THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

## Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

**By reading and using the thesis, the reader understands and agrees to the following terms:**

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

**IMPORTANT**

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

# MULTI-FIDELITY MODELLING BASED ON ADAPTIVE SPARSE POLYNOMIAL CHAOS EXPANSION FOR BRIDGE DAMAGE IDENTIFICATION

**BEIYANG ZHANG**

**PhD**

**The Hong Kong Polytechnic University**

**2022**

The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering

# Multi-fidelity Modelling Based on Adaptive Sparse Polynomial Chaos Expansion for Bridge Damage Identification

**Beiyang ZHANG**

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

March 2022

# CARTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces 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)

Beiyang ZHANG (Name of student)

*Dedicated to my family*

*for their love and support*

# ABSTRACT

For bridge damage identification, two types of methods are popularly used, namely, physics-based methods and data-driven approaches. Physics-based methods identify the structural damage with the aid of a physical model. The damage location and severity can be identified, but such methods are prone to the modelling error. Data-driven approaches directly interpret real observations from a structure of concern by using some statistical methods. The modelling error can be avoided, but a large amount of measurement data is required, and only the existence of damage can be detected upon most occasions. Therefore, this thesis intends to investigate a Multi-Fidelity (MF) modelling technique that can capitalise on the merits from both physics-based and data-driven methods. Accurate surrogate models will be trained to help in damage identification, in which the modelling error caused by temperature can be correctly eliminated. Moreover, the required measurements are far less than those used in data-driven approaches.

In this thesis, Polynomial Chaos Expansion (PCE) method is employed to build the PCE surrogate model owing to its simple model structure and training process. However, challenges also exist in PCE, such as the "curse of dimensionality" issue and adaptive modelling problem. Therefore, we first introduce two novel adaptive modelling techniques to facilitate the PCE method in application to engineering problems. In order to adaptively collect samples for PCE training, a hybrid sequential sampling strategy is developed, which leverages both the input information of

PCE model and the output information from observations to instruct the sampling process. Meanwhile, the sparse regression procedure is used along with this strategy to train a sparse PCE model. As a result, the samples could be collected with high quality and in relatively small quantity. By evaluating on several benchmark functions, it is shown that the proposed strategy outperforms most existing methods. Next, a novel adaptive basis selection strategy is developed to adaptively determine the model structure, which consists of three procedures, basis expansion, pruning and refinement. By using this strategy, the proper truncation degree for PCE modelling can be selected automatically, and the training cost will also be reduced benefitting from removing the insignificant polynomial bases. To reconcile the sequential sampling and the adaptive basis selection in a consistent framework, a stability evaluation process which works in parallel with the sequential sampling process is introduced. As a result, this consistent PCE modelling framework can collect appropriate samples and determine the best model structure all in an automatic way. Through evaluating on several benchmark functions, this PCE modelling framework is demonstrated with satisfactory performance and high efficiency. By using this framework, the PCE models as surrogate to the physical model of the bridge structure are established, in which the pattern of frequency data concerning the structural parameters is of interest in this study.

In real applications, however, it is more practical to predict the responses of real structure rather than the physical model to help in damage identification, since the physical model will inevitably contain modelling errors. As the most influential environmental factor, temperature affects bridge

structures in a complicated way, and such effect is generally difficult to be simulated correctly. Thus, to eliminate the temperature-induced modelling error in the surrogate model, a Transfer Learning (TL) based Multi-Fidelity PCE (MFPCE) modelling technique is presented. The PCE model stemming from a finite element model is regarded as Low-Fidelity (LF) model, and the frequencies collected from the real bridge under healthy condition belong to High-Fidelity (HF) data. By updating the temperature-related polynomial terms in the LF model with HF data, MFPCE model can be formulated, where the temperature effect is considered more accurately.

Based on the formulated MFPCE model, the sparse damage identification is performed and discussed in the last part of the research. Since the PCE model generally has a strong nonlinear property, traditional sparse representation approaches that work for linear systems are unsuitable for this case. Therefore, an approximate $l_0$ sparse damage identification approach is developed by combining a heuristic algorithm, i.e., Cuckoo Search Algorithm (CSA), with the discrepancy principle. We first give assumptions to the number of damages as prior information for the optimisation equation. Then, the optimal solution involving damage locations and severities under each assumption is obtained by using CSA. Through comparison among the optimisation residuals under different assumptions, the discrepancy principle is used to find the correct number of damages, and the damage locations and severities can thus be obtained. Finally, a numerical bridge model and an experimental beam model are explored for verification. Results demonstrate the effectiveness of the proposed MF modelling technique and the damage identification approach.

# LIST OF PUBLICATIONS

**Journal Papers:**

**Zhang, B.Y.**, and Ni, Y.Q. (2021). A hybrid sequential sampling strategy for sparse polynomial chaos expansion based on compressive sampling and Bayesian experimental design. Computer Methods in Applied Mechanics and Engineering, 386(2021), 114130.

**Zhang, B.Y.**, and Ni, Y.Q. (2022). Sparse polynomial chaos expansion with sequential sampling and adaptive basis selection by using Bayesian compressive sensing. Applied Mathematical Modelling, under review.

**Zhang, B.Y.**, and Ni, Y.Q. (2022). A data-driven sensor arrangement strategy for mode shape reconstruction by using recurrent Gaussian process regression. Engineering Structures, under review.

**Conference Papers:**

**Zhang, B.Y.**, and Ni, Y.Q. (2019). An inversed greedy method for information-based optimal sensor placement on bridges. Proceedings of the 12$^{th}$ International Workshop on Structural Health Monitoring, 10-12 September 2019, Stanford, California, USA.

**Zhang, B.Y.**, and Ni, Y.Q. (2021). A hybrid sequential sampling strategy for sparse polynomial

chaos expansion based on compressive sampling and Bayesian experimental design. Proceedings of the 2nd ZHITU Symposium on Advances in Civil Engineering, 28-29 September 2021, Ulsan, South Korea.

# ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Prof. Yi-Qing Ni, for offering me the precious opportunity to pursue the Ph.D. study and further valuable support during the past years in the Hong Kong Polytechnique University. He is knowledgeable, rigorous and diligent, with unique insights and enthusiasm into scientific research. I learnt many from him not only in the research works but also in his conscientious attitude. It is my great honour to study under his supervision.

I would like to give my thanks to the financial support from the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of National Engineering Research Centre on Rail Transit Electrification and Automation (Grant No. K-BBY1) and the grant from the Research Grants Council of the Hong Kong Special Administrative Region (SAR), China (Project No. PolyU 152024/17E). I also want to acknowledge the colleagues, Mr. Wing-Hong Kwan, Mr. Tai-Tung Wai, Mr. Kwok Lung Cheung, Mr. H.Y. Leung and all technicians in Laboratory Y001 for their expertise and kind support on the experimental tests in my study. I am also grateful to other staffs in the centre and general office of the Department of Civil and Environmental Engineering.

I am grateful to all of my colleagues and friends, Dr. You-Wu Wang, Dr. Su-Mei Wang, Dr. Lu Zhou, Dr. Ran Chen, Dr. Chi Xu, Dr. Si-Qi Ding, Dr. Qiu-Hu Zhang, Dr. Si-Xin Chen, Dr. Lin-Fang Li, Mr. Yun-Ke Luo, Mr. Yang Lu, Mr. Yuan-Hao Wei, Mr. Shuo Hao and all colleagues in the centre

for their kind suggestions and help at the Hong Kong Polytechnique University. I also want to thank Dr. Yan-Jie Zhu for her understanding and accompanying during my study period.

Last but not least, I would like to dedicate my special gratitude to my beloved parents, for their everlasting love, care and encouragement.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ANN | Artificial Neural Network |
| ANOVA | ANalysis Of VAriance |
| aPCE | arbitrary PCE |
| ASPCE-BASS | Adaptive SPCE method with Basis Adaptivity and Sequential Sampling |
| BCS | Bayesian Compressive Sensing |
| BED | Bayesian Experimental Design |
| BPDN | Basis Pursuit De-Noising |
| Coh-entro | coherence-entropy |
| CS | Compressive Sensing |
| CSA | Cuckoo Search Algorithm |
| CV | Cross-Validation |
| ED | Experimental Design |
| EI-ELF | Expected Improvement-based Expected Loss Function |
| ELF | Expected Loss Function |
| FE | Finite Element |
| gPCE | generalized PCE |

| | |
|---|---|
| GPR | Gaussian Process Regression |
| HF | High-Fidelity |
| KLD | Kullback-Leibler Divergence |
| LF | Low-Fidelity |
| LFPCE | Low-Fidelity PCE |
| LHS | Latin Hypercube Sampling |
| LOO | Leave-One-Out |
| MC | Monte Carlo |
| MF | Multi-Fidelity |
| MFPCE | Multi-Fidelity PCE |
| MSE | Mean Square Error |
| NDT | Non-Destructive Testing |
| NLHS | Near-LHS |
| NLHS-DP | NLHS with “Doubling Procedure” |
| NP-hard | Non-deterministic Polynomial-time hard |
| OLS | Ordinary Least-Squares |
| OMP | Orthogonal Matching Pursuit |
| PCA | Principal Components Analysis |
| PCE | Polynomial Chaos Expansion |

| | |
|---|---|
| PDF | Probability Density Function |
| PLHS | Progressive-LHS |
| RRMSE | Relative Root Mean Square Error |
| RS | Response Surface |
| RVM | Relevance Vector Machine |
| SBL | Sparse Bayesian Learning |
| SHM | Structural Health Monitoring |
| SNR | signal-to-noise ratio |
| SPCE | Sparse PCE |
| std | standard deviation |
| TL | Transfer Learning |
| TL-MFPCE | TL based MFPCE |
| tMFPCE | traditional MFPCE |
| UQ | Uncertainty Quantification |

# CHAPTER 1
# INTRODUCTION

## 1.1 Research Background and Motivations

Bridge is an important civil infrastructure to cross river, valley and even sea, which is commonly designed with long lifespan and high durability. Due to the operational and environmental effects, bridge will inevitably suffer from structural damages, and these damages may have a fatal impact to its performance during the life cycle period. Without timely maintenance, these damages will accumulate and ultimately lead to bridge collapse. Thus, it is expected that damages could be detected in their early stage to avoid the structural failure (Fan and Qiao, 2011). Currently, there are two primary techniques to assess the bridge condition, namely, the in-situ Non-Destructive Testing (NDT) and Structural Health Monitoring (SHM) (Jayasundara et al., 2020; McKeon et al., 2014; Pérez and Serra-López, 2019). NDT is normally a periodical routine detection to identify the bridge condition. It is generally laborious, and the detection is not real-time. By contrast, SHM proposes to garner real-time measurements from a bridge through a number of permanently installed sensors. By processing measurements with some techniques, the anomalies on the bridge structure can be identified in time. Hence, SHM can work in a more efficient and

economical way. However, challenges still remain for the SHM application. For example, the number of deployed sensors might be limited due to the project budget, and exploring techniques for interpreting and storing massive data could be another challenge. Therefore, many studies have focused on developing the damage identification methods that can be applied based on spatially sparse measurements.

Regarding the damage identification methods for bridges, the vibration-based approaches are a kind of mature technique that can recognise damage in a global way (Hou and Xia, 2021). They evaluate the dynamic features from the measured vibration signals and then assess the bridge condition by comparing the changes in the dynamic characteristics. The vibration measurements can be acceleration, velocity, displacement, or dynamic strain. With the development of sensing technology, measuring vibration signals has become more and more convenient and cheap, and the measuring precision has also been highly improved. Thus, such approaches have been widely implemented in many fields, involving mechanical, aerospace, and civil engineering, etc. (Abdeljaber et al., 2017; Lam et al., 1998; Liu et al., 2014).

In general, there are two kinds of methods that were popularly used in the existing vibration-based damage identification techniques, physics-based methods and data-driven approaches (Hou et al., 2018b, 2020; Huang and Beck, 2015; Jayasundara et al., 2020; Zhu et al., 2018, 2019). Physics-based methods identify damages by comparing the outputs from a physical model (e.g., finite element model) with the observations from real structure (Huang and Beck, 2015). Here, the

physical model can be substituted by a surrogate model for the sake of cost saving. The advantage is that both the location and severity of damage can be identified. But such methods are hard to consider the measurement noise, and they are prone to the modelling error, such as incorrect modelling of the environmental effects. Data-driven approaches employ statistical methods to directly process observations from the real structure in healthy or damaged state (Zhu et al., 2019). The modelling error can be avoided, and the measurement noise can be considered in the modelling process. However, such approaches were usually used to detect the existence of damage. The damage location and severity are difficult to be assessed since collecting comprehensive labelled data from real structure in damaged state is impractical. To summarise, physics-based methods have more challenges in building accurate physical models, and data-driven approaches have more limitations in actual applications. To overcome the drawbacks and capitalise on the merits from both methods, the combinational approaches were developed, such as model updating, digital twin and Multi-Fidelity (MF) modelling (Bigoni and Hesthaven, 2020; Das and Debnath, 2018; Diez-Olivan et al., 2019; Gregory et al., 2019). The principle behind the combinational approaches is to update a physical model or its surrogate model by using a small number of real observations. The modelling error is expected to be eliminated, and the measurement noise can be taken into account. Subsequently, damage identification can be performed based on the updated model to achieve results with higher precision than that based on the original model. At present, exploiting the combinational approaches has become a research hotspot, which is calling for more in-depth

investigations.

Although numerous studies were conducted in developing the damage identification approaches, there are still several challenges that should be paid attention to:

(1) The number of sensors in an SHM system is limited, that is, the available measurement points are finite. The quantity of potential damage locations is typically greater than that of available measurement points, which results in an underdetermined issue (Hou et al., 2018b). In other words, the problem has non-unique solutions. Hence, finding the unique optimal solution is a big challenge. Recently, the sparse representation techniques were utilised to cope with this issue and have received promising results (Ding et al., 2019; Hou et al., 2018b; Huang and Beck, 2013).

(2) The vibration characteristics of a bridge are not only influenced by damage, but also affected by the environmental effects, such as temperature, humidity, and wind, etc., in which the temperature is generally considered the most influential factor (Sun et al., 2018). Response variations caused by the temperature effect in a bridge structure can often exceed those caused by damage (Bao et al., 2012; Erazo et al., 2019; Huang et al., 2018). If the temperature is not well-considered, the damage is not able to be correctly recognised. Nonetheless, it was found that the impact mechanism of temperature on bridge is extremely sophisticated (Han et al., 2021). Identifying damage with consideration of the temperature effect still needs further studies.

This thesis intends to investigate the surrogate modelling technique and damage identification

method for bridge structures considering temperature effect. To achieve the targets, Polynomial Chaos Expansion (PCE) is recommended to build the surrogate model. PCE is an outstanding polynomial regression method, which was broadly implemented in the field of uncertainty quantification, sensitivity analysis and reliability evaluation (Wan et al., 2020; Ng and Eldred, 2012; Zhou et al., 2020), owing to its simple modelling structure and training process. Nonetheless, it suffers from the "curse of dimensionality" issue when applied to problems with high input dimension or high polynomial degree (Lüthen et al., 2020). The PCE model will have an extremely large cardinality in modelling such problem so that massive samples and observations are required for training, and this will result in unacceptable computational cost. Besides, several other challenges in PCE should also draw our attention, such as choosing appropriate sample set, determining a proper sample quantity, and selecting the best truncation degree, etc. (Blatman and Sudret, 2010, 2011; Jakeman et al., 2015). Therefore, we first try to tackle these challenges by developing two adaptive modelling strategies. A novel sequential sampling strategy is developed to adaptively collect training samples. The input information from the PCE model and the output information from the target responses are both leveraged to instruct the sampling process, so the samples can be collected with high quality and in relatively small quantity, and the best sample quantity can be automatically determined. Then, an adaptive basis selection strategy is proposed. Three procedures involving basis expansion, pruning and refinement contribute to selecting the significant basis terms and removing the insignificant basis terms in the PCE model. Not only the

proper truncation degree can be chosen, but also the best model structure can be selected. Furthermore, a framework is introduced to reconcile the sequential sampling strategy and adaptive basis selection strategy into a consistent algorithm, in which a sparse representation method is employed for regression calculation. Benefitting from these adaptive modelling strategies, the PCE model can be adaptively built for a target task. The sparse representation and adaptive basis selection both contribute to reducing the training cost, which enable the implementation of PCE technique for complicated problems.

For the target of damage identification, the surrogate model was generally built depending on a physical model (Gregory et al., 2019). However, the physical model will inevitably differ from the real structure due to the existence of modelling error. As the most influential environmental factor, temperature will affect the bridge structure in a complicated way. Rationally considering the temperature effect in the physical model or its surrogate model is a challenge. Here, we present to use the MF modelling technique to eliminate the temperature-induced modelling error in the PCE surrogate model. Regarding the PCE model as Low-Fidelity (LF) model and the observations from real bridge as High-Fidelity (HF) data, the MF modelling technique proposes to update the LF model by using the HF data. To overcome the difficulty that the observations (HF data) from bridge under damaged condition are commonly deficient in practice, the concept of Transfer Learning (TL) is introduced to the traditional Multi-Fidelity PCE (MFPCE) modelling process. The knowledge of damage that leant from the physical model is retained in the PCE model, and the temperature-related

polynomial bases are updated with the measurements from bridge in healthy condition. Ultimately, the MFPCE model can correctly consider the temperature effect and is further used to help in damage identification.

In the final part of this thesis, an approximate $l_0$ sparse damage identification method is developed based on the PCE model. The principle behind those damage identification approaches which employed data models or surrogate models is to find a set of damage parameter values that can minimise the residual between the model evaluations and real responses. The traditional sparse recovery algorithms will fail in this study since the system (PCE model) is nonlinear. Hence, a heuristic algorithm, Cuckoo Search Algorithm (CSA), is employed to solve the optimisation problem with nonlinear system, and a discrepancy principle is introduced to determine the solution sparsity. Finally, the damage can be recognised with its location and severity. By validating on a simulation bridge model and an experimental beam model, the TL based MFPCE modelling technique will be compared with the traditional MFPCE technique to demonstrate its effectiveness. Also, the performance of the approximate $l_0$ sparse damage identification method will be verified on the two cases.

## 1.2 Research Objectives

This study intends to develop a completely adaptive PCE modelling framework and a novel MF technique for building accurate PCE model as surrogate to the real bridge structure. By introducing a novel sparse damage identification method, the damage occurred on the bridge can be

identified correctly based on the surrogate model. The detail objectives in this study are:

1. To propose a hybrid sequential sampling strategy for sparse PCE modelling. The existing sequential sampling strategies mostly rely on information from single source, such as the information from input distribution, polynomial bases or observations already be collected. To consider information from different sources, a hybrid sequential sampling strategy is explored, in which sparse representation is recommended to further reduce training cost.

2. To develop an adaptive basis selection strategy and combine it with the sequential sampling and sparse representation. Adaptive basis selection is an important strategy to help select the significant basis terms during the training process and determine the best PCE truncation degree automatically. Meanwhile, performing the adaptive basis selection and sequential sampling concurrently also has practical necessity. In this study, a novel adaptive basis selection strategy with high efficiency is pursued, and an adaptive PCE modelling framework is developed to effectively combine the adaptive basis selection and sequential sampling strategies.

3. To capitalise on the concept of TL in the MFPCE framework to eliminate the temperature-induced modelling error. Utilising real observations to update the temperature-induced modelling error in the surrogate model can be realised by MF modelling, whereas the observations from structure under damaged condition are generally unavailable. The concept of TL enables the updating process to reduce its dependence on the real observations. Accurate surrogate model to the

real bridge is expected to be trained without observations from damaged structure.

4. To develop a sparse damage identification approach based on the PCE surrogate model. Sparse representation is an efficient tool in addressing damage identification problems, whereas the traditional sparse representation approaches are generally valid only for linear system. In this study, PCE model is a nonlinear system. A sparse damage identification approach for nonlinear system is required so that the damage can be identified with desired accuracy.

## 1.3 Thesis Outline

This thesis consists of seven chapters, which are briefly introduced as follows:

Chapter 1 gives an overall introduction about this thesis, which involves the background and motivation, the introduction to the vibration-based damage identification methods and challenges, and the primary research objectives.

Chapter 2 reviews the key contents in this study. The PCE technique and its pertinent developments are first introduced, including the brief history, sparse representation and adaptive modelling strategies. Then, the MF technique for PCE modelling is reviewed. After that, two kinds of damage identification methods, physics-based methods and data-driven approaches, are discussed. Finally, the temperature effects on bridge structures are reviewed, and their difficulties in numerical modelling are discussed.

Chapter 3 proposes a hybrid sequential sampling approach for PCE model to adaptively collect

samples with high convergence rate, in which the Bayesian compressive sensing is employed as a sparse regression procedure to train a sparse PCE model. The idea of this sampling approach is to employ both the input information of PCE model and the output information from observations to instruct the sampling process in a sequential way. As a result, samples could be collected with high quality and in relatively small quantity, and the best sample quantity can be automatically determined. As an example, the coherence-optimal sampling method is combined with Bayesian experimental design to build a coherence-entropy method. This method is evaluated on several benchmark functions through comparison with three input-dependent only methods and two output-dependent only methods.

Chapter 4 presents a novel adaptive basis selection strategy, which is then combined with the sequential sampling method to build a fully adaptive PCE modelling framework. The adaptive basis selection strategy has three procedures, basis expansion, pruning and refinement, which can adaptively select the significant polynomial bases of proper degree during the modelling process. To reconcile the sequential sampling and adaptive basis selection into a consistent modelling framework, a stability evaluation process is introduced. As a result, this adaptive PCE modelling framework is able to automatically determine the appropriate truncation degree and training sample set simultaneously. By evaluating on several benchmark functions, the fully adaptive PCE modelling framework is demonstrated to have good performance and high efficiency. Beyond that, this framework is also extended for modelling the multi-output problem.

Chapter 5 investigates a TL based MFPCE modelling technique to eliminate the temperature-induced modelling error in the surrogate model. Firstly, the adaptive PCE modelling framework is used to build LFPCE models depending on the data from finite element model. Then, the HF observations collected from the real structure in healthy condition under different temperatures are collected to update the temperature-related terms in the LFPCE models. The knowledge of damage is transferred from the finite element model to the MF models when the real observations from damaged structure are deficient. Finally, MFPCE models are built, which can predict the responses of real structure with high accuracy. This method is assessed on a simulation bridge and an experimental beam by comparing with the traditional MFPCE technique.

Chapter 6 develops an approximate $l_0$ sparse damage identification method based on the formulated MFPCE models. Discrepancy principle is employed to find the best solution sparsity, and CSA is employed to garner the damage results involving locations and severities. Performance of the proposed damage identification method is evaluated on a simulation bridge model and an experimental beam model.

Chapter 7 gives the conclusions, major findings and potential future works.

# CHAPTER 2
# LITERATURE REVIEW

As introduced in Chapter 1, the purpose of this study is to develop accurate surrogate model for real structure based on the Polynomial Chaos Expansion (PCE) and Multi-Fidelity (MF) modelling techniques. In the surrogate model, the temperature-induced modelling error can be eliminated, and the model can be further used for damage identification. Therefore, the first part of this chapter is to review the PCE technique with some state-of-art modelling techniques such as sparse representation and adaptive modelling. The advantages and disadvantages of PCE compared to other surrogate modelling techniques are also reviewed. After that, the MF modelling techniques that applied on engineering cases are reviewed, following by the review of the historically used physics-based and data-driven damage identification methods. Finally, the overview of temperature effects on bridge is given, so the research gap concerned in this study can be clearly elaborated.

## 2.1 Polynomial Chaos Expansion

### 2.1.1 Introduction

PCE is a kind of method to represent a target variable through a polynomial function of input

variables, and these variables are all random variables (Xiu and Karniadakis, 2002). A significant property of these polynomials is that they are orthogonal to each other with respect to the joint probability density function of the inputs. By the use of PCE model, the uncertainty propagated from the input variables to the target variable can be easily studied. Subsequently, the trained PCE model can be used for further analysis, such as Uncertainty Quantification (UQ) or sensitivity analysis (Ni et al., 2019; Sun et al., 2020).

PCE originated from the so-called 'homogeneous chaos' that was presented by Wiener in 1938, which used Hermite polynomials to model the stochastic process with Gaussian random variables (Wiener, 1938). In applications, however, Wiener's formulation was demonstrated difficult to include high-order terms so that its modelling accuracy cannot be guaranteed. Based on the Wiener-Hermite expansion, a more explicit formulation was developed with high convenience in utility, named Wiener chaos expansion (Cameron and Martin, 1947). The core idea was to discretise the white noise process through its Fourier expansion, so this approach had a wide application in stochastic analysis involving white noise. The above models were proposed based on input variable with Gaussian distribution, but various variables with different distributions exist in reality. Hence, the generalised PCE (gPCE) was developed to summarise those polynomials that are orthogonal with respect to various types of distributions (Xiu and Karniadakis, 2002). In 2012, rigorous proofs of the existence and convergence of gPCE were provided (Ernst et al., 2012). Afterwards, the PCE technique became more and more mature, and it could be applied to many research areas

(Sepahvand et al., 2010; Sudret, 2008; Xu and Kong, 2018). Recently, the PCE technique was further developed to describe arbitrary variables with implicit distributions, which is called arbitrary PCE (aPCE) (Wan and Karniadakis, 2006). For those variables that do not have explicit or known distribution functions, aPCE can be employed to formulate the corresponding orthogonal polynomials through a data-driven process, and reliable PCE model can be built accordingly.

Mathematically, the PCE model is a linear combination of different polynomials with unknown coefficients. These coefficients can be calculated through a training process, and the trained PCE model can then represent the target model. Generally, the training methods can be divided into two categories, intrusive and non-intrusive (Sudret, 2008). The intrusive method, or called Galerkin projection, is a kind of numerical analysis method (Crestaux et al., 2009). It requires the solving path adapted to the numerical code of the target problem, thereby calling intrusive. In the earlier studies, this method has been applied in several research areas, such as the seismic soil-structure interaction, stochastic constitutive relation, etc. (Ghanem and Ghiocel, 1998a, 1998b; Ghiocel and Ghanem, 2002). However, it would be challengeable for this approach to be applied on complicated problems since the modification of numerical code was arduous. Alternatively, non-intrusive methods emerged, which had a wider applicability than the intrusive method.

In principle, the non-intrusive approaches do not require the numerical code or mathematical expression of the target problem. Instead, the evaluations of the target model under different input values are needed (Ng and Eldred, 2012). In other words, the solving path of the target problem is

inessential, and only the input-output data set is required. There are two kinds of methods in the non-intrusive approach, projection method and regression method (Sudret, 2008). Projection method leverages the orthogonal property of PCE bases. By multiplying a basis term to both sides of the PCE expression and taking integrals, the unknown coefficient that corresponds to the multiplied basis term will be retained while others will be eliminated (Blatman, 2009). Hence, the unknown coefficient can be received by calculating the integral. For most problems, however, the integral is analytical unsolvable, and thus the numerical integration techniques (e.g., Monte Carlo (MC), quadrature) were employed to receive the approximate solution (Debusschere et al., 2004). The regression method aims at computing the coefficients by minimising the mean square error of PCE approximations at sampled points (Blatman, 2009). These samples were named Experimental Design (ED) (Blatman and Sudret, 2010; Fajraoui et al., 2017). By collecting a set of ED and corresponding model evaluations, the coefficients can be calculated through regression estimation. It was indicated that the samples and corresponding model evaluations required by the regression method are far less than those needed by the projection method (Blatman, 2009). Thus, the training cost can be largely reduced in case the target model has a high complexity. Accordingly, a challenge of this method lies in the issue of selecting proper sample points so that the regression equations are well-conditioned. Developing the strategies to select samples with high quality has become a research hotspot, which will be elaborated in Section 2.1.4. To summarise, the regression method has a simple calculation process and a low training cost, so it is by far the most popular method in

training the PCE model.

In applications, the PCE technique was generally used for solving UQ problem. Simulation models generated by computer to describe physical phenomena or engineering systems have been made available for decades with the development of computer technology. More and more complex simulation models are built to attain analysis results with high accuracy, and these models can be used as reliable references to help solve practical problems. However, the simulation modelling result will inevitably have discrepancy to the observations from real system. The uncertainty existent in the input parameters of a simulation model, which may induce estimation difficulty or inaccurate estimation of these parameters, is one of the causes, and this may produce erroneous judgement to the problems that we focus on (Blatman, 2009). The PCE technique can help explore the uncertainty propagation from the input parameters to the model output with lower cost than the traditional MC simulation, perturbation method, and first-order/second-order reliability methods (Kareem, 1988; Leng and He, 2006; Liu et al., 1986; Sepahvand et al., 2010; Vishwanathan and Vio, 2019; Zhang and Du, 2010). Ng and Eldred (2012) applied the PCE technique to study the uncertainty propagation in some partially differential functions (e.g., Helmholtz equation in acoustic wave propagation). Wan et al. (2020) employed the aPCE technique to recognise the relation between design parameters and modal frequencies of bridge structures, and then the uncertainty propagated from the design parameters to the modal frequencies were studied. With the aid of UQ, further analysis on the input parameters or target system could be conducted, such as parameter

sensitivity analysis and system reliability analysis (Cheng and Lu, 2020; Ni et al., 2019; Sun et al., 2020). Crestaux et al. (2009) employed the so-called Sobol' indices to study the parameter sensitivity of some benchmark functions. Wan et al. (2020) investigated the sensitivity of each design parameter to the modal frequencies of bridge structures, which can provide reliable reference for structural design and monitoring. Marelli and Sudret (2018) compared the PCE technique with the MC simulation method and first-order/second-order reliability methods in investigating the reliability of a frame strucutre, in which the displacement was the target output. Zhou et al. (2020) investigated the reliability problem in vehicle crashworthiness, and the analysis results were further utilised to instruct the design optimisation. In addition to the research relying on the UQ outcomes, the PCE technique was also employed to build surrogate model in engineering problems. Here, the PCE model was commonly simplified as the response surface model, in which the polynomials are not orthogonal to each other. For instance, Stutz et al. (2018) adopted the response surface model to build a surrogate model of the flexibility matrix of a beam structure. Damages of the beam element could be identified by comparing the model outputs with the responses from structure under damaged state. Umar et al. (2018) utilised the second-order response surface models to describe the relations between the modal parameters (frequencies and mode shapes) and Young's modulus of elements in a beam structure. Then, the model updating strategy was employed to identify damages. In these applications, it is worth noticing that the truncation degree of the model was very small because the training cost of high degree model was unacceptable.

As can be seen, the PCE model has a simple expression form, and its training process is easy to implement. However, the PCE model suffers from the "curse of dimensionality" issue that the quantity of unknown coefficients will grow exponentially with the increase of input dimensionality or truncation degree. The training cost, even with the regression method, will be unacceptable, which impedes the development of the PCE technique for surrogate modelling. In addition, the model selection is another challenge in PCE. The truncation degree for a PCE model should be selected before training but choosing the most appropriate degree value for an unknown problem is unpractical when no prior information is provided. Therefore, the PCE technique is currently popular in UQ and sensitivity analysis due to the low requirement on the truncation degree. For addressing problems with high complexity, more techniques should be introduced in PCE to ease the modelling process, which will be presented in the following sections.

### 2.1.2 Compressive sensing and sparse representation

As is introduced in the previous section, the regression method was a popular approach in PCE training due to its low requirement on the sample quantity. Even though, it still needs $kP$ samples for getting robust solution, where $P$ denotes the number of unknown coefficients and $k$ is a constant which was recommended between 2 and 3 (Hosder et al., 2007). Since $P$ will grow exponentially with the increase of input dimensionality or truncation degree, the quantity of samples and corresponding model evaluations required by the regression method will also grow dramatically.

To facilitate the surrogate modelling via PCE technique, it is likely to use as less samples (model running times) to train a PCE model with as high precision as possible. In the past, one of the research areas in PCE focused on the development of state-of-art training strategies to reduce the dependency on the size of training dataset.

The concept of sparse or compressibility was first introduced to PCE by Blatman and Sudret (Blatman, 2009; Blatman and Sudret, 2008, 2010). The so-called “sparsity of effects” principle stated that most phenomenon-describing models are dominated by the main effects and interactions of low order (Montgomery, 2006), so the real-world problems are generally considered compressible on polynomial chaos (Lüthen et al., 2021). This means that the PCE model can be expressed in a sparse way, in which most coefficients are zero. Based on the “sparsity of effects”, Blatman and Sudret (2011) created a truncation scheme called hyperbolic truncation. This scheme proposed to truncate the basis terms with high interaction effects under a given truncation degree, so the number of unknown coefficients could be largely reduced. Then, they developed an adaptive basis selection strategy based on the least angle regression method. By only retaining the basis terms that are significant to modelling the target model in the training process, the number of coefficients that needed to be solved were further reduced (Blatman and Sudret, 2011). Inspired by this concept, the sparse recovery technique has received considerable attention in PCE training.

In 2006, the concept of Compressive Sensing (CS) was proposed. It is a signal processing technique for efficiently acquiring and reconstructing a signal. By assuming that the original signal

can be sparsely represented on a domain, CS theory demonstrated that the samples required to recover the original signal could be smaller than the samples collected based on the Nyquist/Shannon sampling theory (Candes et al., 2006; Donoho, 2006). As a result, only a small number of observations about the signal are necessary to be collected, and the original signal can be recovered with desired accuracy by using some optimisation tools. These tools were known as sparse representation methods. In CS, several sparse representation methods were proposed for recovering the original signal from sparse samples, such as Orthogonal Matching Pursuit (OMP), Basis Pursuit De-Noising (BPDN), iterative reweighted methods and Sparse Bayesian Learning (SBL), etc. (Chen et al., 1998; Mallat and Zhang, 1993; Needell, 2009; Tipping, 2001). In the last decade, the sparse representation methods have been broadly leveraged for training the Sparse PCE (SPCE) model. Jakeman et al. (2015) employed the OMP method to train the PCE model. In the modelling process, OMP was combined with an adaptive basis selection strategy, thereby further reducing the training cost. Based on the standard BPDN algorithm, a new weighted $l_1$ minimisation algorithm was proposed for PCE modelling by Peng et al. (2014). Comparing with the existing sparse representation methods, this weighted $l_1$ minimisation algorithm was demonstrated to receive PCE model with better accuracy. Furthermore, the SBL method was also employed to train SPCE model (Zhou et al., 2019b). Under the Bayesian framework, a novel sampling strategy was proposed to enhance the sample quality, and the SPCE model was trained with robust performance. To summarise, it is often cheap to train an SPCE model with the aid of

sparse representation, which enables the PCE technique to deal with problems with high complexity.

### 2.1.3 Adaptive modelling

When solving the unknown PCE coefficients by using either regression approaches or sparse representation methods, it is no doubt that the sample quality will affect the PCE modelling precision, especially when the sample quantity is restricted. As a result, it is crucial to choose a set of well-designed samples for obtaining a reliable PCE model. The sample set was named as the Experimental Design (ED) (Blatman and Sudret, 2010; Fajraoui et al., 2017; Ni et al., 2017). In the context of PCE, many sampling strategies have been proposed, such as MC (Migliorati et al., 2014), Latin Hypercube Sampling (LHS) (Blatman and Sudret, 2010; Jakeman et al., 2015), D-optimal design (Burnaev et al., 2017; Diaz et al., 2018), and others (Alemazkoor and Meidani, 2018; Shin and Xiu, 2016; Thapa et al., 2020). Typically, samples will have good quality (e.g., the corresponding observations could discover more information about the target model) when they spread over the input domain in a space filling way (Crombecq et al., 2011; Sheikholeslami and Razavi, 2017). MC is a traditional way to generate random samples from a given distribution. However, when only a few samples are collected, the sample distribution may have a large discrepancy from the target distribution, that is, the samples will not be able to evenly spread across the probability space. LHS can overcome this drawback (Stein, 1987). Even when the sample number is small, the LHS method can collect samples from the target distribution in a space filling

way. Choi et al. (2004) leveraged the LHS method to collect samples for training the PCE model of the buckling eigenvalue of a joined-wing aircraft, which was then used for UQ. With the development of regression methods, some sampling strategies were presented for the target of receiving robust regression solutions. The D-optimal sampling strategy was proposed to make the samples contribute more to the training of the PCE model, which was originally designed for the Ordinary Least-Squares (OLS) method (Zein et al., 2013). To attain the smallest estimation uncertainty of coefficients in using the OLS method, samples are collected at those places which minimise the determinant of the information matrix. Thus, the training of PCE model with these samples will circumvent the ill-conditioned regression matrix. Moreover, there are some other similar criteria, such as S-optimal (Shin and Xiu, 2016), A-optimal (Thapa et al., 2018b) and K-optimal (Loukrezis et al., 2020), etc., which were all proposed to optimise the sample distribution according to the information matrix in regression calculation. Apart from the strategies designed for regression-based PCE, it is also important to design appropriate sampling methods for PCE modelling with sparse representation. Diaz et al. (2018) extended the D-optimal design to collect samples for solving $l_1$ minimisation problem by employing QR factorisation. The quantity limitation of samples in the least-square regression was removed. Based on the CS theory, Hampton and Doostan (2015) proposed a coherence-optimal sampling strategy, in which a lower bound of sample quantity which can cope with the $l_1$ minimisation problem with desired accuracy in a high likelihood was deduced. In their work, the input distribution was modified by multiplying a weight

function to form a coherence-optimal distribution; thus, the regression matrix would have the lowest coherence when sampling from this distribution, and the samples needed for best recovery of sparse representation would be the minimum.

Although the methods mentioned above were originated from different theories, they were designed to collect all samples at once. However, the best size (less but enough) of ED generally remains unknown when the target model information is not known a priori. It is hard to determine the best size of ED in advance. Sequential sampling strategy, or called active learning, has been proposed to cope with this problem in the field of PCE through refinement of the existed non-sequential sampling methods. Subsequently, the concept of adaptive modelling for PCE emerged, and the sequential sampling has become one of the research hotspots. In principle, the sequential sampling strategy begins with a small ED size, and then gradually adds new samples and the corresponding model evaluations to the current selection set until a predefined stop criterion is satisfied (e.g., the precision of the trained PCE model). This kind of strategy is flexible to help determine the ED size because the previously selected samples and corresponding observations will be kept in the subsequent iterations, and the final sample size could be determined automatically. In the previous studies, LHS was coupled with sequential sampling to form a nested LHS method for the purpose of space filling (Blatman and Sudret, 2010). D-optimal design was refined with sequential sampling to achieve a robust evaluation of PCE model (Diaz et al., 2018).

Besides the sampling problem, choosing an appropriate PCE model structure, which is known

as model selection problem, is also critical in getting precise modelling results (Tan, 2015). Basically, the model selection problem in PCE is to determine a proper truncation degree (Blatman and Sudret, 2011), or in a more complicated way, to choose the most significant basis terms (Ni et al., 2017). A PCE model is generally truncated with a given polynomial degree before training so that only a finite number of coefficients are needed to be solved. However, a small truncation degree cannot ensure promising accuracy of the trained PCE model, and an extremely large truncation degree will induce a heavy training cost in collecting massive samples and observations. When the sample quantity is restricted due to limitations in practice, the model cannot be trained with reliable solution. It is hard to determine an appropriate truncation degree in advance when little information about the target problem is available. To address this, adaptive basis selection strategies have been proposed, which aim to find the significant basis terms with low cost (less samples) (Jakeman et al., 2015; Lüthen et al., 2020). The commonly used adaptive basis selection strategies can be summarised as follows: (1) Predefine a candidate set with basis terms of different degrees and add significant basis terms from this candidate set to the PCE model until the required modelling accuracy is achieved (Zhao et al., 2019); (2) Add basis terms of higher degree to the current PCE model iteratively for training while removing the insignificant basis terms until the required modelling accuracy is achieved (Blatman and Sudret, 2011; Jakeman et al., 2015; Loukrezis et al., 2020; Thapa et al., 2020). Hence, if those basis terms which are not decisive in modelling the target problem are removed and only the significant basis terms are retained, a precise PCE model can be

trained with low training cost even if a high truncation degree is used. Blatman and Sudret (2011) predefined a degree upper bound and commenced the modelling process from a small degree value. In the iteration process, the PCE model with gradually increased truncation degree was trained. The Cross-Validation (CV) error was utilised to evaluate the performance of the trained PCE model in each iteration. Finally, the PCE model with the smallest CV error was considered as the best one, and the corresponding truncation degree was regarded as the most appropriate. Zhao et al. (2019) proposed a forward-backward strategy to select significant basis terms. A set of basis terms with different polynomial degrees was chosen as a candidate pool, from which one can choose basis terms to be added into the PCE model. In each iteration, only one basis in the candidate pool which contributes the most to the PCE modelling of the target problem was added, while a basis term which has the smallest coefficient value in the current PCE model was removed. Jakeman et al. (2015) proposed a forward neighbour expansion strategy, in which the $l_1$ minimisation method was employed for regression calculation. After training the PCE model in each iteration, a set of neighbour basis terms were defined based on the basis terms with non-zero coefficients in the current PCE model. Then the PCE model would be re-trained after expanding with these neighbour basis terms. By continually expanding the PCE model and discarding the insignificant basis terms with zero-value coefficients, the PCE model could converge to an accurate solution rapidly.

It deserves to note that using sparse representation methods to train the PCE model is also a kind of adaptive basis selection strategies. Compared with the popularly studied adaptive basis

selection strategies, sparse representation methods require the number of unknown coefficients and samples quantity be kept in an appropriate proportion (Jakeman et al., 2015). In other words, the adaptive basis selection strategies can solve problems with much more unknown coefficients than the sparse representation methods when the samples quantity is limited.

Although the sampling problem and basis selection problem have been thoroughly studied in the past, in general one of the problems was pursued by leaving out the other. In real applications, however, the two problems co-exist, which should be solved simultaneously. Only a few studies have addressed these two problems in a concurrent way. Thapa et al. (2020) proposed an adaptive weighted least-squares PCE modelling method with basis adaptivity and sequential sampling. A two-loop framework was formulated, where the outer loop was to increase the basis terms of higher degree and the inner loop was to sequentially add samples. In parallel, the insignificant basis terms were removed from the PCE model during the iteration process to accelerate the convergence. Loukrezis et al. (2020) combined the alphabetic-optimal design with a basis expansion strategy to build an adaptive PCE modelling framework. The discrete least-squares method was employed for regression calculation, with a threshold being defined according to the alphabetic-optimal design. If the optimisation target value in the alphabetic-optimal design exceeds the threshold, the basis addition operation will be executed; otherwise, the sequential sampling process will be executed. The CV error was calculated to terminate the algorithm when the obtained PCE model is in desired accuracy.

In earlier studies, PCE was primarily applied in the fields of UQ, sensitivity analysis and reliability analysis, which have low requirements on the modelling accuracy, and thus the computational cost is low. Nevertheless, the merits of PCE model, such as simple model structure and convenient training process, facilitated this technique to be applied in more complicated problems. As a result, the adaptive modelling strategies become more and more critical in the PCE modelling, which attracts more and more attention and requires more in-depth investigations in future.

### 2.1.4 Other surrogate modelling techniques

In addition to the PCE technique, there are many other mathematical tools that can help build surrogate model, such as Gaussian Process Regression (GPR), regression trees, Relevance Vector Machine (RVM) and Artificial Neural Network (ANN), etc. (Chakraborty, 2021; De'Ath and Fabricius, 2000; Fricker et al., 2011; Lee et al., 2021; Zhou et al., 2013). Once an accurate surrogate model is obtained, the model responses can be collected with low computational cost. Compared to PCE, these surrogate modelling techniques have different merits and drawbacks in practice. For comparison, two popularly used surrogate modelling techniques in recent years, GPR and ANN, are briefly introduced.

GPR was originated from the so-called Kriging that was proposed in geo-statistics (Matheron, 1967). It belongs to the Bayesian regression method, which can discover the potentially unknown

relationship existed in data while no specific function form should be nominated beforehand (Rasmussen and Williams, 2006). This method supposes that each point of the target output is a random Gaussian variable, and any finite assembly of these variables have a joint Gaussian distribution. Embarking on giving a prior to the unknown function, the likelihood function can be formulated based on the available observations of the target output. By estimating the parameters in the GPR model through the maximum likelihood method, the posterior distribution of the unknown function can be derived, which can then be used to predict the points of interest (Rasmussen and Williams, 2006). It is worth noting that the predictions are all Gaussian distributions. We can not only get predictions at unobserved points (mean function) but also estimate the uncertainty about the predictions (variance function). This property has been leveraged to solve numerous practical problems. For example, Krause et al. (2008) proposed a near-optimal sensor placement strategy based on GPR and information theory for the sake of evaluating the temperature distribution in a room. The thermometers were sequentially placed at the positions where have large prediction uncertainties in the GPR model. As a result, the temperature distribution can be predicted with less and less uncertainty. Xu et al. (2011) employed GPR to model the spatiotemporal physical phenomena, and a mobile sensor network navigation strategy was proposed to minimise the prediction uncertainty from GPR. The active learning was achieved by absorbing new measurements and discarding old measurements so as to make the navigation real-time and effective. Apart from building model based on real observations, GPR was also utilised to build surrogate for

simulation models. For example, a GPR model was built for uncertainty quantification of the frequency response functions based on the Finite Element (FE) model (Fricker et al., 2011). Furthermore, the differential equations to describe physical laws could also be learnt (Gregory et al., 2019; Raissi et al., 2017), and the burden in solving the differential equations can be lightened. Even though GPR performs well in nonlinear modelling, it has nonnegligible drawback when processing data with large size. In the training process, the covariance matrix, which is calculated from the samples at the observed points, should be inverted to learn the hyperparameters. Clearly, the inverse operation has a $\mathcal{O}(N^3)$ complexity (Rasmussen and Williams, 2006), in which $N$ denotes the number of observations. Thus, the computational cost for GPR training will grow exponentially with the increase of data size.

As an outstanding pattern recognition approach, ANN originated from the emulation of how human brain works, which performs well in modelling nonlinear and complex tasks (Gomes et al., 2018; Liu et al., 2018b). Different from the GPR and PCE techniques that have explicit function expressions, ANN organises a neuron structure with several layers, and each layer contains a certain number of nodes that connect to each other. The input data could be transferred from one layer to another layer sequentially with some simple operations and ultimately output from the output layer. Parameters existed in each layer, or called weights, can change their values to record the data features. Using the labelled data to train the network, the information involved in the data will be learnt and reserved by estimating these parameters. Finally, the trained ANN can be utilised

for prediction or classification. Obviously, an ANN structure can involve different number of layers with different number of nodes, so this approach is flexible in dealing with numerous problems. According to the layer number, layer operation and learning mode, ANN can be categorised into various types, such as feedforward network, recurrent networks, modular networks, convolutional neural network and so on (Funahashi and Nakamura, 1993; Ramos and Martínez, 2013; Truong et al., 2020; Zhang and Berardi, 2001), and they have been applied in many areas including but not limited to damage identification, fault diagnosis and time series responses prediction (Guo et al., 2016; Ramos and Martínez, 2013; Ren et al., 2018; Tang et al., 2019; Truong et al., 2020). With long-time studies, the merits of ANN can be summarised as: i) robust noise tolerance, which means that the errors contained in the training data will not affect the solution, ii) flexibility and adaptivity, that is, the model structure has a large design freedom to adapt to different kinds of problems, iii) parallel processing, which means they can handle more than one task at the same time. However, there is a fatal problem in ANN. The network and its behaviour are unexplainable. For a trained ANN model, the features in the target task cannot shown by the model structure. The ANN model can only be treated as a “black box”, which reduce our confidence to the trained model. Moreover, the flexibility and adaptivity of ANN make it hard to determine the model structure. There is no specific rule in choosing the quantity of layer and node so that the most appropriate network is only achieved by experience and trial.

Through comparison among PCE and these two techniques, it can be summarised that PCE

has the following merits and drawback as a surrogate modelling method:

The merits are: (1) PCE has a simpler model structure than GPR and ANN. The expression of a PCE model is a linear combination of polynomials with different degrees, so no intricate mathematical operation is involved. However, the expression of GPR is derived from Bayesian regression and ANN consists of numerous layers and neutrons; both of them are more difficult in formulation. Also, the simple model structure of PCE facilitates manipulations for in-depth analysis. (2) The PCE model has explicit functional expression, which enables direct analysis based on the components in the function. For example, the variance-based sensitivity analysis can be conveniently computed from the coefficient values. (3) The training process of PCE is simple, as the regression-based approaches are easy to be implemented. However, GPR generally employs a gradient descent method to find the optimal solution of the hyperparameters, in which the initial values of the hyperparameters have tremendous impact on the training results. In addition, an inverse calculation is required in GPR, which gives rise to a high computational burden. (4) The model selection is uncomplicated for PCE, where the truncation degree can be adaptively selected. But GPR requires to choose a proper kernel function for representing the target of interest, and ANN also needs an appropriate network structure including the determination of layer number and neutron number. Adaptive selection is not suitable for these two techniques.

The drawback is: PCE only employs polynomials to represent the target of interest. However, GPR and ANN do not define the functional forms, so they have wider applicability than PCE.

## 2.2 Multi-fidelity Modelling Technique

Intuitively, building surrogate model directly based on the model evaluations from the numerical model with high complexity is generally time-consuming and costly. In engineering simulations, the meshing size in the numerical model can be adjusted to achieve balance between the modelling precision and cost. The model with coarse mesh is in low cost to evaluate while has large modelling errors. Such model is known as Low-Fidelity (LF) model, and the corresponding model evaluations are called LF data. By contrast, the numerical model with high precision is named High-Fidelity (HF) model, e.g., the mesh size is small, and the corresponding data is called HF data. These data are time-consuming to collect. To train a reliable surrogate model with acceptable cost, the concept of Multi-Fidelity (MF) modelling was proposed by leveraging the information from both the LF and HF data (Forrester et al., 2007). As a result, the surrogate model can be built with high precision but less reliance on the HF data.

The MF modelling technique originated from the so-called global-local approximation method proposed for Response Surfaces (RS) (Haftka, 1991). Knill et al. (1999) first presented the correction RS model in aerodynamic modelling, which is a prototype of the MF technique in current studies. This research utilised a correction RS function to fit the difference between the LF data and HF data. Subsequently, the MF RS model were established by adding the correction RS function to the LF model. In applying to a supersonic high-speed civil transport case, it was shown that the computational cost in solving problems with high input dimension could be reduced. In 2002,

another type of MF modelling was proposed by Vitali et al. (2002), called scaling operation. The ratio between the HF data and LF data was calculated, and this ratio was multiplied to the LF model to build the MF model.

Apart from the research based on RS, Kenndy and O'Hagan (2000) proposed the MF framework based on Kriging method, which was named co-Kriging. A new modelling operation, autoregressive form, was developed. In the autoregressive modelling, the HF surrogate model is regarded as the addition of a scaled LF surrogate model with a discrepancy term. The LF surrogate model and the discrepancy term are modelled by GPR (Kriging), and the scaling value is called regression parameter. In 2017, Park et al. (2017) compared several different MF modelling frameworks based on co-Kriging. In contrast to the traditional training approach that estimating the regression parameter and the parameters in the discrepancy terms separately, this research demonstrated that the Bayesian calibration which trains the parameters all at once was the best.

In the field of PCE, the MF modelling technique was first proposed by Ng and Eldred (2012) according to the correction and scaling operations developed by Vitali et al. (2002), in which the non-intrusive method (stochastic collocation) was employed to train the PCE model. The correction and scaling operations were named additive correction and multiplicative correction respectively. The additive correction was to model the discrepancy between the HF data and LF data by using a PCE model, and this corrective function was added to the Low-Fidelity PCE (LFPCE) model to build the Multi-Fidelity PCE (MFPCE) model. Similarly, the multiplicative correction was to model

the ratio of the HF data and LF data by using a PCE model and to multiply the ratio to the LFPCE model to build the MF model. Moreover, a combinative correction was proposed to combine the additive and multiplicative corrections. As the multiplicative and combinative corrections in MFPCE is too complicated in practice, the additive correction was the most widely implemented strategy in the previous studies (Palar et al., 2015, 2016, 2018; Wang et al., 2019). Originally, the MFPCE was trained by non-intrusive methods, such as stochastic collocation, sparse grid or regression (Ng and Eldred, 2012; Palar et al., 2016). With the development of CS theory, the sparse representation was recently employed in MF modelling to further save the training cost (Cheng et al., 2019; Eldred et al., 2017; Rumpfkeil and Beran, 2020).

So far, the concept of MF modelling has been utilised in many fields for various targets. Zhou and Tang (2021) employed the MF modelling technique and GPR method for the purpose of uncertainty quantification of structural mode shape. Data from full-scale FE analysis was regarded as HF, and data from order-reduced FE model was deemed as LF. Compared to building surrogate model with LF data alone, the MF model had higher precision while similar computational cost. Yang et al. (2019) presented an MFPCE modelling framework for structural analysis, in which the sparse representation approach was leveraged to train the PCE coefficients. To analyse a damaged plate, the FE model considering material nonlinearities was regarded as HF model to afford HF data. The FE model which only considered linear material property was employed to generate LF data. Results shown that the MF modelling technique can reduce the computational cost while

maintaining the accuracy. Jin et al. (2021) implemented the MF technique to fuse strain data from sensors with different precisions. The point strain gauges could provide accurate monitoring data, but the measurements have a low-spatial resolution. The distributed strain sensor enabled collecting measurements in large areas (a high-spatial resolution) while the data had a relative low accuracy. Thus, the measurements from point strain gauges were LF data, and the observations from distributed sensors were HF data. The strain distribution over the whole structure was predicted with high precision under limited budgets by building the MF surrogate model. Furthermore, Olleak and Xi (2020) proposed to integrate the simulation data with limited experiment data by adopting the MF modelling technique. This so-called physics-based MF modelling approach was applied in the laser melting process to predict melt pool size under intended printing configurations. A similar framework was also adopted by Pepper et al. (2021). To conclude, the concept of MF modelling has potential to be applied in various scenarios for the sake of reducing modelling cost and increasing modelling precision.

Even if the MF modelling technique has been studied for a long time, there are still some limitations: (1) The highest polynomial degree of the correction function should lower than the LF model. Since the HF data is far less than the LF data, a higher degree of the correction function than the LF model may induce overfitting problem (Palar et al., 2016). (2) The LF model should be able to capture the global trends of the target task, otherwise the accuracy of MF modelling cannot be ensured (Liu et al., 2018a).

## 2.3 Vibration-based Damage Identification Methods for Bridges

In early 1990s, Rytter (1993) first suggested four levels of damage identification: detection of damage existence, identification of damage location, quantification of damage severity, and prediction of the remaining life. Since then, research efforts have been devoted into the first three levels and many damage identification methods have been proposed in the bridge engineering. Among those, methods based on the structural vibration features have become the popular ones because the vibration signals are easily to be measured with low cost. The theory of vibration-based methods is that the structural damage would alter the bridge physical properties, which can be reflected in the dynamic characteristics of bridge. Hence, the structural damage or variations of physical properties can be easily detected by merely comparing different stages of those identified characteristics. Furthermore, to recognise the damage location and severity, an inverse problem can be raised and solved if the relationship between the damage parameters and vibration characteristics could be well identified. Basically, the vibration-based damage identification methods were either physics-based or data-driven.

Physics-based methods, also called model-based methods, are to establish a representative physical model of real structure. By comparing the model predictions with real observations, the parameter variations in the physical model can be identified. The physical model is generally established based on the physical laws, such as FE model or partial differential equation. For instance, Hou et al. built a FE model of bridge and solved an inverse problem based on the FE model

to identify damages (Hou et al., 2018b, 2020). According to the equations of motion and the characteristic equations, Huang et al. developed hierarchical SBL approaches for damage identification on structures (Huang et al., 2017a, 2017b; Huang and Beck, 2015). Gregory et al. (2019) employed the Euler-Bernoulli equation to characterise the relation between the external force and structural responses in a sleeper beam, and the GPR technique was then used to help detect damages. To summarise, there are several challenges in the physics-based methods. First of all, the physical model is established based on the physical laws instead of the real structure, so the modelling errors cannot be circumvented. Second, to identify the damage location and severity, an inverse problem is generally required to be addressed. However, the uniqueness and stability of the solution are hard to be guaranteed because the inverse problem may be ill-posed or ill-conditioned. Under this circumstance, the sparse representation methods help settle down the problem to some extent (Hou et al., 2020; Huang et al., 2017a). Last but not least, predictions from the physical model do not contain measurement noise. Ignoring measurement noise may result in unexpected results.

