping ratios increased. Martins et al. (2014) investigated the influence of wind and temperature variations on structural modal properties and responses by monitoring the Braga stadium suspension roof for eight months. The results indicated that the wind in particular directions led to a sag reduction and further resulted in the reduction of the cable tension and structural frequencies.

2.2.1.2 Operational Factors

Operational variabilities such as loadings and boundary conditions also affect structural dynamic responses. Alampalli (2000) showed that the freezing of accumulated moisture at the boundary support increased the stiffness of embedded girders and caused an increase in the natural frequencies of a steel girder bridge. In the long-term monitoring of the Z24 bridge (Peeters et al., 2001), a piecewise linear relationship between temperature and natural frequencies was observed. This phenomenon was attributed to the asphalt layer, which significantly increased the stiffness of the structural boundary condition on cold days.

The influence of traffic loading on the structural responses has also been investigated. Zhang et al. (2002) conducted 24 h vibration monitoring on a cable-stayed bridge under a relative steady wind and temperature environment. The results indicated that the natural frequencies exhibited nearly 1% change due to the traffic variability. The damping ratios increased significantly when the deck vibration exceeded a certain level. Kim et al. (2001) investigated the effect of vehicle loads on the vibration properties of three bridges with varied scales. The traffic-induced vibration tests were conducted individually. It was concluded that the effect of traffic loads on the structural dynamic characteristics depended on the relative magnitude of the traffic loads compared with the structural mass. The frequency varied by 5.4% for the short-span bridge, while no obvious variations were observed for the middle- and long-span bridges.

All these studies indicated that the influence of the environmental and operational variations on the structural dynamic characteristics should not be neglected. Instead of directly adopting the modal parameters for damage detection, data normalization methods should be employed to consider or eliminate the effects of the ambient variations.

2.2.2 Regression Methods with EOV Data

When the EOV data are available, the regression model is built to explore the relationship between the structural dynamic characteristics and EOV data, particularly the temperature. The models can be generally divided into linear and nonlinear ones.

2.2.2.1 Linear Regression Models

Most studies assume that structural natural frequencies vary with temperature linearly and establish a linear regression model to describe their relationship.

Sohn et al. (1999) proposed a linear adaptive filter to describe the frequency changes of the Alamosa Canyon bridge caused by temperature variations. The regression model was optimized by minimizing the mean square error between the model-predicted and measured frequencies. The well-trained regression model was then utilized to discriminate the temperature effect from structural damage and other environmental factors. As mentioned by the authors, a perfect match of the predicted and measured frequencies was not available due to the incompleteness of the model and the measurement noises. In this connection, a confidence interval was defined to account for the existence of uncertainties. For a new temperature profile, the defined confidence interval served as the reference to judge the change in the frequencies was caused by temperature variations or damage.

Peeters et al. (2001) used a dynamic linear Auto-Regressive and eXogeneous (ARX) model to explain the variations in frequencies of the Z24 Bridge due to changing temperature. The advantage of the dynamic ARX model over static regression models was that the input and output at previous time instants were taken into consideration to decide the current output. A confidence interval was estimated for damage prediction. When a new identified modal parameter lay outside the confidence interval, the bridge was deemed to be damaged. A comparative study showed that a single-output ARX model outperformed multiple static linear regression models.

Xia et al. (2006) developed a linear regression model to express the relationship between the natural frequencies of an RC slab and temperature and humidity. The results implied that the linear model could simulate the correlation between the frequency and temperature very well. A quantitative analysis indicated that the variation of the modal properties was primarily attributed to the variation of materials’ elastic modulus.

Bao et al. (2012) proposed a Bayesian-based technique for damage detection of a laboratory tested frame. A linear relation between the frequencies and temperature was assumed. The temperature effect on structural frequencies was then eliminated. The proposed method was applied to a two-story steel frame under changing temperature conditions for damage identification. Hou et al. (2020) later improved the Bayesian method in Bao et al. (2012) by incorporating the sparsity information into the regression model for damage detection.

2.2.2.2 Nonlinear Regression Models

In practice, the environmental effect on structural dynamic characteristics may be nonlinear. Specifically, different materials in the structures have non-uniform sensitivities to the environmental variations. The bridge’s boundary conditions may also change due to the expansion and contraction of joints, leading to a nonlinear effect. In this case, the linear model may be unable to reveal their relation accurately, and nonlinear models need to be explored.

Moaveni and Behmanesh (2012) compared four models in describing the relationship between natural frequencies and temperature of the Dowling Hall Footbridge, including a linear regression model, an ARX model, a bilinear model, and a fourth-order polynomial model. The fourth-order static polynomial model was finally chosen. Two sets of natural frequencies, before and after removing the temperature effects using the developed model, were used for FE model updating. The results indicated that the variations of parameters in the model updating were reduced, and the probability of false identification decreased after the environmental effect decreased.

Ding and Li (2011) proposed an *n*-order polynomial equation to model the correlation between the daily varying frequency of the Runyang Suspension Bridge and temperature. The order *n* increased from 1 to 10, and the sixth-order polynomial model with the least error was selected. The 40-day measured temperature data were then input to the regression model to predict the frequency variations. The result indicated that the proposed model could estimate the correlation between the structural frequency and temperature.

However, as mentioned by many researchers (Yan et al., 2015a, b; Hsu and Loh, 2010), several issues exist in the regression methods. The main limitation is that the model efficiency heavily relies on the accurate measurement of the environmental and operational factors. In contrast, ML-based data normalization methods do not need the EOV data.

2.2.3 ML-based Methods without EOV Data

When EOV data are not available, ML-based methods are developed to eliminate the environmental effects using structural dynamic responses only. The corresponding environmental factors are considered as embedded variables and not required to be measured. These ML-based methods can be further divided into linear and nonlinear dimensionality reduction models, black-box models, and others.

2.2.3.1 Linear Dimensionality Reduction Models

Principal component analysis (PCA) is a linear data dimensionality reduction technique and has been extensively employed in SHM. The latent dimensionality of the measurement data is determined in advance. Through projecting the measurement data onto the principal components, the influences from different sources are decoupled, and then the ambient impacts can be eliminated. Yan et al. (2005a) employed the PCA to discriminate the influence of environmental variations from that of damage. The damage detection was conducted based on the novelty analysis. Debarshi et al. (2019) applied the PCA for damage detection in a continuous beam structure under environmental and operational variations. The results indicated that PCA could detect the discontinuities in the measurement data at the onset of damage.

FA is another linear model that searches for the latent variables and explains the correlation between the variables and measurement data. Kullaa (2011) employed the FA to remove the environmental and operational effects on structural dynamic characteristics. The FA model was constructed using the measurement data from an undamaged structure. The structural health condition was determined according to the reconstructed data from the FA model and the original measurements. The effectiveness of the proposed method was verified by its application to a wooden bridge in the laboratory. Deraemaeker and Worden (2018) further compared the PCA and FA models. The results confirmed the ability of these two methods to filter out environmental influences.

Independent component analysis (ICA), also named the blind source separation, decomposes the given data into a linear combination of statistically independent latent variables using the observations only. Several researchers (Zang et al., 2004; Yang et al., 2014) applied the ICA for structural system identification and damage detection and verified its robustness for feature extraction and pattern recognition.

Singular value decomposition (SVD) is also a linear decomposition method that has been extensively used for dimensionality reduction. Ruotolo and Surace (1999) employed SVD to detect structural damage under environmental variation. A structure was treated as damaged when the number of singular values exceeded a threshold. Liu et al. (2015) developed a novel damage-sensitive feature extraction using SVD. The changes produced by the damage could be separated from those caused by EOVs.

2.2.3.2 Nonlinear Dimensionality Reduction Models

When the ambient influences are nonlinear during a long-term monitoring period, the kernel PCA can be adopted, in which the latent dimensions are projected to reproduce the kernel Hilbert space. The kernel PCA has the advantage that defining a specific nonlinear relationship is unnecessary and the calculation is simplified by using the kernel function.

Oh et al. (2009) employed the kernel PCA as a data normalization technique to remove the environmental effects on measurement data. Data collected from the normal conditions served as the baseline data. A Gaussian kernel PCA was applied to the baseline data to calculate the principal eigenvectors. The discrepancy between the most recent measured modal parameters and the reconstructed counterparts based on the kernel PCA was calculated and quantified by a novelty index. A threshold value was set as a reference to judge the data was normal or not. The application to the Yeongjong Grand Bridge in Korean indicated that the proposed KPCA method was able to detect the abnormality using the output data only. Reynders et al. (2014) proposed an improved kernel PCA to build the relationship between changing environmental factors and structural frequencies. The optimal parameters in the kernel function were determined automatically. The proposed method was applied to a three-span concrete bridge which was monitored for one year and damaged in the end. The proposed kernel PCA was able to detect the onset of damage accurately, whereas the linear PCA could not.

In addition to the kernel PCA, Yan et al. (2005b) used the local PCA models for local regions in the data space. The nonlinear problem was tackled in the way of a piecewise linearization. Kullaa (2014) employed the mixture of FA models, which allows different local linear factor models for different regions in the data space, to compensate for the nonlinear effects between the natural frequencies of the Z24 bridge and temperature.

2.2.3.3 Black-box Models

All techniques above share one common feature, that is, the models are established in an interpretable manner. By contrast, the auto-associative neural network (AANN) deals with the problem in a black-box and separates the environmental influences by training the NN automatically.

Sohn et al. (2001, 2007) adopted the AANN, which actually is equivalent to nonlinear PCA in a black-box form, to reveal the correlation between the measured structural modal parameters and unmeasured latent variables. The AANN network was composed of three hidden layers, namely, the mapping, the bottleneck and de-mapping layers. The number of neurons in the bottleneck layer represented the latent dimensionality of the measurement data. The discrepancy between the AANN-predicted data and measured data was calculated. A threshold value was determined as a reference to evaluate the structural health condition. The proposed method was verified through the application to a computer hard disk.

Hsu and Loh (2010) also used AANN to extract the underlying environmental factors and detect structural stiffness reduction. The effectiveness of the proposed method was verified through the application to a synthetic bridge model under varying ambient conditions, including temperature, humidity, loading and support condition. The results showed that AANN was able to accurately identify the structural stiffness reduction.

2.2.3.4 Other models

Gaussian mixture model (GMM) is a linear mixture of multiple Gaussian components, aiming at simulating a complex distribution rather than a single Gaussian. It has been widely used in data mining, pattern recognition, machine learning and statistical analysis. Kullaa (2014) used the GMM to eliminate the environmental effects from measurement data. The GMM was based on the mixture of multiple linear models, each corresponding to a local region in the input space. The data labels were regarded as latent variables in the GMM model, and the expectation-maximization (EM) algorithm was employed to identify the mixture model. The proposed method was validated using the measured natural frequencies of the Z24 Bridge under varying temperatures.

SVM can be used to learn relationships in data in a supervised manner. Ni et al. (2005) used SVM to model the correlation between structural frequencies of the Ting Kau cable-stayed bridge and temperature. In their study, the measurement data were divided into two groups, one for training the SVM model, and the other for model validation. The results indicated that such a strategy was able to tackle the problem of over-fitting. The well-trained SVM exhibited more superior generalization than the conventional method.

2.3 Bayesian Inference Techniques

Uncertainties exist in the vibration-based damage detection process, including measurement noises, methodology errors, and modeling errors. Most traditional damage identification techniques adopt a deterministic approach and ignore the uncertainties. Besides, the FE model updating for damage identification is usually an ill-posed inverse problem. A small perturbation of the measurement data will lead to a significant variation in the damage identification result. To address this challenge, researchers have incorporated the regularization technique in the objective function of model updating.

However, the regularization parameter needs to be selected in advance. Since the 1990s, the Bayesian theorem has been developed for structural system identification and damage detection. In comparison with the traditional deterministic approach, the Bayesian theorem treats the discrepancy between the model-predicted modal properties and measurement data in a probabilistic manner. Besides, all unknown hyper-parameters can be optimized automatically.

In the Bayesian inference methods, one needs to derive the posterior probability density function (PDF) and calculate the most probable value (MPV) by maximizing the posterior function, which is equivalent to minimizing the objection function in the deterministic methods. The challenging issue is to calculate the high-dimensional integral. A few techniques have been developed to circumvent the problem. These techniques include the Laplace approximation technique, expectation-maximization (EM) technique, VBI, hierarchical model, Markov chain Monte Carlo (MCMC) estimation, and so on. In the following, the application of Bayesian inference for structural damage detection and techniques will be reviewed according to the techniques employed.

2.3.1 Laplace Approximation

The Laplace approximation is an analytical technique that directly solves the computationally prohibited integral in the Bayesian equation. The PDF to be integrated is approximated as a Gaussian distribution, which is centered at the MPV with the covariance matrix equal the inverse Hessian matrix. In contrast to the numerical methodologies, the Laplace approximation results in the analytical solution and possesses a unique solution. It has been utilized to develop asymptotic expansions of posterior interests and expectations, especially in the Bayesian framework.

Beck and Katafygiotis (1998) are pioneers to apply the Laplace approximation to the Bayesian model updating for structural system identification using modal parameters. The denominator in the proposed Bayesian equation, termed as the evidence, contains a complex nonlinear and high-dimensional integral and thus is intractable. To solve this problem, systems were classified into globally identifiable, locally identifiable, and unidentifiable cases based on the measurements (Beck and Katafygiotis, 1998; Katafygiotis and Beck, 1998; Katafygiotis et al., 1998). For the globally identifiable system, the posterior PDF is highly peaked and thus can be approximated as a Gaussian distribution (Beck and Katafygiotis, 1998; Katafygiotis and Beck, 1998). Finding the MPV of the variables is equivalent to calculating the mean of the posterior Gaussian distribution. For the locally identifiable systems, the PDF is derived as a weighted combination of Gaussian distributions centered at the isolated optimal modes (Katafygiotis et al., 1998).

The globally identifiable, locally identifiable, and unidentifiable systems defined above have laid a foundation for later studies. For example, Yuen et al. (2006) utilized the Laplace approximation technique to approximate the posterior PDF in model updating using incomplete modal data without mode matching. Huang et al. (2015, 2017) combined the Laplace approximation technique and the hierarchical model to estimate the full posterior PDF. Yang et al. (2019) utilized the Laplace approximation to evaluate the uncertainties of the posterior PDF in an autoregressive model for system identification. Ni et al. (2016) proposed a fast Bayesian method for modal identification using free vibration data, in which the posterior uncertainties are evaluated based on the Laplace approximation.

2.3.2 EM Algorithm

The EM algorithm can be utilized to solve the posterior PDF in the Bayesian framework with latent variables. The algorithm consists of iterative E and M steps. In the E step, the expectation of the complete-data likelihood function with respect to the latent variable is calculated. Then in the M step, the expectation is maximized to calculate the

MPVs of variables. Through the iterative E and M steps, the variables converge to the definite values.

Ching and Beck (2004) proposed a two-step Bayesian approach for probabilistic structural model updating using incomplete measurement data. Given that the problem belongs to the non-concave and non-linear optimization problem, the EM algorithm was employed to convert the problem into two coupled concave quadratic optimization problems. The effectiveness and robustness of the proposed method were verified using the IASC-ASCE Phase II Simulation Benchmark data. The results indicated that the EM algorithm was more robust than the nonlinear optimization algorithm in searching for the MPV of the stiffness parameter. Hou et al. (2019) proposed an SBL framework for structural damage detection, in which the damage index was contained in the Bayesian equation as the latent variable. The EM algorithm was utilized to calculate the latent damage index via an iterative process.

2.3.3 Variational Inference

VBI provides an approximation solution to the intractable posterior PDF of variables in the Bayesian probabilistic framework. It proposes a tractable PDF to approximate the intractable target PDF by minimizing the Kullback–Leibler (KL) divergence (Bishop, 2006) between the proposed and target PDFs. The mean field theory (Parisi, 1988) is generally utilized to factorize the proposed PDF and independently derive the posterior PDFs of each individual unknown. In combination with the EM algorithm, the variables and parameters in the VBI model can be computed in an iterative approach.

Zhou et al. (2007) developed a structural damage classification method based on the hidden Markov model, where the VBI was employed to determine the model parameters. They pointed out that the VBI was capable to automatically select a model with appropriate complexity and alleviate the overfitting in the traditional maximum likelihood method. Fujimote et al. (2011) developed a state-space model for system identification based on the VBI, which was applicable to a wide category of systems. Baldacchino et al. (2016) developed a mixture of expert models trained by VBI for systems exhibiting nonlinear properties. Li and Der Kiureghian (2017) employed the VBI to approximate the evidence in the Bayesian operational modal analysis. The effectiveness and superiority were demonstrated through application to the One Rincon Hill Tower in San Francisco, USA.

2.3.4 Hierarchical Models

The hierarchical Bayesian model (HBM) expands the intractable posterior PDF as a series of coupled hierarchical linear PDFs. The technique simplifies the calculation and results in the posterior PDF analytically tractable posterior PDFs. The HBM contains a multi-level structure, in which the parameter in the prior PDF relies on other introduced parameters. These parameters are named the hyper-parameters and are assumed to follow designated distributions.

There have been many applications of HBM in system identification and damage detection. For example, Sedehi et al. (2019) developed a hierarchical Bayesian operational modal analysis for system identification using vibration data. Nagel and Sudret (2016) have developed a hierarchical model using noise-free vibration data. Huang et al. (2015, 2017) applied the HBM to structural damage detection and extended the complex nonlinear problem as the product of a series of coupled hierarchical linear PDFs. Behmanesh et al. (2015) applied the HBM to model updating, predicted the uncertainties of the parameters, and then extended this model for structural damage identification. Sedehi et al. (2019) compared the performance of the hierarchical Bayesian probabilistic framework with the non-hierarchical counterpart. The results indicated that the HBM is superior for its reliability and robustness in calculating uncertainty bounds.

2.3.5 MCMC Estimation

For the unidentifiable problem, no matter how complicated the PDF is, if the probabilistic expression proportional to the target PDF is available, the MCMC simulation can generate samples following the target PDF and obtain the corresponding statistical characteristics. However, the basic MCMC technique is extraordinarily time-consuming for models containing a large number of parameters or involving partial differential equations. In recent decades, the basic MCMC algorithm has been developed and improved to promote computational efficiency, including the Metropolis-Hastings (MH) algorithm, Gibbs sampling, transitional MCMC, delayed rejection (DR) technique, importance sampling, and so on.

The MH algorithm is from the MCMC and applies to high-dimensional problems. In the algorithm, the proposal distribution and the stopping criterion for sampling are two key factors that will largely affect the accuracy of the uncertainty evaluation, whereas they are uneasy to ascertain. Yuen et al. (2004) developed an adaptive MCMC simulation based on the MH algorithm for structural damage detection. The numerical integration was solved in conjunction with the adaptive MCMC to identify the important regions of the updated PDF. Lam et al. (2015) developed an enhanced MH algorithm for Bayesian model updating, in which the stopping criterion for sampling was adjusted automatically according to the important region in the posterior PDF.

Gibbs sampling is a popular MCMC technique with a straightforward idea. It carries out sampling by sweeping through the conditional PDF of each variable and keeping the rest variables fixed. Ching et al. (2006) employed the Gibbs sampling for the unidentifiable problem in the Bayesian model updating. The uncertain parameters were classified into three groups, so that the numerical sampling can be directly carried out on any group conditional on the rest. Huang et al. (2017) applied the Gibbs sampling to generate samples following the posterior PDFs of unknowns in the hierarchical Bayesian model and then calculated the statistics from the generated samples. Li et al.

(2018) employed the Gibbs sampling to calculate the posterior PDF of a Bayesian equation using operational modal analysis for system identification.

The transitional MCMC has been developed for the cases when the proposed distribution is difficult to ascertain, and successive intermediate PDFs are used to gradually approach the target PDF. Muto and Beck (2008) employed the transitional MCMC approach to tackle the unidentifiable problem in Bayesian application. Ching and Chen (2007) developed a Bayesian model updating using the transitional MCMC approach, in which a series of intermediate PDFs were used to approach the target posterior PDF step by step, and the direct sampling on the posterior PDF was then avoided.

Importance sampling simulation is a popular variance reduction technique that can be utilized to compute the failure probability of structures. Au and Beck (2003) investigated the applicability of the importance sampling to high-dimensional problems and recommended that the importance sampling could be utilized when the covariance matrix associated with each design point did not deviate obviously from the identity matrix. Au (2004) developed a probabilistic approach for failure analysis based on the importance sampling simulation, based on which the generated samples were used to compute the structural reliability.

Given that the acceptance ratio in the sampling process is usually low in high dimensional models, the delayed rejection (DR) technique was proposed to improve the efficiency. In recent years, the DRAM algorithm, as a combination of adaptive MH algorithm and DR, has been widely used. For example, Zhang et al. (2013) utilized the DRAM algorithm for a Bayesian model updating. With this technique, the responses were quantitatively predicted, and the corresponding confidence interval was obtained. Wan and Ren (2016) demonstrated that the accuracy of the DRMA algorithm for model updating outperforms the Laplace’s approximation when the posterior PDF is non-normal shaped. The validity of the DRAM algorithm to large-scale structures was also verified.

2.3.6 Others

In addition to the aforementioned techniques, other representative techniques are also employed in the Bayesian methods such as particle swarm optimization and Gaussian Process (GP).

Chen et al. (2017) defined a multi-sample objective function for structural damage detection based on the Bayesian theorem. The particle swarm optimization algorithm and an improved Nelder-Mead method were combined to optimize the objective function. The particle swarm optimization technique is a straightforward and efficient convergence algorithm and is prevalent in solving optimization problems. The proposed method was applied to the ASCE 4-storey benchmark frame.

Wan and Ni (2018) adopted GP with a Gaussian prior and Gaussian likelihood for structural response prediction. A moving window was proposed to reduce the size of the input data and further result in a reduced-order GP model. The GP model allows uncertainty evaluation and is not restricted to a definite algebraic structure. The effectiveness of the proposed approach was demonstrated using the measurement data of the Canton Tower.

2.4 Data-driven DL Techniques

DL techniques have a powerful capability to learn from and make predictions on a huge amount of data (LeCun et al., 2015). In recent years, DL techniques have been extensively explored for SHM and structural damage detection. According to the input of the DL models, these studies can be classified into data-driven and visual-based methods (Sun et al., 2020). The former category is the major concern in this study, although the latter has also been extensively explored. We reviewed the data-driven DL studies from two aspects, supervised and unsupervised learning, according to whether the labels of the training dataset are available or not.

2.4.1 Supervised Learning

Supervised learning refers to the task of learning a model that maps inputs to outputs using labeled datasets. The labels act as a “teacher” instructing the model what output is related to the specific input value. Each sample in the training dataset is a pair consisting of the input and the desired output that are both known. In terms of the architectures employed, existing supervised DL studies can be classified into CNN-based, RNN-based, deep autoencoder-based, and deep multilayer perceptron (MLP)-based.

2.4.1.1 CNN Architectures

CNN, as the state-of-the-art deep learning technique, has been a research hotspot since the AlexNet was proposed in 2012 (Krizhevsky et al., 2012). CNN consists of multiple convolutional layers, pooling layers (also named sub-sampling layer), fully-connected layers, and a decision-making layer (classification or regression) (Krizhevsky et al., 2012). The feature of the convolutional layer is local receptive field and weight sharing. The pooling layer, as a sub-sampling layer, plays the role to lower the spatial size of the feature map and improve the computation efficiency (Krizhevsky et al., 2012). The max and average pooling operations are the most frequently used ones. After multiple convolution-pooling operations, the fully-connected layer connects all previous feature maps and refines the features extracted by the convolutional layers. The dropout technique is usually employed to avoid over-fitting in this process. The final layer of CNN, namely, the softmax classifier, is implemented for classification.

CNN has the following advantages (LeCun et al., 2015; Goodfellow and Bengio, 2016). First, the number of parameters in CNN is significantly reduced with the strategy of weight sharing and local receptive field, making the CNN much easier to be trained and less likely to suffer from over-fitting. Second, CNN can extract features from the real-time measurement data without knowing the excitations and without data pre-processing and post-processing. Besides, CNN is efficient and suitable for real-time online applications.

According to the input data type, the CNN studies on structural damage detection can be further divided into time domain (i.e., accelerations, strain), frequency domain (i.e., fast Fourier transform data, transmissibility function, modal parameters), and time-frequency domain.

(1) Time domain data

Abdeljaber et al. (2017) developed an adaptive one dimensional (1D) CNN for damage detection of the Qatar University grandstand simulator using the raw accelerometer data. The damage was simulated experimentally by loosening the bolt in the joints. Each CNN was responsible for one individual joint in the structure. The proposed CNN was composed of an input layer, two hidden CNN layers consisting of the convolutional and sub-sampling layers, two fully-connected layers, and one output layer for decision making. The results demonstrated the effectiveness of the proposed 1D CNN for damage detection. Later, the same grandstand simulator was studied by Avci et al. (2018) with wireless sensors for real-time damage detection using the 1D CNN. Each CNN was assigned to each wireless sensor in the network and processed the local data only. The study processed the vibration data in three directions and identified the most distinctive damaged direction. Zhang et al. (2019) applied the 1D CNN to a short steel girder bridge for state identification using the raw vibration data. The results showed that CNN with a simple architecture was very sensitive to tiny mass and stiffness changes.

Training CNNs for structural damage detection requires a huge amount of training data of multiple DSs. However, sufficient data in all damage cases are difficult to be acquired especially for large-scale structures. In this regard, Abdeljaber et al. (2018) developed an enhanced 1D CNN, which requires only two measurement sets corresponding to the undamaged and the fully damaged cases for training. The results indicated that the well-trained CNN was able to detect all nine separate DSs accurately.

The above studies regard damage detection as the classification task, where the cross-entropy function is used as the loss function. The regression analysis using the mean square error function has also been explored. Lin et al. (2017) proposed a damage identification approach based on a customized 1D CNN. In addition to the common layers including the convolutional, pooling, fully-connected and softmax layers, the batch normalization layer as a regularization technique was also inserted into the CNN before the nonlinear activation stage in a few convolutional layers. For the damage location task, the cross entropy was used as the objective function. Whereas for the damage severity quantification, the error function was a mean square error regression equation that measures the difference between the true and predicted labels. The numerical FE model of a simply supported Euler-Bernoulli beam was utilized as an example. The proposed CNN was verified to have the robustness to noise and excellent performance of damage localization and quantification. Wu et al. (2019) also employed CNN for the regression task to predict dynamic responses of a linear single-degree-of-freedom system, a nonlinear single-degree-of-freedom system, and a 3-story multi-degree-of-freedom steel frame. A comparison study showed that CNN was more robust to noise-contaminated data and more suitable for the response prediction than the MLP.

In addition to the acceleration responses, strain data have also been used as the input. For example, Gulgec et al. (2019) used the normalized strain data as the input of CNN for damage detection and localization. The ABAQUS software was employed to simulate the damage state and generate training data. The detection task was a classification problem, while the localization was treated as a regression problem. For simplicity and efficiency, both tasks utilized the identical layers in the former architecture of CNN. The results verified the high accuracy, high efficiency and robustness of the proposed CNN model for damage diagnosis and localization.

In the 1D CNN, the measurement data should be concatenated and processed into 1D data, which will inevitably increase the length of the input. In this regard, 2D raw data-based CNN has been developed for SHM. This was inspired by the fact that multiple raw responses are generally measured by multiple sensors in practice. Khodabandehlou et al. (2019) developed a 2D deep CNN for structural condition assessment with a limited number of acceleration data. Each time-series response was transformed to a 122×70 matrix as the input of CNN. Results showed that both damage location and severity were accurately predicted using the well-trained network.

(2) Frequency domain data

In some studies, the raw time series data are transformed into frequency-domain features for preliminary feature extraction. A few representative methods are reviewed as follows.

Yu et al. (2019) developed a deep CNN-based method to assess and localize damages of a numerical building with several isolators designed underneath. The time domain signals were transformed into frequency domain features, which were compressed to a 2D matrix (5×2832) as the input of a 2D CNN. The output was a vector consisting of each floor’s health condition. The proposed CNN consisted of ten layers, including an input layer, three convolutional and subsampling layers, two fully connected layers, and an output layer. The network was trained in a supervised manner. The limitation was that only numerical investigations were conducted in this study.

Liu and Zhang (2019) utilized the short time Fourier transformation to transform the measurement data to image format, which is compatible with the visual-based CNN model. The input data were two dimensional, one was corresponding to the time coordinate, and the other was frequency, whose magnitude was represented by varying color shades. Transfer learning was also employed in this study to simplify the design and training of the CNN model.

Cofre-Martel et al. (2019) proposed a novel deep CNN model for damage detection and quantification using the image of structural transmissibility functions. The transmissibility function is the ratio of the output to input in the frequency domain. Its peak, deep and magnitude are all related to the structural health state. In the study, the input of CNN contained the values of the logarithmic magnitude of the transmissibility function and was contaminated with additional noise. The CNN was trained for damage quantification, which was a regression task.

(3) Time-frequency domain data

Some researchers combined the raw time domain data and its transformation in the frequency domain as the input.

Tang et al. (2019) split the raw response data into sections and used the data in both time and frequency domains to form the dual-channel image as the input of CNN for anomaly detection. The whole data set was manually labeled as the training data for supervised learning. Oh et al. (2019) developed a CNN model for wind-induced response estimation using three types of data together as the input, including the time domain responses at the top floor, frequency domain displacement response, and frequency domain wind speed. Wang et al. (2021) processed the vibration data using the Hilbert-Huang transform and then used the time-frequency graphs as the input of the proposed CNN model for damage detection. The proposed method had a better performance as compared with the traditional artificial neural network (ANN) and SVM models.

2.4.1.2 RNN Architectures

The RNN is a kind of ANN where connections between nodes form a temporal sequence. The difference between RNNs and other DL techniques is that the input comes from two sources: the current input data and previously perceived data. In other words, an RNN takes individual input data at a time, and the output is combined with the next input together to produce another output. Such a topology structure allows the RNN to relate the current state to historical states and thus makes it suitable for dynamic systems.

Seker et al. (2003) verified the capability of the RNN for condition monitoring and diagnosis in nuclear power plants and rotating machineries. The study illustrated that RNN was a powerful network for real time applications to grasp the system's dynamic change. Graf et al. (2010) proposed to employ the RNN for structural response prediction. The RNN model trained by time-series measurement data enabled the fuzzy data processed in a temporary sequence, which had a high degree of flexibility and generality due to its independence on structural materials and components. A textile reinforced concrete plate was employed to verify the effectiveness of the proposed RNN in response prediction.

Zhao et al. (2020) developed a strain data-driven model for structural state evaluation based on deep learning and clustering techniques. The long-short term memory network, as a special RNN to alleviate gradient disappearance problems, was employed to classify the non-stationary and stationary vehicle-induced strain. The model was trained by the adaptive moment estimation optimizer. The results indicated that the test accuracy of the classification was over 99%. The proposed RNN can be utilized as an indicator for the bridge inspection, truck-weight-limit and reinforcement work.

Perez-Ramirez et al. (2019) developed a nonlinear ARX based RNN model for predicting dynamic responses of buildings. The methodology integrated three concepts:

the empirical mode decomposition for signal processing, mutual information index to determine the number of neurons, and Bayesian algorithm for uncertainties evaluation. The model was applied to a scaled high-rise residential building and five-story steel frame. The results showed that the proposed methodology was able to predict structural responses despite the material's mechanical properties. The required training data was also reduced attributed to the Bayesian regularization.

2.4.1.3 Deep Autoencoder Architectures

The traditional autoencoder is unsupervised, which consists of two network components, namely, the encoder and decoder. The encoder maps the original input data to feature space by reducing the dimensionality of the input data to preserve the important information, and the decoder reconstructs the data based on the features. In recent years, researchers in SHM employed the autoencoder for supervised learning. The decoder was trained to learn the relationship between the features and the stiffness reduction parameters of the structure. The deep auto-encoder has more layers and is powerful for the feature extraction of complex problems, as compared with the traditional auto-encoders.

Pathirage et al. (2018) used the deep auto-encoder for modal based structural damage detection. The modal parameters were used as the input of the framework and the element stiffness reduction as the output. Two hidden layers were designed for dimensionality reduction with 100 neurons each. The hyperbolic tangent and linear activation functions were employed in the pre-training stage. Afterwards, the hyperbolic tangent function was adopted in the hidden layers. The robustness and effectiveness of the proposed algorithm for damage detection were demonstrated through the numerical and experimental examples.

Wang et al. (2018) developed a novel parallel deep autoencoder framework for damage detection using a multi-scale dataset. The modal parameters as the input were separately processed for dimensionality reduction and then combined for relationship learning. It was demonstrated that the proposed framework was able to learn the relationship between the structural modal parameters and stiffness.

2.4.1.4 Deep MLP Architectures

Deep MLP is a type of ANN with multiple hidden layers between the input and output layers, which enables it to learn high-level features of the input.

Several researchers directly utilized structural raw responses as the input of deep MLP. For example, Arangio and Beck et al. (2012) developed a two-step Bayesian multilayer perceptron network for damage identification and quantification of a long suspension bridge. In the first step, different networks were trained for structural response approximation. Damage location and severity were identified in the second step. The Bayesian model selection and ARD model improved the performance of the proposed neural network.

Raw data responses can also be directly transformed into the image format as the input of the deep MLP. For example, Bao et al. (2019) transformed structural acceleration response data into image vectors and fed them into the autoencoder deep MLP as the training data for damage detection. Three network layouts, including the parallel, fusion and multi-group layouts, were proposed. The proposed method considerably reduced false alarms induced by anomalies in the SHM system.

2.4.2 Unsupervised Learning

In practice, labeled damage data are difficult to obtain as the structural condition is unknown in advance. Therefore, supervised learning is challenging in this case. In contrast, unsupervised learning needs input data only and does not need the corresponding output; that is, the datasets are unlabeled. Unsupervised learning techniques discover hidden patterns in the dataset, cluster the dataset into groups according to similarities, or represent the dataset in a compressed format. Due to the lack of supervision, most unsupervised learning methods are limited to the first level of damage detection, that is, novelty detection or detecting the damage existence. For the second and third levels of damage detection tasks (damage localization and quantification), transfer learning-based unsupervised methods have been developed. These two types of unsupervised damage detection studies are reviewed in the following sections.

2.4.2.1 Novelty Detection

The novelty detection methods aim to discriminate the abnormal or unusual observations from the normal data in an unsupervised manner. The models are estimated using the normal data, and then the consistency of new observations with the estimated model is checked. An inconsistency that exceeds the threshold indicates the abnormality, that is, the emergence of damage. Different unsupervised deep architectures, such as the deep restricted Boltzmann machine (DRBM), deep belief network (DBN), and deep autoencoder, have been developed to detect the existence of damage.

The DRBM is a network composed of symmetrically connected stochastic binary units. No within-layer connection exists in DRBM. The network is trained layer by layer, and then finetuned through the back propagation (Salakhutdinov and Hinton, 2009). Rafiei and Adeli (2018) developed an unsupervised DRBM-based method for the condition assessment of the entire structural and substructure systems. The DRBM was composed of an encoder and decoder for feature extraction. Structural responses in the time domain under ambient excitations were transformed into the frequency domain by the fast Fourier transform and then used as the input of the DRBM. The damage index was defined on the reconstruction error of the network. The effectiveness of the proposed method was validated using the vibration data of a 1:20 scaled 38-story residential concrete high-rise building in Hong Kong.

The DBN is a multi-layered probabilistic generative model constructed by multiple RBMs (Salakhutdinov and Hinton, 2009). Different from the DBRM, the DBN does not incorporate top-down feedback. Each RBM is trained to predict the probabilities of each variable in the present layer from the pre-determined layer. Through stacking such models, the network can probabilistically reconstruct the input data without supervision (Salakhutdinov and Hinton, 2009). Pan et al. (2018) developed a Bayesian DBN for damage detection of a four-story frame. Comparative studies revealed that the deep networks achieved more accurate damage diagnostics than traditional shallow learning techniques.

The deep autoencoder has also been extensively explored for unsupervised damage detection. Ni et al. (2020) developed a deep convolutional autoencoder for anomaly detection through data compression and reconstruction. The effectiveness of the proposed method was verified using the one-month acceleration data of a long-span bridge in China. Shang et al. (2021) proposed a deep denoising autoencoder for damage detection using the cross-correlation functions of acceleration data as the input. The exponentially weighted moving average control charts were employed to determine the threshold for anomaly detection. The proposed method was applied to an experimental continuous beam under changing temperature conditions for damage detection.

2.4.2.2 Transfer Learning-based Methods

To mitigate the dependency on labeled data, transfer learning has been developed for learning from an unlabeled or few-labeled target domain with the help of a source domain where abundant labeled data are available (Pan and Yang, 2009). As a subcategory of transfer learning, unsupervised domain adaptation (DA) aims to handle the data distribution divergence between source and target domains, so that the knowledge learned from the labeled source domain could be intelligently applied to the unlabeled target domain (Pan et al., 2010; Ganin and Lempitsky, 2015). DA seeks to extract domain-invariant and target-discriminate features shared by both domains. In the last few years, DA has achieved considerable progress in computer vision tasks, such as digit recognition and semantic segmentation (Zou et al., 2018; Wang and Deng, 2018). DA methods have also been developed for structural damage detection, which can be classified into the following two categories.

The first category is the moment matching method, which aims to minimize the distribution divergence between the source and target domains in the reproducing kernel Hilbert space. For example, Garden et al. (2020) studied the knowledge transfer between different structures by developing the transfer component analysis and joint DA-based methods, where the maximum mean discrepancy (MMD) distance was utilized as the criterion. The method belongs to the shallow model that has difficulties in feature extraction and adaptation from raw measurement data. Lin et al. (2021) further developed an MMD-based deep DA network for damage localization by using structural dynamic responses as the input. The source domain consisted of labeled damage data generated from a numerical FE model, and the target data to be classified were the real-word experimental data. The domain-invariant damage sensitive features were extracted by minimizing the MMD distance between the distributions of both domains. The results indicated that the damage of the experimental beam could be accurately localized without pre-collecting labeled damage data of the beam.

The second category adopts the deep adversarial DA network. The basic deep adversarial network consists of a feature generator (extractor), classifier and domain discriminator (Ganin et al., 2016). The feature extractor and domain discriminator are trained in an adversarial manner. Thus, features extracted by the generator are shared by both domains and cannot be recognized by the domain discriminator. Xu and Noh (2021) developed a physics-informed domain adversarial network for knowledge transfer between structures with different stories for damage detection and quantification. The network was trained to extract damage-sensitive features that were robust to different structures without any labeled data. A weight parameter was designed in the loss function of DA based on the physical similarities of the building to improve the positive transfer. The proposed method achieved 90.13% and 84.46% accuracy for damage detection and quantification tasks on numerical structures with different stories, respectively, and 100% damage detection accuracy and 69.93% damage quantification accuracy when transferring from numerical to experimental structures.

2.5 Summary and Challenges

The data normalization, Bayesian inference and DL techniques for structural damage detection have been reviewed. In summary, the data normalization techniques (including the regression model, PCA, FA and other machine learning techniques) build effective models to remove the environmental and operational effects on structural vibration properties, and then evaluate the damage condition by comparing the damage-sensitive features before and after damage. The model updating-based Bayesian methods build a probabilistic framework that considers the uncertainties and evaluates the unknown parameters automatically. The data-driven DL techniques use structural dynamic responses as the input for supervised or unsupervised damage-sensitive feature extraction.

Although these techniques have achieved considerable progress, there are several challenges to be addressed:

(1) In the existing dimensionality reduction-based data normalization techniques, the latent dimensionality in the feature space is usually selected by the trial-and-error methods or according to engineers' experience, which is tedious and time consuming in processing high-dimensional data. For example, the kernel PCA involves the data projection to a high-dimensional feature space.

(2) Structural damage usually occurs at limited locations only, particularly in the initial

stage. The l_1 regularization technique can be introduced to the model updating for sparse damage detection. However, the optimal regularization parameter needs to be decided in advance manually. The Bayesian method could optimize unknown parameters automatically. The sparsity can be incorporated into the Bayesian equation for damage detection. However, an analytical posterior PDF is not available due to the computationally prohibited integration in the evidence, which is resulted from the high-dimension and the nonlinear relationship between the measured modal data and structural parameters.

(3) Most DL-based damage detection methods face the challenge of lack of labeled damage data in practical structures, as the damage state is unreversible and the structural condition is unknown in advance. Some researchers built the FE model of a real structure to generate labeled damage data, which could consider all possible DSs for network training. However, the FE model may suffer from modeling errors. The discrepancy between the FE model and real structure may lead to performance degradation when the DL model trained on the basis of the FE model is applied to the real structure. Besides, the well-trained model performs well only in the specific structure that the training data arise from. The performance of the model drops significantly when it is applied to another structure. Labeled data should be re-collected to re-train a DL model for a different structure, which is often expensive and infeasible in practice. Although the unsupervised DA-based methods have been developed, they are limited to the case where the source and target data have the identical label space, which is not realistic as structures may have different DSs.

This study will address the above difficulties and carry out studies in the following chapters.

CHAPTER 3

SPARSE BAYESIAN FACTOR ANALYSIS FOR STRUCTURAL DAMAGE DETECTION UNDER UNKNOWN ENVIRONMENTAL CONDITIONS

3.1 Introduction

The dimensionality reduction techniques are a main branch of data normalization methods without measuring the environment data, as summarized in Section 2.1. However, most methods suffer from the limitation that the number of latent dimensions to be retained is typically selected according to engineers' experiences or by a trial-and-error manner, which is tedious and time-consuming, especially for high-dimensional problems. The FA provides a multi-variate statistical model to interpret the variation of observations in terms of latent variables. In the present FA-based studies on damage detection (Kullaa, 2011; Deraemaeker et al., 2018), a diagonal covariance matrix is adopted for the error vector in the FA model, which means that each dimension in the observations is treated uncorrelated and independent. This assumption is questionable when structural vibration modes are used, given that different modes are extracted from the same set of time series responses, and thus the observed modal data should be correlated. Besides, the number of latent factors was also determined on the basis of the researchers' experience in their studies.

In practice, most physical systems have only a few relevant items that account for the variations. The underlying environmental factors that influence structural vibration properties are also deemed to be limited. In this chapter, a novel sparse Bayesian FA method is developed for structural damage detection method under unknown environmental conditions. The study contains two novelties: 1) A full rather than a diagonal covariance matrix is adopted, and 2) The sparsity is introduced to the Bayesian FA model such that the number of underlying factors can be determined automatically.

3.2 FA in Sparse Bayesian Framework

FA was developed on the basis of the assumption that, for a collection of observations, there are a set of latent variables that can describe the variabilities of the observations. The latent variables are named factors. Denote the observations as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_{2,}, \ldots, \mathbf{D}_N] \in \mathbb{R}^{N_m \times N}$, the basic FA model is expressed as (Bishop, 2006; Bartholomew et al., 2011)

$$\mathbf{D}_n = \boldsymbol{\mathcal{W}}\mathbf{z}_n + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_n \; (n = 1,2, \ldots, N) \tag{3.1}$$

where $\boldsymbol{\mathcal{W}}$ is the factor loading matrix with the dimension of $N_m \times M$; $\mathbf{z}_n$ is an M-dimensional vector whose entries consist of M independent variables with the standard Gaussian prior $\boldsymbol{\mathcal{N}}(\mathbf{z}_n|\mathbf{0}, \mathbf{I})$; $\boldsymbol{\mu}$ is the global mean of all vectors in $\mathbf{D}$; and $\boldsymbol{\varepsilon}_n$ is an N_m-dimensional error vector that accounts for the uncertainties and model prediction errors. A Gaussian prior function $\boldsymbol{\mathcal{N}}(\boldsymbol{\varepsilon}_n|\mathbf{0}, \boldsymbol{\Psi})$ is adopted for $\boldsymbol{\varepsilon}_n$. The conventional FA model restricts the matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N_m \times N_m}$ to be diagonal, which is equivalent to modeling the elements in $\mathbf{D}_n$ as uncorrelated and independent.

In this study, $\mathbf{D}$ comprises the evolution of structural natural frequencies under changing environmental conditions. As different dynamic models in $\mathbf{D}_n$ $(n = 1,2, \ldots, N)$ are correlated and dependent, the covariance matrix $\boldsymbol{\Psi}$ is treated as a full matrix instead of a diagonal matrix. In Eq. (3.1), only $\mathbf{D}_n$ are available while the remaining variables and parameters are all unknown and need to be evaluated. Calculating $\mathbf{z}_n$, $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$ from $\mathbf{D}_n$ is a classical inverse problem from the data space to the latent feature space. To this end, the Bayesian theorem that is an inverse probability function is employed.

The overall Bayesian equation to calculate the latent variable $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N] \in \mathbb{R}^{M\times N}$ is expressed as (Bishop, 2006)

$$p(\mathbf{z}|\mathbf{D}, \boldsymbol{\xi}) = c_0^{-1} p(\mathbf{D}|\mathbf{z}, \boldsymbol{\xi}) p(\mathbf{z}) \tag{3.2}$$

where $p(\mathbf{z}|\mathbf{D}, \boldsymbol{\xi})$ is the posterior PDF of the latent variable $\mathbf{z}$ given the observations $\mathbf{D}$ and parameters $\boldsymbol{\xi} = \{\mathbf{W}, \boldsymbol{\mu}, \boldsymbol{\Psi}\}$; $p(\mathbf{D}|\mathbf{z}, \boldsymbol{\xi})$ is the observed-data likelihood function; $p(\mathbf{z})$ is the prior PDF of $\mathbf{z}$; and $c_0 = \int p(\mathbf{D}|\mathbf{z}, \boldsymbol{\xi}) p(\mathbf{z}) d\mathbf{z} = p(\mathbf{D}|\boldsymbol{\xi})$ is the evidence that serves normalization. Note that in Eq. (3.2), the actual observations are substituted for $\mathbf{D}$, whereas Eq. (3.1) defines a probabilistic FA model that the observations should be.

The overall Bayesian equation of the posterior PDF of the parameters $\boldsymbol{\xi}$ is expressed as

$$p(\boldsymbol{\xi}|\mathbf{D}, \mathbf{z}) = c_1^{-1} p(\mathbf{D}, \mathbf{z}|\boldsymbol{\xi}) p(\boldsymbol{\xi}) \tag{3.3}$$

where $p(\mathbf{D}, \mathbf{z}|\boldsymbol{\xi})$ is the complete-data likelihood function; $p(\boldsymbol{\xi})$ is the prior of $\boldsymbol{\xi}$; and $c_1 = \int p(\mathbf{D}, \mathbf{z}|\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} = p(\mathbf{D}, \mathbf{z})$ is the evidence. All items in Eqs. (3.2) and (3.3) are given as follows.

3.2.1 Prior PDF

Parameters $\boldsymbol{\xi} = \{\boldsymbol{\mathcal{W}}, \boldsymbol{\mu}, \boldsymbol{\Psi}\}$ are assumed to be independent. Therefore, $p(\boldsymbol{\xi})$ is expressed as $p(\boldsymbol{\xi}) = p(\boldsymbol{\mathcal{W}}) p(\boldsymbol{\mu}) p(\boldsymbol{\Psi})$. $\boldsymbol{\mathcal{W}}$ is defined as an $N_m \times M$ matrix, where M denotes the dimension size of each latent variable and is unknown. In this study, M represents the number of environmental factors that significantly influence observations $\mathbf{D}$. In traditional methods, M is usually determined according to engineers' experiences in advance. In this study, a specific ARD prior (Bishop, 1999) is adopted for $\mathbf{W}$ for automatic model selection, enabling the automatic determination of M through an optimization process. An independent standard Gaussian distribution is defined over each column of $\boldsymbol{\mathcal{W}}$ as follows:

$$p(\boldsymbol{\mathcal{W}}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \left(\frac{\alpha_i}{2\pi}\right)^{\frac{N_m}{2}} exp\left\{-\frac{1}{2}\alpha_i \mathbf{w}_i{}^{\mathrm{T}}\mathbf{w}_i\right\} \tag{3.4}$$

where the factor loading $\mathbf{w}_i$ is the ith (i=1, …, M) column of $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M]^{\mathrm{T}}$ and α_i is the hyper-parameter, which governs the precision of $\mathbf{w}_i$. M is initially set to N_m, making $\boldsymbol{\mathcal{W}}$ an $N_m \times N_m$ matrix. In the optimization process, some items in $\boldsymbol{\alpha}$ will tend to infinity, forcing the corresponding columns in $\boldsymbol{\mathcal{W}}$ to zero. Accordingly, M will be optimized to a value denoting the number of underlying factors that significantly influence $\mathbf{D}$. For simplicity, non-informative priors are adopted for $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$. In Eq. (3.2), a standard Gaussian distribution is adopted as the prior $p(\mathbf{z})$, that is,

$$p(\mathbf{z}) = \prod_{n=1}^{N} p(\mathbf{z}_n) = \left(\frac{1}{2\pi}\right)^{\frac{MN}{2}} \prod_{n=1}^{N} exp\left\{-\frac{\mathbf{z}_n{}^{\mathrm{T}}\mathbf{z}_n}{2}\right\} \tag{3.5}$$

and $p(\mathbf{z}, \boldsymbol{\xi}) = p(\mathbf{z})p(\boldsymbol{\xi})$, where $\mathbf{z}$ and $\boldsymbol{\xi}$ are chosen to be independent a priori.

3.2.2 Observed-data Likelihood Function

According to Eq. (3.1), the statistics, including the mean and covariance of the likelihood function, are given by

$$\mathbb{E}[\mathbf{D}_n|\boldsymbol{\xi}, \mathbf{z}_n] = \boldsymbol{\mathcal{W}}\mathbf{z}_n + \boldsymbol{\mu} + \mathbb{E}[\boldsymbol{\varepsilon}_n] = \boldsymbol{\mathcal{W}}\mathbf{z}_n + \boldsymbol{\mu} \tag{3.6}$$

$$\mathrm{Cov}[\mathbf{D}_n|\boldsymbol{\xi}, \mathbf{z}_n] = \mathbb{E}[(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^{\mathrm{T}}] = \mathbb{E}[\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n{}^{\mathrm{T}}] = \boldsymbol{\Psi} \tag{3.7}$$

Thus $p(\mathbf{D}_n|\boldsymbol{\xi}, \mathbf{z}_n)$ follows the Gaussian distribution $\mathcal{N}(\mathbf{D}_n|\boldsymbol{\mathcal{W}}\mathbf{z}_n + \boldsymbol{\mu}, \boldsymbol{\Psi})$. Given that $\boldsymbol{\varepsilon}_n (n = 1,2, \dots, N)$ are mutually independent, the observed-data likelihood function is expressed as

$$\begin{aligned} p(\mathbf{D}|\boldsymbol{\xi}, \mathbf{z}) &= \prod_{n=1}^{N} p(\mathbf{D}_n|\boldsymbol{\xi}, \mathbf{z}_n) \\ &= \left(\frac{1}{2\pi}\right)^{\frac{N_m N}{2}} |\boldsymbol{\Psi}|^{-\frac{N}{2}} \prod_{n=1}^{N} exp\left\{-\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})\right\} \end{aligned} \tag{3.8}$$

where $|\boldsymbol{\Psi}|$ denotes the determinant of $\boldsymbol{\Psi}$.