Data-driven approaches do not rely on physical laws but directly conduct analysis based on real measurements. It is often convenient to discover the regularity in the measurements collected from structure in a reference state through developing statistical models, and the current structural condition can be evaluated by comparing the new measurements with the predictions from the statistical models. Beyond that, exploring damage indicators based on some techniques is another widely used data-driven approach. Obviously, the modelling error is circumvented in the data-

driven approaches, and the measurement noise can also be taken into account. For example, ANN has been widely applied to build the data models that map the relation between the damage parameters and model responses, and the models were then used for damage identification (Jayasundara et al., 2020; Ni et al., 2000; Wang and Ni, 2015). Pathirage et al. (2019) presented a sparse encoder based deep neural network for pattern recognition, and the sparse regularisation was employed to identify sparse damages. Neves et al. (2017) combined ANN with GPR approach to enhance the damage detection ability. In addition to ANN, some other statistical models were also used, such as Kalman filter (Yan et al., 2004), autoregressive model (Mosavi et al., 2012), support vector machine (Worden and Lane, 2001), etc. To build damage indicators, Principal Component Analysis (PCA) was broadly applied by extracting damage sensitive responses from raw measurements (Sen et al., 2019a; Zhu et al., 2018, 2019). Clearly, some data-driven approaches should train statistical models in advance, in which sufficient measurements in both healthy and damaged conditions are required. However, data from the damaged structure is unavailable upon most occasions in practice. Subsequently, data-driven approaches were usually employed to detect the existence of damage, while the damage location and severity are hard to be recognised. Moreover, the statistical model training is also time consuming with the growth of data volume.

It can be found in the past investigations that the modelling error issue in physics-based methods and the data deficiency problem in data-driven approaches were the primary challenges that hindered the progress of damage identification on bridges. To overcome these challenges,

several strategies were proposed to combine the physics-based methods and data-driven approaches, e.g., model updating and digital twin (Behmanesh and Moaveni, 2016; Bigoni and Hesthaven, 2020; Gregory et al., 2019; Schlune et al., 2009). The basic principle is that the physical model or its surrogate model is employed to describe the overall pattern of the bridge system, and the scarce measurements are leveraged to update the physical model (or its surrogate model). Therefore, the modelling error can be eliminated, and the measurement noise can be considered. In the meantime, the requirement on measurements is also reduced benefitting from the use of physical model. In such methods, furthermore, the use of a surrogate model rather than using a physical model can help release the computational burden in damage identification, because the surrogate model is much cheaper to be evaluated than the complex physical model. Besides, in the traditional model updating methods, only the value of structural parameters can be updated. The utilisation of a surrogate model enables the updating of the inaccurate physical law instead of only updating the parameter values. Ultimately, the damage identification can be performed based on the updated model. In recent years, such combined damage identification approaches have become more and more popular. Further research on applications to the bridge engineering is still required.

With a model that can correctly describe the relation between the damage parameters and vibration characteristics, an inverse problem is ready to be solved for getting the damage location and severity. As aforementioned, the inverse problem will be ill-posed since the number of measurement points on the bridge is generally less than the quantity of monitored components. To

overcome this challenge, a reasonable prior was introduced, that is, the number of damages is sparse with respect to the monitored bridge components. Then, the sparse representation algorithms could be applied to garner the sparse solution. In the past decade, various sparse representation algorithms were employed in the field of damage identification for bridges. Zhou et al. and Hou et al. proposed to use $l_1$ regularisation to induce sparsity on the damage solution (Hou et al., 2018b; Zhou et al., 2015). The damage sensitivity matrix was calculated by taking derivative on the vibration responses with respect to the damage parameters, and the $l_1$ regularisation term was added to induce sparsity on the damage parameters. SBL is another sparse representation algorithm that was introduced for damage identification in many studies (Chen et al., 2020; Hou et al., 2019, 2020; Huang et al., 2017a). As a Bayesian regression approach, it can identify the sparse solution with a confidence level, so the damage can be quantified under a probabilistic framework. In addition, Entezami et al. (2017) developed a new iterative regularisation method for solving sparse damage identification problem in truss bridge. The measurement noise and modelling errors were both considered, and the proposed method shown promising performance in identifying the damage location and severity. Even though the sparse representation has been introduced to cope with the damage identification problems with good results, a limitation should be mentioned is that the system of problem should be linear. To handle problems with nonlinear system, the algorithms should be improved.

## 2.4 Temperature Effects Overview

As discussed previously, the environmental effects cannot be ignored in the damage

identification for bridge structures, such as temperature, humidity, and wind, etc. Many studies have demonstrated that temperature is the most influential factor among those environmental effects, and the impacts caused by humidity and wind are negligible (Huang et al., 2018; Kita et al., 2019; Sun et al., 2018; Xia et al., 2011). Therefore, research efforts have been devoted to the damage identification considering temperature variations for a quite long time. Generally, there are two categories in the damage identification methods considering the temperature effect: eliminating the temperature effect and utilising the temperature effect.

As the temperature-induced variations in structural responses can often exceed the variations caused by damage (Bao et al., 2012; Erazo et al., 2019; Huang et al., 2018), a common idea is to eliminate the temperature effect from the observed structural responses. Many studies have discovered that the temperature has an approximately linear relation with the vibration frequencies of a structure (Bao et al., 2012; Magalhães et al., 2012). By investigating this linear relation, the thermal impact can be identified and eliminated from monitoring data. Comparison studies demonstrated that the accuracy of damage identification results was increased by incorporating the temperature variations. Shokrani et al. (2018) introduced a PCA-based method to distinguish the response variations caused by temperature and damages. In case that the relation between temperature and structural responses is linear or weakly nonlinear, it was manifested that PCA-based method can successfully separate the temperature-induced response change, and the damage can be correctly localised. In addition to the linear assumption, some studies found that the relation

between the temperature and frequencies may be nonlinear for large bridge structures (Ni et al., 2005; Zhou et al., 2011). Deng et al. (2010) utilised six-order polynomial models to describe the correlations of frequency-temperature and displacement-temperature. The variations in frequency and displacement caused by temperature could be effectively eliminated, and the damage was successfully detected. Furthermore, some damage indicators that are sensitive to damage but not sensitive to temperature were discovered, for example, frequency ratios (Deraemaeker et al., 2008; Surace and Bovsunovsky, 2020).

However, several studies have found the relation between temperature and structural responses is indeterminate for various structures, which means that recognising such relation is challengeable, especially when data is not sufficient. Therefore, an alternative solution of utilising the temperature effect were developed, which detected damages by comparing the extracted temperature-responses relations before and after damage (Kromanis and Kripakaran, 2016; Xia et al., 2017). Zhu et al. (2019) extracted the temperature-strain relation over time by using a blind source separation technique. Combining with the moving PCA, the damages on a truss bridge were detected and localised. A merit of this study was that the thermal measurements are not required. Yarnold and Moon (2015) discovered a three-dimensional near-flat plane among the temperature, local mechanical strains and global displacements. Such relation was found insensitive to normal operational change, while some scenarios that would result in the structural performance change were able to be detected. To summarise, among the damage identification methods utilising the

temperature effect, damage should induce nonnegligible changes on the temperature-induced responses.

Undeniably, understanding the mechanism of temperature on the bridge structure is crucial for developing damage identification methods. Basically, the influence of temperature on bridge can be described in the following aspects (Han et al., 2021): i) elastic modulus change; ii) thermal expansion; iii) boundary condition change. The elastic modulus of bridge material will change with temperature varying, while this relation is usually easy to be characterised with acceptable accuracy (Sun et al., 2018). The increasing temperature will also cause the expansion of bridge and then affect its dynamic characteristics. Comparing with the elastic modulus change, the mechanism of thermal expansion was found much weaker (Xia et al., 2012). More importantly, the change of boundary condition caused by temperature was deemed much more complicated in previous research (Han et al., 2021; Peeters et al., 2001). Under some circumstances, the boundary condition may exhibit a nonlinear variation with respect to the temperature change, and such variation is commonly difficult to be represented or simulated in a physical model. For instance, if the thermal expansion of bridge is constrained or partially constrained at supports, axial force will be generated on the bridge girder, and thus the structural frequencies will change (Han et al., 2021). Also, freezing of supports will cause a nonlinear change of bridge frequencies (Sen et al. 2019b). Sun et al. (2018) argued that correctly identifying bridge damages without considering the freezing effect is very difficult.

In addition to the above issues, it deserves to note that evaluating the temperature distribution

on a bridge is always challengeable, since the distribution is generally non-uniform and time-varying. Due to the limitation on the number of measurement points, the temperature distribution is difficult to be comprehensively recognised. In the past studies, three simplification methods were presented and used in the field of SHM:

(1) Only one temperature variable is considered. For bridges with small scale, temperatures measured from different positions at the same time might be similar, so the analysis results can have acceptable accuracy if the average temperature is used (Erazo et al., 2019; Wang et al., 2020c). Besides, if a temperature variable is representative in correlating to the model responses, this variable can be regarded as the dominated temperature in the structure, e.g., canonical correlated temperature (Huang et al., 2020). For those bridges with large scale, such simplification will lose its efficacy.

(2) The temperature gradient along the vertical direction of the bridge deck should be considered. Suffering from the solar radiation, temperature gradient will be nonnegligible in the vertical direction of bridge deck, and the temperature variation in the longitudinal direction of bridge is normally insignificant and can be ignored. Various studies considered the temperature gradient in the damage identification problem (Huang et al., 2019; Sun et al., 2019). Wang et al. (2021) studied the temperature gradient distribution in a steel-concrete composite bridge deck. The measurements obtained from an experimental test were compared with the simulation results and the recommendations in specifications. Results demonstrated that the specifications could provide

promising references for simulating the temperature gradient. Therefore, the distribution of temperature gradient can be determined according to the measurements from the bridge deck along the vertical direction and the specifications. In general, the linear or bilinear gradient model (two or three temperature variables) was enough to simulate the temperature gradient distribution with desired accuracy (Xia et al., 2018; Xu et al., 2019).

(3) Considering single temperature variable for individual bridge component. Due to the complicated configuration of large-scale bridges, the temperature distribution on the whole bridge will be extremely complex. Xia et al. (2013) proposed to simplify the temperature distribution on a large suspension bridge by applying a single temperature value instead of a temperature distribution to each bridge component (cable, deck and tower). Results shown that the simulation results with the simplified temperature distribution can provide responses that agree well with the real measurements.

To summarise, the temperature effect on bridge is extremely intricate, and it is rarely possible to completely characterise them in simulation, especially for large-scale bridges. Efforts on structural damage identification considering temperature effect are still inadequate and more investigations are required.

# CHAPTER 3

# HYBRID SEQUENTIAL SAMPLING STRATEGY FOR POLYNOMIAL CHAOS EXPANSION

## 3.1 Introduction

Polynomial Chaos Expansion (PCE) has been widely applied in the field of Uncertainty Quantification (UQ) or sensitivity analysis since a small truncation degree could meet the requirements. Modelling problems with high input dimension and high truncation degree is still a challenge for PCE method. The main reasons behind the problem are the high cost of regression calculation and the demands on large amount of samples and model responses for training. Therefore, increasing research has been concentrated on the sampling techniques to help reduce the samples for training, which could help release the burden of acquiring massive model evaluations. In general, sparse representation and experimental design are two strategies that have been demonstrated to be effective in reducing the number of samples, as introduced in Chapter 2.

Owing to the "sparsity of effect" principle, usually real-world problems are sparsely represented, or at least compressible on polynomial chaos (Montgomery, 2006). The sparse representation methods feature an appealing merit in garnering the PCE coefficients with less

amount of model responses. Bayesian Compressive Sensing (BCS) is one of the state-of-art sparse representation methods, which is used in this study to calculate the PCE coefficients. It introduces hierarchical priors instead of a direct Laplace prior to induce feasible Bayesian inference, and this operation still endows a similar effect to $l_1$ minimisation (Babacan et al., 2010). An appealing benefit from BCS than other regularisation algorithms is that this method is under the Bayesian framework. It can make predictions to unknown points with a distribution instead of a deterministic way, which provides a reference of our confidence on the estimated value at a point. This property can be further utilised for experimental design.

To collect a set of well-designed samples, or called Experimental Design (ED), for PCE modelling, many strategies were proposed to instruct the sampling process based on the distributions of input variables, model truncation degree and the corresponding orthogonal polynomials, which we refer to as input-dependent only approaches, such as Latin Hypercube Sampling (LHS) method, D-optimal sampling strategy and coherence-optimal sampling approach, etc (Blatman and Sudret, 2010; Burnaev et al., 2017; Hampton and Doostan, 2015; Jakeman et al., 2015). The utilisation of input information improves the stability of the regression calculation during the sampling process. In addition to the input dependent sampling strategies, Bayesian Experimental Design (BED) was developed to collect samples in light of the information about experiment outcomes and modelling results. This strategy is capable of constructing suitable sampling processes for problems with distinct complexity. The collected samples will contain the most information

which is needed for accurate modelling. Thus, it will reduce the number of required samples.

In this chapter, a hybrid sequential sampling strategy is proposed to take into account both the input information and the target model feature by combining compressive sampling and BED. Here, the compressive sampling method is named coherence-optimal sampling in the field of PCE, in which a lower bound of sample quantity for $l_1$ minimisation is deduced. Samples are collected that contribute more to the regression accuracy when $l_1$ minimisation methods are used. To make the coherence-optimal samples be collected in a sequential way, the progressive LHS method is adopted. This method is easier to be combined with the coherence-optimal sampling method to generate space-filling samples than the traditional Monte-Carlo methods. BED is the simplest output-oriented approach that can be easily combined to build the hybrid method. The procedures of building the hybrid method are as follows. First, a sequential sampling framework is established to collect samples that approximately match the coherence-optimal distribution, which is derived from the compressive sampling theory, during the iteration process. Then, by resorting to the BCS method and information theory, favourable sampling points in each iteration are determined according to the modelling results, substituting for randomly selecting sampling points. Finally, the performance of the proposed sampling strategy is evaluated on several analytical functions case through comparison with three input-dependent only sampling methods and two output-dependent only sampling methods, and it is also applied on an engineering case for parameter sensitivity analysis.

## 3.2 Polynomial Chaos Expansion and Sparse Representation

### 3.2.1 Polynomial chaos expansion

For completeness of description and unity of symbols, the principle of PCE is briefed here. It represents the scalar model output $y$ as an expansion of a set of orthogonal polynomials of input variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots \xi_d)$ with $d$ dimensions. Here the orthogonal polynomials defined in a probabilistic space can be expressed as:

$$\int \phi_i(\boldsymbol{\zeta})\phi_j(\boldsymbol{\zeta})p(\boldsymbol{\zeta})d\boldsymbol{\zeta} = w_i\delta_{ij} \tag{3.1}$$

where $\phi_i(\boldsymbol{\zeta})$, $\phi_j(\boldsymbol{\zeta})$ are two orthogonal polynomials of random variable $\boldsymbol{\xi}$ with different degrees and $p(\boldsymbol{\zeta})$ is the probability density function of $\boldsymbol{\xi}$; $w_i$ is a constant; $\delta_{ij}$ is a Dirac function, which is equal to one when $i = j$ and otherwise zero. For computational convenience, the orthogonal polynomials are commonly normalised with respect to the probability density function so that $w_i$ is equal to one when $i = j$, namely orthonormal polynomials (Xiu and Karniadakis, 2002). Thus, the PCE can be expressed as:

$$y = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} c_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \tag{3.2}$$

where $c_{\boldsymbol{\alpha}}$ are unknown coefficients; $\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ are multivariate orthonormal polynomials, which can be written as a tensor product of univariate polynomials when the input variables are assumed to be independent from each other:

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{i=1}^{d} \phi_{\alpha_i}(\xi_i) \tag{3.3}$$

and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots \alpha_d)$ represent the indices of the polynomial bases, in which $\alpha_i$ ($i \in [1, d]$) denotes the degree of each independent variable $\xi_i$ in a polynomial term. Therefore, for different distributions, there are different types of orthogonal polynomials with respect to the Probability Density Functions (PDF). For example, the Hermite polynomials are associated with the Gaussian distribution and the Legendre polynomials are orthogonal with respect to the uniform distribution. Some of the commonly used polynomial types are summarised in Table 3-1 (Xiu and Karniadakis, 2002):

Table 3-1 Type of univariate orthogonal polynomials with different continuous variables

| Random variable | Polynomial type | Support |
|---|---|---|
| Uniform | Legendre | $[a, b]$ |
| Gaussian | Hermite | $(-\infty, +\infty)$ |
| Beta | Jacobi | $[a, b]$ |
| Gamma | Laguerre | $[0, +\infty)$ |

Under the circumstance that no standard polynomials are defined for given input variables with known PDF, the isoprobabilistic transform technique can be adopted (Torre et al., 2019). Considering a vector of random variables $\boldsymbol{Z}$ with joint PDF $\boldsymbol{Z} \sim f_{\boldsymbol{Z}}(z)$, an isoprobabilistic transform operator $\boldsymbol{\mathcal{F}}$ exists such that:

$$\boldsymbol{\xi} = \boldsymbol{\mathcal{F}}(\boldsymbol{Z}), \quad \boldsymbol{Z} = \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\xi}) \tag{3.4}$$

where $\boldsymbol{\xi}$ is a random vector with independent components which generally distributes according to one of the distributions in Table 3-1. We can then rewrite Equation (3.2) as:

$$y = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{\mathcal{F}}(\boldsymbol{Z})) \tag{3.5}$$

In practice, to facilitate calculation, PCE representation in Equation (3.2) will be truncated such that only a finite number of coefficients are kept in training. There are two commonly used truncation strategies. The standard truncation strategy is to keep the total degree of PCE not exceeding a given degree $p$:

$$\mathcal{A}^{p,d} \equiv \{\boldsymbol{\alpha} \in \mathbb{N}^d : \|\boldsymbol{\alpha}\|_1 \leq p\} \tag{3.6}$$

and thus, the cardinality in the truncated PCE is:

$$P = \binom{d+p}{d} = \frac{(d+p)!}{d!\,p!} \tag{3.7}$$

The second truncation strategy, which is a modification of the standard truncation strategy, is called hyperbolic truncation as defined in Equation (3.8) (Blatman and Sudret, 2011):

$$\mathcal{A}^{p,d,q} \equiv \left\{\boldsymbol{\alpha} \in \mathbb{Z}^{p,d} : \|\boldsymbol{\alpha}\|_q \leq p, 0 < q \leq 1\right\} \tag{3.8}$$

It is obvious that the hyperbolic truncation strategy corresponds to the standard truncation strategy if $q = 1$. For $q < 1$, smaller $q$ represents truncating more basis terms with interaction effects among different variables. A simple 2D case is shown in Figure 3.1 with different values of $p$ and $q$. Without specification, the standard truncation strategy is used hereinafter.

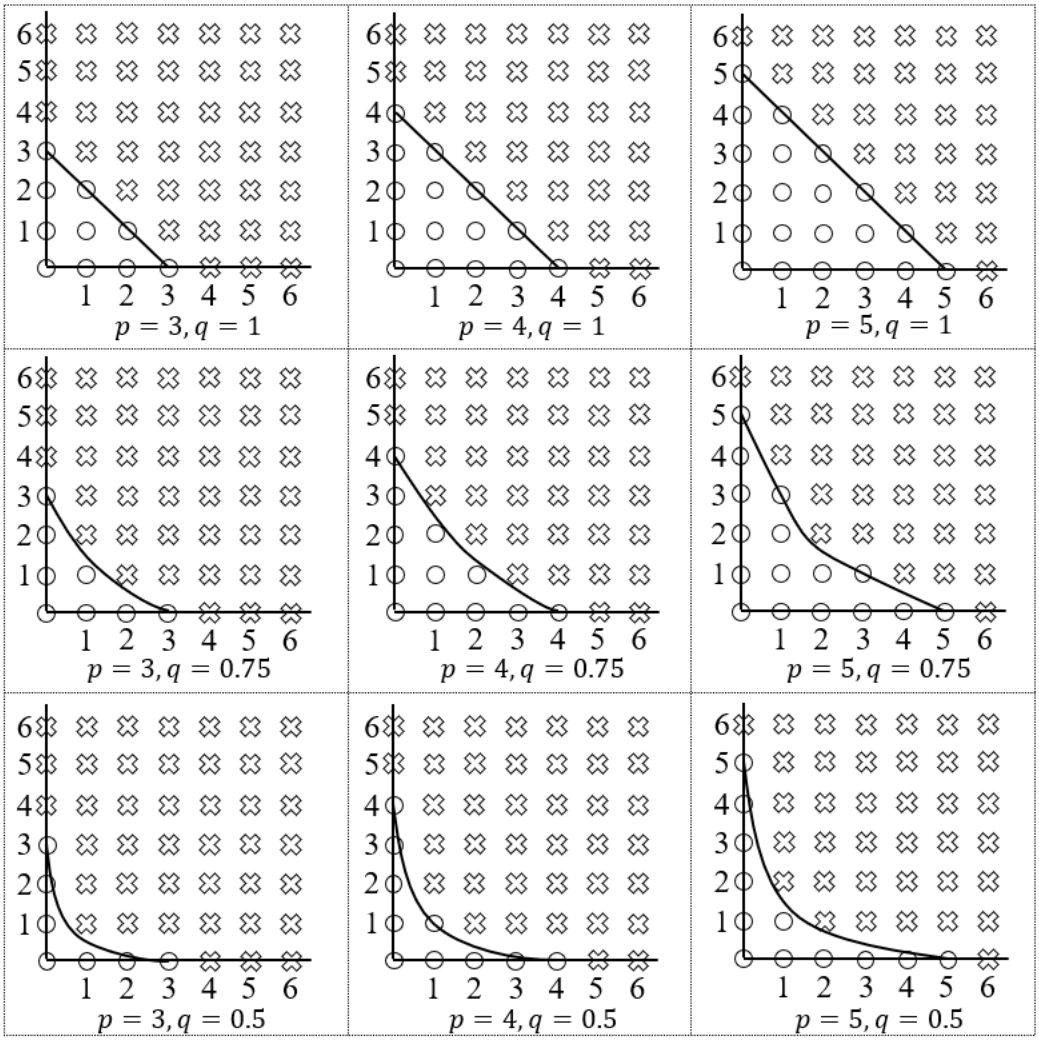



Figure 3.1 Simple 2D case of hyperbolic truncation (circle represents the indices of kept bases; cross denotes the indices of truncated bases)

By using the truncation strategy, the model output $y$ is then expressed as a sum of the truncated PCE and a truncation error $\varepsilon$:

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} c_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) + \varepsilon = \boldsymbol{\Psi}(\boldsymbol{\xi})\boldsymbol{c} + \varepsilon \tag{3.9}$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi}) = [\psi_{\boldsymbol{\alpha}_1}(\boldsymbol{\xi}), \psi_{\boldsymbol{\alpha}_2}(\boldsymbol{\xi}), \cdots, \psi_{\boldsymbol{\alpha}_P}(\boldsymbol{\xi})]$ and $\boldsymbol{c} = [c_{\boldsymbol{\alpha}_1}, c_{\boldsymbol{\alpha}_2}, \cdots, c_{\boldsymbol{\alpha}_P}]^T$. By resorting to the regression methods (e.g., the OLS method), if an ED with $N$ samples from the random variables $\{\boldsymbol{\xi}^{(i)}\}_{i=1}^{N} = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \cdots, \boldsymbol{\xi}^{(N)}\}$ is made and the corresponding model observations $\boldsymbol{Y} = \{y^{(i)}\}_{i=1}^{N}$

at these sampled places are obtained, we can reformulate the above equation as:

$$\boldsymbol{Y} = \boldsymbol{\Psi c} + \varepsilon \tag{3.10}$$

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(1)}\right) \\ \boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(2)}\right) \\ \vdots \\ \boldsymbol{\Psi}\left(\boldsymbol{\xi}^{(N)}\right) \end{bmatrix} \boldsymbol{c} + \varepsilon \tag{3.11}$$

The OLS method seeks to obtain the solution of $\boldsymbol{c}$ by solving:

$$\hat{\boldsymbol{c}} = arg\min_{\boldsymbol{c}} \|\boldsymbol{\Psi c} - \boldsymbol{Y}\|_2 \tag{3.12}$$

To get unique solution, the number of samples *N* should be larger or at least equal to the number of unknown coefficients, and an oversampling rate of 2~3 is recommended for obtaining reliable and robust results (Hosder et al., 2007).

### 3.2.2 Bayesian compressive sensing

From Equation (3.7), it is apparent that the cardinality of PCE will suffer from the "curse of dimensionality" issue. The required model observations by OLS method will grow exponentially with the increase of input dimension or PCE degree, which would be unaffordable for complex target models. Thus, the sparse representation methods that can solve Equation (3.12) with sampling number far less than the number of unknown coefficients become more promising. Some of them belong to $l_1$ -norm regularisation methods and some have equivalent effect of $l_0$ -norm regularisation. By placing penalty on the unknown coefficients, the regression problem in Equation (3.12) can be redefined as the following optimisation problem:

$$\hat{\boldsymbol{c}} = arg\min_{\boldsymbol{c}}\{\|\boldsymbol{\Psi c} - \boldsymbol{Y}\|_2^2 + \lambda\|\boldsymbol{c}\|\} \tag{3.13}$$

where $\lambda$ is a regularisation coefficient which controls the penalty weight to the coefficients, and the norm in $\|\boldsymbol{c}\|$ can be either $l_0$, $l_1$ or $l_p$ $(0 < p < 1)$ norm to induce different sparsity. Among them, $l_1$ minimisation is preferable since it can drive a sparse solution and meanwhile is more tractable than $l_0$- and $l_p$-norm regularisation (Bruckstein et al., 2009). In this study, the BCS method proposed by Babacan et al. (2010), which has been demonstrated to endow a similar effect to $l_1$ minimisation through introducing a hierarchical form of Laplace priors to the coefficients, is recalled as the sparse regression procedure to solve Equation (3.13). The principle of this method is briefly introduced in the following.

Based on Equation (3.10), the likelihood function with model observations can be formulated by assuming the truncation error follows a zero mean Gaussian distribution:

$$p(\boldsymbol{y}|\boldsymbol{c}, \beta) = N(\boldsymbol{y}|\boldsymbol{\Psi c}, \beta^{-1}) \tag{3.14}$$

where $\beta$ is assigned with a Gamma distribution as conjugate prior. To induce $l_1$ minimisation, hierarchical priors are introduced to coefficients $\boldsymbol{c}$ instead of a Laplace prior since this setting is a conjugate prior to the likelihood function for tractable Bayesian inference (Babacan et al., 2010; Jiang et al., 2018). The hierarchical modelling are as follows:

$$p(\boldsymbol{c}|\boldsymbol{\lambda}) = \prod_{i=1}^{P} N(c_i|0, \lambda_i) \tag{3.15}$$

$$p(\lambda_i|\kappa) = \Gamma(\lambda_i|1, \kappa/2) = \frac{\kappa}{2}\exp\left(-\frac{\kappa\lambda_i}{2}\right) \tag{3.16}$$

Thus, the modelling result of the coefficients $\boldsymbol{c}$ becomes

$$p(\boldsymbol{c}|\kappa) = \int p(\boldsymbol{c}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\kappa)d\boldsymbol{\lambda} = \prod_i \int p(c_i|\lambda_i)p(\lambda_i|\kappa)d\lambda_i$$
$$= \frac{\kappa^{P/2}}{2^P} exp\left(-\kappa^{1/2}\sum_i |c_i|\right) \tag{3.17}$$

which is an expression of Laplace distribution. Likewise, $\kappa$ is modelled with a Gamma hyperprior $p(\kappa)$.

Then, the Bayesian inference can be deduced to obtain the posterior distribution of all parameters:

$$p(\boldsymbol{c}, \boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{c}, \beta)p(\boldsymbol{c}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\kappa)p(\beta)p(\kappa)}{p(\boldsymbol{y})} \tag{3.18}$$

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{c}, \beta)p(\boldsymbol{c}|\boldsymbol{\lambda})p(\boldsymbol{\lambda}|\kappa)p(\beta)p(\kappa)d\boldsymbol{c}d\beta d\boldsymbol{\lambda}d\kappa \tag{3.19}$$

Nevertheless, the calculation of the marginal likelihood $p(\boldsymbol{y})$ is analytically intractable. The calculation of optimal parameters values through maximum likelihood estimation becomes inaccessible. Thus, an asymptotic approximation is employed to generate an iteration process (Babacan et al., 2010; Beck and Katafygiotis, 1998). To this end, the posterior distribution is decomposed as:

$$p(\boldsymbol{c}, \boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) = p(\boldsymbol{c}|\boldsymbol{\lambda}, \kappa, \beta, \boldsymbol{y})p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) \tag{3.20}$$

where the posterior distribution of the target coefficients $\boldsymbol{c}$ conditional on all other parameters and observations can be calculated from:

$$p(\boldsymbol{c}|\boldsymbol{\lambda}, \kappa, \beta, \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{c}, \beta)p(\boldsymbol{c}|\boldsymbol{\lambda}) \tag{3.21}$$

which is a Gaussian distribution $N(\boldsymbol{c}|\mu_c, \Sigma_c)$ with

$$\mu_c = \Sigma_c \beta \boldsymbol{\Psi}^T \boldsymbol{y} \tag{3.22}$$

$$\Sigma_c = \left(\beta \boldsymbol{\Psi}^T \boldsymbol{\Psi} + \text{diag}(1/\lambda_i)\right)^{-1} \tag{3.23}$$

Subsequently, the unknown parameters in Equation (3.22) and Equation (3.23) can be estimated through maximising $p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y})$. Since $p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) \propto p(\boldsymbol{y}, \boldsymbol{\lambda}, \kappa, \beta)$, maximising $p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y})$ is alternatively pursued by maximising $p(\boldsymbol{y}, \boldsymbol{\lambda}, \kappa, \beta)$, which can be readily calculated from $p(\boldsymbol{y}, \boldsymbol{c}, \boldsymbol{\lambda}, \kappa, \beta)$ by simply integrating out $\boldsymbol{c}$. That is,

$$p(\boldsymbol{y}, \boldsymbol{\lambda}, \kappa, \beta) = \int p(\boldsymbol{y}|\boldsymbol{c}, \beta) p(\boldsymbol{c}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\kappa) p(\beta) p(\kappa) d\boldsymbol{c} \tag{3.24}$$

Lastly, by taking partial derivatives to the parameters $\boldsymbol{\lambda}, \kappa, \beta$ to maximise the logarithm of Equation (3.24), the optimal values of them can be obtained, and $\mu_c$ and $\Sigma_c$ can be updated iteratively until convergence. However, the gradient algorithm is in low calculation efficiency due to calculation burden in Equation (3.22) and Equation (3.23). To reduce the computational cost and speed up the iteration process, a fast Laplace algorithm was proposed by Babacan et al. (2010) to make the BCS method more applicable to large-scale problems. The central idea is to update only a single $\lambda_i$ instead of updating the whole $\boldsymbol{\lambda}$ in each iteration so that the updates of $\mu_c$ and $\Sigma_c$ become efficient. The details can be found in (Babacan et al., 2010).

At the beginning of the algorithm, the model will be set as empty, namely $\boldsymbol{\lambda} = \boldsymbol{0}$ (Babacan et al., 2010). So only $\beta$ needs to be assigned with a proper value to launch the algorithm. Even though a fixed value, $\beta^{-1} = 0.01\|\boldsymbol{y}\|_2^2$, was suggested by Babacan et al. (2010) for initialisation, it is found

that different values of $\beta$ will affect the solution. Cross-Validation (CV) has been proven a good tool to provide proper choice of the hyperparameters in PCE, such as estimating the polynomial degree and the error tolerance (Huan et al., 2018; Jakeman et al., 2015), when no validation dataset is available. Here we choose CV technique to help determine a proper value for $\beta$ as initialisation, which was also adopted by Lüthen et al. (2021). The principle of the so-called *K*-fold CV is to randomly partition the training data into *K* parts with equal size; one part is regarded as test data while the remaining $K-1$ parts are used as training data. Thus, the average predicted error on the test data could be evaluated. By taking each part as test data in turn, a total of *K* average predicted errors can be obtained, and the average of these *K* values is defined as the CV error. When applying the *K*-fold CV to choose a proper value of the model parameter, a bunch of values of the target parameter should be chosen in advance. For each value with given training data, a CV error will be estimated. The parameter value with the smallest CV error will be viewed as the best choice. In this study, 10-fold CV is used to select the initial value of the parameter $\beta$.

BCS is a sparse representation method that can provide high degree sparsity to solutions (Babacan et al., 2010). As such, the PCE model resulting from BCS can be expressed with a very simple structure. An appealing benefit of BCS compared with other regularisation algorithms is that, based on the modelling result, this method can make predictions to unknown points with a distribution instead of a deterministic way, which provides a reference of our confidence on the estimated value at a point. This property will be utilised in BED as described in the next section.

## 3.3 Sequential Sampling

In this section, two sampling schemes to be utilised in the proposed method are outlined. One is the coherence-optimal sampling strategy originated from the compressive sampling theory, and the other is the BED. It has been proven that the former can achieve the minimum sample quantity while preserving satisfactory recovery performance in solving $l_1$ minimisation problem (Hampton and Doostan, 2015). This sampling method was compared with several input-dependent only sampling methods by verifying on several benchmark function tests by Lüthen et al. (2021) and was shown promising to collect high-quality samples. The latter capitalises on the information provided by model observations through Bayesian modelling techniques, which has a wider applicability to various problems (Chaloner and Verdinelli, 1995). Then, a hybrid sequential sampling method is proposed by combining two approaches. A sampling framework to sequentially sample from the coherence-optimal distribution is first introduced. Then, BED in conjunction with the differential entropy is fused with the coherence-optimal sampling to develop a hybrid sequential sampling strategy, which is expected to achieve high convergence rate and stable modelling performance.

### 3.3.1 Coherence-optimal sampling

The coherence parameter in the context of PCE is defined as (Hampton and Doostan, 2015):

$$\mu(\boldsymbol{\xi}, \mathbb{Z}^{p,d}) = \sup_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} \max |\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|^2 \tag{3.25}$$

which has been testified to be a vital parameter of bounding the number of samples needed for

accurate recovery of $l_1$ minimisation problem. A lower coherence parameter value represents a smaller bound of the sample number. Equation (3.13) can be rewritten in a weighted form:

$$\hat{\boldsymbol{c}} = arg\min_{\boldsymbol{c}}\{\|\boldsymbol{W\Psi c} - \boldsymbol{WY}\|_2^2 + \lambda\|\boldsymbol{c}\|_1\} \tag{3.26}$$

Here we focus on $l_1$ minimisation problem. The weight matrix $\boldsymbol{W}$ is an identity matrix in Equation (3.13) so that $\boldsymbol{W\Psi} = \boldsymbol{\Psi}$. The regression matrix $\boldsymbol{W\Psi}$ is controlled by the truncated polynomials at sampled points. To achieve the lowest sample bound, the concept of isotropy of regression matrix is introduced. The regression matrix in the context of PCE is isotropy if the ED is sampled from the input distribution, and the lowest sampling number could be achieved in a large probability for accurate reconstruction (Hampton and Doostan, 2015; Krahmer and Ward, 2014). To conclude, with the input variables $\boldsymbol{\xi}$ and the corresponding truncated polynomials, a coherence parameter value can be determined to bound the lowest sample number for accurate recovery of $l_1$ minimisation problem in a high likelihood. To attain this bound, the samples should be collected from the input distribution. Therefore, in target to recover an $l_1$ minimisation problem with satisfactory performance and as small number of samples as possible, one should reduce the coherence parameter value and sample from the corresponding distribution to make the regression matrix isotropic.

Now we consider a new polynomial basis which is modified from the original standard polynomials by multiplying a weight coefficient, where the coherence parameter of the new polynomial basis would be the minimum as defined below:

$$\mu_{min}(\boldsymbol{\gamma}, \boldsymbol{\mathcal{A}}^{p,d}) = \sup \max_{\boldsymbol{\alpha}\in\boldsymbol{\mathcal{A}}^{p,d}} |w(\boldsymbol{\zeta})\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|^2 \tag{3.27}$$

in which the weight coefficient is determined by:

$$w(\boldsymbol{\zeta}) = \frac{1}{cB(\boldsymbol{\zeta})} \tag{3.28}$$

where $B(\boldsymbol{\zeta}) = \max_{\boldsymbol{\alpha}\in\boldsymbol{\mathcal{A}}^{p,d}} |\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|$. Then the new variables $\boldsymbol{\gamma}$ with distribution in compliance with the new orthogonal polynomial basis $\psi_{\alpha}^{new}(\boldsymbol{\gamma}) = w(\boldsymbol{\zeta})\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ will have the PDF:

$$f_{\boldsymbol{\gamma}}(\boldsymbol{\zeta}) = c^2 B(\boldsymbol{\zeta})^2 f(\boldsymbol{\zeta}) \tag{3.29}$$

where $c^2 = \int B(\boldsymbol{\zeta})^2 f(\boldsymbol{\zeta}) d\boldsymbol{\zeta}$ is a normalisation constant. Thus, the weight matrix could be calculated from the weight coefficient:

$$\boldsymbol{W}(i,i) = w\left(\gamma^{(i)}\right) \tag{3.30}$$

where $\gamma^{(i)}$ is a sample from the distribution $f_{\boldsymbol{\gamma}}$. The detailed statement of the coherence-optimal sampling and its convergence theorems can be found in (Hampton and Doostan, 2015).

In principle, the coherence-optimal sampling strategy amounts to discovering a weighted orthogonal polynomial system $w(\boldsymbol{\zeta})\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ that achieves the lowest coherence parameter value; then sampling from the corresponding new distribution will make the new regression matrix $\boldsymbol{W\Psi}$ isotropic. As a result, the perfect recovery of $l_1$ minimisation problem will be guaranteed in a large probability while the samples can be controlled to a small quantity (Hampton and Doostan, 2015). The coherence-optimal sampling strategy instructs the sampling process based on the input variables, model degree, and the corresponding orthogonal polynomials. Thus, the regression matrix $\boldsymbol{W\Psi}$ will be well-conditioned to ensure stable computation of the coefficients. One difficulty in this

method is that the new distribution is generally not a standard or known distribution, and even the calculation of the normalisation constant might be difficult. Direct sampling from this distribution will be undoubtedly hard. Hampton and Doostan (2015) proposed to use the Monte Carlo Markov Chain approach to generate samples without calculating the specific expression of the distribution. This sampling method was originally proposed for Hermite and Legendre polynomials. For problems with input distribution that is neither uniform distribution nor Gaussian distribution but has known PDF, the isoprobabilistic transform can be adopted to convert the original input distribution to uniform or Gaussian distribution (Fajraoui et al., 2017). Then the coherence-optimal sampling method can be performed on the transformed distributions.

### 3.3.2 Bayesian experimental design

It is worth mentioning that, in the coherence-optimal sampling method, only input information from the input distribution and the truncated orthogonal polynomials is used; but the model observations, which represent the target model characteristic, are not considered. Most of the commonly used sampling methods such as LHS, D-optimal, etc., are all input-dependent only. For various problems which can be modelled by the PCE technique with the same input dimension, input distribution and total polynomial degree, the formulated coherence-optimal distributions in the coherence-optimal sampling method are identical to each other. However, for problems with different complexity, the target model characteristic will influence the sampling result. BED takes

advantage of this kind of information to instruct the sampling process (Ji et al., 2008; MacKay, 1992; Seeger and Nickisch, 2008). A noticeable metric used in BED to quantify the information inherent in variables is the Shannon entropy in the context of information theory. The core idea in BED with information theory is that samples will be collected from a predefined candidate pool to maximise the information gain about the target model (Papadimitriou et al., 2000). When applying BED for sequential sampling, the unobserved positions in the candidate pool are assumed as random variables, and the trained PCE model by the use of Bayesian modelling techniques can make predictions to the unobserved positions with distributions instead of deterministic evaluations. The unreliability of the trained model can be reflected in the predicted results by evaluating the prediction uncertainty at each unobserved point. Then, we intend to get the next sample at a location which owns a large prediction uncertainty, because the observation collected at this position will provide more information to train the PCE model in the next iteration (Ji et al., 2008). The PCE model trained using such collected samples and observations would reduce its uncertainty to a large extent. The target of interest is in usually assumed as a continuous variable, so the differential entropy which extends from the Shannon entropy is (Ji et al., 2008; Nielsen, 2019):

$$H_D(X) = -\int f(x)\log\big(f(x)\big)\,dx \tag{3.31}$$

where $f(x)$ denotes the PDF of a continuous variable $X$. The unit of the differential entropy depends on the base of the logarithm. The commonly used bases are 2, Euler's number and 10, which generate entropy units of bits, nats and bans, respectively. A large entropy value implies that

less information is known about $x$ and more uncertainty exists in this variable.

In this study by employing the BCS method to calculate the PCE coefficients, the trained PCE model can make predictions to those unknown places with Gaussian distributions. By considering a set of points $\boldsymbol{\mathcal{X}}_{N_t} = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \cdots, \boldsymbol{\xi}^{(N_t)}\}$ to be predicted, the polynomial basis will form a new matrix $\boldsymbol{\Psi}_{new}$ with $N_t$ rows. According to the posterior distribution of the coefficients, $N(\boldsymbol{c}|\mu_c, \Sigma_c)$, the predictions to these points can be obtained as (Tipping, 2001):

$$Y_{pre}(\boldsymbol{\mathcal{X}}_{N_t}) \sim N\left(\boldsymbol{\Psi}_{new}\mu_c, \boldsymbol{\Psi}_{new}\Sigma_c{\boldsymbol{\Psi}_{new}}^T\right) \tag{3.32}$$

Then the prediction $Y_{pre}\left(\boldsymbol{\xi}^{(i)}\right) \sim N(\mu_i, {\sigma_i}^2)$ at each prediction point $\boldsymbol{\xi}^{(i)}$ can be obtained. Here $Y_{pre}^{(i)}$ denotes the variable with its prediction at point $\boldsymbol{\xi}^{(i)}$. $f(Y^{(i)})$ represents the PDF of $Y_{pre}^{(i)}$. Thus, the differential entropy at each predicted point can be calculated by:

$$H_D(Y_{pre}^{(i)}) = -\int f(Y^{(i)}) \ln\left(f(Y^{(i)})\right) dY^{(i)} \tag{3.33}$$

Here we use Euler's number as base of logarithm. A point with high $H_D$ value denotes that the obtained PCE model is less certain at this point. Thus, this point has a higher priority to be selected than other points. Training a PCE model with this observation will reduce the prediction uncertainty to a large extent. For Gaussian distribution, the differential entropy can be elicited as:

$$H_D\left(Y_{pre}^{(i)}\right) = \frac{1}{2}\ln|\,{\sigma_i}^2| + \frac{1}{2}\ln(2\pi) + \frac{1}{2} \tag{3.34}$$

which is related only to the variance value and easy to be calculated. Finally, the new sample and corresponding observation will be added to the training dataset, and the PCE model will be re-

trained to evaluate that whether to add more samples.

Clearly, BED will make better use of the modelling results to make the trained PCE model more and more accurate. In another aspect, however, this strategy is different from the input-dependent approaches that the regression calculation stability in each iteration has no robust guarantee. Moreover, the Bayesian modelling approaches require enough data for training. The uncertainty will keep in a high level at almost all predicted points when scarce data is used. Under this circumstance, the predicted results are not accurate and will provide invalid instruction in the sampling process. The initial sample size for a sequential sampling process is usually arbitrary and in general insufficient for modelling. Purely relying on the BED will get samples with poor quality at the beginning of sampling process. In the next section, a new sequential sampling strategy in connection with BED will be proposed, where the samples will be constrained simultaneously to approximately follow the coherence-optimal distribution and to ensure the quality of samples.

### 3.3.3 Coherence-entropy sampling method

#### 3.3.3.1 Sequential sampling strategy from coherence-optimal distribution

To sequentially sample from a given distribution, a general framework is to first generate an initial sample set of small size from this distribution, and then gradually add samples until the quantity meets the requirement. There are two challenges in forming this framework.

The first one is about the generation of a small size of initial samples from a given distribution.

In principle, the LHS method is amenable to generating small number of samples from a known distribution with good space filling property. In each dimension of a multi-dimensional space, the definition domain of the variable is uniformly divided into $N$ equal probability slices according to the marginal distribution, where $N$ is the target number of samples. By randomly generating one sample in each slice, a total of $N$ scalar samples are generated in this dimension. Then, the scalar samples from different dimensions are randomly matched together; the generated samples in this multi-dimensional space are referred to as LHS samples, and this distribution property is called Latin hypercube property. However, LHS is originally designed for problems with independent variables, while the coherence-optimal distribution may have non-negligible dependence among variables in a multivariable problem. To apply LHS, the dependence among input variables is ignored in this study. As a result, the collected LHS samples will follow a quasi-coherence-optimal distribution instead of the coherence-optimal distribution, but these samples can distribute over the input space in a more space filling way. Besides, the generation of LHS samples from the quasi-coherence-optimal distribution is still hard since each univariate distribution is mostly not a standard distribution. Equally dividing the input space by probability distribution is intractable. In view of this, an approximate operation is proposed in this study. First, a large number of samples are generated from the target distribution by the use of MCMC method or other procedures. Then, according to the initial size of ED, the input space in each dimension is partitioned into slices so that each slice contains the same quantity of samples. As the sample quantity approaches infinity,

these slices will have the same probability. Therefore, with enough samples (e.g., $10^5$) being generated from the target distribution, these divided slices in each dimension could be thought to have approximately equal probability, and random sampling from these slices will achieve a near LHS sample set. We refer to this method as Near-LHS (NLHS) algorithm, which will be implemented to generate LHS samples from a non-standard distribution. This near LHS sample set conforms to the quasi-coherence-optimal distribution in a space filling way.

The second challenge is how to ensure that the samples always possess a promising space filling property during the sampling process, since each new sample will destroy the Latin hypercube property of the overall samples. Here, a "doubling procedure" adapted from the so-called Progressive-LHS (PLHS) algorithm (Sheikholeslami and Razavi, 2017) is employed to maintain the distribution property of samples during the sequential sampling process. When a near LHS sample set of size $n_1$ has been generated by the NLHS algorithm, the input space is uniformly sliced into $n_1$ equal probability slices by the marginal distribution in each dimension. To add samples, the "doubling procedure" is to perform extra $n_2 = n_1$ slicing operations in the input space of each dimension. A total of $2n_1$ equal probability slices in each dimension are generated. In other words, each slice from the last operation is equally divided into two parts in each input dimension. Slice from different dimensions will form an input subspace, named a block. Those blocks, where slices in all dimensions have no samples located, are defined as active blocks. Subsequently, extra $n_1$ samples are randomly sampled from these newly generated active blocks, and the resulting $2n_1$

samples in total are still LHS samples. For easy interpretation, a simple 2-D example is illustrated in Figure 3.2.

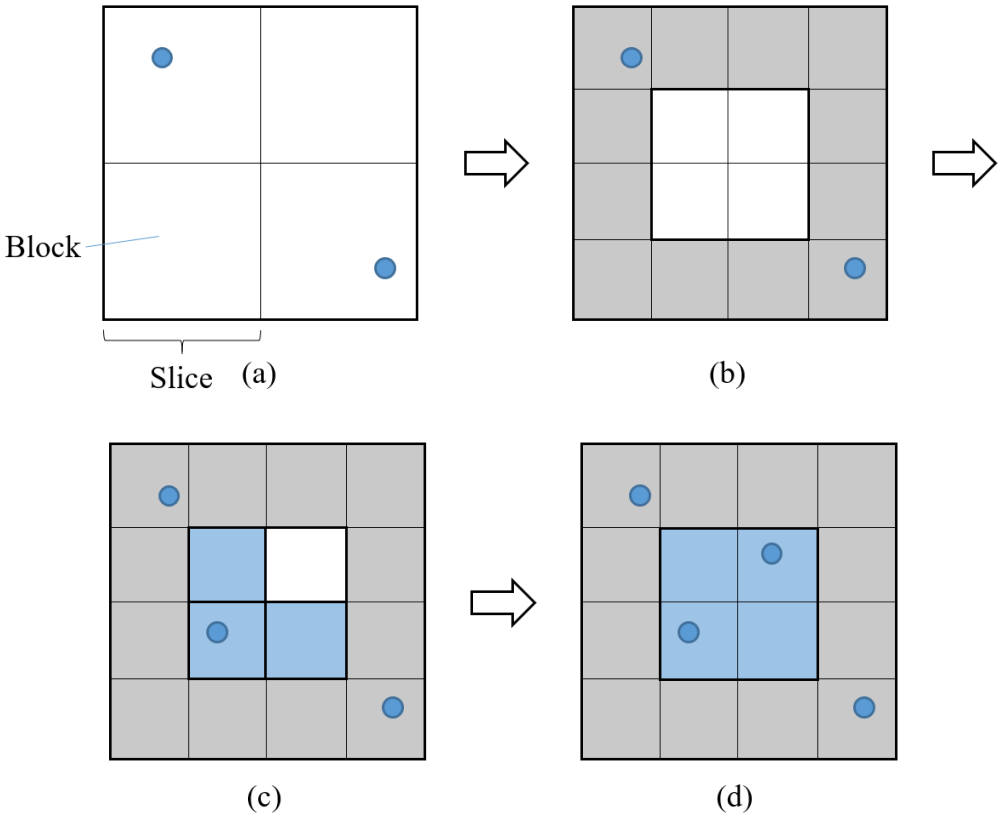



Figure 3.2 Simple 2-D example of PLHS algorithm

Assume that the initial sample set has two sample points, so the marginal probability distributions of two variables in this 2-D plane are uniformly divided into two slices, that is, four blocks, as shown in Figure 3.2(a). Two samples are generated so that each slice has one sample projection in this dimension. Then, to seek for news samples, the two distributions are doubly sliced as depicted in Figure 3.2(b). The newly generated active blocks are marked in white and inactive blocks are in grey. By randomly choosing one active block and sampling once from it, three of the original four active blocks become inactive according to the definition of active block, and they are

marked in blue in Figure 3.2(c). So, the last sample can only be collected in the final active block, as shown in Figure 3.2(d), and all the blocks become inactive. When no more active block exists, the "doubling procedure" will be executed again if more samples are needed. It is apparent that this algorithm imposes a constraint to each sampling operation, and finally the samples will be the near LHS samples. One drawback of this algorithm is that it is less flexible in controlling the sample size. It is known that the samples are near LHS samples only when no active block exists. After each "doubling procedure", the number of samples in need to construct the LHS samples will be doubled (e.g., 2, 4, 8, 16, and so on in the previous case). But this is not a problem in this study since

**Algorithm 1**. NLHS-DP algorithm

**Input:** Coherence-optimal samples $\boldsymbol{X}_N$; Initial sampling number $N_0$; Total sampling number $N_s$.

**Initialisation:**

1. Divide the coherence-optimal samples $\boldsymbol{X}_N$ into $N_0$ slices uniformly according to the input dimension $d$; Totally $d$ sets of slices will be obtained, and in each set, there are $N_0$ slices;
2. In each dimension, randomly collect one sample from each slice; $d$ sets of $N_0$ samples corresponding to the input dimensions will be obtained, and they will be randomly combined in dimension to get the initial $N_0$ samples.

**Iteration:**

1. Check the slices in each dimension; If no block is active, get into step 2; otherwise go to step 3;
2. Under the current sample number $N_k$, re-divide input space in each dimension into $2N_k$ slices so that each slice contains $\frac{N}{2N_k}$ coherence-optimal samples. New $N_k^d$ active blocks are generated;
3. Randomly select one active block, and randomly collect one sample from this active block;
4. With the newly added sample, some active blocks become inactive; re-evaluate the blocks to find the remained active blocks;
5. If the total sampling number is reached, quit iteration.

generating LHS samples in a strict way during the iteration process is not imperative. By using the NLHS algorithm with "Doubling Procedure" (NLHS-DP), a sequential sampling framework can be formulated to sample from the quasi-coherence-optimal distribution. The detailed description of this algorithm is shown in **Algorithm 1**.

#### 3.3.3.2 Hybrid sequential sampling strategy

Under the framework of NLHS-DP, samples are still randomly generated since at least one block is active in each iteration. One should first randomly choose an active block (if there are more than one active block), and then randomly sample once in this block. So, the NLHS-DP algorithm is thought to be less robust and slow to convergence. In view of this, BED is combined to build a hybrid sequential sampling strategy. Assume that a candidate set $\boldsymbol{\mathcal{X}}_{N_t} = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \cdots, \boldsymbol{\xi}^{(N_t)}\}$ with $N_t$ samples is generated in the input space beforehand. In each iteration, the model observations corresponding to the selected samples are collected to train the PCE model. Taking advantage of the BCS method, the obtained PCE model can make predictions to the points which belong to the active blocks in the candidate set. Then the differential entropy values of these points can be obtained from the prediction results, and the point with the maximum entropy value is selected as the next sample. As such, the random sampling operation in sequential sampling framework is replaced by a different sampling criterion which employs the differential entropy to instruct the sampling process. This new sampling strategy is termed coherence-entropy (Coh-entro) algorithm. The algorithmic

procedure is detailed in **Algorithm 2**.

**Algorithm 2.** Coherence-entropy algorithm

**Input:** Problem input dimension $d$; total degree $p$; initial sampling number $N_0$; total sampling number $N_{ts}$ or desired accuracy $\varepsilon$; candidate set $\boldsymbol{\mathcal{X}}_{N_t}$; selection set $S$ and observation set $Y$.

**Initialisation:**

1. Generate $N$ coherence-optimal samples $\boldsymbol{\mathcal{X}}_N$ by employing MCMC sampling method according to the problem dimension and total degree;
2. Generate $N_0$ initial samples which yield the quasi-coherence-optimal distribution and add them into the selection set $S = \boldsymbol{\mathcal{X}}_{N_0}$;
3. Get model observations at sampled points in the selection set $Y = \boldsymbol{Y}(\boldsymbol{\mathcal{X}}_{N_0})$.

**In the $i$th iteration:**

1. Build PCE model according to the selection set $S$ and the corresponding observations $Y$ by employing the BCS method;
2. If the total sampling number $N_{ts}$ is reached or the desired accuracy $\varepsilon$ is achieved, quit iteration; otherwise, get into step 3;
3. Check all the blocks; If no more active block exists, get into step 4; otherwise get into step 5;
4. Under the current sample number $N_S$, re-divide the input space into $2N_S$ slices in each dimension so that each slice contains $\frac{N}{2N_S}$ coherence-optimal samples. $N_S^d$ new active blocks are generated;
5. Abandon the candidate points which do not belong to active blocks temporarily;
6. Make prediction to the remained candidate points and calculate the corresponding differential entropy;
7. Add the candidate point $\boldsymbol{\xi}^{(i)}$ which owns the largest differential entropy value to the selection set $S = S \cup \boldsymbol{\xi}^{(i)}$, and get model observation at this point $Y = Y \cup \boldsymbol{Y}_{\boldsymbol{\xi}^{(i)}}$;
8. According to the current sample set, re-evaluate the active blocks;
9. Update the candidate set by deleting the selected point $\boldsymbol{\mathcal{X}}_{N_t} = \boldsymbol{\mathcal{X}}_{N_t} \backslash \boldsymbol{\xi}^{(i)}$.

In the proposed method, the collected samples are controlled to conform to the quasi-coherence-optimal distribution in each iteration. Even though the samples do not strictly obey the coherence-optimal distribution, they still have a good space filling property to cover the input space, and it is speculated that such distribution property could still help bound the number of samples. The isotropic property of the regression matrix cannot be achieved, but it is expected that the

regression matrix can be in good condition for regression calculation when using $l_1$ sparse representation method. BED based on the BCS method and differential entropy is employed to substitute the random selection operation in sequential sampling framework. It can be deemed that the BED is constrained so that the samples can have a good distribution property while enriching the target model information received from the corresponding observations. Thus, the PCE model can be trained to be more and more accurate with the gradually increased target model information. In addition, even if the collected samples and the corresponding observations at the beginning of the iteration process are not enough for training an accurate PCE model, generating samples from the quasi-coherence-optimal distribution still provides a foundation for further sampling and PCE modelling. The proposed method not only makes full use of the input information, but also earns much information from the observations and modelling results. It is expected to have both better convergence rate and computational stability.

### 3.3.4 Termination criterion

Several criteria have been proposed to terminate the sampling process through evaluating the precision of the obtained PCE model, such as Kullback-Leibler Divergence (KLD) and Leave-One-Out (LOO) error (Blatman and Sudret, 2008; Thapa et al., 2020). Inspired by the idea of comparing the responses obtained from the PCE model in successive iterations for modelling accuracy evaluation, a simple criterion capitalising on the changes of the PCE model mean and standard

deviation (std) in successive iterations is employed in this study to assess convergence of the obtained PCE model. These values can be easily obtained from the PCE coefficients, and they are good statistical measures to represent the PCE model. When the PCE model mean and std values in successive iterations keep stable or the change values of these two statistical measures are smaller than a given threshold, the PCE model is considered to have converged with satisfactory modelling accuracy.

## 3.4 Case Studies

In this section, three analytical benchmark functions are used to validate the proposed method with different input dimensions and degrees. The PCE models are also truncated on different degrees to test the algorithm performance. Three state-of-art input-dependent only sampling methods (Coh-Opt, D-Coh-Opt, and Seq-D-Coh-Opt) proposed by Diaz et al. (2018) and Hampton and Doostan (2015), which are based on coherence-optimal sampling and D-optimal design, are compared with the proposed method (Coh-entro). Among them, the Coh-Opt and D-Coh-Opt methods are non-sequential sampling strategies while the Seq-D-Coh-Opt method is a sequential sampling strategy. In addition, two output-dependent sequential sampling methods are also compared in this study. One is the component of the proposed method, BED with differential entropy. The other is the Expected Improvement-based Expected Loss Function (EI-ELF) criterion which was proposed by Zhou, et al. (2019a). The EI-ELF criterion is based on the PCE modelling results inferred by sparse Bayesian learning to instruct the next sample, which is an output-oriented method. The detailed EI-

ELF criterion is introduced in Appendix A. For convenience of comparison, the termination criterion used for the benchmark tests is set as sample upper limit instead of convergence evaluation. The size of the candidate sample set in the Coh-entro algorithm is set as $N_t = 10^4$. In order to assess the real precision of the obtained PCE models, a validation set with $N_{val} = 10^4$ random samples and the exact model values are used. The Relative Root Mean Square Error (RRMSE) $\varepsilon_{RRMSE}$ (Diaz et al., 2018) is calculated by:

$$\varepsilon_{RRMSE} = \sqrt{\frac{\sum_{i=1}^{N_{val}} (y_{true}^{(i)} - y_{PCE}^{(i)})^2}{\sum_{i=1}^{N_{val}} {y_{true}^{(i)}}^2}} \tag{3.35}$$

After the benchmark study, the proposed method will be applied to an engineering problem to validate its practicability. The convergence is assessed by the termination criterion which automatically determines the ED size with desired accuracy.

The Matlab codes of the Coh-Opt, D-Coh-Opt, and Seq-D-Coh-Seq methods are available online (Diaz et al., 2018). An in-house code of the EI-ELF criterion has been developed according to Zhou et al. (2019a), but the sparse solver there has been changed in this study to BCS with Laplace prior for fair comparison. The code of the BCS method is also available (Jiang et al., 2018). For the sake of clarity, the methods of Coh-Opt, D-Coh-Opt, Seq-D-Coh-Seq and BED with differential entropy are denoted as Coh, D-coh, Coh-seq, and Entro hereafter.

### 3.4.1 Two-dimensional function

The first test function is a low-dimensional analytical function as given in Equation (3.36), which has been studied as a benchmark test function for PCE modelling (Thapa et al., 2018a, 2020). The random input variables are uniformly distributed with a mean of 2.0 and a probability density function height of 0.7222. The analytical results of the mean and standard deviation (std) of this function are 0.079 and 1.124, respectively.

$$y = \ln(1 + {x_1}^2) * \sin(5x_2) \tag{3.36}$$

Two total degrees, 9 and 11, are chosen as truncation degree of the PCE model, which will result in the full PCE expansions of $P = 55$ and $P = 78$ multivariate Legendre polynomials. The initial sampling number is set as 10. Each method is calculated for 30 times to ensure statistical stability. The RRMSE results with respect to the increase of ED size are shown by box plots in Figure 3.3 and Figure 3.4.

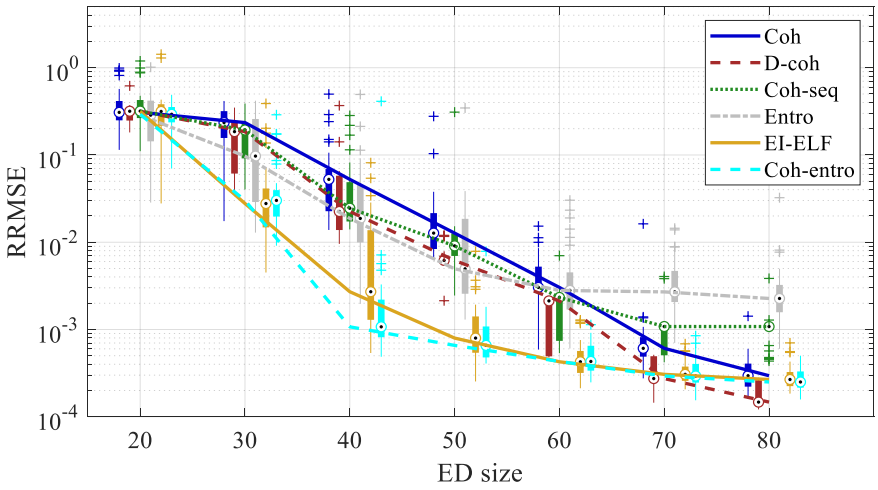



Figure 3.3 RRMSE with different ED size under degree 9

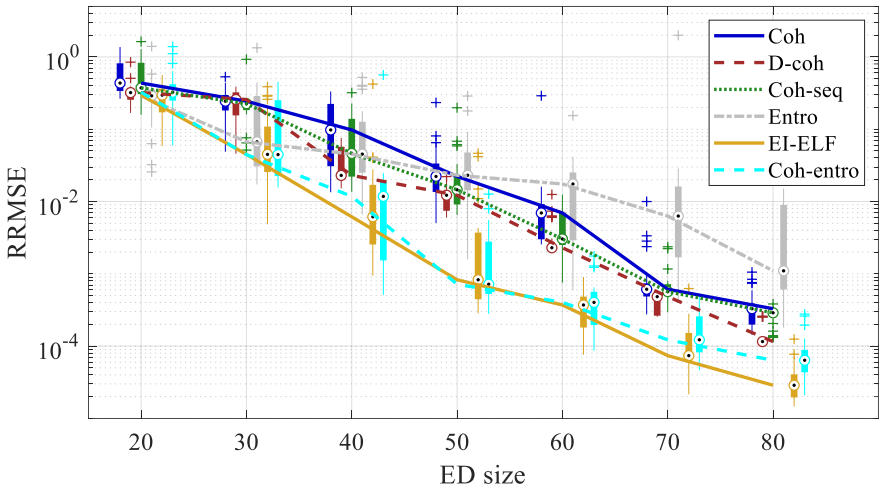



Figure 3.4 RRMSE with different ED size under degree 11

In the box plots, bold vertical lines represent the range between upper and lower quartiles of the 30 results, and fine vertical lines represent the 1.5 times interquartile range which constrains the normal value limitation. Values out of them are regarded as outliers, which are depicted as plus symbols in the plots. The dot inside the white circle denotes the median of RRMSE, and the lines that vary with respect to the ED size represent the variation of the median of RRMSE obtained by different methods. It can be observed that the EI-ELF criterion and the proposed Coh-entro strategy have lower validation error than other three input-dependent only sampling methods and the Entro sampling method under almost all circumstances. In the final converged results, except for the D-coh strategy with PCE degree 9, the PCE models with samples generated by the Coh-entro strategy and the EI-ELF criterion both have smaller validation error than the other four methods. In the circumstance of ED size much larger than the number of unknown coefficients, samples from the D-coh strategy can construct a better regression matrix by sampling at once to form well-conditioned equations which make the solutions stable. For the PCE models with degree 11, the

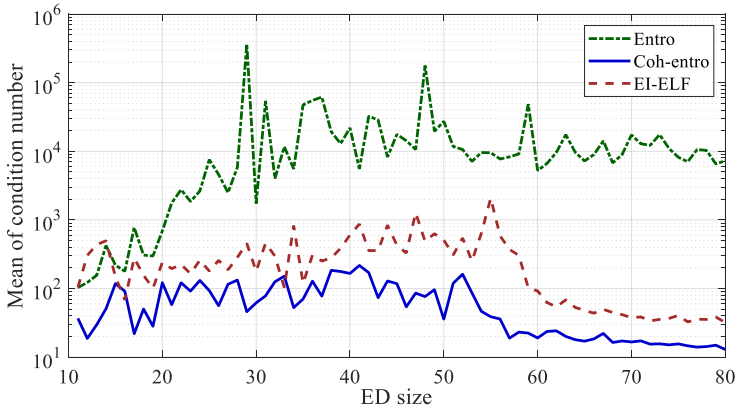
final converged results from the Coh-entro strategy have a bit higher validation error than the EI-ELF criterion, but the Coh-entro strategy still has a better performance than the three input-dependent only methods and the Entro method.

To illustrate the solution stability of the regression calculation during the sampling process, the condition number of the regression matrix and the mean and std values of the obtained PCE model calculated during the iteration process are depicted. The condition number is defined as (Thapa et al., 2020):

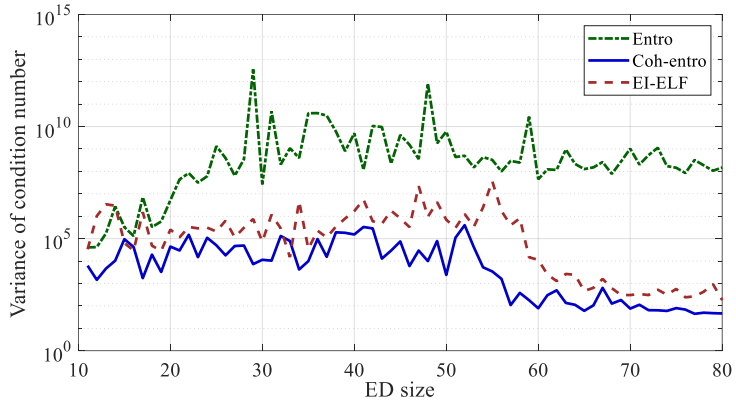
$$condition\ number = \|\boldsymbol{\Psi_c}\| \cdot \left\|\boldsymbol{\Psi_c^{\dagger}}\right\| \tag{3.37}$$

where $\boldsymbol{\Psi_c}$ represents a submatrix of $\boldsymbol{\Psi}$ in Equation (3.10) with columns corresponding to non-zero coefficients; † denotes Moore–Penrose inverse. The norm can be of any form, such as 1-norm, 2-norm, ∞-norm, etc. Here 2-norm is used. A large condition number represents poor performance of the regression matrix and implies that the regression solution becomes more sensitive to changes in the input values and observations, i.e., the correct solution is hard to find. Hence, a large condition number appearing in the sequential sampling process means that the corresponding regression solution may be incorrect. Three strategies, the proposed Coh-entro strategy, the Entro method and the EI-ELF criterion, which are all sequential sampling methods and output-oriented, are compared here. Figure 3.5 and Figure 3.7 show the mean and variance of the condition number during the 30 repeated tests with respect to the increase of ED size. It can be seen that under the PCE degree 9, the condition number of the Entro method is always the largest among the three strategies. The

condition number of Coh-entro has its mean values mostly lower than EI-ELF during the whole iteration process and ultimately becomes stable at a very low value. The variance values of the condition number have a similar variation trend with the mean values, which demonstrates that the condition number is stable in the repeated tests with the increased samples. Under the PCE degree 11, the results of the three methods all show large fluctuation, where the condition number of the Entro method is again the largest during nearly the whole iteration process. In the first half of iteration process before 50 samples, the Coh-entro strategy gives rise to lower condition number than the EI-ELF criterion; but they have similar performance in the second half of iteration process. The iteration stability of the obtained PCE models is represented by the model mean and std values. Figure 3.6 and Figure 3.8 show the statistical properties (mean and variance) of the obtained PCE model mean and std values during the 30 repeated tests against the increase of ED size. The PCE models with samples generated by Coh-entro and EI-ELF have better convergence performance than the Entro method. The PCE model obtained by the Coh-entro strategy is more stable than that resulting from the EI-ELF criterion under the PCE degree 11 as illustrated in Figure 3.8 since several large outliers emerge during the iteration process with the EI-ELF criterion.

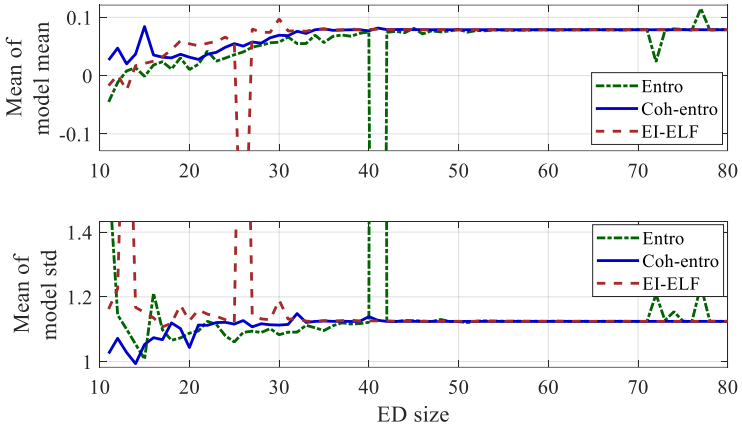



(a)

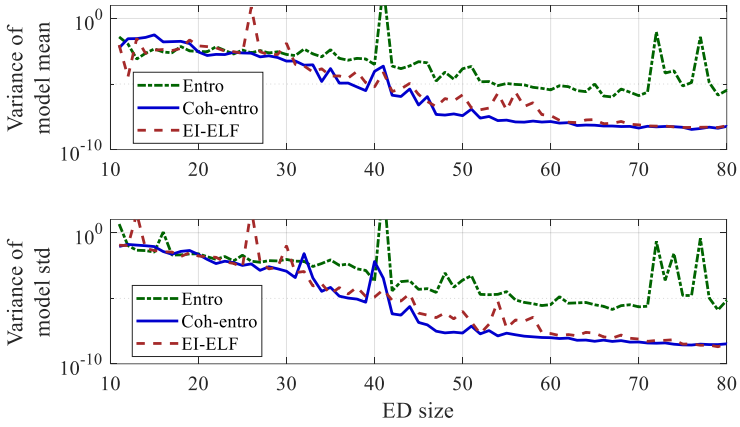



(b)

Figure 3.5 (a) Mean and (b) variance of the condition number with increasing ED size under degree 9

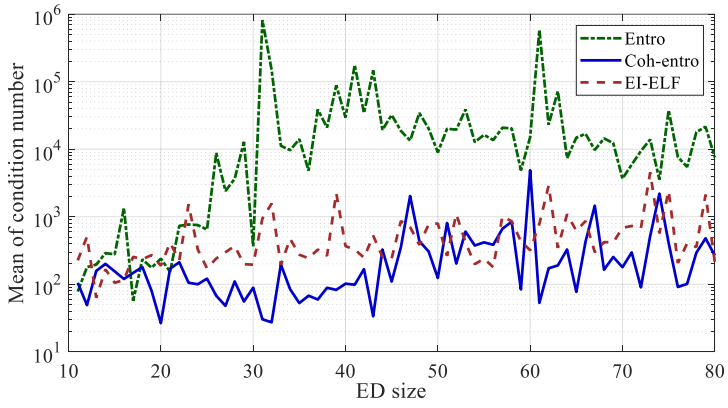



(a)

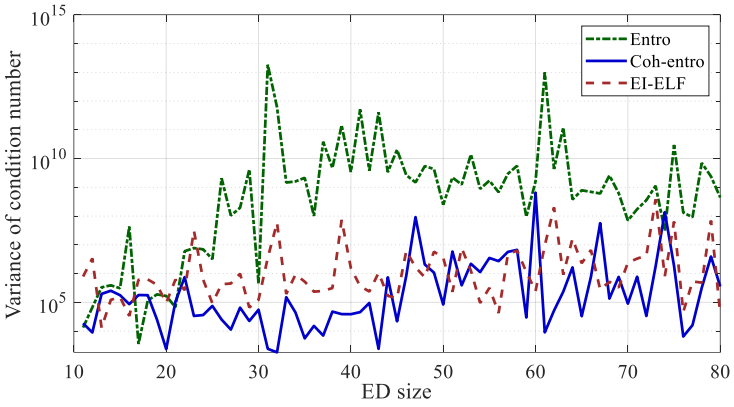



(b)

Figure 3.6 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under degree 9

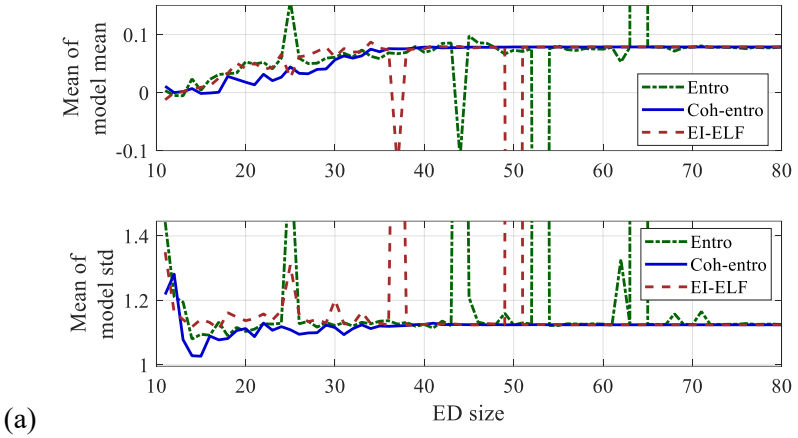



(a)

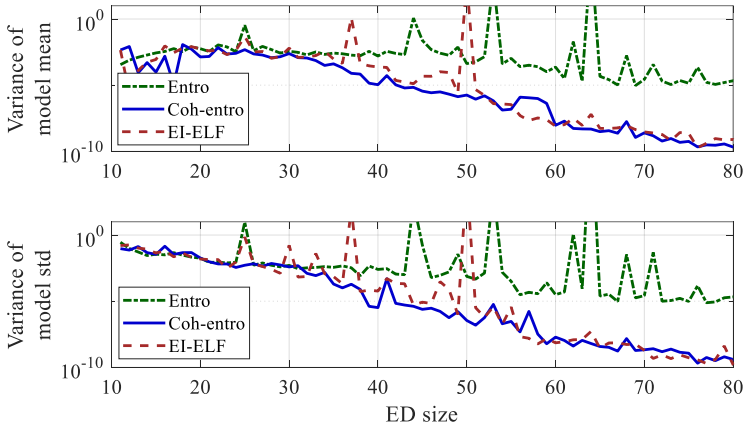



(b)

Figure 3.7 (a) Mean and (b) variance of the condition number with increasing ED size under degree 11



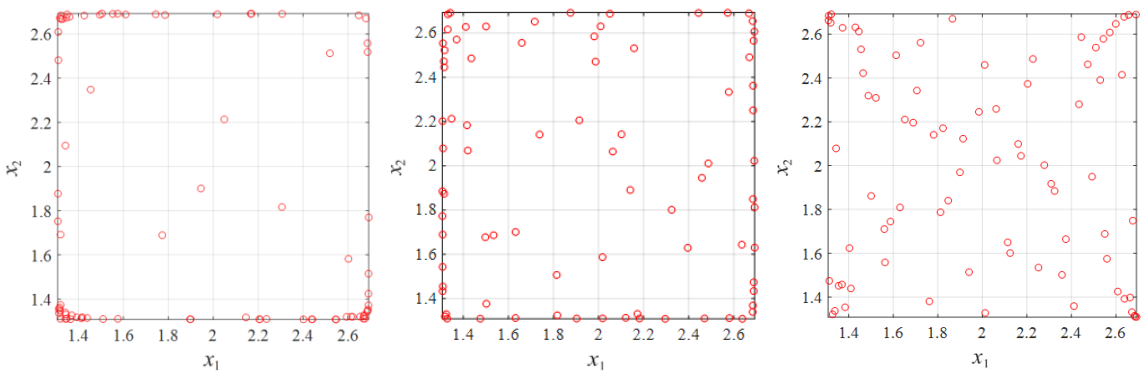

(a)




(b)

Figure 3.8 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under degree 11

The final samples from the three methods in one test under degree 11 are provided in Figure 3.9 for comparison. It is clear that the samples generated by the Entro method almost concentrate on the corners and edges of the input space, and the samples generated by EI-ELF criterion concentrate more on the edges of the input space than Coh-entro. In contrast, the proposed Coh-entro strategy places samples in a more space filling way.




(a) (b) (c)

Figure 3.9 Final samples from (a) Entro; (b) EI-ELF; (c) Coh-entro

In summary, the proposed Coh-entro strategy and EI-ELF criterion outperform the input-dependent only methods and the Entro method in terms of convergence rate, and the Entro method suffers from stability problem during the iteration process. Through constraining the samples distribution to approximately match the coherence-optimal distribution, the proposed strategy overcomes this drawback and improves the solution stability. It not only generates samples in a more space filling way but also results in a low condition number in regression calculation.

### 3.4.2 The Ishigami function

The second benchmark test function is the Ishigami function. This is a highly nonlinear and nonmonotonic function with three input variables which has been extensively studied in the past (Lüthen et al., 2021; Thapa et al., 2020; Zhou et al., 2019a). All the variables conform to uniform distributions on the interval $[-\pi, \pi]$. The expression of this function is as follows:

$$y = \sin x_1 + a(\sin x_2)^2 + b{x_3}^4 \sin x_1 \tag{3.38}$$

where $a, b$ are two parameters which are commonly chosen as 7 and 0.1, respectively. The

analytical mean is 3.5 and std value is around 3.7208. Three total degrees, 9, 12 and 14, are considered in this study, which generate different polynomial items with $P = 220$, $P = 455$, and $P = 680$. The initial sampling number is set as 20. Each method is calculated for 30 times to ensure statistical stability.

The RRMSE results with respect to the increase of ED size are shown by box plots in Figure 3.10, Figure 3.11 and Figure 3.12. It is seen that the validation error decreases with the increase of truncation degree, which implies increased precision of the obtained PCE models. In this case, the EI-ELF criterion and the proposed Coh-entro strategy have much faster convergence rate than the input-dependent only methods in all the cases with different degrees. The Entro method shows discrepant performance in different PCE degrees. It has similar convergence performance with EI-ELF and Coh-entro with degree 9, while it performs worse than EI-ELF and Coh-entro when the PCE degree increases to 12. In degree 14, the validation errors from the Entro method still decrease fast but they cannot converge to a stable result with the increase of samples. Only around 60, 80 and 80 samples are needed for convergence of Coh-entro with PCE degrees of 9, 12 and 14 respectively, and the input-dependent only methods need around 70, 100 and 120 samples respectively for convergence. The Entro method needs around 70 and 120 samples for convergence with PCE degrees of 9 and 12 respectively. The EI-ELF criterion requires 60, 70 and 100 samples for convergence with PCE degrees of 9, 12 and 14 respectively. This demonstrates that by the use of the model observation information under the Bayesian framework, the sampling process will have

a quite fast convergence rate; but using only BED may not be able to achieve stable solutions. Moreover, the proposed Coh-entro strategy has an apparently better convergence performance than the EI-ELF criterion, particularly for the PCE models with degree $p$ = 14.

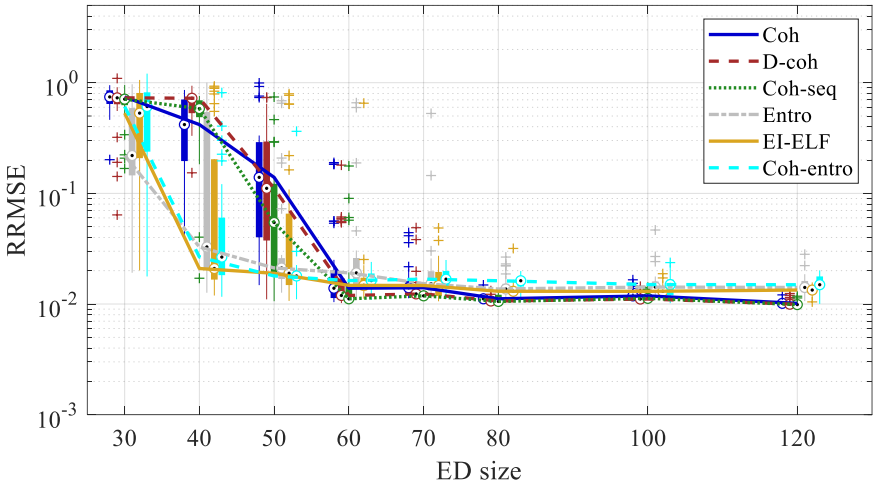



Figure 3.10 RRMSE with different ED size under degree 9

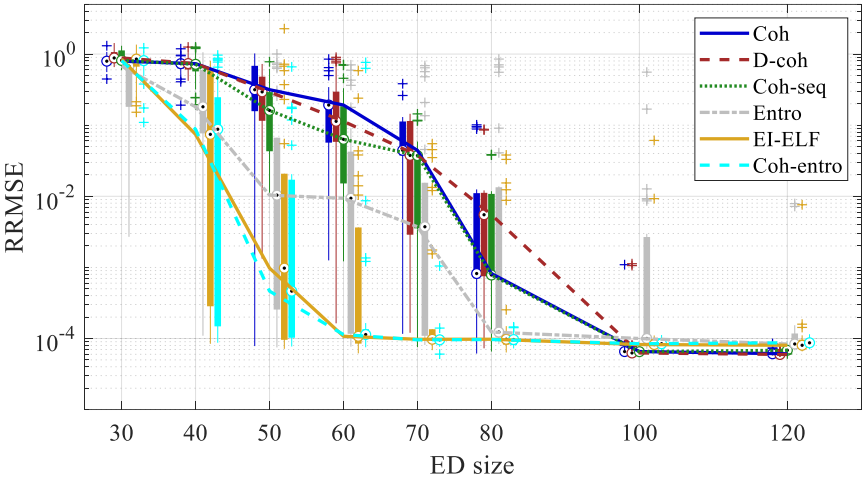



Figure 3.11 RRMSE with different ED size under degree 12

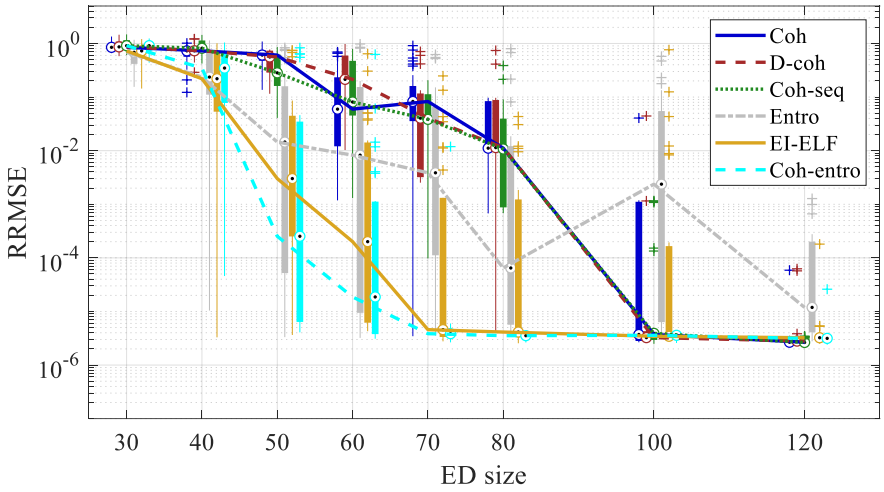
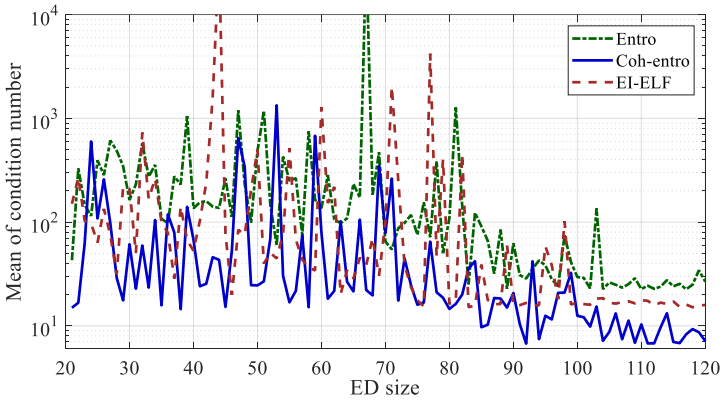
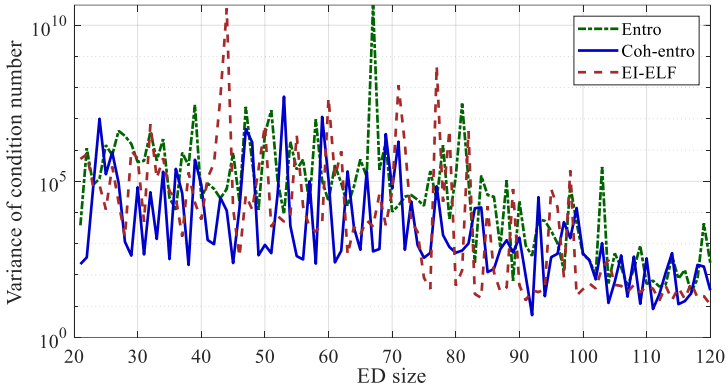



Figure 3.12 RRMSE with different ED size under degree 14

To provide an intuitive insight into the computational stability of the Entro method, the Coh-entro criterion and the EI-ELF method, the condition number of the regression matrix and the statistical properties (mean and variance) of the model mean and std values under different degrees are depicted in Figure 3.13 to Figure 3.18. It should be mentioned that we ignore some extremely large outliers in the statistical properties of the model mean and std values at the beginning of the iteration process, which have been hidden in the figures, to focus on their convergence conditions.

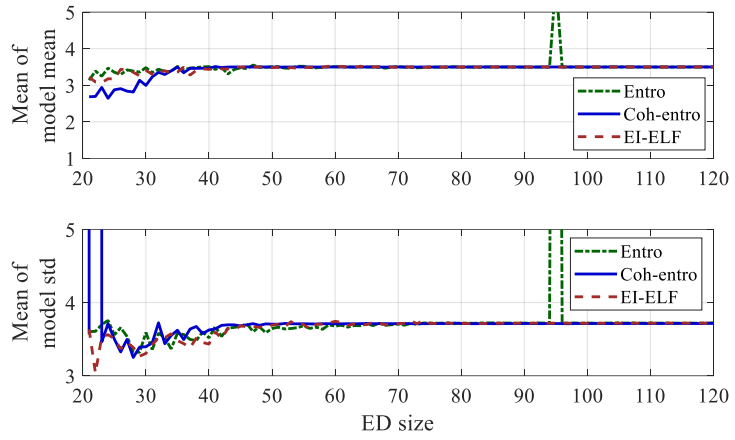



(a)

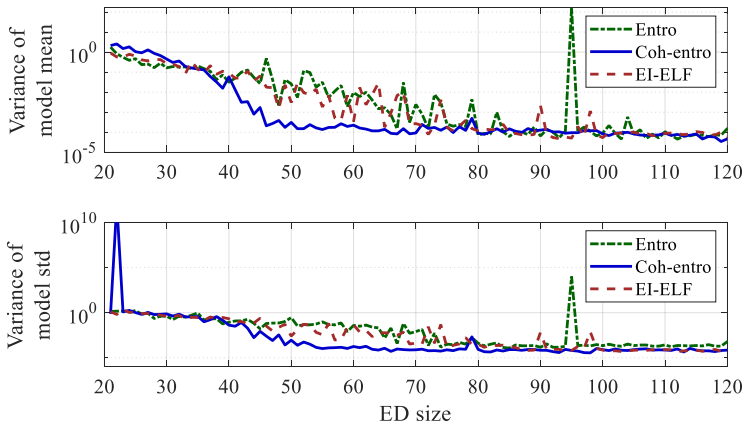



(b)

Figure 3.13 (a) Mean and (b) variance of the condition number with increasing ED size under degree 9

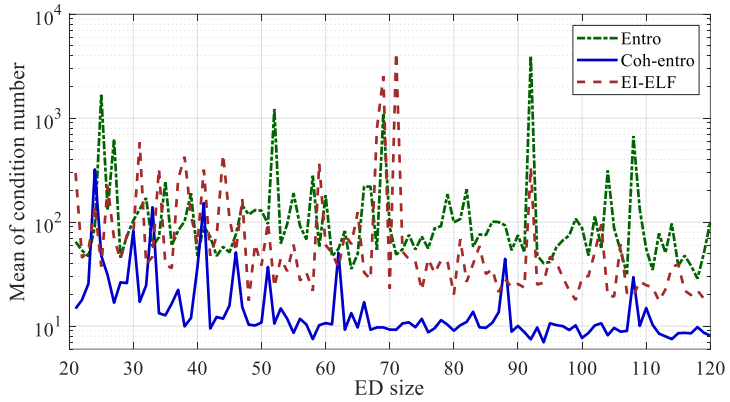



(a)

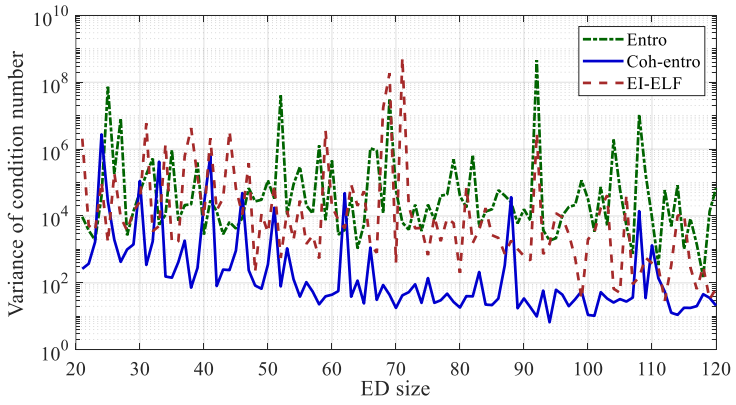



(b)

Figure 3.14 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under degree 9

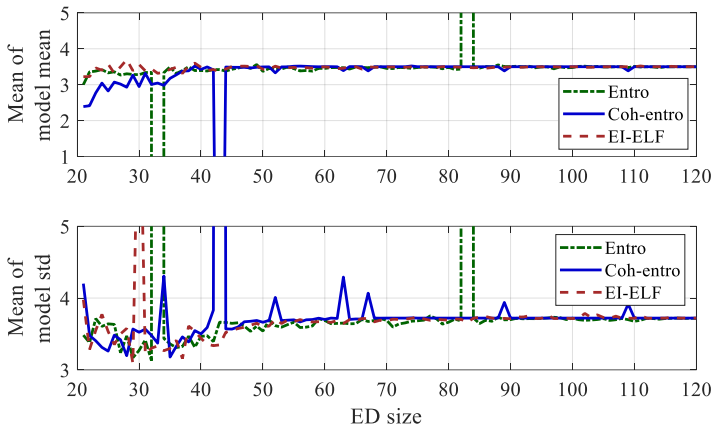



(a)

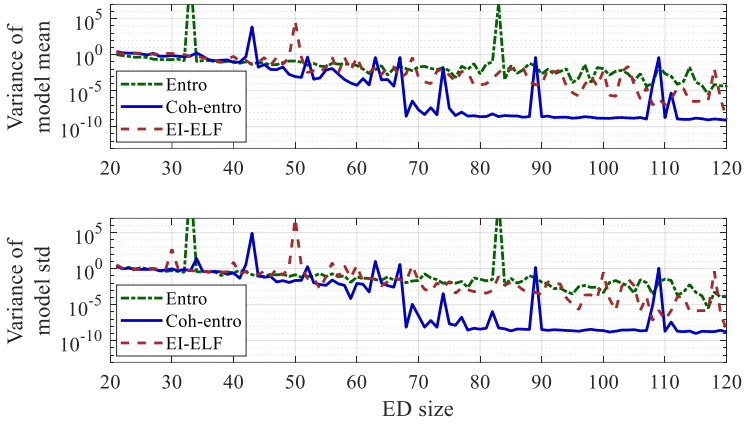



(b)

Figure 3.15 (a) Mean and (b) variance of the condition number with increasing ED size under degree 12

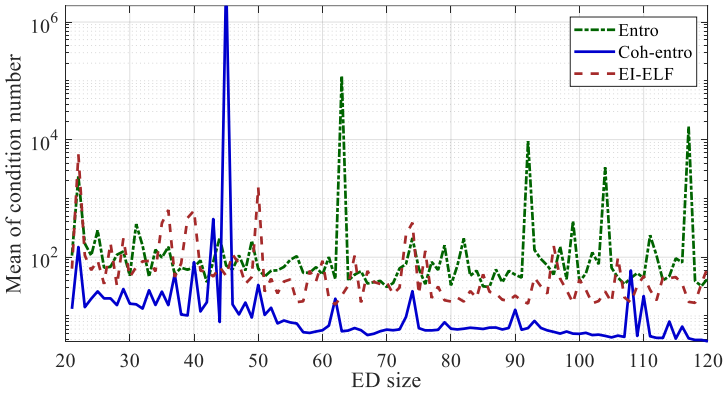



(a)

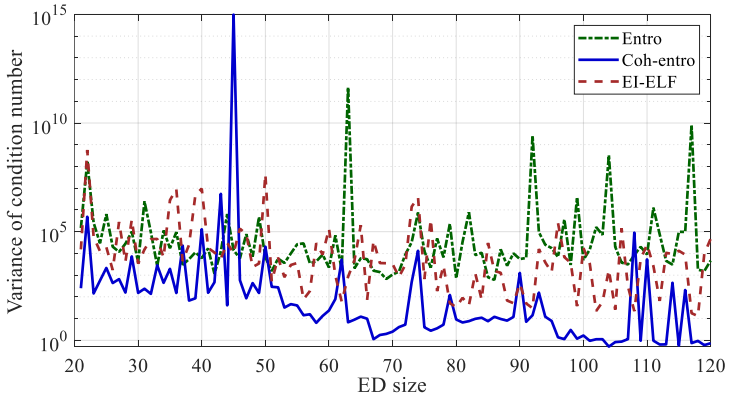



(b)

Figure 3.16 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under degree 12

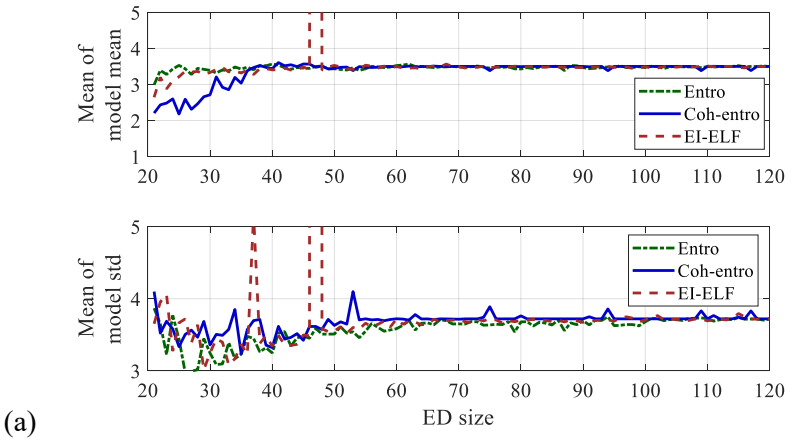



(a)

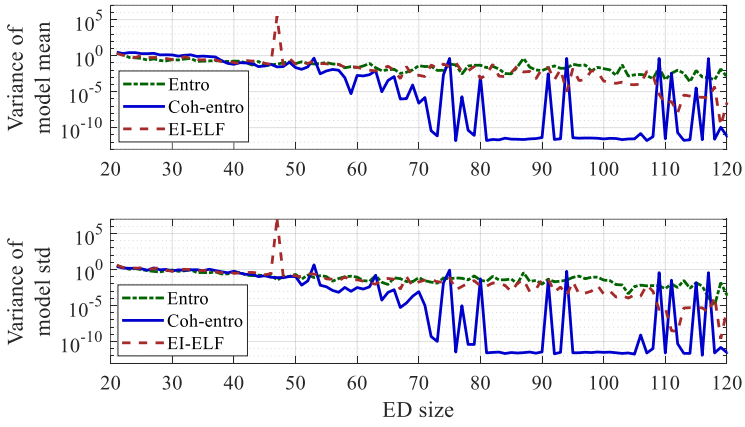



(b)

Figure 3.17 (a) Mean and (b) variance of the condition number with increasing ED size under degree 14




(a)




(b)

Figure 3.18 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under degree 14

As illustrated in Figure 3.13 under the PCE degree 9, the condition numbers of the EI-ELF criterion and the Coh-entro method exhibit similar trends with increasing ED, while the condition numbers of the Entro method show a bit larger values than these two methods. For the PCE models of degrees 12 and 14 as shown in Figure 3.15 and Figure 3.17, the condition numbers from the Coh-entro strategy mostly keep lower than those from the Entro method and the EI-ELF criterion, except for a few emerged peaks. These are influenced by a few extremely large condition numbers in the repeated tests. The variance of the condition numbers in these two cases shows that the Coh-entro strategy generates more stable and lower condition numbers than the Entro method and the EI-ELF

criterion in most instances, especially after convergence. In regard to the model mean and std under the PCE degree 9 in Figure 3.14, all the three methods make the PCE models converge to the target one. As illustrated by the variances of the model mean and std values, the Coh-entro strategy possesses a faster convergence performance than the Entro method and the EI-ELF criterion, and the three methods ultimately show similar trends. When the PCE degree is altered to 12, the Coh-entro strategy can still have a quick convergence rate than EI-ELF, and the variances of model mean and std from Coh-entro mostly keep at low values after convergence, as shown in Figure 3.16(b); whereas the Entro method cannot converge well. It is observed that a few apparent outliers emerge in Coh-entro. These are due to inaccurate model training results that occasionally arise since the iteration processes are repeated 30 times. The outliers from Coh-entro show large amplitudes in the model mean and std, so they can easily be distinguished from the converged training solutions during the iteration process. The same phenomenon can be observed in Figure 3.18 for the PCE degree 14.

The final samples of these three methods in one test under degree 14 are given in Figure 3.19 for comparison. It is seen that the samples from the Entro method still concentrate more on the edges of the input space, and the samples from the Coh-entro strategy are more space filling than those from the EI-ELF criterion.

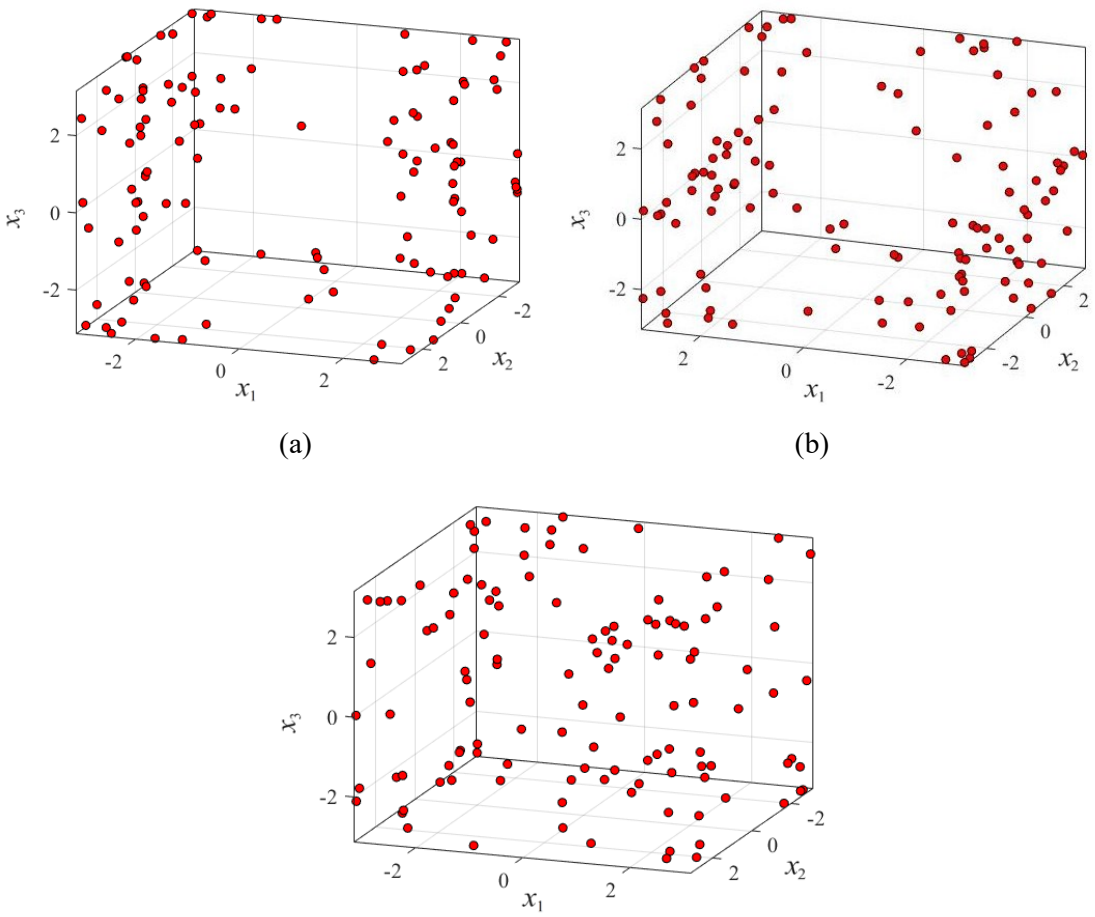



Figure 3.19 Final samples from (a) Entro; (b) EI-ELF; (c) Coh-entro

In summary, the Coh-entro strategy is found to provide samples that can form the regression matrix with low condition number. It can get more stable solution in the regression calculation and lead to faster convergence rate than the Entro method and the EI-ELF criterion. In most circumstances, the proposed method has better convergence stability.

### 3.4.3 High-dimensional function

The third benchmark test function is a high-dimensional function (Lüthen et al., 2021; Thapa et al., 2020). Its expression is as follows:

$$y = 3 - \frac{5}{d}\sum_{i=1}^{d} ix_i + \frac{1}{d}\sum_{i=1}^{d} i\,x_i^3 + ln\left[\frac{1}{3d}\sum_{i=1}^{d} i(x_i^2 + x_i^4)\right] \tag{3.39}$$

where $d$ is the input dimension chosen by user. All the input variables conform to uniform distributions defined on the interval $[1, 2]$. The input dimension is selected as $d$ = 15, and the total degree of the PCE model is chosen as 3, which results in polynomial items with $P$ = 816. The initial sampling number is set as 20. The operation is repeated for 30 times to ensure statistical stability.

The RRMSE results with the increase of ED size are shown in Figure 3.20. The output-oriented methods display similar convergence trends and perform better than the input-dependent only methods in terms of convergence rate. All six methods converge to similar validation errors, and the proposed Coh-entro strategy has the lowest validation error at the end of the sampling process.

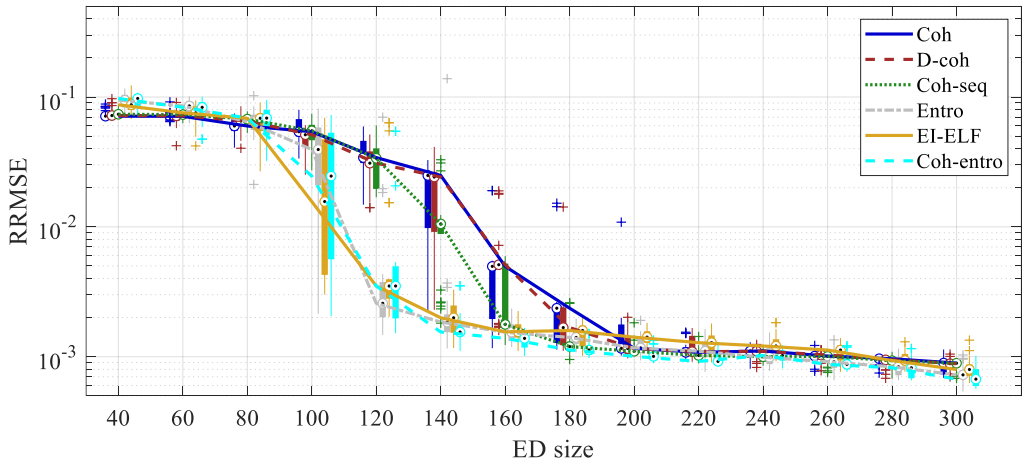
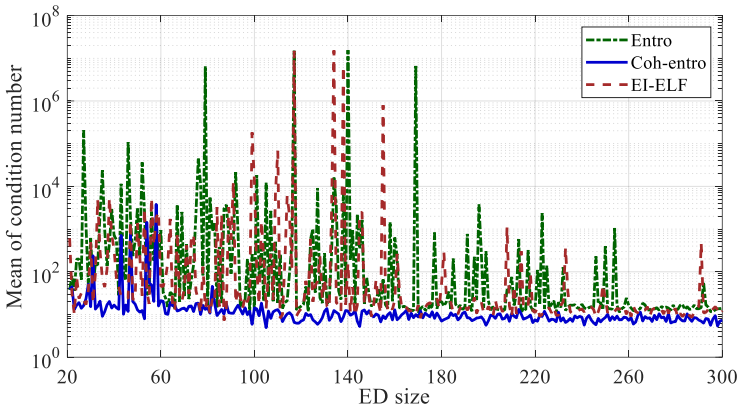
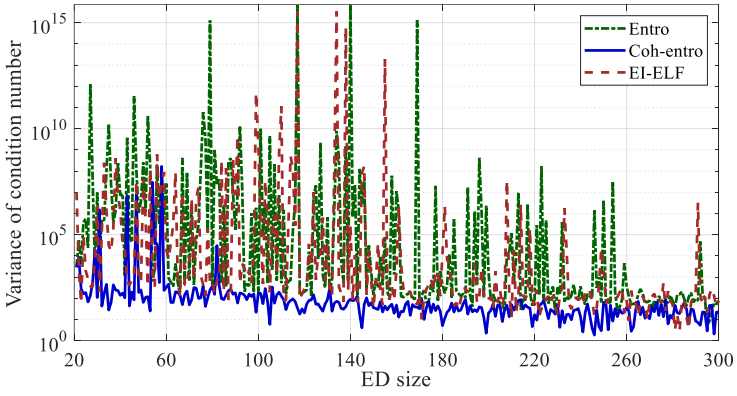



Figure 3.20 RRMSE with different ED size under input dimension 15

The condition number of the regression matrix and the obtained PCE model mean and std values of the three output-oriented sampling methods are depicted in Figure 3.21 and Figure 3.22.

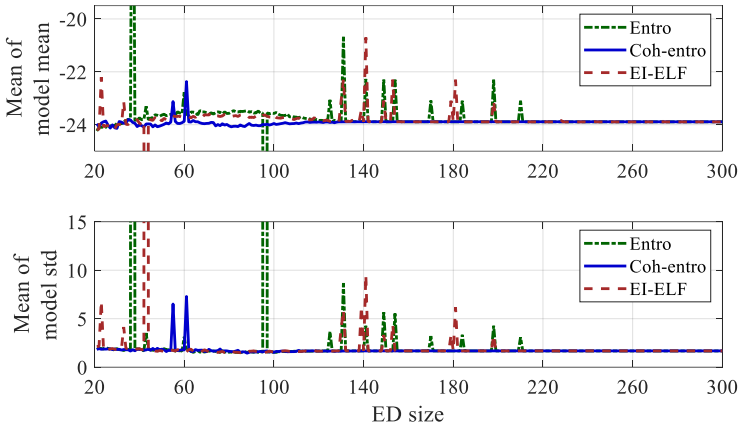



(a)

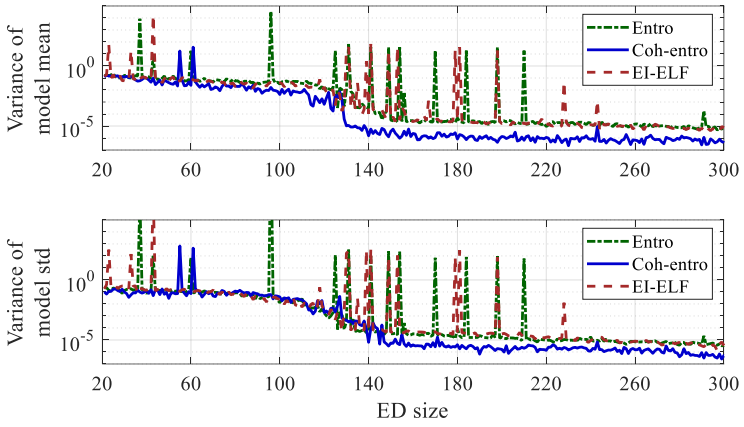



(b)

Figure 3.21 (a) Mean and (b) variance of the condition number with increasing ED size under input dimension 15




(a)




(b)

Figure 3.22 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under input dimension 15

It is apparent that the condition number from the Entro method and the EI-ELF criterion fluctuates more than the Coh-entro strategy during the whole iteration process, especially when the model has 100 to 140 samples. By contrast, the Coh-entro strategy has less extremely large values, which all concentrate at the beginning of the iteration process where ED size is less than 60. When the samples are more than 60, the condition number keeps stable at a very low value. The variance of condition number shows that the proposed strategy is robust in the repeated tests.

The solution stability of the regression calculation during the iteration process is demonstrated through the model mean and std in Figure 3.22. From the variance plot, we can see that all the three methods converge at around 180 samples, which is identical to that shown in Figure 3.20. After convergence, the PCE model from the Coh-entro strategy has no obvious outlier while the Entro method and the EI-ELF criterion show more fluctuations. Moreover, it can be found from the variances of the model mean and std that the Coh-entro strategy converges to a lower value of variance than the Entro method and the EI-ELF criterion, which indicates that the PCE model from

the Coh-entro strategy performs more stable than those from the Entro method and the EI-ELF criterion.

In addition, another input dimension $d = 20$ is tested with polynomial terms $P = 1771$ when the total truncation degree is chosen as 3. Due to the heavy calculation burden, the number of repeated tests in this case is set as 10. The initial sampling number is set as 20 which is the same as the previous case. The RRMSE results with the increase of ED size are shown in Figure 3.23. The three output-oriented methods outperform the input-dependent only methods, and both have similar convergence rate. In the end of iteration, however, the Entro method and the Coh-entro strategy have an equal precision with the D-coh and Coh-seq methods and performs a bit better than the Coh and EI-ELF methods.

The condition number of the regression matrix and the obtained PCE model mean and std values from the output-oriented methods are plotted in Figure 3.24 and Figure 3.25.