3.2.3 Complete-data Likelihood Function

For each observation $\mathbf{D}_n$ there is a latent variable $\mathbf{z}_n$. $\{\mathbf{D},\mathbf{z}\}$ is named the complete-data set. In corresponding, $p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi})$ is termed the complete-data likelihood function. $p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi})$ is a joint distribution that is a product of the observed-data likelihood function $p(\mathbf{D}|\boldsymbol{\xi},\mathbf{z})$ and the prior PDF $p(\mathbf{z})$ (Bishop, 2006), derived as

$$p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi}) = p(\mathbf{D}|\boldsymbol{\xi},\mathbf{z})p(\mathbf{z}) \tag{3.9}$$

Substituting Eqs. (3.5) and (3.8) into Eq. (3.9), one has

$$\begin{aligned} p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi}) = & \left(\frac{1}{2\pi}\right)^{\frac{(M+N_m)N}{2}} |\boldsymbol{\Psi}|^{-\frac{N}{2}} \\ & \times \prod_{n=1}^{N} exp\left\{-\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu}) - \frac{\mathbf{z}_n{}^{\mathrm{T}}\mathbf{z}_n}{2}\right\} \end{aligned} \tag{3.10}$$

3.2.4 Posterior PDF

Based on Eq. (3.2), the posterior PDF of the latent variable $\mathbf{z}$ conditional on parameters $\boldsymbol{\xi}$ is derived as

$$\begin{aligned} & p(\mathbf{z}|\mathbf{D},\boldsymbol{\xi}) = c_0^{-1}p(\mathbf{D}|\boldsymbol{\xi},\mathbf{z})p(\mathbf{z}) = c_0^{-1}p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi}) \\ & \propto \prod_{n=1}^{N} exp\left\{-\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu}) - \frac{\mathbf{z}_n{}^{\mathrm{T}}\mathbf{z}_n}{2}\right\} \\ & \propto \prod_{n=1}^{N} exp\left\{\mathbf{z}_n{}^{\mathrm{T}}\boldsymbol{\mathcal{W}}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mu}) - \frac{1}{2}\mathbf{z}_n{}^{\mathrm{T}}(\mathbf{I} + \boldsymbol{\mathcal{W}}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}})\mathbf{z}_n\right\} \end{aligned} \tag{3.11}$$

which is found to be quadratic in terms of $\mathbf{z}_n$ and follows a standard Gaussian distribution as

$$p(\mathbf{z}|\mathbf{D},\boldsymbol{\xi}) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{z}_n|\bar{\mathbf{z}}_n, \mathbf{C}_\mathbf{z}) \propto \prod_{n=1}^{N} exp\left\{-\frac{1}{2}(\mathbf{z}_n - \bar{\mathbf{z}}_n)^\mathrm{T}{\mathbf{C}_\mathbf{z}}^{-1}(\mathbf{z}_n - \bar{\mathbf{z}}_n)\right\}$$
$$\propto \prod_{n=1}^{N} exp\left\{{\mathbf{z}_n}^\mathrm{T}{\mathbf{C}_\mathbf{z}}^{-1}\bar{\mathbf{z}}_n - \frac{1}{2}{\mathbf{z}_n}^\mathrm{T}{\mathbf{C}_\mathbf{z}}^{-1}\mathbf{z}_n\right\} \tag{3.12}$$

By comparing Eq. (3.11) and (3.12), the posterior mean and covariance matrix of $\mathbf{z}_n$ are seen to be

$$\bar{\mathbf{z}}_n = \mathbf{C}_\mathbf{z}\boldsymbol{\mathcal{W}}^\mathrm{T}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mu}) \in \mathbb{R}^M \tag{3.13}$$

$$\mathbf{C}_\mathbf{z} = (\mathbf{I} + \boldsymbol{\mathcal{W}}^\mathrm{T}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}})^{-1} \in \mathbb{R}^{M\times M} \tag{3.14}$$

The posterior covariance matrix $\mathbf{C}_\mathbf{z}$ in Eq. (3.14) is observed to be independent of $\mathbf{D}_n$, whereas the posterior mean $\bar{\mathbf{z}}_n$ is related to $\mathbf{D}_n$. For a Gaussian distribution, the MAP estimate is equal to the mean. Therefore, the MAP estimate of $\mathbf{z}_n$, denoted as $\hat{\mathbf{z}}_n$, is given by

$$\hat{\mathbf{z}}_n = \mathbb{E}_\mathbf{z}[\mathbf{z}_n|\mathbf{D}_n,\boldsymbol{\xi}] = \bar{\mathbf{z}}_n \tag{3.15}$$

The posterior second moment of $\mathbf{z}_n$ is then computed as

$$\mathbb{E}_\mathbf{z}[\mathbf{z}_n{\mathbf{z}_n}^\mathrm{T}|\mathbf{D}_n,\boldsymbol{\xi}] = \mathbf{C}_\mathbf{z} + \bar{\mathbf{z}}_n\bar{\mathbf{z}}_n^\mathrm{T} \tag{3.16}$$

By substituting Eqs. (3.4) and (3.10) into Eq. (3.3), the posterior PDF of parameters $\boldsymbol{\xi} = \{\boldsymbol{\mathcal{W}}, \boldsymbol{\mu}, \boldsymbol{\Psi}\}$ conditional on $\mathbf{z}$ and $\boldsymbol{\alpha}$ is derived as

$$p(\boldsymbol{\xi}|\mathbf{D},\mathbf{z},\boldsymbol{\alpha}) = c_1^{-1}p(\mathbf{D},\mathbf{z}|\boldsymbol{\xi})p(\boldsymbol{\mathcal{W}}|\boldsymbol{\alpha})$$
$$\propto |\boldsymbol{\Psi}|^{-\frac{N}{2}} \prod_{n=1}^{N} exp\left\{-\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^\mathrm{T}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu}) - \frac{{\mathbf{z}_n}^\mathrm{T}\mathbf{z}_n}{2}\right\}$$
$$\times \prod_{i=1}^{M} exp\left\{-\frac{1}{2}\alpha_i{\mathbf{w}_i}^\mathrm{T}\mathbf{w}_i\right\} \tag{3.17}$$

where the hyper-parameter vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M]^{\mathrm{T}}$ is unknown and needs to be estimated. Different from the latent variable $\mathbf{z}_n$ whose posterior statistics can be directly obtained from Eq. (3.11), Eq. (3.16) is quite complicated, and the MAP estimate of $\boldsymbol{\xi} = \{\boldsymbol{\mathcal{W}}, \boldsymbol{\mu}, \boldsymbol{\Psi}\}$ cannot be directly identified. The intractable posterior PDF is a major challenge in applying the Bayesian theorem to the high-dimensional or nonlinear problems. As summarized in Section 2.3.2, the EM algorithm is a general technique applicable to the Bayesian models with latent variables. In this connection, the EM algorithm is utilized in this chapter.

3.3 EM Algorithm

The EM algorithm consists of two steps, namely, the E step and M step. For the maximum likelihood estimate, the E step calculates the expectation of the logarithm of the complete-data likelihood function with respect to the latent variable, and then the M step maximizes this expectation. The algorithm starts by initializing the parameters, and then proceeds with the iterative E and M steps before the convergence is achieved. For the MAP estimate in this study, the M step maximizes the sum of the expectation and the prior (Bishop, 2006). The procedures are detailed as follows. It is noteworthy that the expectation with respect to **z** is the posterior one throughout this chapter.

In the E step, the expectation of the logarithm of Eq. (3.10) with respect to **z** is calculated as

$$\begin{aligned}&\mathbb{E}_{\mathbf{z}}[ln\, p\,(\mathbf{D}, \mathbf{z}|\boldsymbol{\xi})]\\ &= \frac{(M+N_m)N}{2}\ln\left(\frac{1}{2\pi}\right) - \frac{N}{2}\ln|\boldsymbol{\Psi}|\\ &+\sum_{n=1}^{N}\mathbb{E}_{\mathbf{z}}\left[-\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mathcal{W}}\mathbf{z}_n - \boldsymbol{\mu}) - \frac{{\mathbf{z}_n}^{\mathrm{T}}\mathbf{z}_n}{2}\right]\end{aligned} \tag{3.18}$$

$$= \frac{(M+N_m)N}{2}\ln\left(\frac{1}{2\mathrm{\pi}}\right) - \frac{\mathrm{N}}{2}\ln|\boldsymbol{\Psi}| - \sum_{n=1}^{N}\left[\frac{1}{2}(\mathbf{D_n} - \boldsymbol{\mu})^{\mathbf{T}}\boldsymbol{\Psi}^{\mathbf{-1}}(\mathbf{D_n} - \boldsymbol{\mu})\right]$$

$$+\sum_{n=1}^{N}\left[(\mathbf{D}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n] - \frac{1}{2}\mathbb{E}_{\mathbf{z}}({\mathbf{z}_n}^{\mathrm{T}}\boldsymbol{\mathcal{W}}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbf{z}_n)\right] - \sum_{n=1}^{N}\frac{1}{2}\mathbb{E}_{\mathbf{z}}({\mathbf{z}_n}^{\mathrm{T}}\mathbf{z}_n)$$

The sum of Eqs. (3.18) and (3.4) is calculated as

$$J(\boldsymbol{\xi}) = \mathbb{E}_{\mathbf{z}}[\ln p\,(\mathbf{D},\mathbf{z}|\boldsymbol{\xi})] + \ln p\,(\boldsymbol{\mathcal{W}}|\boldsymbol{\alpha})$$

$$= \frac{(M+N_m)N}{2}\ln\left(\frac{1}{2\pi}\right) - \frac{N}{2}\ln|\boldsymbol{\Psi}| - \sum_{n=1}^{N}\left[\frac{1}{2}(\mathbf{D}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mu})\right]$$

$$+\sum_{n=1}^{N}\left[(\mathbf{D}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n] - \frac{1}{2}\mathbb{E}_{\mathbf{z}}({\mathbf{z}_n}^{\mathrm{T}}\boldsymbol{\mathcal{W}}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbf{z}_n)\right] \tag{3.19}$$

$$-\sum_{n=1}^{N}\frac{1}{2}\mathbb{E}_{\mathbf{z}}({\mathbf{z}_n}^{\mathrm{T}}\mathbf{z}_n) + \frac{N_m M}{2}\ln\left(\frac{\alpha_i}{2\pi}\right) - \sum_{n=1}^{M}\left(\frac{1}{2}\alpha_i{\mathbf{w}_i}^{\mathrm{T}}\mathbf{w}_i\right)$$

The M step maximizes Eq. (3.19) with respect to $\boldsymbol{\xi}$ by keeping $\boldsymbol{\alpha}$ fixed. Setting the derivative of Eq. (3.19) with respect to $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}^{-1}$ separately to zero, we have

$$\frac{\partial J}{\partial \boldsymbol{\mathcal{W}}} = \sum_{n=1}^{N}\boldsymbol{\Psi}^{-1}(\mathbf{D}_n - \boldsymbol{\mu})\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]^{\mathrm{T}} - \sum_{n=1}^{N}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}] - \boldsymbol{\mathcal{W}}\mathrm{diag}(\alpha_i) = 0 \tag{3.20}$$

$$\frac{\partial J}{\partial \boldsymbol{\Psi}^{-1}} = \frac{1}{2}\sum_{n=1}^{N}[\boldsymbol{\Psi} - (\mathbf{D}_n - \boldsymbol{\mu})(\mathbf{D}_n - \boldsymbol{\mu})^{\mathrm{T}}]$$

$$+\sum_{n=1}^{N}(\mathbf{D}_n - \boldsymbol{\mu})\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]^{\mathrm{T}}\boldsymbol{\mathcal{W}}^{\mathrm{T}} - \frac{1}{2}\sum_{n=1}^{N}\mathbf{W}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}]\boldsymbol{\mathcal{W}}^{\mathrm{T}} = 0 \tag{3.21}$$

$$\frac{\partial J}{\partial \boldsymbol{\mu}} = -\sum_{n=1}^{N}[\boldsymbol{\Psi}^{-1}\boldsymbol{\mu} - \boldsymbol{\Psi}^{-1}\mathbf{D}_n] - \sum_{n=1}^{N}\boldsymbol{\Psi}^{-1}\boldsymbol{\mathcal{W}}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n] = 0 \tag{3.22}$$

According to Eq. (3.21) and (3.22), the MAP values of $\boldsymbol{\Psi}$ and $\boldsymbol{\mu}$ can be calculated in the closed-form and expressed as

$$\hat{\boldsymbol{\Psi}} = \frac{1}{N}\sum_{n=1}^{N}\{(\mathbf{D}_n-\boldsymbol{\mu})(\mathbf{D}_n-\boldsymbol{\mu})^{\mathrm{T}} - 2(\mathbf{D}_n-\boldsymbol{\mu})\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]^{\mathrm{T}}\boldsymbol{\mathcal{W}}^{\mathrm{T}} + \boldsymbol{\mathcal{W}}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}]\boldsymbol{\mathcal{W}}^{\mathrm{T}}\} \tag{3.23}$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{D}_n \tag{3.24}$$

where Eq. (3.24) is derived by substituting Eq. (3.14) into (3.22). $\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]$ and $\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}]$ in Eqs. (3.23) and (3.24) are the posterior estimates, which are calculated in Eq. (3.15) and (3.16). However, the analytical solution of $\boldsymbol{\mathcal{W}}$ cannot be directly obtained from Eq. (20). In this regard, Eq. (20) is transformed into the form of the Sylvester equation as

$$\mathbf{A}\widehat{\boldsymbol{\mathcal{W}}} + \widehat{\boldsymbol{\mathcal{W}}}\mathbf{B} = \mathbf{C} \tag{3.25}$$

where $\mathbf{A} = \boldsymbol{\Psi}^{-1}$, $\mathbf{B} = \mathrm{diag}(\alpha_i)(\sum_{n=1}^{N}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}])^{-1}$ and $\mathbf{C} = \boldsymbol{\Psi}^{-1}(\sum_{n=1}^{N}(\mathbf{D}_n - \boldsymbol{\mu})\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]^{\mathrm{T}})(\sum_{n=1}^{N}\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n{\mathbf{z}_n}^{\mathrm{T}}])^{-1}$. The MAP value $\widehat{\boldsymbol{\mathcal{W}}}$ is then calculated using the Sylvester function in MATLAB.

The Hessian matrix, which is the negative second derivative of $J(\boldsymbol{\xi})$ with respect to $\boldsymbol{\xi}$, should be positive definite to ensure that the solution in Eqs. (3.23)–(3.25) is a local maximum instead of a minimum or a saddle point. As defined, $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{N_m \times M}$, $\boldsymbol{\Psi} \in \mathbb{R}^{N_m \times N_m}$ and $\boldsymbol{\mu} \in \mathbb{R}^{N_m}$. Therefore, the Hessian matrix $\mathbf{H} \in \mathbb{R}^{(N_m \cdot M + N_m \cdot N_m + N_m) \times (N_m \cdot M + N_m \cdot N_m + N_m)}$. The high dimension of the Hessian matrix prohibits the explicit form. In this study, the Symbolic Math Toolbox in MATLAB is used to calculate the analytical expression of each entry in the Hessian matrix. In the subsequent experimental examples, the numerical solutions of $\widehat{\boldsymbol{\mathcal{W}}}$, $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\mu}}$ computed in each iteration are substituted into these entries in $\mathbf{H}$, based on which the positive definiteness is examined by calculating all eigenvalues.

A non-informative prior is adopted for hyper-parameter $\boldsymbol{\alpha}$, thus the MAP value of $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \underset{\boldsymbol{\alpha}}{argmax}\, p(\boldsymbol{\alpha}|\mathbf{D},\boldsymbol{\mu},\boldsymbol{\Psi}) = \underset{\boldsymbol{\alpha}}{argmax}\, p(\mathbf{D}|\boldsymbol{\mu},\boldsymbol{\Psi},\boldsymbol{\alpha}) \\ &= \underset{\boldsymbol{\alpha}}{argmax} \int p(\mathbf{D}|\boldsymbol{\mathcal{W}},\boldsymbol{\mu},\boldsymbol{\Psi}) p(\boldsymbol{\mathcal{W}}|\boldsymbol{\alpha}) d\boldsymbol{\mathcal{W}}\end{aligned} \tag{3.26}$$

However, it is difficult to obtain the closed-form solution of the integral. The Laplace approximation is then utilized. The item inside the integral is assumed to follow the Gaussian distribution, then the integral is computed. Assuming that the distribution $p(\mathbf{D}|\boldsymbol{\mathcal{W}},\boldsymbol{\mu},\boldsymbol{\Psi})$ has a unique peak at $\widehat{\boldsymbol{\mathcal{W}}}$, the asymptotic solution of the integral is then obtained as

$$\ln p\,(\mathbf{D}|\boldsymbol{\mu},\boldsymbol{\Psi},\boldsymbol{\alpha}) = \ln p\left(\mathbf{D}|\widehat{\boldsymbol{\mathcal{W}}},\boldsymbol{\mu},\boldsymbol{\Psi}\right) + \ln p\left(\widehat{\boldsymbol{\mathcal{W}}}|\boldsymbol{\alpha}\right) + \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{P}| \tag{3.27}$$

where $\widehat{\boldsymbol{\mathcal{W}}}$ is the MAP value of $\boldsymbol{\mathcal{W}}$ computed from Eq. (25), and $\mathbf{P}$ is the Hessian matrix expressed as

$$\mathbf{P} = -\nabla\nabla \ln p\left(\mathbf{D}|\widehat{\boldsymbol{\mathcal{W}}},\boldsymbol{\mu},\boldsymbol{\Psi}\right) p\left(\widehat{\boldsymbol{\mathcal{W}}}|\boldsymbol{\alpha}\right) \tag{3.28}$$

According to Eq. (3.27), setting the derivative of $\ln p\,(\mathbf{D}|\boldsymbol{\mu},\boldsymbol{\Psi},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ to zero, the analytical solution of $\boldsymbol{\alpha}$ is derived as

$$\hat{\alpha}_i = \frac{N_m}{\mathbf{w}_i{}^{\mathrm{T}}\mathbf{w}_i + [\mathbf{P}^{-1}]_{ii}} \tag{3.29}$$

In the Bayesian PCA, Bishop proposed to treat all parameters as well-determined, Eq. (3.29) is thus further simplified as

$$\hat{\alpha}_i = \frac{N_m}{\mathbf{w}_i{}^{\mathrm{T}}\mathbf{w}_i} \tag{3.30}$$

All solutions of $\hat{\mathbf{z}}_n$ $\widehat{\boldsymbol{\mathcal{W}}}$, $\widehat{\boldsymbol{\Psi}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\alpha}}$ in Eqs. (3.15), (3.23)–(3.25) and (3.30) are mutually coupled and thus should be computed in an iterative way. The EM algorithm starts by initializing $\widehat{\boldsymbol{\mathcal{W}}}$ and $\widehat{\boldsymbol{\Psi}}$ and calculating $\hat{\boldsymbol{\mu}}$, and then proceeds with computing the posterior statistics of $\mathbf{z}$ in Eqs. (3.15) and (3.16), followed by updating $\widehat{\boldsymbol{\mathcal{W}}}$, $\widehat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\alpha}}$ using Eqs. (23)–(3.25) and (3.30). The iteration is stopped until the convergence is achieved, for example, $\left\|\boldsymbol{\Psi}^{(j)} - \boldsymbol{\Psi}^{(j-1)}\right\| / \left\|\boldsymbol{\Psi}^{(j)}\right\| \leq Tol$ (e.g. $Tol = 1\times10^{-5}$) . The number of underlying factors is first assumed to equal the original dimensionality N_m and $\boldsymbol{\mathcal{W}}$ is thus initialized as an $N_m \times N_m$ matrix. As the optimization proceeds, some of α_i will approach infinity, forcing the corresponding column $\mathbf{w}_i$ in $\boldsymbol{\mathcal{W}}$ to zero. To ensure a unique solution of the Sylvester equation in Eq. (3.25), the zero column $\mathbf{w}_i$ is removed in the iterative process. $\boldsymbol{\mathcal{W}}$ converges to an $N_m \times M$ matrix, where M equals the number of latent factors. In this manner, the underlying environmental factors that are significantly related to the variations of structural dynamic modes are identified automatically.

The proposed damage detection method based on the sparse Bayesian factor analysis is summarized as follows:

---

1. Extract structural vibration properties from the measurement data under changing environmental conditions as the entire dataset **D** and divide the entire dataset as the training and test data;
2. Set $M = N_m$ first, initialize parameters $\boldsymbol{\mathcal{W}}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$, and calculate the mean $\boldsymbol{\mu}$ of the training data according to Eq. (3.24);
3. Use the training data to estimate the unknown variables and parameters iteratively. At the jth iteration:

   (1) Set M to be the number of nonzero columns in $\boldsymbol{\mathcal{W}}^{(j-1)}$. Update $\boldsymbol{\mathcal{W}}^{(j-1)}$ as an $N_m \times M$ matrix by removing the zero columns $\mathbf{w}_i$;

   (2) Given $\boldsymbol{\mu}$, $\boldsymbol{\Psi}^{(j-1)}$ and $\boldsymbol{\mathcal{W}}^{(j-1)}$:

   Update $\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n]^{(j)}$ and $\mathbb{E}_{\mathbf{z}}[\mathbf{z}_n\mathbf{z}_n{}^{\mathrm{T}}]^{(j)}$ according to Eqs. (3.15) and (3.16);

   Update $\boldsymbol{\alpha}^{(j)}$ using Eq. (3.30);

   (3) Update $\boldsymbol{\Psi}^{(j)}$ and $\boldsymbol{\mathcal{W}}^{(j)}$ according to Eqs. (3.23) and (3.25), respectively;
4. Let $j = j + 1$, repeat Step 3 until the convergence is achieved, that is, $\left\|\boldsymbol{\Psi}^{(j)} - \boldsymbol{\Psi}^{(j-1)}\right\| / \left\|\boldsymbol{\Psi}^{(j)}\right\| \leq Tol$ (e.g. $Tol = 1\times10^{-5}$).
5. Given the estimated $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Psi}$, calculate $\hat{\mathbf{z}}_n$ of the entire dataset **D** according to

Eq. (3.15), and then re-generate the dataset using $\mathbf{D}' = \boldsymbol{\mathcal{W}}\hat{\mathbf{z}} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$.

6. Compute the Euclidean norm of the error vector $\|\mathbf{e}\|_2 = \|\mathbf{D} - \mathbf{D}'\|_2$ as the damage indicator for structural condition assessment.

The proposed method has the following advantages. First, the number of environmental factors that significantly affect structural dynamic responses could be determined automatically without measuring the environmental data. Second, the EM algorithm can be implemented in an online form, that is, each data sample $\mathbf{D}_n$ can be processed and then discarded before the following data comes in.

3.4 Case Studies

The proposed method is applied to two experimental structures under varying environmental conditions, one is an RC slab (Xia et al., 2006), and the other is a two-story steel frame (Bao et al., 2012).

3.4.1 The RC Slab

The entire RC slab measures 6400 mm × 800 mm × 100 mm with two equal spans of 3000 mm and 200 mm overhang at each end, as shown in Figure 3.1. The structure was monitored for almost two years, from June 2003 to March 2005. A set of hammer tests were conducted on the slab, and the acceleration responses were recorded by 12 accelerometers with a sampling frequency of 500 Hz. The positions of accelerometers are depicted in Figure 3.2. The modal parameters of the slab were extracted using the Rational Fraction Polynomial method (Formenti and Richardson, 2002). The air temperature and humidity were also recorded using a thermohygrometer.

Figure 3.1 RC slab

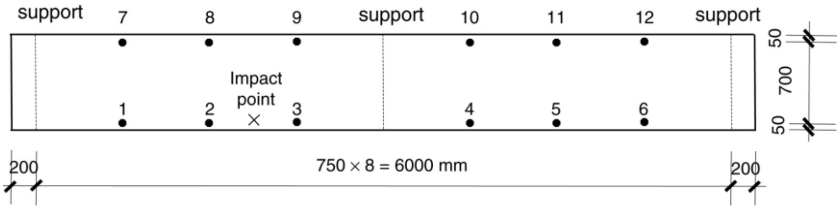


Figure 3.2 Sensor layout on the RC slab

A total of 136 sets of modal properties were collected over the monitoring period. Although the environmental data are available, they are assumed to be unknown and not used in this study. The variations of the first four frequencies are used as the dataset $\mathbf{D}$, which are plotted in Figure 3.3. It is noteworthy that the fourth frequency of Nos. 39~51 data fail to be identified in the hammer test. Thus the first 100 sets of frequencies except the lost data are set to the training data, based on which the unknown parameters are estimated. $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\Psi}$ are initialized, and the mean of the training data is computed first. In this study, the $N_m \times N_m$ entries in $\boldsymbol{\mathcal{W}}^{(0)}$ are randomly generated from the uniform distribution in the interval (0, 1). $\boldsymbol{\Psi}$ is the covariance matrix related to the model prediction error. $\boldsymbol{\Psi}$ is initialized as a semi-positive definite matrix by assuming that the coefficient of variation of each natural frequency is 1% and the correlation coefficient between modes is 0.5. The convergence criterion is set to $\left\|\boldsymbol{\Psi}^{(j)} - \boldsymbol{\Psi}^{(j-1)}\right\| / \left\|\boldsymbol{\Psi}^{(j)}\right\| \leq 1 \times 10^{-5}$ in this study.

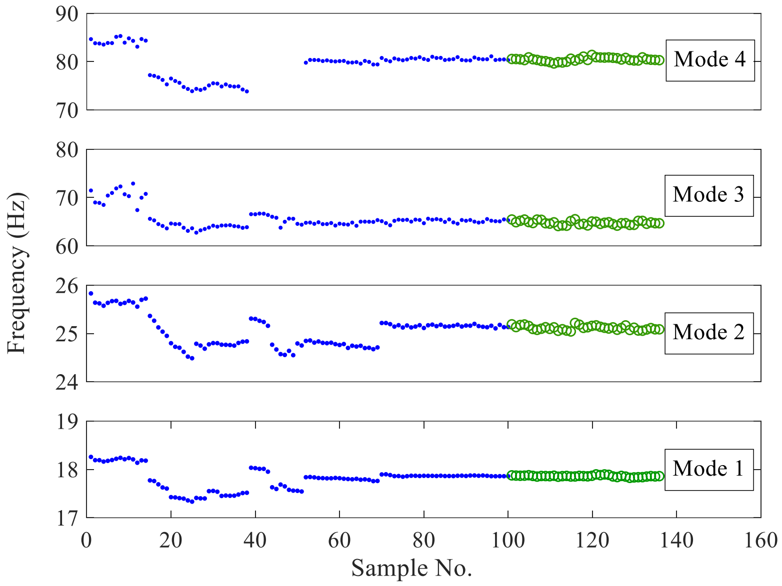


Figure 3.3 Variations of the first four frequencies (blue dots: training data in the undamaged state (N=100); green circles: test data)

$\boldsymbol{\mathcal{W}}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ are shown in Table 3.1. The EM algorithm took fifteen iteration steps before the convergence criterion is satisfied. The variations of the latent variable and parameters are listed in Table 3.1. $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\mu}$ estimated in each step are substituted into the analytical expression of each entry in the Hessian matrix, which is calculated by the Symbolic Math Toolbox in MATLAB. The Hessian matrix keeps positive definite in each step, that is, all eigenvalues are positive, implying that the solutions of parameters correspond to a local maximum of $J(\boldsymbol{\xi})$ in Eq. (3.19). As the iteration progresses, two entries in $\mathbf{z}_n$ become zero. The remaining two nonzero entries in $\mathbf{z}_n$ indicate that two latent factors account for the variations of structural frequencies. This result is consistent with the finding in Xia et al. (2006), where the correlation analysis between structural frequencies and environmental factors were conducted, and two factors, the humidity and temperature, were found to have a strong correlation with the slab's frequencies. This example indicates that the proposed sparse Bayesian FA method can identify the number of environmental factors that significantly affect the structural vibration properties without measuring the environmental factors. This is promising in practical applications since the latent variables are usually unknown.

Table 3.1 Variation of the unknown variable and parameters in the iteration process

Iteration No.	$\boldsymbol{\alpha}$	$\boldsymbol{\mathcal{W}}$	$\mathbf{z}_1$	M	$\boldsymbol{\Psi}$
0		$\begin{bmatrix} 0.19 & 0.44 & 0.58 & 0.18 \\ 0.04 & 0.43 & 0.59 & 0.29 \\ 0.60 & 0.75 & 0.17 & 0.71 \\ 0.96 & 0.43 & 0.81 & 0.26 \end{bmatrix}$		4	$\begin{bmatrix} 0.03 & 0.02 & 0.06 & 0.07 \\ 0.02 & 0.06 & 0.08 & 0.10 \\ 0.06 & 0.08 & 0.43 & 0.26 \\ 0.07 & 0.10 & 0.26 & 0.63 \end{bmatrix}$
1	$\begin{Bmatrix} 3.1 \\ 3.6 \\ 2.9 \\ 5.8 \end{Bmatrix}$	$\begin{bmatrix} 0.12 & 0.04 & 0.06 & 0.03 \\ 0.13 & 0.13 & 0.13 & 0.14 \\ 0.77 & 0.20 & -0.70 & 0.59 \\ 1.55 & -0.09 & 0.17 & 0.10 \end{bmatrix}$	$\begin{Bmatrix} 3.04 \\ 0.46 \\ -1.62 \\ 2.57 \end{Bmatrix}$	4	$\begin{bmatrix} 0.01 & 0.01 & 0.04 & 0.08 \\ 0.01 & 0.04 & 0.08 & 0.11 \\ 0.04 & 0.09 & 0.53 & 0.41 \\ 0.08 & 0.12 & 0.41 & 1.03 \end{bmatrix}$
8	$\begin{Bmatrix} \text{Inf} \\ 0.4 \\ 6.4 \\ 31.8 \end{Bmatrix}$	$\begin{bmatrix} 0.20 & 0.02 & 0.01 \\ 0.21 & 0.02 & 0.07 \\ 1.58 & -0.75 & 0.24 \\ 2.60 & 0.11 & -0.16 \end{bmatrix}$	$\begin{Bmatrix} 2.33 \\ 0.62 \\ 0.05 \end{Bmatrix}$	3	$\begin{bmatrix} 0.01 & 0.02 & 0.11 & 0.16 \\ 0.02 & 0.05 & 0.24 & 0.20 \\ 0.12 & 0.25 & 1.55 & 1.13 \\ 0.16 & 0.20 & 1.10 & 2.10 \end{bmatrix}$
15	$\begin{Bmatrix} \text{Inf} \\ 0.5 \\ 200 \\ \text{Inf} \end{Bmatrix}$	$\begin{bmatrix} 0.19 & 0.03 \\ 0.18 & 0.11 \\ 1.44 & -0.03 \\ 2.51 & 0.01 \end{bmatrix}$	$\begin{Bmatrix} 1.38 \\ 0.32 \end{Bmatrix}$	2	$\begin{bmatrix} 0.02 & 0.02 & 0.11 & 0.19 \\ 0.02 & 0.06 & 0.34 & 0.27 \\ 0.11 & 0.34 & 2.30 & 1.50 \\ 0.19 & 0.27 & 1.50 & 2.60 \end{bmatrix}$

Note: The number of training data N equals 100 in this RC slab example. In corresponding, there are 100 components in $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N] \in \mathbb{R}^{M \times N}$. Only the first component $\mathbf{z}_1$ is listed here due to the space limit. All 100 variable $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{100}$ are initialized to four-dimensional vectors and are optimized to two-dimensional vectors upon convergence.

Upon convergence, $\boldsymbol{\Psi}$ converges to a full matrix where each diagonal entry represents the variance of the prediction errors of each dimension, and the off-diagonal entries reveal the correlations between different dimensions. Specifically, the correlation coefficient between the dimensions i and j can be computed as $\rho_{ij} = \frac{\boldsymbol{\Psi}_{ij}}{\sqrt{\boldsymbol{\Psi}_{ii}}\sqrt{\boldsymbol{\Psi}_{jj}}}$ (i, j =1, 2, 3, and 4). With the estimated $\boldsymbol{\mathcal{W}}^{(15)}$ and $\boldsymbol{\Psi}^{(15)}$, the MAP values of the latent variables corresponding to the entire dataset are calculated using Eq. (15), and then the dataset $\mathbf{D}'$ are re-generated according to Eq. (1). The Euclidean norm of the error vector is then calculated as the damage indicator, that is, $||\mathbf{e}||_2 = ||\mathbf{D} - \mathbf{D}'||_2$. The damage indicators of the entire dataset are then plotted in Figure 3.4. The values of the damage indicators of the test data are comparable with those of the training data, reflecting that the RC slab is in a health condition. This result is consistent with the reality as no damage was introduced to the slab during the monitoring period.

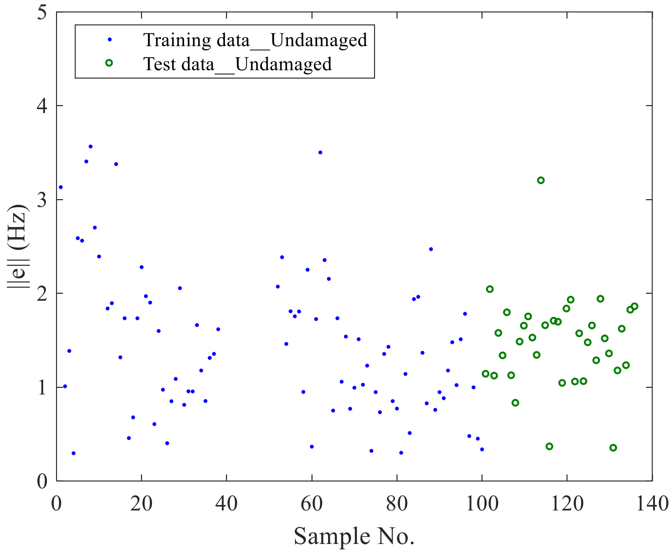


Figure 3.4 Damage indicator in the RC slab example

3.4.2 The Steel Frame

The two-story steel frame in Bao et al. (2012) shown in Figure 3.5 is utilized to demonstrate the effectiveness of the proposed method for damage detection. The frame is 0.5 m wide and 1.0 m high with two equal stories. The young's modulus of the material is 2.0×10^{11} N/m^2, and the mass density is 7.67×10^3 kg/m^3. The cross-sections of the column and beam are 50.0×4.4 mm^2 and 50.0×8.8 mm^2, respectively.

The frame in the health condition was exposed to sunlight from morning to afternoon. The hammer test was conducted on the frame five times every 20 minutes. Fourteen accelerometers were installed on the frame to record acceleration responses with a sampling frequency of 2,048 Hz. The detailed sensor locations are depicted in Figure 3.6. The modal parameters were extracted using the Rational Fraction Polynomial method. The temperature was simultaneously measured by eight thermocouples (T1-T8) shown in Figure 3.6. A total of 140 sets of modal data were collected throughout the day.

Figure 3.5 Experimental steel frame

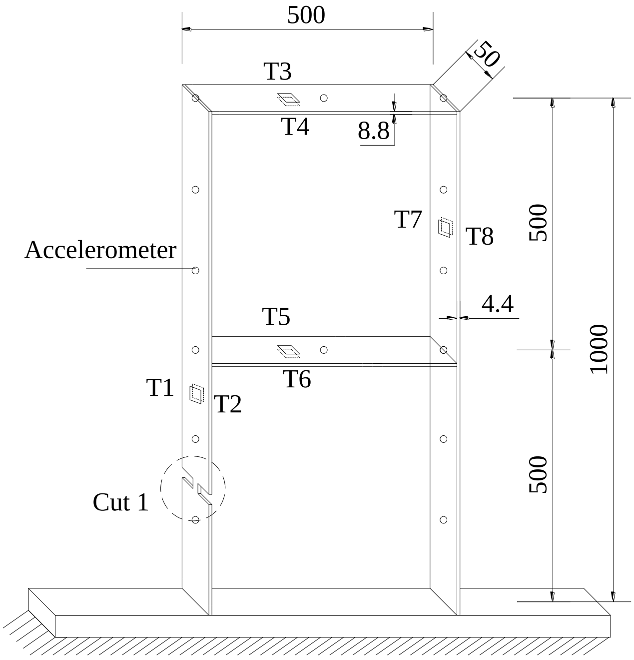


Figure 3.6 Sensor layout and damage location of the steel frame

The damage was then introduced to the steel frame. A saw cut with a length of 20 mm and a width of 4.4 mm was designed at the left column of the first story, as shown in Figure 3.6. The depth of the cut sequentially increased to 5, 10 and 15 mm, resulting in three DSs. In each DS, the frame was exposed to sunlight in one day. The hammer test was conducted on the frame in each DS, similar to that in the health condition. 140 sets of modal data were obtained in each DS.

The frequency variations in DS0–DS3 under varying environmental conditions are plotted in Figure 3.7. The averages of 140 sets of frequencies of the steel frame in undamaged and damaged states are listed in Table 3.2. The intact condition is denoted as DS0. DS1 resulted in an average frequency reduction of 0.26%, and DS2 and DS3 led to 0.64% and 0.79% frequency change on average, respectively. The entire dataset **D** includes 560 sets of the first six frequencies, of which the first 140 sets in health condition are used as the training data for parameters estimation, and the rest 420 sets are the testing data. The temperature variations were also recorded in this example as plotted in Figure 3.8, but they are assumed to be unknown, and only the frequencies are used.

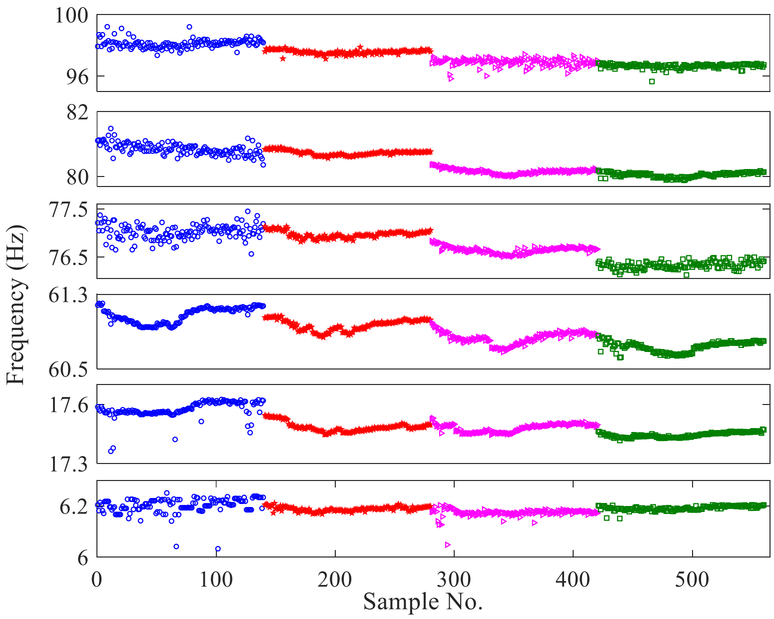


Figure 3.7 Variations of structural first six frequencies (blue circles: DS0; red pentagrams: DS1; carmine triangles: DS2; and green squares: DS3)

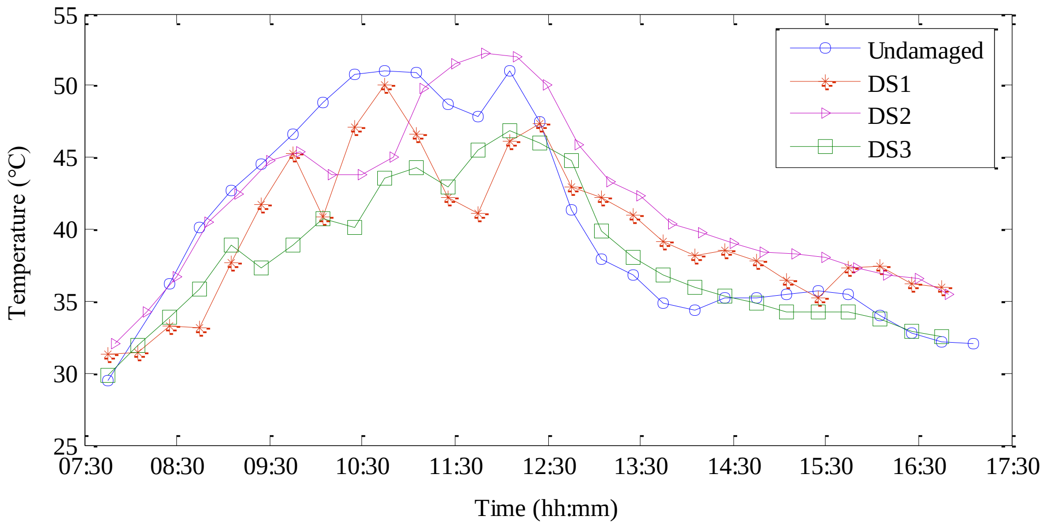


Figure 3.8 Temperature variations in each DS (Hou et al., 2020)

Table 3.2 Average of frequencies in undamaged and damaged states (unit: Hz)

Mode	DS0	DS1	DS2	DS3
1	6.20	6.19(−0.19)	6.17(−0.43)	6.19(−0.13)
2	17.57	17.49(−0.49)	17.48(−0.52)	17.45(−0.72)
3	61.07	60.97(−0.17)	60.83(−0.39)	60.74(−0.55)
4	77.01	76.97(−0.05)	76.65(−0.47)	76.32(−0.89)
5	80.83	80.73(−0.13)	80.16(−0.83)	80.05(−0.97)
6	98.09	97.58(−0.52)	96.91(−1.20)	96.66(−1.46)
Average (%)		(−0.26)	(−0.64)	(−0.79)

Note: Values in the parentheses are the frequency change ratios (%) after damage relative to the undamaged state.

The proposed method is then applied. The parameters $\mathcal{W}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$ are similarly initialized as described in Section 3.5.1. The method took nine iteration steps before convergence. The variations of the latent variable and parameters are listed in Table 3.3. Their values in each step are also substituted into the Hessian matrix for the positive definiteness examination. In the optimization process, some entries in $\boldsymbol{\alpha}$ get larger and approach to infinity, enforcing the corresponding columns in $\mathcal{W}$ to zero. Upon convergence, five columns in $\mathcal{W}$ become zero, and only one column remains nonzero, representing that only one environmental factor is strongly related to the frequency variations.

Table 3.3 Variations of the latent variable and parameters in the steel frame example

No.	$\boldsymbol{\alpha}$	$\boldsymbol{w}$	$\mathbf{z}_1$	M	$\boldsymbol{\Psi}$
0		$\begin{bmatrix} 0.43 & 0.67 & 0.54 & 0.18 & 0.08 & 0.53 \\ 0.24 & 0.55 & 0.24 & 0.18 & 0.80 & 0.19 \\ 0.71 & 0.73 & 0.06 & 0.53 & 0.66 & 0.75 \\ 0.41 & 0.22 & 0.89 & 0.85 & 0.99 & 0.75 \\ 0.80 & 0.61 & 0.32 & 0.82 & 0.14 & 0.30 \\ 0.57 & 0.32 & 0.46 & 0.81 & 0.67 & 0.68 \end{bmatrix}$		6	$10^{-3}\times\begin{bmatrix} 4 & 5 & 19 & 24 & 25 & 30 \\ 5 & 31 & 54 & 68 & 71 & 86 \\ 19 & 54 & 373 & 235 & 247 & 300 \\ 24 & 68 & 235 & 593 & 311 & 378 \\ 25 & 71 & 247 & 311 & 653 & 396 \\ 30 & 86 & 300 & 378 & 396 & 962 \end{bmatrix}$
1	$\begin{Bmatrix} 3.2 \\ 3.3 \\ 4.1 \\ 2.5 \\ 2.4 \\ 3.0 \end{Bmatrix}$	$\begin{bmatrix} 0.11 & 0.67 & 0.54 & 0.18 & 0.08 & 0.53 \\ 0.03 & 0.55 & 0.24 & 0.18 & 0.80 & 0.19 \\ 0.02 & 0.73 & 0.06 & 0.53 & 0.66 & 0.75 \\ -0.01 & 0.22 & 0.89 & 0.85 & 0.99 & 0.75 \\ -0.01 & -0.61 & -0.32 & 0.82 & -0.14 & -0.30 \\ -0.02 & -0.32 & -0.46 & 0.81 & -0.67 & 0.68 \end{bmatrix}$	$\begin{Bmatrix} 0.11 \\ -0.06 \\ -0.06 \\ 0.01 \\ 0.13 \\ 0.02 \end{Bmatrix}$	6	$10^{-4}\times\begin{bmatrix} 6 & 2 & 5 & 5 & -7 & 3 \\ 2 & 15 & 16 & 7 & -22 & 8 \\ 4 & 16 & 66 & 49 & -41 & 94 \\ 6 & 8 & 49 & 288 & 17 & 108 \\ -2 & -18 & -37 & 18 & 294 & 19 \\ 8 & 11 & 94 & 108 & 19 & 820 \end{bmatrix}$
5	$\begin{Bmatrix} \mathrm{Inf} \\ 176 \\ \mathrm{Inf} \\ \mathrm{Inf} \\ 450 \\ 641 \end{Bmatrix}$	$10^{-4}\times\begin{bmatrix} 185 & 113 & 93 \\ 5 & -11 & -10 \\ -18 & -6 & -5 \\ -1 & -1 & -1 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$	$\begin{Bmatrix} -0.32 \\ 0.02 \\ 0.08 \end{Bmatrix}$	3	$10^{-4}\times\begin{bmatrix} 3 & 2 & 5 & 4 & -8 & 2 \\ 2 & 16 & 17 & 5 & -25 & 7 \\ 5 & 17 & 67 & 49 & -43 & 93 \\ 4 & 5 & 49 & 293 & 16 & 118 \\ -8 & -25 & -43 & 16 & 302 & 20 \\ 3 & 7 & 93 & 118 & 20 & 823 \end{bmatrix}$
9	$\begin{Bmatrix} \mathrm{Inf} \\ 798 \\ \mathrm{Inf} \\ \mathrm{Inf} \\ \mathrm{Inf} \\ \mathrm{Inf} \end{Bmatrix}$	$10^{-4}\times\begin{bmatrix} 270 \\ 18 \\ -29 \\ -1 \\ 4 \\ 2 \end{bmatrix}$	$\{-0.18\}$	1	$10^{-4}\times\begin{bmatrix} 2 & 3 & 6 & 5 & -8 & 3 \\ 3 & 17 & 18 & 6 & -25 & 7 \\ 6 & 18 & 68 & 49 & -43 & 93 \\ 5 & 6 & 49 & 299 & 16 & 120 \\ -8 & -25 & -43 & 16 & 302 & 21 \\ 3 & 7 & 93 & 120 & 21 & 835 \end{bmatrix}$

The estimated parameters are then used to calculate the latent variables corresponding to the entire dataset, and then re-generate the dataset. The damage indicators are then calculated based on the error vector and plotted in Figure 3.9. The damage-induced deviations in the damage indicators can be observed from the figure. When the damage severity is slight (i.e. DS1), the deviation is relatively small but still noticeable. When the damage gets severe (i.e. DS2 and DS3), the deviation gets larger significantly. A normal distribution test on the reconstruction errors of each DS is conducted. The chi-squared goodness-of-fit test method (Press et al., 1992) is used. The 140 points in each DS are ranked and divided into 15 nonoverlapped groups (Press et al., 1992). Figure 3.10 shows the probability histogram in each DS. The test results are shown in Table 3.4 when the level of significance $\alpha = 0.05$. Note that the group with observed numbers less than 5 is merged with others, thereby the critical bounds are different for each DS.

Table 4 shows that the chi-square statistic is less than the critical bound, indicating that the normal distribution hypothesis is accepted. Based on the Gaussian distribution, the threshold can be calculated at a significance level of $\alpha = 0.05$. The threshold is 0.89 as shown in Figure 3.9. The probabilities of three DSs in the damaged state can be then calculated. Specifically, the probability of DS1 being damaged is around 50%, and the probabilities of DS2 and DS3 being damaged are larger than 95%.