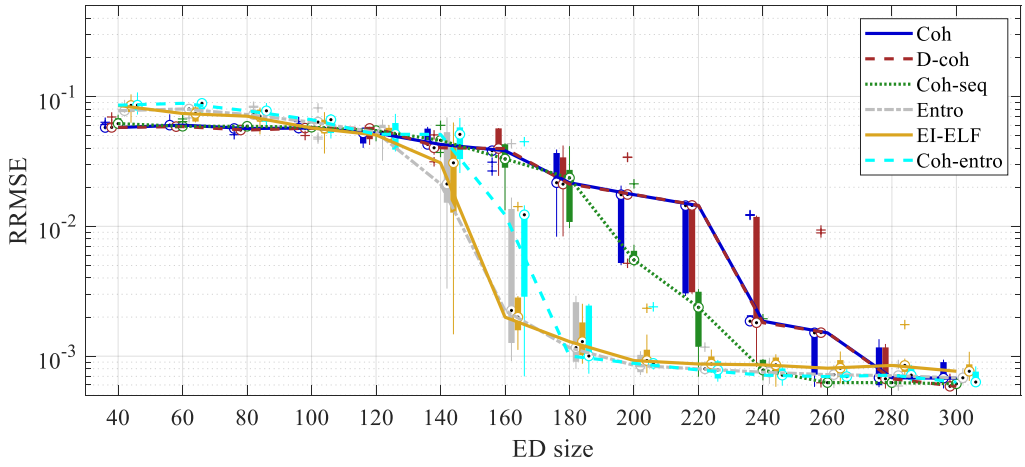
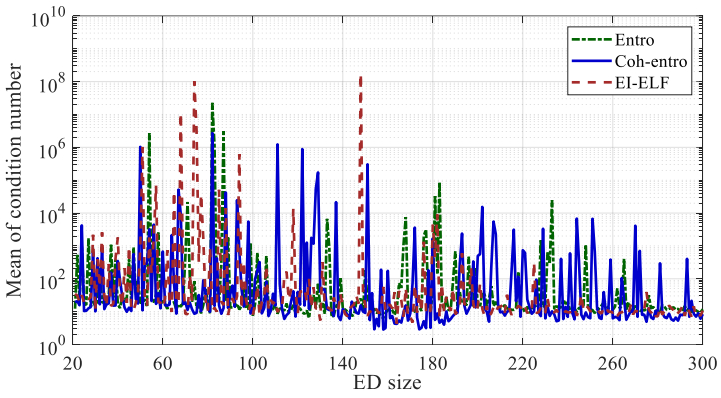
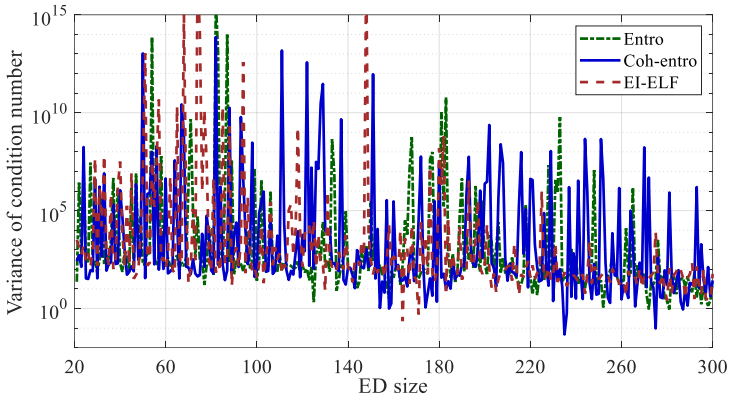



Figure 3.23 RRMSE with different ED size under input dimension 20

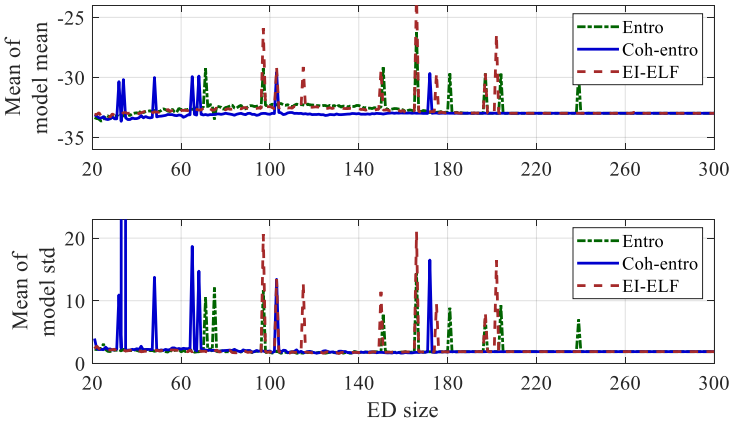



(a)

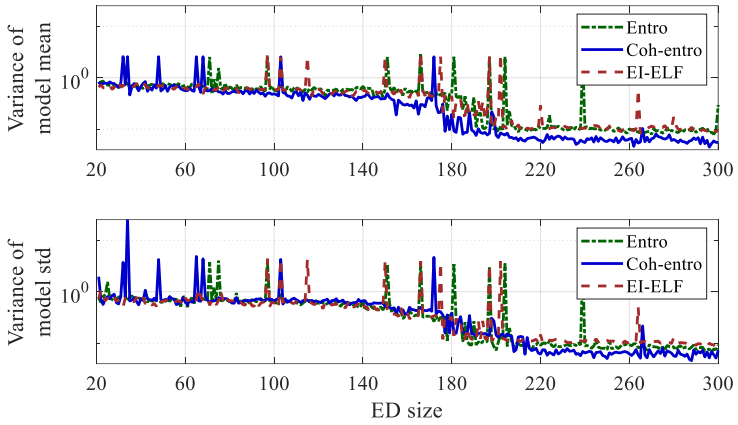



(b)

Figure 3.24 (a) Mean and (b) variance of the condition number with increasing ED size under input dimension 20




(a)




(b)

Figure 3.25 (a) Mean and (b) variance of the model means and standard deviations with increasing ED size under input dimension 20

It can be seen from the condition number that the three methods have similar performance. All of them give rise to large fluctuations during the iteration process. The reason might be that the number of samples is too small compared with the polynomial terms, making the calculation unstable with sparse representation. However, even with large fluctuation on the condition number, the convergence performance of Coh-entro is still a bit better than the Entro method and the EI-ELF criterion, as illustrated in Figure 3.25. After convergence, the Coh-entro strategy shows less outliers than the Entro method and the EI-ELF criterion. Moreover, the variances of the model mean and std obtained by the Coh-entro strategy finally converge to lower values than those from the Entro method and the EI-ELF criterion, demonstrating that the proposed method can help obtain more stable PCE models during the sampling process.

### 3.4.4 A large bridge structure

Since the PCE technique is generally applied in the engineering field for UQ and sensitivity

analysis, the parameter sensitivity of an extradosed cable-stayed bridge is analysed to verify the applicability of the proposed sampling method. Here, the ANalysis Of VAriance (ANOVA, or called Sobol' indices) method is used to analyse the parameter sensitivity in this case (Alış and Rabitz, 2001; Sobol′, 2001; Sudret, 2008). The introduction to the ANOVA method can be found in Appendix B.

In this case, the dynamic characteristics of the bridge structure with respect to structural material properties is modelled, and the sensitivity of the structural material parameters is analysed. The target structure is an extradosed cable-stayed bridge with three spans of 460 meters long in total. The girders are designed as single box with three rooms of 33.5 meters wide. Two towers of 40 meters high each are rigidly consolidated with the girders. The girders are continuously supported on piers. A finite element model of the superstructure is built as the target model without modelling the piers, which is displayed in Figure 3.26. The girders and towers use C55 concrete, and displacements at the second support counted from the left end are constrained.

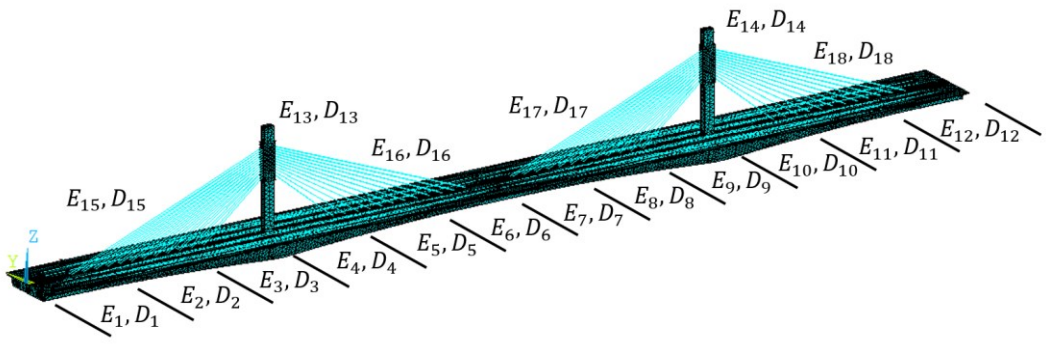



Figure 3.26 Finite element model and partitioned substructures

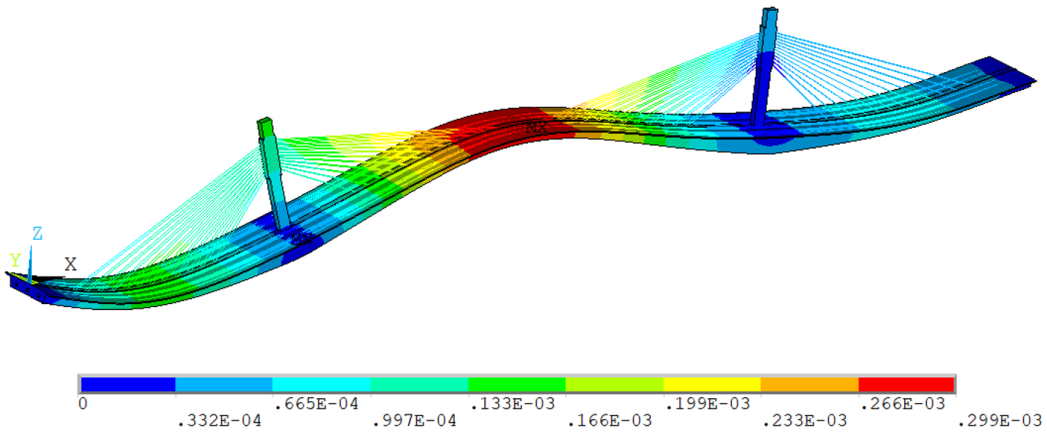



Figure 3.27 First vertical mode

Table 3-2 Input distributions of the extradosed cable-stayed bridge

| Variable | Distribution | Mean | Standard deviation |
|---|---|---|---|
| Elastic modulus $E_1 \sim E_{12}$ (girder) (Pa) | Lognormal | 3.55e10 | 3.55e9 |
| Elastic modulus $E_{13}, E_{14}$ (tower) (Pa) | Lognormal | 3.55e10 | 3.55e9 |
| Elastic modulus $E_{15} \sim E_{18}$ (cable) (Pa) | Lognormal | 1.95e11 | 1.95e10 |
| Density $D_1 \sim D_{12}$ (girder) (kg/m$^3$) | Weibull | 2549 | 254.9 |
| Density $D_{13}, D_{14}$ (tower) (kg/m$^3$) | Weibull | 2549 | 254.9 |
| Density $D_{15} \sim D_{18}$ (cable) (kg/m$^3$) | Weibull | 8005 | 800.5 |

Our target on the bridge dynamic characteristics is the first vertical natural frequency. The corresponding mode shape is shown in Figure 3.27. To identify the impact of the structure material properties on the target characteristic, each side span of the girder is divided into three equal-length regions and the middle span is divided into six equal-length regions. Moreover, the cables on the same side of each tower are grouped into one group, as shown in Figure 3.26. Together with the two towers, eighteen portions (substructures) are considered. The elastic moduli and densities of these eighteen substructures, 36 input variables in total, are taken as input variables in this study. Distributions of these variables are reported in Table 3-2, and they are assumed to be mutually independent. The mean values of these variables are their nominal values, and the variance is set as

0.1 coefficient of variation, which is defined as the ratio of the std to the mean (Wan et al., 2020). The input variables are transformed into standard Gaussian variables to build a Hermite PCE model. The total degree to truncate the PCE model is chosen as 3 to ensure the calculation accuracy, and this will generate a PCE model with $P = 9139$ terms. The initial sampling number is set as 20. The threshold to terminate the sampling process is set as $10^{-5}$. Finally, a total of 192 samples are obtained to attain the precision. The obtained PCE model is then used to quantify the parameter sensitivity by calculating the ANOVA indices. Moreover, due to the heavy calculation burden in generating massive MC simulations in this case, the reference solution of the ANOVA indices is obtained by building a sparse PCE model trained with 5000 LHS samples for comparison. The first-order indices and the total indices of the 36 variables are shown in Figure 3.28 and Figure 3.29, in which $S_F$ and $S_T$ represent the ANOVA indices calculated with samples collected by the proposed Coh-entro strategy; $S_{F_{ref}}$ and $S_{T_{ref}}$ denote the reference values of the ANOVA indices.

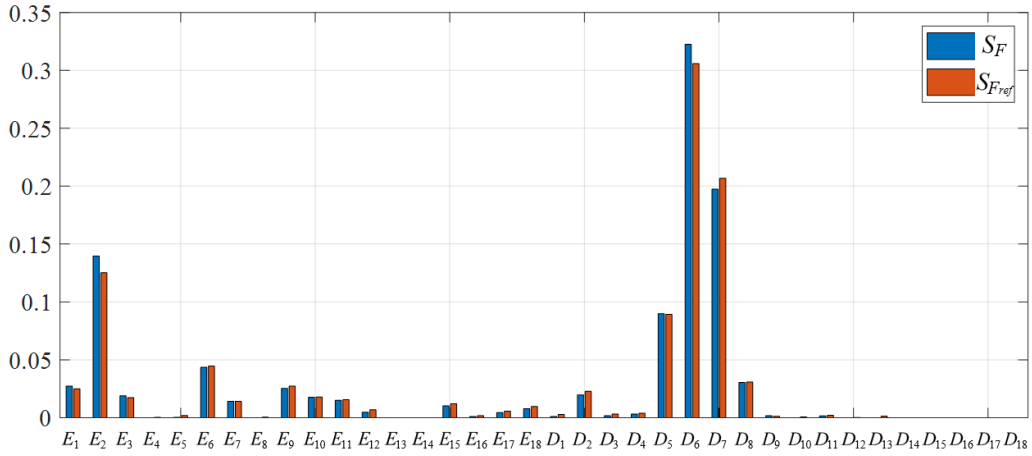



Figure 3.28 First-order indices of input variables of the extradosed cable-stayed bridge

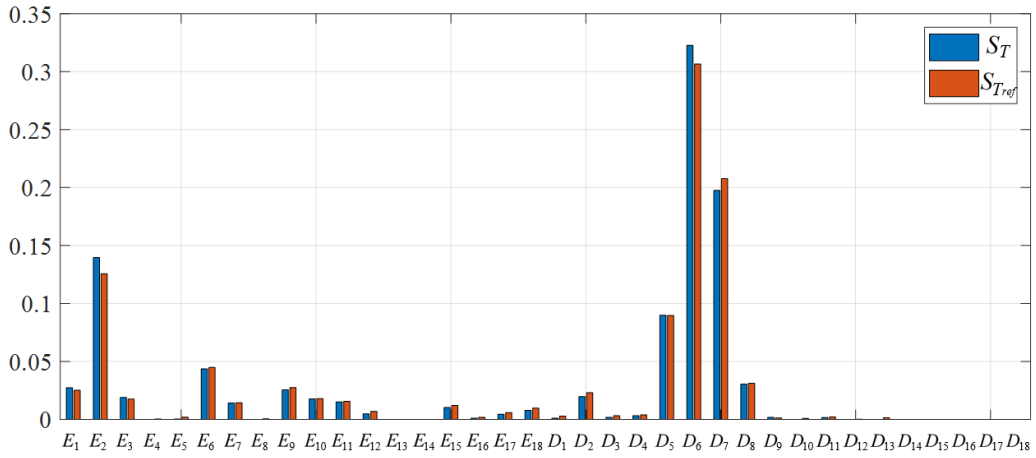



Figure 3.29 Total indices of input variables of the extradosed cable-stayed bridge

From the results of sensitivity analysis, it can be concluded that the ANOVA indices calculated with the proposed Coh-entro strategy have very tiny discrepancy compared with the reference solutions, which demonstrates that the PCE model trained with 192 samples collected by Coh-entro is reliable for sensitivity analysis. In 36 input variables, the material properties of the two towers, $E_{13}, E_{14}$ and $D_{13}, D_{14}$, are insensitive to the first vertical nature frequency, which is in line with our knowledge that the towers have less influence on the girder vibration. Moreover, the densities of the stay cables, $D_{15}$~$D_{18}$, are insensitive to the first vertical nature frequency; and the elastic moduli of the cables, $E_{15}$~$E_{18}$, are also in low sensitivity. $E_2$ and $D_5$~$D_7$ are the most sensitive parameters, and $E_1, E_3, E_6, E_7, E_9, E_{10}, E_{11}$ and $D_2, D_8$ have the second most influence on the frequency. The remaining parameters have extremely small sensitivity values, which means that these parameters are less influential. Moreover, it is worth noting that the first-order indices and the total indices merely have tiny discrepancy, which implies that there are slight interaction effects among these parameters on the first vertical natural frequency.

## 3.5 Summary

In this chapter, a new sampling theory for PCE modelling is developed, which proposes to take advantage of both the input information and model feature to instruct a sequential sampling process. Based on the new sampling theory, a novel sequential sampling method termed coherence-entropy strategy that comprises of two popular sampling approaches, coherence-optimal sampling and BED, is proposed. The BCS is employed as a sparse regression procedure to calculate the unknown coefficients associated with a simple representation of PCE model, which also provides the foundation for BED. The input information is first utilised to form a coherence-optimal distribution in line with compressive sampling theory. Sampling in this distribution ensures a lower bound on sample quantity for accurate recovery of the PCE coefficients. In order to build a sequential sampling framework, the LHS method is instead employed to collect samples from the quasi-coherence-optimal distribution, which ignores the correlation among input variables in the coherence-optimal distribution. The output information in the modelling results is identified by using the differential entropy, and it is leveraged by BED, which is encompassed in the sequential sampling framework, to expedite convergence during the iteration process. The BED also benefits sampling in the quasi-coherence-optimal distribution at the beginning of iteration process in that the collected samples will always have an impressive space filling property.

To validate the proposed method, three analytical functions with different complexity were studied, and the results obtained by the proposed method were compared with those from three

input-dependent only methods and two output-oriented methods, in which the Coh-Opt method and the Entro method are the components of the proposed method. It is shown that the proposed approach and the output-oriented methods generally outperform the input-dependent only methods in convergence rate, and meanwhile the modelling results have good accuracy after convergence. The proposed method generally has the fastest convergence rate. Among the output-oriented methods, the proposed strategy draws samples in a more space filling way than the Entro method and the EI-ELF criterion, and the obtained PCE models are in more stable convergence performance as shown in the repeated tests. For the problems with relatively less unknown PCE coefficients, the proposed strategy gives rise to the condition number of the regression matrix in a low value, thereby performing better than the Entro method and the EI-ELF criterion. For the problems with the number of unknown coefficients much higher than the required ED size, the condition number is hardly kept at a low value, but the proposed sampling strategy still ensures a fast convergence rate and a high PCE modelling accuracy after convergence. Also, it is observed that the proposed method has a faster convergence rate than the Coh-Opt method and gets more stable modelling results than the Entro method, which demonstrates that the combination of BED and coherence-optimal sampling could help improve both methods.

# CHAPTER 4

# ADAPTIVE SPARSE PCE WITH BASIS ADAPTIVITY AND SEQUENTIAL SAMPLING

## 4.1 Introduction

In the previous chapter, a sequential sampling strategy which leverages both input and output information has been exploited. Based on this concept, samples are collected sequentially with high quality and in relatively small quantity. Besides the sampling problem, the model selection problem is also critical in getting precise modelling results (Tan, 2015). Basically, the model selection problem in PCE is to determine a proper truncation degree (Blatman and Sudret, 2011), or in a more complicated way, to choose significant basis terms (Ni et al., 2017). A PCE model is generally truncated with a given polynomial degree before training so that only a finite number of coefficients need to be solved. However, it is hard to determine an appropriate truncation degree in advance when little information about the target problem is available. To address this, adaptive basis selection strategies were proposed, which aim to adaptively find the significant basis terms with low cost (less samples) (Jakeman et al., 2015; Lüthen et al., 2020).

As is introduced in Chapter 2, the commonly used adaptive basis selection strategy is: add

basis terms of higher degree to the current PCE model iteratively for training while removing insignificant basis terms until the required modelling accuracy is achieved (Blatman and Sudret, 2011; Jakeman et al., 2015; Loukrezis et al., 2020; Thapa et al., 2020). If those basis terms which are not decisive in modelling the target problem are removed and only the significant basis terms are retained, a precise PCE model can be trained with low training cost even if a high truncation degree is used. Although many methods were developed to solve the adaptive basis selection problem, few studies proposed to solve it and the sequential sampling problem in a concurrent way. Generally, the truncation degree was assigned with a given value when solving the sampling problem, and sufficient samples were provided in addressing the basis selection problem. To integrate the sequential sampling and adaptive basis selection in one framework, the core is to determine whether to add basis terms or samples in each iteration since either insufficient training data (samples and observations) or inadequate basis terms will lead to inaccurate PCE modelling result. In the past, least-squares methods were employed for regression calculation, so the sample quantity requirement, i.e., the sample quantity was required to be more than the number of coefficients and preferably twice more than the number of coefficients (Hosder et al., 2007), was employed to settle this problem. When the sample quantity is less than the number of coefficients, the samples have to be enriched. When the sample quantity is twice more than the number of coefficients and the modelling result does not have the desired accuracy, more basis terms with higher degrees would be added (Thapa et al., 2020). Nevertheless, the efficiency of these strategies

is limited due to the “curse of dimensionality” issue. To author’s knowledge, there is a paucity of research on the sparse representation in PCE modelling with adaptive basis selection and sequential sampling strategies.

In this chapter, a method of Adaptive Sparse PCE with Basis Adaptivity and Sequential Sampling (ASPCE-BASS) is proposed to fill in this gap, in which the BCS method is introduced for sparse regression. A novel adaptive basis selection strategy is developed, which involves three kernel operations, basis expansion, basis pruning and basis refinement. The basis expansion operation adds in the current PCE model with more basis terms of higher degree; the basis pruning operation removes the insignificant basis terms from the current PCE model with a view to expedite computation speed and efficiency; and the basis refinement operation is a remedy process, which can resume the significant basis terms that were incorrectly expunged in history. This new basis selection strategy can not only add basis terms with different degrees to the PCE model, but also remove the insignificant basis terms to speed up the algorithm and thus reduce the demand on sample quantity. To combine with the basis selection strategy, a hybrid sequential sampling method that distinct from the coherence-entropy strategy, which also capitalises on the input and output information in PCE, is introduced. A two-loop iteration framework is adopted, in which the inner loop is the sequential sampling process, and the outer loop is the adaptive basis selection process. Moreover, a stability evaluation method is presented to determine which loop should be performed in each iteration. To illustrate the efficiency and effectiveness, the proposed ASPCE-BASS

algorithm will be compared with another adaptive PCE modelling algorithm and a non-basis-adaptive algorithm (sequential sampling-only algorithm) by validating on three analytical functions with different complexities.

## 4.2 Method

In pursuit of adaptive basis selection for PCE, a practical way is embarking on a simple PCE model with a small truncation degree and gradually adding bases with higher degree until the precision of the trained PCE model meets predefined requirement. When taking sequential sampling into account as well, the problem becomes more challenging: it starts with a simple PCE model with a small truncation degree and a small set of samples, and then proceeds with alternately adding samples and basis terms until the PCE model is trained with desired accuracy. Notice that either insufficient training dataset or inadequate basis terms would lead to inaccurate PCE modelling results. The key is to determine when more samples are needed and when more bases should be added. As mentioned before, this issue can be solved by the least-squares method in the existence of massive samples and observations. Rather, we aim to build an accurate PCE model with high truncation degree by use of as few samples as possible. Through analysing the intrinsic nature of PCE modelling in consideration of both basis adaptivity and sequential sampling, four issues are found to affect the sample quantity required in PCE modelling to a specific problem.

(1) Sparse representation: The least-squares method has strict requirements on the sample quantity for PCE training. The sample quantity should be at least more than the number of unknown

coefficients and 2~3 times oversampling rate is recommended. The sparse property of the target problem is ignored. For problems that can be sparsely represented by PCE model, much fewer samples are needed for sparse representation than the least-squares method. Since $l_1$ minimisation is employed for regression calculation in this study, we focus our discussions on the $l_1$ minimisation problem hereinafter.

(2) Sample quality: Intuitively, the sample quality would significantly affect the solution quality when the sample quantity is restricted. If the samples are collected with high quality, the number of samples required for accurate PCE modelling can be reduced. The sequential sampling strategy proposed in the previous chapter has addressed this issue.

(3) The highest polynomial degree required by the target model: It has been shown in Equation (3.7) that the cardinality of a PCE model is controlled by the input dimension and truncation degree. If the target problem seeks to be represented by a PCE model with high truncation degree, usually a large number of unknown coefficients should be solved.

(4) Sparsity of the target problem: The sparsity $K$ is defined as the number of non-zero coefficients in a trained PCE model. For two problems having the same highest polynomial degree but disparate sparsity when expanding on PCE model, the one with larger $K$ generally requires more samples for training than the other one if sparse representation is employed for regression calculation (Hampton and Doostan, 2015).

Among the above issues, (3) and (4) are the inherent characteristics of a problem, which are neither known in advance nor can be changed. As shown in the previous studies (Loukrezis et al., 2020; Hampton and Doostan, 2015), the lower bound of sample quantity required to get accurate solution by use of $l_1$ minimisation is controlled by the cardinality $P$. If insignificant basis terms are removed from the current PCE model in each iteration, the number of unknown coefficients in the trained PCE model, namely cardinality $P$, will reduce; in parallel the lower bound of samples required for accurate PCE training will decrease. Some efforts in this research line have recently been made with promising results achieved (Lüthen et al., 2020; Zhao et al., 2019). In short, both sparse representation and the above strategy play an important role in reducing the sample quantity, which in combination are expected to generate better results. In this regard, a new hybrid sequential sampling method, which can be combined with the adaptive basis selection operation easier than the coherence-entropy method, is first introduced. Then, an adaptive basis selection strategy which involves three key operations, namely basis expansion, basis pruning and basis refinement, will be developed in this section. Finally, these two strategies will be combined with the sparse representation to build a fully adaptive PCE framework, which is expected to afford the dual merit of fast convergence rate and high computational accuracy.

### 4.2.1 A new hybrid sequential sampling method

As is shown in the previous chapter, the coherence-entropy sampling method requires to

sample from the quasi-coherence-optimal distribution which is obtained from the input information including the input distribution and truncated polynomial basis terms. When the adaptive basis selection operation is concurrently applied, the number of basis terms in the PCE model will vary during the iteration process, thereby changing the quasi-coherence-optimal distribution. Evaluating the distribution each time as the basis terms change is a heavy burden. To lighten this burden, another hybrid sequential sampling method which combines the EI-ELF criterion with LHS method is introduced.

It can be seen from Section 3.4 that the modelling results by using EI-ELF criterion has similar convergence rate and accuracy with coherence-entropy method, but the samples are more likely to concentrate on the edge of the input space. Therefore, the PLHS method is introduced to restrict the sampling area in each iteration of EI-ELF so that the distribution property of the collected samples can be improved. This algorithm is called EI-PLHS method, and the detailed steps of this algorithm are shown in **Algorithm 3**.

In summary, this algorithm leverages the LHS method to improve the distribution quality of samples that collected by the EI-ELF criterion. The samples can be collected in a more space filling way even if a small number of samples are collected, and the convergence rate of EI-ELF is expected to be retained in the new strategy. Compared to the coherence-entropy algorithm, the burden of evaluating and updating the coherence-optimal distribution is eliminated, so the EI-PLHS algorithm is more practical to be combined in the adaptive PCE modelling framework.

**Algorithm 3.** EI-PLHS algorithm

**Input:** Initial sampling number $N_0$; total sampling number $N_{ts}$ or desired accuracy $\varepsilon$; candidate set $\boldsymbol{\mathcal{X}}_{N_t}$; selection set $S$ and observation set $Y$.

**Initialisation:**

1. Generate $N_0$ initial samples which yield the input distribution by LHS method and add them into the selection set $S = \boldsymbol{\mathcal{X}}_{N_0}$;
2. Get model observations at sampled points in the selection set $Y = \boldsymbol{Y}(\boldsymbol{\mathcal{X}}_{N_0})$.

**In the $i$th iteration:**

1. Build PCE model according to the selection set $S$ and the corresponding observations $Y$ by employing the BCS method;
2. If the total sampling number $N_{ts}$ is reached or the desired accuracy $\varepsilon$ is achieved, quit iteration; otherwise, get into step 3;
3. Check all the blocks; If no more active block exists, get into step 4; otherwise get into step 5;
4. Under the current sample number $N_S$, re-divide the input space into $2N_S$ slices in each dimension. $N_S^d$ new active blocks are generated;
5. Abandon the candidate points which do not belong to active blocks temporarily;
6. Make prediction to the remained candidate points and calculate the corresponding $ELF_{EI}$;
7. Add the candidate point $\boldsymbol{\xi}^{(i)}$ which owns the largest $ELF_{EI}$ value to the selection set $S = S \cup \boldsymbol{\xi}^{(i)}$, and get model observation at this point $Y = Y \cup \boldsymbol{Y}_{\boldsymbol{\xi}^{(i)}}$;
8. According to the current sample set, re-evaluate the active blocks;
9. Update the candidate set by deleting the selected point $\boldsymbol{\mathcal{X}}_{N_t} = \boldsymbol{\mathcal{X}}_{N_t} \backslash \boldsymbol{\xi}^{(i)}$.

### 4.2.2 Basis expansion strategy

Two issues are involved in the stage of basis expansion: (i) when should the current basis terms in the PCE model be expanded? and (ii) how many extra basis terms are appropriate to be added? On the first issue, the sample quantity requirement in the least-squares method cannot be leveraged when the sparse representation approach is applied. Hence, a new stability evaluation method is proposed to address this issue. After doing so, the number of extra basis terms to be added is recommended.

#### 4.2.2.1 Stability evaluation method

Assume that the problem of interest can be completely represented by PCE with degree $p$, which is:

$$y(\boldsymbol{\xi}) = \sum_{i=1}^{P} c_i \, \psi_i(\boldsymbol{\xi}) \tag{4.1}$$

where $P$ is the cardinality of PCE model with degree $p$; $c_i$ is the $i^{th}$ true coefficient which is unknown. Now the initial PCE degree and sample number are nominated as $p_0$ and $N_0$, respectively. Thus, the target problem can be expressed by the current PCE model as follows:

$$y(\boldsymbol{\xi}) = \sum_{i=1}^{P_0} c_i \, \psi_i(\boldsymbol{\xi}) + \varepsilon \tag{4.2}$$

wherein $P_0$ is the cardinality of the current PCE model with degree $p_0$; $\varepsilon$ represents the truncation error. Here the truncation error can be expressed as:

$$\varepsilon = \sum_{i=P_0+1}^{P} c_i \, \psi_i(\boldsymbol{\xi}) \tag{4.3}$$

Then the initial samples and corresponding observations could be used as training data to estimate the coefficients in Equation (4.2) via BCS. It is in general unknown whether the samples and the corresponding observations are enough or not for training. Without loss of generality, $N_0$ samples are assumed not enough, that is to say, these samples and observations cannot fully characterise the target model. By calculating the coefficients, we can write the mean function of the trained PCE

model as:

$$Y_{pre} = \sum_{i=1}^{P_0} \mu_{c_i} \psi_i(\boldsymbol{\xi}) \tag{4.4}$$

where $\mu_{c_i}$ is the $i^{th}$ value of results calculated by Equation (3.22). The regression error of the trained PCE model can be represented as:

$$\delta = Y_{pre} - y = \sum_{i=1}^{P_0} \mu_{c_i} \psi_i(\boldsymbol{\xi}) - \sum_{i=1}^{P_0} c_i \psi_i(\boldsymbol{\xi}) - \varepsilon \tag{4.5}$$

Due to the existence of truncation error, the obtained coefficients mean value $\mu_{c_i}$ will not be equal to the true value $c_i$ even if sufficient samples are used, because the features of the target problem which should be expressed by polynomial terms of high degree may be incorrectly expressed on the polynomial terms of low degree (Jakeman et al., 2015). Therefore, Equation (4.5) can be re-written as following:

$$\begin{aligned} \delta &= \sum_{i=1}^{P_0} \mu_{c_i} \psi_i(\boldsymbol{\xi}) - \sum_{i=1}^{P_0} c_i \psi_i(\boldsymbol{\xi}) - \varepsilon \\ &= \left( \sum_{i=1}^{P_0} \mu_{c_i} \psi_i(\boldsymbol{\xi}) - \sum_{i=1}^{P_0} \tilde{\mu}_{c_i} \psi_i(\boldsymbol{\xi}) \right) \\ &+ \left( \sum_{i=1}^{P_0} \tilde{\mu}_{c_i} \psi_i(\boldsymbol{\xi}) - \sum_{i=1}^{P_0} c_i \psi_i(\boldsymbol{\xi}) \right) + (-\varepsilon) = \delta_t + \delta_{\varepsilon 1} + \delta_{\varepsilon 2} \end{aligned} \tag{4.6}$$

in which $\tilde{\mu}_{c_i}$ denotes the mean value of the $i^{th}$ estimated coefficient when sufficient training data

are afforded. $\delta_t$ implies the regression error due to insufficient training data. $\delta_{\varepsilon 1}$ denotes the error aroused by incorrect expression due to truncation, and $\delta_{\varepsilon 2}$ represents the truncation error excluded in $\delta_{\varepsilon 1}$. By increasing the samples through the sequential sampling strategy without changing the basis terms, the features of the target problem will be gradually discovered in the corresponding observations. As a result, the regression error $\delta_t$ will reduce while $\delta_{\varepsilon 1}$ and $\delta_{\varepsilon 2}$ keep at their values. We can infer that new basis terms with higher polynomial degree can be added to the current PCE model if the regression error $\delta$ does not decrease with the increased samples.

Since the real model response $y$ is unknown, it is hard to directly evaluate the regression error $\delta$. Apparently, the variance of the real model response $y$ is a constant and independent of the regression errors $\delta$. Transforming Equation (4.6) into (4.7) and taking variance to both sides yield Equation (4.8).

$$Y_{pre} = y + \delta_t + \delta_{\varepsilon 1} + \delta_{\varepsilon 2} \tag{4.7}$$

$$Var(Y_{pre}) = Var(y) + Var(\delta_t + \delta_{\varepsilon 1} + \delta_{\varepsilon 2}) \tag{4.8}$$

With the increase of samples, the variance of the trained PCE model will approach the variance of the real model responses due to the decrease of $\delta_t$, but it will deviate from the real variance value to a certain extent because of the existence of $\delta_{\varepsilon 1}$ and $\delta_{\varepsilon 2}$. The variance can be changed as standard deviation (std) as well. The stability of modelling results can be leveraged to help determine when more basis terms can be added. To evaluate the fluctuation of the model variance (or std), we take

variance to both sides of Equation (4.8):

$$Var\left(Var\left(Y_{pre}\right)\right) = Var(Var(\delta)) \tag{4.9}$$

Hence, the variance of the model output variance (or std) is related only to the regression error. To enable evaluating the variance of the model output variance (or std) under the sequential sampling framework, a moving-window strategy is introduced. An illustration example is shown in Figure 4.1. The top plot shows the change of model std with the increase of samples. In each iteration with one new sample added, the PCE model is trained with the updated training dataset. A window of length $L_a$ ($L_a$ samples) is defined from the beginning of the sequential sampling process (red rectangle in the top plot). The variance of the model std in the window is calculated. When a new sample is drawn by the sequential sampling strategy, the window moves a step forward (green rectangle), and the variance of the model std in this new window is calculated. The variances alternation with respect to the window moving is recorded in the bottom plot. When the regression error of the trained model cannot decrease with the increase of samples, the variance of the model std will hold at a value with small fluctuation, as depicted in the bottom plot of Figure 4.1.

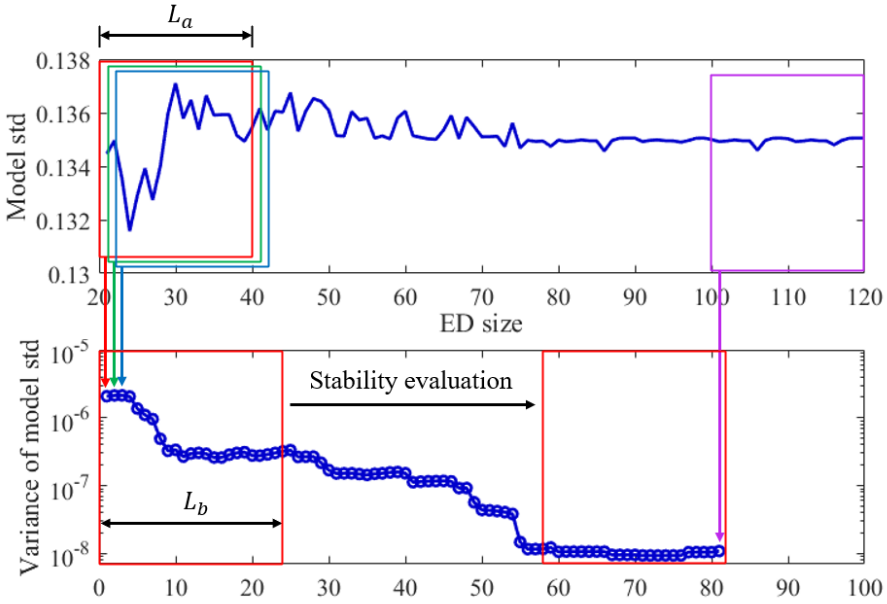



Figure 4.1 Stability evaluation process

Subsequently, we can evaluate the stability of the trained PCE models according to the bottom plot. This is an open-mind issue. In this study, a real-time stability evaluation method is presented. As shown in the red rectangles in the bottom plot, a stability evaluation range of length $L_b$ is introduced from the beginning and it moves forward with the increased data points. When the window in the top plot moves $L_b - 1$ steps from the starting position, $L_b$ variance values will be recorded in the bottom plot and fill up the range. The minimum value in this range is then recorded. With a new added sample, the range will move one step forward by including the new variance value and excluding the first value in this range. The newly included variance value will then be compared to the recorded minimum value. If the new value is smaller than the recorded value and the decrement exceeds a threshold, it can be deemed that the regression error $\delta_t$ has a nonnegligible reduction with the added sample, and the recorded value should be replaced by the new value. Here,

the threshold is used to speed up the evaluation process, which is recommended as one-tenth of the minimum variance value in the current range. If the new value is not smaller than the recorded value or the decrement does not exceed the threshold, the sequential sampling process will step into the next iteration. Once the recorded minimum value becomes the first value in the range, it is considered that the variance of model std cannot have an obvious decrease with the increased $L_b - 1$ samples, and in this regard the PCE model should be expanded with more bases. In summary, if the variance of model std is keeping near a low value with very small fluctuations with the increase of $L_b - 1$ samples, the PCE model with the current basis terms is deemed to have a stable training result. Then, more bases should be added.

There are several points of caution in conducting the stability evaluation: (1) Inaccurate regression solutions may occasionally emerge during the sequential sampling process, which may give rise to outliers in model variance (or std) and have adverse impact to the stability evaluation. Such outliers should be excluded before the stability evaluation. (2) The precision of the trained PCE model may temporarily degenerate when adding too many new bases. At the juncture just after new bases have just been added, the variance values may ascend with the increased samples. In such a case, the stability evaluation should cease until the variances are in downtrend again. (3) If the sample quantity exceeds the number of basis terms, more basis terms can be added.

#### 4.2.2.2 Basis expansion

Premised upon the proposed stability evaluation method, the time point to carry out basis expansion operation can be determined. The other issue is regarding how many basis terms of higher degree need to be added to the current PCE model. The previous studies have shown that the sparse representation approach could work even if the unknown coefficients are more than the training samples, but naively adding too many basis terms of high degree might result in the degradation of modelling accuracy (Jakeman et al., 2015). Jakeman et al. (2015) clarified that the sparse representation via $l_1$ minimisation can ensure stable computation when the quantity of the unknown coefficients is no more than 10 times the number of samples. Therefore, from the current PCE model with degree $p$, basis terms with polynomial degrees $p+1,\ p+2, \ldots$ will be added to the current PCE model in turn until the total number of basis terms in the PCE model is 10 times the current sample quantity. If all the basis terms of the same degree cannot be added to the PCE model at once due to the quantity limitation, basis terms with low interaction effect have the priority to be added than those with high interaction effect, according to the "sparsity of effects" principle. The remaining basis terms with high interaction effect will be added to the PCE model in the next basis expansion operation.

### 4.2.3 Basis pruning strategy

By the use of the basis expansion strategy, we can add massive basis terms to the current PCE model at once. However, the basis terms which can be added to the current PCE model would be

few if the basis terms already included in the current PCE model are in large quantity. Due to the sparse representation capability, insignificant basis terms in the PCE model will be trained with their coefficients equal (or nearly equal) to zero, thus having no or negligible influence on the modelling results. By removing these insignificant basis terms, the efficiency of the training process is improved while retaining the model precision. We name this operation as basis pruning, which will be implemented between the stability evaluation and the basis expansion. When it comes to tell that more basis terms are required, the basis terms with zero-value coefficients in the latest trained PCE model will be removed. Then new bases can be added to the PCE model in line with the basis expansion strategy. By performing the basis pruning strategy, not only the computational stability is expected to be enhanced, but also an opportunity to add more basis terms of higher degree is created.

### 4.2.4 Basis refinement strategy

With the basis expansion and pruning strategies, the basis terms can be adaptively selected to constitute the PCE model. However, some significant basis terms may be incorrectly removed in the basis pruning step since the samples and observations are perhaps insufficient to train a reliable PCE model during the iteration process. With the increase of samples, the hidden model features gradually appear in the observations, while those significant basis terms which have been permanently pruned cannot contribute to the building of accurate PCE model. In this regard, a basis

refinement strategy, which is inspired from the neighbour basis expansion (Jakeman et al., 2015), is introduced to resume the incorrectly removed basis terms. This strategy is performed in compliance with the basis pruning operation.

Assume that the PCE model after pruning has $P_k$ basis terms and the number of historically removed basis terms is $P_r$. The cluster of polynomial indices of these $P_k$ bases are defined as:

$$\boldsymbol{\alpha}_{P_k} = \left\{\boldsymbol{\alpha}: \boldsymbol{\alpha} \in \boldsymbol{\mathcal{A}}^{p,d}, \mu_{c_{\boldsymbol{\alpha}}} \neq 0\right\} \tag{4.10}$$

wherein $\boldsymbol{\mathcal{A}}^{p,d}$ has been defined in Equation (3.6); $\mu_{c_{\boldsymbol{\alpha}}}$ denotes the coefficient estimated by BCS, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots \alpha_d)$ is a vector which represents the index of a polynomial basis as explained in Equation (3.3). Now define a set $\boldsymbol{\alpha}_n$ which represents the neighbours of the indices from $\boldsymbol{\alpha}_{P_k}$:

$$\boldsymbol{\alpha}_n = \left\{\boldsymbol{\alpha}_i \pm \boldsymbol{e}_j: \boldsymbol{\alpha}_i \in \boldsymbol{\alpha}_{P_k}, \boldsymbol{\alpha}_i \pm \boldsymbol{e}_j \in \boldsymbol{\alpha}_{P_r} \text{ for } 1 \leq i \leq P_k, 1 \leq j \leq d\right\} \tag{4.11}$$

where $\boldsymbol{e}_j = \{0, \dots, 1, \dots, 0\}$ is a unite vector with non-zero value at the $j^{th}$ position; $\boldsymbol{\alpha}_{P_r}$ denotes the cluster of polynomial indices of historically removed basis terms. The proposed refinement strategy is to resume the historically removed basis terms with indices belonging to the neighbours of $\boldsymbol{\alpha}_{P_k}$. Then the PCE model constituted by the basis terms of indices $\boldsymbol{\alpha}_{P_k}$ and their neighbours $\boldsymbol{\alpha}_n$ is re-trained to update the coefficients solution. Those basis terms with zero-value coefficients will be removed again. This basis refinement strategy expands the PCE model with the basis neighbours by one polynomial degree, but it might be necessary to expand more than once. To explore those basis terms that are far away from the current basis terms, this strategy could be

applied recursively up to a fixed number $T$. A simple case with two input variables is illustrated in Figure 4.2. As shown in Figure 4.2(a), the abscissa and ordinate represent the degree of each variable in the PCE model. Each circle denotes a basis term, in which grey ones have non-zero coefficients and white ones are what have been removed. The current basis terms are then expanded to their neighbours as shown in Figure 4.2(b). The expanded basis terms are marked in red. The basis terms in both grey and red will constitute a new PCE model, which will be trained using BCS method. After that, only basis terms with non-zero coefficients are retained, and others are removed again, as illustrated in Figure 4.2(c).

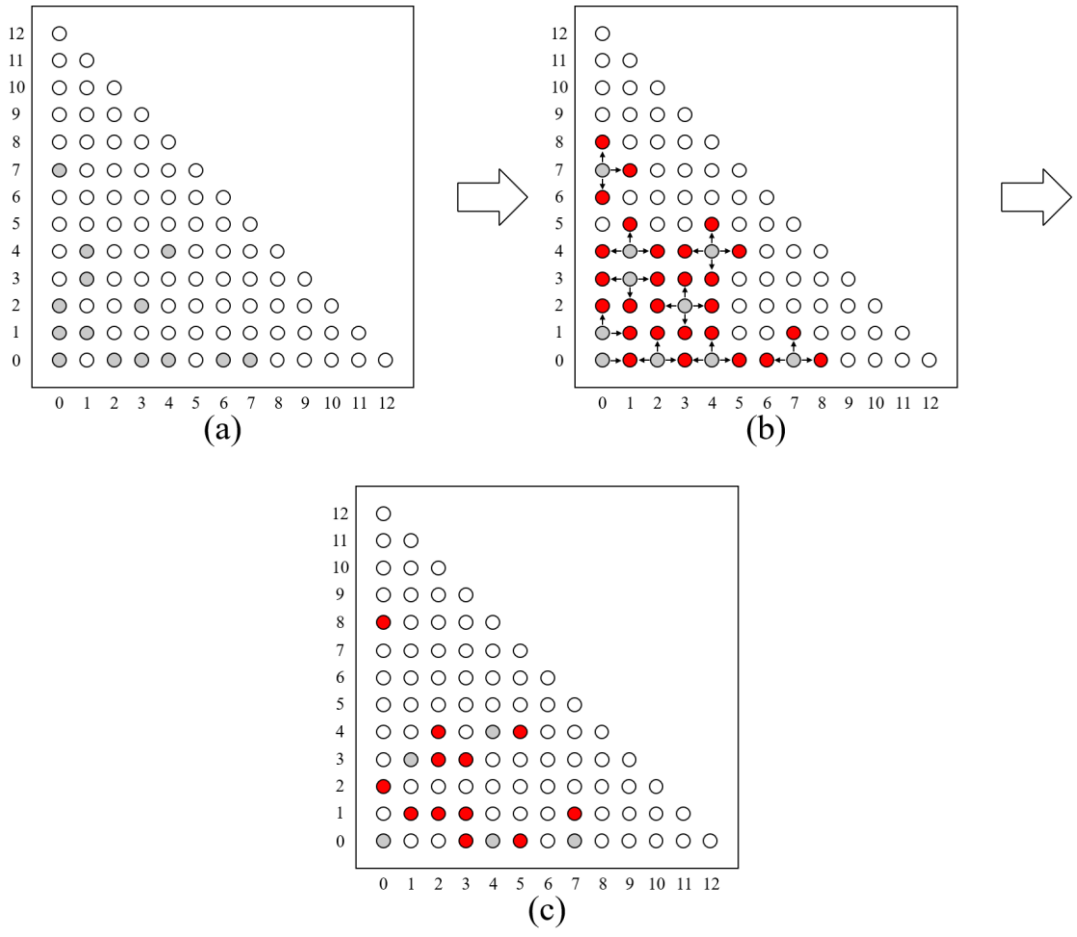



Figure 4.2 Simple 2-D case of basis refinement strategy

It has been classified in the past study that the target model feature on a high degree polynomial basis may be expressed on its neighbours of lower polynomial degree when high degree polynomial basis is not yet available in the current PCE model (Jakeman et al., 2015). When the training dataset is sufficient, this strategy was demonstrated to be effective in selecting basis terms of higher degree than those in the current PCE model. In this study, the training dataset is generally inadequate during the iteration process, so some features of the target model may not be involved or have incorrect manifestations in the observations. We think that these incorrectly represented features are probably to be presented on the basis terms which are in high relevance to the real significant bases. As the samples increase, the hidden model features are gradually discovered in the observations, and the significant basis terms which are wrongly removed in the basis pruning stage are able to be recovered. The implementation procedures of this strategy are given in **Algorithm 4**.

| **Algorithm 4**. Basis refinement strategy |
| --- |
| **Input:** Current PCE model; refinement times $T$; training dataset (current samples and observations). |
| **In $i$th iteration:**<br>1. Expand the basis terms with non-zero coefficients in the current PCE model to their neighbours;<br>2. Implement BCS to compute the coefficients;<br>3. Remove the basis terms with zero-value coefficients;<br>4. Calculate the CV error; If the error increases compared with the previous iteration step, resume to the last PCE model and quit iteration; If the error decreases, go to step 5;<br>5. If $i > T$, quit iteration. |

### 4.2.5 Adaptive PCE modelling

Leveraging the proposed adaptive basis selection strategy with the sequential sampling

approach, an Adaptive SPCE method with Basis Adaptivity and Sequential Sampling (ASPCE-BASS) is put forward. The flowchart of ASPCE-BASS is illustrated in Figure 4.3. This algorithm embarks upon a low truncation degree and a small number of samples. Then, the samples are enriched by EI-PLHS algorithm, along with assessing the stability of the trained PCE model. If the samples are sufficient to train a stable PCE model, the basis pruning strategy is executed to remove the insignificant basis terms with zero-value coefficients. Subsequently, the basis refinement strategy is performed $T$ times to recover incorrectly removed bases, along with pursuing the basis expansion strategy to expand the current PCE model with basis terms of higher degree up to 10 times of the quantity of the current samples. Eventually, the algorithm falls into the next loop to draw more samples.

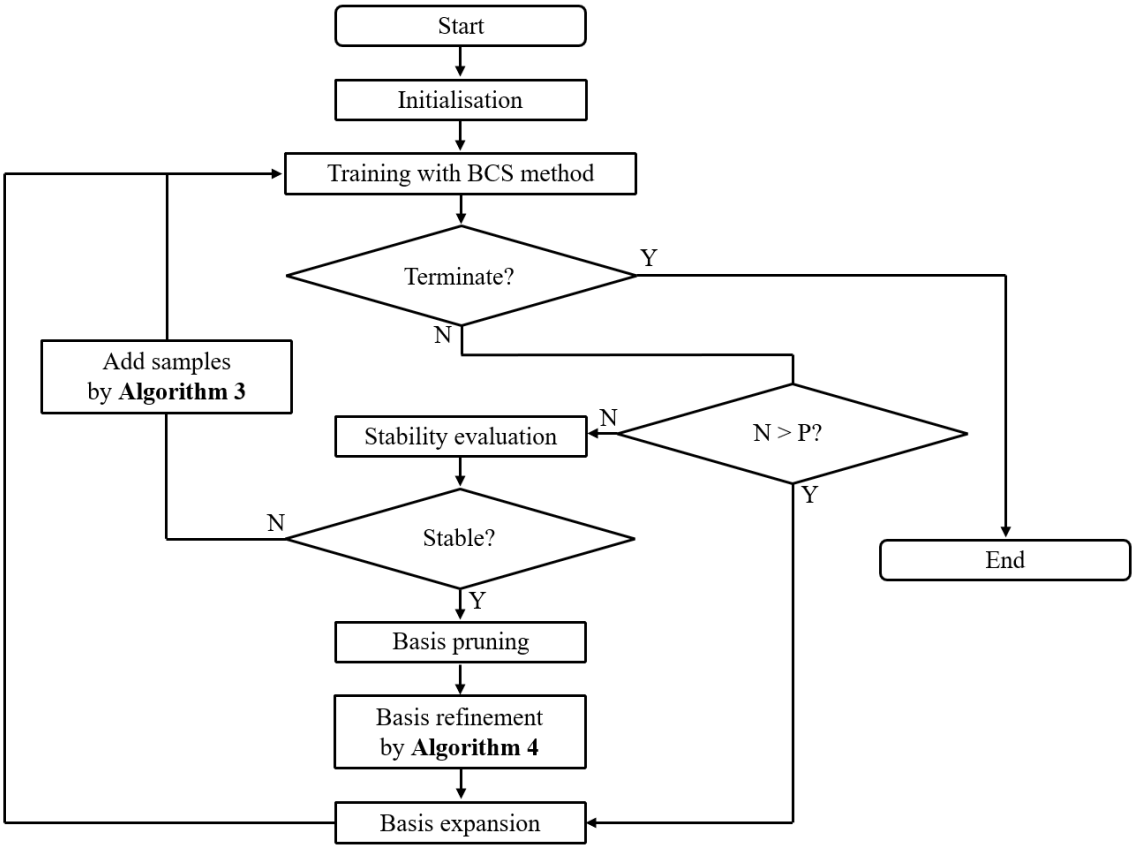



Figure 4.3 Flowchart of ASPCE-BASS

Initialisation of the sample number and truncation degree and specification of the window length and refinement times should be determined in implementing the proposed algorithm. Owing to the sparse representation, the number of initial samples $N_0$ can be a small value. Here, 10 to 20 samples are recommended for initialisation. To collect the initial samples, the LHS method is employed in accordance with **Algorithm 3**. As is explained in Section 4.2.2.2, the sparse representation via $l_1$ minimisation can ensure stable computation when the quantity of the unknown coefficients is no more than 10 times the sample number. Thus, the best choice of the initial truncation degree is that the cardinality of the initial PCE model is close to but no more than 10 times the initial sample number, and any lower values are feasible as well. Since the truncation

degree is contingent on both sample quantity and input dimension as shown in Equation (3.7), the choice of the initial sample number and the initial truncation degree should be considered concurrently to ensure the above criterion is satisfied.

In general, the precision of the trained PCE model which meets a predefined requirement is considered as the termination criterion of the proposed algorithm. Similar to the termination criterion shown in Section 3.3.4, the changes of the PCE model mean and std in successive iterations can be used to assess the convergence of the obtained PCE model. When the PCE model mean and std values in successive iterations keep stable or the change of these two statistical measures is less than a given threshold, the PCE model is viewed as having converged with satisfactory modelling accuracy, and the algorithm terminates. Otherwise, the stability of the PCE model during the sampling process will be evaluated. Appropriate values should be assigned to the window length $L_a$ and stability evaluation range $L_b$ in the stability evaluation process. Large values of these two parameters can provide more precise stability evaluation results, but more samples and higher training cost are required. Small values of the parameters have the potential to generate incorrect stability evaluation results, which will result in a low convergence speed. Through testing on several benchmark functions, they are both recommended as 15. Once the trained PCE model is evaluated as stable during the sequential sampling process, the unnecessary basis terms are removed, while historically removed basis terms have the opportunity to be recovered via $T$ times basis refinement. In this study, the parameter $T$ is chosen as 3.