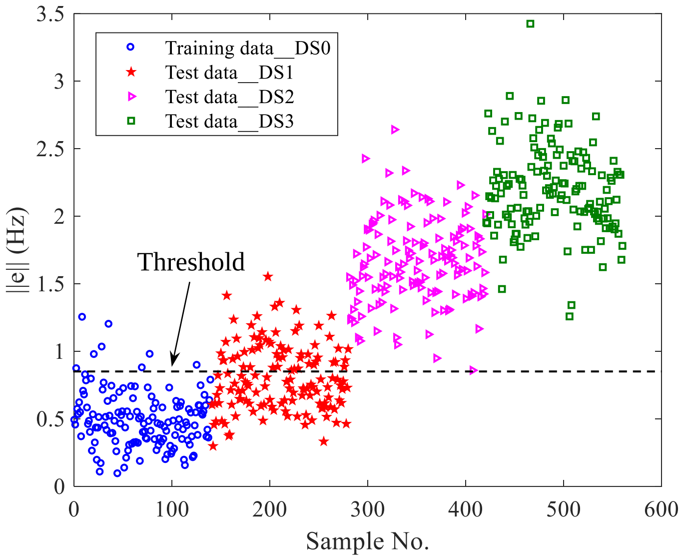
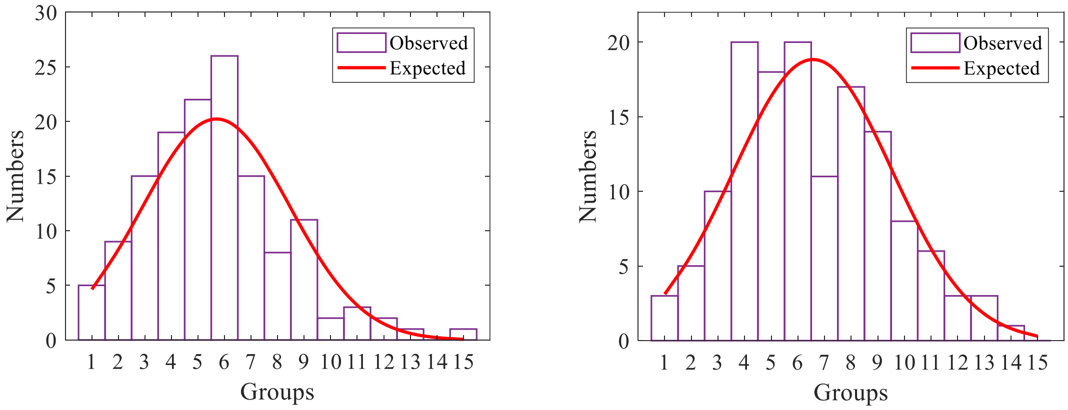
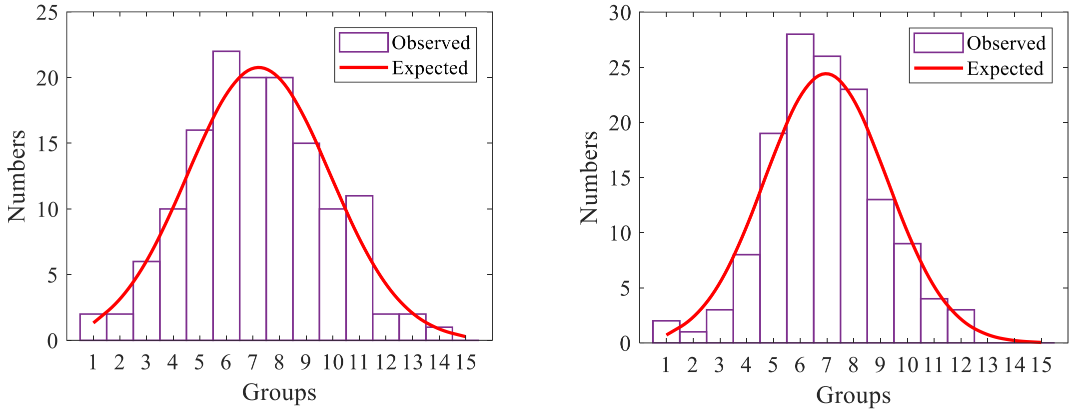


Figure 3.9 Damage indicator in each DS



(a) DS0

(b) DS1



(c) DS2 (d) DS3

Figure 3.10 Chi-square goodness-of-fit test for four DSs

Table 3.4 Results of goodness-of-fit test for normality (α = 0.05)

DS No.	**DS0**	**DS1**	**DS2**	**DS3**
Chi-square value	6.21	7.84	3.29	3.36
Critical bound	18.31	19.68	18.31	16.92

Structural higher frequency modes are generally difficult to measure in practice. In this connection, we further investigate the performance of the proposed method when higher modes are not available. Only the first five frequencies are used as the dataset. The proposed method is then applied, and the calculated damage indicators are plotted in Figure 3.11. As the figure turns out, the damage indicators in DS1 have no noticeable deviation compared with those in DS0 when the sixth mode is removed. This result reflects that the sixth mode is more sensitive to the minor damage.

When only the first five modes are used for analysis, a false-negative damage diagnosis happens, that is, DS1 is falsely assessed as undamaged. In this situation, the measurement data in DS1 may be falsely used as the training data. To investigate the performance of the proposed method when the damage data are falsely used for model training, the first five frequencies in DS0 and DS1 are both used as the training data. The estimated parameters and damage indicators are shown in Table 3.5 and Figure

3.12. The nonzero columns in $\mathcal{W}$ converge to one after 11 iterations, and the number of factors is also one. Significant deviations of the damage indicators are observed in the test data. The result implies that although the minor damage was falsely recognized as undamaged and incorporated into the training data, the proposed method can still detect damage correctly at the later stage as the damage gets severe. The results verify the effectiveness of the proposed method for structural damage detection under varying temperature conditions.

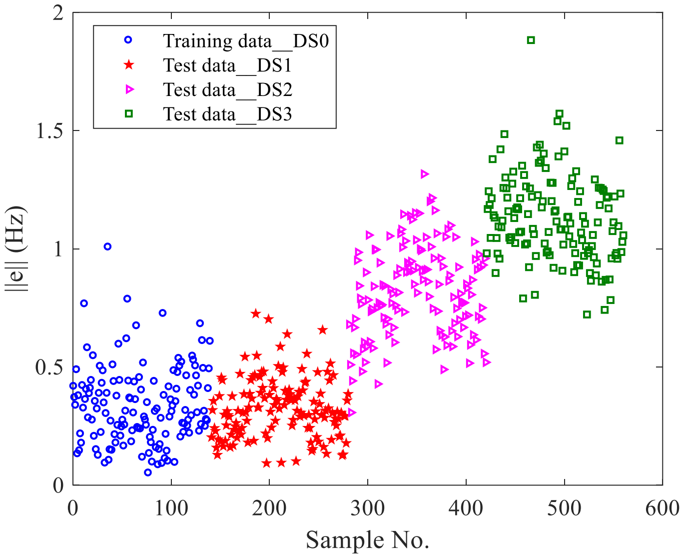


Figure 3.11 Damage indicator using the first five frequencies as the dataset

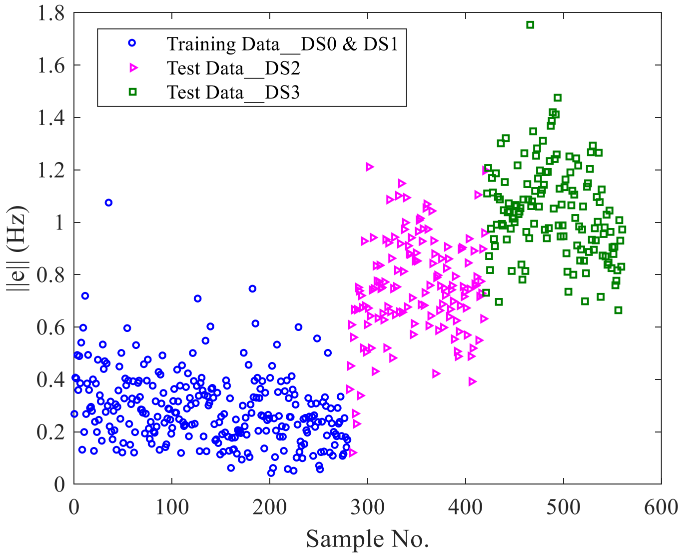


Figure 3.12 Damage indicator using both DS0 & DS1 as training data

Table 3.5 Variations of variables and parameters using DS0 & DS1 as training data

No.	$\boldsymbol{\alpha}$	$\boldsymbol{W}$	$\mathbf{z}_1$	M	$\boldsymbol{\Psi}$
0		$\begin{bmatrix} 0.37 & 0.03 & 0.88 & 0.32 & 0.38 \\ 0.58 & 0.01 & 0.44 & 0.55 & 0.86 \\ 0.15 & 0.80 & 0.09 & 0.48 & 0.69 \\ 0.10 & 0.68 & 0.63 & 0.42 & 0.51 \\ 0.75 & 0.77 & 0.11 & 0.80 & 0.42 \end{bmatrix}$		5	$10^{-3}\times\begin{bmatrix} 4 & 5 & 19 & 24 & 25 \\ 5 & 31 & 54 & 68 & 71 \\ 19 & 54 & 373 & 235 & 247 \\ 24 & 68 & 235 & 593 & 311 \\ 25 & 71 & 247 & 311 & 653 \end{bmatrix}$
11	$\begin{Bmatrix} \text{Inf} \\ 1561 \\ \text{Inf} \\ \text{Inf} \\ \text{Inf} \end{Bmatrix}$	$10^{-4}\times\begin{bmatrix} 177 \\ -23 \\ -6 \\ -2 \\ 2 \end{bmatrix}$	$\{-0.11\}$	1	$10^{-4}\times\begin{bmatrix} 2 & 5 & 7 & 6 & 1 \\ 5 & 30 & 36 & 18 & 18 \\ 7 & 36 & 77 & 51 & 23 \\ 6 & 18 & 51 & 181 & 37 \\ 1 & 18 & 23 & 37 & 204 \end{bmatrix}$

3.5 Summary

A sparse Bayesian FA-based method is developed in this chapter for structural damage detection under changing environmental conditions without knowing the environment data. By adopting the ARD prior in the SBL framework, the number of underlying environment factors is automatically identified. The uncertainties are evaluated upon convergence. The residual error matrix is then calculated as the damage indicator according to the generative FA model.

The proposed method is applied to two experimental examples under varying environmental conditions. The RC slab example demonstrates that the sparse Bayesian FA-based method can automatically identify the number of environmental factors that influence the structural vibration properties, without measuring the corresponding varying factors. The steel frame example verifies that the method can successfully detect the existence of damage under varying environmental conditions.

CHAPTER 4

STRUCTURAL DAMAGE DETECTION CONSIDERING NONLINEAR ENVIRONMENT EFFECTS BY PROBABILISTIC KERNELIZED METHOD

4.1 Introduction

The Bayesian FA proposed in Chapter 3 is a linear model, which is limited to the cases where the environmental influence is linear or weakly nonlinear. In some applications, nonlinearities may arise since the multiple materials in bridges may have varied temperature-sensitive mechanical properties (Peeters and De Roeck, 2001). The bridge’s boundary conditions may also change due to the expansion and contraction of joints, leading to nonlinear correlations (Peeters et al., 2001). For the nonlinear applications, the kernel PCA has been developed, which maps the measurement data onto a high-dimensional feature space wherein a hyperplane could separate samples (Oh et al., 2019; Reynders et al., 2014). The dimension reduction is then performed on the high-dimensional feature space. However, the non-probabilistic kernel PCA suffers from limitations. First, the covariance matrix needs to be computed as an intermediate variable, followed by the eigenvalue decomposition. Second, the optimal kernel parameters need to be determined in advance, which requires high computational costs for high-dimensional problems, given that the kernel PCA involves the data projection to a high-dimensional space.

This chapter develops an improved probabilistic kernelized model to eliminate the nonlinear environmental influence. The Gaussian kernel is employed to introduce nonlinearity. The unknown kernel parameters and the latent variables are estimated automatically in the Bayesian probabilistic framework. The proposed method is applied to the benchmark Z24 bridge for damage detection. Comparisons between the proposed method and non-probabilistic kernel PCA are presented.

4.2 Background of Z24 bridge

The Z24 bridge is an example wherein the environment factor has a nonlinear effect on the vibration properties. It provides the rare long-term monitoring data of a full-scale bridge under the undamaged and realistic damaged states. The dataset of the bridge has served as the benchmark in numerous studies, including vibration-based system identification (Peeters and De Roeck, 2001), damage detection (Kullaa, 2003), and data normalization to remove environmental effects on structural vibration properties (Peeters et al., 2001; Kullaa, 2011).

4.2.1 Description of the Bridge

The Z24 bridge is a post-tensioned concrete box-girder bridge consisting of a 30 m long main span and two 14 m side spans, as depicted in Figure 4.1. The bridge is located in Switzerland and was monitored from 11 November 1997 to 11 September 1998 (Peeters and De Roeck, 2001). A total of 49 sensors were installed on the bridge to record the variations of environmental conditions, including the temperature, wind, humidity, etc. Another 16 accelerometers were installed to record acceleration responses. In the later monitoring period, progressive damages were artificially introduced to the bridge in a controlled manner. The first damage was introduced by lowing the pier on 10 August 1998, followed by several other sequential DSs, including foundation inclination, concrete spalling, anchor head failure, and tendon rupture. More specific information about the experiment can be found in Peeters and De Roeck (2001).

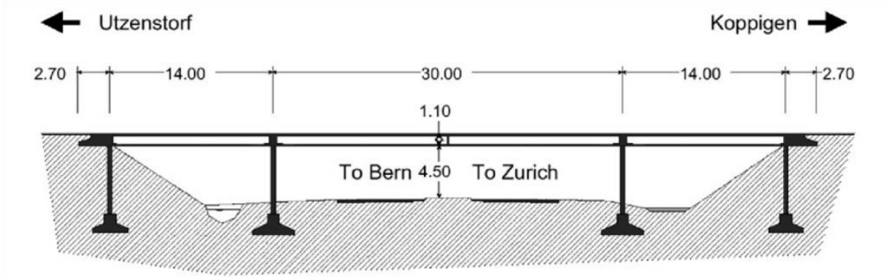
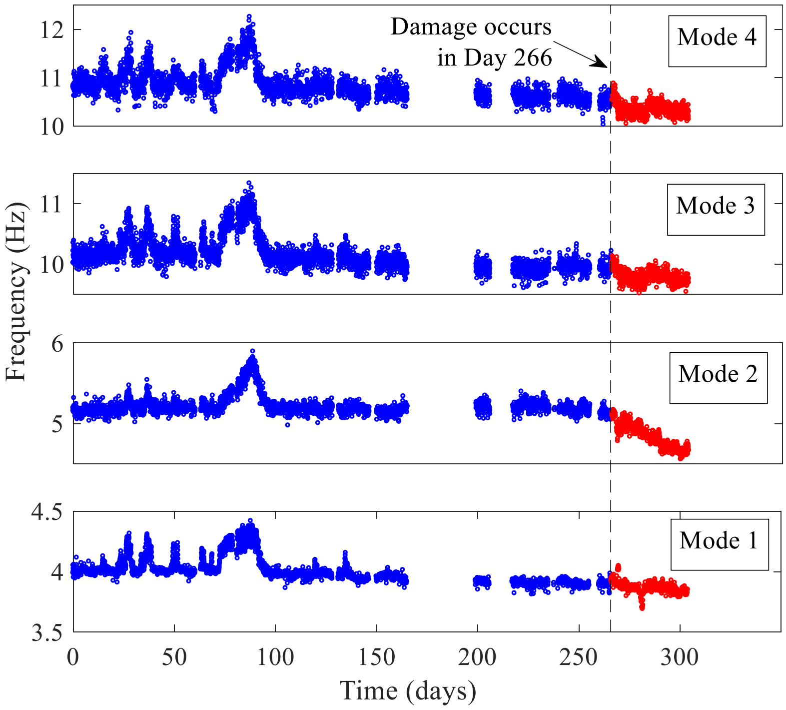


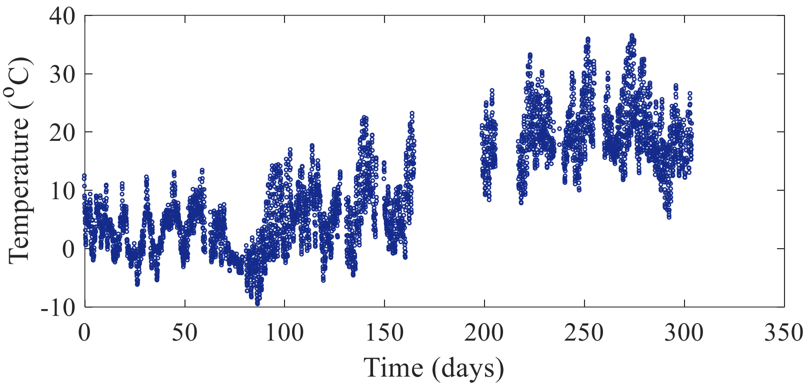
Figure 4.1 Side view of the Z24 bridge (unit: meter) (Peeters and De Roeck, 2001)

The bridge's responses under the ambient excitations were recorded, and the stochastic subspace identification method was used to extract the modal parameters (Peeters and De Roeck, 2001). 5624 sets of the first four frequencies were obtained during the entire 304 monitoring days. The variations of the first four frequencies of the bridge and the temperature over time are plotted in Figure 4.2. The first damage was artificially introduced around day 266, which corresponds to No. 4789 data. The relation between the temperature and natural frequencies is plotted in Figure 4.3.



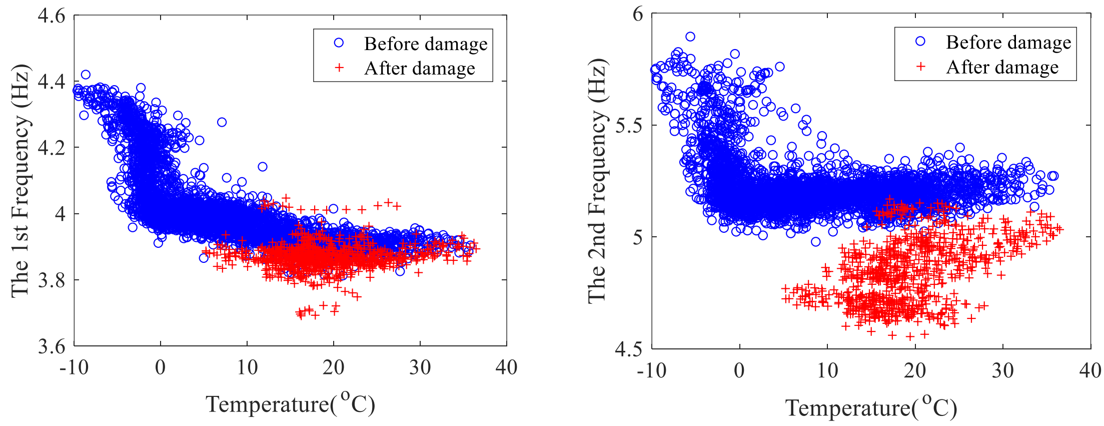
(a) Variation of frequencies

(Blue dots: data in the health state; red dots: data in the damaged state)



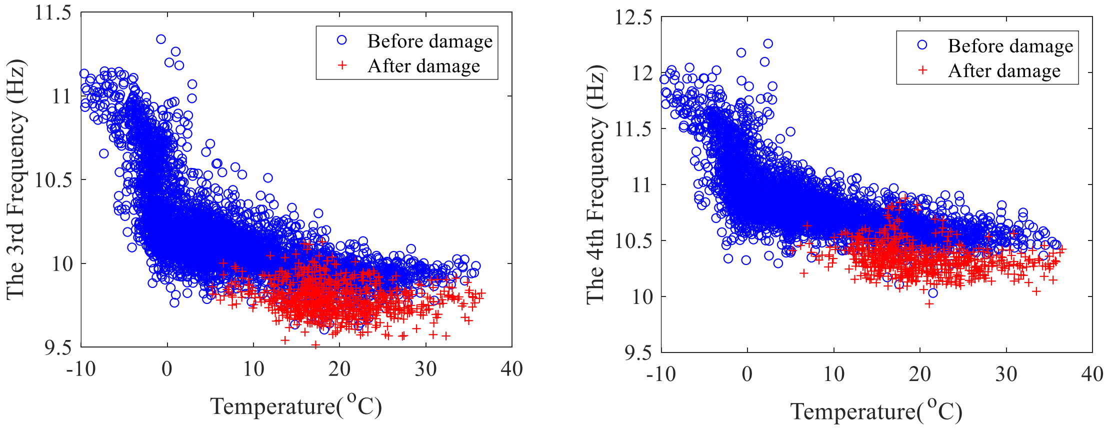
(b) Variation of the temperature

Figure 4.2 Frequency and temperature variations over time



(a) First frequency

(b) Second frequency



(c) Third frequency

(d) Fourth frequency

Figure 4.3 The relation between the frequency and temperature

Figure 4.3 reflects that the natural frequencies have a nonlinear relation with the temperature. The nonlinearity is attributed to the asphalt layer on the bridge's surface, which froze on cold days (lower than 0 ^{o}C) and significantly increased the stiffness of the structure (Peeters and De Roeck, 2001).

4.2.2 The Application of the Linear Model

The linear Bayesian FA method proposed in Chapter 3 is first applied to the Z24 bridge for damage detection. As mentioned, the first 4788 data points are in the health state, and the rest 836 are in the damaged state. To investigate the influence of training data, the first 3000 data points are set as the training data, which accounts around 50% of the entire data set, and the rest 1788 data points in the health state and 836 data points in the damaged state are set as the validation data and test data, respectively, as shown in Figure 4.4.

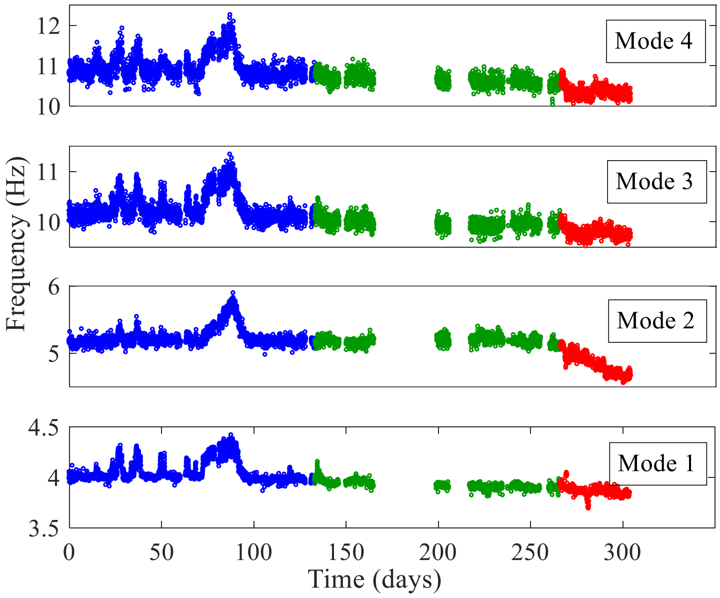


Figure 4.4 Dataset division

(blue dots: training data; green dots: validation data; red dots: test data)

The parameters in the Bayesian FA model are initialized as follows: M is set to the dimension of the original measurement data, that is $M = N_m = 4$; The factor loading matrix $\boldsymbol{\mathcal{W}}^{(0)}$ is randomly initialized from the uniform distribution $\mu(0,1)$; $\boldsymbol{\Psi}$ is similarly initialized as described in Chapter 3, that is, the entries are initialized by assuming that the coefficient of variation of each natural frequency is 1% and the correlation coefficient between modes is 0.5; and the convergence criterion is set to $\|\boldsymbol{\Psi}^{(j)} - \boldsymbol{\Psi}^{(j-1)}\| / \|\boldsymbol{\Psi}^{(j)}\| \leq 1\times10^{-5}$.

The EM algorithm is applied, and the convergence is achieved after 23 iterations. The variations of the parameters are listed in Table 4.1. Upon convergence, two columns in $\boldsymbol{\mathcal{W}}$ are penalized to zero, and the rest two keep nonzero.

Table 4.1 Variation of the unknown variable and parameters in the iteration process

Iteration No.	$\boldsymbol{\alpha}$	$\boldsymbol{\mathcal{W}}$	M	$\boldsymbol{\Psi}$
0		$\begin{bmatrix} 0.47 & 0.43 & 0.88 & 0.59 \\ 0.76 & 0.32 & 0.81 & 0.24 \\ 0.53 & 0.63 & 0.37 & 0.15 \\ 0.85 & 0.16 & 0.53 & 0.57 \end{bmatrix}$	4	$10^{-2}\times\begin{bmatrix} 0.16 & 0.11 & 0.21 & 0.22 \\ 0.11 & 0.27 & 0.27 & 0.29 \\ 0.21 & 0.27 & 1.05 & 0.56 \\ 0.22 & 0.29 & 0.56 & 1.20 \end{bmatrix}$
10	$\begin{Bmatrix} 2.21 \\ 5.58 \\ 2.15 \\ 5.28 \end{Bmatrix}$	$10^{-2}\times\begin{bmatrix} 0.21 & 0.12 & 0.18 & 0.11 \\ 0.25 & 0.14 & 0.22 & 0.13 \\ 0.53 & 0.31 & 0.45 & 0.28 \\ 0.60 & 0.31 & 0.49 & 0.33 \end{bmatrix}$	4	$10^{-2}\times\begin{bmatrix} 0.08 & 0.06 & 0.19 & 0.42 \\ 0.06 & 0.08 & 0.18 & 0.30 \\ 0.19 & 0.17 & 0.60 & 1.02 \\ 0.42 & 0.29 & 1.00 & 3.15 \end{bmatrix}$
23	$\begin{Bmatrix} 32 \\ \text{Inf} \\ 128 \\ \text{Inf} \end{Bmatrix}$	$\begin{bmatrix} 0.08 & 0.05 \\ 0.09 & 0.06 \\ 0.19 & 0.15 \\ 0.27 & 0.04 \end{bmatrix}$	2	$10^{-2}\times\begin{bmatrix} 0.03 & 0.04 & 0.07 & 0.12 \\ 0.03 & 0.06 & 0.10 & 0.17 \\ 0.07 & 0.10 & 0.26 & 0.31 \\ 0.12 & 0.17 & 0.31 & 0.70 \end{bmatrix}$

With the estimated $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\Psi}$, the latent variables $\mathbf{z}_n$ corresponding to the training data, validation data and test data are estimated. The measurement data are then regenerated from the trained Bayesian FA model. The squared prediction error (SPE) statistic is then defined as

$$\mathrm{SPE}_n = (\boldsymbol{e}_n)^{\mathrm{T}}\boldsymbol{e}_n \tag{4.1}$$

where $\boldsymbol{e}_n$ denotes the regeneration error, which is the discrepancy between the model predictions and measurement data. The SPE statistic can be used as the index for novelty detection. Denoting the mean and variance of the SPEs of training data as m and v, the SPE statistic can be well approximated by a weighted χ^2 distribution (Box, 1954a, b; Nomikos and MacGregor, 1995), and the confidence limit is

$$\mathrm{SPE}_\alpha = g \cdot \chi^2_{h,\alpha} \tag{4.2}$$

where $g = v/2m$; $h = 2m^2/v$; and $\chi^2_{h,\alpha}$ is the critical value of the χ^2 distribution with h degree of freedom at a significance level of α. The control limit is a threshold to judge whether the process is under the normal condition. The SPE statistic and the corresponding threshold with $\alpha = 0.01$ are calculated and plotted in Figure 4.5.

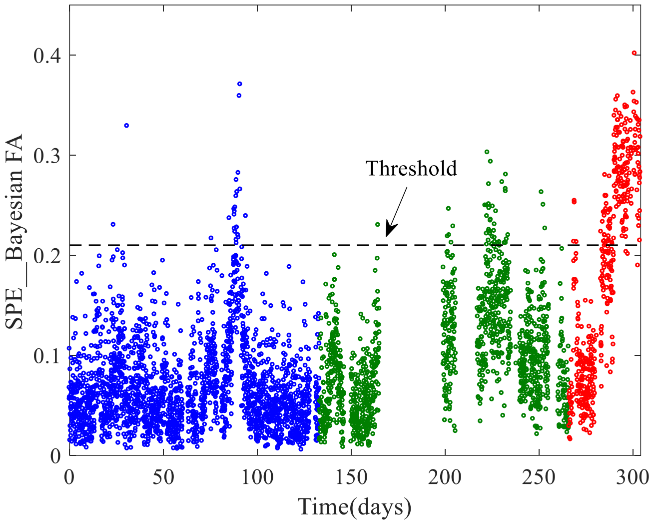


Figure 4.5 SPE statistic of Bayesian FA

As Figure 4.5 turns out, the SPEs of the test data in the damaged condition (red points) are not significantly larger than those of the training data and validation data, making it difficult to timely detect the damage. An abnormality occurs in the health data around days 80~90, where the SPEs are comparable with those in the damaged state. As mentioned previously, the temperature in these days went below zero and exerted a bilinear influence on the structural frequencies. The nonlinear relation cannot be well explained by the linear Bayesian FA. As a result, a false-positive damage indication occurs when the temperature goes below zero, although the structure remains intact. In addition, the SPEs of the validation data also exceed the threshold. The reason is that the training data (the first 3000 data points) only cover part of the environmental variations. The temperature of the validation data goes beyond the range of the training data, as shown in Figure 4.2(b), resulting in the increased SPEs in the validation data. Consequently, the linear Bayesian FA cannot timely and correctly detect the damage in the Z24 bridge example.

4.3 Nonlinear Environmental Model

This section studies the nonlinear environmental effects using the probabilistic kernelized model. First, the nonlinear probabilistic PCA, also named the Gaussian process latent variable model (GPLVM) (Lawrence, 2005), is discussed. We will then explain how the damage can be detected. Finally, the proposed method is compared with the non-probabilistic kernelized model.

4.3.1 GPLVM

GPLVM provides a nonlinear probabilistic interpretation of standard PCA by employing the Gaussian process. The model between the data sample and latent variable is defined as (Lawrence, 2005)

$$\mathbf{D}_n = \boldsymbol{f}(\mathbf{z}_n) + \boldsymbol{\varepsilon}_n \tag{4.3}$$

where $\boldsymbol{f}$ is the nonlinear mapping function; and $\mathbf{z}_n$ and $\boldsymbol{\varepsilon}_n$ are the same as the definition of the Bayesian FA model in Eq. (3.1). Different from the Bayesian FA, GPLVM marginalizes the parameters and optimizes with respect to the latent variable. Specifically, the prior is defined over $\boldsymbol{\mathcal{W}}$, and the likelihood function is marginalized with respect to $\boldsymbol{\mathcal{W}}$. For simplicity, a Gaussian prior conjugate to the likelihood function is adopted as $p(\boldsymbol{\mathcal{W}}) = \prod_i^D \mathcal{N}(\boldsymbol{w}_i|\mathbf{0},\mathbf{I})$, where $\boldsymbol{w}_i$ is the ith row of $\boldsymbol{\mathcal{W}}$. Then the marginalization with respect to $\boldsymbol{\mathcal{W}}$ is

$$p\left(\mathbf{D}_{d,:}|\mathbf{z},\beta\right) = \int p\left(\mathbf{D}_{d,:}|\mathbf{w}_i,\beta,\mathbf{z}\right)p(\boldsymbol{w}_i)d\boldsymbol{w}_i = \mathcal{N}\left(\mathbf{D}_{d,:}|\mathbf{0},\mathbf{z}^{\mathrm{T}}\mathbf{z} + \beta^{-1}\mathbf{I}\right) \tag{4.4}$$

where $\mathbf{D}_{d,:}$ represents the *d*th row of $\mathbf{D} \in \mathbb{R}^{N_m \times N}$. In corresponding, the objective function of the likelihood function in the logarithm form is obtained as

$$L = -\frac{N_m N}{2}\ln 2\pi - \frac{D}{2}\ln|\mathbb{K}| - \frac{1}{2}\mathrm{tr}(\mathbb{K}^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{D}) \tag{4.5}$$

where $\mathbb{K} = \mathbf{z}^{\mathrm{T}}\mathbf{z} + \beta^{-1}\mathbf{I}$ is the covariance matrix or kernel function. The kernel function here is defined on the latent variable **z**. By replacing the inner product $\mathbb{K}$ with nonlinear kernel functions, GPLVM achieves the nonlinear projection. The advantage of the kernel function is that the nonlinear mapping function is not required to be known explicitly, thereby efficiently reducing the complexity of searching for the high-dimensional feature space. According to the Mercer's theorem, the kernel function must be continuous and symmetric, and the corresponding kernel matrix must be positive semi-definite. The representative kernel functions include the polynomial, sigmoid and radial basis function (RBF) kernel. For example, the RBF kernel is adopted as

$$k\left(\mathbf{z}_i,\mathbf{z}_j\right) = \theta_{\mathrm{rbf}}\exp\left[-\frac{\gamma}{2}\left(\mathbf{z}_i - \mathbf{z}_j\right)^{\mathrm{T}}\left(\mathbf{z}_i - \mathbf{z}_j\right)\right] + \theta_{\mathrm{bias}} + \theta_{\mathrm{white}}\delta_{ij} \tag{4.6}$$

where θ_{rbf} is the process variance and influences the scale of the output functions; γ is the inverse width parameter; θ_{bias} corresponds to the prior variance; θ_{white} is the white noise term; and δ_{ij} is the Kronecker delta function.

After the kernel function has been selected, the next step of GPLVM is to optimize the latent variables $\mathbf{z}_n$ and parameters of the kernel functions, including γ, θ_{rbf}, θ_{bias} and θ_{white}. The likelihood of the latent variables and parameters are estimated by setting the gradient of the log-likelihood in Eq. (4.6) to zero. In particular, the gradient with respect to **z** is computed through the chain rule as

$$\frac{\partial L}{\partial \mathbf{z}} = \frac{\partial L}{\partial \mathbb{K}} \times \frac{\partial \mathbb{K}}{\partial \mathbf{z}} = 0 \tag{4.7}$$

where $\frac{\partial L}{\partial \mathbb{K}} = \mathbb{K}^{-1}\mathbf{D}\mathbf{D}^{\mathrm{T}}\mathbb{K}^{-1} - N_m\mathbb{K}^{-1}$; and $\frac{\partial \mathbb{K}}{\partial \mathbf{z}}$ can be directly obtained from Eq. (4.6). However, it is not easy to solve Eq. (4.7) as the log-likelihood L has a highly nonlinear relationship with the latent variable and kernel parameters. To solve this problem, the scaled conjugate gradient approach (Lawrence, 2005) is utilized for the gradient-based optimization. Another problem is the computational complexity of GPLVM, which makes it inefficient in dealing with a large number of data samples. Lawrence et al. (2003) developed the informative vector machine as a sparsification mechanism to improve the computational efficiency. The active set, denoted as $\mathbf{D}_I$, is sequentially selected from the overall training dataset and then used to optimize the kernel parameters and latent variables. The informative vector machine reduces the dominant computational cost to $O(d^2 \cdot N)$, where d is the number of data in the active set ($d < N$).

With the parameters of the kernel function and the latent variables of the active data set optimized, the latent variables $\mathbf{z}_{\mathrm{new}}$ of the new measured data samples $\mathbf{D}_{\mathrm{new}}$, including the inactive subset of the training data and the test data, are obtained from the scaled conjugate gradient optimization. Then the likelihood of $\mathbf{D}^*_{\mathrm{new}}$ generated from $\mathbf{z}_{\mathrm{new}}$ can be calculated using the GP (Lawrence, 2005):

$$p(\mathbf{D}^*_{\mathrm{new}}|\mathbf{z}_{\mathrm{new}}) = \mathcal{N}(\mathbf{D}^*_{\mathrm{new}}|\boldsymbol{\mu}_{\mathrm{new}}, \boldsymbol{\sigma}^2_{\mathrm{new}}\mathbf{I}) \tag{4.8}$$

where $\boldsymbol{\mu}_{\mathrm{new}} = \mathbf{D}_I\mathbb{K}_I^{-1}\Bbbk(\mathbf{z}_I, \mathbf{z}_{\mathrm{new}})$, $\mathbb{K}_I^{-1}$ is the kernel matrix calculated from the latent variables of the active set, $\Bbbk(\mathbf{z}_I, \mathbf{z}_{\mathrm{new}})$ is a column vector with elements of kernels between each sample in the active set and the new data sample; and $\boldsymbol{\sigma}^2_{\mathrm{new}} = \Bbbk(\mathbf{z}_{\mathrm{new}}, \mathbf{z}_{\mathrm{new}}) - \Bbbk^{\mathrm{T}}(\mathbf{z}_I, \mathbf{z}_{\mathrm{new}})\mathbb{K}_I^{-1}\Bbbk(\mathbf{z}_I, \mathbf{z}_{\mathrm{new}})$, $\Bbbk(\mathbf{z}_{\mathrm{new}}, \mathbf{z}_{\mathrm{new}})$ is the value of the kernel function between $\mathbf{z}_{\mathrm{new}}$ and itself. As the likelihood of $\mathbf{D}^*_{\mathrm{new}}$ follows the Gaussian distribution, whose mean is the most probable value, $\boldsymbol{\mu}_{\mathrm{new}}$ in Eq. (4.8) can be used as the prediction result, that is, $\mathbf{D}^*_{\mathrm{new}} = \boldsymbol{\mu}_{\mathrm{new}}$.

GPLVM also has a Bayesian interpretation that could automatically determine the latent dimensions by using the ARD prior. An ARD prior is defined on the kernel function of the latent variable $\mathbf{z}$

$$k_{f(\mathrm{ard})}(\mathbf{z}_i, \mathbf{z}_k) = \sigma_{\mathrm{ard}}^2 \exp\left[-\frac{1}{2}\sum_{j=1}^{M}\frac{\left(z_{i,j} - z_{k,j}\right)^2}{l_j^2}\right] \tag{4.9}$$

where $1/l_j^2$ is a weight parameter. Similar to the Bayesian factor analysis, some items of $1/l_j^2$ will approach zero and automatically penalize the surplus dimensions, thereby realize the dimensionality selection. A full Bayesian analysis is conducted to avoid the risk of overfitting, which implements an Ockham's Razor in the model selection process by integrating out all irrelevant variables. Then the posterior PDF of predictions $p(\mathbf{D}_{\mathrm{new}}^*) = \mathcal{N}(\mathbf{D}_{\mathrm{new}}^* | \boldsymbol{\mu}_{\mathrm{new}}, \boldsymbol{\Sigma}_{\mathrm{new}})$ is derived using the variational inference. More detailed are presented in Lawrence (2005).

The reconstruction errors of both inactive training data and test data are then calculated as

$$\boldsymbol{e}_n = \mathbf{D}_{\mathrm{new}} - \mathbf{D}_{\mathrm{new}}^* \tag{4.10}$$

where $\mathbf{D}_{\mathrm{new}}$ is the measurement data, and $\mathbf{D}_{\mathrm{new}}^*$ is the prediction result of GPLVM. The SPE statistic is defined as

$$\mathrm{SPE}_n = (\boldsymbol{e}_n)^{\mathrm{T}}\boldsymbol{e}_n = (\mathbf{D}_{\mathrm{new}} - \boldsymbol{\mu}_{\mathrm{new}})^{\mathrm{T}}(\mathbf{D}_{\mathrm{new}} - \boldsymbol{\mu}_{\mathrm{new}}) + \boldsymbol{\sigma}_{\mathrm{new}}^2 \tag{4.11}$$

where $\boldsymbol{\sigma}_{\mathrm{new}}^2$ should be replaced by $\boldsymbol{\Sigma}_{\mathrm{new}}$ when the Bayesian GPLVM is adopted. As the nonlinear projection is involved in this process, the kernel density estimation (KDE) (Silverman, 2018) is applied here to compute the threshold of the SPE statistic.

4.3.2 Damage Detection Results

Parameters of the kernel functions need to be initialized first. The RBF kernel in Eq. (4.6) is employed. As suggested in Lawrence (2005), θ_{rbf}, γ, θ_{bias} and θ_{white} are initialized as $\theta_{\mathrm{rbf}} = \gamma = 1$ and $\theta_{\mathrm{bias}} = \theta_{\mathrm{white}} = \exp(-1)$. The active set d in the informative vector machine algorithm is set to d =400 in this example. The latent dimension M is set to $M = N_m = 4$ in the beginning.

Using the Bayesian GPLVM (Damianou et al., 2011), two of four parameters $1/l_j^2$ (j=1, 2, 3, 4) are automatically optimized to zero. The rest two nonzero parameters indicate that the latent dimension of the latent space is 2. M is then set to 2 in GPLVM, and kernel parameters are optimized iteratively using the the scaled conjugate gradient approach. Parameters of the kernel function are optimized as $\theta_{\text{rbf}} = 1.5905$, $\gamma = 1.9598$, $\theta_{\text{bias}} = 0.1360$ and $\theta_{\text{white}} = 1.1487\times10^{-4}$ when the first 3000 data are set as the training data. With the estimated parameters, the measurement data are regenerated using the Gaussian process in Eq. (4.8). The SPE statistic is then calculated according to Eq. (4.11), and the corresponding threshold with $\alpha = 0.01$ is computed and plotted in Figure 4.6.

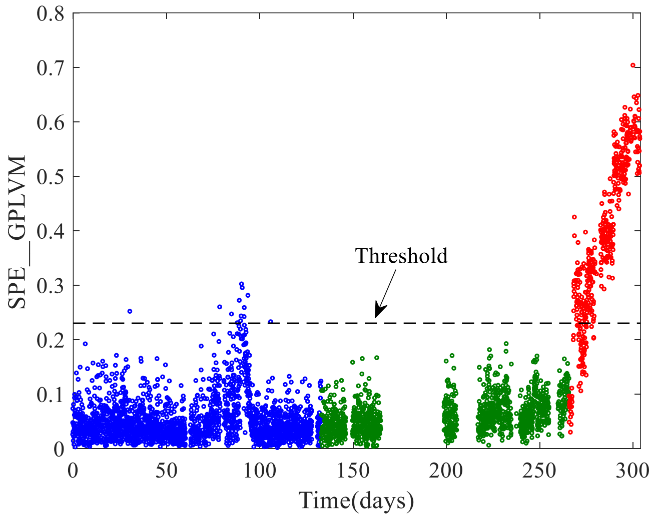


Figure 4.6 SPE statistic of GPLVM

As Figure 4.6 turns out, the SPEs of the damaged data are obviously larger than those of the health data, and the phenomenon of false-positive damage identification is eliminated by using GPLVM. In addition, the validation data are correctly identified to be in the health state, indicating that the proposed method performs well even when the training data have not covered the temperature variation. The entire process, including the automatic dimensionality determination, parameter optimization and SPE statistic computation, takes around 8 minutes only in a personal desktop computer with Intel Core I7-8700 CPU and 20 GB RAM. Consequently, the damage detection method based on GPLVM is efficient, accurate and robust.

4.3.3 Comparison with Non-probabilistic Kernelized Model

A non-probabilistic kernel PCA is also studied for comparison. The data vectors $\mathbf{D}_n$ are mapped onto a high-dimensional feature space wherein a hyperplane could separate samples, and then eigenvalue decomposition is performed on the feature space. The non-probabilistic kernelized model is expressed as

$$\mathbf{\Phi}(\mathbf{D}_n) = \boldsymbol{\mathcal{W}}\mathbf{z}_n + \boldsymbol{\varepsilon}_n \tag{4.12}$$

where $\mathbf{\Phi}$ is a nonlinear function that maps the measurement data onto a high-dimensional feature space $\mathcal{F}$; and $\boldsymbol{\mathcal{W}}$, $\mathbf{z}_n$ and $\boldsymbol{\varepsilon}$ are the same as the definition in Eq. (3.1). The eigenvalue decomposition is performed on the covariance matrix, which is the dot product of $\mathbf{\Phi}(\mathbf{D}_n)$.

The inner product $\mathbf{\Phi}(\mathbf{D}_m)^{\mathrm{T}}\mathbf{\Phi}(\mathbf{D}_n)$ is directly defined in the reproducing kernel Hilbert space by the kernel function $\mathbb{k}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n)$. The Gaussian kernel function, which involves an infinite-dimensional feature space using merely a unique parameter, is adopted here and defined as

$$\mathbb{k}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n) = \exp\left(-\frac{1}{2\sigma^2}(\mathbf{D}_m - \mathbf{D}_n)^{\mathrm{T}}(\mathbf{D}_m - \mathbf{D}_n)\right) \tag{4.13}$$

where $\mathbb{k}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n)$ is the element in the *m*th row and the *n*th column of the kernel matrix $\mathbb{K}_{\mathbf{D}}$; and σ^2 is the variance parameter controling the bandwidth of $\mathbb{K}_{\mathbf{D}}$. Notably, the kernel function in Eq. (4.13) is defined on the measurement data **D** rather than on the latent variable **z** in Eq. (4.6).

After determining the kernel function, the eigenvalue decomposition is conducted on the decentralized kernel matrix $\widetilde{\mathbb{k}}_{\mathbf{D}}$ with the entries $\widetilde{\mathbb{k}}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n)$ as

$$\begin{aligned}\widetilde{\mathbb{k}}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n) = \mathbb{k}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n) - \frac{1}{n}\sum_{r=1}^{N} \mathbb{k}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_r) \\ - \frac{1}{n}\sum_{r=1}^{N} \mathbb{k}_{\mathbf{D}}(\mathbf{D}_r, \mathbf{D}_n) + \frac{1}{n^2}\sum_{r,s=1}^{N} \mathbb{k}_{\mathbf{D}}(\mathbf{D}_r, \mathbf{D}_s)\end{aligned} \tag{4.14}$$

The eigenvectors $\boldsymbol{v}_l(l = 1, \dots, M)$ corresponding to the first M eigenvalues $\lambda^l(l = 1, \dots, M)$ of $\widetilde{\mathbb{k}}_{\mathbf{D}}(\mathbf{D}_m, \mathbf{D}_n)$ are retained for dimensionality reduction. The principal component $\boldsymbol{t}_l$ is then obtained by projecting $\boldsymbol{\Phi}(\mathbf{D})$ onto the eigenvector $\boldsymbol{v}_l$

$$\boldsymbol{t}_l = \langle \boldsymbol{v}_l, \boldsymbol{\Phi}(\mathbf{D}) \rangle = \sum_{i=1}^{N} \alpha_i^l \widetilde{\mathbb{k}}_{\mathbf{D}}(\mathbf{D}_i, \mathbf{D}) \tag{4.15}$$

where α_i^l is the ith element of the vector $\boldsymbol{\alpha}^l$, which is the normalized eigenvector $\boldsymbol{v}_l$ and satisfies that $\|\boldsymbol{\alpha}^l\|^2 = 1/\lambda^l$.

Since the explicit form of $\boldsymbol{\Phi}(\mathbf{D}_n)$ is not defined, it is not clear how to re-project back to the measurement data space. Therefore, the SPE statistic is defined in the reproducing kernel Hilbert space as (Lee et al., 2004)

$$\mathrm{SPE} = \left[\boldsymbol{\Phi}(\mathbf{D}) - \sum_{l=1}^{M} \boldsymbol{t}_l \boldsymbol{v}_l\right]^{\mathrm{T}} \cdot \left[\boldsymbol{\Phi}(\mathbf{D}) - \sum_{l=1}^{M} \boldsymbol{t}_l \boldsymbol{v}_l\right] = \sum_{j=1}^{n} \boldsymbol{t}_j^2 - \sum_{j=1}^{M} \boldsymbol{t}_j^2 \tag{4.16}$$

where n is the number of nonzero eigenvalues of the decentralized kernel matrix $\widetilde{\mathbb{k}}_{\mathbf{D}}$ in Eq. (4.14). However, the SPE statistic may not strictly follow the standard χ^2 distribution due to the nonlinear projection. The KDE is employed to compute the threshold of the SPE statistic. Assume that the SPE statistic follows the unknown PDF $p(x)$:

$$p(x) = \frac{1}{hL}\sum_{n=1}^{N} \varphi\left(\frac{x - \mathrm{SPE}_n}{h}\right) \tag{4.17}$$

where h is the kernel width and $\varphi(\cdot)$ denotes the selected kernel function. The Gaussian kernel function $\varphi(y) = \frac{e^{\frac{-y^2}{2}}}{\sqrt{2\pi}}$ is generally used. The control limit SPE_α is then computed by

$$\int_{-\infty}^{\mathrm{SPE}_\alpha} p(x)dx = 1 - \alpha \tag{4.18}$$

where α is the significance level.

The parameter σ^2 of the kernel function needs to be determined. Widjaja et al. (2012) proposed two different criteria for the optimal value selection of σ^2, namely, the entropy criterion and the eigenvalue criterion. Here the second one is employed. The range of σ^2 is set to be from $\widehat{\sigma^2}/100$ to $100 \cdot \widehat{\sigma^2}$ (Widjaja et al., 2012), where $\widehat{\sigma^2} = N_m \cdot \mathrm{mean}(\mathrm{var}(\mathbf{D}_n))$. The optimal value of σ^2 can be estimated by maximizing the difference between the first eigenvalue and the sum of the remaining eigenvalues of the kernel matrix (Widjaja et al., 2012). $N = 3000$ in this example. Therefore, the eigenvalue decomposition is conducted on the 3000×3000 kernel matrix $\widetilde{\mathbb{K}}_{\mathbf{D}}$. In corresponding, a total of 3000 eigenvalues and eigenvectors will be generated. The optimal value of σ^2 is found by trial-and-error and is determined to be $\sigma^2 = 0.58$ here. The number of retained principal subspaces (eigenvectors) preserves the 95% variance, leading to a value $M = 16$. The SPE statistic is then calculated according to Eq. (4.16). The corresponding threshold at the significance level $\alpha = 0.01$ is computed by the KDE (Silverman, 2018). The result is plotted in Figure 4.7.

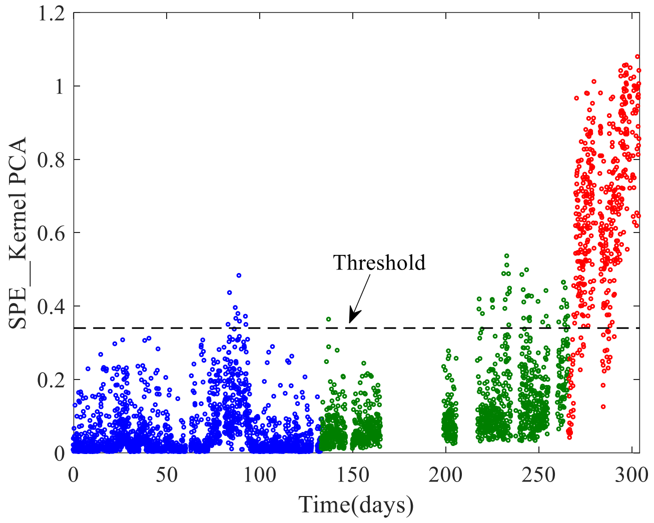


Figure 4.7 SPE statistic of kernel PCA

In Figure 4.7, the SPEs of the health data in cold days are much smaller than those in the damage state, and the false-positive damage identification is eliminated. However, when the temperature variations exceed the range of those in the training data, the SPEs increase and the validation data is falsely identified to be in the damaged state. Besides, the kernel PCA is at the expense of increased computational cost. The selection of the optimal value of σ^2 is conducted by trial-and-error, which is time-consuming. The entire process, including the selection of the optimal kernel parameters, eigenvalue decomposition of the covariance matrix in the reproducing kernel Hilbert space, selection of dimension to be retained and SPE statistic computation, takes 135 minutes approximately in the same computer, which is around 17 times of that using GPLVM.

4.3.4 Summary of Models

The models of the Bayesian FA, kernel PCA and GPLVM can be unified as

$$\boldsymbol{\Phi}(\mathbf{D}_n) = \boldsymbol{f}(\mathbf{z}_n) + \boldsymbol{\varepsilon} \tag{4.19}$$

The model corresponds to the linear Bayesian FA when both $\boldsymbol{\Phi}$ and $\boldsymbol{f}$ are linear functions, that is, $\boldsymbol{\Phi}(\mathbf{D}_n) = \mathbf{D}_n$ and $\boldsymbol{f}(\mathbf{z}_n) = \boldsymbol{\mathcal{W}}\mathbf{z}_n$; it becomes kernel PCA when $\boldsymbol{\Phi}$

changes into nonlinear function and $\boldsymbol{f}$ remains linear, and it becomes GPLVM when $\boldsymbol{f}$ is nonlinear while $\boldsymbol{\Phi}$ remains linear. The comparisons among these three models are summarized in Table 4.2. Compared with the non-probabilistic kernel PCA, the method based on GPLVM could automatically determine the optimal kernel parameters and is more efficient.

Table 4.2 Comparison of different models

Models	Probabilistic	Nonlinear	Dominant Computational Cost	SPE Threshold
Bayesian FA	YES	NO	$O(N_m \cdot M \cdot N)$	$g \cdot \chi^2_{h,\alpha}$
Kernel PCA	NO	YES	$O(N^3)$	KDE
GPLVM	YES	YES	$O(d^2 \cdot N)$	KDE

Note: N_m is the dimension of the original measurement data, M is the dimensions to be retained for data compression ($M \leq N_m$), N is the number of training data samples, and d is the number of data in active set ($d < N$).

4.4 Summary

This chapter develops a structural damage detection method on the basis of the probabilistic kernelized model to eliminate the nonlinear environmental effects. The nonlinear model is estimated automatically in the Bayesian probabilistic framework without knowing the environment data. The method is validated with the monitoring data from the benchmark Z24 bridge, whose frequencies have a nonlinear relationship with the temperature. The results indicate that the method could discriminate damage from the varying environment factors, and accurately detect the onset of damage. Besides, the proposed method could eliminate the false-positive damage identification when the training data have not covered the temperature variation. As compared with the non-probabilistic kernel PCA, the method proposed in this chapter is more efficient and robust.

CHAPTER 5

SPARSE BAYESIAN LEARNING FOR STRUCTURAL DAMAGE IDENTIFICATION BASED ON LAPLACE APPROXIMATION

5.1 Introduction

In Chapters 3 and 4, probabilistic ML-based methods are developed for the first level damage detection, that is, to determine the existence of the damage. This chapter aims at a higher level of damage identification, that is, damage localization and quantification based on the Bayesian inference. The sparse damage information is incorporated into the Bayesian probabilistic framework, forming the SBL model. One major challenge is that the denominator in the SBL equation, named the evidence, involves a too complex integral to be calculated analytically. Specifically, this item is high dimensional and involves a nonlinear relationship. As a result, the posterior PDF of the damage index cannot be obtained explicitly. This study addresses this issue and develops a Laplace approximation-based SBL method for structural damage identification. The unknown variables and hyper-parameters are optimized automatically without the need to be determined in advance.

5.2 Bayesian Probabilistic Framework

Assuming that the first N_m vibration modes of a structure are available, namely, the eigenvalue vector $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_{N_m}] \in \mathbb{R}^{1\times N_m}$ and the mode shape matrix $\hat{\boldsymbol{\Psi}} = [\hat{\boldsymbol{\phi}}_1, \hat{\boldsymbol{\phi}}_2, \cdots, \hat{\boldsymbol{\phi}}_{N_m}] \in \mathbb{R}^{N_P\times N_m}$, where $\hat{\boldsymbol{\phi}}_r \in \mathbb{R}^{N_p}$ is the *r*th mode shape vector corresponding to N_p measured degrees of freedoms.

5.2.1 Definition of A Model Class

A linear structural model is assumed. The mass matrix is regarded as known, and the stiffness matrix $\mathbf{K}$ is parameterized as a combination of n element stiffness matrix $\mathbf{K}_i$ (i=1, 2, …, n) (Zhou et al., 2015)

$$\mathbf{K} = \sum_{i=1}^{n} s_i \, \mathbf{K}_i \tag{5.1}$$

where s_i is the stiffness parameter of the ith element.

Suppose that the structural mass matrix $\mathbf{M}$ keeps unchanged in the damaged state while the stiffness matrix reduces to $\bar{\mathbf{K}}$, expressed as:

$$\bar{\mathbf{K}} = \sum_{i=1}^{n} \bar{s}_i \, \mathbf{K}_i \tag{5.2}$$

where $\bar{s}_i$ is the ith element stiffness parameter in the damaged condition. The structural stiffness reduction factor (SRF) is then defined as (Zhou et al., 2015):

$$\theta_i = \frac{\bar{s}_i - s_i}{s_i} \tag{5.3}$$

where θ_i is larger than -1 and no more than 0, particularly, a value of 0 represents the intact state of the element and -1 denotes a complete damage. Thus θ_i is treated as the damage index that reflects both the damage localization and severity.

The analytical eigenvalues and mode shapes calculated from the FE model are governed by the eigenequation, defined as

$$(\bar{\mathbf{K}} - \lambda_r \mathbf{M})\boldsymbol{\phi}_r = \mathbf{0} \tag{5.4}$$

where λ_r and $\boldsymbol{\phi}_r$ represent the rth analytical eigenvalue and mode shape, respectively.

5.2.2 Bayesian Model Updating Framework

With the selected model class $\mathcal{M}$ and measurement data, the overall Bayesian equation can be expressed as (Beck and Katafygiotis, 1998)

$$p(\boldsymbol{\theta}|\boldsymbol{\mathcal{D}},\mathcal{M}) = c^{-1}p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{M}) \tag{5.5}$$

where $\boldsymbol{\mathcal{D}}$ denotes the measured structural modal parameters including frequencies and mode shapes $\langle\hat{\lambda}_r,\widehat{\boldsymbol{\phi}}_r\rangle$; $\boldsymbol{\theta}$ is the damage index defined in Eq. (5.3); $p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta},\mathcal{M})$ denotes the posterior PDF of $\boldsymbol{\theta}$ given $\boldsymbol{\mathcal{D}}$ and $\mathcal{M}$; $p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta},\mathcal{M})$ is the likelihood function; $p(\boldsymbol{\theta}|\mathcal{M})$ is the prior PDF of $\boldsymbol{\theta}$; and $c = p(\boldsymbol{\mathcal{D}}|\mathcal{M})$ is the normalization constant, termed evidence.

5.2.3 Likelihood Function for Structural Modal Parameters

For simplicity, items dependent on $\mathcal{M}$ are herein omitted, and the likelihood function is expressed as $p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta})$. The modal parameters are deemed independent from frequencies and mode shapes and from mode to mode (Vanik, 2020) as

$$p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}) = p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta})p(\widehat{\boldsymbol{\Psi}}|\boldsymbol{\theta}) = \prod_{r=1}^{N_m} p(\hat{\lambda}_r|\boldsymbol{\theta})p(\widehat{\boldsymbol{\phi}}_r|\boldsymbol{\theta}) \tag{5.6}$$

The discrepancy between analytical and measured modal parameters, denoted as $\boldsymbol{\varepsilon}_r$ and $\boldsymbol{e}_r$, is assumed to follow the Gaussian distributions with a zero mean and diagonal variance matrix, and respectively expressed as

$$\boldsymbol{\varepsilon}_r = \frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r} \sim N(0,\beta^{-1}) \tag{5.7}$$

$$\boldsymbol{e}_r = \widehat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta}) \sim N(\boldsymbol{0},\, \gamma^{-1}\boldsymbol{I}) \tag{5.8}$$

where $\lambda_r(\boldsymbol{\theta})$ and $\boldsymbol{\phi}_r(\boldsymbol{\theta})$ represent the model-predicted eigenvalues and mode shapes, respectively; $\hat{\lambda}_r$ and $\widehat{\boldsymbol{\phi}}_r$ are the measured counterparts, and the hyper-parameters β and

γ equal the reciprocal of the variance, reflecting the precision of the measured eigenvalue and mode shapes, respectively.

The likelihood functions are then given by

$$p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta},\beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{N_m}{2}} exp\left\{-\frac{\beta}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\} \tag{5.9}$$

$$p(\hat{\boldsymbol{\phi}}|\boldsymbol{\theta},\gamma) = \left(\frac{\gamma}{2\pi}\right)^{\frac{N_p \cdot N_m}{2}} exp\left\{-\frac{\gamma}{2}\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right\} \tag{5.10}$$

5.2.4 Prior PDF of Damage Index

In the Bayesian framework, the prior is generally determined according to engineering judgements or objectives. In this study, the prior PDF of $\boldsymbol{\theta}$ serves as the regularization item for the ill-posed damage detection problem, and the ARD prior (Tipping, 2001) following the Gaussian distribution is defined. Moreover, each damage index θ_i is defined by an individual hyper-parameter α_i associated with the uncertainty of θ_i. Consequently, the prior PDF of $\boldsymbol{\theta}$ is the product of *n* independent Gaussian distributions as follows, each corresponding to one damage index,

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} p(\theta_i|\alpha_i) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \prod_{i=1}^{n}\left[\alpha_i^{\frac{1}{2}} exp\left\{-\frac{1}{2}\alpha_i\theta_i^2\right\}\right] \tag{5.11}$$

where α_i reflects the precision of θ_i.

5.2.5 Posterior PDF of Damage Index

The posterior PDF is then derived according to Eq. (5.5), that is,

$$p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\beta,\gamma) = c^{-1}p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta},\beta)p(\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\gamma)p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \tag{5.12}$$

where the evidence c is calculated by integrating the numerator $p\big(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta}\big|\boldsymbol{\alpha}, \beta, \gamma\big)$ in Eq. (5.5) with respect to $\boldsymbol{\theta}$, expressed as

$$
\begin{aligned}
c = \; & p(\boldsymbol{\mathcal{D}}|\, \mathcal{M}) = p\big(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}\big|\boldsymbol{\alpha}, \beta, \gamma\big) = \int p\big(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta}\big|\boldsymbol{\alpha}, \beta, \gamma\big) d\boldsymbol{\theta} \\
= & \int p\big(\hat{\boldsymbol{\lambda}}\big|\boldsymbol{\theta}, \beta\big) p\big(\hat{\boldsymbol{\psi}}\big|\boldsymbol{\theta}, \gamma\big) p(\boldsymbol{\theta}|\boldsymbol{\alpha}) d\boldsymbol{\theta} \\
= & \int \left(\frac{\beta}{2\pi}\right)^{\frac{N_m}{2}} \left(\frac{\gamma}{2\pi}\right)^{\frac{N_p \cdot N_m}{2}} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\prod_{i=1}^{n} \alpha_i^{\frac{1}{2}}\right) exp\left\{-\frac{\beta}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2 \right. \\
& \left. -\frac{\gamma}{2}\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\alpha_i \theta_i^2\right)\right\} d\boldsymbol{\theta}
\end{aligned}
\tag{5.13}
$$

5.3 Bayesian Inference based on Laplace Approximation

The MPV of the damage index $\boldsymbol{\theta}$ equals the MAP estimate of $p\big(\boldsymbol{\theta}\big|\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\alpha}, \beta, \gamma\big)$, that is

$$
\begin{aligned}
\overline{\boldsymbol{\theta}} & = \arg\max \; p\big(\boldsymbol{\theta}\big|\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\alpha}, \beta, \gamma\big) \\
& = \arg\max \; c^{-1} p\big(\hat{\boldsymbol{\lambda}}\big|\boldsymbol{\theta}, \beta\big) p\big(\hat{\boldsymbol{\psi}}\big|\boldsymbol{\theta}, \gamma\big) p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \\
& = \arg\max \; c^{-1} \left(\frac{\beta}{2\pi}\right)^{\frac{N_m}{2}} \left(\frac{\gamma}{2\pi}\right)^{\frac{N_p N_m}{2}} \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\prod_{i=1}^{n} \alpha_i^{\frac{1}{2}}\right) \times \\
& \qquad exp\left\{-\frac{\beta}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2 - \frac{\gamma}{2}\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\alpha_i \theta_i^2\right)\right\}
\end{aligned}
\tag{5.14}
$$

For the convenience of calculation, the negative natural logarithm of the right-hand side in Eq. (5.14) is used, that is,

$$
\begin{aligned}
\overline{\boldsymbol{\theta}} & = \arg\;\min\; J(\boldsymbol{\theta}) \\
& = \arg\;\min\; \beta\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2 + \gamma\sum_{r=1}^{N_m}\sum_{j=1}^{N_p}\left[\hat{\phi}_{j,r} - \phi_{j,r}(\boldsymbol{\theta})\right]^2 + \sum_{i=1}^{n}\left(\alpha_i \theta_i^2\right)
\end{aligned}
\tag{5.15}
$$

where the items unrelated with $\boldsymbol{\theta}$ are dropped for simplicity.

In addition to the damage index $\boldsymbol{\theta}$, the hyper-parameters $\{\boldsymbol{\alpha}, \beta, \gamma\}$ in Eq. (5.15) are also unknown, which need to be estimated first. The MPVs of $\{\boldsymbol{\alpha}, \beta, \gamma\}$ can be estimated by maximizing their posterior PDF $p\left(\boldsymbol{\alpha}, \beta, \gamma \middle| \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}\right)$ with respect to $\boldsymbol{\alpha}$, β and γ, respectively. In particular,

$$p\left(\boldsymbol{\alpha}, \beta, \gamma \middle| \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}\right) = \frac{p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) p(\boldsymbol{\alpha}, \beta, \gamma)}{p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}\right)} \propto p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) \tag{5.16}$$

where a uniformly distributed (non-informative) prior PDF is adopted for $\{\boldsymbol{\alpha}, \beta, \gamma\}$. According to Eq. (5.16), $\{\boldsymbol{\alpha}, \beta, \gamma\}$ can be estimated by maximizing $p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right)$, which is the evidence in Eq. (5.13).

Due to the high dimensional integral involved in the evidence and the nonlinear relationship between $\langle \lambda_r, \boldsymbol{\phi}_r \rangle$ and $\boldsymbol{\theta}$, a closed-form solution of the evidence is difficult to obtain. Therefore, asymptotic or numerical techniques are required. In Chapter 3, the EM algorithm is utilized. However, the posterior statistics of $\boldsymbol{\theta}$ cannot be directly identified from Eq. (5.14). Hence the expectation in the E step cannot be calculated analytically here. The Laplace approximation is a desirable technique that can be applied to obtain the asymptotic solution of the integral, provided that the system is globally or locally identifiable based on the available measurements (Beck and Katafygiotis, 1998), as explained in Section 2.3.1. Assuming that the problem is globally identifiable with the measured modal parameters, the item within the integral in Eq. (5.13), $p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right)$ can be approximated as a Gaussian distribution based on the Laplace approximation, that is,

$$p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) \cong p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \overline{\boldsymbol{\theta}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) exp\left\{-\frac{1}{2}(\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{A} (\boldsymbol{\theta} - \overline{\boldsymbol{\theta}})\right\} \tag{5.17}$$

where $\overline{\boldsymbol{\theta}}$ denotes the mode of $p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right)$ and equals the MAP estimate calculated by Eq. (5.15). The matrix $\mathbf{A}$ denotes the Hessian matrix, which equals the second derivatives of the objective function with respect to the damage index, that is,

$$\mathbf{A} = -\nabla\nabla \ln p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) |_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = \frac{\partial}{\partial\theta_i \partial\theta_j} J(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = \mathbf{W} + \beta\mathbf{H} + \gamma\mathbf{P} \tag{5.18}$$

In Eq. (5.18), the Hessian matrix **A** is decomposed into three items, i.e., **W** is a diagonal matrix with entries $\mathbf{W}_{ii} = \alpha_i$, and **H** and **P** are computed as

$$\begin{aligned} \mathbf{H}_{ij} &= \frac{\partial}{\partial\theta_i \partial\theta_j} \frac{1}{2} \sum_{r=1}^{N_m} \left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r} \right]^2 \\ &= \sum_{r=1}^{N_m} \frac{1}{\hat{\lambda}_r^{\ 2}} \cdot \frac{\partial \lambda_r(\bar{\boldsymbol{\theta}})}{\partial\theta_i} \cdot \frac{\partial \lambda_r(\bar{\boldsymbol{\theta}})}{\partial\theta_j} - \sum_{r=1}^{N_m} \left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r} \right] \cdot \frac{1}{\hat{\lambda}_r} \cdot \frac{\partial \lambda_r(\bar{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j} \end{aligned} \tag{5.19}$$

$$\begin{aligned} \mathbf{P}_{ij} &= \frac{\partial}{\partial\theta_i \partial\theta_j} \frac{1}{2} \sum_{r=1}^{N_m} \sum_{k=1}^{N_p} \left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}}) \right]^2 \\ &= \sum_{r=1}^{N_m} \sum_{k=1}^{N_p} \frac{\partial \phi_{k,r}(\bar{\boldsymbol{\theta}})}{\partial\theta_i} \cdot \frac{\partial \phi_{k,r}(\bar{\boldsymbol{\theta}})}{\partial\theta_j} - \sum_{r=1}^{N_m} \sum_{k=1}^{N_p} \left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}}) \right] \cdot \frac{\partial \phi_{k,r}(\bar{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j} \end{aligned} \tag{5.20}$$

In the above equations, the first and second derivatives of the eigenvalues and eigenvectors with respect to the stiffness parameter can be calculated by employing the Nelson's method (Nelson, 1976) or the substructural approach (Weng et al., 2011, 2013).

A normalized Gaussian distribution proportional to $p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right)$ is expressed as

$$q(\boldsymbol{\theta}) = \frac{|\mathbf{A}|^{1/2}}{(2\pi)^{n/2}} exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{A} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \right\} = \mathcal{N}\left(\boldsymbol{\theta} \middle| \bar{\boldsymbol{\theta}}, \mathbf{A}^{-1}\right) \tag{5.21}$$

where $|\mathbf{A}|$ denotes the determinant of the matrix **A**. On the basis that the integral of a normalized Gaussian distribution equals one, the evidence $p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right)$ can be computed as

$$\begin{aligned} p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) &= \int p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \boldsymbol{\theta} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) d\boldsymbol{\theta} \\ &\cong p\left(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \bar{\boldsymbol{\theta}} \middle| \boldsymbol{\alpha}, \beta, \gamma\right) \int exp\left\{ -\frac{1}{2} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{A} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) \right\} d\boldsymbol{\theta} \end{aligned} \tag{5.22}$$

$$= p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \overline{\boldsymbol{\theta}} | \boldsymbol{\alpha}, \beta, \gamma) \frac{(2\pi)^{n/2}}{|\mathbf{A}|^{1/2}}$$

Hence the asymptotic analytical solution of the evidence is obtained. For the convenience of optimization, the logarithm of Eq. (5.22) is adopted, that is

$$\begin{aligned}
&\ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma) = \ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}}, \overline{\boldsymbol{\theta}} | \boldsymbol{\alpha}, \beta, \gamma) + \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{A}| \\
&= \ln p(\hat{\boldsymbol{\lambda}} | \overline{\boldsymbol{\theta}}, \beta) + \ln p(\hat{\boldsymbol{\psi}} | \overline{\boldsymbol{\theta}}, \gamma) + \ln p(\overline{\boldsymbol{\theta}} | \boldsymbol{\alpha}) + \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{A}| \\
&= \frac{N_m}{2}\ln\left(\frac{\beta}{2\pi}\right) + \frac{N_p N_m}{2}\ln\left(\frac{\gamma}{2\pi}\right) + \frac{n}{2}\ln\left(\frac{1}{2\pi}\right) + \frac{1}{2}\sum_{i=1}^{n}\ln\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\left(\alpha_i \bar{\theta}_i^{\,2}\right) \\
&\quad - \frac{\beta}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\overline{\boldsymbol{\theta}})}{\hat{\lambda}_r}\right]^2 - \frac{\gamma}{2}\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\overline{\boldsymbol{\theta}})\right\|_2^2 + \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{A}|
\end{aligned} \tag{5.23}$$

Based on Eq. (5.16), the MPVs of $\{\boldsymbol{\alpha}, \beta, \gamma\}$ can be calculated by maximizing $p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)$. Setting the derivative of $\ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)$ in Eq. (5.23) with respect to $\boldsymbol{\alpha}$ to zero, we have

$$\begin{aligned}
&\frac{\partial \ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)}{\partial \alpha_i} = \frac{1}{2\alpha_i} - \frac{\overline{\boldsymbol{\theta}}_i^{\,2}}{2} - \frac{1}{2}\frac{\partial \ln|\mathbf{W} + \beta\mathbf{H} + \gamma\mathbf{P}|}{\partial \alpha_i} \\
&= \frac{1}{2\alpha_i} - \frac{\overline{\boldsymbol{\theta}}_i^{\,2}}{2} - \frac{1}{2}Tr\left[(\mathbf{W} + \beta\mathbf{H} + \gamma\mathbf{P})^{-1}\frac{\partial(\mathbf{W} + \beta\mathbf{H} + \gamma\mathbf{P})}{\partial \alpha_i}\right] \\
&= \frac{1}{2\alpha_i} - \frac{\overline{\boldsymbol{\theta}}_i^{\,2}}{2} - \frac{1}{2}[(\mathbf{W} + \beta\mathbf{H} + \gamma\mathbf{P})^{-1}]_{ii} \\
&= 0
\end{aligned} \tag{5.24}$$

where $Tr[\cdot]$ denotes the trace of a matrix, which equals the sum of the leading diagonal entries. In Eq. (5.24), the derivative equation of a matrix is used:

$$\frac{\partial \ln|\mathbf{A}|}{\partial \boldsymbol{\alpha}} = Tr\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}}\right) \tag{5.25}$$

Since $\mathbf{W}$ is a diagonal matrix with $\mathbf{W}_{ii} = \alpha_i$, $\frac{\partial(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})}{\partial \alpha_i}$ is a diagonal zero matrix except a nonzero item at the *i*th entry. Based on Eq. (5.24), the MPV of α_i is calculated as

$$\alpha_i = \frac{1}{\bar{\boldsymbol{\theta}}_i{}^2 + [(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}]_{ii}} \tag{5.26}$$

Similarly, the MPV of β is obtained by setting the derivative of $\ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)$ with respect to β to zero, that is,

$$\begin{aligned}
\frac{\partial \ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)}{\partial \beta} &= \frac{N_m}{2\beta} - \frac{1}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r}\right]^2 - \frac{1}{2}\frac{\partial \ln |\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P}|}{\partial \beta} \\
&= \frac{N_m}{2\beta} - \frac{1}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r}\right]^2 - \frac{1}{2} Tr[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{H}] \\
&= \frac{N_m}{2\beta} - \frac{1}{2}\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r}\right]^2 - \frac{1}{2}\sum_{i=1}^{n}[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{H}]_{ii} \\
&= 0
\end{aligned} \tag{5.27}$$

According to Eq. (5.27), β is calculated as

$$\beta = \frac{N_m}{\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r - \lambda_r(\bar{\boldsymbol{\theta}})}{\hat{\lambda}_r}\right]^2 + \sum_{i=1}^{n}[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{H}]_{ii}} \tag{5.28}$$

Setting the derivative of $\ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)$ with respect to γ to zero, we have

$$\begin{aligned}
&\frac{\partial \ln p(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\psi}} | \boldsymbol{\alpha}, \beta, \gamma)}{\partial \gamma} \\
&= \frac{N_m N_p}{2\gamma} - \frac{1}{2}\sum_{r=1}^{N_m}\sum_{k=1}^{N_p}\left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}})\right]^2 - \frac{1}{2}\frac{\partial \ln |\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P}|}{\partial \gamma} \\
&= \frac{N_m N_p}{2\gamma} - \frac{1}{2}\sum_{r=1}^{N_m}\sum_{k=1}^{N_p}\left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}})\right]^2 - \frac{1}{2} Tr[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{P}] \\
&= \frac{N_m N_p}{2\gamma} - \frac{1}{2}\sum_{r=1}^{N_m}\sum_{k=1}^{N_p}\left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}})\right]^2 - \frac{1}{2}\sum_{i=1}^{n}[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{P}]_{ii} \\
&= 0
\end{aligned} \tag{5.29}$$

γ is then computed as

$$\gamma = \frac{N_m N_p}{\sum_{r=1}^{N_m}\sum_{k=1}^{N_p}\left[\hat{\phi}_{k,r} - \phi_{k,r}(\bar{\boldsymbol{\theta}})\right]^2 + \sum_{i=1}^{n}[(\mathbf{W}+\beta\mathbf{H}+\gamma\mathbf{P})^{-1}\mathbf{P}]_{ii}} \tag{5.30}$$

Note that the MPVs of $\bar{\boldsymbol{\theta}}$ in Eq. (5.15) and $\{\boldsymbol{\alpha}, \beta, \gamma\}$ in Eqs. (5.26), (5.28) and (5.30) are coupled with each other. Therefore, they can be calculated in an iterative manner. The algorithm starts with initializing the parameters, and proceeds with estimating the damage index using the current parameters and updating the parameters based on the damage index, until the convergence criterion is satisfied.

The procedure of the proposed algorithm is summarized as:

Algorithm based on Laplace Approximation

```
1. Initialize $\boldsymbol{\alpha}^{(0)}, \beta^{(0)}, \gamma^{(0)}$ and $\boldsymbol{\theta}^{(0)}$;
2. For $j = 1, 2, \ldots$, update $\bar{\boldsymbol{\theta}}^{(j)}$ using Eq. (5.15) with $\boldsymbol{\alpha}^{(j-1)}, \beta^{(j-1)}, \gamma^{(j-1)}$;
3. Update ${\alpha_i}^{(j)}$ using Eq. (5.26) with $\bar{\boldsymbol{\theta}}^{(j)}, \beta^{(j-1)}, \gamma^{(j-1)}$;
4. Update $\beta^{(j)}$ using Eq. (5.28) with $\bar{\boldsymbol{\theta}}^{(j)}, {\alpha_i}^{(j)}, \gamma^{(j-1)}$;
5. Update $\gamma^{(j)}$ using Eq. (5.30) with $\bar{\boldsymbol{\theta}}^{(j)}, {\alpha_i}^{(j)}, \beta^{(j)}$;
6. Let $j = j + 1$, repeat Steps 2~5 until the convergence criterion is met (e.g.
   $\|\bar{\boldsymbol{\theta}}^{(j)} - \bar{\boldsymbol{\theta}}^{(j-1)}\|_2 / \|\bar{\boldsymbol{\theta}}^{(j)}\|_2 \le Tol$).
```

5.4 Case Study

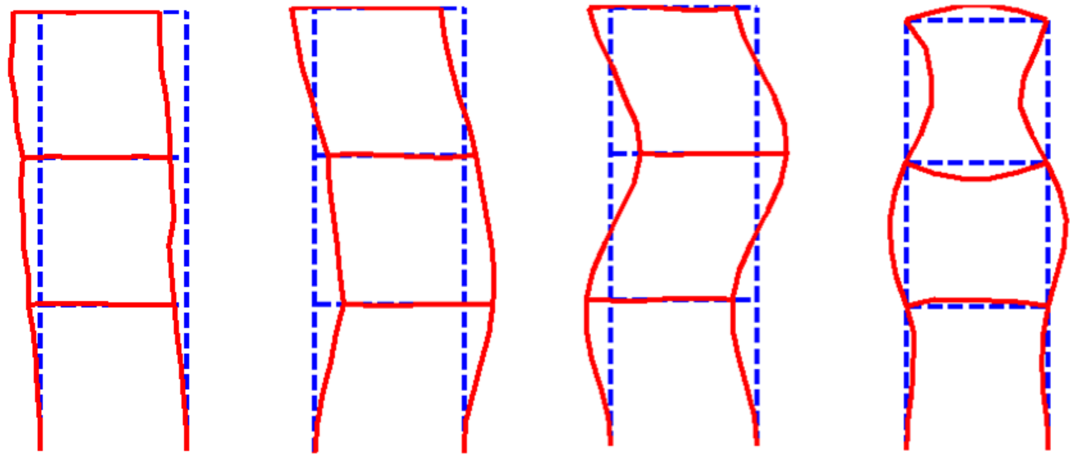
5.4.1 Model Descriptions

The proposed method is applied to a laboratory-tested three-story steel frame (Hou et al., 2018). The structure was firmly fixed on the strong floor through a steel plate. Therefore, the boundary condition of the frame is regarded as fixed. The frame has a 0.5

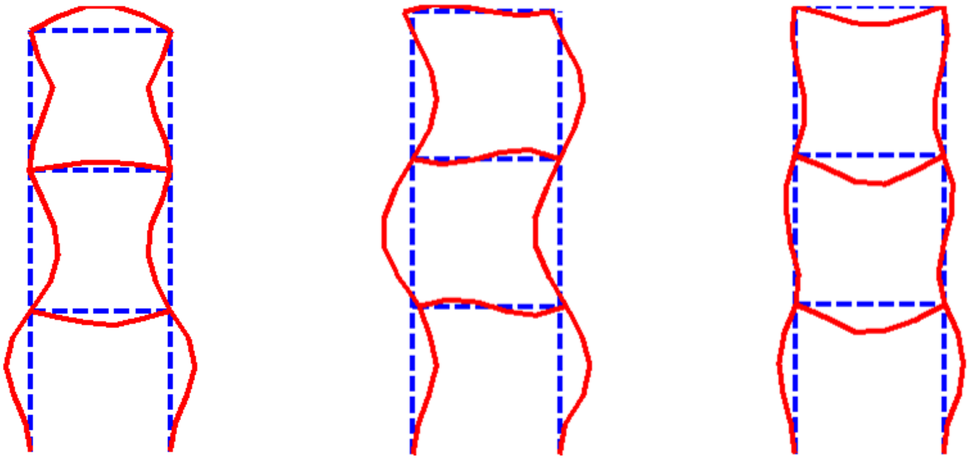
m span and 1.5 m height with three similar stories. The beam and column have an identical cross-section of 75.0 × 5.0 mm^2. The Young’s modulus of the steel material is assumed to be 2.0×10^{11} N/m^2, and the mass density is 7.92×10^3 kg/m^3. The entire frame is divided into 225 Euler-Bernoulli beam elements, each being 20 mm long. In this study, the damage is identified at the element level.

Figure 5.1 Overview of the three-story steel frame

A total of 39 measurement points with a space of 100 mm was determined to obtain the modal parameters (i.e., frequencies and mode shapes) of the frame through the hammer test. Accelerometers were installed to record the acceleration responses under the excitation of the hammer with a rubber tip. The rational fraction polynomial method (Formenti and Richardson, 2002) is utilized to extract the first seven frequencies and mode shapes of the frame in the intact state. The measured first seven mode shapes are plotted in Figure 5.2.



Mode 1 (4.23Hz) Mode 2 (14.03Hz) Mode 3 (25.45 Hz) Mode 4 (44.81 Hz)



Mode 5 (58.12Hz) Mode 6 (68.36Hz) Mode 7 (72.27 Hz)

Figure 5.2 The first seven frequencies and mode shapes in the undamaged state measured in the hammer test

Afterwards, three saw cuts in different locations were introduced to the frame in sequence, as shown in Figure 5.3. Cuts 1–3 were located at elements No. 1, 12 and 176. Three cuts have the same length of 20 mm and different depths of 22.5, 22.5, and 30 mm. The cut has the same length as the beam element. Hence, the damage severity quantified by the SRF in Eq. (5.3) equals the reduction in the moment of inertia of the cross-section. That is, the SRFs of the three damaged elements are −60%, −60% and −80%, respectively. The detailed information of the three DSs is listed in Table 5.1.

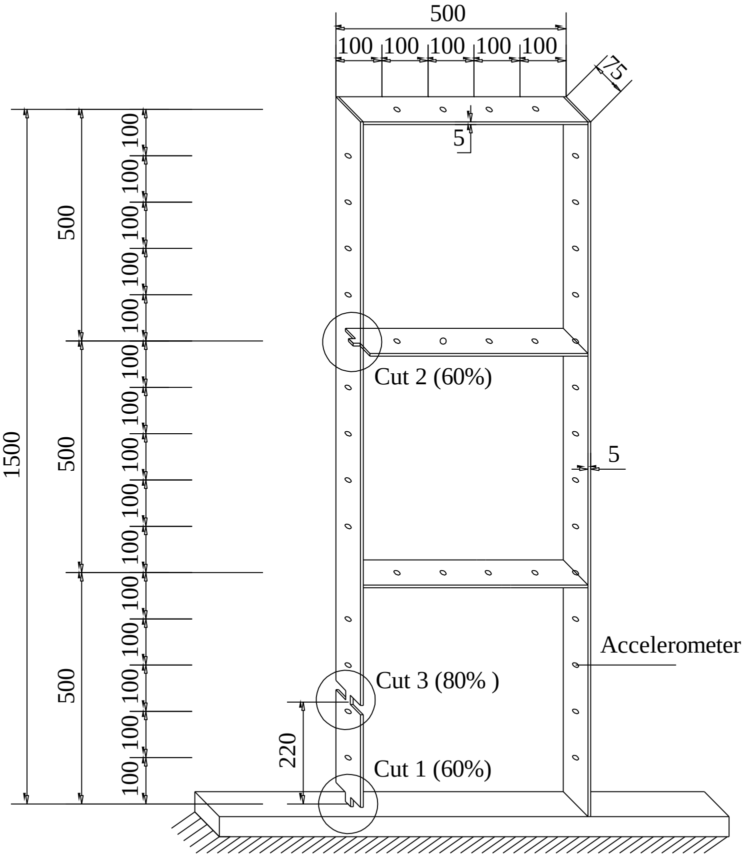


Figure 5.3 Location of accelerometers and damages (unit: mm)

Table 5.1 Damage locations and severities in three DSs

Damage Scenario	Cut No.	Element No.	Cut depth (mm)	Damaged component	SRF (θ)
DS1	Cut 1	1	22.5	Column	$\theta_1 = -60\%$
DS2	Cut 1	1	22.5	Column	$\theta_1 = -60\%$
	Cut 2	176	22.5	Beam	$\theta_{176} = -60\%$
DS3	Cut 1	1	22.5	Column	$\theta_1 = -60\%$
	Cut 2	176	22.5	Beam	$\theta_{176} = -60\%$
	Cut 3	12	30	Column	$\theta_{12} = -80\%$

The number of damaged elements in three DSs is one, two and three, respectively, which is much less than the total number of the elements (e.g. 225). Therefore, the SRF vector is very sparse. That is, most entries are zero, except several entries corresponding to the damage locations are nonzero. The hammer test mentioned previously was similarly conducted in each DS, and the modal data (frequencies and mode shapes) were extracted. Table 5.2 compares the first seven frequencies in the undamaged state and three DSs. The structural frequencies decrease with the increase of damage severity. The modal assurance criterion (MAC) is utilized for mode shape matching, so that the measured mode shapes match the analytical counterparts of the FE model (Allemang, 2003).

Table 5.2 Frequencies of the frame in undamaged and damaged states (units: Hz)

Mode	Undamaged	DS1	DS2	DS3
1	4.23	4.13 (−2.31)	4.08 (−3.53)	4.06 (−3.88)
2	14.03	13.75(−1.96)	13.45(−4.11)	13.42(−4.32)
3	25.45	25.14(−1.19)	25.13(−1.23)	25.09(−1.41)
4	44.81	44.70(−0.23)	44.69(−0.27)	44.62(−0.42)
5	58.12	57.39(−1.24)	57.28(−1.44)	56.55(−2.69)
6	68.36	67.34(−1.49)	66.11(−3.29)	65.31(−4.46)
7	72.27	72.06(−0.28)	71.42(−1.18)	70.74(−2.12)
Average (%)		(−1.44)	(−2.32)	(−3.37)

Note. Values in parentheses are the frequency change ratios (%) between the damaged and undamaged states.

5.4.2 Damage Identification

The algorithm starts with initializing the parameters. The modal testing experiences indicate that structural frequencies may contain 1% noise level and measured mode shapes are generally less accurate than the frequencies (Mottershead and Friswell, 1993). Therefore, a noise level of 1% and 5% are assigned to the frequencies and mode shapes, respectively, namely, $\beta^{(0)}=1/(0.01)^2=1\times10^4$ and $\gamma^{(0)}=1/(0.05)^2=400$. The noise level of the damage index in the prior is estimated to be 10% of the true value, that is, $\alpha_i^{(0)}=1/(10\%)^2=100$ $(i=1, 2, \ldots, 225)$. The actual damage location and severity are unknown, thus the initial damage index is assumed to be a zero vector, that is, $\boldsymbol{\theta}^{(0)}=\{0, \ldots, 0\}^{\mathrm{T}}$. The convergence criterion of the proposed algorithm is set to be

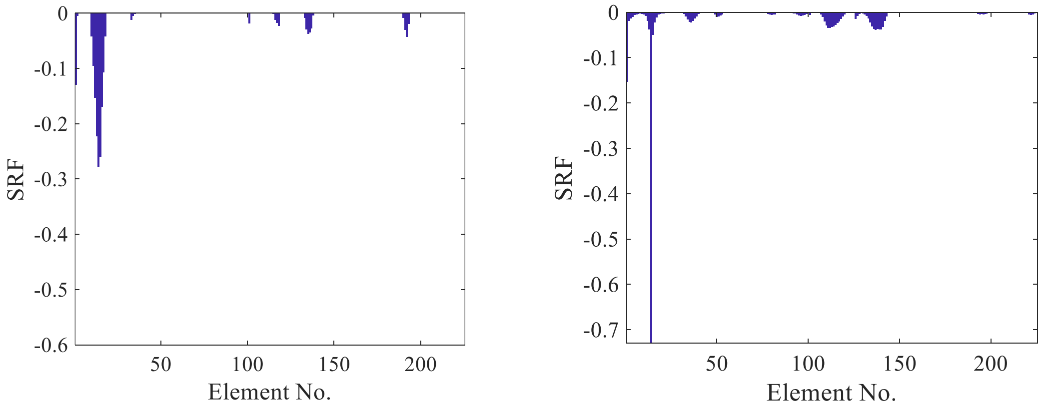
$\left\|\overline{\boldsymbol{\theta}}^{(j)} - \overline{\boldsymbol{\theta}}^{(j-1)}\right\|_2 / \left\|\overline{\boldsymbol{\theta}}^{(j)}\right\|_2 \leq 0.01$. To evaluate the accuracy of damage quantification, an error index is defined as

$$\delta = \sqrt{\frac{\left\|\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}\right\|_2^2}{n}} \tag{5.31}$$

where $\overline{\boldsymbol{\theta}}$ is the damage index identified by the proposed algorithm.

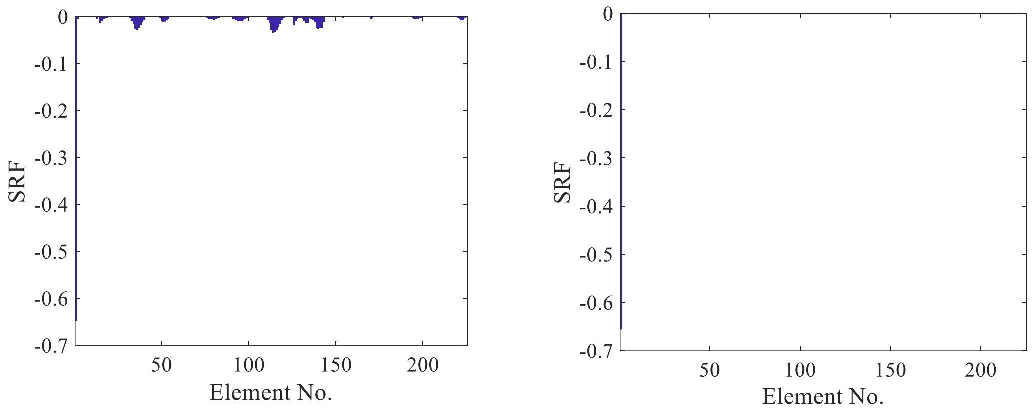
DS1. In DS1, only the first element was damaged. The damage identification results in the iterative process are shown in Figure 5.4. Due to the space limitation, only the results in the first two and last two iterations are presented here. In the first two iterations, multiple elements are mistakenly identified as damaged. After six iterations, the identification results converge, and the actual damage in the first element is identified. The identification error is 0.37%.

DS2. In DS2, No. 1 and 176 elements were damaged. The proposed algorithm is applied, and the identification results converge after five iterations. Figure 5.5 shows the results in the first two and last two iterations. Two damage are accurately localized upon convergence, with the severity identification error of 1.29%.



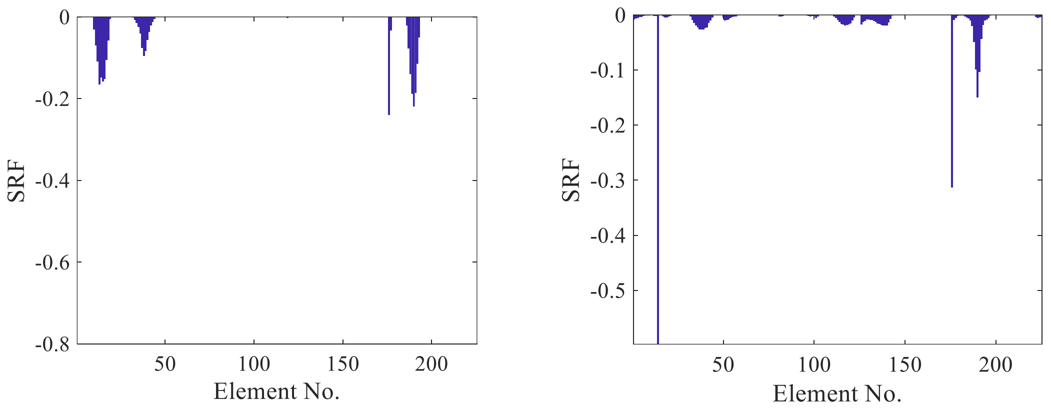
(a) Iteration No.1

(b) Iteration No.2

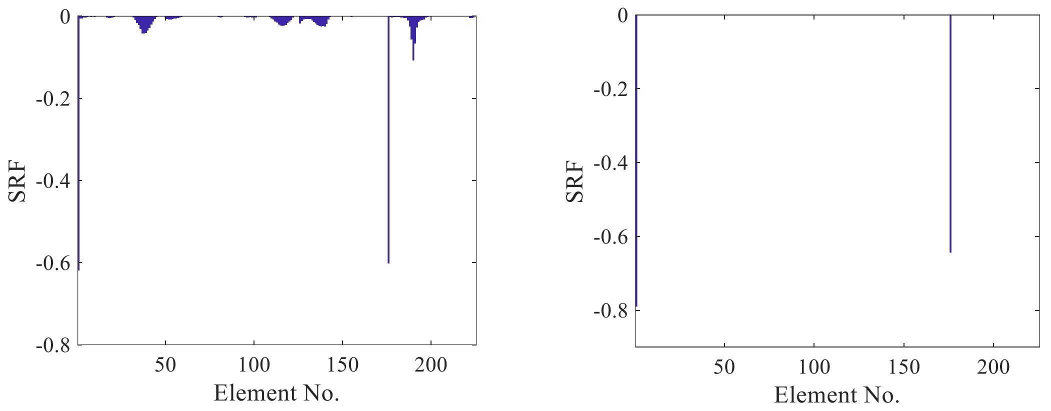


(c) Iteration No.5 (d) Iteration No.6

Figure 5.4 Damage identification results of DS1



(a) Iteration No.1 (b) Iteration No.2

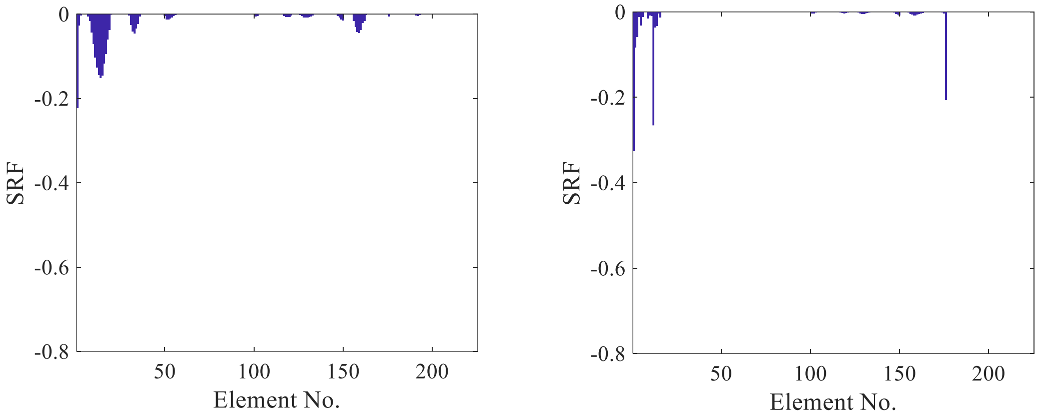


(c) Iteration No.4 (d) Iteration No.5

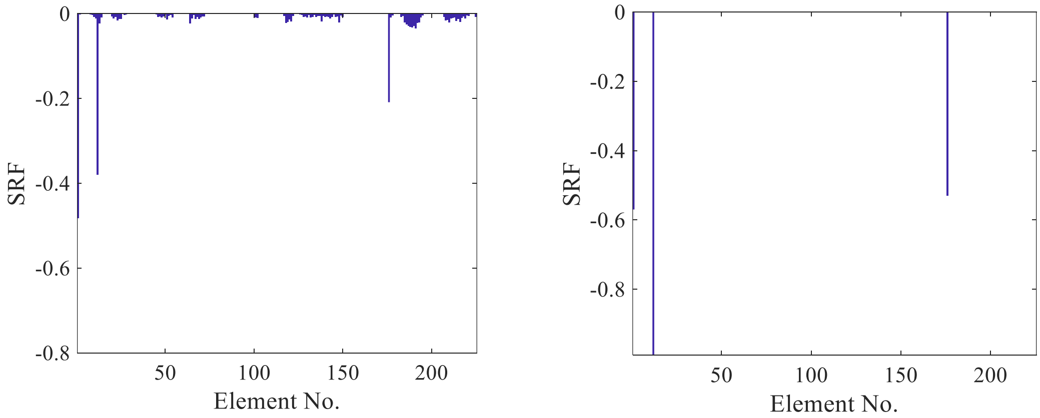
Figure 5.5 Damage identification results of DS2

DS3. In DS3, three elements (No. 1, 12 and 176) are damaged. The convergence is attained after four iterations. The damage identification results are presented in Figure 5.6. Again, all damages are localized correctly with the severity identification error of 1.36%.

All damaged elements in three DSs are detected accurately with minor quantification errors, all of which are less than 2%. It can be concluded that the proposed Laplace approximation-based sparse Bayesian method is able to locate and quantify the sparse damage.



(a) Iteration No.1 (b) Iteration No.2

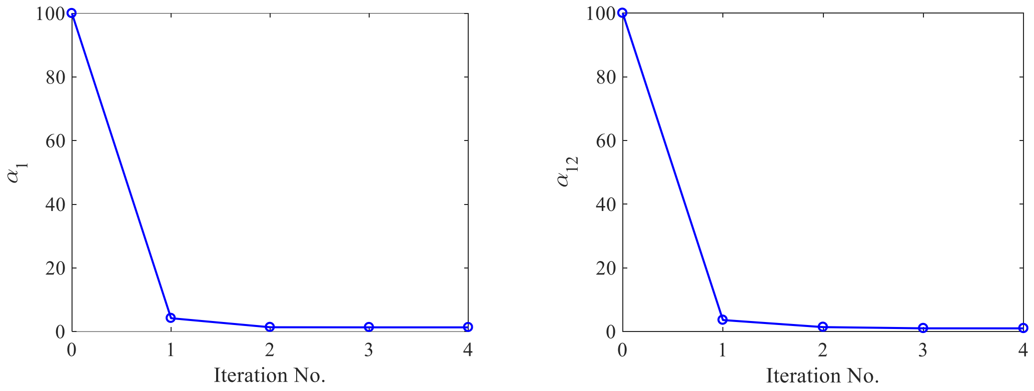


(c) Iteration No.3 (d) Iteration No.4

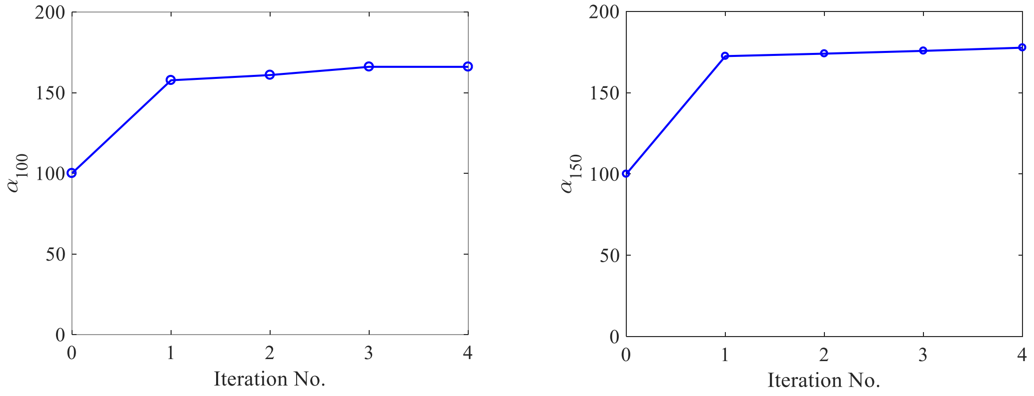
Figure 5.6 Damage identification results of DS3

5.4.3 Variation of Hyper-parameters

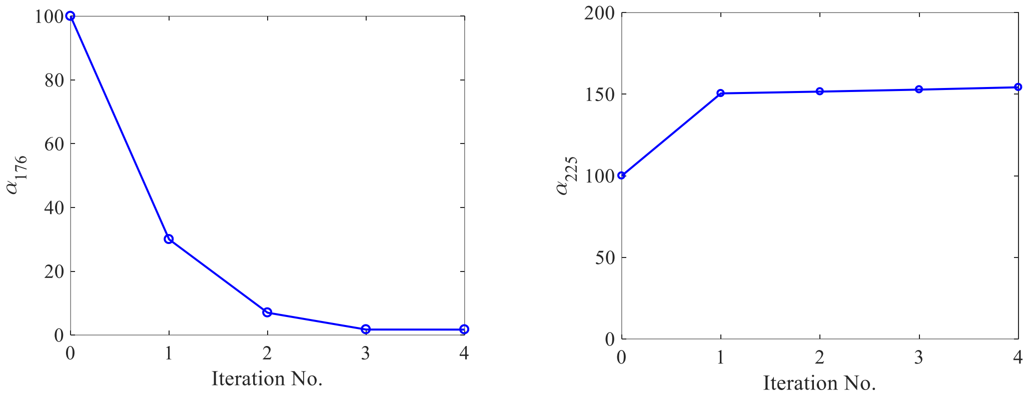
The hyper-parameters also alter as the iterations proceed. Figures 5.7 shows the variations of a few parameters in DS3. Although all α_i are initialized identically to be 100, they converge to different values. Specifically, α_i corresponding to the undamaged elements (for example, $\alpha_{100}, \alpha_{150}$ and α_{225}) converge to a significant value, as compared with α_i associated with the damaged elements (α_1, α_{12} and α_{176}). This result is consistent with the mechanism of the ARD model, where each variable is assigned with an individual hyper-parameter. Most hyper-parameters in the optimization process become significant. Hence, the associated variables are penalized more, enforcing these variables to zero and inducing the sparse results. In DS3, most α_i are significantly larger than the rest three counterparts. In corresponding, all damage indices except the three ones corresponding to damaged locations (θ_1, θ_{12} and θ_{176}) are optimized to zero, realizing the sparse damage detection.