## 4.3 Case Studies

In this section, three analytical benchmark functions with different input dimensions and complexities are used to validate the proposed method. An adaptive PCE approach (VARPCE) with basis adaptivity and sequential sampling presented by Thapa et al. (2020) is compared. The main difference between the proposed ASPCE-BASS method and the VARPCE method is that VARPCE employs a least-squares method for regression while ASPCE-BASS utilises BCS. The core idea behind VARPCE is adding a group of basis terms to the PCE model for training in each iteration and truncating the insignificant basis terms by evaluating their variance contributions to the whole model. A crucial parameter, named truncation threshold $\varepsilon_{trunc}$, was defined in VARPCE to control the amount of basis terms which will be truncated. A large value of truncation threshold will help to truncate a large amount of basis terms. In this setting, the algorithm can converge fast, but the model precision is sacrificed. On the other hand, a small value of truncation threshold helps to retain some basis terms with small variance contributions, but the algorithm will be in low convergence speed, which implies that more samples and model responses for PCE training are required. The truncation threshold was recommended between $\{10^{-4}, 10^{-3}\}$ by Thapa. Besides the truncation threshold, some other parameters should also be prescribed, such as the upper bound of the PCE degree $p_{max}$ to terminate the algorithm and the oversampling rate $t$ to stop the sampling process. In this study, different truncation threshold values will be tested. The upper bound of the truncation degree will be determined according to the problem complexity. The oversampling rate $t$ was

recommended between $\{2,4\}$ (Thapa et al., 2020), which is set as 3 in this study. In addition to VARPCE, a sequential sampling strategy without basis adaptivity, EI-ELF, is also compared to illustrate the performance of the proposed adaptive basis selection strategy. A predefined candidate sample pool is required for both ASPCE-BASS and EI-ELF. To keep the sufficiency in candidate samples, a large size $10^4$ is adopted in this study. These samples are generated by LHS method in advance.

To facilitate comparison, the maximum sample quantity is restricted in the case studies beyond which the algorithms terminate. To assess the real precision of the obtained PCE models, a validation set with $N_{val} = 10^4$ random samples and the exact model responses is used. The Relative Root Mean Square Error (RRMSE) $\varepsilon_{RRMSE}$ (Diaz et al., 2018) is still calculated by Equation (3.35). The Matlab codes of the EI-ELF criterion and VARPCE method are developed in-house according to references (Thapa et al., 2020; Zhou et al., 2019a).

### 4.3.1 The Ishigami function

The first test function is the Ishigami function, which is the same to Section 3.4.2. The equation is re-written here, and all the variables conform to uniform distributions on the interval $[-\pi, \pi]$.

$$y = \sin x_1 + a(\sin x_2)^2 + b{x_3}^4 \sin x_1 \tag{4.12}$$

where $a, b$ are two parameters which are commonly chosen as 7 and 0.1, respectively, for benchmark test. The analytical mean is 3.5 and the std value is around 3.7208. The comparison is

first performed between ASPCE-BASS and VARPCE. Since an upper bound of truncation degree is required in VARPCE, we set the degree upper bound $p_{max} = 14$ in both methods for fair comparison, which generates $P = 680$ basis terms. In initialisation, VARPCE starts from degree 1, and the ED size is automatically determined as 8 according to Thapa et al. (2020). ASPCE-BASS starts with 20 samples, and the degree is chosen as 5 for initialisation. To avoid poor outcomes from VARPCE caused by the improper selection of truncation threshold parameter, four truncation threshold values, $10^{-3}, 10^{-4}, 10^{-5}$ and $10^{-6}$, are compared in this study. Each method is calculated for 30 times to ensure the statistical stability. The RRMSE results with respect to the increase of ED size are shown by box plots in Figure 4.4.

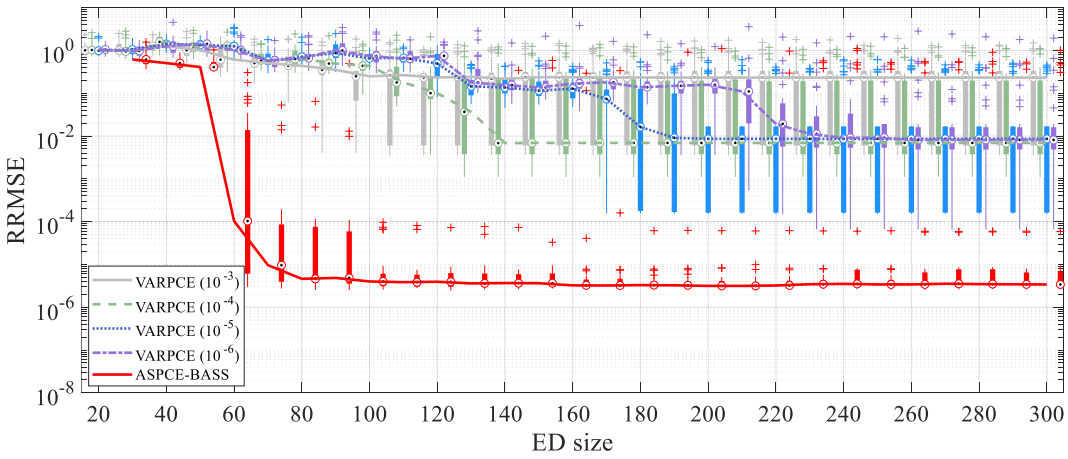



Figure 4.4 RRMSE with increasing ED size

In the box plots, bold vertical lines represent the range between upper and lower quartiles of the 30 results, and fine vertical lines represent the 1.5 times interquartile range which constrains the normal value limitation. Values out of them are regarded as outliers, which are depicted as plus symbols in the figure. The dot inside the white circle denotes the median of RRMSE, and the lines

that vary with respect to the ED size represent the variation of the median of RRMSE obtained by different approaches (the plot is truncated at 300 ED size). In VARPCE, the algorithm will terminate when the basis terms of degree $p_{max}$ are all explored even if the sample quantity is smaller than 300. Such results are hard to be shown in the plot since the algorithm may terminate with different ED sizes in the repeated tests. To facilitate comparison, those early terminated results in VARPCE are retained in the box plot as shown in Figure 4.4. It can be observed that VARPCE converges more and more slowly with the decrease of truncation threshold value, but the precision of the trained PCE model increases. ASPCE-BASS has the lowest validation error after convergence, and also, the convergence rate is the fastest as compared with the four VARPCE models.

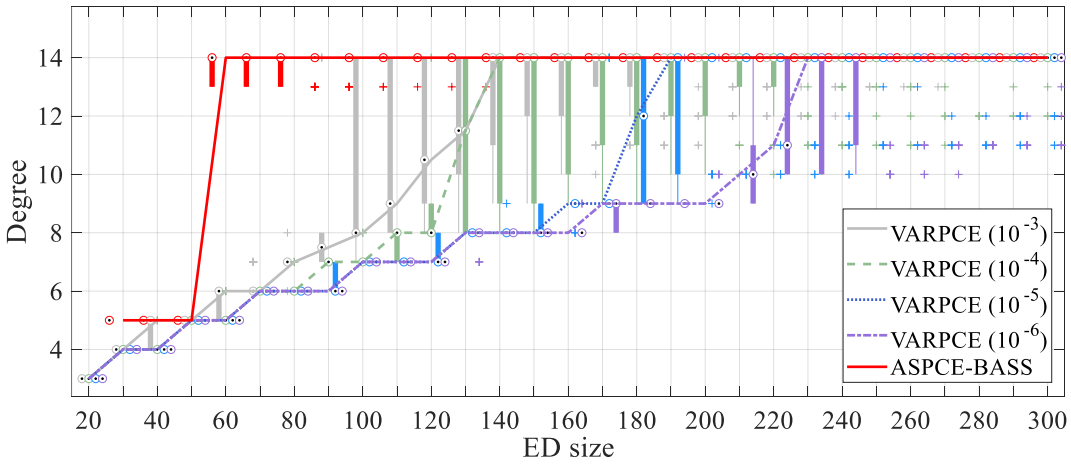



Figure 4.5 Degree changes with increasing ED size

The PCE degree changes with the increase of ED size are depicted in Figure 4.5. Compared to VARPCE, ASPCE-BASS can quickly approach the highest degree. With the decrease of truncation threshold value, VARPCE becomes more and more slow in adding basis terms of higher degree. Some of the repeated tests cannot even reach the degree upper bound when the ED size is 300, since

too many insignificant basis terms are retained.

The sample quantities used for terminating the VARPCE in the repeated tests are recorded as shown in Figure 4.6. With the decreasing of the threshold value, the samples needed for termination increase. The RRMSE results of VARPCE after termination are compared with results from ASPCE-BASS under different ED sizes, which are portrayed in Figure 4.7. It can be concluded that ASPCE-BASS with ED size larger than 100 performs better than VARPCE under all threshold values.

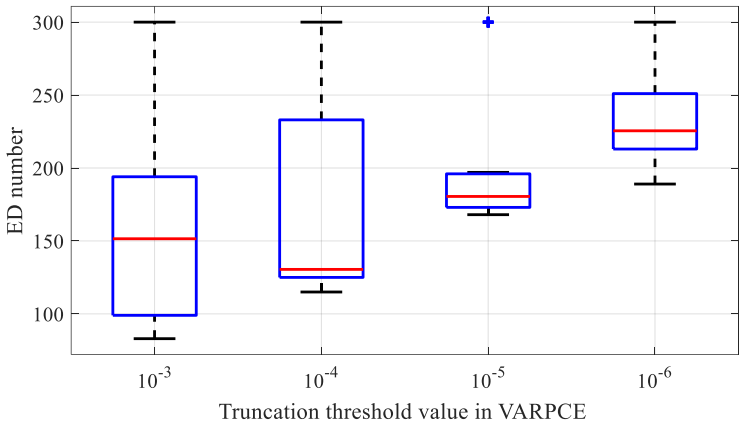



Figure 4.6 Sample quantities for automatic termination of VARPCE

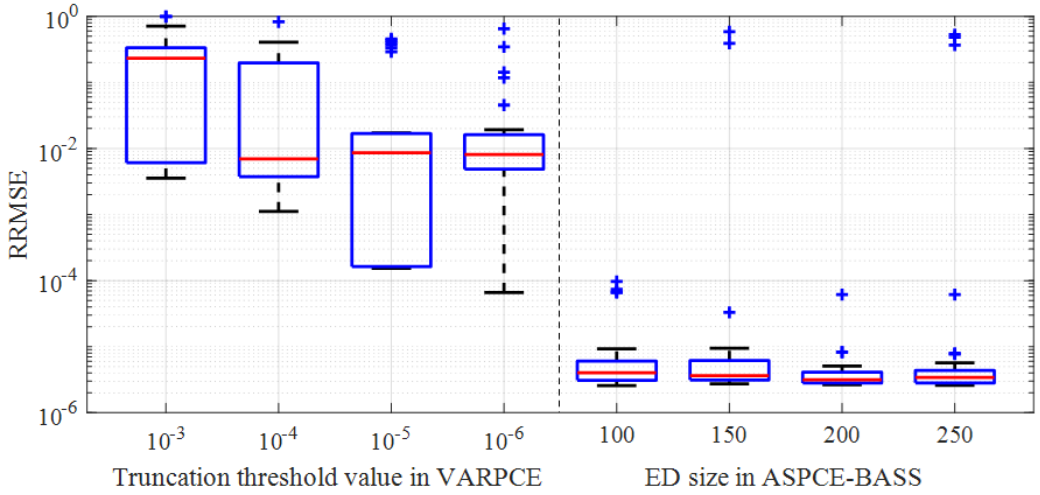



Figure 4.7 RRMSE comparison between VARPCE and ASPCE-BASS with different ED sizes

Then, the proposed method is compared with the EI-ELF method with different truncation

degrees. Four degrees, 9, 12, 14 and 16, are chosen in EI-ELF for comparison, which give rise to full PCE expansions of $P = 220$, 455, 680 and 969, respectively. Here, we do not provide upper bound on the truncation degree in ASPCE-BASS. The initial sample number for EI-ELF is set as 20. Each method is calculated for 30 times. The RRMSE results of the two methods are shown in Figure 4.8, and the degree changes of ASPCE-BASS with respect to ED size are given in Figure 4.9. With the increase of truncation degree, EI-ELF converge more and more slow, which need 40, 50, 60 and 70 samples for each truncation degree, respectively, but the precision of the obtained PCE model increase. Compared with EI-ELF, ASPCE-BASS shows a relatively slow convergence speed but has a similar convergence trend. It can automatically select the significant basis terms with different degrees without largely sacrificing the convergence speed and modelling accuracy. As shown in Figure 4.9, all the repeated tests can explore basis terms with degree 13 after having 60 samples. Most of the repeated tests can explore degree 19 to 20 after having 110 samples.

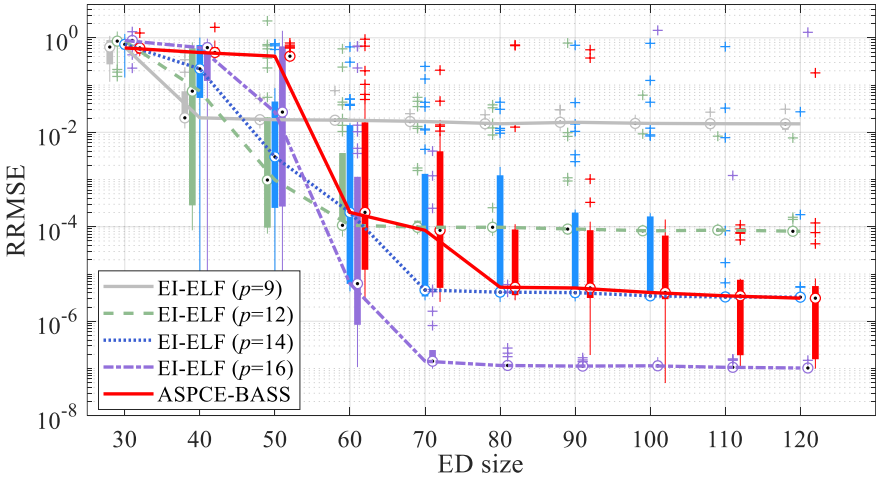



Figure 4.8 RRMSE comparison between EI-ELF and ASPCE-BASS with increasing ED size

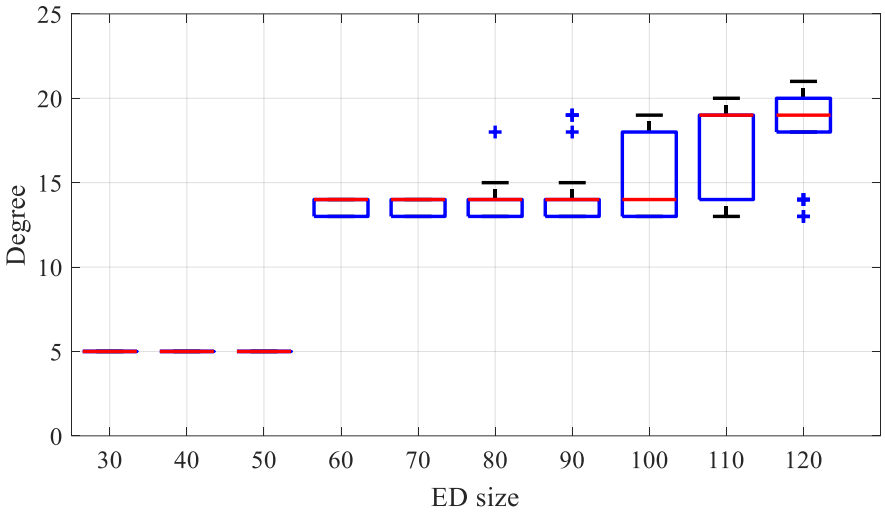


Figure 4.9 Degree changes of ASPCE-BASS with increasing ED size

The model mean and std values of 30 repeated tests with respect to the increase of ED size are shown in Figure 4.10. It is obvious that the PCE models have an evident fluctuation before 60 samples and they converge to a stable state after collecting 70 samples. The stability evaluation process of 30 repeated tests is depicted in Figure 4.11. There are obviously two stable regions. The first region is before 30 iterations, and the second region is after 50 iterations. At 30 iterations, the PCE model is trained with 50-65 samples. This is in consistence with the time point that RRMSE decreases and degree increases as depicted in Figure 4.8 and Figure 4.9. The PCE models under initial truncation degree 5 are evaluated to be stable when the samples are increased to around 50, and then more basis terms of high degrees are added. At around 40 iterations in Figure 4.11, 60-75 samples are collected for training. The variance of model std decreases rapidly ruing this period. After 50 iterations, the variance values converge to two clusters. The upper cluster is constituted by results with the highest polynomial degree 13 and the bottom cluster is constituted by results with the highest polynomial degree 14. Several outliers result in the phenomena in the red oval, but the

stability evaluation process still works when these outliers are excluded in the stability evaluation process.

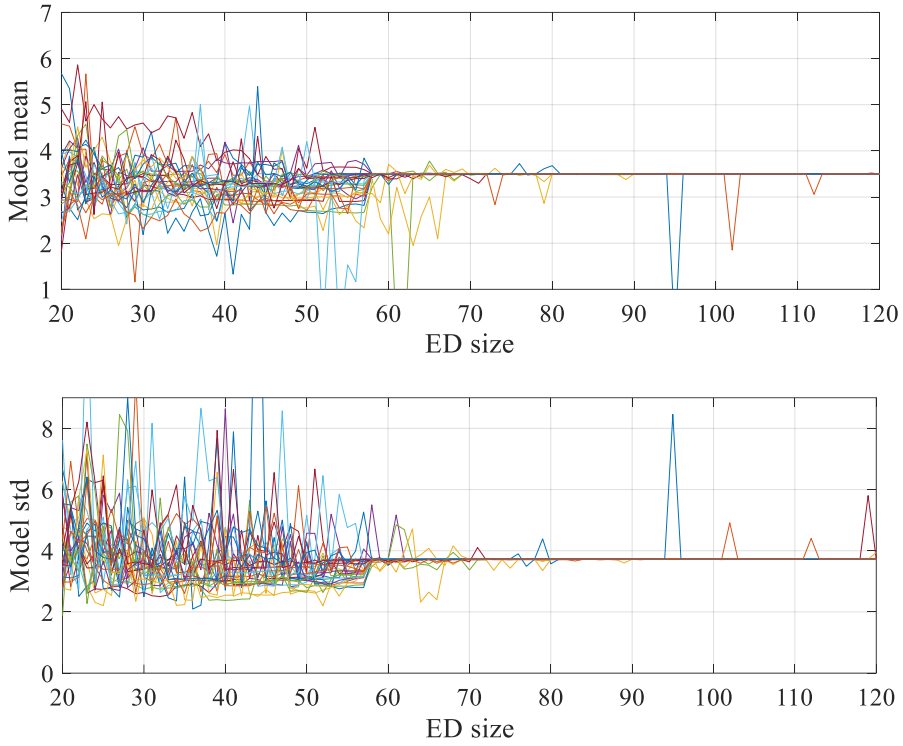



Figure 4.10 Model means (left) and standard deviations (right) with increasing ED size

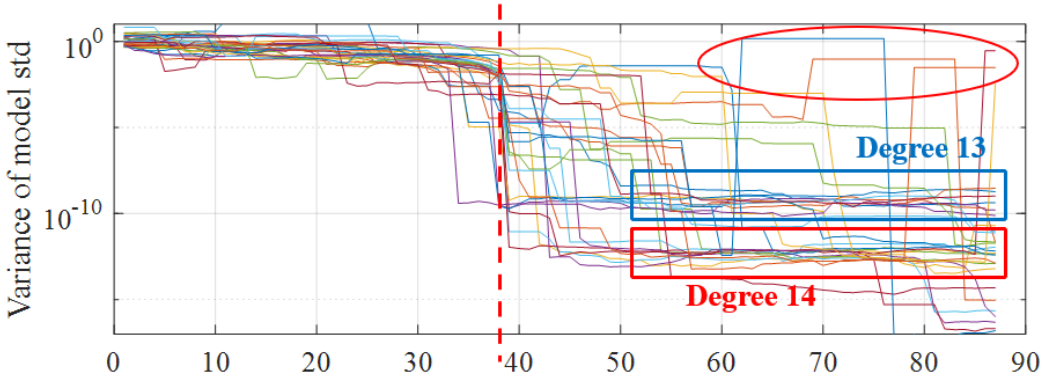



Figure 4.11 Stability evaluation process

In summary, the proposed ASPCE-BASS which employs sparse representation outperforms the VARPCE method which employs the least-squares method. The former has a high computational efficiency in exploring basis terms with high polynomial degree while the modelling accuracy can

be kept. Compared to sequential sampling-only method, ASPCE-BASS can adaptively select basis terms without largely sacrificing convergence rate, which confirms the capability of the proposed adaptive basis selection strategy.

### 4.3.2 High-dimensional function

The second benchmark test function is a high-dimensional function, which is the same to Section 3.4.3. Its expression is re-written as follows:

$$y = 3 - \frac{5}{d}\sum_{i=1}^{d} i x_i + \frac{1}{d}\sum_{i=1}^{d} i\, x_i^3 + ln\left[\frac{1}{3d}\sum_{i=1}^{d} i(x_i^2 + x_i^4)\right] \tag{4.13}$$

where $d$ is the input dimension chosen by user. All the input variables conform to uniform distributions defined on the interval $[1, 2]$. The input dimension is selected as $d$ = 30 in this study. The comparison is first performed between ASPCE-BASS and VARPCE. In initialisation, VARPCE starts from degree 1 and 62 samples according to Thapa et al. (2020), and ASPCE-BASS starts with degree 1 and 20 samples. Again, four truncation threshold values, $10^{-3}, 10^{-4}, 10^{-5}$ and $10^{-6}$, are tested in VARPCE. Due to the heavy computational burden, each method is repeated for 5 times. The RRMSE results and PCE degree changes with respect to the increase of ED size are depicted in Figure 4.12 and Figure 4.13.

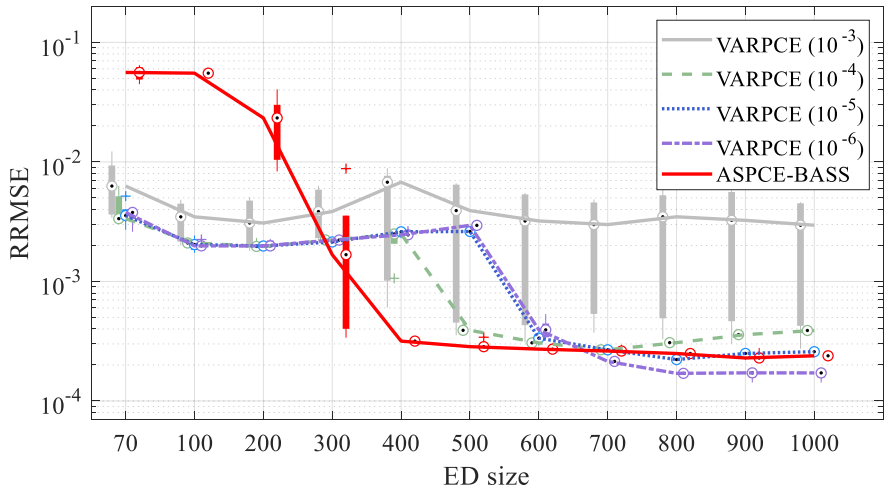



Figure 4.12 RRMSE comparison between VARPCE and ASPCE-BASS with increasing ED size

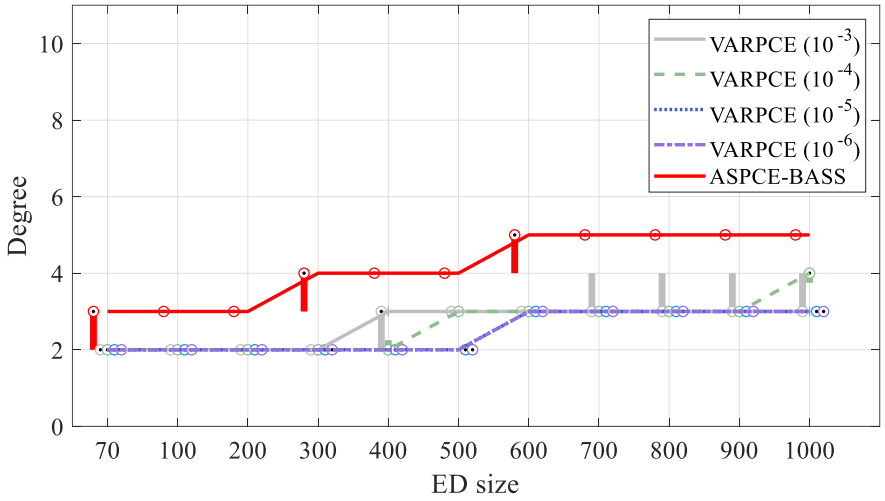



Figure 4.13 Degree changes with increasing ED size

It can be seen from Figure 4.12 that the PCE model obtained by ASPCE-BASS show good performance after having 400 samples. In VARPCE, the PCE models obtained under truncation threshold value $10^{-3}$ show the worst performance. The PCE models obtained under other three threshold values all converge after collecting 600 samples. The PCE model obtained by ASPCE-BASS converges to a similar validation error value with those obtained by VARPCE. As illustrated in Figure 4.13, ASPCE-BASS can explore more basis terms of higher degree than VARPCE.

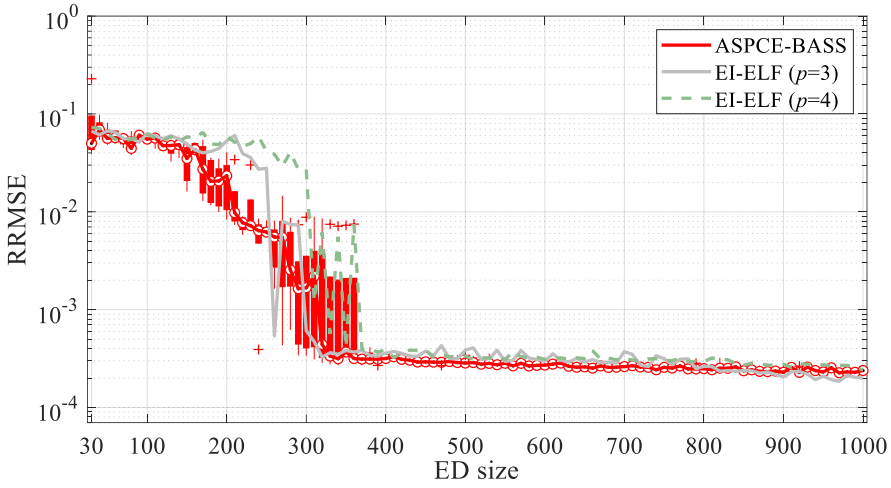



Figure 4.14 RRMSE comparison between EI-ELF and ASPCE-BASS with increasing ED size

Then, the proposed method is compared with EI-ELF with two truncation degrees, 3 and 4, which result in full PCE expansions with cardinality of $P = 5456$ and 46376, respectively. The initial sample number for EI-ELF is set as 20. Only one calculation for each degree is provided due to the heavy regression calculation burden. The RRMSE results of ASPCE-BASS and EI-ELF are shown in Figure 4.14. The results of ASPCE-BASS here are the same as those depicted in Figure 4.12, and the degree changes can refer to Figure 4.13. As is shown in Figure 4.14, the PCE models obtained from EI-ELF under different degrees have similar accuracy. The PCE model trained by ASPCE-BASS also have the similar precision to those obtained by EI-ELF during the whole iteration process, while the computational cost of ASPCE-BASS is lower than EI-ELF.

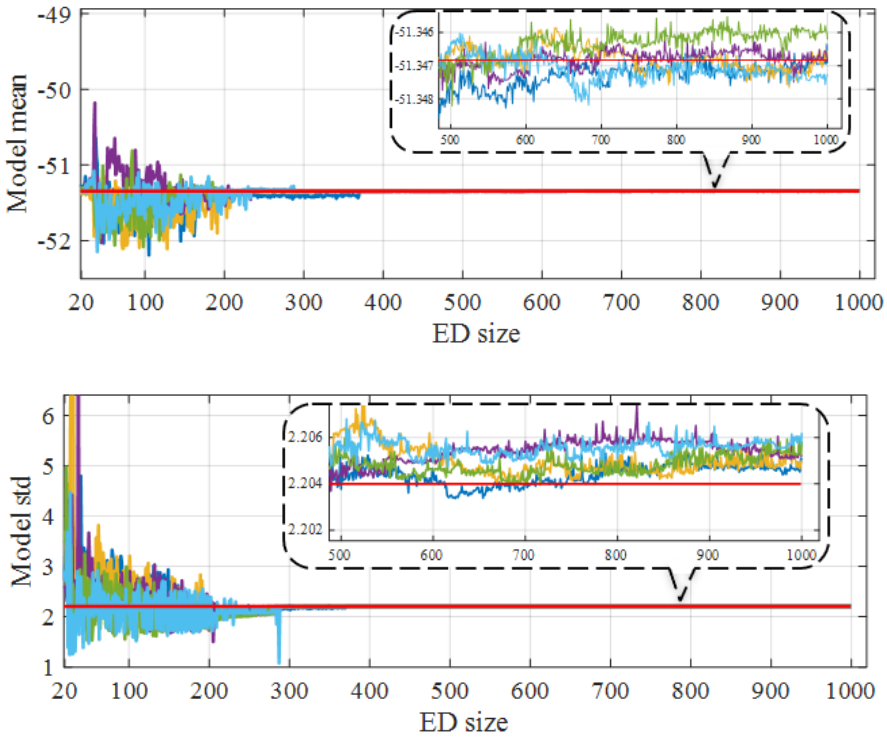



Figure 4.15 Model means (left) and standard deviations (right) with increasing ED size

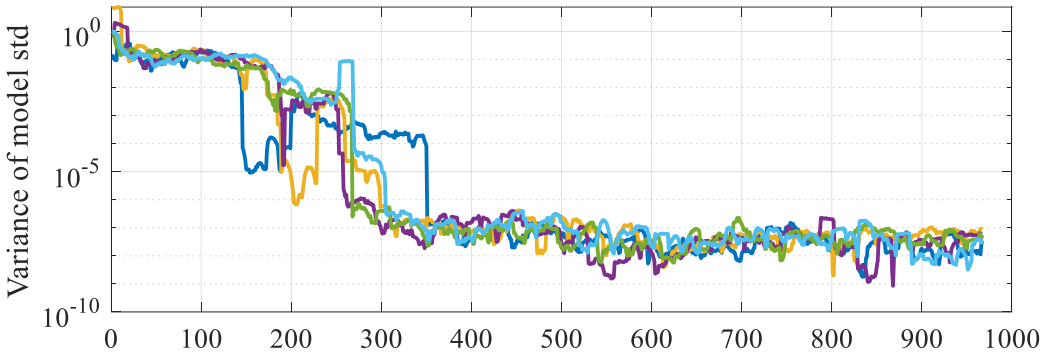



Figure 4.16 Stability evaluation process

The model mean and std values of 5 repeated tests from ASPCE-BASS during the iteration process are shown in Figure 4.15. The red straight lines are the reference values of model mean and std calculated from $10^6$ function evaluations. The reference values are -51.3468 and 2.2040 for model mean and std, respectively. It is shown in Figure 4.15 that the trained PCE models have large fluctuations when samples are less than 300. With the increase of samples, the trained PCE models gradually converge to the true value. The stability evaluation process is shown in Figure 4.16. The

variance of model std declines rapidly after collecting 300 samples, and they become stable with more than 400 samples. These results demonstrate that the adaptive basis selection strategy and the stability evaluation process work well in the proposed method, which make the PCE model more and more accurate.

### 4.3.3 Algebraic test function

The third benchmark test function is the algebraic corner-peak test function, which is expressed in Equation (4.14):

$$y = \left(1 + \sum_{k=1}^{d} c_k x_k\right)^{-(d+1)} \tag{4.14}$$

where $\boldsymbol{x} = \{x_1, \dots, x_d\}$ are uniformly distributed on $[0,1]^d$, and $d$ is the dimensionality chosen by user. This function is flexible in controlling the effective dimensionality and compressibility by assigning different values to the coefficients $\boldsymbol{c} = (c_1, c_2, \dots, c_d)$. In this study, three different choices are provided (Jakeman et al., 2015):

$$c_k^{(1)} = e^{(-8k/d)}, \qquad c_k^{(2)} = \frac{1}{k^2}, \qquad c_k^{(3)} = \frac{k - 0.5}{d} \qquad k = 1,2, \dots, d$$

which are normalised such that $\sum_{k=1}^{d} c_k = 0.25$. The coefficients $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}$ and $\boldsymbol{c}^{(3)}$ imply an increase of effective dimensionality, which means that Equation (4.14) with coefficients $\boldsymbol{c}^{(3)}$ has the most significant polynomial terms, and Equation (4.14) with coefficients $\boldsymbol{c}^{(1)}$ has the minimum

number of effective polynomial terms. The dimensionality $d$ is set as 10 in this case. The modelling results by different methods are compared under different $\boldsymbol{c}$. The comparison is first conducted between ASPCE-BASS and VARPCE. The degree upper bound is set as 10 for both methods, which generates $P = 184756$ polynomial basis terms. In initialisation, VARPCE starts from degree 1 and 22 samples, and ASPCE-BASS starts with degree 2 and 20 samples. Similarly, four truncation threshold values, $10^{-3}, 10^{-4}, 10^{-5}$ and $10^{-6}$, are set in VARPCE for comparison. Due to the heavy computational burden, each method is repeated for 5 times. Then, the proposed method is compared with EI-ELF. Three truncation degrees, 5, 7, and 9, are set for EI-ELF, which result in full PCE expansions of $P = 3003$, 8976 and 92378, respectively. The initial sample number for EI-ELF is set as 20, and the maximum sample quantity is 1000. Due to the computational burden for non-basis-adaptive method in this case, only one calculation for each degree is made.

#### 4.3.3.1 $\boldsymbol{c^{(1)}}$

The RRMSE results of ASPCE-BASS and VARPCE with respect to the increase of ED size are shown by box plots in Figure 4.17, and the PCE degree changes with the increase of ED size are depicted in Figure 4.18.

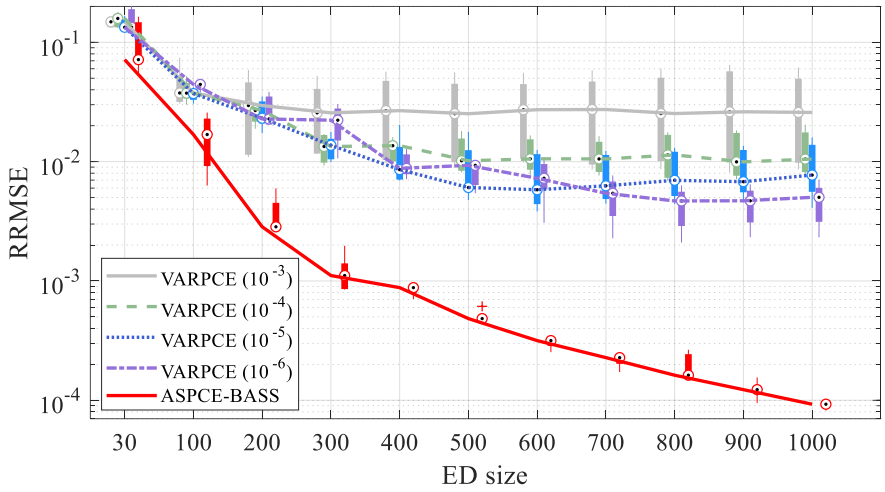



Figure 4.17 RRMSE with increasing ED size

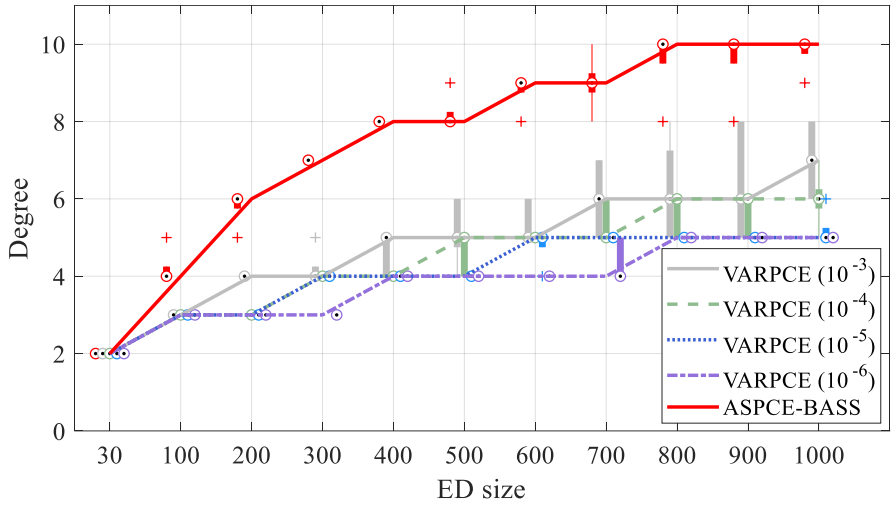



Figure 4.18 Degree changes with increasing ED size

As shown in Figure 4.18, when the ED size is increased to 1000, VARPCE with any truncation threshold value cannot explore basis terms with polynomial degree over 8, while ASPCE-BASS can explore basis terms with the highest degree 10. Under the same sample quantity, ASPCE-BASS can explore more basis terms of higher degree than VARPCE. In fact, not all the basis terms of degree 10 are explored by ASPCE-BASS. Both algorithms terminate by ED size in this case, and the degree upper bound does not work. With the decrease of threshold value, the highest degree explored by VARPCE gradually decreases. The reason is that the increase of the retained basis terms of low

degree impedes the further addition of basis terms of high degree. From Figure 4.17, it can be seen that the RRMSE values under 1000 ED size decrease with the decrease of threshold value in VARPCE. Although VARPCE with small threshold value explores less basis terms than with large threshold value, the trained PCE models with small threshold value are of higher accuracy than those with large threshold value since more significant basis terms can be retained. Compared to VARPCE, the PCE models obtained from ASPCE-BASS have the lowest validation error during the whole iteration process.

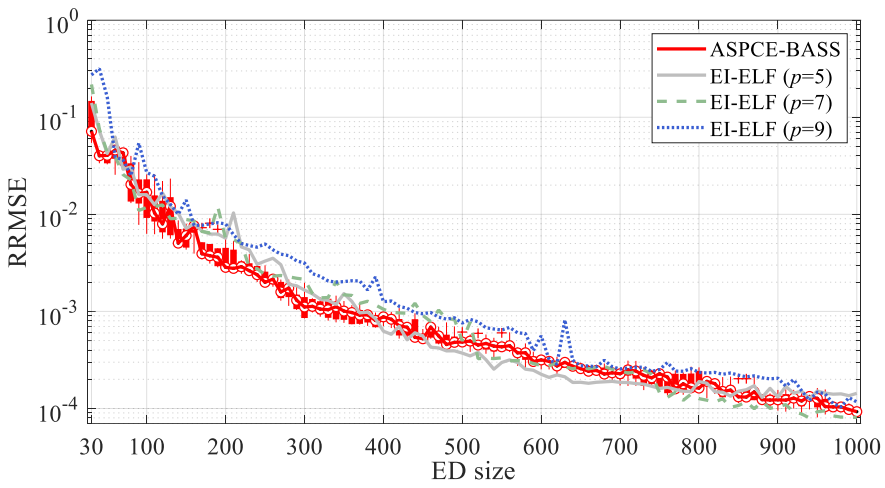



Figure 4.19 RRMSE comparison between EI-ELF and ASPCE-BASS with increasing ED size

The RRMSE results of ASPCE-BASS and EI-ELF with respect to the increase of ED size are compared in Figure 4.19. The RRMSE results of ASPCE-BASS here are the same as those depicted in Figure 4.17 since the degree upper bound does not work. The degree change of ASPCE-BASS can refer to Figure 4.18 as well. It can be observed that results from ASPCE-BASS show similar performance with those from EI-ELF, which means that the proposed method does not sacrifice the convergence rate of the sequential sampling process while the basis terms can be adaptively selected.

More importantly, the number of unknown coefficients to be solved in ASPCE-BASS in each iteration is always less than that in EI-ELF with the same truncation degree. For example, ASPCE-BASS could explore basis terms over degree 9 in this case with 1000 samples, so more than 92378 basis terms have been explored. But the number of unknown coefficients solved in each regression calculation is always smaller than 10000 because it cannot exceed 10 times the sample quantity. In EI-ELF with truncation degree 9, however, the number of unknown coefficients to be solved is always 92378 in each iteration. So, the calculation cost in EI-ELF is much higher than ASPCE-BASS. The proposed ASPCE-BASS method can not only adaptively select basis terms but also alleviate the computational burden in regression calculation.

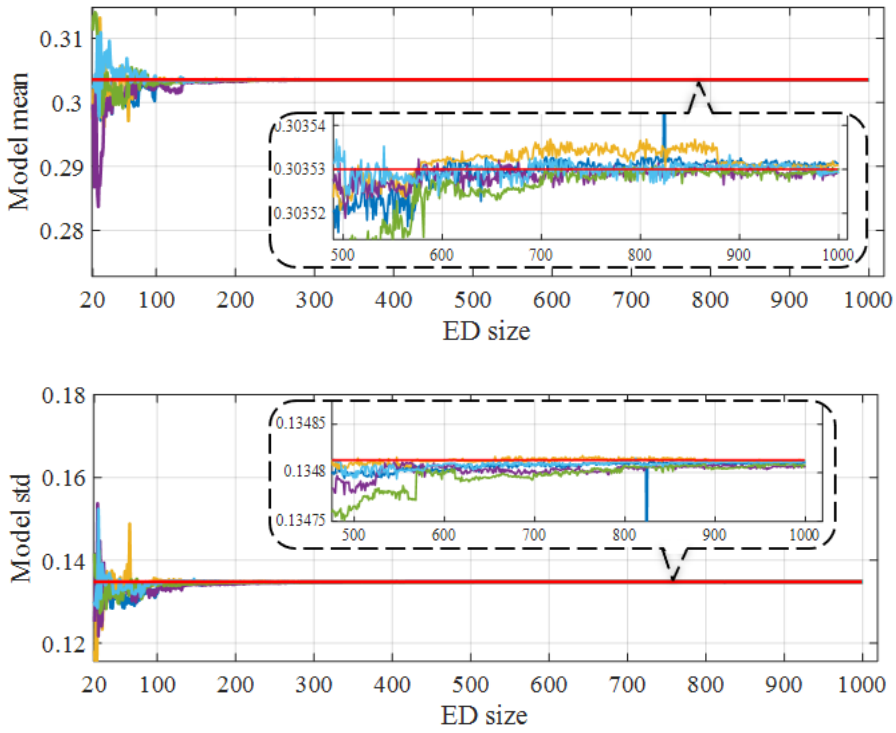



Figure 4.20 Model means (left) and standard deviations (right) with increasing ED size

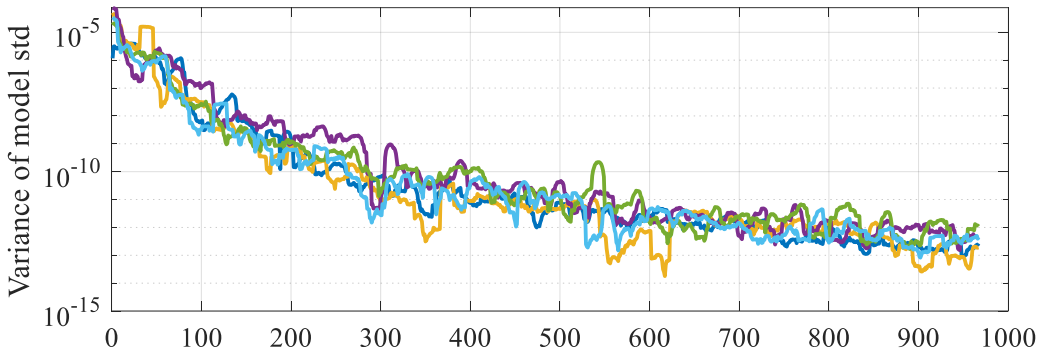



Figure 4.21 Stability evaluation process

To illustrate the iteration process of the ASPCE-BASS method, the model mean and std values of 5 repeated tests are shown in Figure 4.20, and the stability evaluation process is depicted in Figure 4.21. The red straight lines in Figure 4.20 are the reference model mean and std values which are calculated from $10^6$ function evaluations. The reference mean and std values are 0.30351 and 0.13481, respectively. As is shown in Figure 4.20, the trained PCE model fluctuates more and more weakly with the increased samples, and the model mean and std values gradually converge to the reference values. In Figure 4.21, the variance of model std shows stable downward trend, which denotes that the adaptive basis selection strategy and stability evaluation process work well.

#### 4.3.3.2 $\boldsymbol{c}^{(2)}$

In the case with coefficients $\boldsymbol{c}^{(2)}$, the number of effective components in Equation (4.14) is more than the preceding case with coefficients $\boldsymbol{c}^{(1)}$. More polynomial coefficients in the PCE model are non-zero. The RRMSE results and PCE degree changes of ASPCE-BASS and VARPCE with respect to the increase of ED size are portrayed by box plots in Figure 4.22 and Figure 4.23. The results show similar performance as to the case with $\boldsymbol{c}^{(1)}$. ASPCE-BASS has the fastest

convergence rate, and the PCE model obtained by ASPCE-BASS has the minimum validation error compared to those obtained by VARPCE. Most of the repeated tests by ASPCE-BASS can explore basis terms with the highest degree 10, while VARPCE can only explore basis terms of degree no more than 9.

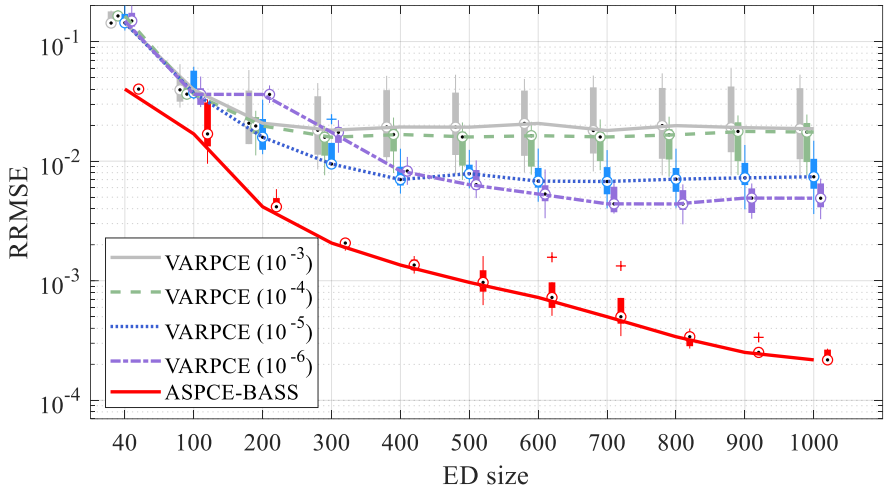



Figure 4.22 RRMSE comparison between VARPCE and ASPCE-BASS with increasing ED size

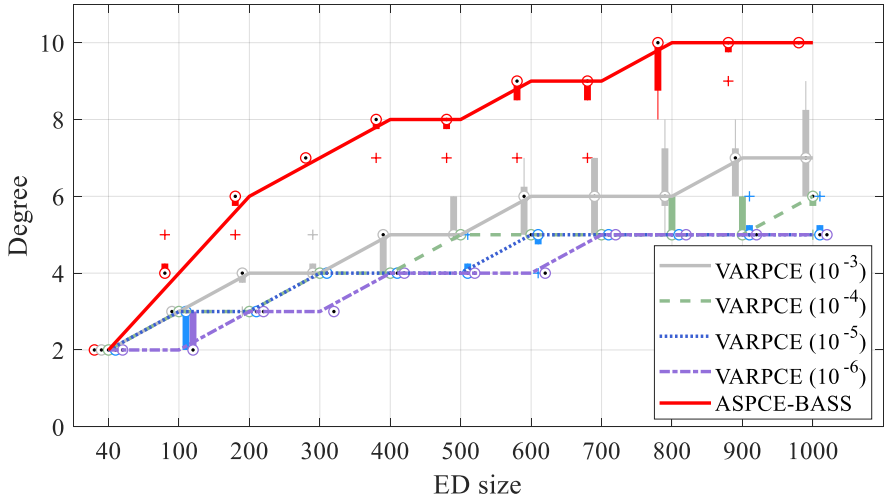



Figure 4.23 Degree changes with increasing ED size

The RRMSE results of ASPCE-BASS and EI-ELF with respect to the increase of ED size are shown in Figure 4.24. The degree changes of ASPCE-BASS can refer to Figure 4.23. We can observe from Figure 4.24 that results obtained from EI-ELF under different truncation degrees show

similar performance within 1000 ED size. Again, the results from ASPCE-BASS show quite similar performance with those from EI-ELF.

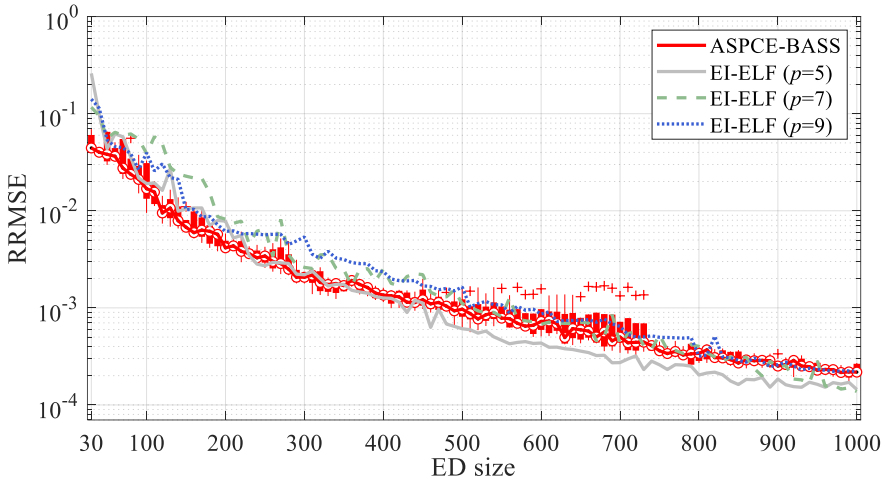



Figure 4.24 RRMSE comparison between EI-ELF and ASPCE-BASS with increasing ED size

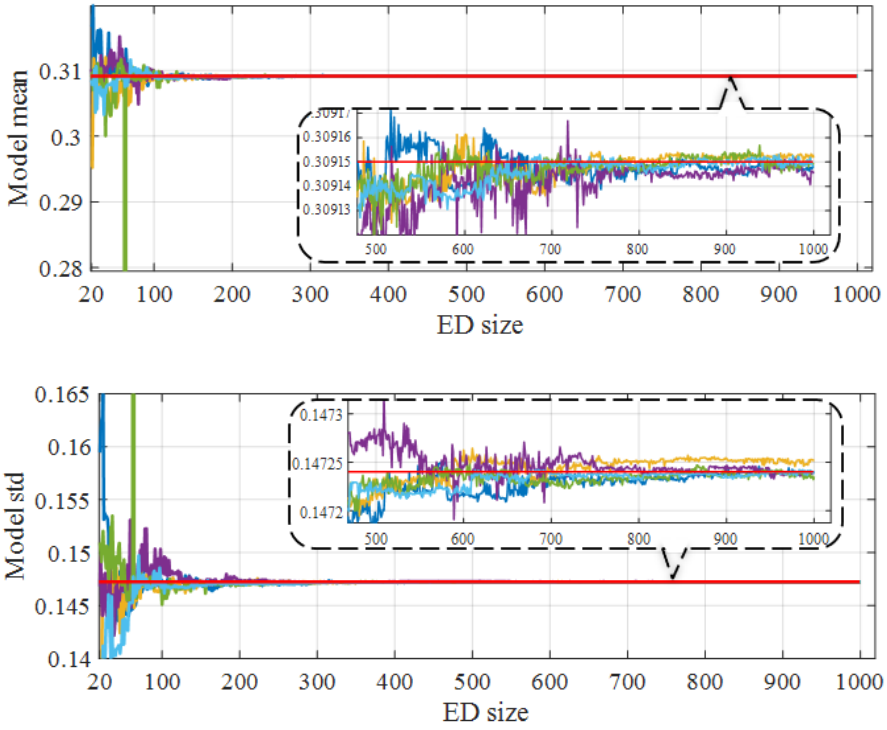



Figure 4.25 Model means (left) and standard deviations (right) with increasing ED size

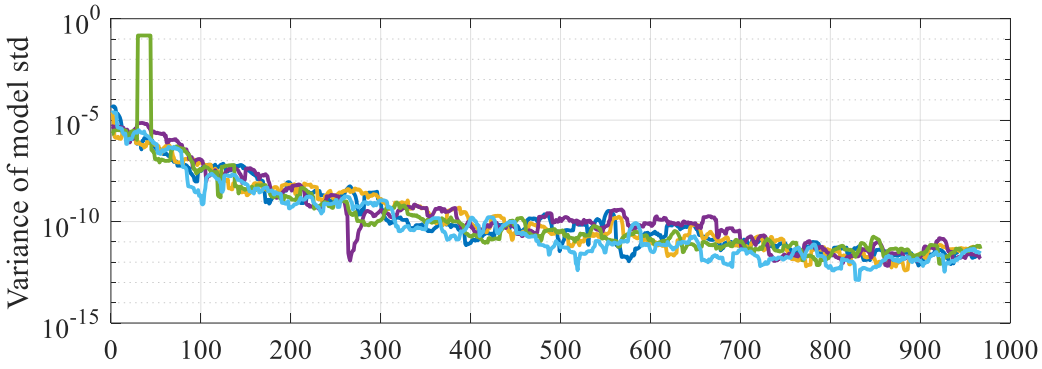



Figure 4.26 Stability evaluation process

The model mean and std values of 5 repeated tests with respect to the increase of ED size by using ASPCE-BASS are shown in Figure 4.25. The red straight lines are also the reference values calculated from $10^6$ function evaluations. The reference values are 0.30915 and 0.14724 for model mean and std, respectively. With the increase of samples, the mean and std values of the trained PCE model gradually converge to the true values. The stability evaluation processes of 5 repeated tests are depicted in Figure 4.26, which shows that the stability evaluation process is globally in steady downward trend.

#### 4.3.3.3 $\boldsymbol{c}^{(3)}$

Among three values of $\boldsymbol{c}$ in this section, $\boldsymbol{c}^{(3)}$ gives Equation (4.14) the most effective polynomial terms than the other two, that is to say, Equation (4.14) with $\boldsymbol{c}^{(3)}$ will be expressed by PCE model with the most non-zero coefficients. The RRMSE results and PCE degree changes of ASPCE-BASS and VARPCE with respect to the increase of ED size are displayed in Figure 4.27 and Figure 4.28.