(a) α_1 (b) α_{12}



(c) α_{100} (d) α_{150}



(e) α_{176} (f) α_{225}

Figure 5.7 Variation of hyper-parameters during the iteration process of DS3

5.5 Comparison with l_p (p=0, 1 and 2) Regularization Technique

The SBL developed in this study is compared with the l_p (p=0, 1, and 2) regularization technique in the sparse recovery.

Eq. (5.21) is similar to the objection function in regularization methods. The first two terms correspond to the weighted model fitting terms, and the third term corresponds to the regularization term. However, the ARD prior assigns an independent "regularization parameter" α_i to each damage index θ_i. Some α_i converge to a significant value in the optimization process, enforcing the corresponding damage index to zero and induce sparsity to the results. Eq. (5.26) reveals that the regularization parameter is at the level of $1/\theta^2$, reflecting that Eq. (5.21) closely resembles the l_0 regularization technique (Daubechies et al., 2010). If all α_i are set to the same value, Eq. (5.26) is equivalent to the l_2 regularization technique. And if the Laplace prior is defined on the damage index rather than the Gaussian prior, Eq. (5.26) is equivalent to the l_1 regularization technique (Babacan et al., 2009). It is noteworthy that l_p (p=0, 1, and 2) regularization parameters need to be determined in advance. Consequently, the proposed SBL method is preferred given that all unknown parameters are optimized automatically.

5.6 Summary

The chapter proposes a sparse Bayesian model for structural damage detection based on the Laplace approximation. The ARD model is adopted as the sparsity prior in the Bayesian probabilistic framework, which equivalently functions as the regularization item to tackle the ill-posed inverse problem in the structural damage detection. For the problem of the intractable posterior PDF resulted by the nonlinear relationship between the damage index and modal parameters, the Laplace approximation is utilized. The item to be integrated within the evidence is approximated as a Gaussian distribution. In this manner, the posterior PDFs of the damage index and hyper-parameters are expressed in an analytical form, based on which the MPVs are solved iteratively.

The proposed method is applied to a three-story steel frame, to which three damages were sequentially introduced. The results indicate that the sparse damage in all three DSs can be accurately localized with minor severity quantification errors. The effectiveness of the proposed method is thus verified.

CHAPTER 6

STRUCTURAL DAMAGE IDENTIFICATION BASED ON VARIATIONAL BAYESIAN INFERENCE AND DELAYED REJECTION ADAPTIVE METROPOLIS ALGORITHM

6.1 Introduction

The Laplace approximation technique in Chapter 5 only works well when the model updating problem is globally or locally identifiable. In most cases, the condition may not be satisfied due to the limitation of measurement data. In locally identifiable and unidentifiable cases, finding all optimal modes or eligible points is computationally challenging and nontrivial, especially in the high-dimensional and nonconvex optimizations. Besides, the Hessian matrix in high-dimensional problems also requires a high computation cost. In this chapter, an improved SBL is developed for damage identification based on VBI and DRAM, which is applicable to both standard and nonstandard distributions and is highly efficient for high-dimensional problems. A comparative study between the method in Chapter 5 and the VBI-DRAM algorithm is conducted.

6.2 Sparse Bayesian Method for Model Updating

The definitions of the Bayesian equation and the prior PDF are the same as Eqs. (5.5) and (5.11). The difference lies in the likelihood function, which considers the different uncertainties of each mode. In corresponding, the posterior PDF is also different. The likelihood function and posterior PDF are detailed as follows.

6.2.1 Likelihood Function

The discrepancy between the model predictions and measurement data arises from the measurement noises and modelling errors. The two sources do not exhibit an identical error for each frequency and mode shape. In Section 3.2.3, the uncertainties of different modes are assumed equally, which may affect the accuracy of the uncertainty assessment. In this Chapter, the uncertainty of each modal parameter is individually evaluated as

$$\boldsymbol{\varepsilon}_r = \frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r} \quad \sim\mathcal{N}\left(0, {\beta_r}^{-1}\right) \tag{6.1}$$

$$\boldsymbol{e}_r = \hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta}) \quad \sim\mathcal{N}(\mathbf{0},\, {\gamma_r}^{-1}\boldsymbol{I}) \tag{6.2}$$

where β_r and γ_r reflect the uncertainty level of each mode.

Then the likelihood function of each modal parameter is individually expressed as

$$p\left(\hat{\lambda}_r \middle| \boldsymbol{\theta}, \beta_r\right) = \left(\frac{\beta_r}{2\pi}\right)^{\frac{1}{2}} exp\left\{-\frac{1}{2}\beta_r \cdot \left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\} \tag{6.3}$$

$$p\left(\hat{\boldsymbol{\phi}}_r \middle| \boldsymbol{\theta}, \gamma_r\right) = \left(\frac{\gamma_r}{2\pi}\right)^{\frac{N_p}{2}} exp\left\{-\frac{1}{2}\gamma_r \cdot \left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right\} \tag{6.4}$$

The modal parameters are assumed to be independent from mode to mode. Therefore, the resulting likelihood functions of $\boldsymbol{\theta}$ based on measured modal parameters $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\psi}}$ are formulated as

$$\begin{aligned} p\left(\hat{\boldsymbol{\lambda}} \middle| \boldsymbol{\theta}, \boldsymbol{\beta}\right) &= \prod_{r=1}^{N_m} p\left(\hat{\lambda}_r \middle| \boldsymbol{\theta}, \beta_r\right) \\ &= \left(\prod_{r=1}^{N_m} \frac{\beta_r}{2\pi}\right)^{\frac{1}{2}} exp\left\{-\frac{1}{2}\sum_{r=1}^{N_m}\left(\beta_r \cdot \left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)\right\} \end{aligned} \tag{6.5}$$

$$
\begin{aligned}
p(\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\boldsymbol{\gamma}) &= \prod_{r=1}^{N_m} p(\hat{\boldsymbol{\phi}}_r|\boldsymbol{\theta},\gamma_r) \\
&= \left(\prod_{r=1}^{N_m}\frac{\gamma_r}{2\pi}\right)^{\frac{N_p}{2}} exp\left\{-\frac{1}{2}\sum_{r=1}^{N_m}\left(\gamma_r\cdot\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)\right\}
\end{aligned}
\tag{6.6}
$$

where N_m is the number of measured modes.

6.2.2 Posterior PDF

The posterior PDF of the damage index is then formulated as

$$
\begin{aligned}
&p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\gamma})p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} = \frac{p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta},\boldsymbol{\beta})p(\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\boldsymbol{\gamma})p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{\int p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})d\boldsymbol{\theta}} \\
&= c^{-1}\, p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta},\boldsymbol{\beta})p(\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\boldsymbol{\gamma})p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \\
&= c^{-1}\left(\prod_{r=1}^{N_m}\frac{\beta_r}{2\pi}\right)^{\frac{1}{2}}\left(\prod_{r=1}^{N_m}\frac{\gamma_r}{2\pi}\right)^{\frac{N_p}{2}}\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\left(\prod_{i=1}^{n}\alpha_i^{\frac{1}{2}}\right) exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\alpha_i\theta_i^2)\right. \\
&\qquad\left.-\frac{1}{2}\sum_{r=1}^{N_m}\left(\beta_r\cdot\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)-\frac{1}{2}\sum_{r=1}^{N_m}\left(\gamma_r\cdot\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)\right\}
\end{aligned}
\tag{6.7}
$$

Similarly, the MPVs of the variables cannot be directly obtained from Eq. (6.7) due to the high-dimensional evidence and unknown parameters. The VBI and DRAM algorithms is integrated to solve this problem, as detailed in the following section.

6.3 VBI-DRAM Algorithm

6.3.1 VBI

The mechanism of VBI is to approximate the target PDF by a proposed PDF. In this study, the target PDF to be approximated is the posterior PDF $p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, and the proposed PDF is symbolized as $\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$. For simplicity, the proposed PDF is denoted as $\boldsymbol{Q}$.

According to Eq. (5.13), the evidence can be calculated as

$$c = p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \frac{p(\hat{\boldsymbol{\lambda}}|\boldsymbol{\theta},\beta)p(\hat{\boldsymbol{\psi}}|\boldsymbol{\theta},\gamma)p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} = \frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} \tag{6.8}$$

Taking the logarithm of both sides in Eq. (6.8), one has

$$\begin{aligned} \ln p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) &= \ln\frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})} \\ &= \ln\frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{\boldsymbol{Q}} - \ln\frac{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{\boldsymbol{Q}} \end{aligned} \tag{6.9}$$

Taking the expectation of both sides in Eq. (6.9) with respect to $\boldsymbol{Q}$, we have

$$\begin{aligned} &\int \boldsymbol{Q}\ln p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\,d\boldsymbol{\theta} \\ &= \int \boldsymbol{Q}\ln\frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{\boldsymbol{Q}}\,d\boldsymbol{\theta} - \int \boldsymbol{Q}\ln\frac{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{\boldsymbol{Q}}\,d\boldsymbol{\theta} \end{aligned} \tag{6.10}$$

Considering that $\ln p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$ is irrelevant to $\boldsymbol{\theta}$, the left-hand side of Eq. (6.10) equals $\ln p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$, as the integral of $\boldsymbol{Q}$ with respect to $\boldsymbol{\theta}$ equals 1. Hence, Eq. (6.10) can be further simplified as

$$\ln p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \boldsymbol{\mathcal{L}}(\boldsymbol{Q}) + D_{KL}\{\boldsymbol{Q}||p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\} \tag{6.11}$$

where

$$\boldsymbol{\mathcal{L}}(\boldsymbol{Q}) = \mathbb{E}_{\boldsymbol{Q}}\left[\ln\frac{p(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}{\boldsymbol{Q}}\right] \tag{6.12}$$

$$D_{KL}\{\boldsymbol{Q}||p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})\} = \int \boldsymbol{Q}\ln\frac{\boldsymbol{Q}}{p(\boldsymbol{\theta}|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})}\,d\boldsymbol{\theta} \tag{6.13}$$

where $\mathbb{E}_{\boldsymbol{Q}}$ denotes the expectation with respect to $\boldsymbol{Q}$, $\boldsymbol{\mathcal{L}}(\boldsymbol{Q})$ represents the lower bound of $\boldsymbol{Q}$ and D_{KL} represents the KL divergence (Bishop, 2006) between $\boldsymbol{Q}$ and posterior PDF $p\left(\boldsymbol{\theta}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$.

It is demonstrated that $D_{KL} \geq 0$ and $D_{KL} = 0$ when $\boldsymbol{Q} = p\left(\boldsymbol{\theta}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$. Therefore, increasing the proximity of $\boldsymbol{Q}$ to $p\left(\boldsymbol{\theta}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$ can be attained by minimizing D_{KL}. However, the minimization of D_{KL} cannot be realized directly as the posterior PDF $p\left(\boldsymbol{\theta}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$ is unknown. According to Eq. (6.11), minimizing D_{KL} is equivalent to maximizing $\boldsymbol{\mathcal{L}}(\boldsymbol{Q})$. To decouple the posterior PDF of the damage index $\boldsymbol{\theta}$ and parameters $\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$, $\boldsymbol{Q}$ is further factorized into two items according to the mean field theory (Parisi, 1988), expressed as

$$\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = q(\boldsymbol{\theta})q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) \tag{6.14}$$

The maximization of $\boldsymbol{\mathcal{L}}(\boldsymbol{Q})$ can be attained by optimizing each factor in turn through calculating the expectation of the numerator $p\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}\middle|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$ in $\boldsymbol{\mathcal{L}}(\boldsymbol{Q})$ with respect to other factors. Therefore, $q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})$ can be obtained by computing the expectation of $\ln p\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}\middle|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$ with respect to $\boldsymbol{\theta}$, shown as

$$\begin{aligned}
&\ln q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \mathbb{E}_{\boldsymbol{\theta}}\left[\ln p\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}\middle|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)\right] + \text{const} \\
&= \mathbb{E}_{\boldsymbol{\theta}}\left[\ln p\left(\hat{\boldsymbol{\lambda}}\middle|\boldsymbol{\theta},\beta\right) + \ln p\left(\hat{\psi}\middle|\boldsymbol{\theta},\gamma\right) + \ln p(\boldsymbol{\theta}|\boldsymbol{\alpha})\right] + \text{const} \\
&= \frac{1}{2}\sum_{r=1}^{N_m}\ln\beta_r - \frac{1}{2}\sum_{r=1}^{N_m}\left\{\beta_r\cdot\mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)\right\} + \frac{N_p}{2}\sum_{r=1}^{N_m}\ln\gamma_r \\
&-\frac{1}{2}\sum_{r=1}^{N_m}\left\{\gamma_r\cdot\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|^2\right)\right\} + \frac{1}{2}\sum_{i=1}^{n}\ln\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\{\alpha_i\mathbb{E}(\theta_i^2)\} + \text{const}
\end{aligned} \tag{6.15}$$

Parameters $\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$ are assumed to be independent from each other and from mode to mode, resulting in the factorization as

$$q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = q(\boldsymbol{\alpha})q(\boldsymbol{\beta})q(\boldsymbol{\gamma}) \tag{6.16}$$

$$q(\boldsymbol{\alpha}) = \prod_{i=1}^{n} q(\alpha_i) \tag{6.17}$$

$$q(\boldsymbol{\beta}) = \prod_{r=1}^{N_m} q(\beta_r) \tag{6.18}$$

$$q(\boldsymbol{\gamma}) = \prod_{r=1}^{N_m} q(\gamma_r) \tag{6.19}$$

Therefore, we have

$$\ln q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}) = \sum_{i=1}^{n} \ln q(\alpha_i) + \sum_{r=1}^{N_m} \ln q(\beta_r) + \sum_{r=1}^{N_m} \ln q(\gamma_r) \tag{6.20}$$

According to Eq. (6.20), the logarithm of each posterior PDF is then derived as

$$\ln q(\alpha_i) = \frac{1}{2}\ln \alpha_i - \frac{1}{2}\alpha_i \mathbb{E}(\theta_i^2) + \text{const} \tag{6.21}$$

$$\ln q(\beta_r) = \frac{1}{2}\ln \beta_r - \frac{\beta_r}{2}\mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right) + \text{const} \tag{6.22}$$

$$\ln q(\gamma_r) = \frac{N_p}{2}\ln \gamma_r - \frac{\gamma_r}{2}\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right) + \text{const} \tag{6.23}$$

All parameters are found to follow the gamma distribution, expressed as

$$q(\alpha_i) \propto (\alpha_i)^{\frac{1}{2}} \cdot \exp\left\{-\frac{\alpha_i}{2} \cdot \mathbb{E}(\theta_i^2)\right\} \tag{6.24}$$

$$q(\beta_r) \propto (\beta_r)^{\frac{1}{2}} \cdot \exp\left\{-\frac{\beta_r}{2} \cdot \mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)\right\} \tag{6.25}$$

$$q(\gamma_r) \propto (\gamma_r)^{\frac{N_p}{2}} \cdot \exp\left\{-\frac{\gamma_r}{2} \cdot \mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)\right\} \tag{6.26}$$

Therefore, the statistics of each parameter are obtained as

$$\mathbb{E}(\alpha_i) = \frac{3}{E(\theta_i^2)};\ \text{Var}(\alpha_i) = \frac{6}{[\mathbb{E}(\theta_i^2)]^2} \tag{6.27}$$

$$E(\beta_r)=\frac{3}{\mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)}\ ;\ \mathrm{Var}(\beta_r)=\frac{6}{\left\{\mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)\right\}^2} \tag{6.28}$$

$$E(\gamma_r)=\frac{N_p+2}{\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)};\ \mathrm{Var}(\gamma_r)=\frac{2N_p+4}{\left[\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)\right]^2} \tag{6.29}$$

Factor $q(\boldsymbol{\theta})$ can be similarly optimized by computing the expectation of $\ln p\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}\middle|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)$ with respect to parameters $\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\}$, that is

$$\begin{aligned}
&\ln q(\boldsymbol{\theta})=\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}\left[\ln p\left(\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}},\boldsymbol{\theta}\middle|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}\right)\right]+\mathrm{const}\\
&=\mathbb{E}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}}\left[\ln p\left(\hat{\boldsymbol{\lambda}}\middle|\boldsymbol{\theta},\boldsymbol{\beta}\right)p\left(\hat{\boldsymbol{\psi}}\middle|\boldsymbol{\theta},\boldsymbol{\gamma}\right)p(\boldsymbol{\theta}|\boldsymbol{\alpha})\right]+\mathrm{const}\\
&=\mathbb{E}_{\boldsymbol{\beta}}\left[\ln p\left(\hat{\boldsymbol{\lambda}}\middle|\boldsymbol{\theta},\boldsymbol{\beta}\right)\right]+\mathbb{E}_{\boldsymbol{\gamma}}\left[\ln p\left(\hat{\boldsymbol{\psi}}\middle|\boldsymbol{\theta},\boldsymbol{\gamma}\right)\right]+\mathbb{E}_{\boldsymbol{\alpha}}[p(\boldsymbol{\theta}|\boldsymbol{\alpha})]+\mathrm{const}\\
&=-\frac{1}{2}\sum_{r=1}^{N_m}\left\{\mathbb{E}(\beta_r)\cdot\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\}-\frac{1}{2}\sum_{r=1}^{N_m}\left\{\mathbb{E}(\gamma_r)\cdot\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right\}\\
&\quad-\frac{1}{2}\sum_{i=1}^{n}\{\mathbb{E}(\alpha_i)\cdot\theta_i^2\}+\mathrm{const}
\end{aligned} \tag{6.30}$$

where items irrelevant to $\boldsymbol{\theta}$ are merged into the item of constant. Therefore, we have

$$\begin{aligned}
q(\boldsymbol{\theta})\propto\ \exp\Bigg\{&-\sum_{r=1}^{N_m}\left[\frac{\mathbb{E}(\beta_r)}{2}\cdot\left(\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right)^2\right]\\
&-\sum_{r=1}^{N_m}\left[\frac{\mathbb{E}(\gamma_r)}{2}\cdot\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right]-\sum_{i=1}^{n}\left[\frac{\mathbb{E}(\alpha_i)}{2}\cdot\theta_i^2\right]\Bigg\}
\end{aligned} \tag{6.31}$$

Eqs. (6.27)–(6.29) and (6.31) are coupled and thus can be calculated recursively. The iterative procedures correspond to the variational Bayesian expectation and maximization step and converge to determined values. Given that the EM algorithm does not guarantee the convergence of global optimum, it is recommended to assign different initializations to the parameters. Three expectations are required to be calculated, namely, $\mathbb{E}(\theta_i^2)$, $\mathbb{E}_{\boldsymbol{\theta}}\left(\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)$ and $\mathbb{E}_{\boldsymbol{\theta}}\left(\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)$, which is a full Bayesian treatment by taking posterior uncertainties of $\boldsymbol{\theta}$ into consideration. However, the statistics of $\boldsymbol{\theta}$ cannot be directly recognized from Eq. (6.31) due to the nonlinear relationship between $\boldsymbol{\theta}$ and the modal parameters, which hinders the iteration progress in the VBI. In next section the numerical DRAM algorithm is employed to obtain the statistics of $\boldsymbol{\theta}$.

6.3.2 DRAM Algorithm

The MH algorithm is an improved MCMC simulation that can be employed to generate numerical samples of the target PDF whose statistics cannot be obtained directly. Based on the proposed PDF S and the formulation $q(y)$ that is proportional to the target PDF, the basic MH algorithm are described as follows (Chib and Greenberg, 1995):

(1) Generate the candidate sample $y \sim S(\cdot\,|x^j)$, where x^j is the current sample,

(2) Compute the acceptance ratio of the candidate sample,

$$\xi(x^j,y)=\min\left\{1,\ \frac{q(y)S(y|x^j)}{q(x^j)S(x^j|y)}\right\}$$

(3) The next sample is

$$x^{j+1}=\begin{cases} y & \text{with the acceptance ratio } \xi(x^j,y) \\ x^j & \text{with the rejection ratio } 1-\xi(x^j,y) \end{cases}$$

The samples will converge to their equilibrium distribution after the burn-in period (Andrieu et al., 2003). However, when the MH algorithm is applied to high-dimensional problems, the rejection ratio is usually high due to the large sampling space. This will lead to a long burn-in period and a low sampling efficiency. To solve this problem, the DR algorithm (Green and Mira, 2001) has been developed to improve the acceptance ratio of sampling. When the candidate sample is supposed to be rejected, a secondary candidate sample is generated according to an adjusted proposal PDF instead of getting a repeated sample, and the acceptance ratio is recalculated. This process is termed DR and the procedures are shown as (Green and Mira, 2001):

(1) Generate the candidate sample $y_1 \sim S(\cdot\,|x^j)$, where x^j is the current sample,

(2) Compute the acceptance ratio of the candidate sample,

$$\xi_1(x^j, y_1) = \min\left\{1,\ \frac{q(y_1)S(y_1|x^j)}{q(x^j)S(x^j|y_1)}\right\}$$

(3) The next sample is

$$x^{j+1} = \begin{cases} y_1 \text{ and go to Step (7)} & \text{with the acceptance ratio } \xi_1(x^j, y_1) \\ \text{go to Step (4)} & \text{with the rejection ratio } 1 - \xi_1(x^j, y_1) \end{cases}$$

(4) Generate the candidate sample $y_2 \sim S(\cdot\,|x^j, y_1\,)$,

(5) Compute the acceptance ratio of the secondary candidate sample,

$$\xi_2(x^j, y_1, y_2) = \min\left\{1,\ \frac{q(y_2)S(y_1|y_2)S(\mathrm{x}^j|y_1, y_2)(1 - \xi_1(y_2, y_1))}{q(x^j)S(y_1|x^j)S(y_2|x^j, y_1)(1 - \xi_1(x^j, y_1))}\right\}$$

(6) The next sample is

$$x^{j+1} = \begin{cases} y_2 & \text{with the acceptance ratio } \xi_2(x^j, y_1, y_2) \\ x^j & \text{with rejection ratio } \xi_2(x^j, y_1, y_2) \end{cases}$$

(7) End

In addition, the proposed PDF also will influence the sampling efficiency. The higher the proximity of the proposed PDF to the target PDF, the higher the acceptance ratio will be. The Gaussian distribution is adopted as the proposed PDF due to its operability and large entropy. In the MH algorithm, the mean of the Gaussian distribution is consistently adjusted with the generated sample, whereas the variance is generally determined according to experience. The value of the variance is also important. A small variance will lead to a slow convergence, especially when the sample severely deviates from the equilibrium distribution, and a large variance will result in a high sampling rejection ratio. In this connection, the AM algorithm is developed to automatically adjust the variance using the generated samples. A certain number of samples are required before the adjustment to ensure the semi-positive definite property of the covariance matrix. The adjustment rule is (Haario et al., 1999, 2001)

$$C = \begin{cases} C, j + 1 < N_t \\ S_d \operatorname{cov}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(j+1)}) + S_d \varepsilon \boldsymbol{I}_d, j + 1 \geq N_t \end{cases} \tag{6.32}$$

where C is the covariance matrix that is initialized based on experience, S_d is a parameter related to the dimension of the covariance matrix and is generally set to $2.4^2/d$ (d is the dimension of $\boldsymbol{\theta}$) (Haario et al., 2006), ε is an extremely small constant, $\boldsymbol{I}_d$ is an identity matrix, and N_t is the bound number of samples before adjusting the covariance matrix.

In view of the advantages of DR and AM algorithms, researchers have combined them to generate a novel algorithm named the DRAM algorithm (Haario et al., 2006). The DRAM algorithm is applicable to all kinds of complicated probabilistic distributions (etc. standard and non-standard probabilistic distributions), provided that the formulation proportional to the target PDF is available. The DRAM algorithm has a high efficiency for high-dimensional problems (Wan and Ren, 2016).

In this study, the DRAM algorithm is employed to generate samples of $\boldsymbol{\theta}$ using Eq. (6.31). A two-layer delayed rejection scheme is adopted in this study. Specifically, if the secondary sample is rejected, a repeated sample is adopted and the third-layer sampling is not carried out any more. A Gaussian distribution $\mathcal{N}(\mu, C)$ is adopted as the proposed PDF. To accelerate the convergence and shorten the burn-in period, the mean of the proposed PDF, μ, is initially set as $\overline{\boldsymbol{\theta}}$ that is calculated by minimizing the objective function in Eq. (5.14). Both the values of the mean μ and covariance matrix C are adjusted with the progress of sampling. The procedures are described as follows:

DRAM algorithm

Given the proposed sampling PDF $\mathcal{N}(\mu, C)$, the first sample $\boldsymbol{\theta}^{(1)}$, the scale factor ρ to adjust the covariance matrix ($\rho < 1$), the bound N_t, the trivial constant ε, and the number of samples n_s

1. Generate the candidate sample $x_1 \sim \mathcal{N}(\mu, C)$.
2. Compute the acceptance ratio of the candidate sample,

$$\xi_{j1}\left(\boldsymbol{\theta}^{(j)}, x_1\right) = \min\left\{1, \frac{q(x_1)\mathcal{N}_{\boldsymbol{\theta}^{(j)}}(x_1, C)}{q(\boldsymbol{\theta}^{(j)})\mathcal{N}_{x_1}(\boldsymbol{\theta}^{(j)}, C)}\right\}$$

```
3. Randomly generate $\boldsymbol{\mu}$ from the uniform distribution $\mathcal{U}(0,1)$.
4. If $\boldsymbol{\mu} < \xi_{j1}(\boldsymbol{\theta}^{(j)}, x_1)$, let $\boldsymbol{\theta}^{(j+1)} = x_1$ and go to Step 8. Otherwise, go to Step 5.
5. Generate the secondary candidate sample $x_2 \sim \mathcal{N}(\mu, \rho C)$.
6. Compute the acceptance ratio of the secondary candidate sample,
   $$\xi_{21}(x_2, x_1) = \min\left\{1, \frac{q(x_1)\mathcal{N}_{x_2}(x_1, C_0)}{q(x_2)\mathcal{N}_{x_1}(x_2, C_0)}\right\}$$
   $$\xi_{j2}(\boldsymbol{\theta}^{(j)}, x_2) = \min\left\{1, \frac{q(x_2)\mathcal{N}_{x_1}(x_2, C)\mathcal{N}_{\boldsymbol{\theta}^{(j)}}(x_2, C)[1 - \xi_{21}(x_2, x_1)]}{q(\boldsymbol{\theta}^{(j)})\mathcal{N}_{x_1}(\boldsymbol{\theta}^{(j)}, C)\mathcal{N}_{x_2}(\boldsymbol{\theta}^{(j)}, C)[1 - \xi_{j1}(\boldsymbol{\theta}^{(j)}, x_1)]}\right\}$$
7. If $\boldsymbol{\mu} < \xi_{j2}(\boldsymbol{\theta}^{(j)}, x_2)$, then $\boldsymbol{\theta}^{(j+1)} = x_2$. Otherwise, $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)}$.
8. Adjust the sampling covariance
   $$C = \begin{cases} C, j + 1 < N_t \\ S_d \operatorname{cov}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(j+1)}) + S_d \varepsilon \boldsymbol{I}_d, j + 1 > N_t \end{cases}$$
9. Let $\mu = \boldsymbol{\theta}^{(j+1)}$.
10. Let $j = j + 1$, repeat Steps 1–8 until $j = (n_s - 1)$.
11. The samples following the posterior PDF of $\boldsymbol{\theta}$ are obtained, $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(n_s)})$. The MPV and uncertainty of $\boldsymbol{\theta}$ are calculated according to these samples.
```

6.3.3 Summary of the VBI-DRAM Method

The procedure of the proposed method is illustrated as Figure 6.1, which begins with the initialization of parameters and progresses by iteratively updating all unknowns. The damage index $\boldsymbol{\theta}$ is sampled using the DRAM algorithm according to Eqs. (5.14) and (6.31), and $\boldsymbol{\theta}_{\text{MAP}}$ is calculated from the generated samples. The parameters $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ are updated according to Eqs. (6.27)–(6.29). These procedures are carried out iteratively until the defined convergence criterion is satisfied.

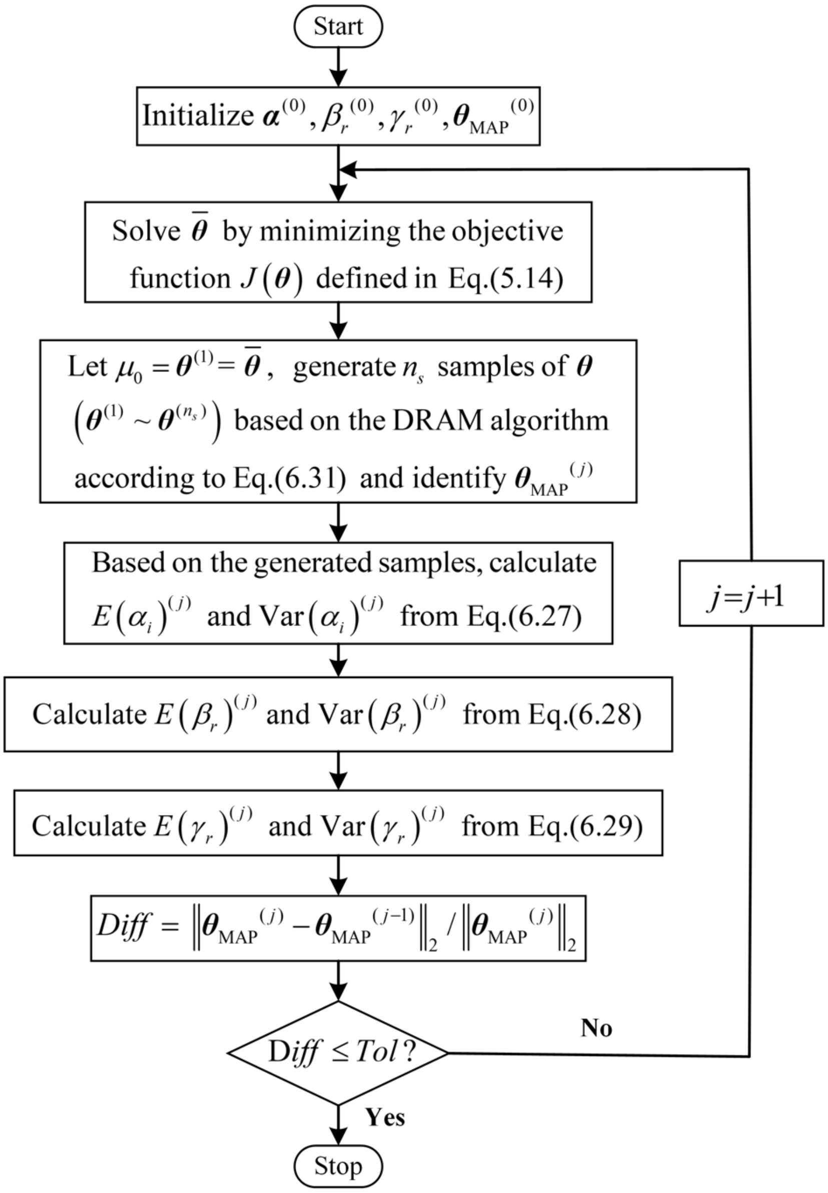


Figure 6.1 Flowchart of the proposed VBI-DRAM method

6.4 Case Study

6.4.1 Model Description

The proposed algorithm is applied to a laboratory-tested two-story steel frame for damage identification. As shown in Figure 6.2, the frame is fixed on the strong floor through a steel plate. The frame is 1.0 m high with two equal stories of 0.5 m high. The span is 0.5 m. The cross-section of all beams and columns is 50.0×5.0 mm^2. The mass density measures 7.92×10^3 kg/m^3, and the Young's modulus is estimated to be 2.0×10^{11} N/m^2. The frame is modeled by 150 Euler–Bernoulli beam elements, each being 20 mm long.

Figure 6.2 Overview of the two-story steel frame

The frame was excited by a hammer with a rubber tip. Accelerometers were installed at 26 locations to collect the vibration response, as shown in Figure 6.3. The first eight frequencies and mode shapes were then extracted using the rational fraction polynomial method (Formenti and Richardson, 2002).

Three cuts were then introduced to the frame at different locations in sequence. Specifically, Cut 1 was located at the clamped end, that is, element No.1. Cuts 2 and 3 were introduced to the beam-column joint and the middle of the bottom column, that is, elements No. 126 and No. 12, respectively. Three cuts had the same length of 20 mm and different depths of 10 mm, 15 mm and 15 mm. Considering that the cuts have identical length as the element, the damage severity is estimated to be the reduction in the moment of inertia of the cross-section. Hence, the SRFs of elements No. 1, 126 and 12 are 40%, 60% and 60%, respectively, that is, $\theta_1 = -40\%$, $\theta_{126} = -60\%$ and $\theta_{12} = -60\%$. The damage locations and severities in the three DSs are summarized in Table 6.1.

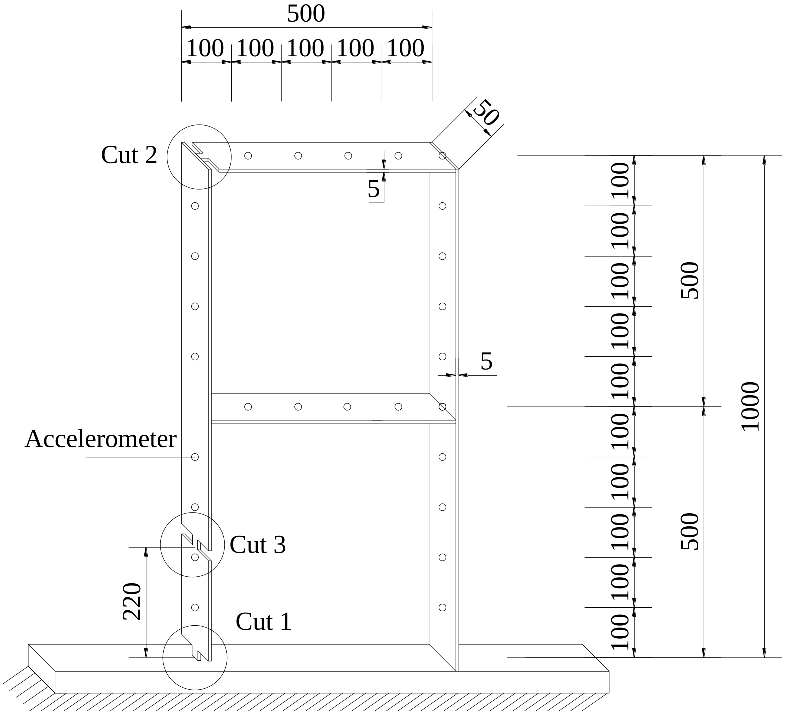


Figure 6.3 Damage locations and sensor arrangements

Table 6.1 Damage locations and severities in three DSs

Damage Scenario	Cut No.	Element No.	Cut depth (mm)	Damaged component	SRF ($\boldsymbol{\theta}$)
DS1	Cut 1	1	10	Column	$\theta_1 = -40\%$
DS2	Cut 1	1	10	Column	$\theta_1 = -40\%$
	Cut 2	126	15	Beam	$\theta_{126} = -60\%$
DS3	Cut 1	1	10	Column	$\theta_1 = -40\%$
	Cut 2	126	15	Beam	$\theta_{126} = -60\%$
	Cut 3	12	15	Column	$\theta_{12} = -60\%$

The hammer test was similarly conducted in each DS to obtain the first eight modal frequencies and mode shapes, as compared in Table 6.2. The sixth mode was observed to be very close to the seventh mode, which may be due to modal identification errors. In this connection, these two modes were not used. Only the first five and eighth modes were used in the subsequent damage identification.

Table 6.2 Modal data of the frame in undamaged state and three DSs (units: Hz)

Mode	Undamaged	DS1		DS2		DS3	
		Freq. (HZ)	MAC	Freq. (HZ)	MAC	Freq. (HZ)	MAC
1	6.27	6.19(−1.39)	0.99	6.17(−1.60)	0.91	6.16(−1.78)	0.96
2	20.64	20.45(−0.92)	0.99	20.16(−2.33)	0.99	20.17(−2.28)	0.87
3	44.35	44.17(−0.40)	0.98	43.89(−1.03)	0.96	43.94(−0.90)	0.93
4	62.63	62.41(−0.36)	0.99	61.69(−1.51)	0.99	61.40(−1.96)	0.98
5	71.29	70.62(−0.93)	0.97	69.98(−1.84)	0.97	69.60(−2.37)	0.95
8	106.52	105.6(−0.90)	0.94	105.1(−1.32)	0.91	104.2(−2.21)	0.88
Average (%)		(−0.82)	0.98	(−1.61)	0.96	(−1.92)	0.93

Note. (1) Values in parentheses are the frequency change ratios (%) between the damaged and undamaged states.

(2) MAC refers to the Modal Assurance Criterion of the mode shapes in the damaged and undamaged states

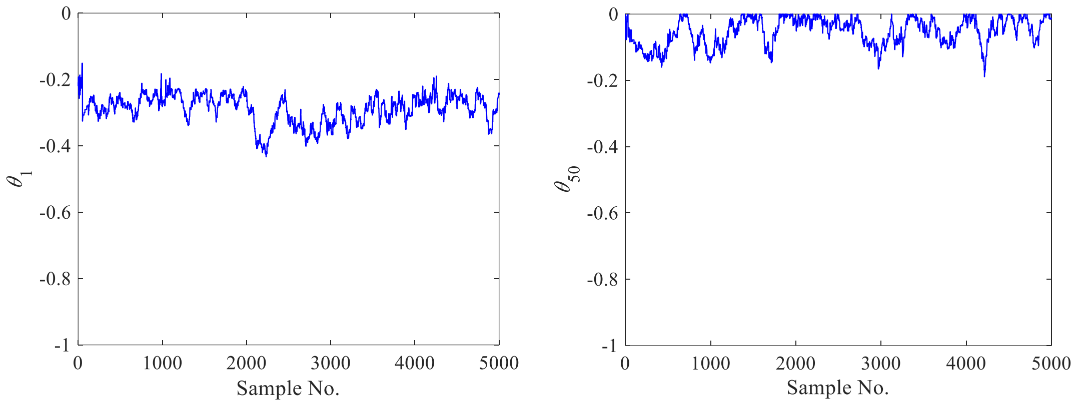
6.4.2 Damage Identification

The proposed VBI-DRAM algorithm requires initializing the parameters first. As previously mentioned, the global optimum upon convergence is not ensured in the VBI, thereby different initializations of parameters are recommended. Nevertheless, the DRAM algorithm is a global technique which compensates for the limitation of VBI. Specifically, the DRAM algorithm generates samples from the entire sampling space and is applicable to both standard and non-standard distributions. Therefore, $\boldsymbol{\theta}_{\mathrm{MAP}}$ computed on the basis of the samples generated by the DRAM algorithm is supposed to be the global maximum of the posterior PDF theoretically (more rigorously, near the global maximum due to the randomness of the samples). As defined, parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are associated with the uncertainties of the measured modal parameters. Therefore, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are supposed to converge to determined values under different initializations, which will be investigated later. The uncertainty levels of the damage index, frequency and mode shape are initially assumed to be 10%, 1% and 5%, respectively. All entries in one vector are initialized identically, that is, $\alpha_i^{(0)}$=1/(10%)^2 =100 (i =1, 2,…, 150),

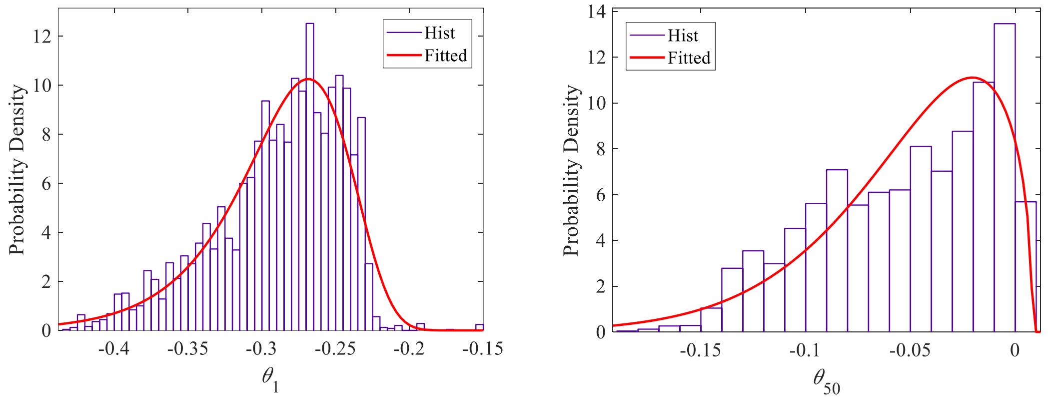
$\beta_r{}^{(0)}$=1/(1%)^2 =1×10^4 (r =1, 2,…, 5, 6), and $\gamma_r{}^{(0)}$=1/(5%)^2 =400 (r =1, 2,…, 5, 6). $\boldsymbol{\theta}_{\text{MAP}}$ is initialised to $\mathbf{0}$, that is, $\boldsymbol{\theta}_{\text{MAP}}{}^{(0)} = \mathbf{0}$.

$\overline{\boldsymbol{\theta}}$ is firstly calculated by substituting the initialized parameters to Eq. (6.14). The DARM algorithm is then applied to sample $\boldsymbol{\theta}$ according to the target PDF in Eq. (6.31). The initial sample $\boldsymbol{\theta}^{(1)}$ and mean μ of the proposed PDF are set to $\overline{\boldsymbol{\theta}}$, that is, $\boldsymbol{\theta}^{(1)} = \mu = \overline{\boldsymbol{\theta}}$. The covariance matrix C in Step 1 is initialized to a diagonal matrix with the diagonal entry of 0.05, which will be adjusted in Step 8 as the sampling process proceeds. The bound N_t is set to 100. ε is set to 10^{-5} to ensure the semi-definiteness of the covariance matrix. The number of samples to be generated is set to 5000, that is, $n_s = 5000$. $\boldsymbol{\theta}_{\text{MAP}}$ and three expectations $E(\theta_i^2)$, $E\left(\left[\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right)$ and $E\left(\left\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right)$ are then calculated on the basis of the generated samples. According to Eqs. (6.27)–(6.29), parameters $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ are then updated. The recursive procedures are then repeated until the convergence criterion is met, that is, $Tol \leqslant$0.02. Upon convergence, the damage index and parameters are valued as $\boldsymbol{\theta} = \boldsymbol{\theta}_{\text{MAP}}$, $\alpha_i = E(\alpha_i)$, $\beta_r = E(\beta_r)$ and $\gamma_r = E(\gamma_r)$.

DS1. In DS1, the actual damage is $\theta_1 = -40\%$. In the first iteration step, 5,000 samples of $\boldsymbol{\theta}$ are generated by the DRAM algorithm, and the probability histograms are plotted in Figure 6.4. Due to the space limitation, only the samples of θ_1 (the damaged element) and θ_{50} (the undamaged element) are presented here. As observed in Figures 6.4(a) and 6.4(b), the burn-in period in the sampling process is very short, and the samples quickly converge to the region nearby the actual damage severity. The may be because the mean of the proposed PDF and the initial sample in the sampling process are set to $\overline{\boldsymbol{\theta}}$, which is computed by minimizing the objective function in Eq. (6.14), thus the effective sampling is accelerated. The corresponding probability histograms and fitted PDF curves are plotted in Figures 6.4(c) and 6.4(d), respectively. The posterior PDFs of the damage indices are found to be non-standard distributions.



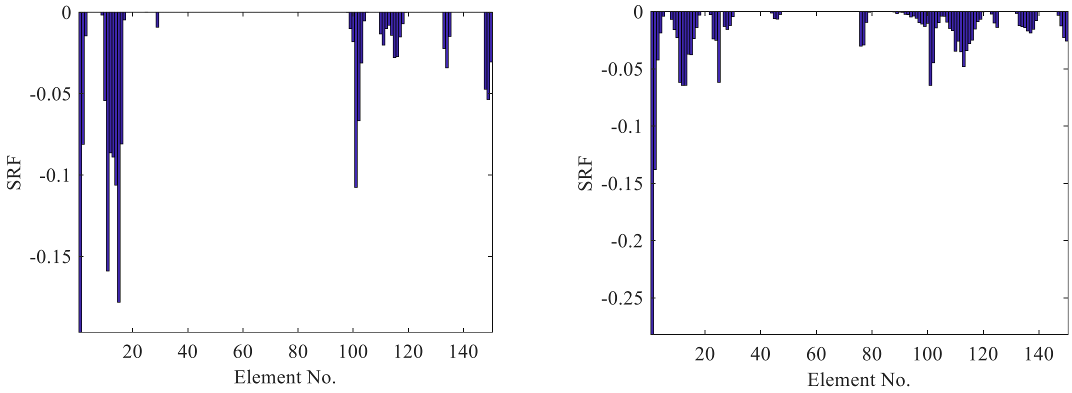
(a) Samples of θ_1 (b) Samples of θ_{50}



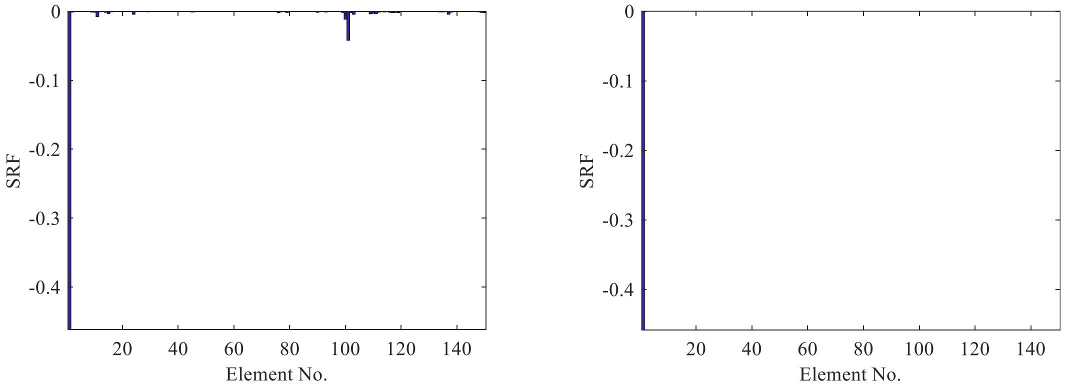
(c) Posterior PDF of θ_1 (d) Posterior PDF of θ_{50}

Figure 6.4 Samples and posterior PDFs in the first iteration (DS1)

As the iteration proceeds, the damage detection results are shown in Figure 6.5. For brevity, only the results in the first two and last two iterations are presented. A few elements are mistakenly identified as damaged in the beginning, as shown in Figure 6.5(a). The convergence is attained after five iterations. The final damage detection results are shown in Figure 6.5(d). The damage is accurately identified with no false identification. The samples of θ_1 and θ_{50} and their probability distributions in the final iteration are plotted in Figure 6.6. Again the posterior PDFs are not standard Gaussian distributions.

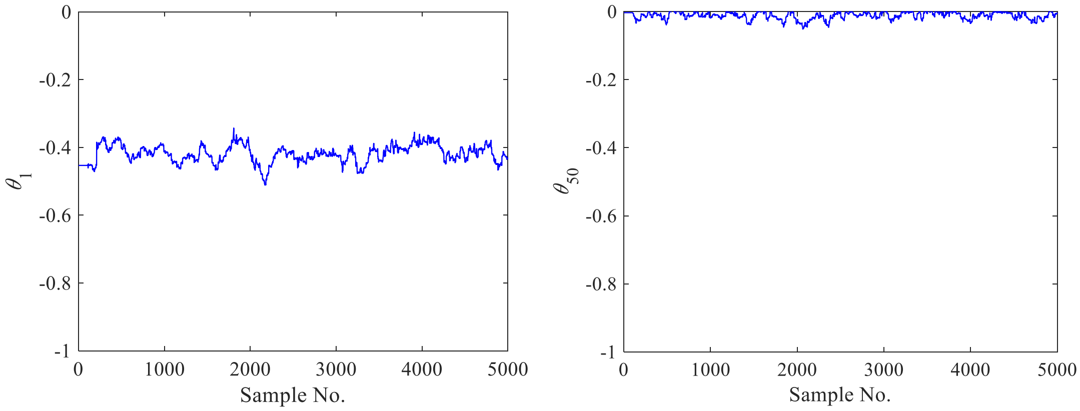


(a) Iteration No. 1 (b) Iteration No. 2

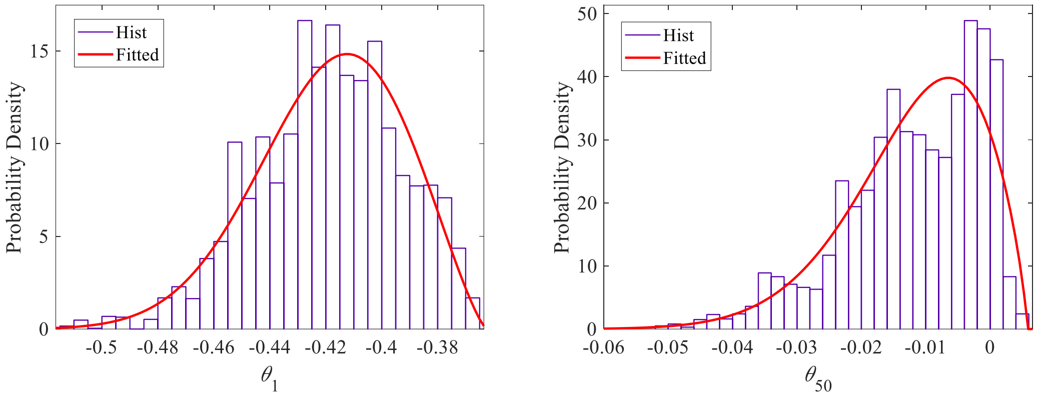


(c) Iteration No. 4 (d) Iteration No. 5

Figure 6.5 Damage identification results in DS1



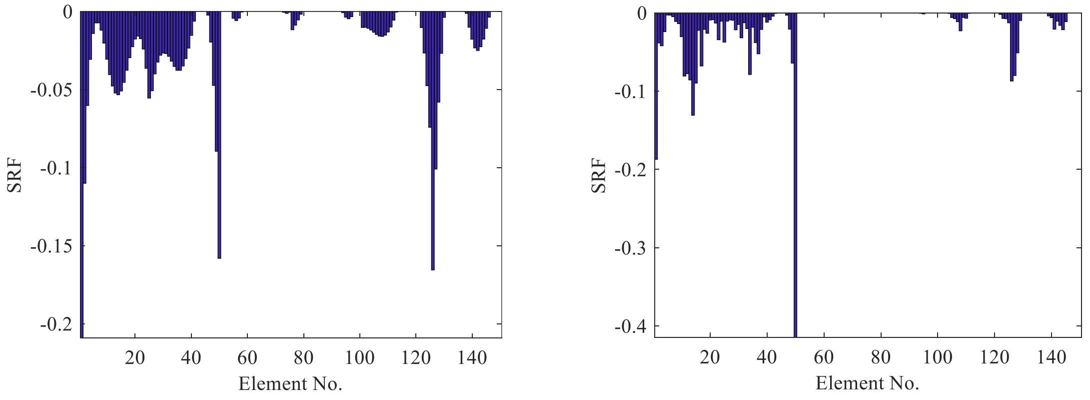
(a) Samples of θ_1 (b) Samples of θ_{50}



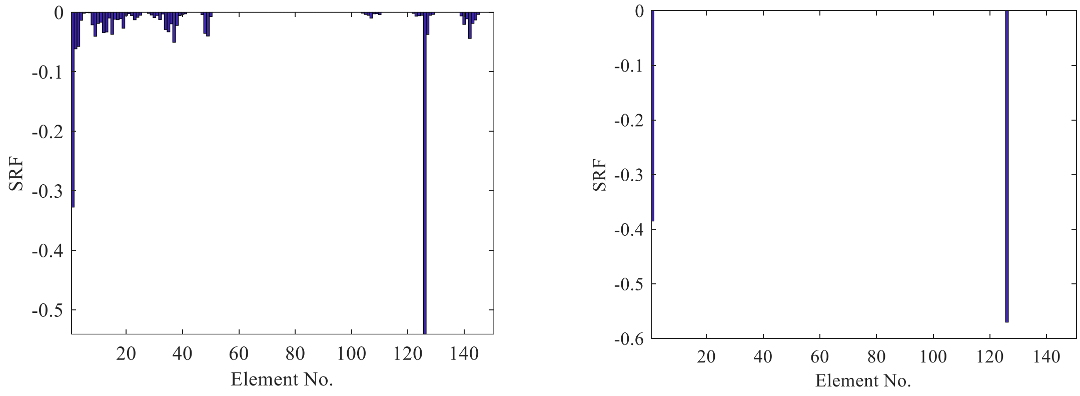
(c) Posterior PDF of θ_1 (d) Posterior PDF of θ_{50}

Figure 6.6 Samples and posterior PDFs in the final iteration (DS1)

DS2. The actual damage in DS2 are $\theta_1 = -40\%$ and $\theta_{126} = -60\%$. The proposed VBI-DRAM algorithm is similarly applied to the data in DS2. The samples and probability distributions are not presented here for brevity. The damage detection results during the iteration process are plotted in Figure 6.7. The converge is achieved after six iterations. Notably, element No. 50 is falsely identified as damaged in the second iteration. The reason may be that the element is located at the top of the left column and is adjacent to the damage element No. 126. This error disappears upon convergence. The two damages are accurately identified with no false identification, as shown in Figure 6.7 (d).



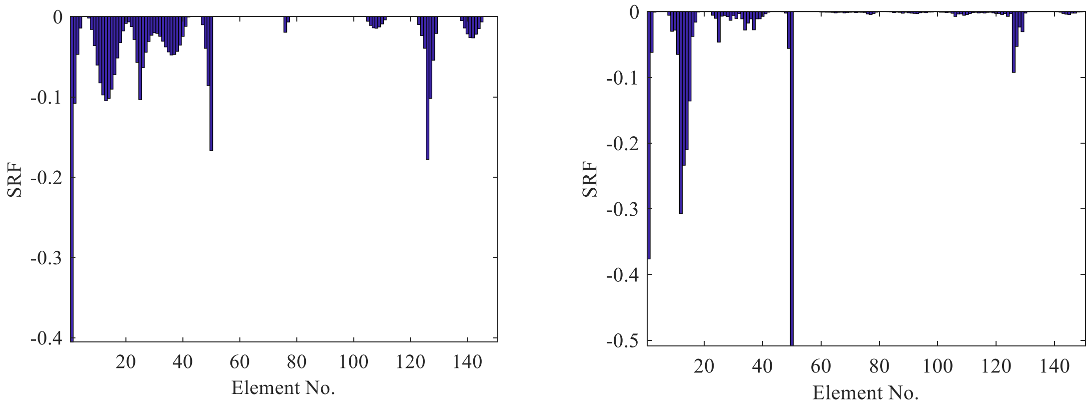
(a) Iteration No. 1 (b) Iteration No. 2



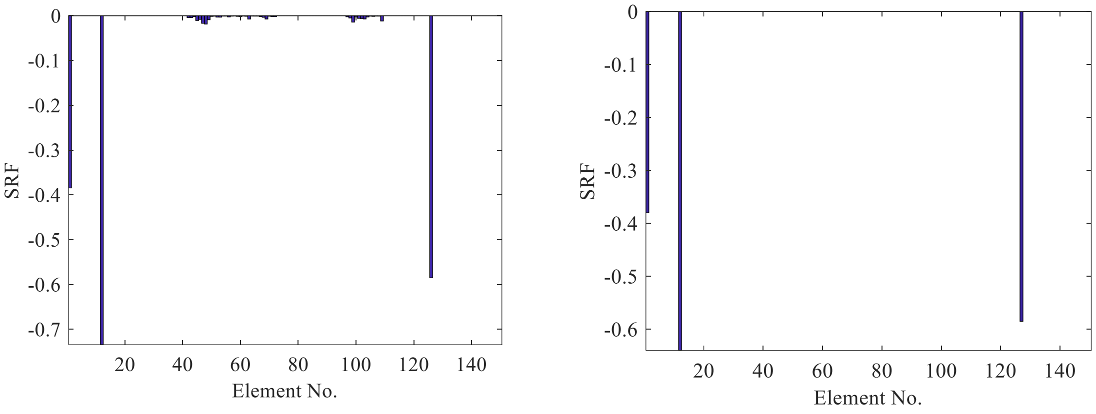
(c) Iteration No. 5 (d) Iteration No. 6

Figure 6.7 Damage identification results in DS2

DS3. Three damages exist in DS3, that is, $\theta_1 = -40\%$, $\theta_{126} = -60\%$ and $\theta_{12} = -60\%$. The damage identification results in each iteration are presented in Figure 6.8. The identification results converge after seven iterations. Three damage locations are correctly identified, and the damage severities are accurately identified, as depicted in Figure 6.8(d). Figure 6.9 shows the samples of θ_1 and θ_{50} in the fourth iteration. Although the initial sample $\overline{\boldsymbol{\theta}}$ deviates far from the global minimum of the objective function in Figure 6.9 (a), the samples converge to the equilibrium distribution after a short burn-in period. The DRAM sampling technique is efficient in the high-dimensional problem.

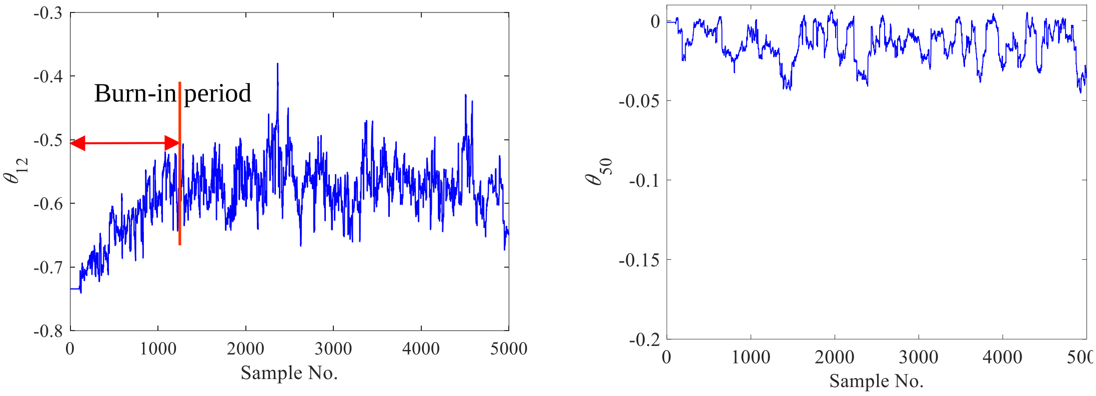


(a) Iteration No. 1 (b) Iteration No. 2



(c) Iteration No. 6
(d) Iteration No. 7

Figure 6.8 Damage identification results in DS3



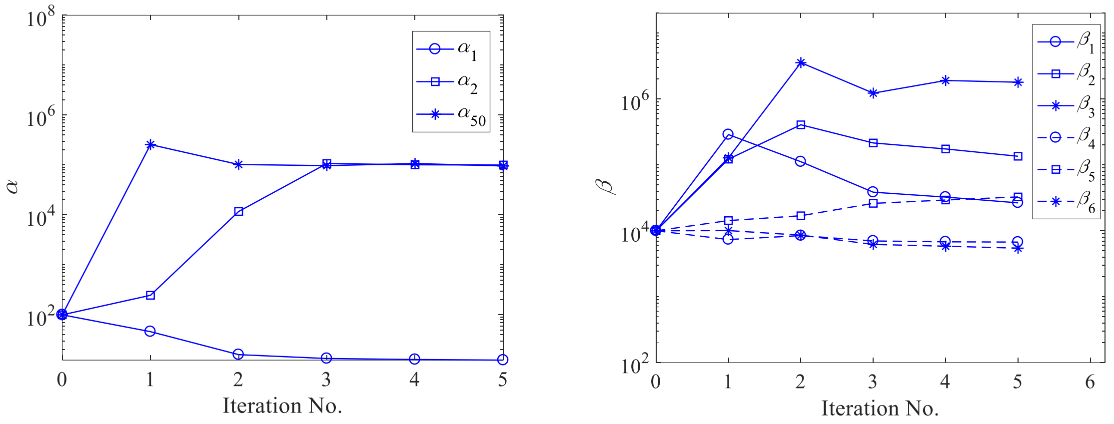
(a) Samples of θ_{12}
(b) Samples of θ_{50}

Figure 6.9 Samples in the fourth iteration (DS3)

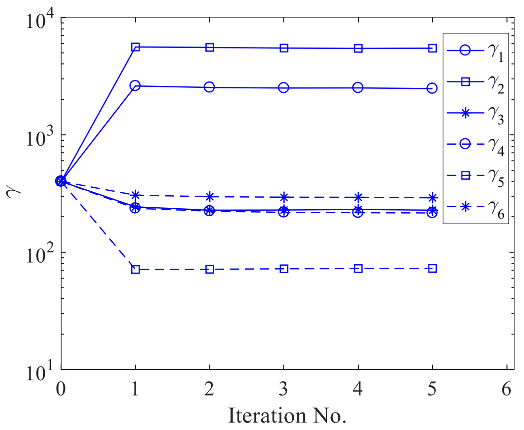
6.4.3 Variations of Parameters

The parameters also change as the iteration process progresses. Figure 6.10 shows the variations of $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ in DS1. Particularly, although all α_i (i = 1, 2, …, 150) are initialized with the identical value, α_1 converges to a small value close to 1, whereas the remaining α_i (i =2, 3, …, 150) increase significantly, enforcing the corresponding damage indices of the undamaged elements to zero in Eq. (5.14) and resulting in a sparse SRF result. The variations of parameters β_r and γ_r in DS1 are plotted in Figures

6.9(b) and 6.9(c). These parameters are different, reflecting the different uncertainties of each mode.



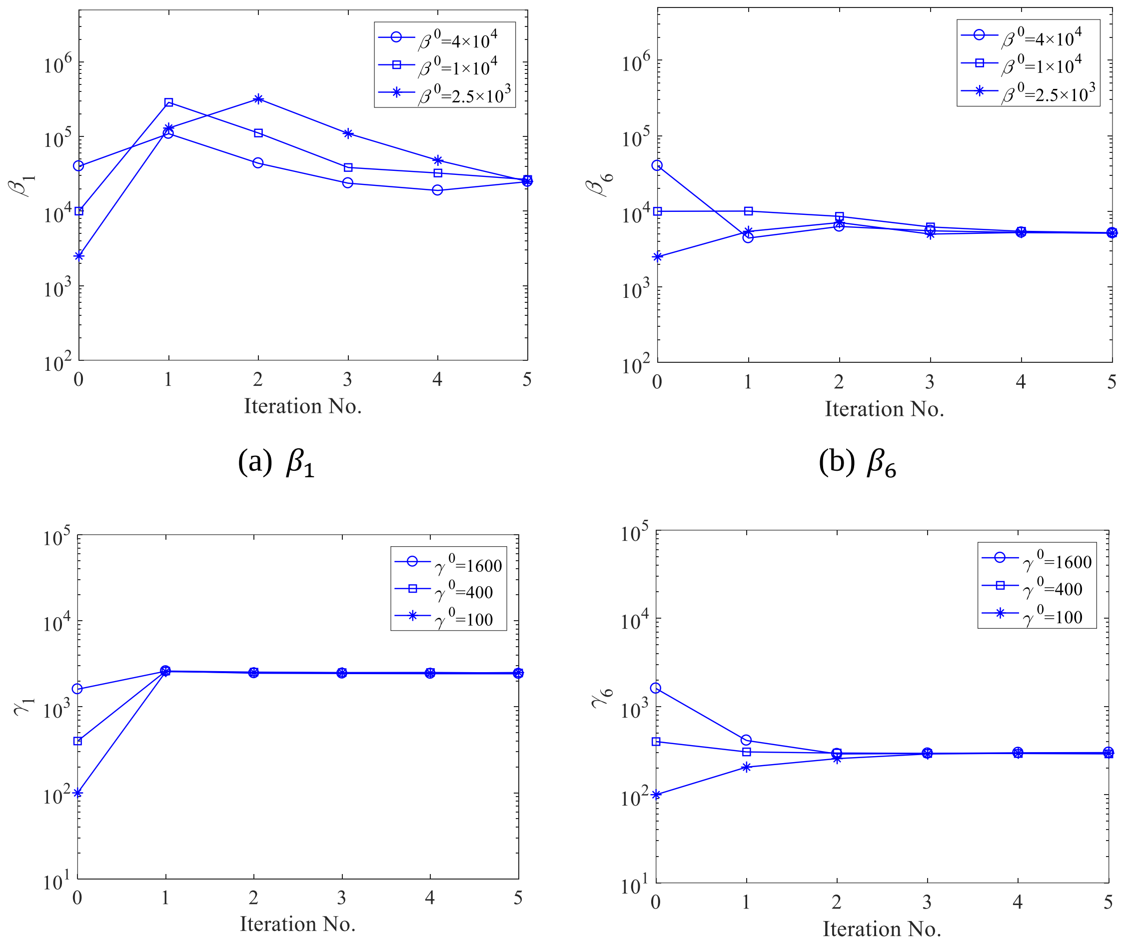
(a) α (b) β



(c) γ

Figure 6.10 Variation of parameters in DS1 during the iteration process

In addition, the performance of the proposed VBI-DRAM algorithm with different initializations is also investigated. Two extra groups of initializations are assigned to the parameters in DS1, that is, $\beta^{(0)} = 1/(2\%)^2 = 2.5\times10^3$, $\gamma^{(0)} = 1/(10\%)^2 = 100$ and $\beta^{(0)} = 1/(0.5\%)^2 = 4\times10^4$, $\gamma^{(0)} = 1/(2.5\%)^2 = 1600$. The iterative values of $\beta_1, \beta_6, \gamma_1$ and γ_6 under different initializations are compared in Figure 6.11. The results indicate that the parameters converge to the identical values even being initialized differently. Specifically, $\beta_1 = 2.8\times10^4$, $\beta_6 = 5.2\times10^3$, $\gamma_1 = 2.5\times10^3$ and $\gamma_6 = 2.9\times10^2$. In corresponding, the uncertainties of the first and eighth eigenvalues are 0.6% and 1.4%, respectively, and those of modal shapes are 2.0% and 5.9%, respectively.



(a) β_1 (b) β_6

(c) γ_1 (d) γ_6

Figure 6.11 Parameters with different initializations

6.5 Comparison with EM and Laplace Approximation Techniques

A SBL method based on the EM technique is proposed by Hou et al. (2019) where the evidence is maximized through expectation and maximization steps. In short, in the E step the expectation of the logarithm of the complete-data likelihood function is solved as

$$E\{\ln p\left(\boldsymbol{\theta},\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}\middle|\boldsymbol{\xi}\right)\} \\ = \frac{N_m}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}E\left\{\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\} + \frac{N_p\cdot N_m}{2}\ln\left(\frac{\gamma}{2\pi}\right) \tag{6.33}$$

$$-\frac{\gamma}{2}E\left\{\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right\}+\frac{n}{2}\ln\left(\frac{1}{2\pi}\right)+\frac{1}{2}\sum_{i=1}^{n}\ln\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\alpha_i E(\theta_i^2)$$

In the M step, Eq. (6.33) is maximized with respect to $\boldsymbol{\alpha}$, β, and γ. By setting the derivative of Eq. (6.33) with respect to each parameter to zero, one has

$$\frac{\partial E\left\{\ln p\left(\boldsymbol{\theta},\boldsymbol{\xi}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}\right)\right\}}{\partial\alpha_i}=\frac{1}{2\alpha_i}-\frac{1}{2}E(\theta_i^2)=0 \tag{6.34}$$

$$\frac{\partial E\left\{\ln p\left(\boldsymbol{\theta},\boldsymbol{\xi}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}\right)\right\}}{\partial\beta}=\frac{N_m}{2}\frac{1}{\beta}-\frac{1}{2}\,E\left\{\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\}=0 \tag{6.35}$$

$$\frac{\partial E\left\{\ln p\left(\boldsymbol{\theta},\boldsymbol{\xi}\middle|\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\psi}}\right)\right\}}{\partial\gamma}=\frac{N_p\cdot N_m}{2}\frac{1}{\gamma}-\frac{1}{2}E\left\{\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|_2^2\right\}=0 \tag{6.36}$$

The MPV of each parameter is then solved as

$$\alpha_i=\frac{1}{E(\theta_i^2)} \tag{6.37}$$

$$\beta=\frac{N_m}{E\left\{\sum_{r=1}^{N_m}\left[\frac{\hat{\lambda}_r-\lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r}\right]^2\right\}} \tag{6.38}$$

$$\gamma=\frac{N_p\cdot N_m}{E\left\{\sum_{r=1}^{N_m}\left\|\hat{\boldsymbol{\phi}}_r-\boldsymbol{\phi}_r(\boldsymbol{\theta})\right\|^2\right\}} \tag{6.39}$$

The damage index $\boldsymbol{\theta}$ is similarly calculated by minimizing the objective function Eq. (5.14). Since the expectations in Eqs. (6.37)–(6.39) cannot be calculated directly, Hou et al. (2020) proposed to use the likelihood numerical sampling technique.

The performances of the Laplace approximation method in Chapter 5, the VBI-DRAM algorithm in this chapter, and the EM algorithm in Hou et al. (2019) are compared by applying them to a numerical beam and the experimental frame in Section 5.4.

6.5.1 Numerical Example

A numerical cantilever beam in Figure 6.12 is first utilized for comparison. The mass density and Young's modulus are 7.67×10^3 kg/m^3 and 7.0×10^{10} N/m^2, respectively. The beam is modeled with 45 equal Euler–Bernoulli beam elements (i.e., n=45), 20-mm long each. Damage is simulated by the reduction of the bending stiffness while mass remains unchanged. Element 1 at the clamped end and Element 23 at the mid-span are damaged by 50% (i.e., $\theta_1 = \theta_{23} = -0.5$). The first six natural frequencies are used for damage identification. Table 6.3 lists the frequencies in the undamaged and damaged conditions. A noise level of 1% is assigned to the frequencies.

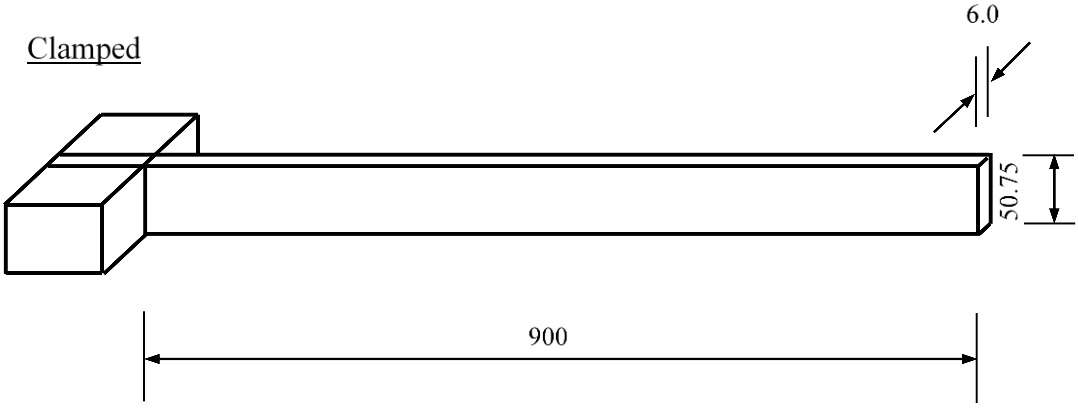
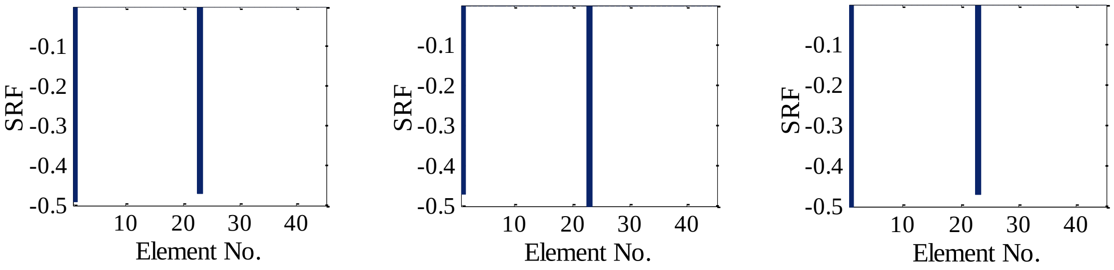


Figure 6.12 Geometric configuration of the beam structure (unit: mm)

Table 6.3 Frequencies of the beam in the undamaged and damaged states

Mode no.	Undamaged Freq. (Hz)	Damaged Freq. (Hz)	Change ratio (%)
1	6.02	5.75	−4.56
2	37.75	35.67	−5.50
3	105.73	102.44	−3.11
4	207.25	197.69	−4.61
5	342.70	333.96	−2.55
6	512.07	492.45	−3.83
Average of frequency change (%)			−4.03

Damage index and hyper-parameters should be initialized first. The initializations are the same, that is, $\beta^{(0)}=1/(1\%)^2=1\times10^4$ and $\alpha_i^{(0)}=1/(10\%)^2=100\ (i=1,2,\ldots,45)$. Initial damage indexes are set as $\boldsymbol{\theta}^{(0)}=\{0,\ldots,0\}^T$, indicating a lack of damage. Given that only frequencies are used for model updating, the items associated with mode shapes in the objective function Eq. (5.14) are dropped. The convergence criterion is set as $Tol=0.01$. The damage identification results of three methods are plotted in Figure 6.13.



(a) Laplace approximation (b) VBI-DRAM algorithm (c) EM algorithm

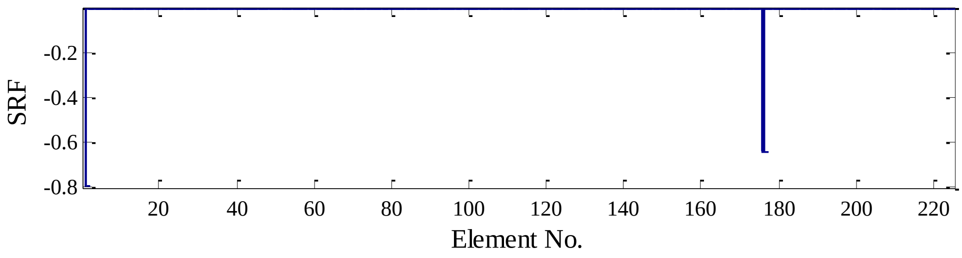
Figure 6.13 Damage identification results of the beam

The identification errors of Laplace approximation, VBI-DRAM algorithm and EM algorithm are calculated to be 0.45%, 0.42% and 0.41% according to Eq. (5.31), respectively. The corresponding computation times are 116 s, 319 s and 5082 s, respectively, which are obtained using a PC with Intel Core I7-8700 CPU and 20 GB RAM. Laplace approximation costs the least computational time since no sampling is required. The VBI-DRAM algorithm that conducts the posterior sampling is efficient than the EM technique that conducts the likelihood sampling.

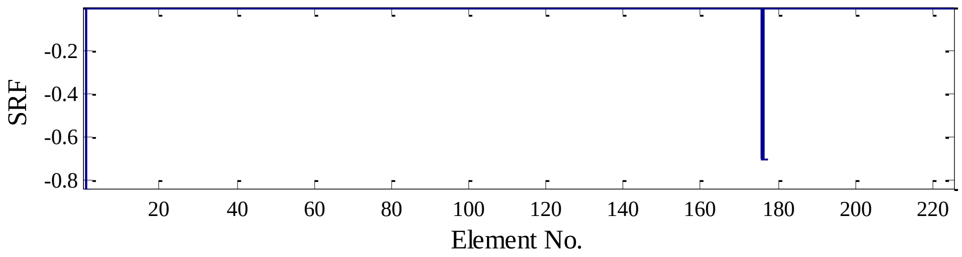
6.5.2 Experimental Study

The three-story steel frame in Section 5.4 is employed here as the experimental study. The frame is modeled by 225 Euler-Bernoulli beam elements as described in Section

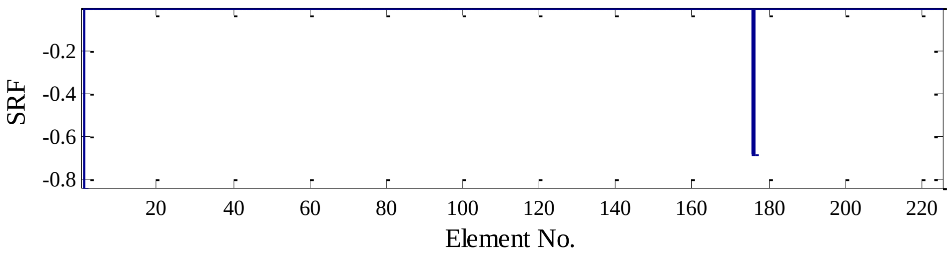
5.4. The damage identification results of three models are plotted in Figure 6.14.



(a) Laplace approximation



(b) VBI-DRAM algorithm



(c) EM algorithm

Figure 6.14 Damage identification results of the three-story frame

The identification errors of the Laplace approximation, VBI-DRAM and EM algorithm are 1.29%, 1.73% and 1.71%, respectively. The corresponding computation times are 246 mins, 73 min and 657 mins, respectively. The three techniques have good accuracy even though damage identification errors increase compared with those in the numerical example. The possible reason is that the number of unknowns in the experimental example is substantially larger than that in the numerical one. Similarly, the EM technique requires the most computational time. However, the VBI-DRAM algorithm is more efficient than the Laplace approximation for this case (n=225) because the calculation of the large-scale Hessian matrix in high-dimensional problem is extremely time-consuming.

In summary, both the numerical and experimental examples show that the damage detection results using the three techniques are accurate. The EM technique is the least efficient. Laplace approximation is markedly efficient for low-dimensional problems because no sampling is required. The VBI-DRAM algorithm is substantially efficient in dealing with high-dimensional problems. Therefore, the VBI-DRAM algorithm is recommended for large-scale structures with thousands of elements, in which case substructuring methods may be combined.

6.6 Summary

This chapter proposes an improved Bayesian method for structural damage detection based on the combination of VBI and DRAM algorithms. The posterior PDFs of the damage index and parameters are derived, and a full Bayesian treatment of all posterior uncertainties is conducted. The damage index and parameters are iteratively solved.

The proposed VBI-DRAM method was applied to a laboratory-tested frame. The results demonstrate that the damage location and severity can be accurately detected. The uncertainty of each mode is different, verifying the independence of each mode assumed in this study. The proposed method has the advantages of high accuracy and wide applicability for structural damage detection. The proposed method is applicable to standard and nonstandard probability distributions. Besides, the VBI-DRAM algorithm is more efficient for high-dimensional problems as compared with the Laplace approximation technique.

CHAPTER 7

KNOWLEDGE TRANSFER FOR STRUCTURAL DAMAGE IDENTIFICATION THROUGH RE-WEIGHTED ADVERSARIAL DOMAIN ADAPTATION

7.1 Introduction

The damage identification methods developed in Chapters 5 and 6 are based on modal parameters, which requires pre-processing the time-domain data and extracting the modal properties. This chapter explores DL-based damage identification techniques using the time series data directly. As summarized in Chapter 2, despite DL methods have been developed extensively, data-driven DL-based structural damage identification is far from mature, particularly in damage localization and quantification tasks, which belong to supervised learning. The major challenge is the lack of labels in the measurement data of practical structures, as the structural condition is unknown in advance. A few researchers used the FE model of a real structure to simulate labeled damage data, which could comprise all possible DSs for DL network training. However, the FE model may inevitably suffer from modeling errors. The discrepancy between the FE model and the real structure may lead to performance degradation when the AI model trained on the FE model is applied to the real structure. Another challenge is that the well-trained DL network only performs well on the structure where the training data arise from. When the network is applied to the other structure, the performance degrades significantly. Thus, the network should be re-trained from scratch by re-collecting labeled damage data, which is expensive and infeasible in SHM.

The reason behind the performance degradation is that the datasets from different datasets have varied distributions in feature spaces. To mitigate the dependency on labeled data, transfer learning has been developed for learning in an unlabeled or few-labeled target domain with the help of a source domain where abundant labeled data are available. DA, as a subcategory of unsupervised transfer learning, explores the knowledge transfer from labeled source domain to unlabeled target domain for label prediction. Section 2.4.2.2 summarizes the DA-based structural damage detection method that have been developed.

Despite the progress, two major challenges exist in the application of DA to structural damage identification. First, the common DA assumes that the source and target domains share an identical label space, that is, the label categories are the same. However, damage in the practical structure is a rare and irreversible event, which generally occurs at limited positions. Consequently, the DSs of a real structure are always limited and may be one or a small subset of the source label space that comprises all possible DSs. This label space inconsistency may lead to a negative transfer in DA (Cao et al., 2018). Second, the generator in the adversarial DA merely focuses on extracting features that could fool the domain discriminator while neglects the task-specific classifier boundaries between classes in the target domain (Saito et al., 2018). Hence, features generated near the classifier boundaries may be ambiguous and further deteriorating the classification accuracy.

This chapter develops an RADA method that considers the label space inconsistency and classifier boundaries. A re-weighting mechanism is introduced to down-weigh the importance of irrelevant source DSs in the DA and promote the positive transfer by matching the feature distributions in the shared label space.

7.2 Preliminary Knowledge

The definitions and basic operations about CNN, DenseNet, and DA are briefly introduced.

7.2.1 CNN Architecture

CNN, as the state-of-the-art DL technique, has the features of the receptive field, weight sharing, and spatial down-sampling (Krizhevsky et al., 2012). The basic operations in CNN consist of convolution, nonlinear activation, pooling, fully-connection, and decision-making. Besides, some techniques are particularly proposed to facilitate the network training, such as dropout (DP) (Srivastava et al., 2014) and batch normalization (BN) (Ioffe and Szegedy, 2015).

Conv-ReLU-Pooling module. The convolutional layer is the core of a CNN. The convolutional kernel slides across the height and weight of the input in the two-dimensional CNN, calculating the dot product between the kernel parameters and kernel-sized region over the feature map. In the one-dimensional CNN, the kernel slides across the width of the input whereas extends through the height and channel. The kernel size affects the width and height of the output feature map, and the number of kernels decides the channels of the output. After the convolution operation, the features are further processed by the nonlinear activation functions for nonlinear transformation. The representative nonlinear activation functions include the hyperbolic tangent function, sigmoid function, and rectified linear units (ReLU) (Nair and Hinton, 2010) activation function. Among them, the ReLU is most widely used as it mitigates the risk of gradient vanishing in the DL network training. After the nonlinear transformation, the features pass through the pooling layer, which can be regarded as a special convolutional layer. Different from the common convolution operations whose results are related to the kernel parameters, the pooling layer has no kernel parameters. Thus the pooling operations are deterministic when the size is chosen. The pooling layer serves a dual purpose of spatially down-sampling features maps and improving the robustness of the extracted features. The average and max pooling are two representative pooling techniques, which compute the average and maximum of the non-overlapping fixed-shape region, respectively. The Conv-ReLU-Pooling is a crucial module of a CNN.

Fully-connection and decision module. The fully-connected layers are generally placed after multiple Conv-ReLU-pooling modules, forming the last few layers of a CNN. The fully-connected operation is the same as that of a traditional multilayer perceptron. For the classification task, the softmax function that computes the probabilities of the output belonging to each category is usually utilized for decision-making.

DP and BN. DP (Srivastava et al., 2014) is a regularization technique that alleviates the overfitting problems in the DL network training. Neurons are randomly and temporarily dropped with the pre-designed probability to prevent the network from co-adapting too much. BN (Ioffe and Szegedy, 2015) is generally placed after the convolutional layer and before the nonlinear activation operation. A simple CNN architecture designed for the classification task is depicted in Figure 7.1.

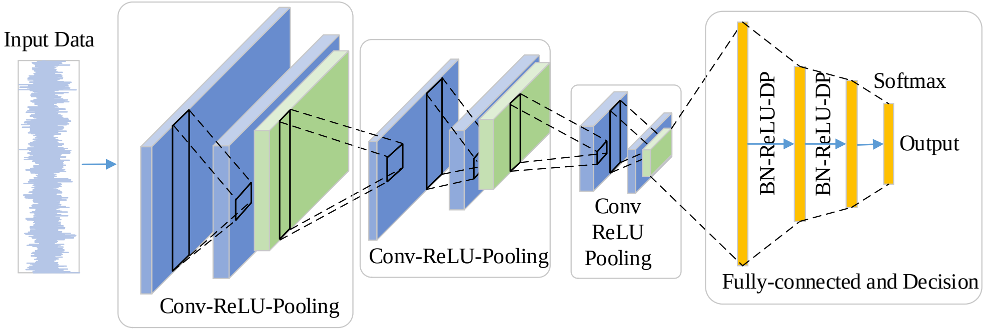


Figure 7.1 Architecture of a simple CNN for classification task

7.2.2 DenseNet Architecture

The densely connected convolutional network (DenseNet) proposed by Huang et al. (2015) is a special CNN, in which the convolutional layers are connected in a special skipping manner, enhancing the information flow between layers. The basic operations and components of a DenseNet are introduced as follows.

Dense block (DB). A DB is composed of multiple convolutional layers with the identical size of feature maps. Each convolutional layer connects to other layers in a feed-forward pattern and receives features from all preceding layers. Hence, the feature maps of all layers with one DB should maintain the same size for feature concatenation, and no pooling layer that conducts the down-sampling is involved in one DB. As compared with traditional DL networks that pass the information flow from layer to layer, the dense connection alleviates the gradient vanishing problem and strengthens the features of feed-forward and back-propagation.

Transition layer (TL). Down-sampling is crucial in reducing the size of feature maps for feature extraction. However, the convolutional layers within one DB are required to maintain the same size of feature maps. In this connection, the entire DenseNet is divided into several DBs, and the adjacent two are connected by the transition layer, which typically conducts the pooling operation for feature down-sampling.

Growth rate. Supposed that each convolutional layer produces k feature maps, then the lth layer in one DB receives $k_0 + (l - 1) \times k$ feature maps and outputs k feature maps, where k_0 denotes the channels (number of feature maps) of the input to one DB, and k is the growth rate. If a DB has n convolutional layers, then the TL behind it will receive $k_0 + n \times k$ feature maps. Figure 7.2 shows a simple DenseNet composed of two DBs and one TL.

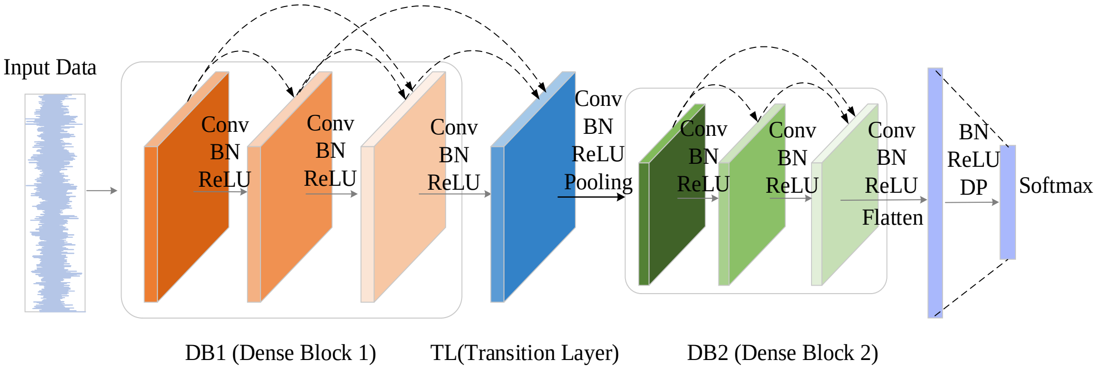


Figure 7.2 Architecture of a simple DenseNet

7.2.3 Adversarial DA

Some symbols and definite about the DA are first introduced. Denote the data and labels in the source domain as $D_s = \{X_i^s, Y_i^s\}(i = 1, \ldots, n_s)$, which consists of n_s labeled samples with N_s categories. The target domain is denoted as $D_t = \{X_i^t\}$ $(i = 1, \ldots, n_t)$ that consists of n_t unlabeled samples with N_t classes. The data distribution of the source and target domains $P_s(X) \neq P_t(X)$ because of the domain dataset bias (Pan and Yang, 2009). The target of the DA is to design a feature extractor *G* and classifier *C* to minimize the classification error $\epsilon = \mathbb{E}_{(X_i^t, Y_i^t) \sim P_t(X)}[C(G(X^t)) \neq Y^t]$ in the target domain, where Y^t is unknown in the model training process.

Ganin et al. (2015) first proposed the adversarial DA network, which consists of a feature generator *G*, a label predictor *C*, and a domain discriminator *D*, as depicted in Figure 7.3. All three components are composed of neural networks. The entire network is trained in a minimax manner. Specifically, *D* is trained to discriminate whether the features extracted by *G* come from the source or target domain, whereas *G* is trained to extract features that couldn't be discriminated by *D*. Meanwhile, *G* is also trained to extracted source class-sensitive features that could be classified by *C*. Such an adversarial training methodology attempts to match feature distributions of both domains, such that the model trained on source domain generalizes well to the target domain. Denote the parameters in *G*, *C* and *D* as θ_c, θ_d, and θ_f, respectively. The adversarial DA network is trained to minimize the objective function (Ganin et al., 2015):

$$J(\theta_f, \theta_c, \theta_d) = \frac{1}{n_s} \sum_{x_i \epsilon X_s} L_y(G_y(G_f(x_i)), y_i) - \frac{\lambda}{n_s + n_t} \sum_{x_i \epsilon X_s \cup X_t} L_d(G_d(G_f(x_\mathrm{i})), d_i)$$

(7.1)

where d_i denotes the domain labels of x_i, that is, $d_i = 0$ when x_i comes from the source domain and $d_i = 1$ otherwise; L_y is the loss of label prediction of the source domain; L_d

is the loss of domain discrimination; and λ is a trade-off parameter between two losses that changes from 0 to 1 incrementally as the training process proceeds (Ganin et al., 2015). The parameters θ_c, θ_d, and θ_f are trained in a two-player minimax manner. Specifically, G is trained to maximize L_d, and D is trained to minimize L_d. Mathematically, the parameters are optimized as:

$$(\hat{\theta}_f, \hat{\theta}_c) = \underset{\theta_f, \theta_c}{\operatorname{argmin}} J(\theta_f, \theta_c, \theta_d) \tag{7.2}$$

$$(\hat{\theta}_d) = \underset{\theta_d}{\operatorname{argmax}} J(\theta_f, \theta_c, \theta_d) \tag{7.3}$$

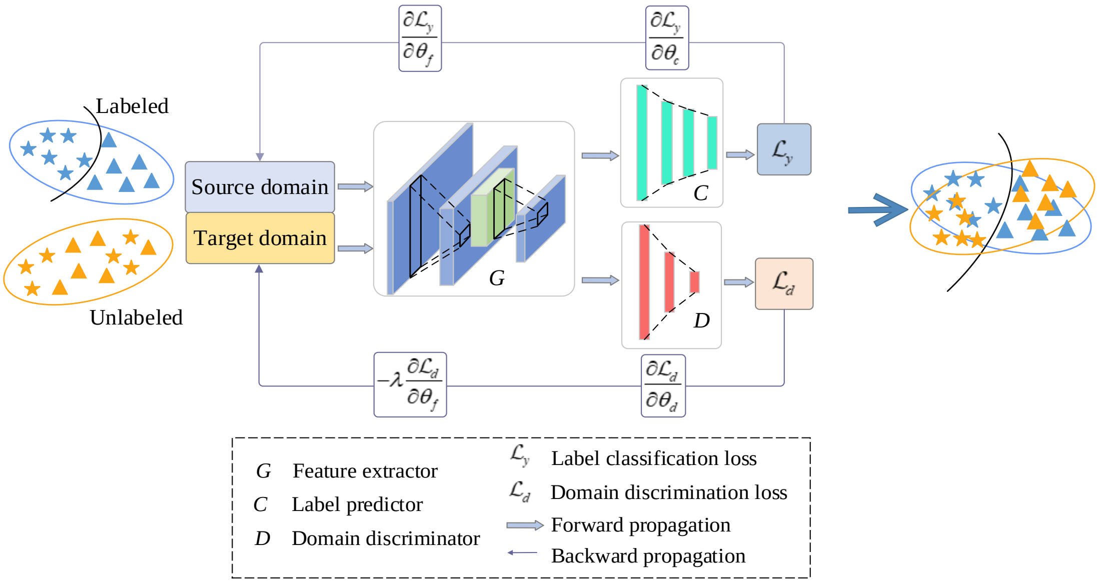


Figure 7.3 Domain adversarial training of neural network

However, the basic domain-adversarial DA network suffers from limitations. First, the above training methodology merely aims to extract domain-invariant and class-discriminative features while neglects the classifier boundaries (Saito et al., 2018). As a result, the extracted features may be ambiguous near the classifier boundary, as shown in Figure 7.4, deteriorating the classification accuracy on the target domain. Second, the source and target domains are assumed to have an identical label space. This assumption does not hold in structural damage identification because the DSs (labels) in the target domain are always less than those in the source domain that comprise all possible DSs. This label space inconsistency may cause negative transfer due to the significant mismatch between the label sets (Cao et al., 2018), as depicted in Figure 7.5.

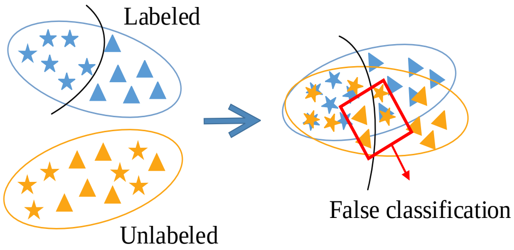


Figure 7.4 Ambiguous features near classifier boundary

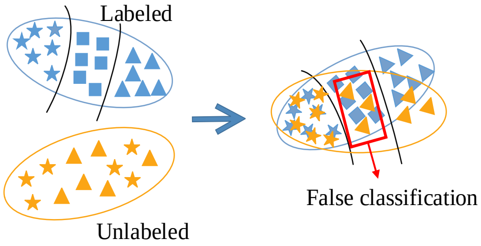


Figure 7.5 Classifier inconsistency

7.3 RADA Network for Damage Detection

This study develops an RADA network that maximizes the classifier discrepancy and considers the label space inconsistency. The method is inspired by the study of Saito et al. (2018), where a network with two task-specific classifiers was designed to maximize the classifier discrepancy. The training process of their study is illustrated in Figure 7.6, where red lines represent classifier boundaries. The discrepancy between the two classifiers was measured. The generator and two classifiers were trained in an adversarial manner. Specifically, the two classifiers attempted to maximize the discrepancy based on the extracted features, while the generator was trained to extract features from the target domain that minimize the discrepancy. Such an adversarial training methodology avoids generating ambiguous features outside the support of the source classifier boundary (Saito et al., 2018), thus instructs the network to learn class-separate and domain-invariant features. However, the source and target domains are assumed to have identical label spaces (Saito et al., 2018), which is inconsistent with the present damage identification problem, as explained previously. This study will improve the network for the problem with inconsistent label spaces. The main idea is to identify the label space shared by both domains, and then adapt the features from both domains in the consistent label spaces for knowledge transfer, as depicted in Figure 7.7.

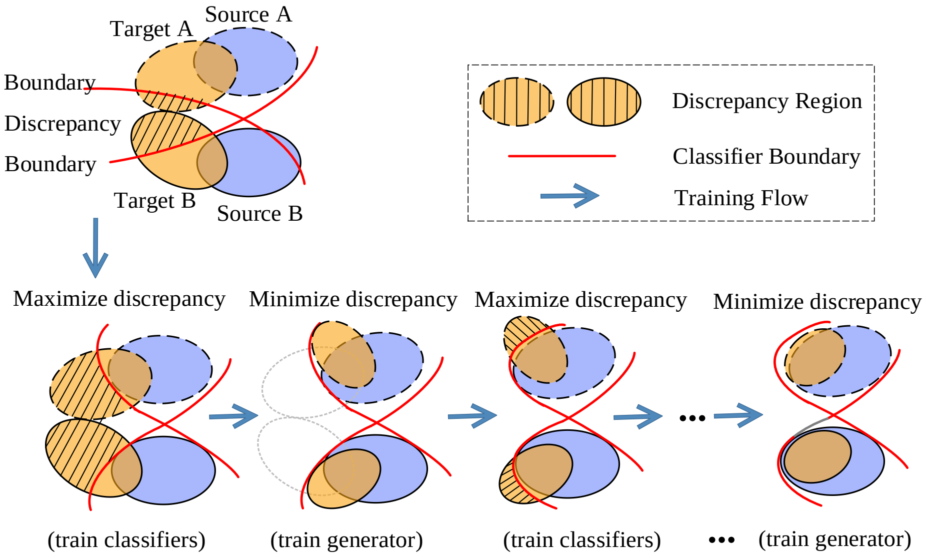


Figure 7.6 DA method proposed by Saito et al. (2018)

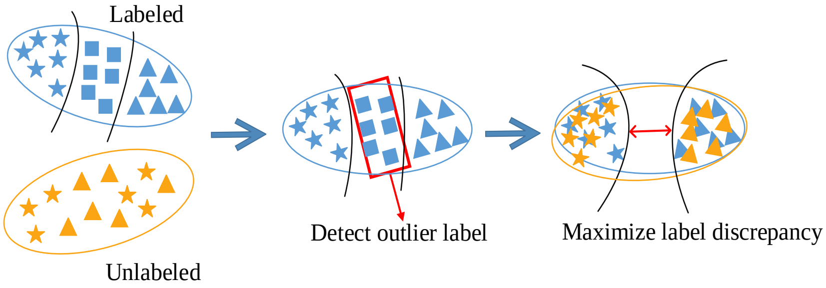


Figure 7.7 Proposed RADA method

7.3.1 RADA Network Architecture

The architecture of the proposed RADA network (Figure 7.8) consists of a generator G and two classifiers C_1 and C_2. The specific connection patterns between layers are introduced as follows.

Generator G adopts the architecture of DenseNet with three DBs and two TLs. Each DB has three convolutional layers. The growth rate of the three DBs is 32, 64 and 126. Three consecutive operations are carried out in each convolutional layer of one DB, that is, BN, ReLU, and conv operation. The size of the convolutional kernel is set to 7×1 with a stride of 1. The zero-padding strategy is adopted to ensure the unchanged size of feature maps after the conv operation. Two TLs are designed with a down-sampling rate of 0.5. Similar to Huang et al. (2017), a convolutional layer is designed before the first DB for data pre-processing.

C_1 and C_2 consist of two fully-connected layers. The DP with a rate of 0.5 is introduced to alleviate overfitting. Notably, the global average pooling (GAP) (Lin et al., 2013) is applied before the features extracted by G enters the classifier. The advantage of GAP lies in that the average of each feature map is directly calculated without defining any parameters. In addition, the size of the feature map after the GAP operation changes to 1. Thus the input size to the classifier is related to the channels only, and not to the input size. The network gets rid of re-designing parameters when the input size changes. We suppose that the size of data samples is 1000×1 and the output size is $m \times 1$. Table 1 lists the detailed configuration of each layer. The input shape to C_1 and C_2 depends on the channels only and is not related to the size of data samples.