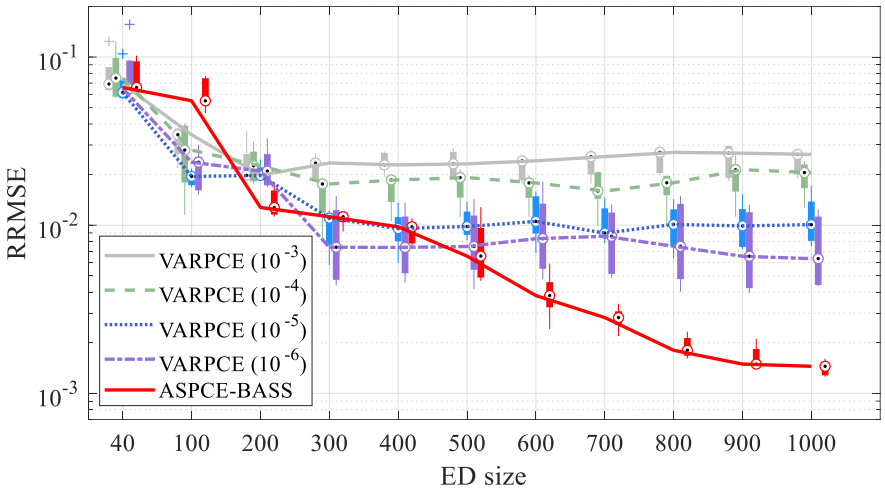



Figure 4.27 RRMSE comparison between VARPCE and ASPCE-BASS with increasing ED size

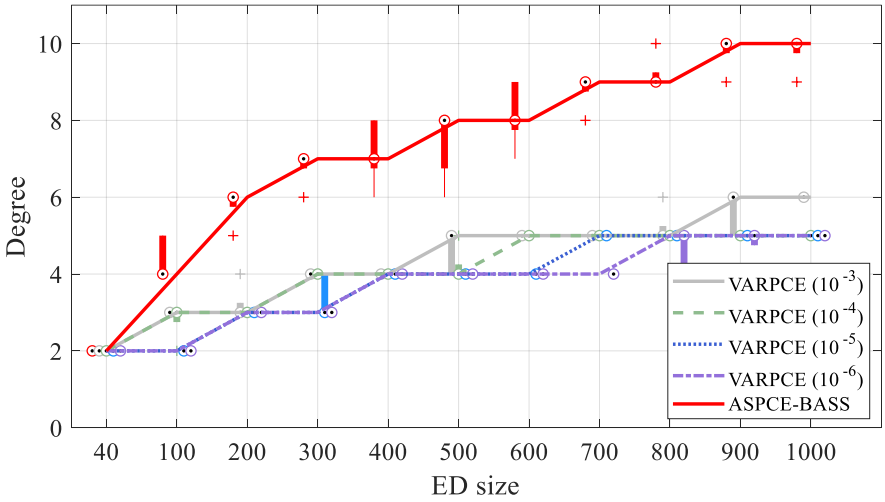



Figure 4.28 Degree changes with increasing ED size

As shown in Figure 4.28, some of the repeated tests by ASPCE-BASS under ED size 1000 can only explore basis terms of degree 9. This is because the number of effective polynomial terms increases compared to the previous cases with $\boldsymbol{c}^{(1)}$ and $\boldsymbol{c}^{(2)}$. More basis terms are retained during the iteration process, thereby decreasing the number of basis terms of higher degree that can be added. However, ASPCE-BASS can still explore basis terms of higher degree in comparison with VARPCE. In Figure 4.27, the PCE models obtained by ASPCE-BASS during the iteration process cannot always have lower validation errors than those from VARPCE, but ASPCE-BASS can still

obtain better results than the VARPCE after having 600 samples.

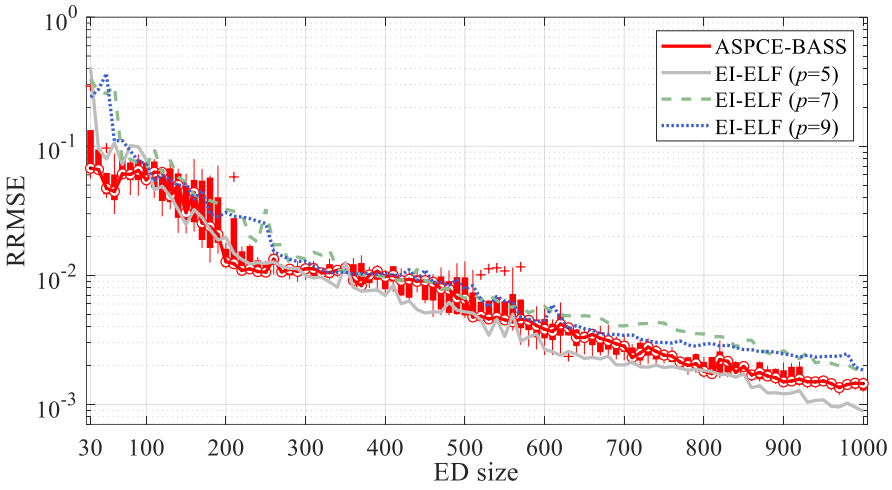



Figure 4.29 RRMSE comparison between EI-ELF and ASPCE-BASS with increasing ED size

The RRMSE results of ASPCE-BASS and EI-ELF with respect to the increase of ED size are compared in Figure 4.29. The results from ASPCE-BASS still show similar convergence trends to those from EI-ELF.

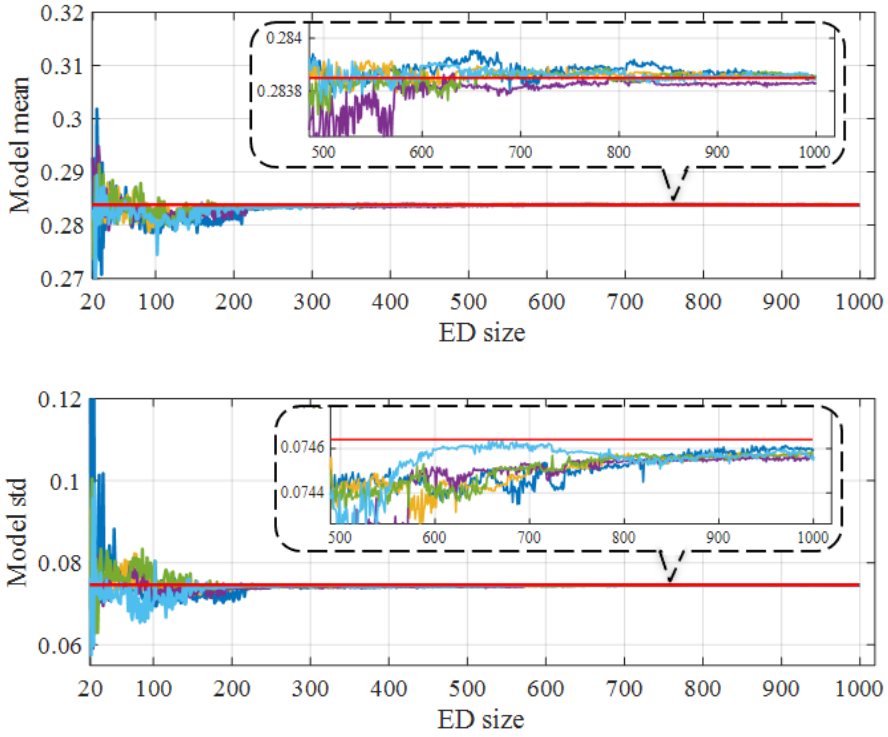



Figure 4.30 Model means (left) and standard deviations (right) with increasing ED size

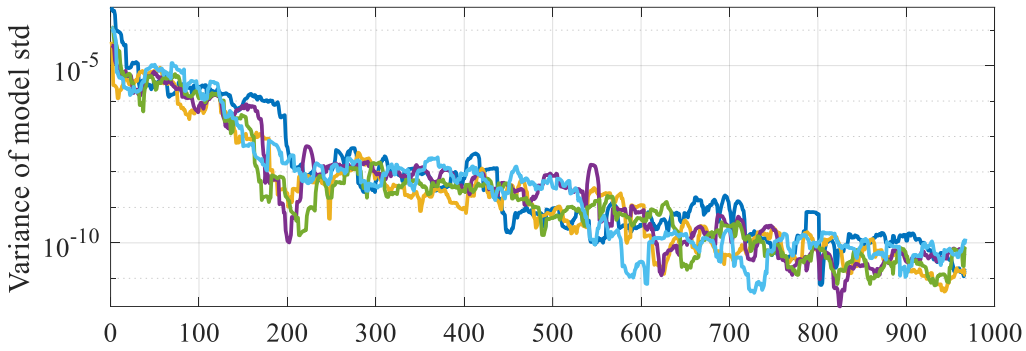



Figure 4.31 Stability evaluation process

The model mean and std values of 5 repeated tests with respect to the increase of ED size by using ASPCE-BASS are shown in Figure 4.30. The red straight lines are the reference values of model mean and std, which are 0.28385 and 0.07465, respectively. With the increase of samples, the trained PCE models gradually converge to the true values, but the converged precision is not as good as the previous two cases. The stability evaluation processes are shown in Figure 4.31. Globally, a downward trend is observed, but the trend is obviously slower than in the previous two cases.

#### 4.3.3.4 Discussions

Here we provide a discussion and comparison over the results associate with the three equation coefficients $\boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)}, \boldsymbol{c}^{(3)}$. Premised on the fact that the compressibility of Equation (4.14) on PCE model decreases with the increased coefficient number $i$ in $\boldsymbol{c}^{(i)}$, the following observations can be made by comparing the results from the three cases.

(1) With the increased coefficient order $i$ in $\boldsymbol{c}^{(i)}$, the PCE models obtained by ASPCE-BASS and EI-ELF with the same ED size show decreased modelling accuracy, but no obvious declination

in modelling precision is observed in the PCE model obtained by VARPCE. The reason is that the modelling accuracy by using sparse representation is affected by the target model complexity. For problem with low compressibility or high complexity, the sparse representation method requires more samples and observations to obtain results with high accuracy.

(2) With the increase of coefficient order $i$, the PCE model obtained by ASPCE-BASS has closer precision to those obtained by VARPCE, but comparable to EI-ELF. This indicates that the sparse representation approach will have no worse performance than the least-squares method in adaptively modelling problems with disparate complexities. Therefore, for a problem with unknown complexity, ASPCE-BASS is the most favourable choice.

## 4.4 Adaptive PCE Technique for Multi-output Problem

### 4.4.1 Method

The proposed ASPCE-BASS method can adaptively build a simple and precise PCE model as surrogate to a target task with high complexity. However, this method is only proposed for single output problem. In real engineering applications, multi-output problem is usually needed to be considered. For instance, the change of structural parameters (e.g., the material elastic modulus and density) will induce the changes of multiple responses (e.g., the displacements at different locations or multiple orders of frequency). Therefore, the sampling method should be able to capitalise on information from multiple outputs. Subsequently, multiple PCE models with desired precisions can

be built simultaneously, and the required samples will less than those needed in modelling outputs separately. In EI-PLHS algorithm, the EI-ELF strategy leverages the output information to instruct the sampling process. Hence, this strategy is modified so that the EI-PLHS algorithm can fit the multi-output problem.

For the sake of simple expression, Equation (A.6) is re-written here.

$$ELF_{EI}(\boldsymbol{\zeta}) = \left(Y(\boldsymbol{\zeta_m}) - \hat{y}(\boldsymbol{\zeta})\right)^2 T(N, P_s) + \sigma^2(\boldsymbol{\zeta}) \tag{4.15}$$

Generally, Equation (4.15) from different outputs can be simply added as the optimisation target to instruct the sampling process. However, the optimisation should achieve balance among different outputs. As is shown, Equation (4.15) comprises of two parts. The first one is the regression error, which is calculated from the predicted mean function, and the second one is the variance function. Basically, the regression error and the variance function will be affected by the scaling effect and not affected by the translation effect. Therefore, by scaling the regression results of different outputs to the same level (e.g., the difference between the maximum value and minimum value of each output is scaled up to the same value), the optimisation is expected to achieve balance among different outputs. As a result, a normalisation coefficient is introduced.

$$NC^j = \hat{y}^j_{max}(\boldsymbol{\zeta}) - \hat{y}^j_{min}(\boldsymbol{\zeta}) \tag{4.16}$$

and the objective function in EI-ELF can be modified as

$$ELF_{multi}(\boldsymbol{\zeta}) = \sum_{j}^{J} \left( \left[ \left( Y^j(\boldsymbol{\zeta_m}) - \hat{y}^j(\boldsymbol{\zeta}) \right) / NC^j \right]^2 T(N, P_s) + \left( \sigma^j(\boldsymbol{\zeta}) / NC^j \right)^2 \right) \tag{4.17}$$

in which $J$ represents the number of modelling tasks in the multi-output problem. By replacing the optimisation function in Equation (4.17) with $ELF_{multi}$, the collected samples and corresponding observations are expected to contribute to the modelling of multiple outputs.

In summary, to apply ASPCE-BASS in multi-output problem, the sample collected in each iteration will take into account the information from all the outputs. Then, each output will correspond to a unique PCE model, and the adaptive basis selection operation will be executed separately for each PCE model. Similarly, the termination criterion is also evaluated separately for each PCE model. If a PCE model is assessed to have desired precision, the training process for this model will be terminated and the output information corresponds to this model will be excluded in the following modelling process. Finally, the algorithm will terminate if the modelling processes for all the outputs finish.

### 4.4.2 Case study

To verify the proposed strategy for multi-output problem, a 2-D truss structure is introduced here, which has been widely studied as a benchmark test in the past (Blatman and Sudret, 2008, 2010; Marelli and Sudret, 2018). As depicted in Figure 4.32, this structure comprises 23 bars and 13 nodes, and the targets are the deflections $u_1, u_2, u_3$ under 6 vertical forces. The multi-output

ASPCE-BASS can build three PCE models to $u_1, u_2, u_3$ simultaneously, and the modelling results will be compared to those obtained by ASPCE-BASS for single output to show the algorithm efficiency.

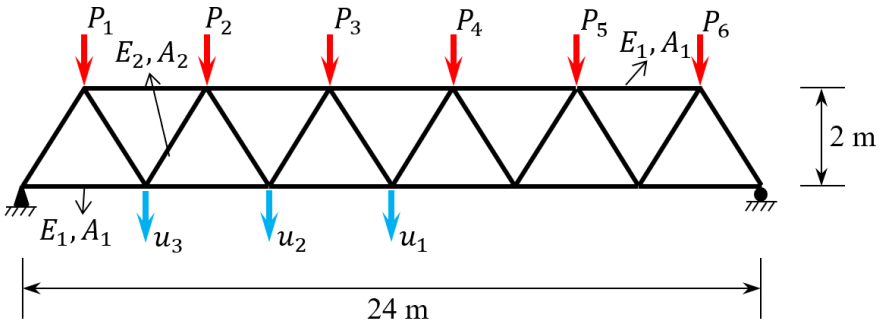



Figure 4.32 Layout of the truss structure

Table 4-1 Input distributions of the truss structure

| Variable | Distribution | Mean | Standard deviation |
|---|---|---|---|
| Elastic moduli $E_1, E_2$ (Pa) | Lognormal | 2.10e11 | 2.10e10 |
| Cross-section area $A_1$ (m2) | Lognormal | 2.0e-3 | 2.0e-4 |
| Cross-section area $A_2$ (m2) | Lognormal | 1.0e-3 | 1.0e-4 |
| Vertical forces $P_1 \sim P_6$ (N) | Gumbel | 5.0e4 | 7.5e3 |

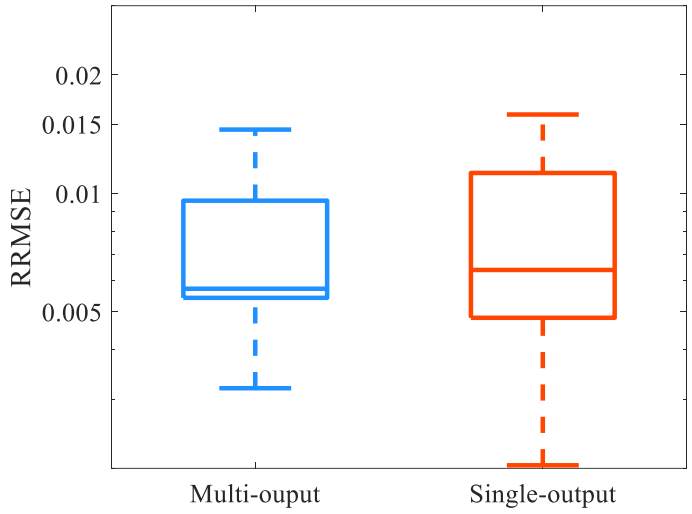
A total of ten input variables are considered in this case, which include six vertical forces $(P_1, P_2, P_3, P_4, P_5, P_6)$, the elastic moduli $(E_1, E_2)$ and the cross-section areas $(A_1, A_2)$ of the horizontal bars and diagonal bars. The distributions of these variables are reported in Table 4-1 (Blatman and Sudret, 2010), and they are assumed to be mutually independent. Three PCE models with Hermite polynomials are built to investigate the relations between three vertical deflections and the ten variables by using ASPCE-BASS. The input variables are transformed into standard Gaussian variables to ease computational burden. The values of each parameter in ASPCE-BASS algorithm are given as shown in Table 4-2.

Table 4-2 Parameter values in ASPCE-BASS

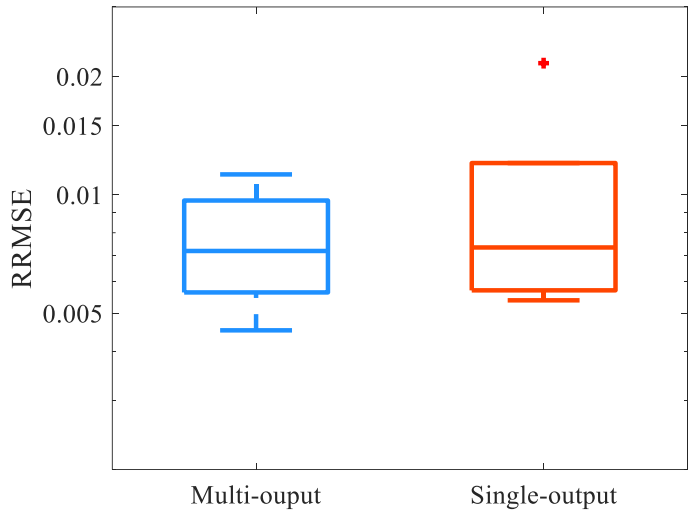
| Name | Initial sample number $N_0$ | Initial truncation degree $P_0$ | Window length $L_a$ | Stability evaluation range $L_b$ | Recovery times $T$ |
|---|---|---|---|---|---|
| Value | 20 | 2 | 15 | 15 | 3 |

In addition, the changes of the PCE model mean and std in successive iterations is employed as the termination criterion, and the threshold to terminate the sampling process is set as $10^{-5}$. To test the algorithm robustness, the modelling process is repeated 10 times. All the settings in the ASPCE-BASS algorithm for single-output and multi-output are the same.

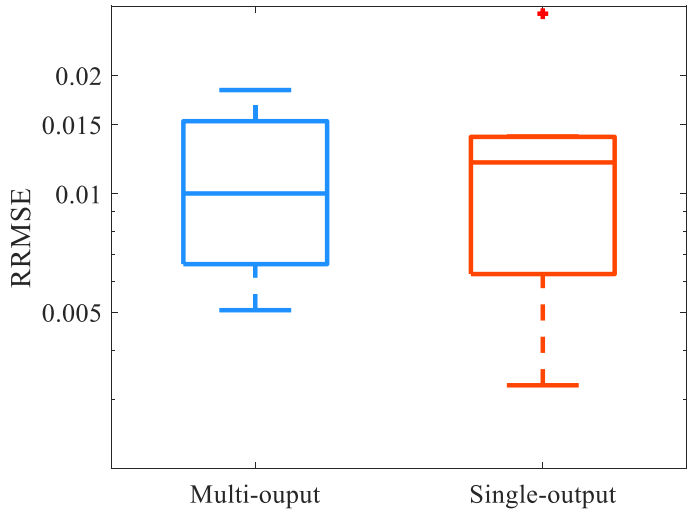
The RRMSE of the obtained PCE models after termination are shown by boxplots in Figure 4.33. To show the modelling efficiency, the number of samples used for single-output modelling and multi-output modelling are compared in Figure 4.34 and Figure 4.35, and the highest polynomial degrees in the trained PCE models are depicted in Figure 4.36.




(a)

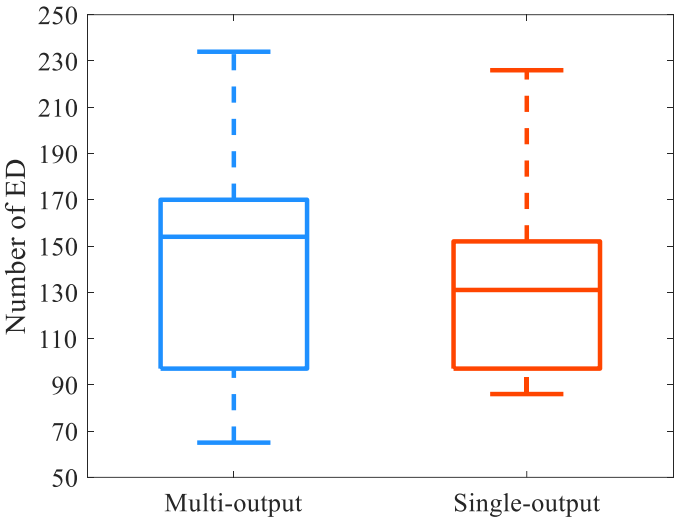



(b)

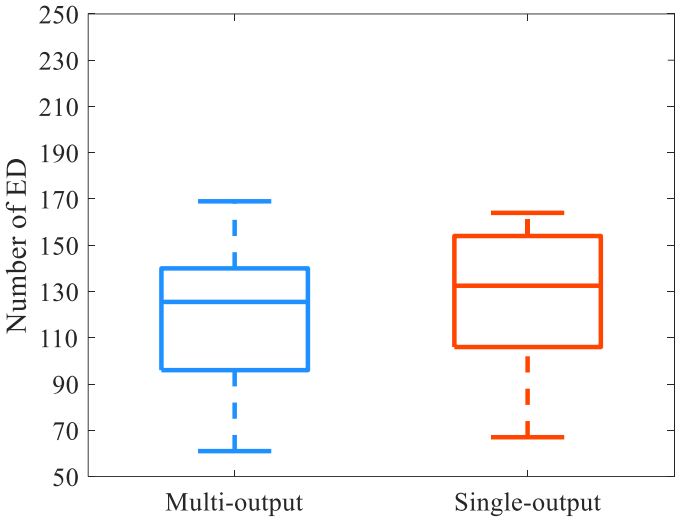



(c)

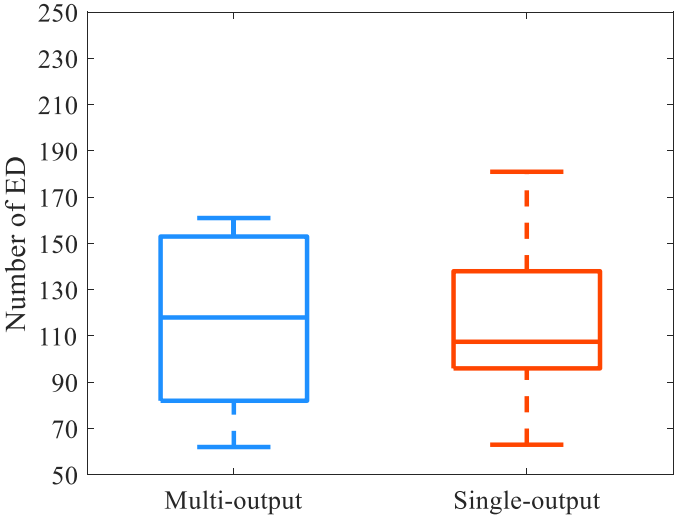
Figure 4.33 Comparison of RRMSE between multi-output and single-output ASPCE-BASS after termination (a) $u_1$; (b) $u_2$; (c) $u_3$




(a)




(b)




(c)

Figure 4.34 Comparison of ED size between multi-output and single-output ASPCE-BASS after termination (a) $u_1$; (b) $u_2$; (c) $u_3$

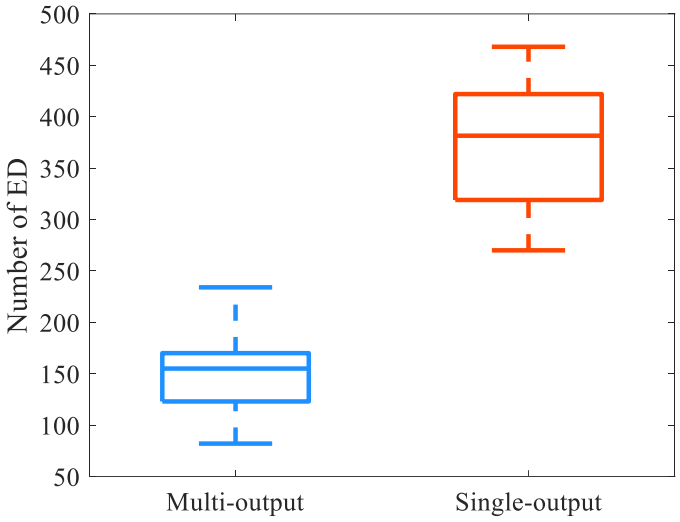
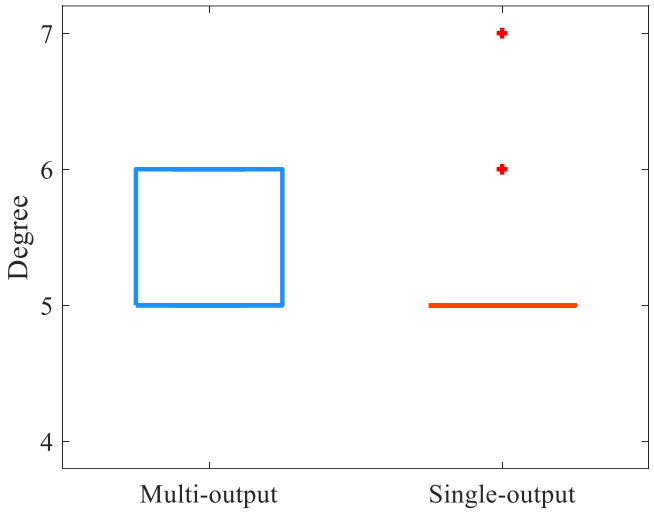
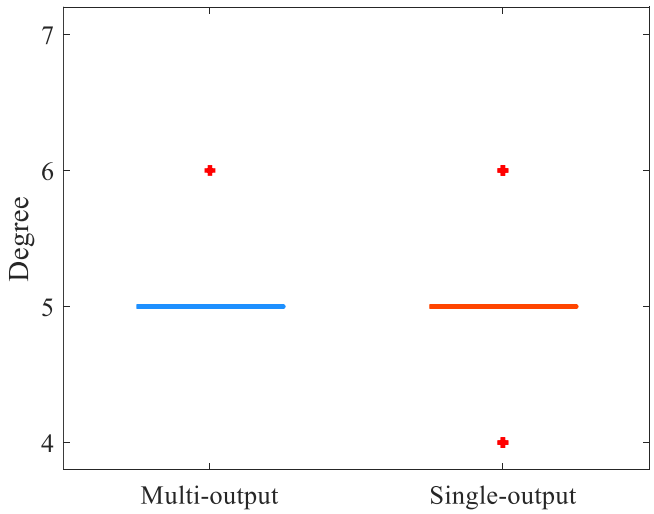


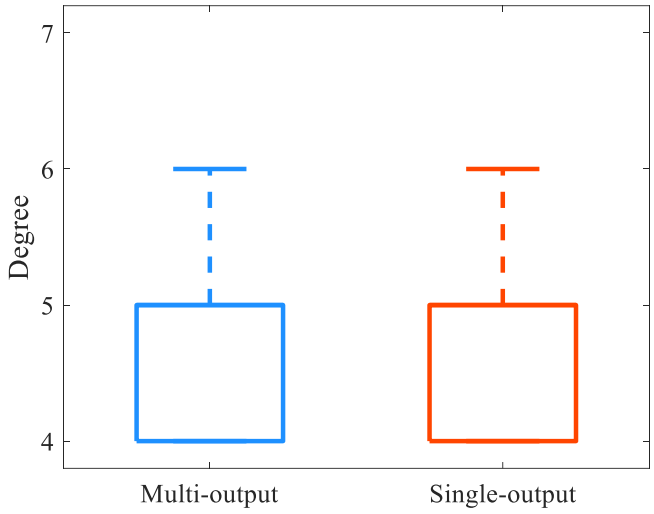

Figure 4.35 The total size of ED required in PCE modelling for three outputs




(a)




(b)




(c)

Figure 4.36 Comparison of degree between multi-output and single-output ASPCE-BASS after termination (a) $u_1$; (b) $u_2$; (c) $u_3$

Obviously, the comparison of RRMSE of PCE models for three outputs indicates that the modelling precisions by using ASPCE-BASS for single-output and multi-output are similar. For each output, the number of samples to terminate the algorithms and the highest polynomial degrees in the trained models are also similar as shown in Figure 4.34 and Figure 4.36. Such phenomena implies that the multi-output modelling technique would not sacrifice the performance of ASPCE-BASS algorithm in modelling each task. Seemingly, the quantity of samples required by ASPCE-BASS for single-output and multi-output are similar. In multi-output modelling, however, the collected samples and corresponding model evaluations are partially shared in the multi-output modelling. The single-output modelling technique trains PCE models for three target outputs separately, so the total number of samples required in modelling three outputs is the summation of sample quantity required for each output. The total numbers of samples in multi-output modelling and single-output modelling are compared in Figure 4.35. It is clear that the samples used by multi-

output modelling are far less than those required by single-output modelling in addressing the multi-output problem. To summarise, the multi-output modelling strategy has desired efficiency than the single-output modelling in solving the multi-output problem while the modelling precision can still be retained.

## 4.5 Summary

In this chapter, a novel adaptive basis selection strategy is developed to automatically choose significant basis terms in the PCE model, which comprises of three core operations, basis expansion, pruning and refinement. The PCE model will begin with a low truncation degree, and basis terms with higher degree can be gradually added to the PCE model to achieve higher training precision by performing basis expansion. Then, the insignificant basis terms will be prudently truncated during the modelling process to lighten the training burden through the employment of basis pruning and basis refinement operations. By combining this basis selection strategy with the sparse representation method and the hybrid sequential sampling approach, an adaptive PCE modelling framework, named ASPCE-BASS, is built. It is worth mentioning that a new hybrid sequential sampling approach, EI-PLHS, is developed in this chapter, which is more appropriate to be used for building the fully adaptive modelling framework than the coherence-entropy method. In the proposed ASPCE-BASS, a two-loop framework is built, in which the sequential sampling process is introduced as the inner loop and the adaptive basis selection process is developed as the outer loop. A novel stability evaluation approach is presented to determine which loop should be

performed in each iteration, and this strategy enables the sparse representation to be integrated in the adaptive modelling framework. Finally, the proper training sample set and truncation degree (or significant basis terms) are both automatically determined.

To illustrate the efficiency and effectiveness, the proposed ASPCE-BASS is compared with two existing approaches by validating on three benchmark functions. One of the approaches is an adaptive PCE modelling technique named VARPCE, in which a least-squares regression method rather than the sparse representation is used for regression calculation. The other one is a non-basis-adaptive algorithm (sequential sampling-only algorithm) called EI-ELF. It can be seen from the results compared between ASPCE-BASS and VARPCE that the proposed method shows better convergence rate and modelling accuracy. ASPCE-BASS requires less samples to explore more basis terms with high degree than VARPCE, which demonstrates that the sparse representation approach is impressive in dealing with problem with high input dimension or high truncation degree. Compared with the sequential sampling-only method, EI-ELF, the proposed ASPCE-BASS shows similar performance under most circumstances, but ASPCE-BASS can automatically determine the best truncation degree. Meanwhile, the computational burden of the regression calculation is lightened by removing the insignificant basis terms in ASPCE-BASS. Thus, the effectiveness of the proposed adaptive basis selection strategy is demonstrated. Moreover, through validating on functions with disparate complexity, it is shown that the proposed ASPCE-BASS which employs sparse representation outperforms on problems with high compressibility than those with low

compressibility. Thus, the proposed method has a better practicability than the other two methods when there is no prior knowledge about the problem complexity.

Furthermore, the proposed ASPCE-BASS is extended to solving multi-output problem. The proposed multi-output modelling strategy is able to build a PCE model for each output of interest simultaneously. By validating on an engineering case with comparison to building each PCE model separately, it is demonstrated that the multi-output modelling can help reduce the size of training dataset while retaining the modelling accuracy.

# CHAPTER 5
# TRANSFER LEARNING BASED MULTI-FIDELITY PCE MODELLING METHOD

## 5.1 Introduction

By using the adaptive modelling approaches proposed in the previous chapters, we can build PCE surrogate models as alternative to the complex simulation model for analysis with low cost. In structural engineering, however, the simulation model will inevitably deviate from the actual structure due to the existence of modelling error. It is preferred to train surrogate models of the real structure rather than the simulation model to achieve more promising results for further analysis. In the damage identification for bridge structures, for example, the dynamic characteristics (frequency and mode shape) reflect the inherence properties of a bridge. Any damage on the bridge will cause the variations of frequencies, which can be leveraged to identify the damage. Nevertheless, the variations of structural responses will be affected by not only damages but also environmental effects, and the response variations caused by environmental effects can often exceed the variations caused by damage (Bao et al., 2012; Erazo et al., 2019). It is easy to receive incorrect damage identification results depending on the physical model without simulating the environmental effects

accurately. Data-driven approaches can overcome this challenge, but they usually require massive observations. It was argued that data from the damaged structure is scarce (Zhu et al., 2019). As a result, the combinational methods emerged, which propose to update the physical model using scarce measurements, such as model updating, digital twin and Multi-Fidelity (MF) modelling (Bigoni and Hesthaven, 2020; Das and Debnath, 2018; Diez-Olivan et al., 2019; Gregory et al., 2019). Among them, the MF modelling technique has attracted great attentions recently.

In brief, MF modelling is a technique to improve the traditional surrogate modelling method by leveraging training data with different levels of fidelity. A small number of high-cost High-Fidelity (HF) data is utilised to update a Low-Fidelity (LF) surrogate model trained by LF data. Subsequently, the updated model will achieve balance between the training cost and modelling accuracy (Ng and Eldred, 2012). In this study, the physical model (or its surrogate model) is regarded as LF model, and observations from the real structure are the HF data. However, there are also limitations/premises in MF modelling. The LF model should at least capture the global trends of the target system. If the LF model only captures a part of the local trends, the other local trends which are mistakenly represented by the LF model cannot be precisely updated with insufficient HF data. Besides, the HF data should spread over the input domain. Under some specific circumstances, the HF samples are required to be a subset of the LF samples (Liu et al., 2018a).

As discussed in Chapter 2, temperature is the most influential environmental factor on the bridge structure, and it is hard to be precisely simulated in the physical model due to its intricate

mechanism in reality. Spontaneously, we would like to eliminate the temperature-induced modelling error in the LF physical model by using the MF modelling technique. However, traditional MF modelling technique may fail to help solve the damage identification problem because the HF measurements are hard to spread over the input domain, i.e., measurements from damaged bridge are difficult to be collected. To overcome this difficulty, the concept of Transfer Learning (TL) is introduced to the Multi-Fidelity PCE (MFPCE) modelling technique to develop a TL-based MFPCE (TL-MFPCE) approach. The idea is to freeze the coefficients of basis terms associated with the damage parameters in the Low-Fidelity PCE (LFPCE) model and to update the coefficients of temperature-related basis terms to eliminate the temperature-induced modelling error. Firstly, the LFPCE models to describe the bridge frequencies concerning damage parameters and temperature are established with data from the Finite Element (FE) model by using ASPCE-BASS. Regarding the measurements from real structure in healthy condition as HF data, the coefficients of basis terms associated with temperature are then updated by using an MF modelling function. Ultimately, the formulated MFPCE models can not only learn from the FE model about the knowledge of damage but also study from the real measurements to correct the temperature-induced modelling error. For validation, the proposed TL-MFPCE method is compared with the traditional MFPCE method on a numerical bridge model and an experimental beam model.

## 5.2 The proposed method

### 5.2.1 MFPCE method

Traditionally, there are three different modelling forms in the MF technique, the additive correction, the multiplicative correction and the combinational correction (Ng and Eldred, 2012). By assuming that the relation between the LF data and HF data can be represented by a PCE function, the additive correction function and multiplicative correction function are represented as (Ng and Eldred, 2012):

$$\Delta PCE_1(\boldsymbol{\xi}) = R_{high}(\boldsymbol{\xi}) - PCE_{low}(\boldsymbol{\xi}) \tag{5.1}$$

and

$$\Delta PCE_2(\boldsymbol{\xi}) = \frac{R_{high}(\boldsymbol{\xi})}{PCE_{low}(\boldsymbol{\xi})} \tag{5.2}$$

respectively. Then

$$PCE_{multi}(\boldsymbol{\xi}) = PCE_{low}(\boldsymbol{\xi}) + \Delta PCE_1(\boldsymbol{\xi}) \tag{5.3}$$

or

$$PCE_{multi}(\boldsymbol{\xi}) = PCE_{low}(\boldsymbol{\xi}) \cdot \Delta PCE_2(\boldsymbol{\xi}) \tag{5.4}$$

where $PCE_{low}(\boldsymbol{\xi})$ and $R_{high}(\boldsymbol{\xi})$ represent the LFPCE model and the unknown HF model, respectively; $\Delta PCE_1(\boldsymbol{\xi})$ and $\Delta PCE_2(\boldsymbol{\xi})$ are two corrective PCE functions. There might be several levels of fidelity in data, but the case with only two levels of fidelity is considered in this study. With the HF data, the MF model can be obtained by evaluating the unknown coefficients in the

corrective PCE function. A crucial criterion in MFPCE modelling is that the polynomial terms in the corrective function must be a subset of those in the LFPCE model (Palar et al., 2016). The reason is that the HF data can be used for training is basically far less than the LF data. The overfitting issue may occur if the corrective function has a higher truncation degree than the LFPCE model. In addition, a combinational form was defined based on the above two forms (Ng and Eldred, 2012):

$$PCE_{multi}(\boldsymbol{\xi}) = \gamma(PCE_{low}(\boldsymbol{\xi}) + \Delta PCE_1(\boldsymbol{\xi})) + (1-\gamma)\Delta PCE_2(\boldsymbol{\xi}) \cdot PCE_{low}(\boldsymbol{\xi}) \tag{5.5}$$

where $\gamma \in [0,1]$ denotes a weight parameter that determines the proportion of the additive correction and the multiplicative correction. The combinational correction form will outperform other correction forms if there are large discrepancy between data with different fidelity, while the number of parameters and coefficients that should be solved increase and the model structure is much more complicated. Therefore, the combinational correction form was generally not recommended.

Apart from the above-mentioned three correction forms, an autoregressive modelling form originated from the co-Kriging technique is also introduced (Kennedy and O'Hagan, 2000), which can be regarded as a special form of Equation (5.5):

$$PCE_{multi}(\boldsymbol{\xi}) = \rho PCE_{low}(\boldsymbol{\xi}) + \Delta PCE_3(\boldsymbol{\xi}) \tag{5.6}$$

It can be deemed that the weighted multiplicative correction function in Equation (5.5) is simplified as a single weight coefficient $\rho$, so the number of unknown coefficients to be solved decreases. In

this study, the autoregressive form is used for MFPCE modelling.

To train an MFPCE model, several procedures were developed (Park et al., 2017). A basic method is to train the corrective function based on the discrepancy or ratio between the LF and HF data at the overlapped sample points. The corrective function and LF surrogate model are then combined to constitute the MF model. Clearly, a limitation of this method is that the HF samples should be a subset of LF samples. In 2016, an improvement strategy was introduced to remove the limitation (Berchier, 2016). Taken the additive correction (Equation (5.3)) as an example, the steps of the improved strategy are shown in Figure 5.1:

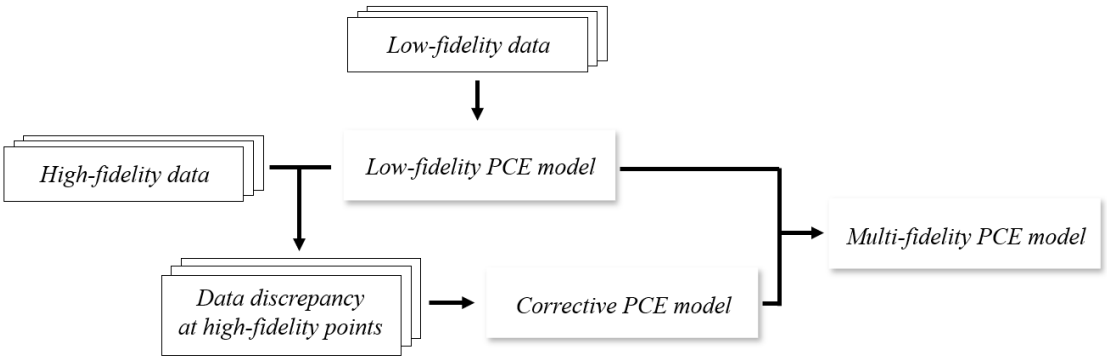



Figure 5.1 Procedure of training MFPCE model (additive correction)

As can be seen, the LF data is first generated to train an LFPCE model. Then the discrepancies between the HF data and evaluations from the LFPCE model at the HF sample points are calculated, which are used to train the additive correction PCE model. Finally, by adding the corrective PCE model to the LFPCE model, an MFPCE model is obtained. Berchier (2016) also demonstrated that this approach has a higher modelling accuracy than the basic method. Similarly, this approach can

be extended to solve Equation (5.6) with autoregressive form. The training process is shown in Figure 5.2.

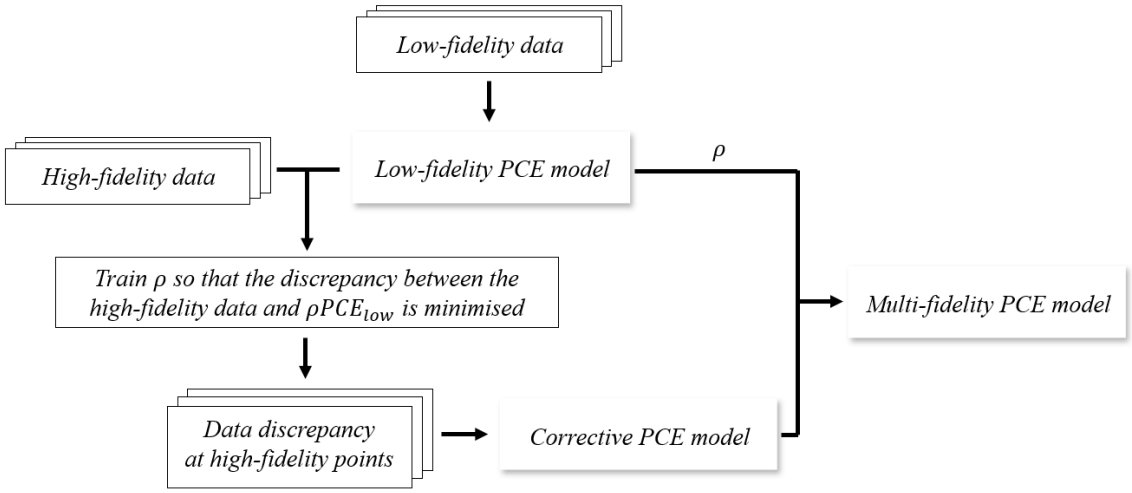



Figure 5.2 Procedure of training MFPCE model (autoregressive form)

Compared to the training process of additive correction, only one extra step is added to train the weight coefficient $\rho$. Here, the weight coefficient $\rho$ and the corrective model $\Delta PCE_3(\boldsymbol{\xi})$ are trained separately. Yet, it was demonstrated by Park et al. (2017) that training $\rho$ with the additive correction function concurrently can achieve a more accurate MF modelling result. Even if this conclusion was drawn based on the Kriging method in their study, we believed that this operation is also valid when the PCE technique is used. Inspired by this, an improved training process for MFPCE with autoregressive form is proposed in this study. By expanding the corrective function, Equation (5.6) can be re-written as:

$$PCE_{multi}(\boldsymbol{\xi}) = \rho PCE_{low}(\boldsymbol{\xi}) + \boldsymbol{\Psi}(\boldsymbol{\xi})\boldsymbol{c}_3 = [\boldsymbol{\Psi}(\boldsymbol{\xi}), PCE_{low}(\boldsymbol{\xi})] \cdot \begin{bmatrix} \boldsymbol{c}_3 \\ \rho \end{bmatrix} \tag{5.7}$$

Hence, the unknown coefficients $\boldsymbol{c}_3$ and regression parameter $\rho$ can be solved via the BCS method in parallel.

### 5.2.2 TL-based MFPCE modelling

Machine learning technologies have been presented for many engineering tasks including classification, regression and clustering. However, the data limitation in some practical circumstances hinders the application of many of them, e.g., the training and test data in a task may change their feature space and distribution over time. When the distribution varies, most of the previously trained statistical models are needed to be re-built to match the new data, so more cost and time will be paid. For the sake of cost saving, the concept of TL was developed (Pan and Yang, 2010). If some tasks have no sufficient data for training precise model while other different but similar tasks have trained models, TL proposes to re-train a model from an existing task for a new task. So, the training data required for the new task can be reduced. To facilitate understanding, the definition of TL is provided: Given a source domain $\mathcal{D}_S = \{\mathcal{X}_S, P(\mathcal{X}_S)\}$ and source task $\mathcal{T}_S$, a target domain $\mathcal{D}_T = \{\mathcal{X}_T, P(\mathcal{X}_T)\}$ and learning task $\mathcal{T}_T$, TL aims to help handle the target task $\mathcal{T}_T$ with data from the target domain in $\mathcal{D}_T$ using the knowledge from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$. Currently, TL has been applied in many areas, especially in image recognition using convolutional neural network (Pan and Yang, 2010). In a typical convolutional neural network, for example, low-level layers can be trained on an existing large database to extract low-level features

such as edges, corners and shapes, which are not specific for task $\mathcal{T}_S$. Therefore, these layers constitute a general feature extractor with well-trained parameters, which can be frozen and transferred to a new model. Then, the parameters in high-level layers are re-trained for the target task based on the data from target domain $\mathcal{D}_T$. This kind of TL is called "parameter-based TL" (Pan and Yang, 2010; Chen et al., 2021).

As aforementioned, we would like to use MF modelling technique to eliminate the modelling error caused by temperature in the LFPCE model, but the HF measurement data cannot spread over the input domain, that is, the observations from bridge under damaged condition are generally deficient. Subsequently, the MFPCE model most likely fails to represent the real bridge structure under damaged condition. As a solution, a TL-based MFPCE (TL-MFPCE) approach is proposed. From the perspective of TL, the physical model can be regarded as the source domain, from which the target responses of interest are generated to train the LFPCE model for the source task, i.e., the relation among the damage parameters, temperature, and the bridge frequency in the physical model. The real structure will be the target domain, and the real observations, or called HF data, are collected from the target domain to combine with the LFPCE model to handle the target task, i.e., the relation among the damage parameters, temperature, and bridge frequency in the real structure. The LFPCE model has been trained to describe the relation between the responses and damage parameters. This knowledge can be retained and transferred to the MF model. Then, the incorrect relation between the responses and temperature in the LFPCE model can be updated by using the

HF data. In other words, the LFPCE model and the MFPCE model can share the coefficients of basis terms that related to the damage parameters, and the HF data can be used to update the coefficients of basis terms associated with temperature. As a result, the modelling error can be updated, while the knowledge of damage from the source domain can be kept and leveraged. In the next section, the mathematical principle of TL-MFPCE will be provided.

### 5.2.3 TL-MFPCE for surrogate modelling under varying temperature

The LFPCE model built by ASPCE-BASS is first written here:

$$F_{low}(\boldsymbol{q},\boldsymbol{T}) = \sum_{i=1}^{P} c_i \psi_i(\boldsymbol{q},\boldsymbol{T}) \tag{5.8}$$

where $F_{low}$ represents the response of interest that is built as the LF model; $\boldsymbol{q}$ denotes a vector of damage parameters; $\boldsymbol{T}$ is a vector of temperature variables. In this study, $F_{low}$ is the frequency, and $\boldsymbol{q} = (\boldsymbol{X} - \boldsymbol{X_d})/\boldsymbol{X}$, in which $\boldsymbol{X}, \boldsymbol{X_d}$ are the parameters of bridge components under healthy and damaged conditions, respectively. The traditional autoregressive MF technique intends to update all the basis terms with expression as given in Equation (5.9):

$$F_{multi}(\boldsymbol{q},\boldsymbol{T}) = \rho F_{low}(\boldsymbol{q},\boldsymbol{T}) + \Delta PCE(\boldsymbol{q},\boldsymbol{T}) + \epsilon \tag{5.9}$$

in which $\epsilon$ represents the measurement noise, which is assumed as a zero mean Gaussian distribution. Due to the lack of data from structure under damaged condition, the estimations of the corrective function $\Delta PCE(\boldsymbol{q},\boldsymbol{T})$ and the weight coefficient $\rho$ would be inaccurate. Therefore, we

re-written Equation (5.8) by re-organising the basis terms with different input variables:

$$F_{low}(\boldsymbol{q},\boldsymbol{T}) = f(\boldsymbol{q}) + h(\boldsymbol{q},\boldsymbol{T}) + g(\boldsymbol{T}) \tag{5.9}$$

in which $f(\boldsymbol{q})$ denotes the cluster of basis terms that are only related to $\boldsymbol{q}$; $g(\boldsymbol{T})$ is the cluster of basis terms that are only associated with $\boldsymbol{T}$; and $h(\boldsymbol{q},\boldsymbol{T})$ represents the cluster of basis terms that have interaction effect between $\boldsymbol{q}$ and $\boldsymbol{T}$. Since $\boldsymbol{q}$ is calculated from the parameters of the bridge components, $h(\boldsymbol{q},\boldsymbol{T})$ implies the part of responses dominated by the parameters change caused by temperature, e.g., the temperature-induced elastic modulus change. $g(\boldsymbol{T})$ denotes the part of responses dominated by the thermal expansion and temperature-induced boundary condition change. Generally, the elastic modulus can be measured by ultrasonic methods or stress-strain diagram (Y. Xia et al., 2012). But the thermal expansion and the boundary condition change are difficult to be recognised because they will affect each other (Han et al., 2021). Thus, the basis terms in $f(\boldsymbol{q})$ and $h(\boldsymbol{q},\boldsymbol{T})$ with trained coefficients can be frozen and transferred to the MF model, and $g(\boldsymbol{T})$ should be updated with the HF data to eliminate the modelling error. By employing Equation (5.6), the TL-MFPCE model is:

$$F_{multi}(\boldsymbol{q},\boldsymbol{T}) = f(\boldsymbol{q}) + h(\boldsymbol{q},\boldsymbol{T}) + \rho \cdot g(\boldsymbol{T}) + \Delta PCE(\boldsymbol{T}) + \epsilon \tag{5.11}$$

With the HF data $\{(\widetilde{\boldsymbol{q}},\widetilde{\boldsymbol{T}}),\widetilde{\boldsymbol{F}}\}$, this model can be trained:

$$\widetilde{\boldsymbol{F}} - f(\widetilde{\boldsymbol{q}}) - h(\widetilde{\boldsymbol{q}},\widetilde{\boldsymbol{T}}) = \left[g(\widetilde{\boldsymbol{T}}), \boldsymbol{\Psi}(\widetilde{\boldsymbol{T}})\right] \cdot \begin{bmatrix}\rho \\ \boldsymbol{c}_c\end{bmatrix} + \epsilon \tag{5.12}$$

which has the same expression form to Equation (3.10). Here, $\boldsymbol{c}_c$ is the unknown coefficients of $\Delta PCE(\boldsymbol{T})$. Hence, the BCS method can be employed to garner the solutions of $\rho$ and $\boldsymbol{c}_c$. Moreover,

by assuming that the measurement noise follows the zero mean Gaussian distribution $\epsilon \sim N(0, \beta^{-1})$ like Equation (3.14), the variance of the noise, $\sigma^2 = \beta^{-1}$, can be estimated by using BCS delineated in Section 3.2.2.

So far, an issue is still worthy to note that the truncation degree of the corrective PCE function should be properly selected to avoid overfitting. Here, a popularly used Cross Validation (CV) approach is employed to help choose the degree. The leave-one-out CV errors of the MF models under different degrees will be recorded. When the CV error cannot decrease with the increase of degree, the corresponding degree value will be treated as the best one. To summarise, the procedure of TL-MFPCE is given in **Algorithm 5**:

---

**Algorithm 5.** TL-MFPCE

---

**Input:** LF data from FE model (source domain); HF data from real structure (target domain) in healthy condition with different temperatures;

---

**Initialisation:** Train LFPCE models by using ASPCE-BASS (trained models from source domain); Record the highest degree of the trained LF model as $P_h$.

---

**In $i$th iteration:**

1. Define the truncation degree of the corrective function $\Delta PCE(\boldsymbol{T})$ as $i$;
2. Divide the basis terms in LFPCE models according to the variables in the basis term;
3. Freeze the structural parameter-related basis terms according to the parameter-based TL;
4. Update the basis terms that are not frozen based on Equation (5.11) and use BSC to calculate the unknown parameters;
5. Calculate the CV error and recorded.
6. If $i > P_h$, quit iteration, and compare the CV error under different degree values defined for $\Delta PCE(\boldsymbol{T})$ and choose the one with no more declined CV error.

---

To summarise, the merits of this TL-MFPCE technique are: (1) The requirement on the HF data is reduced so that only data from structure in the healthy condition are needed; (2) Freezing

some of the trained basis terms in the LFPCE model ensures that the features learnt from the FE model can be retained. Thus, the MF surrogate model benefits from the knowledge in FE model when no measurement data from the damaged structure is available. (3) The variance of the measurement noise can be evaluated by using the BCS method, which is useful in the next chapter.

Compared with the traditional multi-fidelity modelling methods based on other surrogate modelling techniques such as GPR and ANN, the advantage of the proposed method based on PCE is: The trained PCE can be easily divided according to the variables contained in each basis term, so the parameter-based TL can be used. On the contrary, the variables in GPR are coupled with each other in the expressions and the hyperparameters are learnt in a global way, which is not appropriate for addressing the problem in this thesis that the real measurements are absent in some portions of the definition domain. For ANN, it is convenient to employ TL by choosing the layers in ANN to be frozen or updated, but updating the model based on different input variables is still tricky because the trained ANN model is complex and unexplainable.

## 5.3 Case Studies

In this section, a numerical bridge and an experimental beam are explored to verify the proposed TL-MFPCE modelling method under different temperature distributions. The numerical bridge is applied with gradient temperature along vertical direction, and the experimental beam works under the uniform temperature.

### 5.3.1 Brief descriptions of two cases

#### 5.3.1.1 Numerical case of a three-span continuous beam bridge

The numerical case is a three-span continuous beam bridge with single-box cross section. The layout of this bridge is given in Figure 5.3 and Figure 5.4. As can be seen, the second support counted from the left end constrains the lateral displacement of the beam. For analysis simplicity, the beam is equally divided into 26 parts with each of 5 m long, and the elastic modulus of each part is regarded as a structural parameter, as shown in Figure 5.3. To note, the structural material is temperature-related in this simulation. The elastic modulus under 20℃ is $3.55\times10^{10}$ $Pa$, and the relation between the elastic modulus and temperature are depicted in Figure 5.5 (Jiao et al., 2014). Besides, the mass density and the thermal expansion coefficient of the structural material are 2549 $kg/m^3$ and $1\times10^{-5}$, respectively. The simulation is realised by building the FE model with ANSYS.

In this case, the temperature field is applied on this structure with vertical gradient, and the temperature along the lateral direction of bridge is assumed identical. Given the temperatures of the top and bottom surfaces, the distribution of temperature field along the vertical direction of the structure is simply assumed as linear (Xia et al., 2018). Therefore, the temperatures of the top and bottom surfaces are two independent variables in this case, which will vary between 0℃ and 50℃. In addition, the temperature of the bottom surface will always lower than that of the top surface.

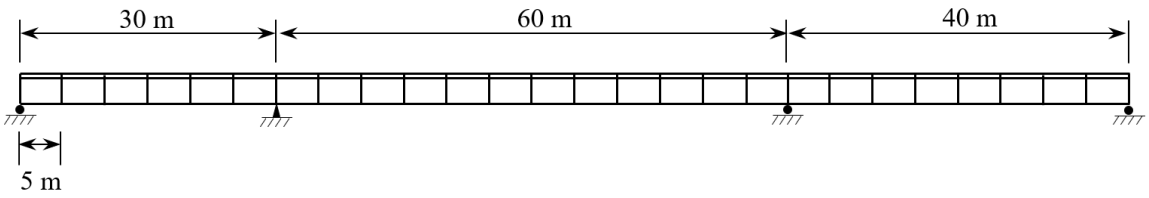



Figure 5.3 Layout of three-span continuous beam bridge

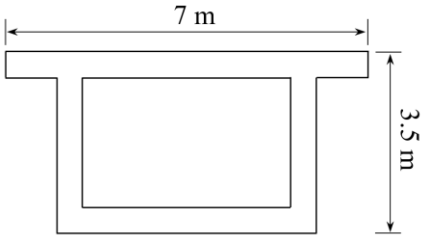



Figure 5.4 Layout of the cross section

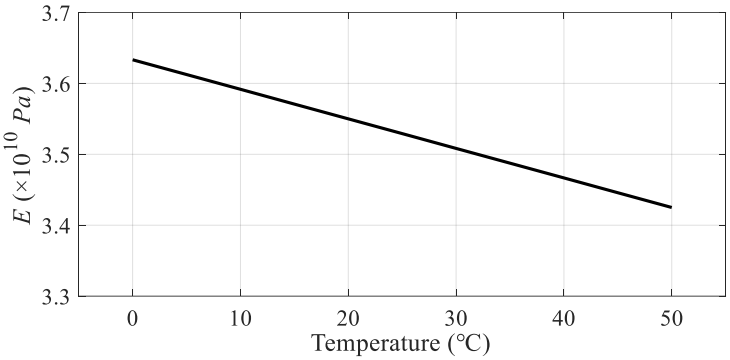
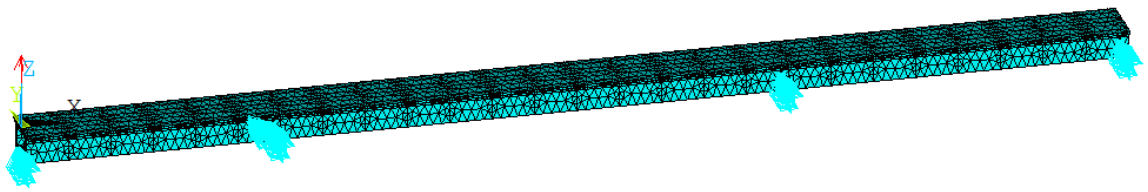


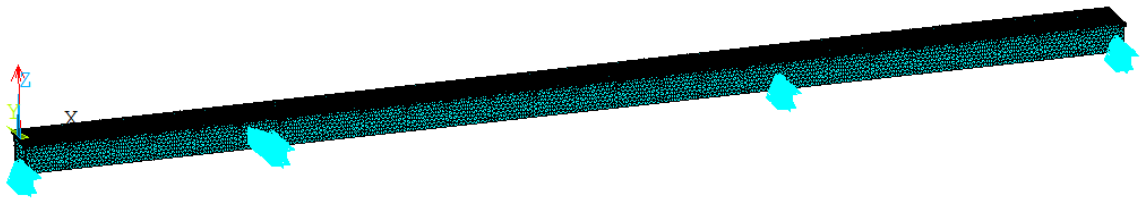

Figure 5.5 Elastic modulus versus temperature

To validate the proposed TL-MFPCE method, the LF model and HF model should be established separately. Here, we use different mesh densities to build models with different fidelities, as shown in Figure 5.6. The LF model is meshed with element size of 1.0 m, while the HF model is meshed with element size of 0.5 m. In addition, the HF model is added with temperature-related axial forces to simulate the change of boundary condition, while the LF model has no axial force under any temperature to simulate its discrepancy with the HF model. As is portrayed in Figure 5.7, two pairs of concentrated forces are applied at both ends of the HF model. One pair is imposed at the central axis of the top surface and the other one is applied at the central axis of the bottom

surface. The values of two forces, $F_t$ and $F_b$, are supposed to have functional relations to the temperature of top and bottom surfaces, respectively. In the past studies, it was found that the temperature-induced boundary condition change can cause a nonlinear variation of the model responses, and the frequencies may have changes over 10% (Peeters et al., 2001; Zhou and Song, 2018). Hence, a nonlinear temperature-force function is designed in this case, which is displayed in Figure 5.8. Such relation may not exist in practical, but it can be used to verify the proposed method without loss of generality. By applying this force relation, the frequencies of this bridge could have a maximum variation of 10% with temperature changing.



(a)



(b)

Figure 5.6 FE model of the bridge with mesh size (a) 1.0 m; (b) 0.5 m

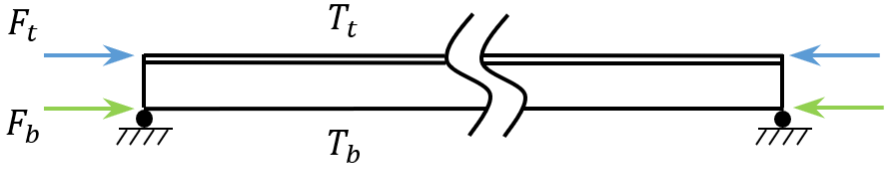



Figure 5.7 Simple drawing of the axial forces on the bridge

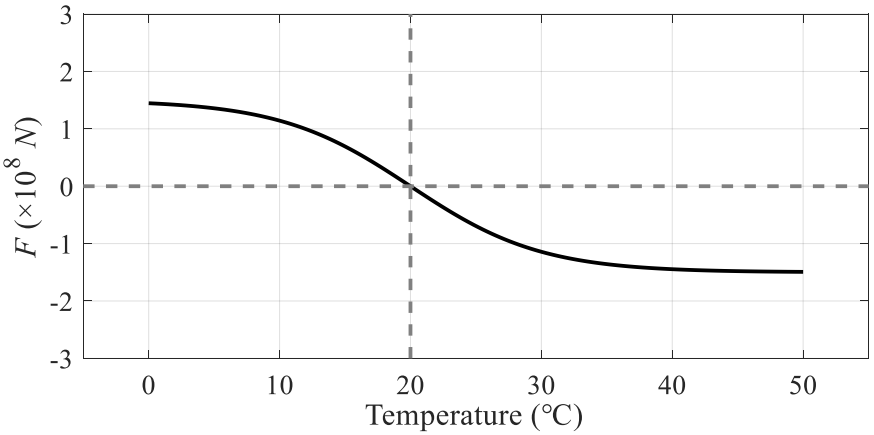
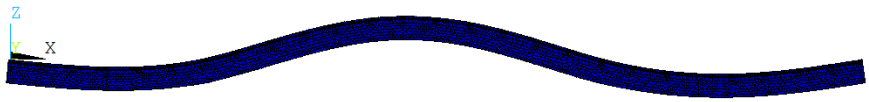
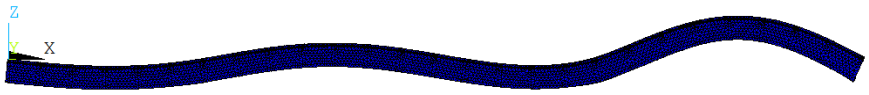



Figure 5.8 Boundary condition change (axial force) versus temperature

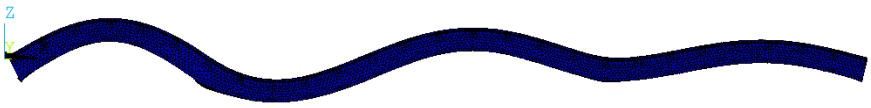
The first 5 frequencies corresponding to the vertical mode shapes are of interest. The frequencies and the mode shapes from the HF model at 20℃ (no axial force) are depicted in Figure 5.9.



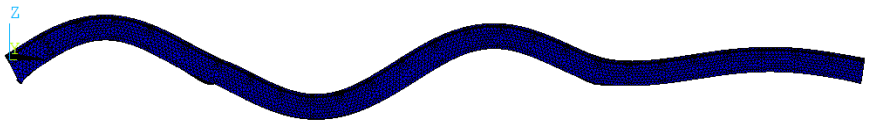

(a) Mode 1 (3.0150 Hz)



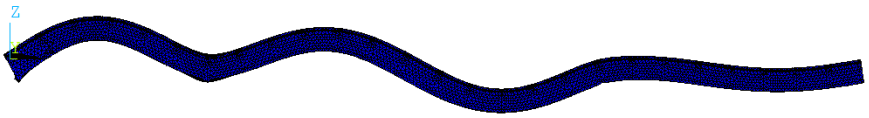

(b) Mode 2 (5.6041 Hz)




(c) Mode 3 (7.4741 Hz)




(d) Mode 4 (9.6388 Hz)



(e) Mode 5 (10.5007 Hz)

Figure 5.9 First 5 frequencies corresponding to the vertical mode shapes

#### 5.3.1.2 Experimental case of a two-span continuous beam

The experimental case is a two-span continuous beam with a 38.02 mm×9.55 mm rectangular cross section, which is shown in Figure 5.10. The lengths of two spans are 900 mm and 1100 mm, respectively. For simplicity in analysis, the beam is equally divided into 20 elements, and the inertia moment of each element is a structural parameter that can be used to characterise structure damage. The material of the beam is aluminium, and the mass density and thermal expansion coefficients are estimated as 2669.7 $kg/m^3$ and 2.34×$10^{-5}$, respectively. Moreover, the elastic modulus of the material is temperature-related, and the relation is depicted in Figure 5.11.

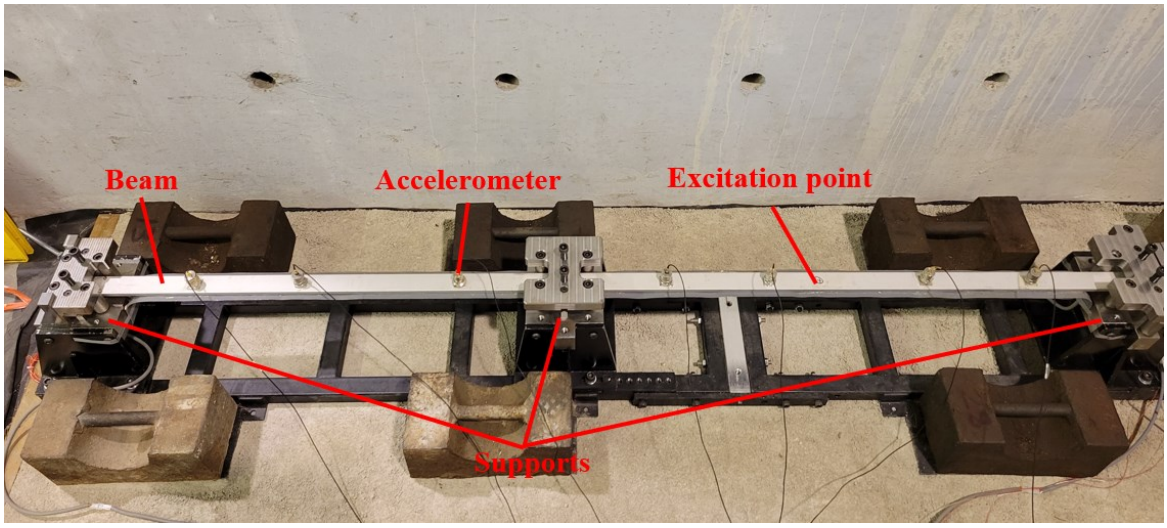



Figure 5.10 Setup of the experimental case

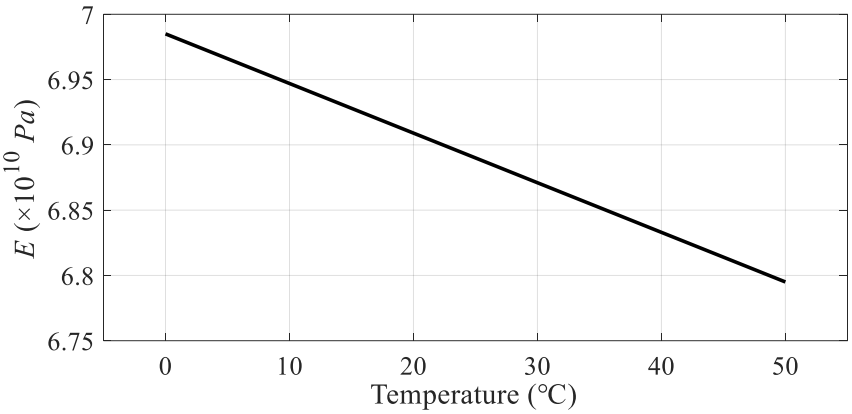
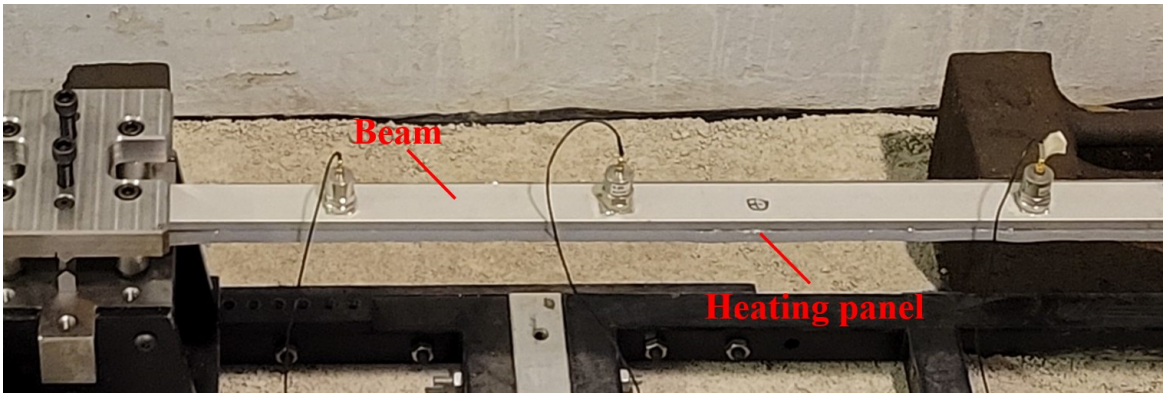



Figure 5.11 Elastic modulus versus temperature

To apply different temperatures, a heating panel was sticked on the bottom of the beam. The temperature was controlled between 25℃~50℃ by a control box, and the temperature on the beam was measured by a resistance thermometer PT100. The heating panel and the control equipment is shown in Figure 5.12.




(a)


(b)

Figure 5.12 (a) Installed heating panel and (b) Temperature control equipment

Since the frequencies were the targets of interest, the impact hammer modal test was applied, and the accelerometers were used to measure the structural signals. In this test, we used 7 accelerometers to ensure the robustness of the identified results, and the sensor installation positions have been shown in Figure 5.10. There is a magnetic base in each accelerometer to ease the installation. As the material of the beam used in this experiment is aluminium, several small steel plates are sticked on the beam so that the accelerometers can be easily mounted. The operating temperature range of these accelerometers is -74℃~250℃, so they are workable in this test. The data acquisition and control system as well as the impact hammer are shown in Figure 5.13.


(a)


(b)


(c)

Figure 5.13 (a) Data acquisition system (EDX-100A and computer); (b) Hammer; (c) Signal amplifier

According to the above information, an FE model of this beam is built as the LF model, which is shown in Figure 5.14. The first 5 frequencies corresponding to the vertical mode shapes are of interest. The frequencies and the mode shapes from the FE model at 25℃ are depicted in Figure 5.15.

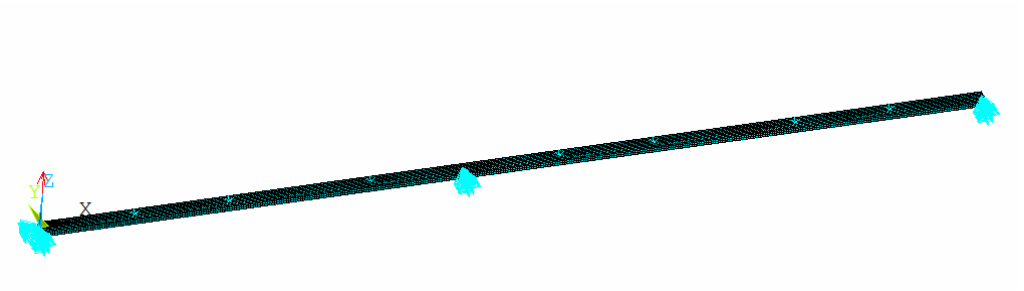
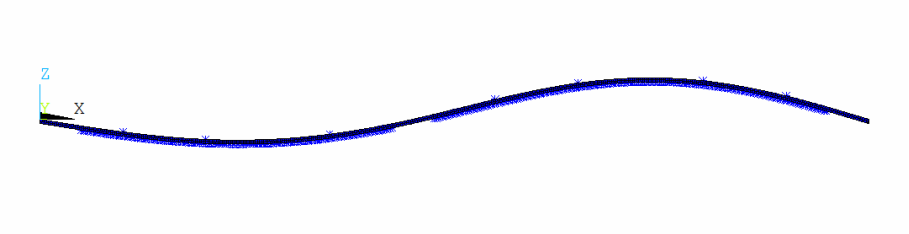
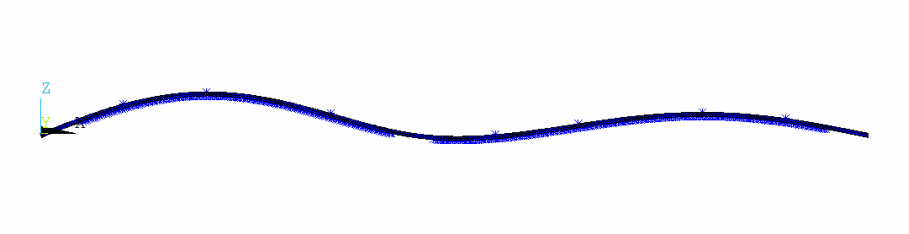


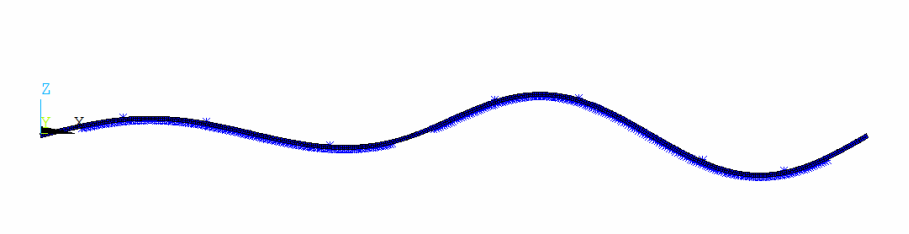

Figure 5.14 Finite element model of the experimental case



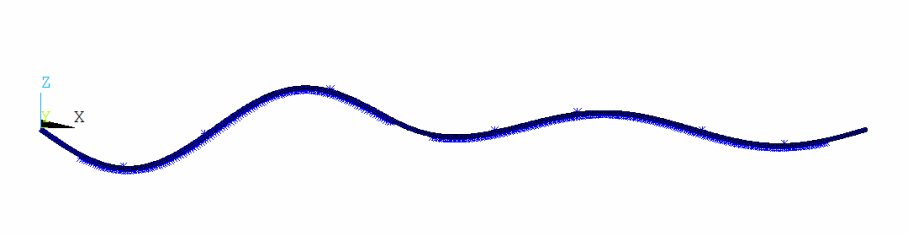

(a) Mode 1 (18.8397 Hz)



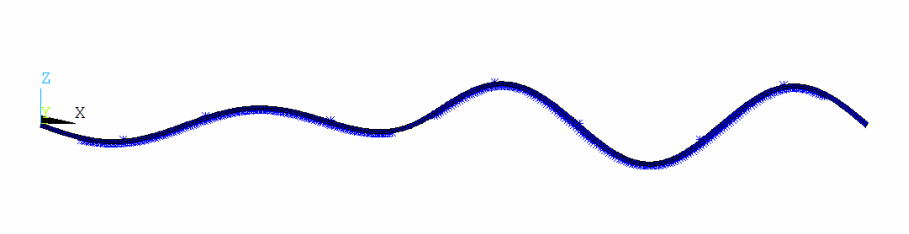

(b) Mode 2 (32.7243 Hz)




(c) Mode 3 (65.6139 Hz)




(d) Mode 4 (103.5837 Hz)




(e) Mode 5 (148.6948 Hz)

Figure 5.15 First 5 frequencies corresponding to the vertical mode shapes

Similar to the previous case, the experimental beam which represents the HF model is applied with a temperature-related axial force to simulate the boundary condition change caused by temperature, and such change is not considered in the FE model to represent the modelling error in the LF model. The axial force is controlled by using a jack and a load cell. The setting of the jack and load cell is displayed in Figure 5.16. The axial force is designed to have a nonlinear relation with temperature, which is given in Figure 5.17.

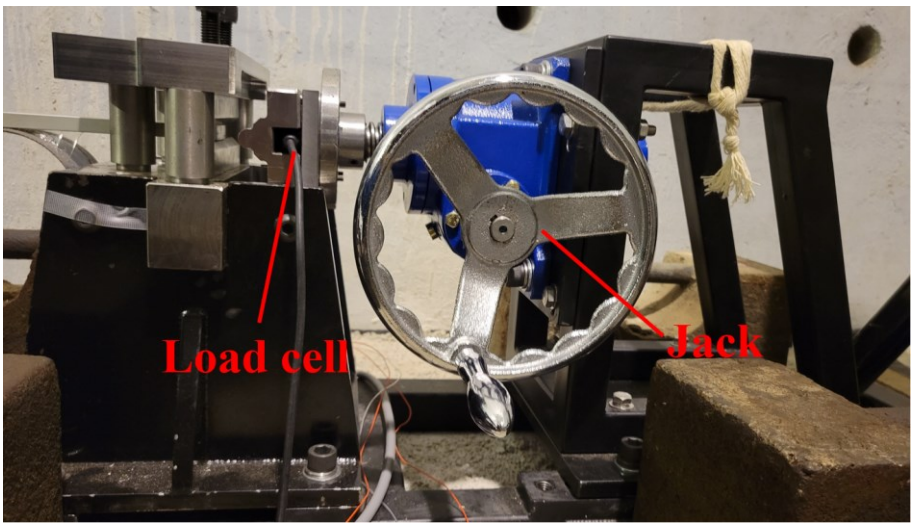



Figure 5.16 Deployment of jack and load cell

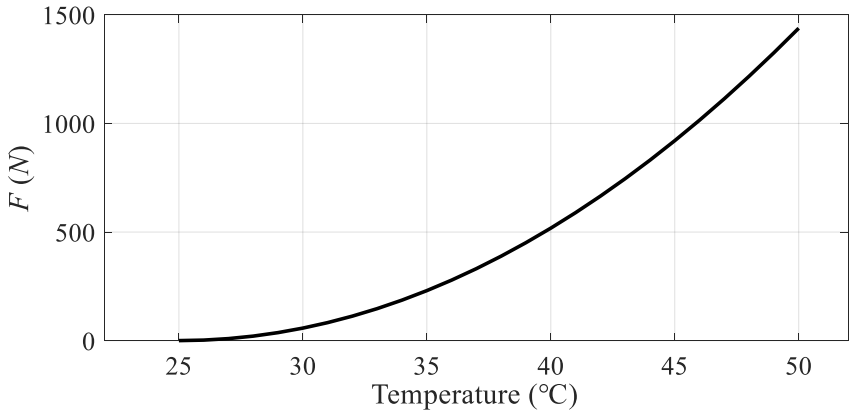



Figure 5.17 Boundary condition change (axial force) versus temperature

### 5.3.2 Numerical case of a three-span continuous beam bridge

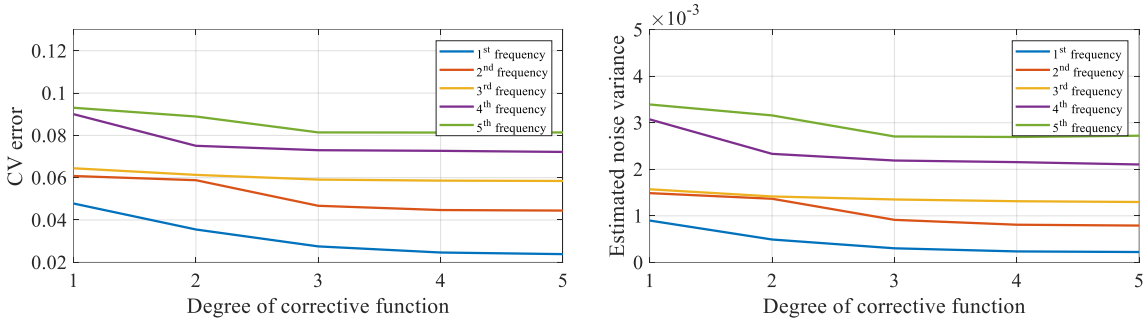
With the LF and HF FE models built in Section 5.3.1.1, the frequency data that vary with the structural parameters and temperatures can be provided. Following **Algorithm 5**, the LF data are first used to build the LFPCE models by employing the multi-output ASPCE-BASS algorithm. The input variables comprise of 26 structural parameters (elastic moduli of 26 parts) and 2 temperature variables ($T_t, T_b$). For building the surrogate models, all the input variables are assumed to conform uniform distribution. The distribution information is given in Table 5-1.

Table 5-1 Input distributions of the three-span continuous beam bridge

| Variable | Distribution | Lower bound | Upper bound |
|---|---|---|---|
| Elastic modulus $E_1 \sim E_{26}$ (Pa) (under 20 ℃) | Uniform | 1.42e10 | 3.55e10 |
| Temperature $T_t, T_b$ (℃) | Uniform | 0 | 50 |

The target outputs are the first 5 frequencies corresponding to the vertical mode shapes, so the number of outputs in ASPCE-BASS is 5. The threshold to terminate the algorithm is set as $10^{-5}$. With trained LFPCE models, the HF data are then leveraged to build the TL-MFPCE models. Regarding the HF data as the real observations, a zero-mean Gaussian white noise is added to the data of each frequency mode to simulate the measurement noise, where the signal-to-noise ratio (SNR) is 1%. For each output, the number of HF data is 150, which are randomly sampled in the input domain. In the training process, the CV error of each model with respect to the degree of corrective PCE function is shown in Figure 5.18(a), and the best degree can then be determined. Also, the estimated noise variances of the trained model for each frequency are shown in Figure

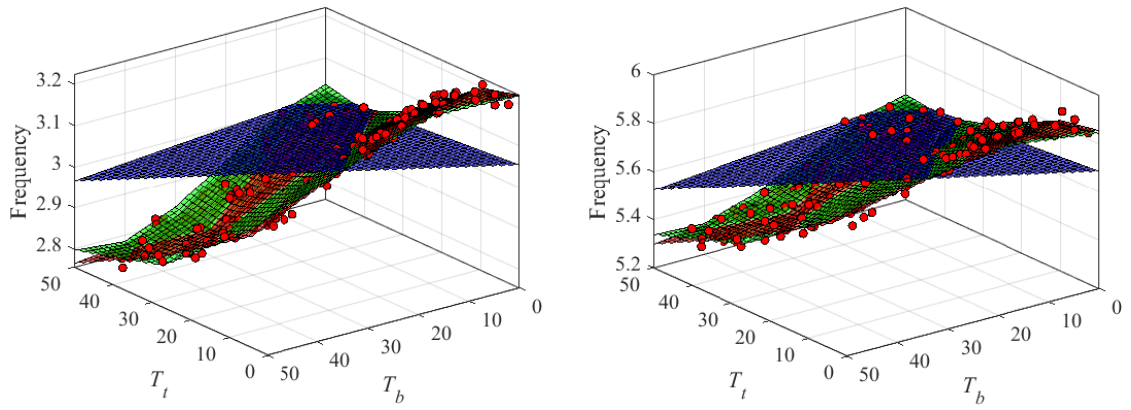
5.18(b), which could help choose the best degree.



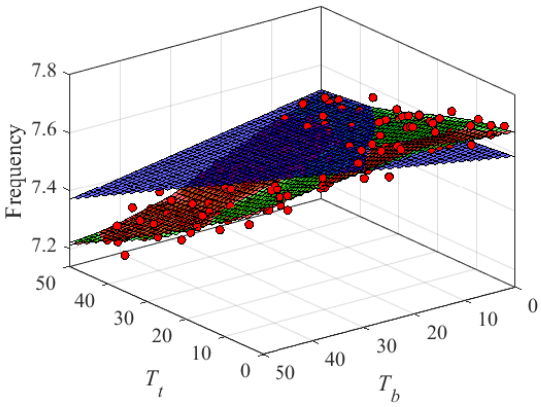

(a) (b)

Figure 5.18 (a) CV error versus degree of corrective function; (b) Estimated noise variance versus degree of corrective function

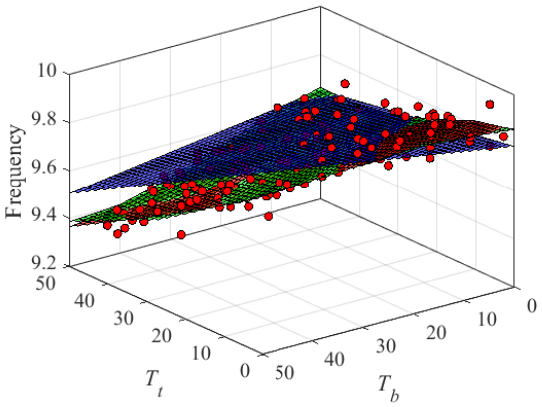
As can be seen, the CV errors of the trained models representing the first two frequencies do not decrease when the truncation degree exceeds 4, and similar trends can also be observed in Figure 5.18(b), so the best truncation degree for the corrective PCE models of the first two frequencies is 4. Similarly, the best truncation degree of corrective function for the last three frequencies is 3. Then, the trained MF models are compared with the LF models and the HF models as shown in Figure 5.19. It is obvious that the MF models approach more to the HF models than the LF models.



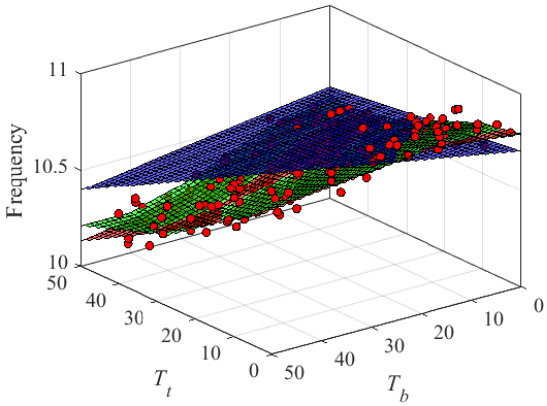

(a) 1st frequency (b) 2nd frequency

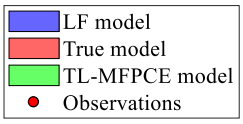



(c) 3rd frequency




(d) 4th frequency




(e) 5th frequency

Figure 5.19 Training results of TL-MFPCE

Furthermore, the quantity of HF data affecting the training accuracy is studied. The Mean Square Error (MSE) that defined in Equation (5.13) is utilised to evaluate the training accuracy. The size of HF data varies from 50 to 250, and the MSE results with respect to the data size are shown in Figure 5.20.

$$\varepsilon_{MSE} = \sum_{i=1}^{N_{val}} (F_{true}^{(i)} - F_{TL-MFPCE}^{(i)})^2 / N_{val} \tag{5.13}$$

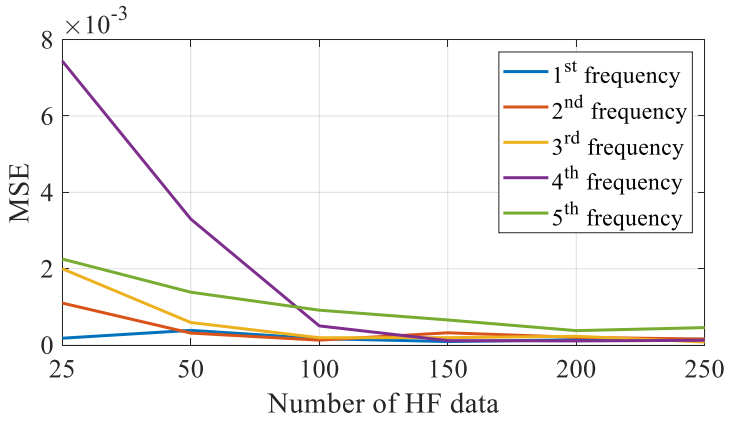


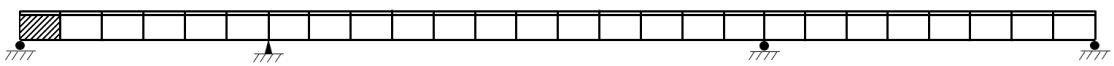

Figure 5.20 MSE with respect to the number of HF data

Clearly, the training accuracies of 5 models become stable when the number of HF data exceeds 200. Hence, the TL-MFPCE models that trained with 200 HF data are used in this case.

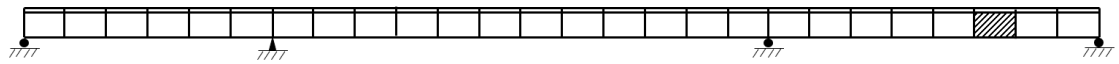
In order to validate the trained TL-MFPCE models under the circumstance that the structure is damaged, three damage scenarios are defined, and the responses from the HF model and the TL-MFPCE models are compared. The damage scenarios are simulated by decreasing the elastic modulus of the structure segments, and the detailed information can be found in Table 5-2 and Figure 5.21.

Table 5-2 Damage locations and severities of three damage scenarios

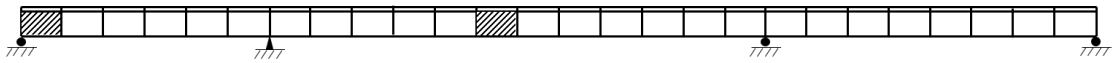
| Scenario | | Description |
|---|---|---|
| Single damage | DS1 | Part No. 1 is damaged with elastic modulus reducing 30% |
| | DS2 | Part No. 24 is damaged with elastic modulus reducing 50% |
| Double damages | DS3 | Parts No. 1 & 12 are damaged with elastic modulus reducing 20% and 40% respectively |



(a) DS1



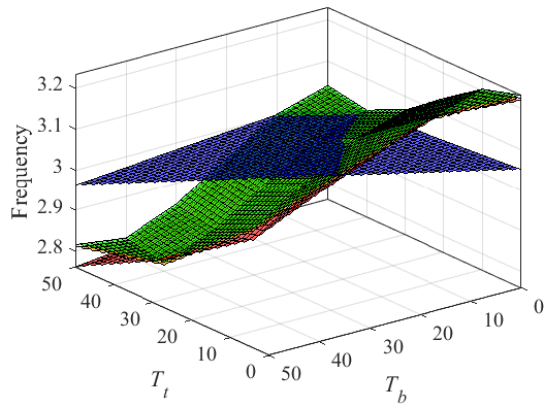
(b) DS2



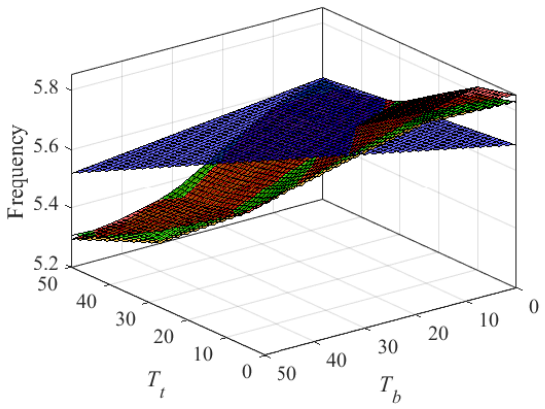
(c) DS3

Figure 5.21 Damage locations of three scenarios

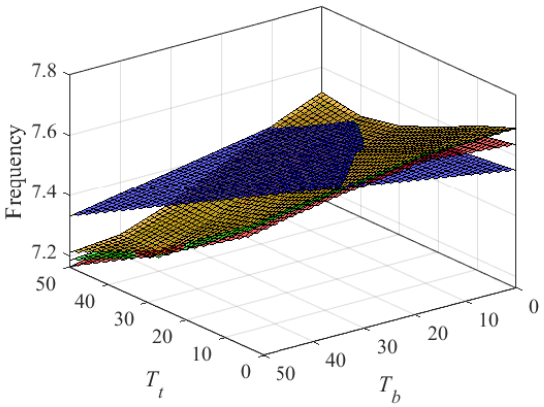
To demonstrate the improvement of the proposed TL-MFPCE method, the traditional MFPCE (tMFPCE) technique is employed for comparison. For the sake of fairness, the autoregressive form is used in tMFPCE. Here, the truncation degree of the corrective function in tMFPCE is chosen as 3 because larger truncation degree will result in instable regression solution with only 200 HF training data. The comparison of modelling results for three damage scenarios are displayed in Figure 5.22 to Figure 5.24.



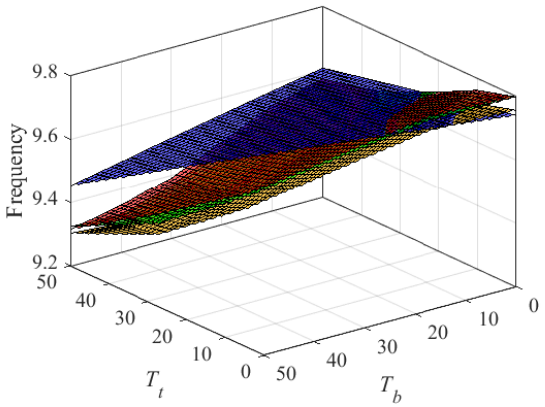

(a) 1st frequency



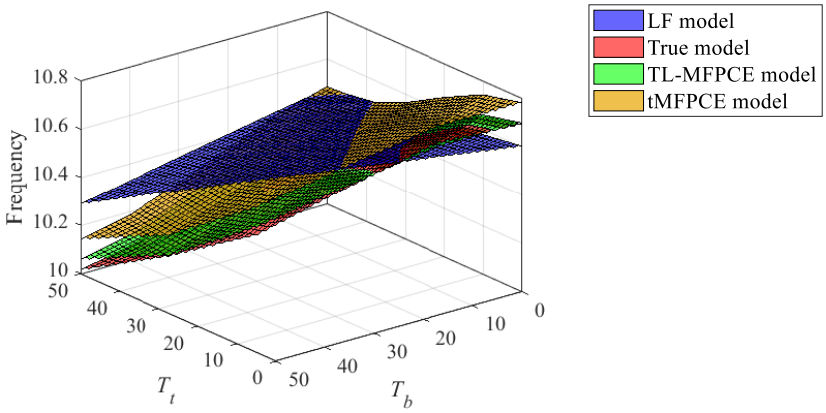

(b) 2nd frequency




(c) 3rd frequency

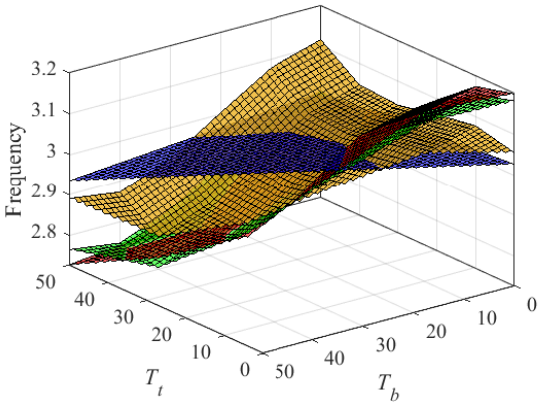



(d) 4th frequency

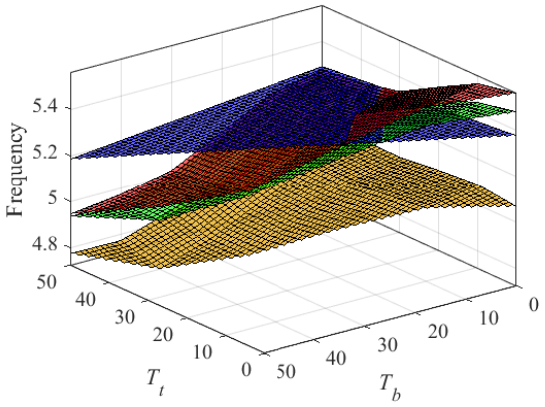



(e) 5th frequency

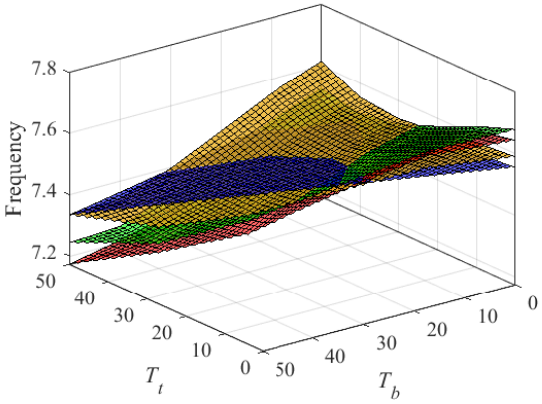
Figure 5.22 Comparison of validation results under damage scenario DS1



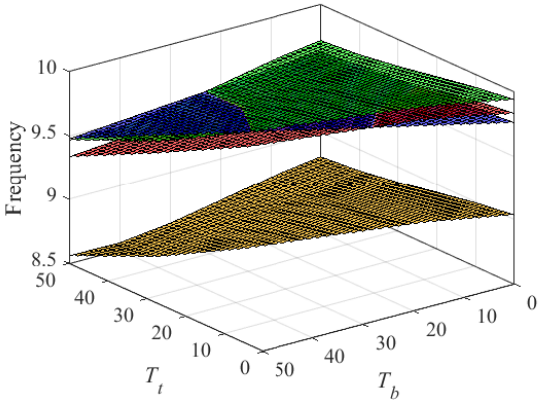

(a) 1st frequency



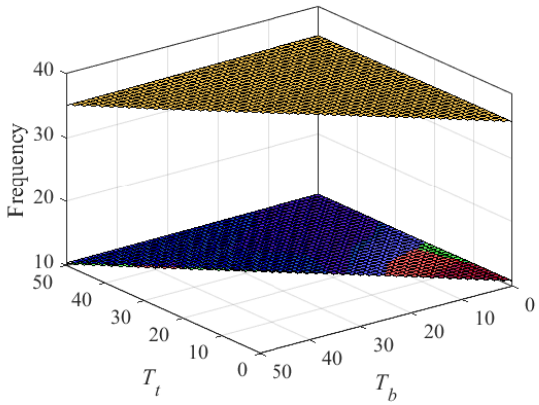

(b) 2nd frequency

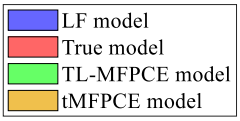



(c) 3rd frequency



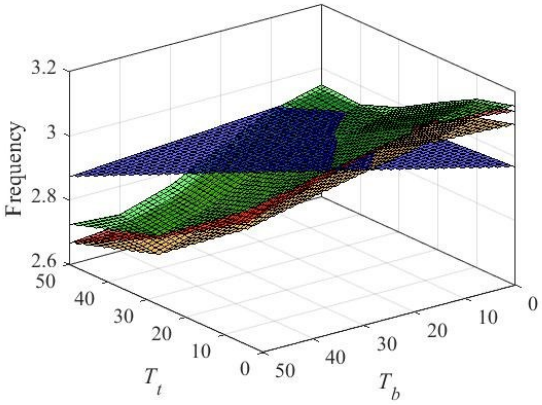

(d) 4th frequency



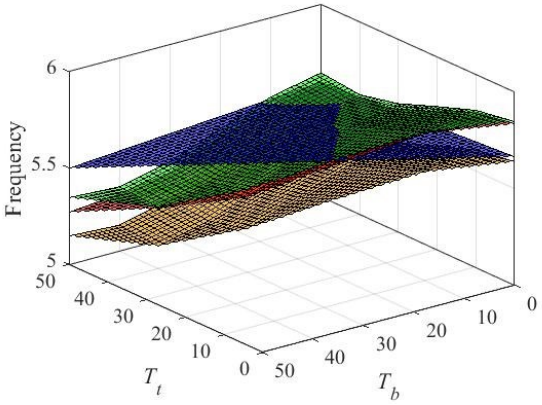




(e) 5th frequency

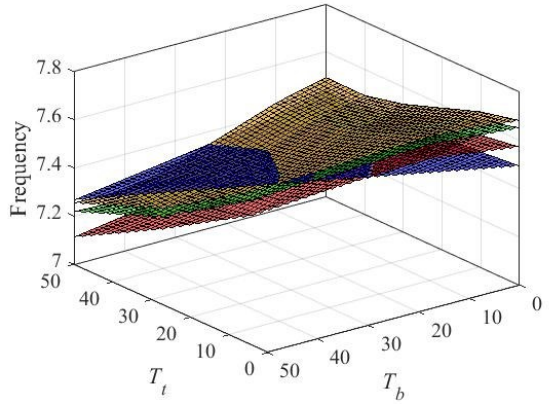
Figure 5.23 Comparison of validation results under damage scenario DS2



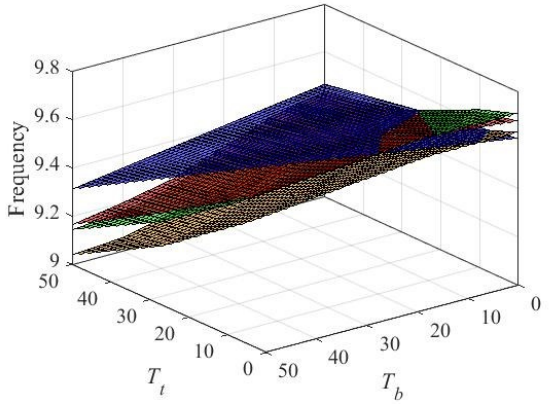

(a) 1st frequency



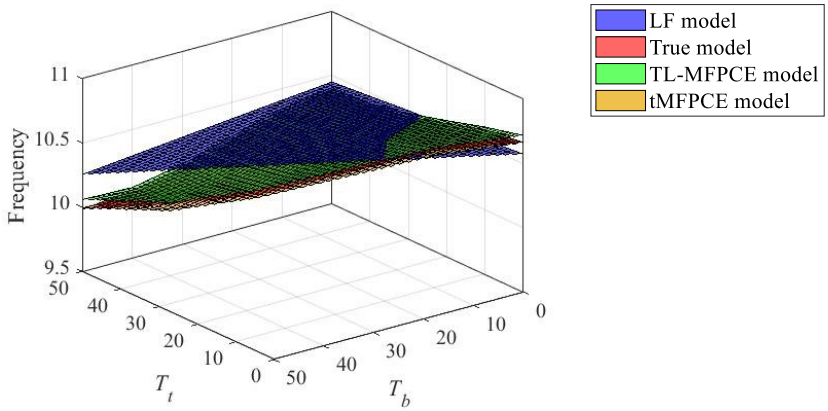

(b) 2nd frequency




(c) 3rd frequency




(d) 4th frequency




(e) 5th frequency

Figure 5.24 Comparison of validation results under damage scenario DS3

As is shown in Figure 5.22, the predictions from the TL-MFPCE models and tMFPCE models for damage scenario DS1 almost match well with those from HF models, except for the tMFPCE model of the 5th frequency. For damage scenario DS2 as depicted Figure 5.23, the predictions from TL-MFPCE models of 5 frequencies still close to the results from HF models, while the predictions from the tMFPCE models have large discrepancies, especially the predictions of the 4th and 5th frequencies. For damage scenario DS3, the tMFPCE models show a bit worse performance than the TL-MFPCE models. To make it clear, the MSE values of these predictions are given in Figure 5.25 to Figure 5.27.

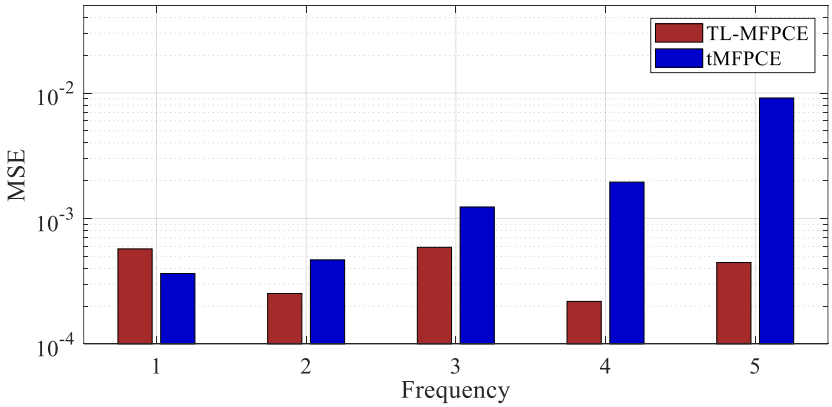



Figure 5.25 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS1

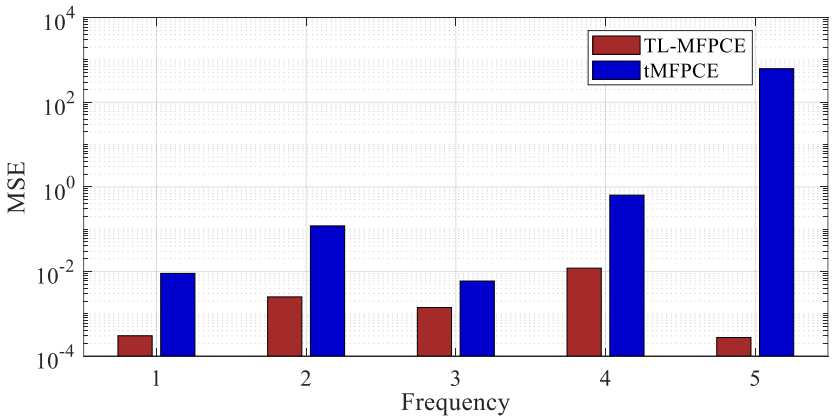



Figure 5.26 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS2

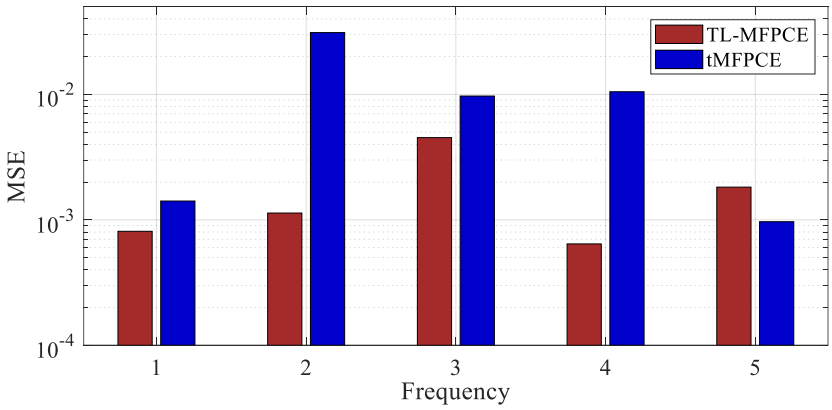



Figure 5.27 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS3

Intuitively, the MSE values of the TL-MFPCE models are mostly smaller than those of the tMFPCE models. Only the tMFPCE models of the 1st frequency in DS1 and the 5th frequency in

DS3 show a bit better performance than the TL-MFPCE models. It deserves to note that the tMFPCE models for DS2 have extremely large prediction errors. The reason might be that the basis terms related to the damaged structural segment in this damage scenario are updated with incorrect coefficient values. To conclude, the TL-MFPCE models outperforms the tMFPCE models when the structure is damaged.

### 5.3.3 Experimental case of a two-span continuous beam

With the LF FE model and the HF experimental beam structure, the LF and HF frequency data that vary with the structural parameters and temperature can be collected. The LF data is first used to build the LFPCE models by using the multi-output ASPCE-BASS algorithm. In these models, the input variables comprise of 20 structural parameters (section inertia moment of 20 parts) and 1 temperature variable. All the input variables conform to uniform distributions, and the distribution information is given in Table 5-3.

Table 5-3 Input distributions of the two-span continuous beam

| Variable | Distribution | Lower bound | Upper bound |
|---|---|---|---|
| Section inertia moment $I_1 \sim I_{20}$ ($m^4$) | Uniform | 1.1038e-9 | 2.7596e-9 |
| Temperature $T$ (℃) | Uniform | 25 | 50 |

The target outputs are the first 5 frequencies corresponding to the vertical mode shapes. The threshold to terminate the multi-output ASPCE-BASS algorithm is set as $10^{-5}$. With the trained LFPCE models, the measurements collected from the healthy experimental beam under different

temperatures are then leveraged as the HF data to train the TL-MFPCE models. Here, 135 measurement data for each frequency are collected for MF modelling. In the training process, the CV error of each frequency with respect to the degree of corrective PCE function is given in Figure 5.28(a), and the estimated noise variances are shown in Figure 5.28(b).

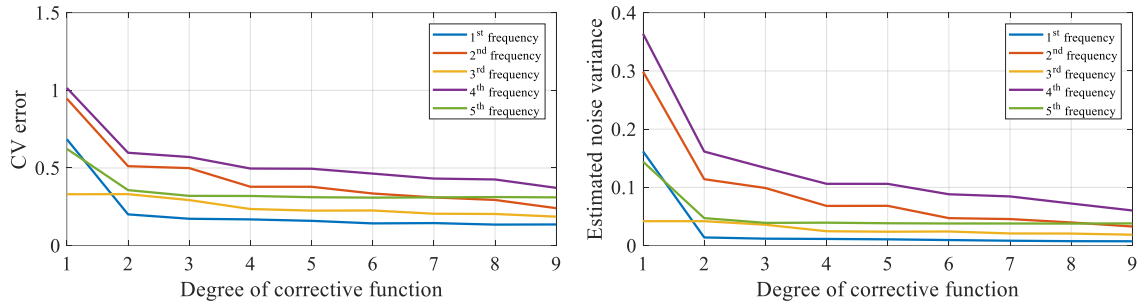
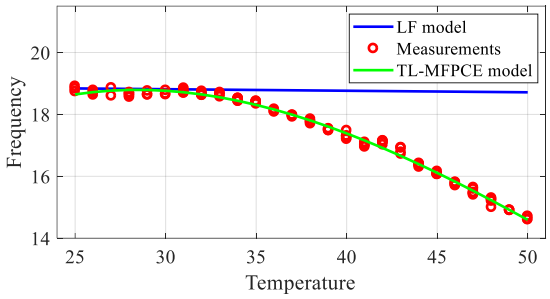


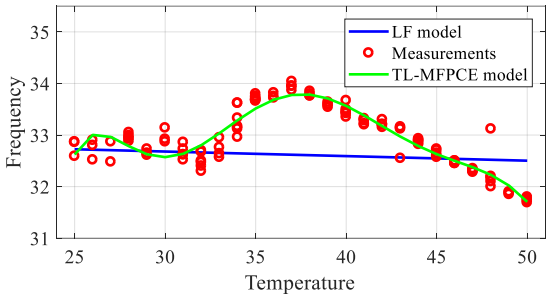

Figure 5.28 (a) CV error versus degree of corrective function; (b) Estimated noise variance versus degree of corrective function

It can be seen in Figure 5.28 that the CV errors and the estimated noise variances of the 1st, 3rd, and 5th frequencies do not decrease when the truncation degree exceeds 4, so the best truncation degree for the corrective PCE models of these frequencies is 4. By contrast, the CV errors and the estimated noise variances of the 2nd and 4th frequencies keep decreasing with the increase of the degree of corrective function. However, purely increasing the truncation degree will cause overfitting issue. As the CV errors and estimated noise variances have relatively low values when the truncation degree exceeds 7, we choose 7 for modelling these two frequencies. Under the selected truncation degrees, the trained MF models are shown in Figure 5.29. Apparently, the MF models can update the discrepancy between the LF models and the HF data when the structure is in healthy condition. To note, it can be seen from Figure 5.29(b)&(d) that the measurements of 2nd and

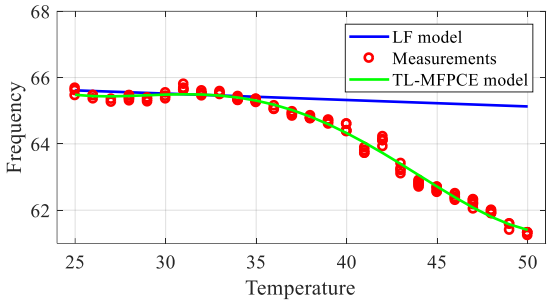
4th frequencies show distinct regularity to others, and the TL-MFPCE models need larger truncation degree in the corrective function for updating. We infer that it is caused by the stiffness change of boundary with the increasing of axial force, which further affects these two frequencies.