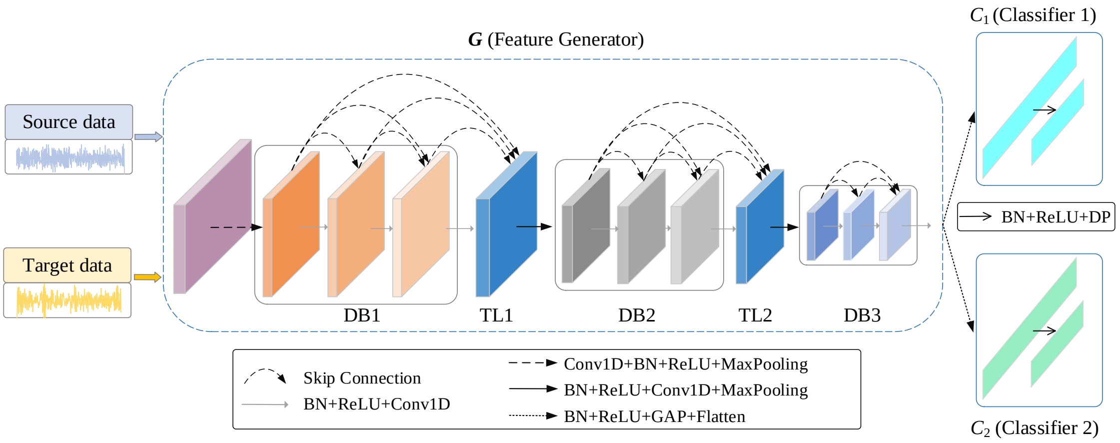


Figure 7.8 Architecture of the proposed RADA network

Table 7.1 Configuration of the network (input size 1000×1 and output size m×1)

Type	Layer	Kernel size	Kernel number	Stride	Pad-ding	Input size	Output size	BN	DP
Conv	Conv1D	5	64	2	2	1000×1	64-[500×1]	YES	NO
	Max-Pooling	3	None	2	1	64-[500×1]	64-[250×1]	NO	NO
DB1	Conv1D	7	32	1	3	64-[250×1]	32-[250×1]	YES	NO
	Conv1D	7	32	1	3	96-[250×1]	32-[250×1]	YES	NO
	Conv1D	7	32	1	3	128-[250×1]	32-[250×1]	YES	NO
TL1	Conv1D	1	80	1	0	160-[250×1]	80-[250×1]	YES	NO
	Max-Pooling	2	None	2	0	80-[250×1]	80-[125×1]	NO	NO
DB2	Conv1D	7	64	1	3	80-[125×1]	64-[125×1]	YES	NO
	Conv1D	7	64	1	3	144-[125×1]	64-[125×1]	YES	NO
	Conv1D	7	64	1	3	208-[125×1]	64-[125×1]	YES	NO
TL2	Conv1D	1	136	1	0	272-[125×1]	136-[125×1]	YES	NO
	Max-Pooling	2	None	2	0	136-[125×1]	136-[62×1]	NO	NO
DB3	Conv1D	7	126	1	3	136-[62×1]	126-[62×1]	YES	NO
	Conv1D	7	126	1	3	262-[62×1]	126-[62×1]	YES	NO
	Conv1D	7	126	1	3	388-[62×1]	126-[62×1]	YES	NO
	GAP	None	None	None	None	514-[62×1]	514×1	NO	NO
C_1/ C_2	Fully-connected	None	None	None	None	514×1	256×1	YES	0.5
	Fully-connected	None	None	None	None	256×1	m×1	NO	NO
	Softmax	None	None	None	None	m×1	m×1	NO	NO

7.3.2 Loss Function

Two kinds of loss functions are used. The first one is the categorical cross-entropy function, which calculates the classification error of the source data:

$$L(X_S,Y_S) = -\mathbb{E}_{(x_s,y_s)\sim(X_S,Y_S)} \sum_{k=1}^{K} 1_{[k=y_s]} \log p(y|x_s) \tag{7.4}$$

where $1_{[k=y_s]}$ equals 1 if $k = y_s$ and 0 otherwise; K denotes the number of categories

($K = N_s$ in the source domain); and $p(y|x_s)$ is the probability of the source data x_s belonging to each category, which is estimated by the softmax function.

The second loss function used is the discrepancy loss between the two classifiers. With the *i*th input, the discrepancy between the probabilistic outputs of the two classifiers is defined as (Saito et al., 2018):

$$d\big(p_1(y|x_i), p_2(y|x_i)\big) = \|p_1(y|x_i) - p_2(y|x_i)\|_1 = \frac{1}{K}\sum_{k=1}^{K}\|p_{1k} - p_{2k}\|_1 \tag{7.5}$$

where $p_1(y|x_i)$ and $p_2(y|x_i)$ denote the softmax output of C_1 and C_2, respectively; and p_{1k} and p_{2k} represent the probabilistic outputs for class k of the two classifiers. The discrepancy loss of the entire target dataset is defined as the average of all samples' discrepancies, that is:

$$L_{adv} = \mathbb{E}_{(x_t)\sim(\mathrm{X}_t)} d\big(p_1(y|x_t), p_2(y|x_t)\big) = \frac{1}{n_t}\sum_{i=1}^{n_t} d\big(p_1(y|x_i), p_2(y|x_i)\big) \tag{7.6}$$

7.3.3 Network Training

Denote the parameters in networks G, C_1, and C_2 as θ_f, θ_{c_1}, and θ_{c_2}, respectively. Although two classifiers have identical architecture, their parameters are initialized differently to result in different classifier boundaries and generate a classification discrepancy from the beginning of the training. The entire training process of the RADA network consists of four steps.

Step A: Categorize the source data. In the first step, the entire network, including G, C_1, and C_2, is trained using the labeled source data to classify the categories. As shown in Figure 7.9, all parameters θ_f, θ_{c_1}, and θ_{c_2} are updated to minimize the loss function:

$$\min_{\theta_f, \theta_{c_1}, \theta_{c_2}} L_y = L_{src} = L_1(X_S, Y_S) + L_2(X_S, Y_S) \tag{7.7}$$

where L_{src} denotes the classification error of the source data, which is computed by the cross-entropy function in Eq. (7.4) on the basis of the outputs of C_1 and C_2. The Adam optimization algorithm (Kingma and Ba, 2014) is adopted throughout this chapter, with the initial learning rate of 1×10^{-4} and the weight decay of 1×10^{-6}. The first step is crucial to evaluate the weight vector in the next step.

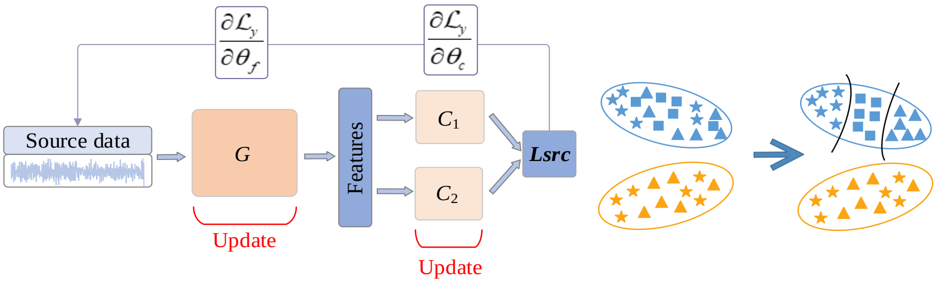


Figure 7.9 First step in the training process

Step B: Calculate the weight vector. When the source and target domains have inconsistent label spaces, the core is to identify the label space shared by both domains and lower the importance of the outlier source classes in the DA. To attain this, we attempt to compute a weight vector $\mathbf{W}$, whose elements reveal the probability that each category in the source domain belongs to the label space shared by both domains. We take $\mathbf{W} = [1, 0.1, 1]$ as an example with three categories. This vector indicates that the first and third categories have a high probability of belonging to the share label space, and the second category probably belongs to the outlier label space. Then the weight vector can be introduced to the cross-entropy function to decrease the importance of the outlier source classes in the DA.

However, the challenge in estimating $\mathbf{W}$ is that the labels of the target data are unknown. After Step A, the network can evaluate the probability of source data samples belonging to each category. In the process of DA, features extracted from the source and target domains are adapted to follow the identical distribution. Thus, when the target data are used as the input, the network can also predict the probability of each data sample in the target domain belonging to each category, although the result may be inaccurate in the preliminary stage. By summarizing all target data samples and comparing the predicted classes with the source classes, the probability of each category belonging to the shared label space or outlier space. Therefore, the average of the outputs of all data samples in the target domain can be calculated as the weight vector:

$$\mathbf{W} = \frac{1}{2n_t}\sum_{i=1}^{n_t}[p_1(y|x_i^t) + p_2(y|x_i^t)] \tag{7.8}$$

where $\mathbf{W}$ is an N_s-dimensional vector, and $p_1(y|x_i^t)$ and $p_2(y|x_i^t)$ denote the softmax output of C_1 and C_2 corresponding to the ith target samples, respectively. Notably, $\sum_j^{N_s} \mathrm{w}_j = 1$ according to the properties of the softmax function. In the preliminary stage, $\mathbf{W}$ may not be accurate. To alleviate the problem that some underestimated elements with trivial values may have low participation in the early stage of network training, $\mathbf{W}$ is normalized as (Cao et al., 2018):

$$\mathbf{W} = \mathbf{W}/\max\left(\mathrm{w}_1, w_2, \ldots, w_{N_s}\right) \tag{7.9}$$

All parameters of the generator and two classifiers are fixed in this step, and there is no model training, as presented in Fig. 10.

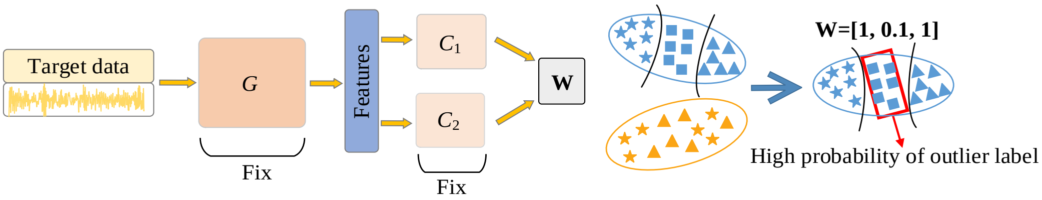


Figure 7.10 Second step in the training process

<u>Step C: Train classifiers C_1 and C_2.</u> In this step, two classifiers C_1, and C_2 are trained to maximize the discrepancy between their outputs, and thus, the ambiguous features out of source classifier boundaries (the discrepancy region in Figure 7.6) are detected. Saito et al. (2018) considered the classification error of the source domain in this step to prevent significant performance dropping. The weight vector $\mathbf{W}$ calculated in Step B is able to evaluate the importance of each source category in DA. Thus, $\mathbf{W}$ is introduced to instruct the DA to focus on the shared label space. The loss function used in this step is defined as:

$$\min_{\theta_{c_1},\,\theta_{c_2}} L_y = L_{wsrc} - L_{adv}$$

$$= -\mathbb{E}_{(x_s,y_s)\sim(X_s,Y_s)} \sum_{k=1}^{K} \left[w_k\, 1_{[k=y_s]} \log p_1(y|x_s) + w_k\, 1_{[k=y_s]} \log p_2(y|x_s) \right] - L_{adv} \tag{7.10}$$

where L_{wsrc} denotes the re-weighted classification loss of the source data. Take $\mathbf{W} = [1, 0, 1]$ as an example, only the classification errors corresponding to the first and third categories participate in L_{wsrc}. The parameters in G keep fixed, thereby the features extracted by G remain unchanged. Only the parameters in C_1, and C_2 are updated, as shown in Figure 7.11.

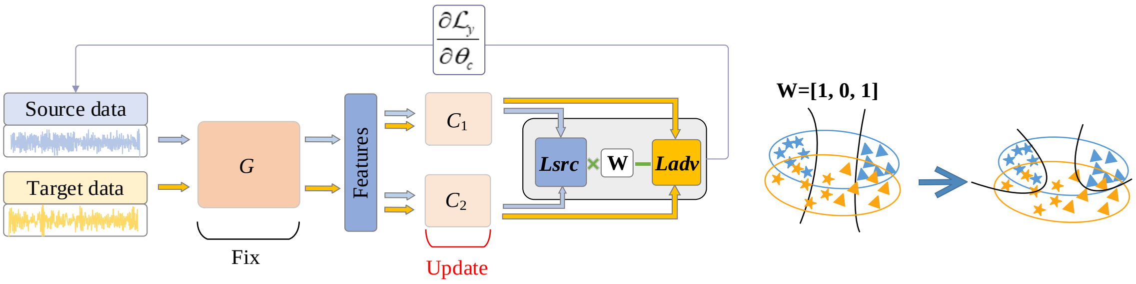


Figure 7.11 Third step in the training process

<u>Step D: Train generator *G*.</u> In the last step, G is trained to minimize the classifier discrepancy, that is, to extract features within the support of the source classifier boundary. The parameters in G are updated to minimize the loss function as:

$$\min_{\theta_f} L_y = L_{adv} \tag{7.11}$$

Notably, the parameters in C_1, and C_2 are fixed, meaning that the classifier boundaries remain unchanged, and only the parameters in G are updated, as depicted in Figure 7.12.

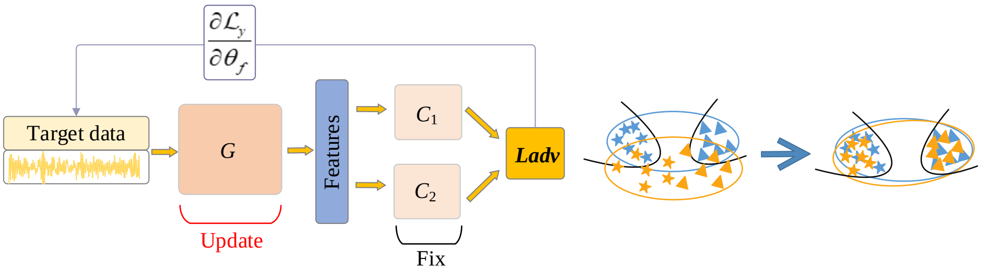


Figure 7.12 Fourth step in the training step

These four steps are iterated to extract the classifier-discriminate and domain-invariant features for the classification task of the unlabeled target data. In this study, *G* is deep with many layers as listed in Section 7.3, so *G* may be more difficult to be trained than C_1 and C_2. In this regard, Step D is designed to repeat four times for each mini-batch before proceeding to the next step.

7.3.4 Summary of the proposed RADA Method

The training process is summarized below.

Algorithm: Training of the RADA network

```
Set the mini-batch size and epochs. The Adam optimization algorithm is utilized with
the initial learning rate of $1 \times 10^{-4}$ and the weight decay of $1 \times 10^{-6}$.
For $i = 1$ to No. of epochs (i.e., 200):
  For $j = 1$ to No. of batches ( = No. of data samples /batch size)
    Train $G$, $C_1$, and $C_2$ to classify the source data by minimizing Eq. (7.7);
    Estimate the weight vector $\mathbf{W}$ according to Eqs. (7.8) and (7.9);
    Train $C_1$ and $C_2$ to maximize the discrepancy by minimizing Eq. (7.10);
    For $k = 1$ to $n$ (i.e., $n = 4$)
      Train $G$ to minimize the discrepancy in Eq. (7.11);
    End
  End
End
Use the well-trained $G$, $C_1$, and $C_2$ for the classification of unlabeled target data.
```

7.4 Knowledge Transfer between Structures

7.4.1 Knowledge Transfer between Structures with Different Sizes

This subsection studies the knowledge transfer between structures of different sizes. A simply-supported steel beam shown in Figure 7.13 is taken as an example,. Beam A is assumed to have labeled damage data under all possible DSs, whereas Beams B and C have damage data only without any labels. The target is to transfer the DL model trained on Beam A to B or C for damage detection. The length of Beam A is 1 m and the cross-section is 50 mm × 5 mm. The Young's modulus and mass density of the steel material are 2.0×10^{11} N/m^2 and 7.67×10^3 kg/m^3, respectively. Beams B and C have the same material properties and cross-sections as Beam A but different lengths, i.e., 0.7 m and 1.3 m long, respectively, as shown in Figure 7.13. The Rayleigh damping is adopted:

$$C = \alpha\boldsymbol{K} + \beta\boldsymbol{M} \tag{7.12}$$

where $\alpha = 2\xi\omega_i\omega_j/(\omega_i + \omega_j)$ and $\beta = 2\xi/(\omega_i + \omega_j)$. The first and second frequencies are used for damping estimation, that is, $\omega_i = \omega_1$, $\omega_j = \omega_2$ and ξ is set to 0.05.

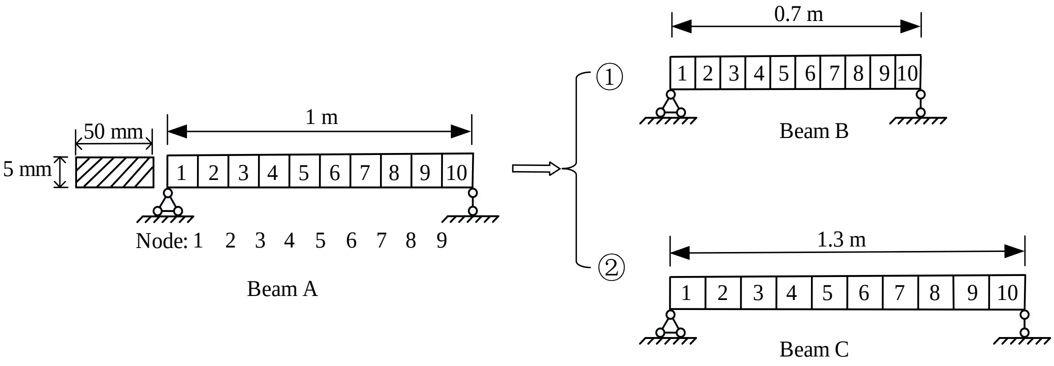


Figure 7.13 Configuration of beams with different lengths

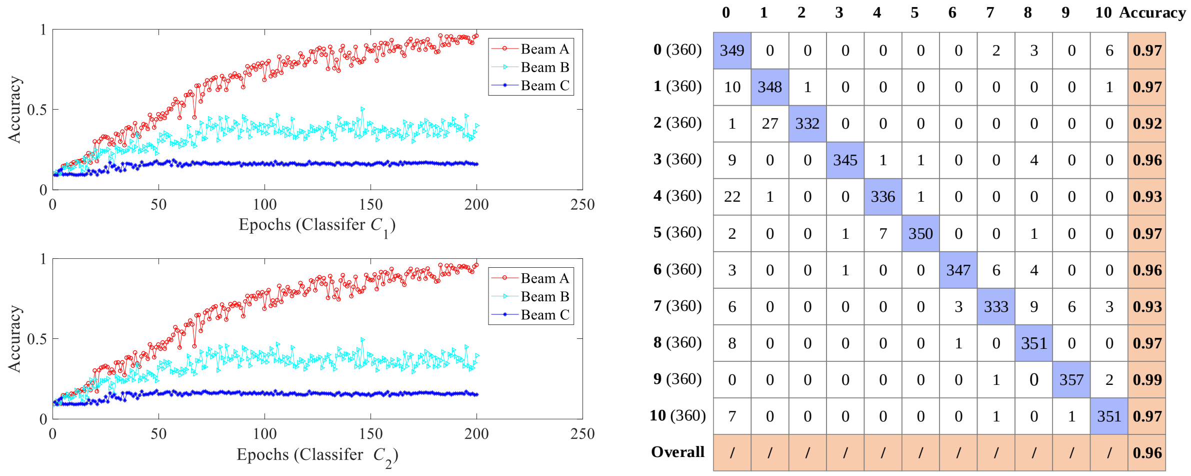
The source data are collected from Beam A. The beam is divided into 10 equal Euler-Bernoulli beam elements. A random white Gaussian noise is applied to beam supports in the vertical direction. Each excitation sample has 1,000 data points at the time step of 5×10^{-4} s, accounting for a duration of 0.5 s. The Newmark's method is employed to calculate the structural dynamic responses. The acceleration responses of Nodes 1 to 9 in the vertical direction are collected as the input to the RADA network. The responses of the nine nodes with a duration of 0.5 seconds are concatenated into one-dimensional data. Consequently, each data sample has a size of $9{,}000 \times 1$. For comparison, the data samples are directly used in the two-dimensional form in Section 7.4, showing that the proposed RADA network can avoid re-designing parameters when the input size changes. In practice, the measurement noise may exist. The accelerations are additionally contaminated with 2% white Gaussian noise. The target data are similarly collected from Beam B or C. The labels of the target data are treated as unknown in the network training stage and are only used for estimating the accuracy of the damage identification results.

7.4.1.1 Damage localization

Damage is introduced by decreasing the elemental stiffness. The source data come from Beam A. Nine levels of damage are introduced to each element of Beam A in sequence by reducing the elemental stiffness 10%, 20%, and up to 90%. At each damage level, 40 excitations are applied repeatedly. Hence, 360 data samples are generated for each damaged element and a total of 3,600 samples for all damaged scenarios. Besides, 360 more samples are also generated in the intact state of the structure so that the intact state has the identical number of samples as one DS. For the damage localization task, all 3960 samples are classified into 11 categories (10 damage locations + 1 intact condition) labeled by the one-hot encoding method, where only one entry is one whereas the rest are all zero. Thereby, the size of the RADA network output is 11×1, where the first entry corresponds the probability of the undamaged condition, and the rest ten correspond to the damage locations from the first element to the tenth. The target data are similarly collected from Beam B or C. Consistent and inconsistent label spaces between the source and target domains are designed to investigate the performance of the proposed network. All calculations are conducted on a desktop with a CPU of Intel Core i7-8700, 48 GB RAM, and GPU acceleration of NVIDE GeForce RTX 2080Ti.

(1) Consistent label spaces

First, consistent label spaces between the source and target domains are investigated. 3960 target samples with 11 label categories are collected from Beam B or C. The network is first trained to classify the labeled source data from Beam A without DA, that is, only Step A in Section 3.3 is conducted whereas the rest three steps are neglected. The batch size and number of epochs are set to 60 and 200, respectively. The two classifiers C_1 and C_2 are trained in parallel. The training accuracies over epochs and final classification results are plotted in Figure 7.14.

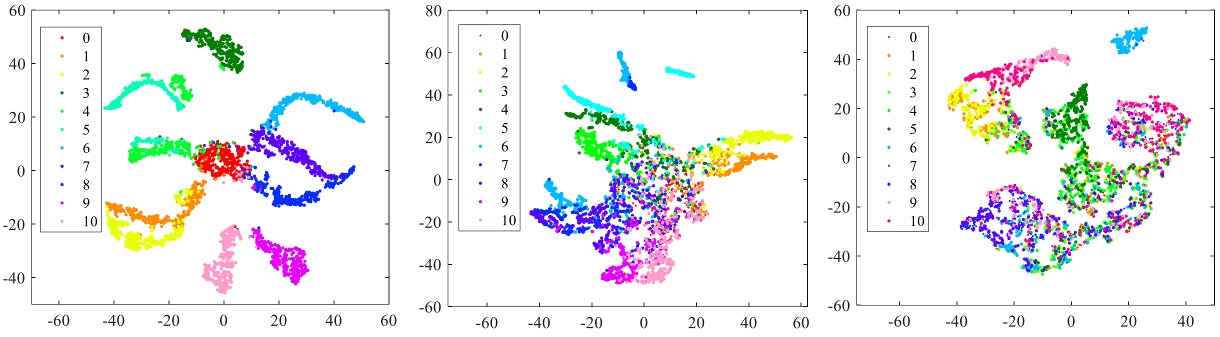


	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Accuracy**
0 (360)	349	0	0	0	0	0	0	2	3	0	6	**0.97**
1 (360)	10	348	1	0	0	0	0	0	0	0	1	**0.97**
2 (360)	1	27	332	0	0	0	0	0	0	0	0	**0.92**
3 (360)	9	0	0	345	1	1	0	0	4	0	0	**0.96**
4 (360)	22	1	0	0	336	1	0	0	0	0	0	**0.93**
5 (360)	2	0	0	1	7	350	0	0	1	0	0	**0.97**
6 (360)	3	0	0	1	0	0	347	6	4	0	0	**0.96**
7 (360)	6	0	0	0	0	0	3	333	9	6	3	**0.93**
8 (360)	8	0	0	0	0	0	1	0	351	0	0	**0.97**
9 (360)	0	0	0	0	0	0	0	1	0	357	2	**0.99**
10 (360)	7	0	0	0	0	0	0	1	0	1	351	**0.97**
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.96**

(a) Training accuracies over epochs (b) Classification results of Beam A

Figure 7.14 Accuracies trained on source data using CNN without DA

Two classifiers achieve an accuracy of 95.8% and 95.9% on the source data. However, when the model trained on the source data are applied to the unlabeled target data, the classification accuracies are 37.2% and 37.2% only for Beam B, and 15.9% and 16.1% for Beam C, respectively. The t-distributed stochastic neighbor embedding (t-SNE) technology (Van der Maaten and Hinton, 2008) is employed to visualize the features extracted from the last layer of the network. Different categories are clearly separated in the source data but mixed in the target data, as shown in Figure 7.15.



(a) Source data from Beam A (b) Target data from Beam B (c) Target data from Beam C

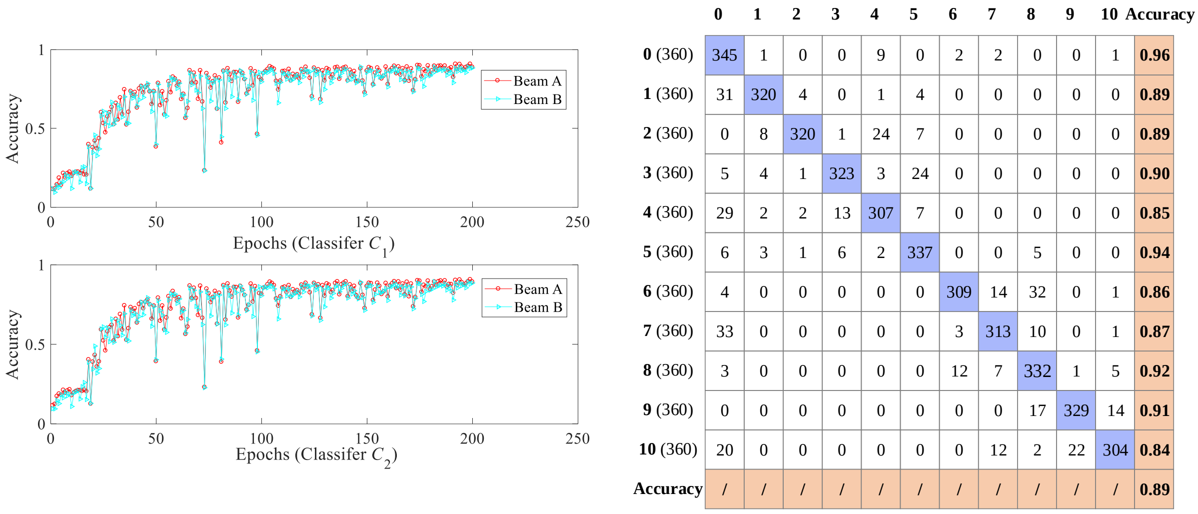
Figure 7.15 Feature visualization through t-SNE using CNN without DA (0 refers to the undamaged state and 1–10 denote damage at elements 1–10)

This performance degradation indicates that the network only performs well on the structure where the training data arise from. The proposed RADA is then carried out. Four steps in Section 7.3.3 are iterated for 200 epochs. Within one epoch, Step D is repeated four times. The weight vector **W** is initialized to an 11-dimensional vector with all entries being one. The variations of **W** in the DA process between Beams A and C are listed in Table 2. Upon convergence, all entries in **W** are larger than zero, indicating that the label space of the target domain is consistent with that of the source domain.

Table 7.2 Variations of **W** in RADA with inconsistent label spaces (from Beam A to C)

Epochs	**W**	Accuracies of C_1 /C_2
0	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]	15.8% / 15.7%
1	[0.99, 1.00, 1.00, 0.97, 0.93, 0.99, 0.93, 0.98, 0.92, 0.96, 0.97]	16.1% / 15.1%
50	[0.33, 0.47, 0.47, 1.00, 0.49, 0.31, 0.66, 0.56, 0.70, 0.49, 0.19]	34.8% / 33.4%
100	[1.00, 0.65, 0.75, 0.13, 0.84, 0.23, 0.89, 0.87, 0.73, 0.97, 0.68]	77.3% / 76.8%
150	[1.00, 0.71, 0.85, 0.90, 0.46, 0.81, 0.92, 0.82, 0.55, 0.77, 0.80]	80.1% / 80.2%
200	[1.00, 0.86, 0.79, 0.96, 0.76, 0.89, 0.94, 0.81, 0.79, 0.91, 0.89]	84.9% / 85.3%

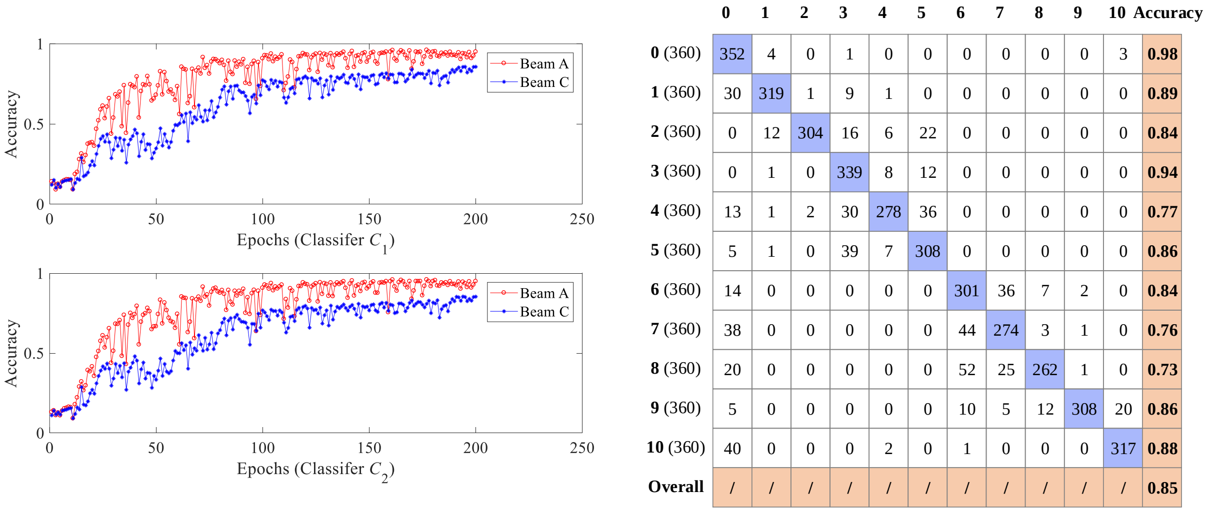
With the proposed RADA technique, the classification results of the unlabeled target data of Beam B and C are shown in Figure 7.16 and 7.17, respectively. Upon convergence, Beam B and C achieves approximately 89% and 85% accuracy, respectively, which are much more accurate than those directly using the model trained on the source data.



	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (360)	345	1	0	0	9	0	2	2	0	0	1	**0.96**
1 (360)	31	320	4	0	1	4	0	0	0	0	0	**0.89**
2 (360)	0	8	320	1	24	7	0	0	0	0	0	**0.89**
3 (360)	5	4	1	323	3	24	0	0	0	0	0	**0.90**
4 (360)	29	2	2	13	307	7	0	0	0	0	0	**0.85**
5 (360)	6	3	1	6	2	337	0	0	5	0	0	**0.94**
6 (360)	4	0	0	0	0	0	309	14	32	0	1	**0.86**
7 (360)	33	0	0	0	0	0	3	313	10	0	1	**0.87**
8 (360)	3	0	0	0	0	0	12	7	332	1	5	**0.92**
9 (360)	0	0	0	0	0	0	0	0	17	329	14	**0.91**
10 (360)	20	0	0	0	0	0	0	12	2	22	304	**0.84**
Accuracy	/	/	/	/	/	/	/	/	/	/	/	**0.89**

(a) Training accuracies over epochs (b) Classification results of Beam B

Figure 7.16 Accuracy and classification results of Beam B using RADA

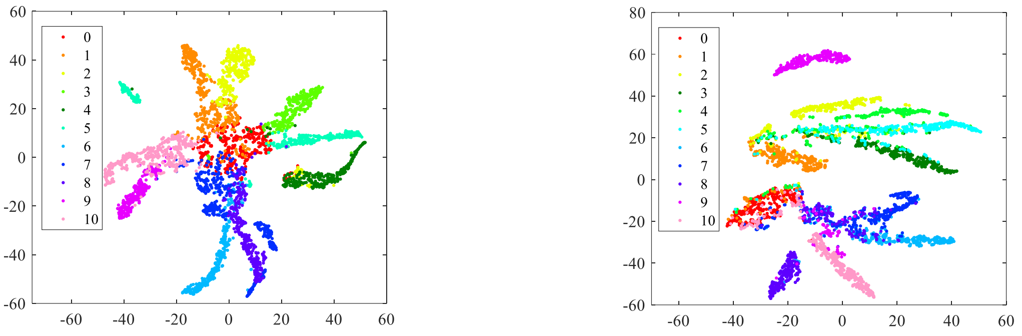


	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (360)	352	4	0	1	0	0	0	0	0	0	3	**0.98**
1 (360)	30	319	1	9	1	0	0	0	0	0	0	**0.89**
2 (360)	0	12	304	16	6	22	0	0	0	0	0	**0.84**
3 (360)	0	1	0	339	8	12	0	0	0	0	0	**0.94**
4 (360)	13	1	2	30	278	36	0	0	0	0	0	**0.77**
5 (360)	5	1	0	39	7	308	0	0	0	0	0	**0.86**
6 (360)	14	0	0	0	0	0	301	36	7	2	0	**0.84**
7 (360)	38	0	0	0	0	0	44	274	3	1	0	**0.76**
8 (360)	20	0	0	0	0	0	52	25	262	1	0	**0.73**
9 (360)	5	0	0	0	0	0	10	5	12	308	20	**0.86**
10 (360)	40	0	0	0	2	0	1	0	0	0	317	**0.88**
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.85**

(c) Training accuracies over epochs (d) Classification results of Beam C

Figure 7.17 Accuracy and classification results of Beam C using RADA

Fig. 7.18 shows the t-SNE visualization (Van der Maaten and Hinton, 2008) results of Beam B and Bean C. Compared with Figure 7.15, different categories are well separated when the RADA technique is adopted, indicating that the proposed method could attain accurate classification of unlabeled target data.



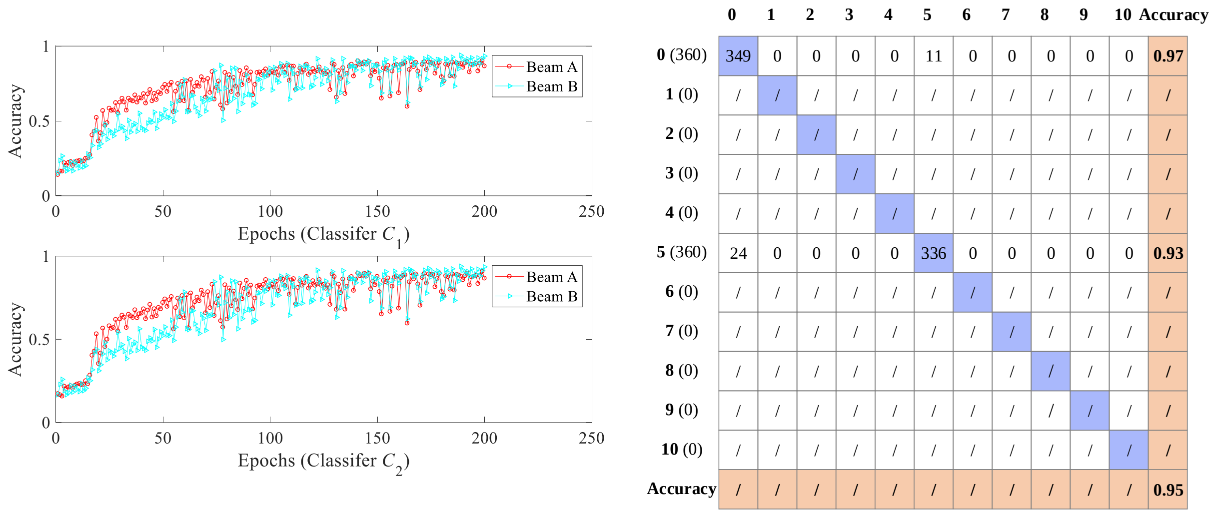
(a) Target data from Beam B (b) Target data from Beam C

Figure 7.18 Feature visualization through t-SNE using RADA

(2) Inconsistent label spaces

In practice, structural damage is a rare and irreversible process. Therefore, the target data do not comprise all DSs. For example, Beam B has a damage in No. 5 element. Then, the target data only contain the intact and damaged states with different damage levels from 10% to 90% at element 5. They do not contain damage at other elements. In this case, the target data contain two categories with 720 data samples only (40 excitations $\times$ 9 damage levels $\times$ 1 damage locations + 360 excitations $\times$ 1 intact condition). By contrast, the source data comprise 11 categories and cover the categories in the target data.

If the model trained on the source data from Beam A is directly applied to the target data from Beam B or C, the classification accuracies of the two classifiers C_1 and C_2 are 17.1% and 16.7% for Beam B, and 1.94% and 1.39% for Beam C, respectively. The RADA is then carried out between the source and target data. **W** is similarly initialized to an 11-dimensional vector with all entries being one. Four steps in Section 7.3.3 are iterated for 200 epochs, and step D repeats four times within one epoch. Figures 7.19 and 7.20 plot the training accuracy and final classification results of Beam B and Beam C, respectively. Upon convergence, the classification accuracies of C_1 and C_2 are 97.1% and 96.9% for Beam C, and 94.6% and 95.3% for Beam B, respectively, which are significantly improved compared with those of using the network trained on the source data directly.

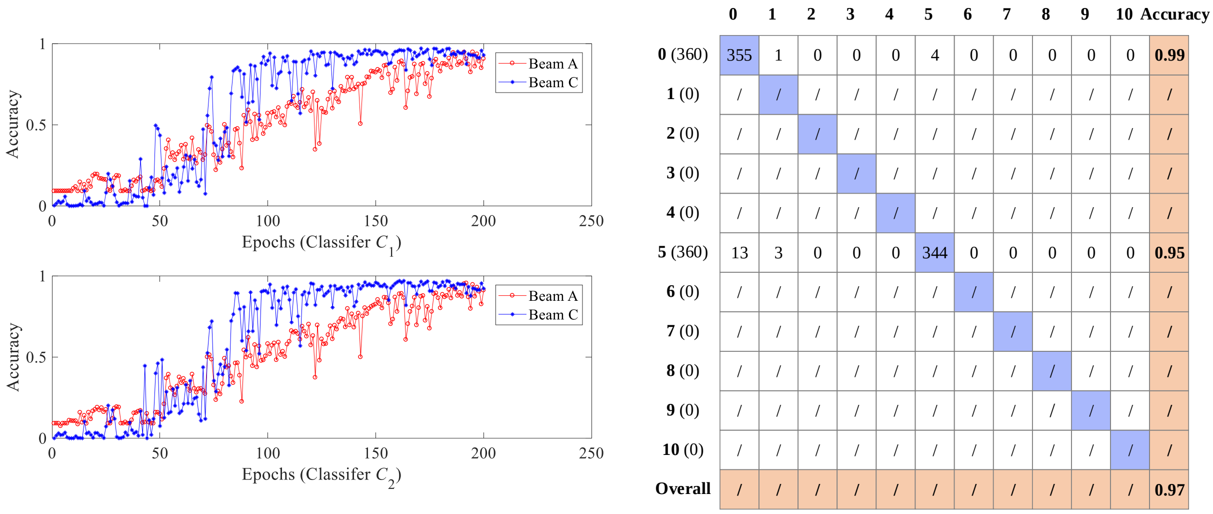


	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (360)	349	0	0	0	0	11	0	0	0	0	0	**0.97**
1 (0)	/	/	/	/	/	/	/	/	/	/	/	/
2 (0)	/	/	/	/	/	/	/	/	/	/	/	/
3 (0)	/	/	/	/	/	/	/	/	/	/	/	/
4 (0)	/	/	/	/	/	/	/	/	/	/	/	/
5 (360)	24	0	0	0	0	336	0	0	0	0	0	**0.93**
6 (0)	/	/	/	/	/	/	/	/	/	/	/	/
7 (0)	/	/	/	/	/	/	/	/	/	/	/	/
8 (0)	/	/	/	/	/	/	/	/	/	/	/	/
9 (0)	/	/	/	/	/	/	/	/	/	/	/	/
10 (0)	/	/	/	/	/	/	/	/	/	/	/	/
Accuracy	/	/	/	/	/	/	/	/	/	/	/	**0.95**

(a) Training accuracies over epochs

(b) Classification results of Beam B

Figure 7.19 Accuracy and classification results of Beam B using the RADA



	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (360)	355	1	0	0	0	4	0	0	0	0	0	**0.99**
1 (0)	/	/	/	/	/	/	/	/	/	/	/	/
2 (0)	/	/	/	/	/	/	/	/	/	/	/	/
3 (0)	/	/	/	/	/	/	/	/	/	/	/	/
4 (0)	/	/	/	/	/	/	/	/	/	/	/	/
5 (360)	13	3	0	0	0	344	0	0	0	0	0	**0.95**
6 (0)	/	/	/	/	/	/	/	/	/	/	/	/
7 (0)	/	/	/	/	/	/	/	/	/	/	/	/
8 (0)	/	/	/	/	/	/	/	/	/	/	/	/
9 (0)	/	/	/	/	/	/	/	/	/	/	/	/
10 (0)	/	/	/	/	/	/	/	/	/	/	/	/
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.97**

(c) Training accuracies over epochs

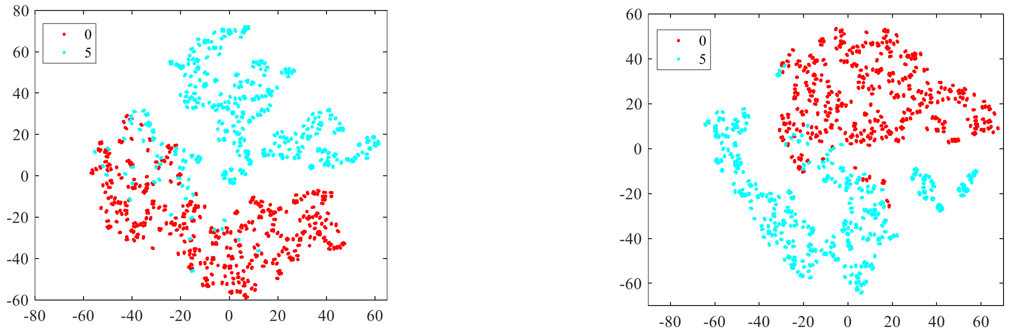
(d) Classification results of Beam C

Figure 7.20 Accuracy and classification results of Beam C using the RADA

Table 7.3 presents the variations of the weight parameter in the iteration process in the DA between beam A and C. The first and sixth entries in **W** converge to near one, whereas the rest elements approach zero, reflecting that the label space of the target domain is a small subset of the source domain. Particularly, the first entry in **W** corresponds to the undamaged condition, and the sixth corresponds to the damage in the fifth element. Therefore, the optimization results are consistent with the actual DS. The t-SNE feature visualization results are shown in Figure 7.21, where two categories in the target data are separated.

Table 7.3 Variations of **W** in RADA with inconsistent label spaces (from Beam A to C)

Epochs	**W**	Accuracies of C_1 /C_2
0	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]	0.00% / 0.04%
1	[0.84, 0.91, 0.91, 1.00, 0.86, 0.84, 0.95, 0.85, 0.80, 0.89, 0.81]	1.53% / 1.67%
50	[0.22, 0.34, 0.05, 0.13, 0.36, 0.22, 0.00, 0.05, 0.01, 0.10, 1.00]	43.5% / 7.50%
100	[0.94, 0.02, 0.00, 0.04, 0.01, 1.00, 0.00, 0.03, 0.03, 0.00, 0.00]	89.6% / 89.6%
150	[1.00, 0.01, 0.00, 0.00, 0.00, 0.97, 0.00, 0.00, 0.00, 0.00, 0.00]	95.7% / 96.0%
200	[1.00, 0.00,0.00, 0.00, 0.00, 0.99, 0.00, 0.00, 0.00, 0.00, 0.00]	97.1% / 96.9%



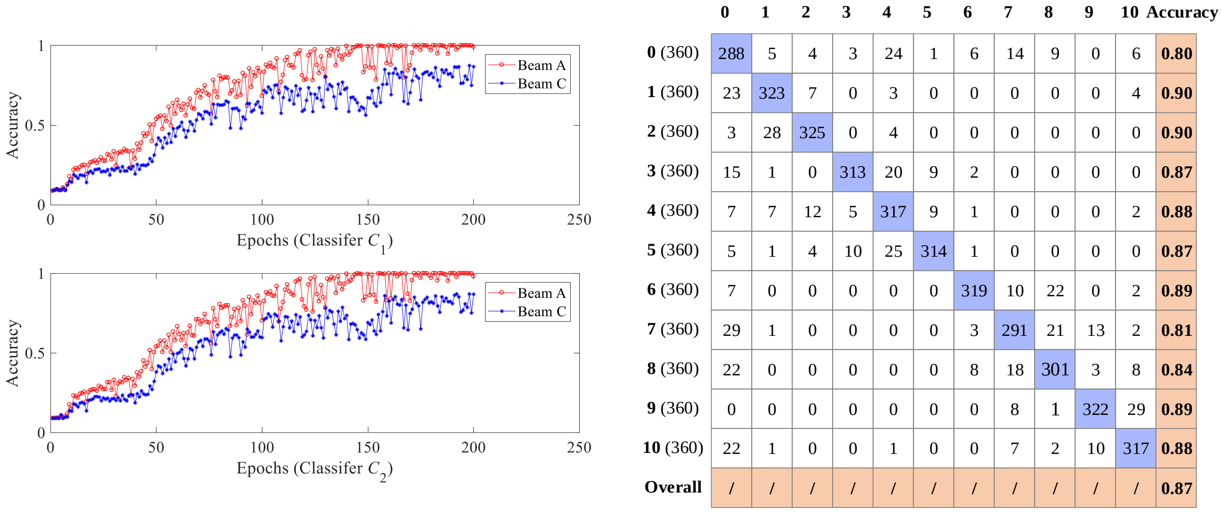
(c) Target data from Beam B

(d) Target data from Beam C

Figure 7.21 Feature visualization via t-SNE using the RADA

(3) <u>Comparison with DA without the re-weighting mechanism</u>

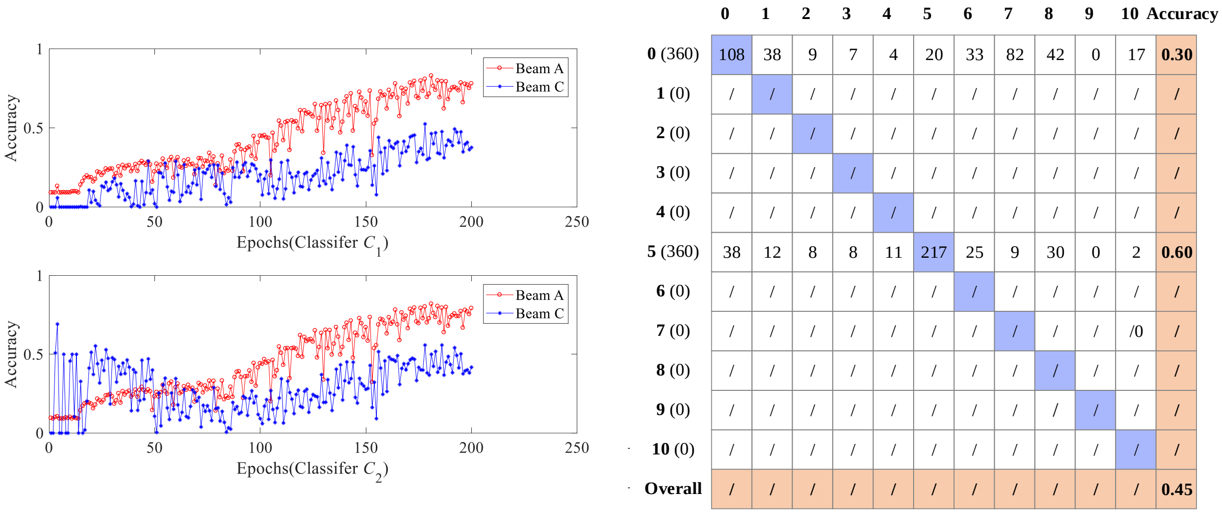
For comparison, the performance of the adversarial DA network without the re-weighting mechanism is also investigated. The network is referred to as the common adversarial DA (CADA). All parameters are set identically to the RADA network, except that Step B and the weight parameter **W** in Step C in Section 3.3 are neglected. The results of the CADA on Beam C for the problem with consistent label spaces are shown in Fig. 22. The accuracies are comparable to the results using the RADA network. By contrast, the results for the problem with inconsistent label spaces are depicted in Fig. 23. The accuracies of C_1 and C_2 drop to 45.3% and 44.7%, respectively. Table 4 compares the classification results using different methods, where the ensemble accuracy is calculated from the average of the outputs of C_1 and C_2. The RADA network with the re-weighting mechanism performs much better than the CNN without DA and CADA for problems with inconsistent label spaces.



	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Accuracy**
0 (360)	288	5	4	3	24	1	6	14	9	0	6	**0.80**
1 (360)	23	323	7	0	3	0	0	0	0	0	4	**0.90**
2 (360)	3	28	325	0	4	0	0	0	0	0	0	**0.90**
3 (360)	15	1	0	313	20	9	2	0	0	0	0	**0.87**
4 (360)	7	7	12	5	317	9	1	0	0	0	2	**0.88**
5 (360)	5	1	4	10	25	314	1	0	0	0	0	**0.87**
6 (360)	7	0	0	0	0	0	319	10	22	0	2	**0.89**
7 (360)	29	1	0	0	0	0	3	291	21	13	2	**0.81**
8 (360)	22	0	0	0	0	0	8	18	301	3	8	**0.84**
9 (360)	0	0	0	0	0	0	0	8	1	322	29	**0.89**
10 (360)	22	1	0	0	1	0	0	7	2	10	317	**0.88**
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.87**

(a) Training accuracies over epochs (b) Classification results of Beam C

Figure 7.22 Accuracy and classification results of Beam C using CADA for problems with consistent label spaces



	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Accuracy**
0 (360)	108	38	9	7	4	20	33	82	42	0	17	**0.30**
1 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
2 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
3 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
4 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
5 (360)	38	12	8	8	11	217	25	9	30	0	2	**0.60**
6 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
7 (0)	/	/	/	/	/	/	/	/	/	/	/0	**/**
8 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
9 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
10 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
Overall	**/**	**/**	**/**	**/**	**/**	**/**	**/**	**/**	**/**	**/**	**/**	**0.45**

(a) Training accuracies over epochs (b) Classification results of Beam C

Figure 7.23 Accuracy and classification results of Beam C using CADA for problems with inconsistent label spaces

Table 7.4 Identification accuracy for damage localization task

Label	Method	Beam B			Beam C		
		C_1	C_2	Ensemble	C_1	C_2	Ensemble
Consistent label spaces	RADA	88.6%	89.2%	89.1%	84.9%	85.3%	85.2%
	CNN	37.2%	37.2%	37.2%	15.9%	16.1%	16.0%
	CADA	88.0%	87.2%	87.6%	86.6%	86.8%	86.6%
Inconsistent label spaces	RADA	94.6%	95.3%	95.1%	97.1%	96.9%	97.0%
	CNN	17.1%	16.7%	16.9%	1.94%	1.39%	1.53%
	CADA	51.1%	49.7%	50.6%	45.3%	44.7%	45.1%

Note: CNN refers to the network trained on source data only without DA, and CADA refers to the common adversarial DA network without a re-reweighting mechanism.

7.4.1.2 Damage quantification

The damage severity is quantified on the basis of the damage localization results. After the damage has been localized, the source data involved in the DA process contain those with specified locations merely. The irrelevant source data are removed in advance, which improves the efficiency of the DA and mitigates the risk of negative transfer. In practice, the damage severity may be minor in the initial stage. Detecting the minor damage and issuing alarming are of practical importance and challenging as the minor damage causes small changes in the modal data. Here we suppose that damage occurs in the fifth element and only consider five damage levels from 10% to 50%. At each damage level, 400 excitations are applied repeatedly. Accordingly, the source data comprise 2,000 data samples (400 excitations × 5 damage levels). Five patterns (5 damage levels) are labeled by the one-hot encoding method. Correspondingly, the output of the RADA network has a size of 5×1. The inconsistent label spaces are designed between the source and target data. The target data are assumed to contain three categories only, as shown in Table 7.5. The target data contain 1200 data samples (400 excitations × 3 damage levels) with the potential damage levels to be quantified. The batch size is set to be 50.

Table 7.5 Damage levels and labels of the target data

Damaged element	Category no.	Elemental damage level	Label of target data
5	0	[−10%],	[1,0,0,0,0]
	2	[−30%],	[0,0,1,0,0]
	3	[−40%],	[0,0,0,1,0]

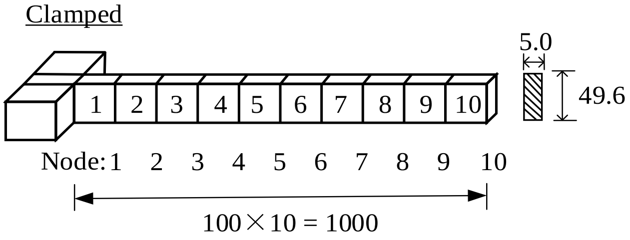
The damage quantification accuracies of Beam B and Beam C adopting the RADA network are summarized in Table 7.6. Although the damage quantification results are not as accurate as those in the single-damage localization task, the proposed RADA technique obviously improves the accuracy compared with that directly using the network without DA. The results reflect that damage severity-sensitive features are more difficult for the network to directly learn from structural acceleration responses under ambient excitations. Data pre-processing or physics-guided model training methods (Zhang et al., 2020) may be explored in the future to further improve the damage quantification results.

Table 7.6 Identification accuracy for damage quantification task

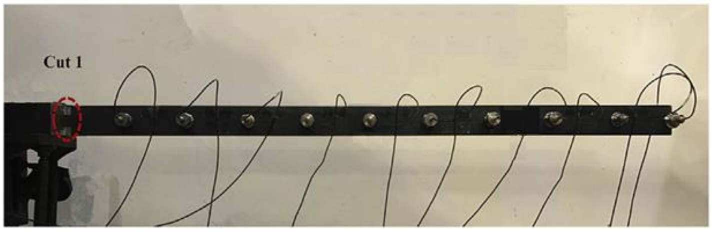
Method	Beam B			Beam C		
	C_1	C_2	Ensemble	C_1	C_2	Ensemble
RADA	62.8%	61.3%	61.7%	67.6%	67.2%	67.3%
CNN	2.11%	2.11%	2.11%	34.6%	34.6%	34.6%

7.4.2 Knowledge Transfer from Numerical to Experimental Structures

Many DL-based damage identification techniques work well on numerical simulations while fail on real structures, as the numerical FE model may have a discrepancy from the real structure. The main challenge of transferring knowledge from a numerical FE model to a real structure lies in the fact that the measurement data from the real structure are unlabeled and the DSs are always one or a small subset of all DSs. Consequently, the target data from the real structure have much fewer label categories than the source data from the FE model. In this section, the proposed RADA technique is applied to migrate the network trained from the numerical FE model to the experimental structure. A laboratory-tested steel cantilever beam (Hou et al., 2018) shown in Figure 7.24 is studied. The length and cross-section of the beam are 1.0 m and 49.6 mm × 5.0 mm, respectively. The Young's modulus of the material is estimated to 2.0×10^{11} N/m^2, and mass density is 7.67×10^3 kg/m^3. The beam is divided into 10 equal Euler-Bernoulli beam elements.



(a) FE model (unit: mm)



(b) Laboratory-tested beam

Figure 7.24 Configuration of the cantilever beam

The target data are collected from the experiment. The beam was excited by an instrumented hammer with a rubber tip. Ten accelerometers were evenly installed on the beam to record the acceleration responses with the sampling frequency of 2,000 Hz. Ten-second structural responses (20,000 data points) to the hammer excitation were recorded. Twenty sets of hammer excitations were conducted. To increase the data samples of the experiment, each data segment is divided into 20 samples with a size of 1000 × 10, resulting in 400 sets of data samples in the undamaged state. Damage was then simulated by introducing a cut in the first element in sequence, as listed in Table 7.7. Correspondingly, the equivalent stiffness reduction of the first element is 10% and 22% for the two DSs. The beam was tested similarly in each DS. Correspondingly, the experimental data have 1200 samples (400 samples × 3 DSs) in total.

Table 7.7 Damage locations and severities for three DSs

DS	Element no.	Cut width (mm)	Cut depth (mm)	Element stiffness reduction	Sample no.
Undamaged	–	–	–	–	400
DS1	1	10	20	−10%	400
DS2	1	10	30	−22%	400

The source data are generated from the FE model. To minimize the modeling error, the FE model is first updated using the measured first six frequencies and mode shapes by minimizing the discrepancy between the measured modal data and model predictions in the intact condition (Zhou et al., 2015):

$$J = \sum_{r=1}^{N_m} \left(\frac{\hat{\lambda}_r - \lambda_r(\boldsymbol{\theta})}{\hat{\lambda}_r} \right)^2 - \sum_{r=1}^{N_m} \left\| \hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r(\boldsymbol{\theta}) \right\|_2^2 \tag{7.13}$$

where $N_m = 6$. The updated FE model is then utilized to generate the source data.

7.4.2.1 Damage localization

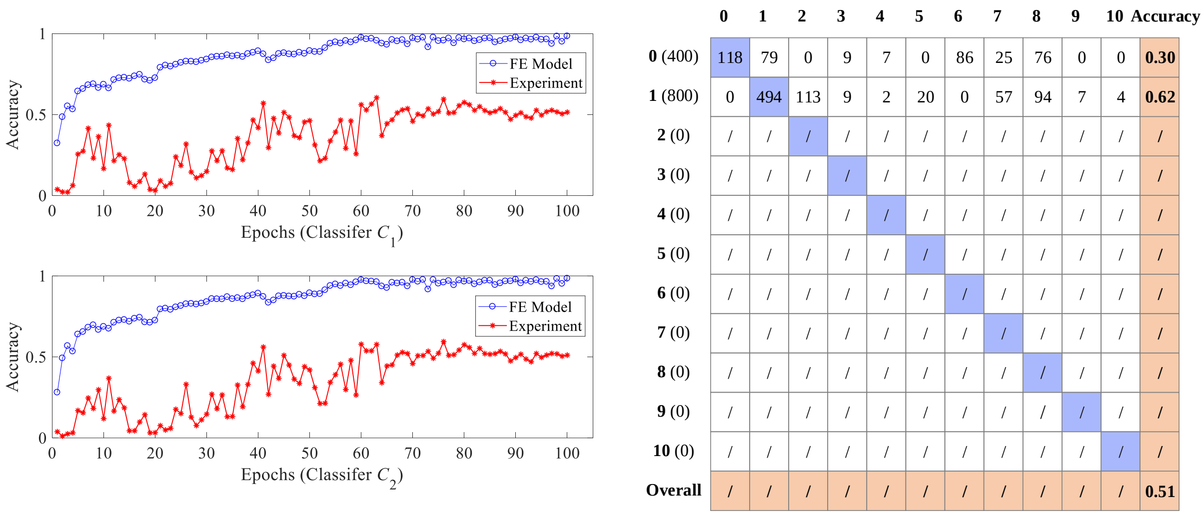
To generate the source data in the undamaged state, an impact force simulating the hammer excitation is applied to nine locations of the first element of the FE model and repeated four times at each location. Newmark's method is utilized to calculate 0.5-sec structural responses with a time step of 5×10^{-4} s. A total of 25 damage levels are then simulated at each element by reducing the elemental stiffness by 2%, 4%, and up to 50%. The structural responses to the impact force are then similarly calculated using Newmark's method. The source data thus contain 9,900 data samples in total (4 excitations × 9 excitation positions × (25 damage levels × 10 damage locations + 25 × 1 intact condition). All data samples are labeled according to the damage locations. The label has a size of 11×1, where the first entry being 1 represents the intact condition, and the rest represent the damage locations.

Three DSs in the target data with 1,200 samples are also labeled, as shown in Table 7.8. Data samples are directly used in the two-dimensional form without any concatenation. Consequently, the size of the network input is 1000×10. The proposed RADA network can be directly used without re-designing parameters when the input size changes, as explained in Section 3.1. The labels of the target data are assumed to be unknown in the model training process and will be detected using the proposed RADA technique.

Table 7.8 Labels of target data in the tasks of damage localization and quantification

Damage scenario	Damage localization		Damage quantification	
	Sample no.	Labels (size: 11×1)	Sample no.	Labels (size: 10×1)
Undamaged	400	[1,0,0,0,0,0,0,0,0,0,0]	—	—
DS1	800	[0,1,0,0,0,0,0,0,0,0,0]	400	[0,1,0,0,0,0,0,0,0,0]
DS2			400	[0,0,0,0,1,0,0,0,0,0]

The network is first trained on the source data without DA. The batch size is set to 50. The convergence is achieved within 50 epochs, as shown in Figure 7.25, which is much faster than that in Section 4.1. The reason may be that the structural responses to the impact hammer force in this example are easier for the network to learn than those under the white Gaussian noise excitation. The accuracy is 51.1% when the model trained on the source data directly applied to the target data.



	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (400)	118	79	0	9	7	0	86	25	76	0	0	**0.30**
1 (800)	0	494	113	9	2	20	0	57	94	7	4	**0.62**
2 (0)	/	/	/	/	/	/	/	/	/	/	/	/
3 (0)	/	/	/	/	/	/	/	/	/	/	/	/
4 (0)	/	/	/	/	/	/	/	/	/	/	/	/
5 (0)	/	/	/	/	/	/	/	/	/	/	/	/
6 (0)	/	/	/	/	/	/	/	/	/	/	/	/
7 (0)	/	/	/	/	/	/	/	/	/	/	/	/
8 (0)	/	/	/	/	/	/	/	/	/	/	/	/
9 (0)	/	/	/	/	/	/	/	/	/	/	/	/
10 (0)	/	/	/	/	/	/	/	/	/	/	/	/
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.51**

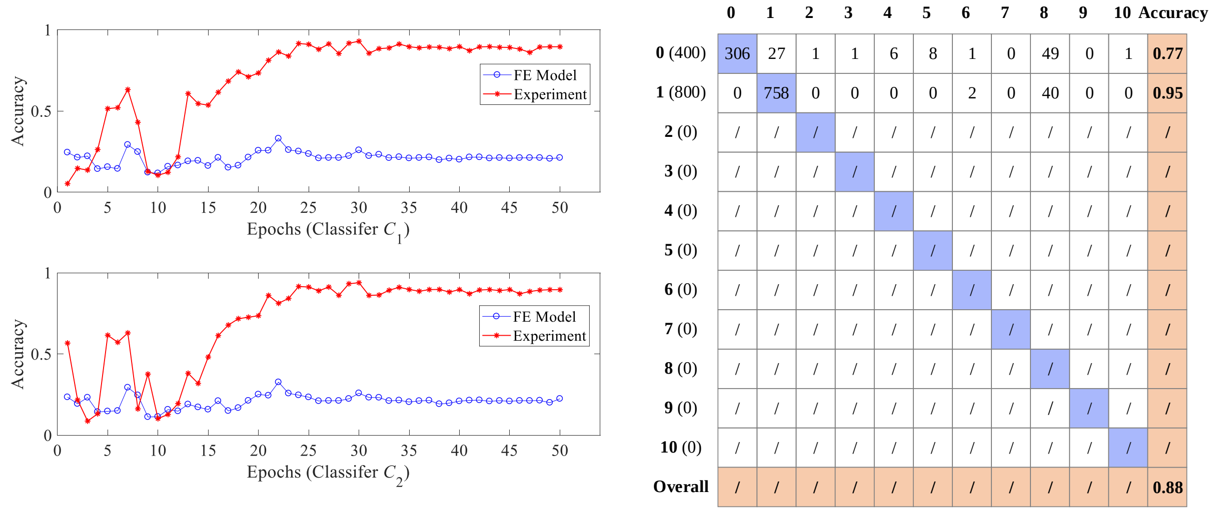
(a) Accuracies over epochs (b) Classification result without DA

Figure 7.25 Accuracy and classification results on the experimental data using CNN without DA

The DA between the FE model and real beam is then conducted using the proposed RADA technique. Table 7.9 presents the variation of **W** in the DA process. Upon convergence, the classification accuracies of two classifiers C_1 and C_2 improve to 88.6% and 88.7%, respectively, as shown in Figure 7.24. We note that the classification accuracy of the source data in Figure 7.26 is relatively low, which is only 21%. The reason may be that the network trained in Step A is re-updated in Step C in the training process, and then the network focuses on the shared label spaces while neglects the outlier label space in the source domain. Nevertheless, this does not affect the classification accuracy of the unlabeled target data, which is the focus of this study.

Table 7.9 Variations of **W** in RADA for damage localization task

Epochs	**W**	Accuracies of C_1 /C_2
0	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]	5.33% / 56.8%
10	[0.84, 0.91, 0.91, 1.00, 0.86, 0.84, 0.95, 0.85, 0.80, 0.89, 0.81]	10.5% / 10.3%
30	[0.46, 1.00, 0.00, 0.00, 0.02, 0.00, 0.00, 0.03, 0.23, 0.00, 0.00]	91.0% / 90.9%
50	[0.44, 1.00,0.00, 0.00, 0.01, 0.01, 0.00, 0.00, 0.26, 0.00, 0.00]	88.6% /88.7%



	0	1	2	3	4	5	6	7	8	9	10	Accuracy
0 (400)	306	27	1	1	6	8	1	0	49	0	1	**0.77**
1 (800)	0	758	0	0	0	0	2	0	40	0	0	**0.95**
2 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
3 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
4 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
5 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
6 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
7 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
8 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
9 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
10 (0)	/	/	/	/	/	/	/	/	/	/	/	**/**
Overall	/	/	/	/	/	/	/	/	/	/	/	**0.88**

(a) Accuracies over epochs

(b) Classification result without DA

Figure 7.26 Accuracy and classification results of damage localization using RADA

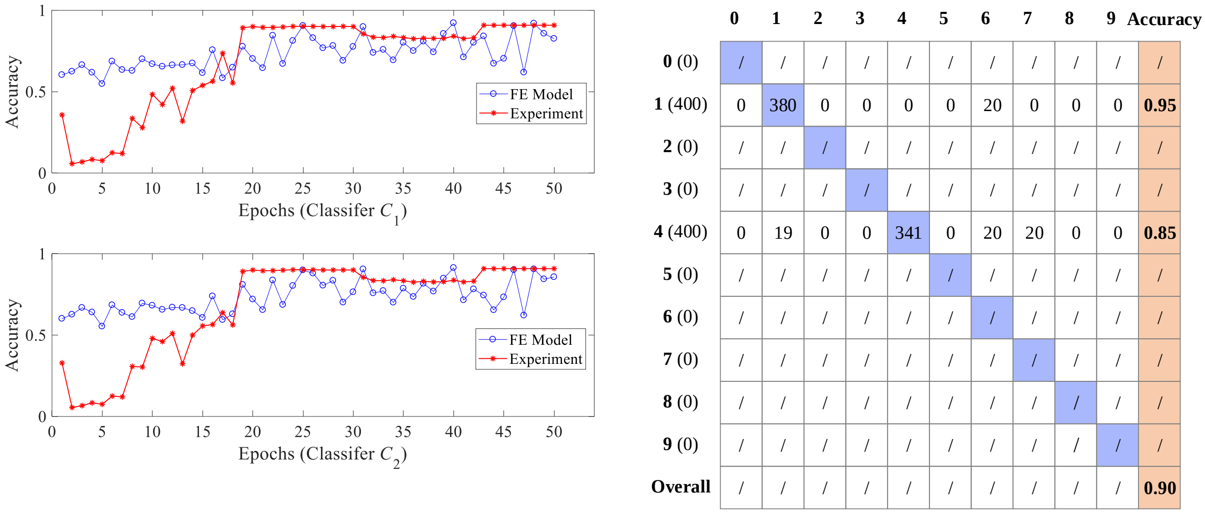
7.4.2.2 Damage quantification

After the damage has been localized, the damage severity is then quantified. The source data merely consider the case where the damage occurs in the first element. Fifty levels of elemental stiffness reduction from 1%, 2%, and up to 50% are contained in the source data. The hammer excitation is similarly applied to nine locations between the fixed boundary and the first node and repeats 20 times at each location. Thus, we obtain 9,000 data samples (50 damage levels × 9 excitation positions × 20 excitations) in total. All samples are classified into 10 categories in sequence, each having five damage levels and 900 samples. For example, the first category contains the damage levels from 1% to 5%, and the second contains the damage levels from 6% to 10%. Ten categories are labeled by the one-hot encoding method. Correspondingly, the output of the RADA network has a size of 10×1. The target data only contain two DSs with 800 samples, of which 400 samples correspond to 10% damage severity and the rest 400 samples correspond to 22% damage severity. The labels of the target data to be predicted are listed in the last column of Table 7.10.

The proposed RADA technique is then applied, and the variation of **W** is listed in Table 7.10. The accuracy and classification results are shown in Figure 7.27. DS1 is labeled "1", which is corresponding to the damage severity from 6% and 10%. DS2 is labeled "4", corresponding to the damage severity from 21% and 25%. Upon convergence, two DSs are quantified with the accuracy of 95% and 85%. For comparison, the CNN trained on the FE model without DA is also applied. The damage identification results are compared in Table 7.11. The RADA method significantly improves the damage identification accuracy compared with the CNN without DA, indicating that the proposed RADA technique enhances the application of the DL technique to the actual structural damage detection. The damage data with labels can be collected easily and economically from the FE model, and the knowledge can be then transferred to practical structures for damage detection without collecting the labeled data from the structures in advance.

Table 7.10 Variations of W in RADA for damage quantification task

Epochs	**W**	Accuracies of C_1 / C_2
0	[1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00]	46.8% / 44.1%
1	[0.11, 0.62, 0.14, 0.14, 1.00, 0.22, 0.36, 0.47, 0.14, 0.37]	33.6% / 30.8%
20	[0.00, 1.00, 0.00, 0.00, 0.85, 0.00, 0.10, 0.05, 0.00, 0.00]	90.1% / 90.1%
31	[0.00, 0.96, 0.00, 0.00, 1.00, 0.00, 0.12, 0.27, 0.00, 0.00]	85.5% / 85.5%
50	[0.00, 0.99,0.00, 0.00, 1.00, 0.00, 0.10, 0.03, 0.00, 0.00]	90.1% / 90.0%



	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Accuracy**
0 (0)	/	/	/	/	/	/	/	/	/	/	/
1 (400)	0	380	0	0	0	0	20	0	0	0	**0.95**
2 (0)	/	/	/	/	/	/	/	/	/	/	/
3 (0)	/	/	/	/	/	/	/	/	/	/	/
4 (400)	0	19	0	0	341	0	20	20	0	0	**0.85**
5 (0)	/	/	/	/	/	/	/	/	/	/	/
6 (0)	/	/	/	/	/	/	/	/	/	/	/
7 (0)	/	/	/	/	/	/	/	/	/	/	/
8 (0)	/	/	/	/	/	/	/	/	/	/	/
9 (0)	/	/	/	/	/	/	/	/	/	/	/
Overall	/	/	/	/	/	/	/	/	/	/	**0.90**

(a) Accuracies over epochs

(b) Classification result using RADA

Figure 7.27 Accuracy and classification results of damage quantification using RADA

Table 7.11 Identification accuracy in tasks of damage localization and quantification

Task	Method	C_1	C_2	Ensemble
Damage localization	RADA	88.6%	88.7%	88.7%
	CNN	51.4%	51.1%	51.0%
Damage quantification	RADA	90.1%	90.0%	90.0%
	CNN	18.6%	18.5%	18.5%

7.5 Summary

This chapter develops a RADA network to transfer the knowledge between structures for damage detection, so that the DL network trained on one structure can be applied to other structures without re-collecting labeled data. The network is composed of a feature generator with the DenseNet architecture and two classifiers with fully connected layers. In view of inconsistent DSs between different structures, a re-weighting mechanism is designed to mitigate negative transfer. The entire network is trained in an adversarial manner to extract domain-invariant and damage-sensitive features for damage detection.

The proposed network is applied to two types of knowledge transfer, namely, from one structure to the other with different sizes and from a numerical model to an experimental structure. Both examples indicate that the RADA network significantly improves the classification accuracy of the unlabeled target data in comparison with that of directly using the network without DA. The proposed RADA technique enhances the application of the DL technique to damage detection of practical structures.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

Structural damage identification methods based on probabilistic ML and Bayesian inference have been developed in this study. The Bayesian theorem is first introduced to the data-driven data normalization techniques. The environmental effects on structural vibration properties are discriminated from the damage detection using the sparse Bayesian FA and an improved probabilistic kernelized method. Next, an SBL technique is developed for structural damage detection using the normalized damage sensitive modal parameters. The damage index is estimated by employing the Laplace approximation and the combination of VBI and DRAM algorithms. Subsequently, An RADA method is developed for unsupervised damage localization and quantification. The results and findings are summarized as follows.

The Bayesian inference-based linear and nonlinear damage detection methods have been developed.

1. The sparse Bayesian FA technique is able to assess the structural condition under unknown and changing environmental conditions. The ARD prior defined on the factor loading matrix can automatically determine the number of underlying environmental factors that significantly affect the structural vibration properties. All parameters and unknowns in the Bayesian probabilistic framework are optimized automatically by the EM method.
2. The probabilistic kernelized method is able to eliminate nonlinear environmental effects by defining the Gaussian kernel to introduce nonlinearity. The nonlinear

mapping function does not need to be defined explicitly. The probabilistic method outperforms the non-probabilistic kernel PCA-based method in two aspects. First, the unknown kernel parameters are estimated automatically in the Bayesian probabilistic framework. Second, the computation efficiency is significantly improved.

The SBL-based model updating has been developed for damage localization and quantification at the element level.

1. The SBL method based on the Laplace approximation technique has the advantage that the variable and hyper-parameters associated with the uncertainties are derived in a closed form without numerical sampling.
2. The VBI-DRAM method is not limited to the Gaussian distribution but applies to all kinds of distributions. Besides, the computational efficiency is significantly improved for high-dimensional problems, as compared with the Laplace approximation technique.

An RADA method is developed to transfer knowledge from the FE model to the experimental structure and from one structure to the other with different sizes for damage detection using time series responses.

1. The proposed RADA network is able to learn damage-sensitive and domain-invariant features for damage detection by exploiting the labeled source data and unlabeled target data. The damage identification accuracy of the unlabeled target data is significantly improved after the DA process, as compared with that of directly using the DL model trained on the source data without DA. By using the proposed RADA method, re-collecting labeled data is not needed for a new structure. The damage knowledge can be transferred even the structure discrepancy exists.
2. The weight parameter designed in the loss function increases the importance of shared label space and downplays the outlier label space in the DA process,

facilitating the positive transfer. The proposed method outperforms the DA without the re-weighting mechanism when the structures have inconsistent label spaces.

8.2 Future Work

Despite the improvements made, the methods developed in this thesis can be improved and extended in the following aspects:

1. The RADA method developed for damage knowledge transfer is limited to homogeneous structures with identical topology and material. For the heterogeneous structures with significant differences, for example, from a two-story frame to a three-story frame or from a two-span bridge to a three-span bridge, how to attain the effective and positive knowledge transfer will be further explored.
2. The damage localization in the RADA method is processed by a single-label classification task. However, structural damage may occur in multiple locations. Accordingly, the damage localization problem becomes a multi-label supervised classification task. The multiple damage localization task can be simplified to the damage existence determination of each location by using the binary entropy function. However, such processing treating the label of each location independent may not be realistic. The damage localization using DL techniques needs further improvement.
3. The global damage-sensitive features (i.e., frequency and mode shape) are not sensitive to minor damages in large-scale structures. The RADA method adopts the black-box DL model, thus the damage-sensitive feature extracted by the feature generator are not interpretable. New interpretable damage-sensitive features that are sensitive to minor damages may be explored in the future with the aid of interpretable ML methods.
4. For large-scale civil structures whose FE models may contain thousands of elements, the methods developed in this thesis for damage localization and

quantification in the elemental level need further studies. For example, the damage is sparser, and the problem indeterminacy may increase in the model-based methods. Besides, the label indicating the damage location will be of a large size in the DL-based method and the model training process will be very time consuming. The methods should be further improved for the damage identification of real large-scale structures.

REFERENCES

Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M, Inman DJ. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration* 2017; 388: 154–70.

Abdeljaber O, Avci O, Kiranyaz MS, Boashash B, Sodano H, Inman DJ. 1D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing* 2018; 275: 1308–17.

Alampalli S. Effects of testing, analysis, damage, and environment on modal parameters. *Mechanical Systems and Signal Processing* 2000; 14: 63–74.

Allemang RJ. The modal assurance criterion–twenty years of use and abuse. *Sound and Vibration* 2003; 37(8): 14–23.

Andrieu C, de Freitas N, Doucet A, Doucet A, Jordan MI. An introduction to MCMC for machine learning. *Machine Learning* 2003; 50: 5–43.

Arangio S, Beck JL. Bayesian neural networks for bridge integrity assessment. *Structural Control and Health Monitoring* 2012; 19: 3–21.

Au SK. Probabilistic failure analysis by importance sampling Markov chain simulation. *Journal of Engineering Mechanics* 2004; 130: 303–11.

Au SK, Beck JL. Important sampling in high dimensions. *Structural Safety* 2003; 25: 139–63.

Avci O, Abdeljaber O, Kiranyaz S, Hussein M, Inman DJ. Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks.

Journal of Sound and Vibration 2018; 424: 158–72.

Babacan SD, Molina R, Katsaggelos AK. Bayesian compressive sensing using Laplace priors. *IEEE Transactions on Image processing* 2009; 19(1): 53–63.

Baldacchino T, Cross EJ, Worden K, Rowson J. Variational Bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems. *Mechanical Systems and Signal Processing* 2016; 66: 178–200.

Bao YQ, Tang ZY, Li H, Zhang YF. Computer vision and deep learning-based data anomaly detection method for structural health monitoring. S*tructural Health Monitoring* 2019; 18: 401–21.

Bao YQ, Xia Y, Li H, Xu YL, Zhang P. Data fusion-based structural damage detection under varying temperature conditions. *International Journal of Structural Stability and Dynamics* 2012; 12(06) 1250052.

Bartholomew D J, Knott M, Moustaki I. *Latent Variable Models and Factor Analysis: A Unified Approach*. John Wiley & Sons, 2011.

Beck JL, Katafygiotis LS. Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics* 1998; 124: 455–61.

Behmanesh I, Moaveni B, Lombaert G, Papadimitriou C. Hierarchical Bayesian model updating for structural identification. *Mechanical Systems and Signal Processing* 2015; 64: 360–76.

Bendat JS, Piersol AG. *Engineering Applications of Correlation and Spectral Analysis*. Wiley: New York, 1980.

Bishop CM. Bayesian PCA. *Advances in Neural Information Processing Systems* 1999; 382–388.

Bishop CM. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006.

Box GE. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics* 1954a; 25: 290–302.

Box GE. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics* 1954b; 25: 484–498.

Cao ZJ, Ma LJ, Long MS, Wang JM. Partial adversarial domain adaptation, *Proceedings of the European Conference on Computer Vision (ECCV)* 2018; 135–150.

Chen ZP, Yu L. A novel PSO-based algorithm for structural damage detection using Bayesian multi-sample objective function. *Structural Engineering and Mechanics* 2017; 63: 825–35.

Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. *American Statistician* 1995; 49: 327–335.

Ching J, Beck JL. New Bayesian model updating algorithm applied to a structural health monitoring benchmark. *Structural Health Monitoring* 2004; 3(4): 313–332.

Ching JY, Chen YC. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *Journal of Engineering Mechanics* 2007; 133: 816–32.

Ching JY, Muto M, Beck JL. Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. *Computer-Aided Civil and Infrastructure Engineering* 2006; 21: 242–57.

Cofre-Martel S, Kobrich P, Droguett EL, Meruane V. Deep Convolutional Neural

Network-Based Structural Damage Localization and Quantification Using Transmissibility Data. *Shock and Vibration* 2019; 2019: 9859281.

Damianou A, Titsias MK, Lawrence ND. Variational Gaussian process dynamical systems. *In Advances in Neural Information Processing Systems* 2011; 2510–2518.

Daubechies I, DeVore R, Fornasier M, Güntürk CS. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 2010, 63(1): 1–38.

Sen D, Erazo K, Zhang W, Nagarajaiah S, Sun LM. On the effectiveness of principle component analysis for decoupling structural damage and environmental effects in bridge structures. *Journal of Sound and Vibration* 2019; 457: 280–298.

Deraemaeker A, Worden K. A comparison of linear approaches to filter out environmental effects in structural health monitoring. *Mechanical Systems and Signal Processing* 2018; 105: 1–15.

Ding YL, Li AQ. Temperature-induced variations of measured modal frequencies of steel box girder for a long-span suspension bridge. *International Journal of Steel Structures* 2011; 11: 145–55.

Doebling SW, Farrar CR, Prime MB. A summary review of vibration-based damage identification methods. *Shock and Vibration Digest* 1998; 30(2): 91–105.

Doebling SW, Farrar CR, Prime MB, Shevitz DW. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review, *Los Alamos National Laboratory Report* 1996.

Farrar CR, Doebling SW, Cornwell PJ, Straser EG. Variability of modal parameters measured on the Alamosa Canyon bridge. *Los Alamos National Laboratory* 1996; NM (United States).

Farrar CR, Worden K. *Structural Health Monitoring: A Machine Learning Perspective*. John Wiley & Sons, 2012.

Formenti D, Richardson M. Parameter estimation from frequency response measurements using rational fraction polynomials (twenty years of progress). *Proceeding of International Modal Analysis Conference* 2020; 4753: 373–382.

Fujimoto K, Satoh A, Fukunaga S. System identification based on variational Bayes method and the invariance under coordinate transformations. *IEEE Conference on Decision and Control and European Control Conference, IEEE* 2011; 3882–3888.

Fujino Y, Abe M, Shibuya H, Yanagihara M, Sato M. Monitoring of Hakucho suspension bridge using ambient vibration. *Proceedings of Workshop on Research and Monitoring of Long Span Bridges* 2000; Hong Kong, 142–149.

Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation. *International conference on machine learning, PMLR,* 2015; 1180–1189.

Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky V. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 2016, 17(1): 2096–2030.

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

Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT press, 2016.

Graf W, Freitag S, Kaliske M, Sickert JU. Recurrent neural networks for uncertain time-dependent structural behavior. *Computer-Aided Civil and Infrastructure Engineering* 2010; 25(5): 322–333.

Green PJ, Mira A. Delayed rejection in reversible jump Metropolis-Hastings.

Biometrika 2001; 88: 1035–1053.

Gulgec NS, Takac M, Pakzad SN. Convolutional Neural Network Approach for Robust Structural Damage Detection and Localization. *Journal of Computing in Civil Engineering* 2019; 33(3): 04019005.

Haario H, Laine M, Mira A, Saksman E. DRAM: Efficient adaptive MCMC. *Statistics and Computing* 2006; 16: 339–354.

Haario H, Saksman E, Tamminen J. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics* 1999; 14: 375–395.

Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm, *Bernoulli* 2001; 7: 223–242.

Hjelmstad KD, Shin S. Damage detection and assessment of structures from static response. *Journal of Engineering Mechanics* 1997; 123(6): 568–576.

Hsu TY, Loh CH. Damage detection accommodating nonlinear environmental effects by nonlinear principal component analysis. *Structural Control and Health Monitoring* 2010; 17: 338–354.

Hou R, Xia Y, Bao Y, Zhou X. Selection of regularization parameter for *l*1-regularized damage detection. *Journal of Sound and Vibration* 2018, 423: 141–160.

Hou R, Xia Y, Zhou X, Huang Y. Sparse Bayesian learning for structural damage detection using expectation-maximization technique. *Structural Control and Health Monitoring* 2019; 26(5): e2343.

Hou R, Wang X, Xia Q, Xia Y. Sparse Bayesian learning for structural damage detection under varying temperature conditions. *Mechanical Systems and Signal Processing* 2020, 145: 106965.

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2017; 4700–4708.

Huang Y, Beck JL. Hierarchical sparse Bayesian learning for structural health monitoring with incomplete modal data. *International Journal for Uncertainty Quantification* 2015; 5:139–169.

Huang Y, Beck JL, Li H. Hierarchical sparse Bayesian learning for structural damage detection: theory, computation and application. *Structural Safety* 2017; 64: 37–53.

Huang Y, Beck JL, Li H. Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment. *Computer Methods in Applied Mechanics and Engineering* 2017; 318: 382–411.

Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning, PMLR*, 2015; 448–456.

Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science* 2015; 349(6245): 255–260.

Katafygiotis LS, Beck JL. Updating models and their uncertainties. II: Model identifiability. *Journal of Engineering Mechanics* 1998; 124(4): 463–467.

Katafygiotis LS, Papadimitriou C, Lam HF. A probabilistic approach to structural model updating. *Soil Dynamic Earthquake Engineering* 1998; 17: 495–507.

Keerthi SS, Lin CJ. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* 2003, 15(7): 1667–1689.

Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? *ArXiv* 2017; 1703: 04977.

Khodabandehlou H, Pekcan G, Fadali MS. Vibration-based structural condition assessment using convolution neural networks. *Structural Control and Health Monitoring* 2019; 26(2): e2308.

Kim CY, Jung DS, Kim NS, Yoon JG. Effect of vehicle mass on the measured dynamic characteristics of bridges from traffic-induced vibration test. *Proceedings-SPIE the International Society for Optical Engineering* 2001; 4359:1106–1111.

Kingma DP, Ba J. Adam: A method for stochastic optimization. *ArXiv* 2014; 1412.

Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 2012, 25: 1097–1105.

Kullaa J. Damage detection of the Z24 bridge using control charts. *Mechanical Systems and Signal Processing* 2003, 17(1): 163–170.

Kullaa J. Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring. *Mechanical Systems and Signal Processing* 2011; 25: 2976–89.

Kullaa J. Structural Health Monitoring under Nonlinear Environmental or Operational Influences. *Shock and Vibration* 2014; 863494.

Lam HF, Yang JH. Bayesian structural damage detection of steel towers using measured modal parameters. *Earthquakes and Structures* 2015; 8: 935–56.

Lam HF, Yang JH, Au SK. Bayesian model updating of a coupled-slab system using field test data utilizing an enhanced Markov chain Monte Carlo simulation algorithm. *Engineering Structures* 2015; 102: 144–55.

Lawrence N. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *Journal of Machine Learning Research* 2005; 6(11) 1783–1816.

Lawrence N, Seeger M, Herbrich R. Fast sparse Gaussian process methods: The informative vector machine. *Proceedings of the 16th Annual Conference on Neural Information Processing Systems* 2003; 609–616.

LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521: 436–444.

Lee JM, Yoo CK, Choi SW, Vanrolleghem PA, Lee IB. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science* 2004; 59: 223–234

Li B, Der Kiureghian A. Operational modal identification using variational Bayes. *Mechanical Systems and Signal Processing* 2017; 88: 377–398.

Li B, Der Kiureghian A, Au SK. A Gibbs sampling algorithm for structural modal identification under seismic excitation. *Earthquake Engineering and Structural Dynamics* 2018; 47(14): 2735–2755.

Lin M, Chen Q, Yan S. Network in network. *ArXiv* 2013; 1312.

Lin YZ, Nie ZH, Ma HW. Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering* 2017; 32: 1025–1046.

Lin YZ, Nie ZH, Ma HW. Dynamics-based cross-domain structural damage detection through deep transfer learning, *Computer-Aided Civil and Infrastructure Engineering* 2021.

Liu C, Harley JB, Berges M, Greve DW, Oppenheim IJ. Robust ultrasonic damage detection under complex environmental conditions using singular value decomposition. *Ultrasonics* 2015; 58: 75–86.

Liu H, Zhang YF. Deep learning-based brace damage detection for concentrically braced frame structures under seismic loadings. *Advances in Structural Engineering* 2019; 22: 3473–3486.

Mahmoud M, Abe M, Fujino Y. Analysis of suspension bridge by ambient vibration measurement using the time domain method and its application to health monitoring. *In Proceedings of SPIE, the International Society for Optical Engineering* 2001, 4359: 504–510.

Martins N, Caetano E, Diord S, Magalhães F, Cunha A. Dynamic monitoring of a stadium suspension roof: Wind and temperature influence on modal parameters and structural response. *Engineering Structures* 2014, 59: 80–94.

Mira A. On Metropolis-Hastings algorithms with delayed rejection. *Metron* 2001; 59: 231–241.

Moaveni B, Behmanesh I. Effects of changing ambient temperature on finite element model updating of the Dowling Hall Footbridge. *Engineering Structures* 2012; 43: 58–68.

Moser P, Moaveni B. Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mechanical Systems and Signal Processing* 2011; 25: 2336–2357.

Mottershead JE, Friswell MI. Model updating in structural dynamics: a survey. *Journal of Sound and Vibration* 1993; 167: 347–375.

Murphy KP. *Machine Learning: A Probabilistic Perspective*. 2012, MIT press.

Muto M, Beck JL. Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control* 2008; 14: 7–34.

Nagayama T, Abe M, Fujino Y, Ikeda K. Structural identification of a nonproportionally damped system and its application to a full-scale suspension bridge. *Journal of Structural Engineering* 2005; 131(10): 1536–1545.

Nagel JB, Sudret B. A unified framework for multilevel uncertainty quantification in Bayesian inverse problems. *Probabilistic Engineering Mechanics* 2016; 43: 68–84.

Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. *The International Conference on Machine Learning (ICML)* 2010.

Nelson RB. Simplified calculation of eigenvector derivatives. *AIAA Journal* 1976; 14:1201–1205.

Ni FT, Zhang J, Noori MN. Deep learning for data anomaly detection and data compression of a long-span suspension bridge. *Computer-Aided Civil and Infrastructure Engineering* 2020, 35(7): 685–700.

Ni YC, Zhang FL, Lam HF, Au SK. Fast Bayesian approach for modal identification using free vibration data, Part II-Posterior uncertainty and application. *Mechanical Systems and Signal Processing* 2016; 70: 221–244.

Ni YQ, Hua XG, Fan KQ, Ko JM. Correlating modal properties with temperature using long-term monitoring data and support vector machine technique. *Engineering Structures* 2005; 27: 1762–1773.

Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; 37(1): 41–59.

Oh BK, Glisic B, Kim Y, Park HS. Convolutional neural network-based wind-induced response estimation model for tall buildings. *Computer-Aided Civil and Infrastructure*

Engineering 2019; 34: 843–858.

Oh CK, Sohn H, Bae IH. Statistical novelty detection within the Yeongjong suspension bridge under environmental and operational variations. *Smart Materials and Structures* 2009; 18(12): 125022.

Pan H, Gui G, Lin Z, Yan, C. Deep BBN learning for health assessment toward decision-making on structures under uncertainties. *KSCE Journal of Civil Engineering* 2018, 22(3): 928–940.

Pan SJ, Tsang IW, Kwok JT, Yang Q. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks* 2010, 22: 199–210.

Pan SJ, Yang Q. A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering* 2009, 22(10): 1345–1359.

Parisi G. *Statistical Field Theory*. Addison-Wesley, 1988.

Pathirage CSN, Li J, Li L, Hao H, Liu WQ. Application of deep autoencoder model for structural condition monitoring. *Journal of Systems Engineering and Electronics* 2018; 29: 873–880.

Pathirage CSN, Li J, Li L, Hao H, Liu WQ, Ni PH. Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures* 2018; 172: 13–28.

Peeters B, De Roeck G. One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics* 2001; 30: 149–171.

Peeters B, Maeck J, De Roeck G. Vibration-based damage detection in civil engineering: excitation sources and temperature effects. *Smart Materials and Structures* 2001; 10: 518–527.

Perez-Ramirez CA, Amezquita-Sanchez JP, Valtierra-Rodriguez M, Adeli H, Dominguez-Gonzalez A, Romero-Troncoso RJ. Recurrent neural network model with Bayesian training and mutual information for response prediction of large buildings. *Engineering Structures* 2019; 178: 603–615.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in FORTRAN: the Art of Scientific Computing*. Cambridge University Press, 1992.

Rafiei MH, Adeli H. A novel unsupervised deep learning model for global and local health condition assessment of structures. *Engineering Structures* 2018; 156: 598–607.

Reynders E, Wursten G, De Roeck G. Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring* 2014; 13: 82–93.

Roberts GP, Pearson AJ. Health monitoring of structures-towards a stethoscope for bridges. *Proceedings of the 23rd International Seminar on Modal Analysis, Leuven, Belgium* 1999; 2: 947–952.

Ruotolo R, Surace C. Using SVD to detect damage in structures with different operational conditions. *Journal of Sound and Vibration* 1999; 226: 425–439.

Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2018; 3723–3732.

Salakhutdinov R, Hinton G. Deep boltzmann machines. *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, PMLR*, 2009; 5: 448–455.

Salawu OS. Detection of structural damage through changes in frequency: a review.

Engineering structures 1997; 19(9): 718–723.

Sedehi O, Papadimitriou C, Katafygiotis LS. Probabilistic hierarchical Bayesian framework for time-domain model updating and robust predictions. *Mechanical Systems and Signal Processing* 2019; 123: 648–673.

Seker S, Ayaz E, Turkcan E. Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery. *Engineering Applications of Artificial Intelligence* 2003; 16: 647–656.

Shang Z, Sun L, Xia Y, Zhang W. Vibration-based damage detection for bridges by deep convolutional denoising autoencoder. *Structural Health Monitoring* 2021; 20(4): 1880–1903.

Silverman BW. *Density Estimation for Statistics and Data Analysis*. Routledge, 2018.

Siringoringo DM, Fujino Y. System identification of suspension bridge from ambient vibration response. *Engineering Structures* 2008; 30: 462–477.

Sohn H. Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* 2007; 365(1851): 539–560.

Sohn H, Dzwonczyk M, Straser EG, Kiremidjian AS, Law KH, Meng T. An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge. *Earthquake Engineering & Structural Dynamics* 1999; 28: 879–897.

Sohn H, Farrar CR, Hemez FM, Shunk DD, Stinemates DW, Nadler BR, Czarnecki JJ. A review of structural health monitoring literature: 1996-2001. *Los Alamos National Laboratory Report* 2003.

Sohn H, Worden K, Farrar CR. Novelty detection under changing environmental conditions. *Smart Structures and Materials 2001: Smart Systems for Bridges, Structures,* *and Highways* 2001; 4330: 108–118.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* 2014, 15: 1929–1958.

Sun LM, Shang ZQ, Xia Y, Bhowmick S, Nagarajaiah S. Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection. *Journal of Structural Engineering* 2020, 146: 04020073.

Tang Z, Chen Z, Bao Y, Li H. Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. *Structural Control and Health Monitoring* 2019; 26(1): e2296.

Tipping ME. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 2001; 1: 211–244.

Tipping ME, Bishop CM. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1999; 61: 611–622.

Van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research* 2008; 9.

Vanik MW, Beck JL, Au SK. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics* 2000, 126(7): 738–745.

Wan HP, Ni YQ. Bayesian Modeling Approach for Forecast of Structural Stress Response Using Structural Health Monitoring Data. *Journal of Structural Engineering* 2018; 144.

Wan HP, Ren WX. Stochastic model updating utilizing Bayesian approach and

Gaussian process model. *Mechanical Systems and Signal Processing* 2016; 70: 245–268.

Wang M, Deng W. Deep visual domain adaptation: A survey. *Neurocomputing* 2018; 312: 135-153.

Wang RH, Li L, Li J. A Novel Parallel Auto-Encoder Framework for Multi-Scale Data in Civil Structural Health Monitoring. *Algorithms* 2018; 11(8): 112.

Wang X, Shahzad MM. A novel structural damage identification scheme based on deep learning framework. *Structures* 2021; 29: 1537–1549.

Weng S, Xia Y, Xu YL, Zhu HP. An iterative substructuring approach to the calculation of eigensolution and eigensensitivity. *Journal of Sound and Vibration* 2011, 330: 3368–3380.

Weng S, Zhu HP, Xia Y, Zhou XQ, Mao L. Substructuring approach to the calculation of higher-order eigensensitivity. *Computers and Structures* 2013; 117: 23–33.

Weng S, Zhu H, Xia Y, Li J, Tian W. A review on dynamic substructuring methods for model updating and damage detection of large-scale structures. *Advances in Structural Engineering* 2020, 23(3): 584–600.

Widjaja D, Varon C, Dorado AC, Suykens JAK, Van Huffel S. Application of kernel principal component analysis for single-lead-ECG-derived Respiration. *IEEE Transactions on Biomedical Engineering* 2012, 59: 1169–1176.

Wu RT, Jahanshahi MR. Deep Convolutional Neural Network for Structural Dynamic Response Estimation and System Identification. *Journal of Engineering Mechanics* 2019; 145(1): 04018125.

Xia Y, Hao H. Statistical damage identification of structures with frequency changes. *Journal of Sound and Vibration* 2003; 263(4): 853–870.

Xia Y, Hao H, Zanardo G, Deeks A. Long term vibration monitoring of an RC slab: Temperature and humidity effect. *Engineering Structures* 2006; 28: 441–452.

Xu S, Noh H Y. PhyMDAN: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning. *Mechanical Systems and Signal Processing* 2021, 151: 107374.

Yan AM, Kerschen G, De Boe P, Golinval JC. Structural damage diagnosis under varying environmental conditions-Part I: A linear analysis. *Mechanical Systems and Signal Processing* 2005a; 19: 847–864.

Yan AM, Kerschen G, De Boe P, Golinval JC. Structural damage diagnosis under varying environmental conditions-part II: local PCA for non-linear cases. *Mechanical Systems and Signal Processing* 2005b; 19: 865–880.

Yang JH, Lam HF. An innovative Bayesian system identification method using autoregressive model. *Mechanical Systems and Signal Processing* 2019, 133: 106289.

Yang YC, Nagarajaiah S. Data Compression of Structural Seismic Responses via Principled Independent Component Analysis. *Journal of Structural Engineering* 2014; 140(7): 04014032.

Ye XW, Jin T, Yun CB. A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems* 2019; 24(5): 567–586.

Yu Y, Wang C, Gu X, Li J. A novel deep learning-based method for damage identification of smart building structures. *Structural Health Monitoring* 2019; 18: 143–163.

Yuen KV, Beck JL, Au SK. Structural damage detection and assessment by adaptive Markov chain Monte Carlo simulation. *Structural Control and Health Monitoring* 2004; 11: 327–347.

Yuen KV, Beck JL, Katafygiotis LS. Efficient model updating and health monitoring methodology using incomplete modal data without mode matching. *Structural Control and Health Monitoring* 2006; 13: 91–107

Zang Z, Friswell MI, Imregun M. Structural damage detection using independent component analysis. *Structural Health Monitoring* 2004; 3: 69–83.

Zhang J, Wan CF, Sato T. Advanced Markov chain Monte Carlo approach for finite element calibration under uncertainty. *Computer-Aided Civil and Infrastructure Engineering* 2013; 28: 522–530.

Zhang QW, Fan LC, Yuan WC. Traffic-induced variability in dynamic properties of cable-stayed bridge. *Earthquake Engineering & Structural Dynamics* 2002; 31: 2015–2021.

Zhang R, Liu Y, Sun H. Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling. *Engineering Structures* 2020; 215: 110704.

Zhang Y, Miyamori Y, Mikami S, Saito T. Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering* 2019; 34: 822–839.

Zhao HW, Ding YL, Li AQ, Ren ZZ, Yang K. Live-load strain evaluation of the prestressed concrete box-girder bridge using deep learning and clustering. *Structural Health Monitoring* 2020; 19(4): 1051–1063.

Zhou C, Chase JG, Rodgers GW. Degradation evaluation of lateral story stiffness using HLA-based deep learning networks. *Advanced Engineering Informatics* 2019; 39: 259–268.

Zhou W, Chakraborty D, Kovvali N, Papandreou-Suppappola A, Cochran D, Chattopadhyay A. Damage classification for structural health monitoring using time-frequency feature extraction and continuous hidden Markov models. *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, IEEE* 2007; 848–852.

Zhou XQ, Xia Y, Weng S. *l*1 regularization approach to structural damage detection using frequency data. *Structural Health Monitoring* 2015; 14: 571–582.

Zou Y, Yu Z, Kumar B, Wang J. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. *Proceedings of the European Conference on Computer Vision* 2018; 289–305.