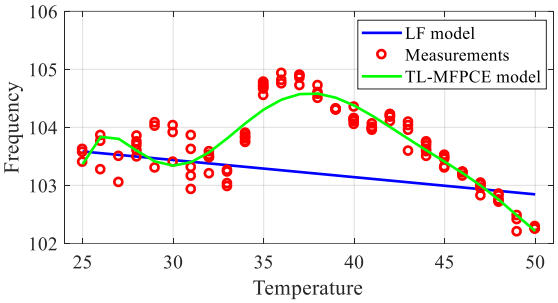

(a) 1st frequency



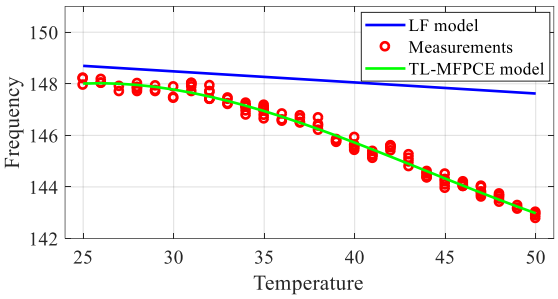

(b) 2nd frequency




(c) 3rd frequency




(d) 4th frequency




(e) 5th frequency

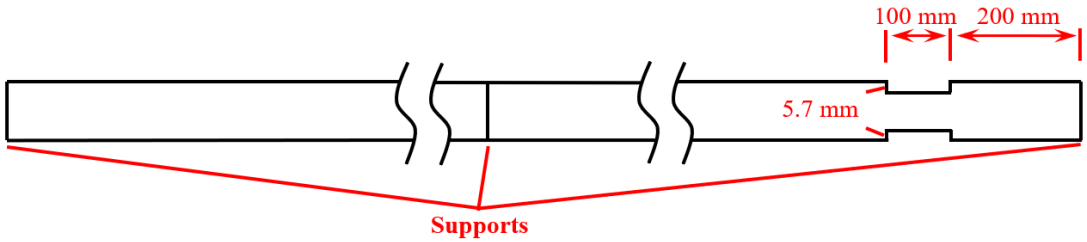
Figure 5.29 Training results of TL-MFPCE

In the validation stage, three damage scenarios are imposed on the experimental beam, and the responses from the experimental model and the TL-MFPCE models are compared. The damage scenarios are generated by cutting a part of the beam, and their detailed information is in Table 5-4

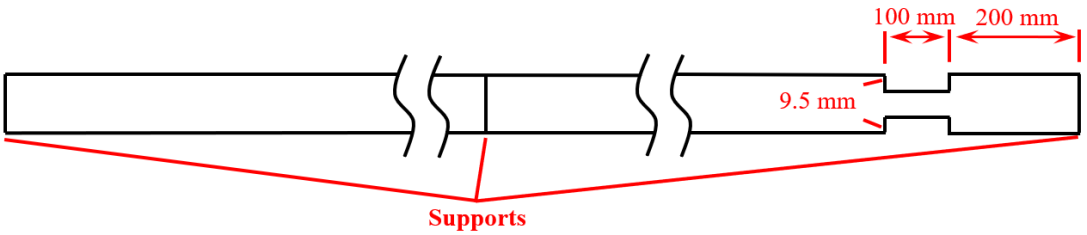
and Figure 5.30. To facilitate understanding, an example of damage scenario DS1 is shown in Figure 5.31.

Table 5-4 Damage locations and severities of three damage scenarios

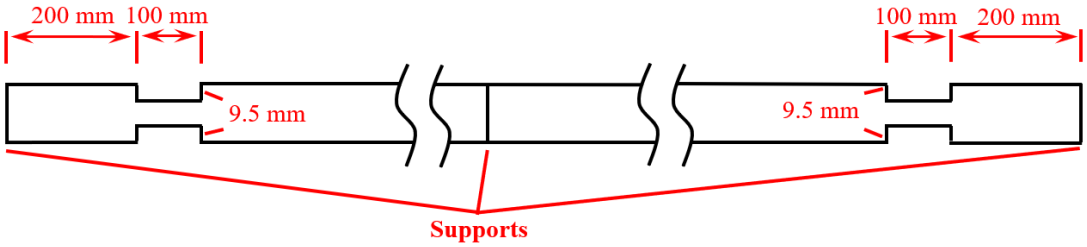
| Scenario | | Description |
|---|---|---|
| Single damage | DS1 | Part No. 18 is damaged with 11.4 mm cutting depth (30% damage) |
| | DS2 | Part No. 18 is damaged with 19.0 mm cutting depth (50% damage) |
| Double damages | DS3 | Parts No. 3 & 18 are all damaged with 19.0 mm cutting depth (50% damage) |




(a) DS1




(b) DS2




(c) DS3

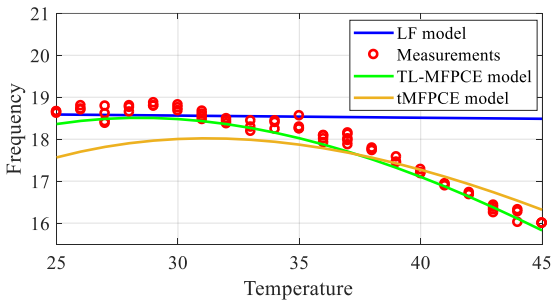
Figure 5.30 Three damage scenarios of the experimental beam

Figure 5.31 An example of 30% damage (DS1)

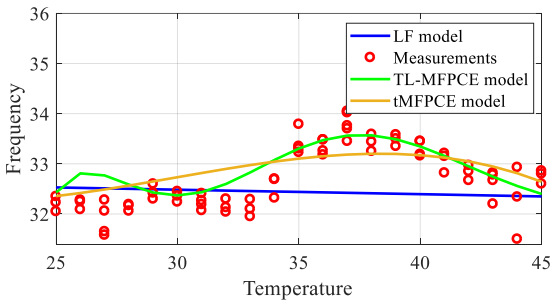
Table 5-5 Information of measurements under different damage scenarios

| Damage scenarios | Temperature range | Data quantity |
|---|---|---|
| DS1 | [25℃, 45℃] | 71 |
| DS2 | [33℃, 45℃] | 47 |
| DS3 | [35℃, 40℃] | 23 |

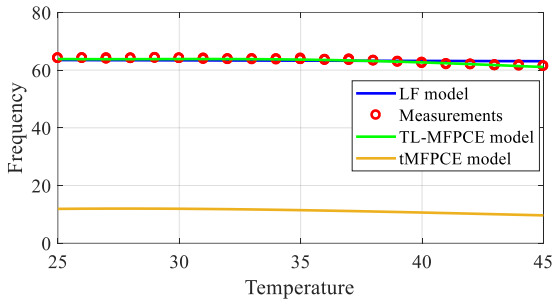
To simulate the real situations, measurements are collected within different temperature ranges in three damage scenarios, and the quantity of measurements are also different in these damage scenarios. The information of measurements in three damage scenarios are summarized in Table 5-5. The model predictions from two methods, tMFPCE and TL-MFPCE, are compared with these measurements to verify their performance. Similar to the previous case, the truncation degree of the corrective function in tMFPCE cannot exceeds 3 due to the requirement on robust regression calculation. The comparison results are displayed in Figure 5.32 to Figure 5.34.



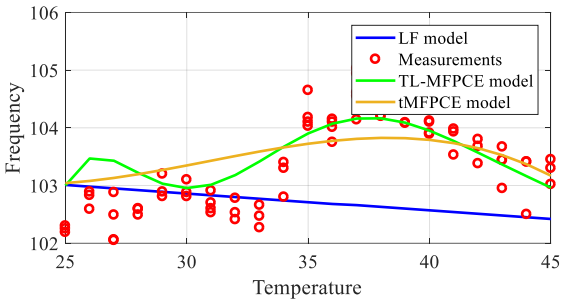

(a) 1st frequency



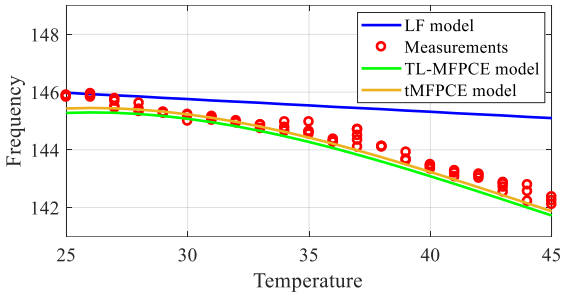

(b) 2nd frequency




(c) 3rd frequency

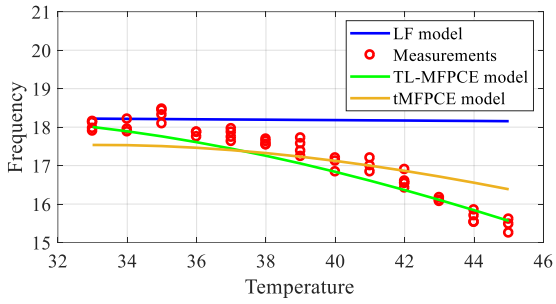



(d) 4th frequency

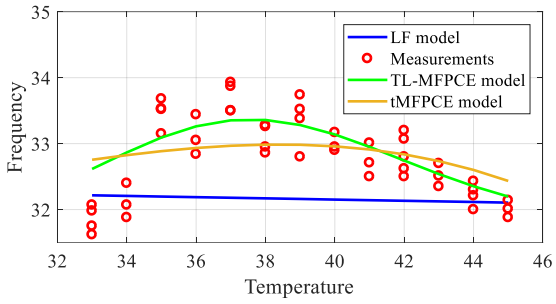



(e) 5th frequency

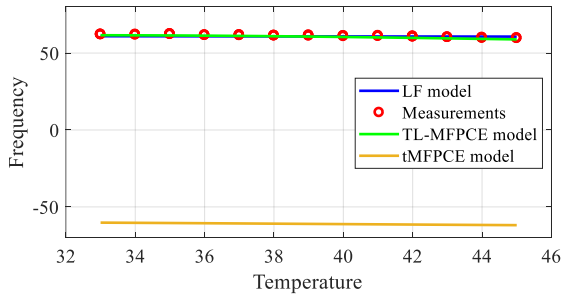
Figure 5.32 Comparison of validation results under damage scenario DS1



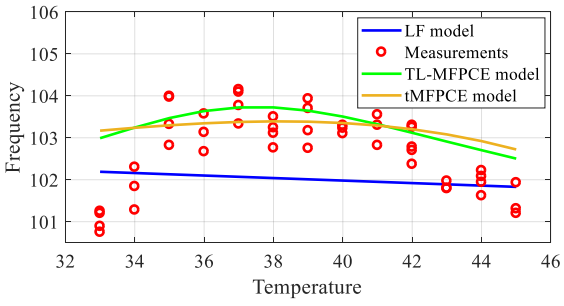

(a) 1st frequency



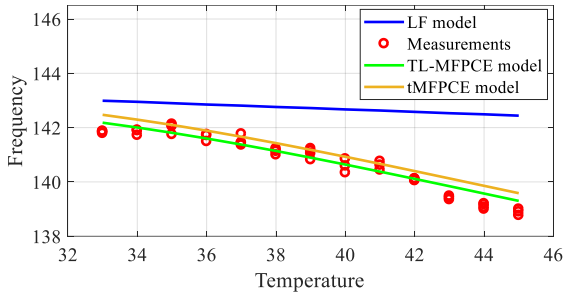

(b) 2nd frequency




(c) 3rd frequency

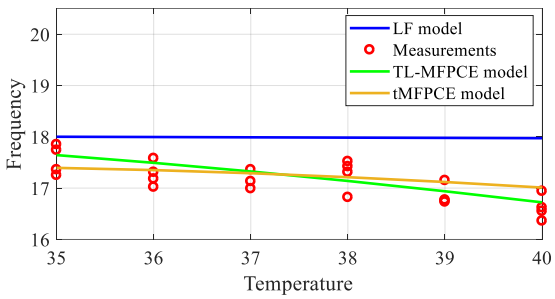



(d) 4th frequency

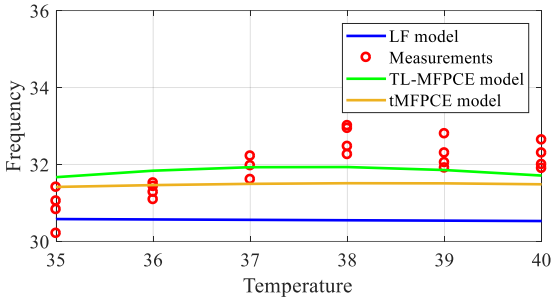



(e) 5th frequency

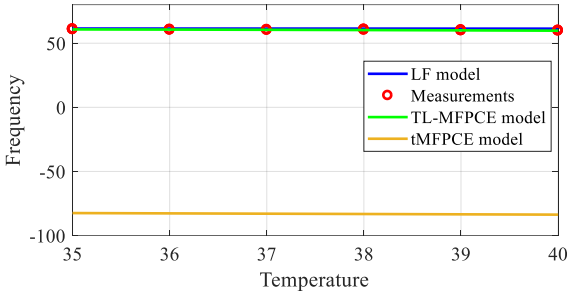
Figure 5.33 Comparison of validation results under damage scenario DS2



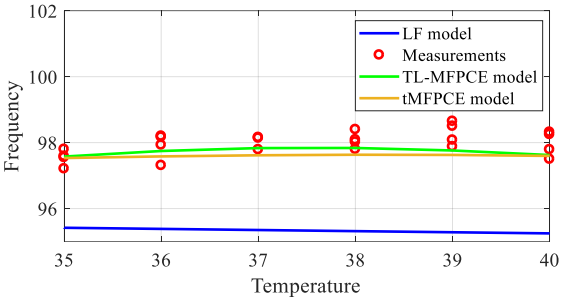

(a) 1st frequency



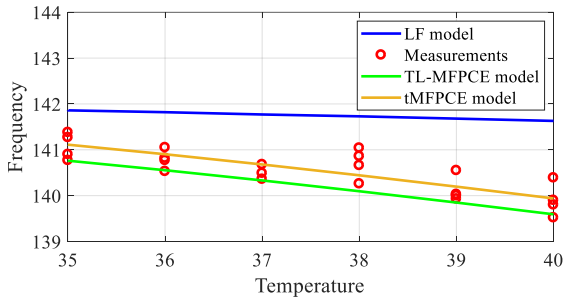

(b) 2nd frequency




(c) 3rd frequency




(d) 4th frequency




(e) 5th frequency

Figure 5.34 Comparison of validation results under damage scenario DS3

Upon most circumstances, the tMFPCE and TL-MFPCE models can both predict the global trends of the structural responses with desired accuracy, while the TL-MFPCE models match more

with the measurements than the tMFPCE models because higher truncation degrees in the corrective function were chosen in the training stage of TL-MFPCE. More importantly, for predicting the 3rd frequency in three scenarios, the tMFPCE models have large deviations to the measurements and the TL-MFPCE models. The reason is that the weight coefficient $\rho$ in this tMFPCE model is trained with improper value. Then, the MSE values of these models under three damage scenarios are compared in Figure 5.35 to Figure 5.37.

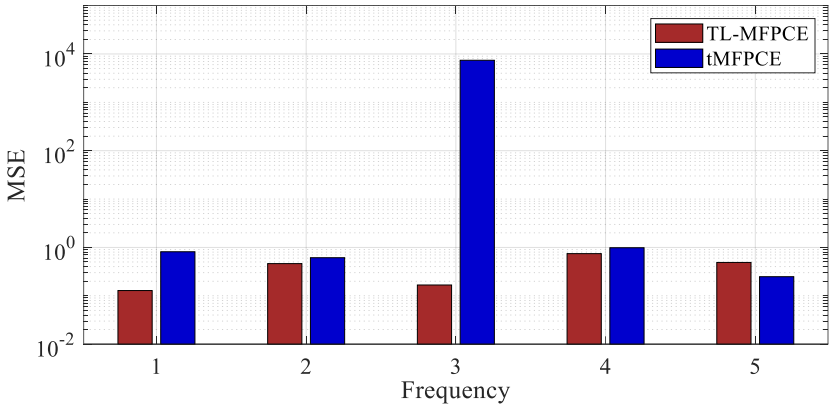



Figure 5.35 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS1

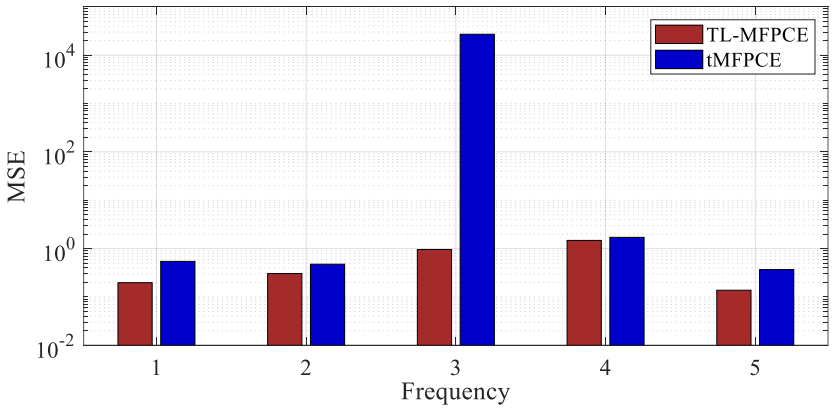



Figure 5.36 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS2

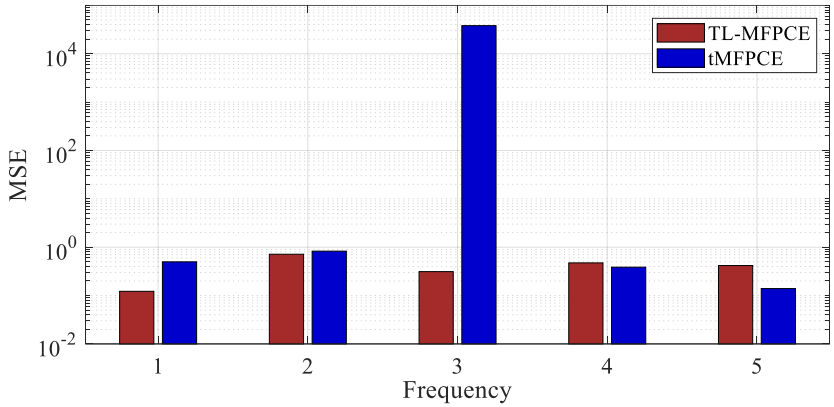



Figure 5.37 MSE comparison between TL-MFPCE and tMFPCE: damage scenario DS3

Intuitively, the tMFPCE models for the 3rd frequency have extremely large prediction errors in three damage scenarios, and the TL-MFPCE models have a higher accuracy than the tMFPCE models upon most occasions. However, a problem of TL-MFPCE is exposed in the results. As shown in Figure 5.32(b)&(d), the predictions from TL-MFPCE models show better performance at temperature over 35°C than below. This is because the models overfit the training data as shown in Figure 5.29(b)&(d). Fortunately, the prediction errors of models for these two frequencies are tiny.

## 5.4 Summary

In this chapter, a Transfer Learning (TL)-based MFPCE (TL-MFPCE) modelling technique is proposed to build accurate surrogate model for bridge structures, which can further help in damage identification. Low-Fidelity PCE (LFPCE) surrogate models are first established based on the LF physical model of a bridge structure by using the ASPCE-BASS algorithm. Then, the concept of parameter-based TL and the MFPCE modelling technique are combined to eliminate the temperature-induced modelling error in the LFPCE models with measurements from the High-

Fidelity (HF) model/real structure in healthy condition. Finally, the updated MFPCE models can be used to predict the structural responses with desired accuracy even if the structure is damaged. In principle, the parameter-based TL enables the retaining of information related to the structural parameters from the physical model, so the requirement on the real measurements can be reduced.

For validation, a simulation case of a three-span continuous beam bridge is built, and an experimental model of a two-span continuous beam is constructed. In the training stage, it is shown that the trained TL-MFPCE models have high accuracy when the structure is in healthy condition, which indicates that the training is successful. In the validation stage, three damage scenarios involving single damage and multiple damages are defined for both cases, and the proposed method is compared with the traditional MFPCE (tMFPCE) method. Results manifest that the proposed method outperforms the tMFPCE method when the structure is damaged.

It is worth mentioning that the proposed TL-MFPCE approach can work together with the popularly used model updating technique, as the modelling error may also exist in the structural parameters of a physical model, e.g., the incorrect estimation of element stiffness. To improve the modelling precision, model updating can be performed by defining a reference state to the structure. The real measurements obtained in the reference state can be used for model updating. Then, the TL-MFPCE modelling technique is applied with data collected in the operational stage.

# CHAPTER 6
# SPARSE DAMAGE IDENTIFICATION BASED ON MFPCE MODEL

## 6.1 Introduction

In the past decades, numerous structural damage identification methods have been developed for bridges, where frequency was one of the most widely used structural responses (Fang and Perera, 2011; Peeters et al., 2001; Zhou et al., 2011). In the previous chapter, reliable PCE surrogate models have been constructed to describe the pattern of structural frequencies concerning the damage parameters and temperature, and such models can be further employed to help identify possible damages. Basically, the principle behind those damage identification approaches which employ the data models or surrogate models is to find a set of damage parameter values that can minimise the residual between the model evaluations and real responses (Wang et al., 2020a; Zhang and Xu, 2016; Zhou et al., 2015):

$$\widehat{\boldsymbol{q}} = arg\min_{\boldsymbol{q}} \left\{ \left\| \boldsymbol{F}(\boldsymbol{q}) - \widehat{\boldsymbol{F}} \right\|_2^2 \right\} \tag{6.1}$$

in which $\boldsymbol{F}(\boldsymbol{q})$ denotes the model evaluations (frequencies) and $\widehat{\boldsymbol{F}}$ represents the measurements perturbed by noise; $\boldsymbol{q} = (\boldsymbol{X} - \boldsymbol{X_d})/\boldsymbol{X}$ is a vector of damage parameters calculated by the structural

parameters $\boldsymbol{X}, \boldsymbol{X_d}$ before and after damage, which indicates both the damage location and damage severity. Typically, it is often reasonable to deem that the damage occurs on a bridge in a sparse way, that is, the number of damages is very small compared with the quantity of total components (Hernandez, 2014; Huang et al., 2017a). So, $\boldsymbol{q}$ is a sparse vector, and the sparse representation methods could help solve this problem. Similar to Equation (3.13), a regularisation term is added to induce sparsity on $\boldsymbol{q}$.

$$\widehat{\boldsymbol{q}} = arg\min_{\boldsymbol{q}} \left\{ \left\| \boldsymbol{F}(\boldsymbol{q}) - \widehat{\boldsymbol{F}} \right\|_2^2 + \lambda \|\boldsymbol{q}\| \right\} \tag{6.2}$$

Generally, $l_1$-norm is selected for the regularisation term since it is the most tractable term to induce sparse solution (Bruckstein et al., 2009). By contrast, $l_0$-norm can induce the most sparsity, but it has been demonstrated that Equation (6.2) with $l_0$ regularisation is a Non-deterministic Polynomial-time hard (NP-hard) problem, which means that the time for getting the optimal solution is unacceptable (Natarajan, 1995). Therefore, $l_1$ regularisation has become the dominant technique in the sparse representation algorithms for addressing structural damage identification problem (Guo et al., 2020; Hou et al., 2018a; Lai and Nagarajaiah, 2019; Wang et al., 2020b).

However, solving Equation (6.2) by using the traditional sparse representation methods with $l_1$ regularisation is still complicated in this study. The reason is that the PCE model $\boldsymbol{F}(\boldsymbol{q})$ is possibly not linearly expressed by $\boldsymbol{q}$, which contradicts to the premise in the traditional sparse representation methods that the system should be linear (Zhang et al., 2015). In this study, an

approximate $l_0$ sparse representation method for damage identification is developed based on the Discrepancy Principle (DP) and the Cuckoo Search Algorithm (CSA). DP was usually used for choosing a proper value of regularisation parameter for Tikhonov regularisation, and it was further applied in the field of structural damage identification with $l_1$ minimisation (Hou, et al., 2018a). CSA is a kind of heuristic algorithm, which is superior in solving the NP-hard optimisation problem. In the proposed method, we first give assumptions to the number of damages, that is, the optimisation is performed with known damage number. Then, the optimisation problem is transferred into an approximate $l_0$ regularisation problem. Regarding the damage location and severity as the optimisation objectives, CSA is then adopted to find the optimal solutions under each assumption. According to DP, the solutions calculated under different assumptions are compared to find the most possible damage scenario, and the corresponding solution obtained by CSA is the correct damage result. Finally, the proposed damage identification method is validated on a numerical bridge model and an experimental beam model.

## 6.2 Theory

### 6.2.1 Discrepancy principle for damage identification

DP was commonly utilised for selecting proper regularisation parameter value in Tikhonov regularisation (Dong et al., 2018; Hämarik and Raus, 2006; Lukas, 1995). In 2018, Hou et al. (2018a) further extended this approach for parameter selection in $l_1$ regularisation and demonstrated that

DP was promising in the field of structural damage identification. Compared with other sparse damage identification methods, this method is simple to be implemented, and it has been demonstrated to be effective in addressing similar issues (Hou et al., 2018a). In this section, the DP for parameter selection in $l_1$ regularisation is introduced.

Taking the norm of regularisation term in Equation (6.2) as $l_1$, the optimisation problem can be expressed as:

$$\widehat{\boldsymbol{q}} = arg\min_{\boldsymbol{q}} \left\{ \left\| F(\boldsymbol{q}) - \hat{F} \right\|_2^2 + \lambda \|\boldsymbol{q}\|_1 \right\} \tag{6.3}$$

Here, $\hat{F}$ represents the measurement data that is contaminated by noise, which is $\hat{F} = F_t + \epsilon$; and $F_t$ is the system true value. Under an unknown damage $\tilde{\boldsymbol{q}}$, the real residual between the model evaluations and measurements is

$$\left\| F(\tilde{\boldsymbol{q}}) - \hat{F} \right\|_2^2 = \|F(\tilde{\boldsymbol{q}}) - F_t - \epsilon\|_2^2 = \|\epsilon\|_2^2 \tag{6.4}$$

when the system true value can be precisely predicted by the surrogate model $F(\boldsymbol{q})$. DP is to find a regularisation parameter $\lambda \geq 0$ so that the solution $\boldsymbol{q}_\lambda$ of Equation (6.3) satisfies:

$$\left\| F(\boldsymbol{q}_\lambda) - \hat{F} \right\|_2^2 = \|\epsilon\|_2^2 \tag{6.5}$$

Then, $\boldsymbol{q}_\lambda$ will be equal to the actual damage $\tilde{\boldsymbol{q}}$. Generally, the condition in Equation (6.5) is too rigorous due to the existence of uncertainties, so Hou et al. (2018a) relaxed it by

$$\left| \left\| \boldsymbol{F}(\boldsymbol{q}_\lambda) - \widehat{\boldsymbol{F}} \right\|_2^2 - \|\epsilon\|_2^2 \right| \leq Tol \tag{6.6}$$

Assuming the measurement noise follows a zero-mean Gaussian distribution, the optimal value of $\lambda$ can be selected by comparing between the calculated residual value and the variance of measurement noise.

Generally, the noise in the frequency data is assumed as a stochastic process:

$$\widehat{\boldsymbol{F}} = (1 + \boldsymbol{\epsilon})\boldsymbol{F}_t \tag{6.7}$$

where $\widehat{\boldsymbol{F}}$ represents a vector of measured frequency data of different orders; $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. Then, Equation (6.3) can be re-organised as:

$$\widehat{\boldsymbol{q}} = arg\min_{\boldsymbol{q}} \left\{ \frac{1}{m \times N_T} \sum_{i=1}^{m} \sum_{j=1}^{N_T} \left( \frac{F_{ij}(\boldsymbol{q}) - \widehat{F_{ij}}}{\widehat{F_{ij}}} \right)^2 + \lambda \|\boldsymbol{q}\|_1 \right\} \tag{6.8}$$

in which $F_{ij}(\boldsymbol{q})$ and $\widehat{F_{ij}}$ are the $i^{th}$ frequency under the $j^{th}$ temperature value from the surrogate model and experimental model, respectively. As a result, we would like to select a value of $\lambda$ so that the residual value is:

$$residual = \frac{1}{m \times N_T} \sum_{i=1}^{m} \sum_{j=1}^{N_T} \left( \frac{F_{ij}(\boldsymbol{q}_\lambda) - \widehat{F_{ij}}}{F_{ij}(\boldsymbol{q}_\lambda)} \right)^2 = E(\|\boldsymbol{\epsilon}\|_2^2) = \sigma^2 \tag{6.9}$$

To summarise, the procedure of DP is: (1) Predefine a set of candidate values of $\lambda$, and solve Equation (6.8) for each given value of $\lambda$; (2) Compute the residual value corresponding to each $\lambda$; (3) Compare the calculated residual values with the variance of the measurement noise to find the optimal $\lambda$. As can be seen, the noise level is necessary in DP. Fortunately, the BCS method employed for TL-MFPCE modelling in the previous chapter has estimated the noise level, as

explained in Equation (5.12), which can be used in this study.

### 6.2.2 Cuckoo search algorithm

Heuristic algorithm is one that designed to solve a problem in a faster and more efficient fashion than traditional methods by sacrificing optimality, accuracy, precision or completeness for speed (Pearl, 1984). One advantage of heuristic algorithm is that it does not require the explicit expression function of the target problem, thus it is usually used to address NP-hard problems (Kesavan et al., 2020). Starting with any feasible solutions, heuristic algorithms will apply random changes on the values of variables in these feasible solutions for the sake of evolution. By employing the objective function and evaluation criteria, solutions with in relatively better objective value will be kept. With the termination criterion, the algorithm will stop the circulation and output the best solution. In order to solve Equation (6.8), an outstanding heuristic algorithm, Cuckoo Search Algorithm (CSA), is introduced.

Cuckoo is a kind of birds which does not know how to build nest and brood for next generation (Yang and Suash Deb, 2009). They propagate their species over the course of history by a special brood strategy, brood parasitism, which is to lay their eggs in the nests of other birds. Some of the host birds may discover the cuckoo's egg and get rid of them by throwing them away or abandoning the nest and building a new one elsewhere, and some of the host birds may bring them up. In the CSA theory, the cuckoo's egg or the nest it located is regarded as a feasible solution. The criterion

to evaluate the solution is that whether cuckoo can normally grow up in this nest or not. Therefore, the process that cuckoos searching nests for laying eggs implies the process to search for better feasible solutions, and the survival of cuckoo denotes that the nest is a good solution.

For the simplicity of application, four idealised assumptions were proposed in CSA (Yang, 2010): (1) Each cuckoo only dumps one egg $x^{(t)}$ in generation $t$; (2) The egg laid by a cuckoo will be discovered by the host bird in a generation with a fixed probability value $Pd$ (which is pre-set before the calculation), and others will survive; (3) The number of cuckoos $a$ is always fixed during the iteration process; (4) The best nest with high objective value will be preserved in the next generation. At the beginning, initial $a$ nests are randomly generated. The performance of each cuckoo individual (or each nest) in this generation will then be evaluated according to the objective function. The cuckoos with high objective values can survival, while those will low objective values have a probability to be abandoned. To keep the population size, some new nests will be randomly generated via a bias random walk strategy for supplementing. This procedure also makes sure that the optimisation process would not fall into the local optima. Subsequently, all the cuckoos in this generation will search for nests to lay eggs and get into the next generation. In CSA, a searching strategy called Lévy flights is employed, which is defined by the following equation (Yang, 2010):

$$x_i{}^{(t+1)} = x_i{}^{(t)} + \alpha \otimes \text{Lévy}(\tau) \tag{6.10}$$

where $x_i$ represents the $i^{th}$ cuckoo in this generation; $\alpha > 0$ is the step size which relates to the scales of the problem of interest; $\otimes$ is the entry-wise multiplications; $\text{Lévy}(\tau)$ is the Lévy random

route which samples from the Lévy distribution, and $\tau$ is a parameter in this distribution.

$$\text{Lévy} \sim u = t^{-\tau}, \quad (1 < \tau < 3) \tag{6.11}$$

This distribution has an infinite variance with an infinite mean, so it is hard to sample from this distribution. To resolve this issue, an approximate calculation called Mantegna method was utilised to generate random samples under Lévy distribution (Yang and Suash Deb, 2009). Sampling under the Lévy fights will make the random walk have an alternate change between long distance move and short distance move so that the searching behaviour is in both local and global ways. Finally, the algorithm will terminate if the iteration time reaches the maximum generation. The best nest (cuckoo) kept in the last generation will be the final optimal solution.

| **Algorithm 6.** CSA (for minimum optimisation) |
|---|
| **Input:** Objective function; measurements from damaged bridge structure; population size; discovery rate; maximum generation; other parameters required. |
| **Initialisation:** Generate a population of feasible solutions; calculate the objective function value of each solution. |
| **In $i$th iteration:**<br>1. Choose a solution randomly among the population;<br>2. Generate a new solution by Lévy flights and evaluate the objective function value of the new solution;<br>3. Select a nest among the population randomly; replace the old solution with the new one if the objective function value of the new solution is smaller than that of the old solution; otherwise the old solution is retained.<br>4. Abandon a fraction of the worse solutions and generate new solutions via bias random walk for supplementing;<br>5. Evaluate the objective function values of all the solutions;<br>6. Record the best solution in the present and past iterations;<br>7. If $i = maximum\ generation$, quit iteration. |

To solve Equation (6.8) via CSA, the objective function is:

$$Objective\ function = \frac{1}{m \times N_T} \sum_{i=1}^{m} \sum_{j=1}^{N_T} \left( \frac{F_{ij}(\boldsymbol{q}) - \widehat{F_{ij}}}{\widehat{F_{ij}}} \right)^2 + \lambda \|\boldsymbol{q}\|_1 \tag{6.12}$$

Our target is to find the nest $\boldsymbol{q}_{best}$ such that the objective function has the minimum value. Each value of $\boldsymbol{q}$ in the definition domain is a feasible solution. The procedure of CSA is detailed in **Algorithm 6**. It is important to note that several parameters in CSA should be assigned proper values before starting the algorithm, such as the discovery rate, population size, step size in Lévy flights and maximum generation, etc. These parameters values are all chosen by experience in this study (Yang and Suash Deb, 2009).

### 6.2.3 Approximate $l_0$ sparse damage identification approach

By using CSA to solve Equation (6.8), the optimal solution can be obtained even though there is a nonlinear PCE function in the optimisation equation. However, dozens of candidate values of $\lambda$ are necessary to be tested in DP for determining the most appropriate one. Since the computational cost of performing CSA in DP is much higher than the traditional sparse recovery algorithms, the computational cost of performing CSA will become unacceptable. In order to improve the practicability of this damage identification approach, a modified DP that can work more efficient together with CSA is proposed in this section.

In the past studies on damage identification for bridge structures, it was commonly assumed that the number of damages does not exceed three due to the sparsity assumption (Hou et al., 2019;

Magalhães et al., 2012; Xu et al., 2018; Zhang and Xu, 2016). In Equation (6.8), the regularisation term is added to induce sparsity on the solution, and parameter $\lambda$ controls the solution sparsity. Choosing a proper regularisation parameter value by DP is, in fact, choosing the best solution sparsity. Thus, if predefining a set of candidate values of the regularisation parameter is changed as assuming the number of damages, the number of CSA calculations can be largely reduced, and the computational cost can be saved. Subsequently, the optimisation problem to be addressed becomes

$$\widehat{\boldsymbol{q}} = arg\min_{\boldsymbol{q}} \left\{ \frac{1}{m \times N_T} \sum_{i=1}^{m} \sum_{j=1}^{N_T} \left( \frac{F_{ij}(\boldsymbol{q}) - \widehat{F_{ij}}}{\widehat{F_{ij}}} \right)^2 \right\} s.t. \|\boldsymbol{q}\|_0 = N_d \tag{6.13}$$

in which $N_d \in \mathbb{Z}$, and $0 \leq N_d \leq N_k$. To ease the comparison, $N_k$ could be a bit larger value than 3 (e.g., 5). As can be seen, Equation (6.13) can be regarded as an approximate $l_0$ minimisation problem, which is NP-hard. Fortunately, CSA is superior in solving the NP-hard problem. Hence, by employing CSA to solve Equation (6.13) under each $N_d$, different residual values will be recorded. Then, the recorded residual values will be compared with the measurement noise to find the best $N_d$. Note that even if Equation (6.12) and Equation (6.13) are not equivalent in mathematics, they are both proposed to constrain the solution sparsity for the regression-based problem described by Equation (6.1). In essence, the target of identifying sparse damage can still be achieved, and thus solving Equation (6.12) here can be converted to solving Equation (6.13).

As an NP-hard problem, $l_0$ minimisation was not recommended for application in the past studies. Even if the heuristic algorithm can be used to address the NP-hard problem, the

computational cost will be unaffordable when the problem has a large scale. In this study, the damage identification problem can be deemed as a small-scale issue (the number of damages is extremely small), so the cost of applying heuristic algorithm to the $l_0$ minimisation problem is relatively low. To summarise, the steps of this approximate $l_0$ sparse damage identification approach are: (1) Predefine the maximum number of damages $N_k$; (2) Solve Equation (6.13) by using CSA under each $N_d$; (3) Compute the residual value under each $N_d$; (4) Compare the calculated residual values with the variance of the measurement noise to find the optimal $N_d$.

## 6.3 Case Studies

To verify the proposed damage identification approach, two cases that have been introduced in the previous chapter are used.

### 6.3.1 Numerical case of a three-span continuous beam bridge

The detailed information about this bridge can be found in Section 5.3.1.1. Moreover, the TL-MFPCE models that can precisely describe the relation between the frequencies and structural parameters of this bridge have been established in the previous chapter. Under the damage scenarios that defined in Table 5-2, the proposed approximate $l_0$ sparse damage identification approach is applied to identify the damages. For each damage scenario, parameter $N_k$ is set as 5, so CSA will perform 5 times. Without loss of generality, 20 frequency data under the temperature range $T = \{T_t, T_b : T_t, T_b \in [15℃, 25℃], T_t \geq T_b\}$ are used.

For damage scenario DS1, the residual change with respect to the number of damages is depicted in Figure 6.1.

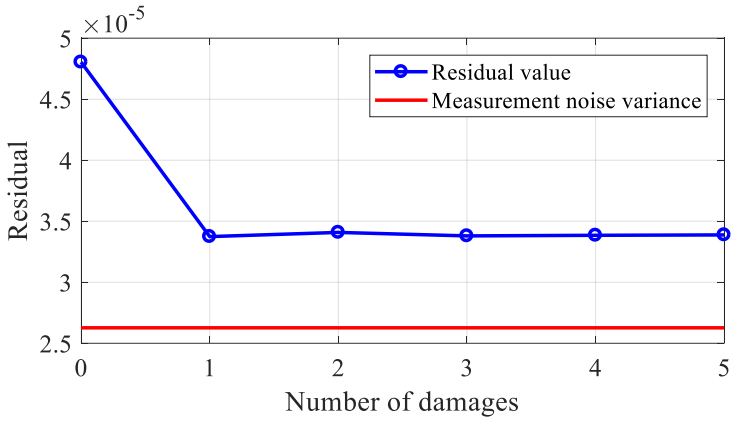
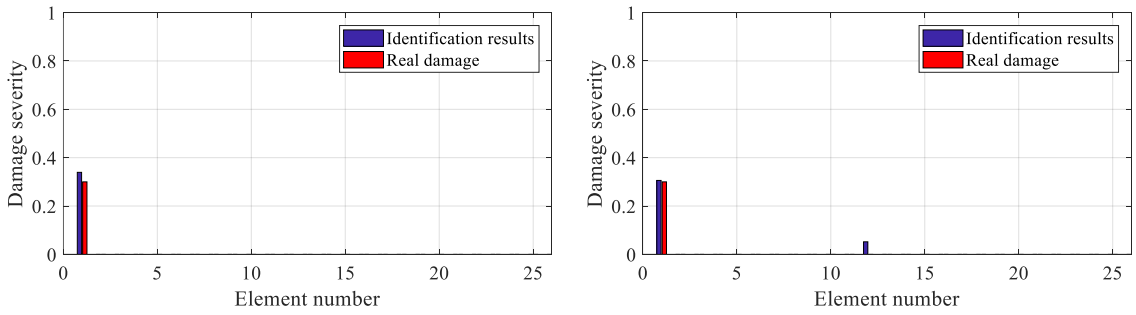



Figure 6.1 Residual values of DS1

The blue line with circles represents the calculated residuals, and the red line denotes the estimated noise variance when training the TL-MFPCE model. As can be seen, when assuming no damage occurs, the residual value is the largest and differs from the measurement noise variance. When the number of damages is assumed as 1, the calculated residual value decreases. Then, the increase of damage number will not contribute to the decrease of residual value. Therefore, the true damage number is 1. Figure 6.2 shows the optimal solutions obtained by CSA under each assumption. Since the true damage number is 1, the corresponding identification result is portrayed in Figure 6.2(a). The damage location and severity are correctly recognised.



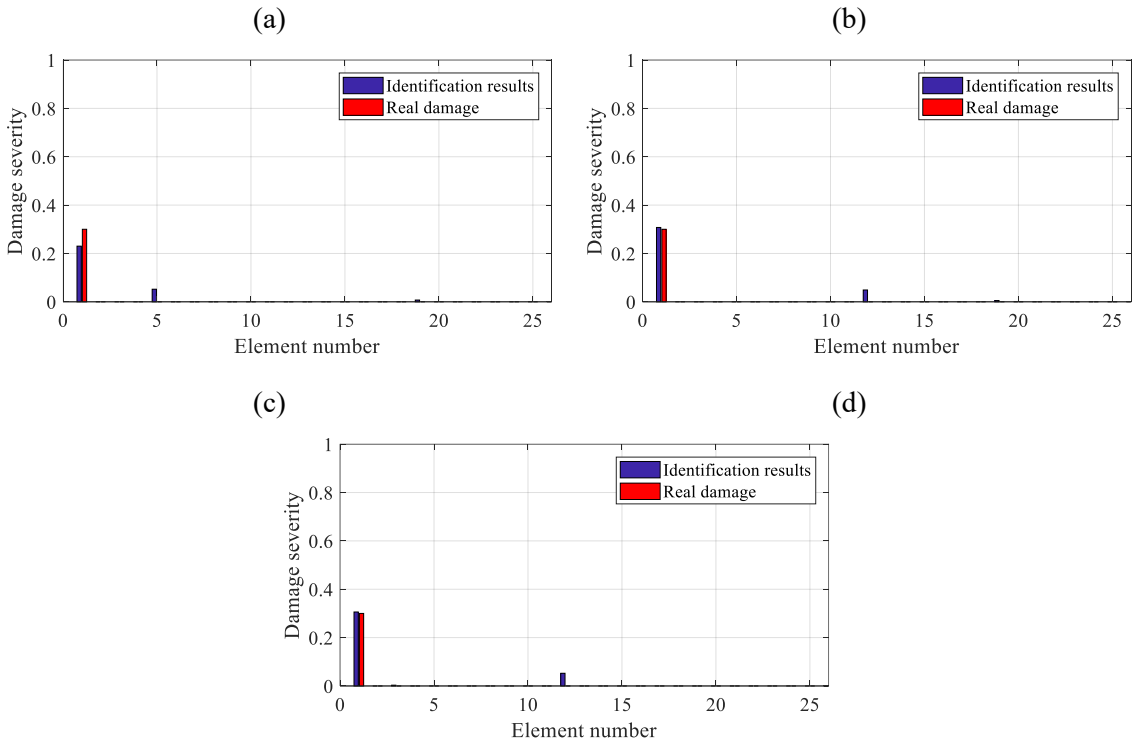



Figure 6.2 Damage identification results (DS1) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

For damage scenario DS2, the residuals compared to the estimated noise variance are depicted in Figure 6.3. Similar to Figure 6.1, the residual value is the largest when assuming no damage occurs. When the number of damages is assumed as 1, the residual value decreases and becomes very close to the estimated noise variance. After that, the residual value has no more fluctuation when more damage is assumed. Hence, the true damage number is recognised as 1. Then, the damage identification results from CSA under each assumption are given in Figure 6.4. As shown in Figure 6.4(a), the damage location and severity are almost correctly identified.

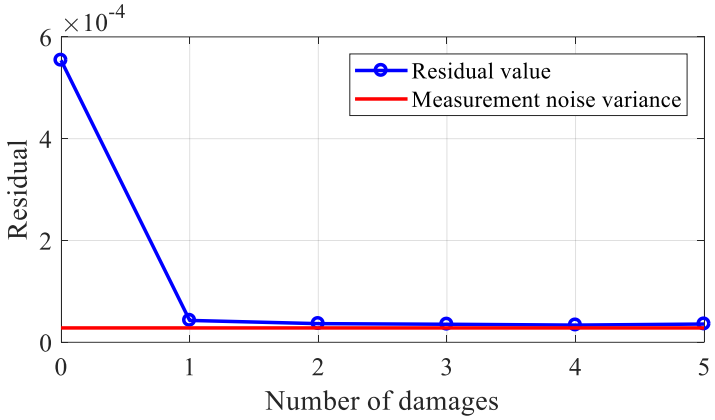



Figure 6.3 Residual values of DS2

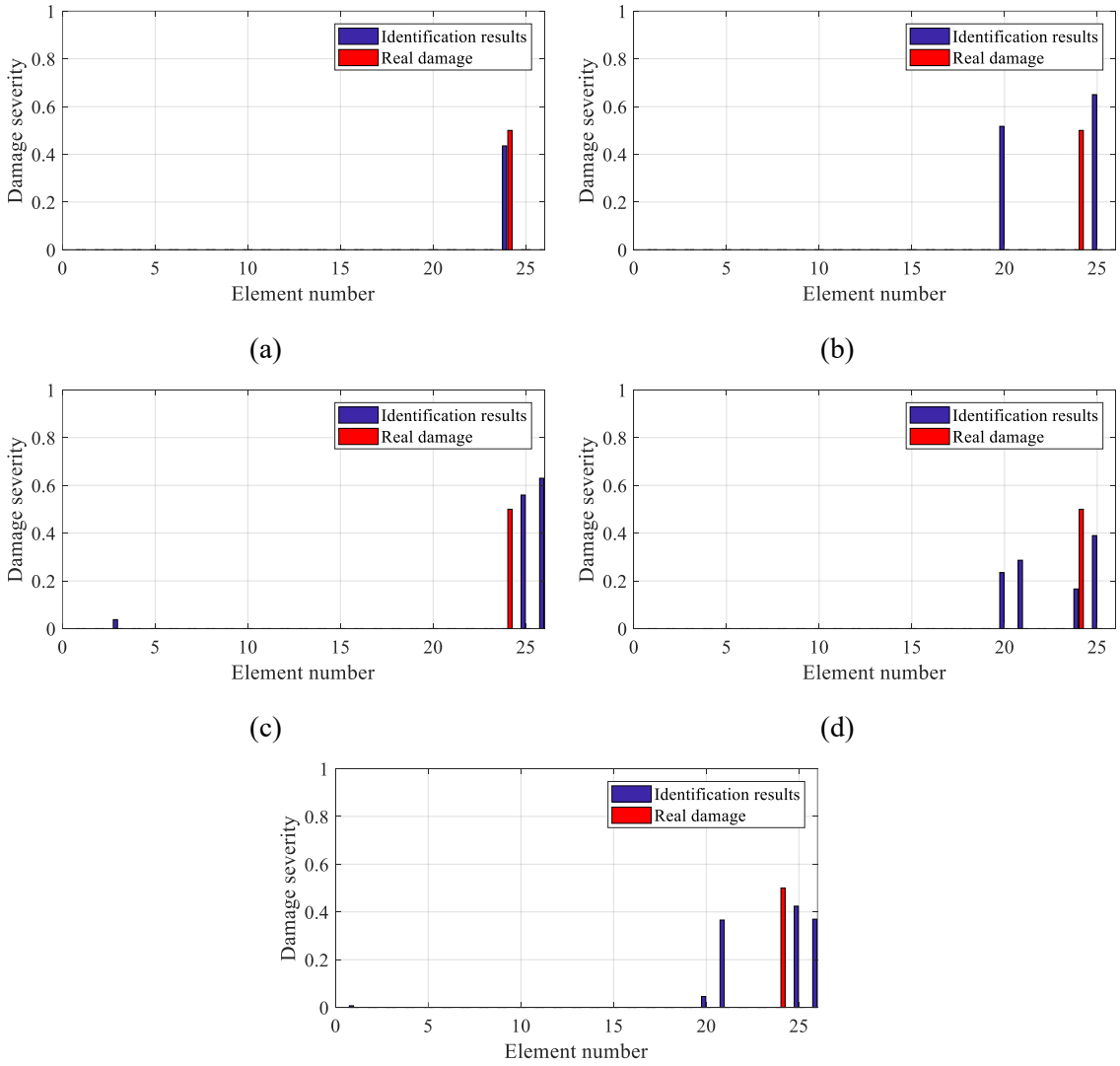



Figure 6.4 Damage identification results (DS2) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

For damage scenario DS3 with two damages, the residuals under different assumptions are

depicted in Figure 6.5. In this damage scenario, the residual value become close to the estimated noise variance as the assumed damage number exceeds 2, so the true damage number is 2. The corresponding damage identification result is shown in Figure 6.6. As displayed in Figure 6.6(b), the damage locations and severities are both obtained with desired accuracy.

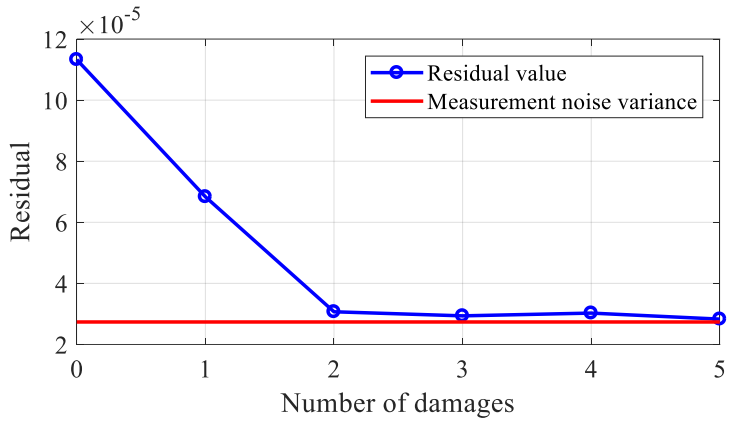
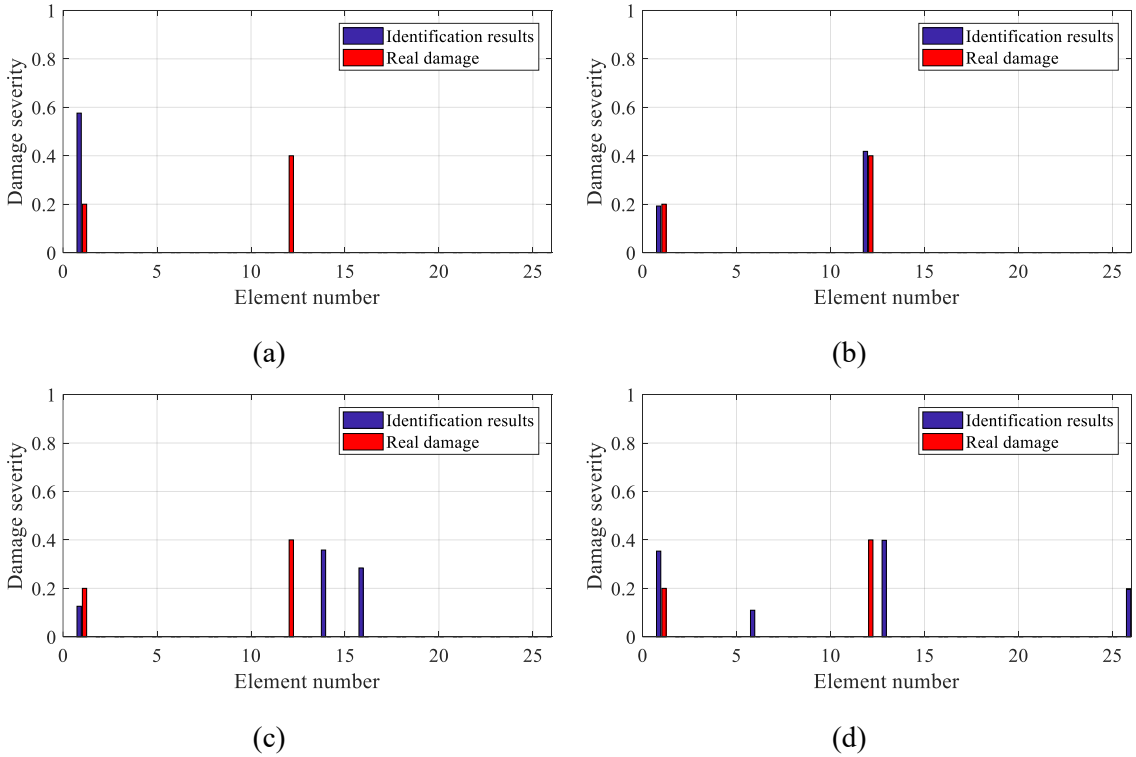
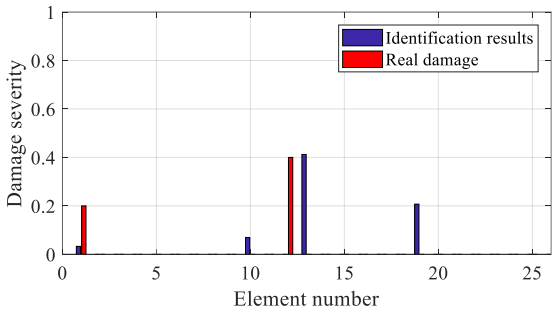



Figure 6.5 Residual values of DS3






(e)

Figure 6.6 Damage identification results (DS3) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

### 6.3.2 Experimental case of a two-span continuous beam

The detailed information about this experimental model has been elaborated in Section 5.3.1.2, and the TL-MFPCE models to describe the pattern between frequencies and the structural parameters have also been established in the previous chapter. Then, the proposed approach is applied to identify the damage in the scenarios as given in Table 5-4. For each damage scenario, parameter $N_k$ is set as 5, so CSA will perform 5 times. The measurements from structure under different damage scenarios are the same to those shown in Section 5.3.3 (Table 5-5).

For damage scenario DS1, the variation of residual with respect to the assumed number of damages is displayed in Figure 6.7.

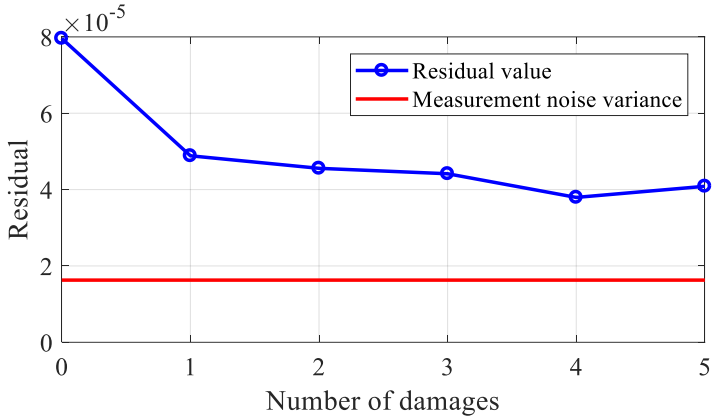
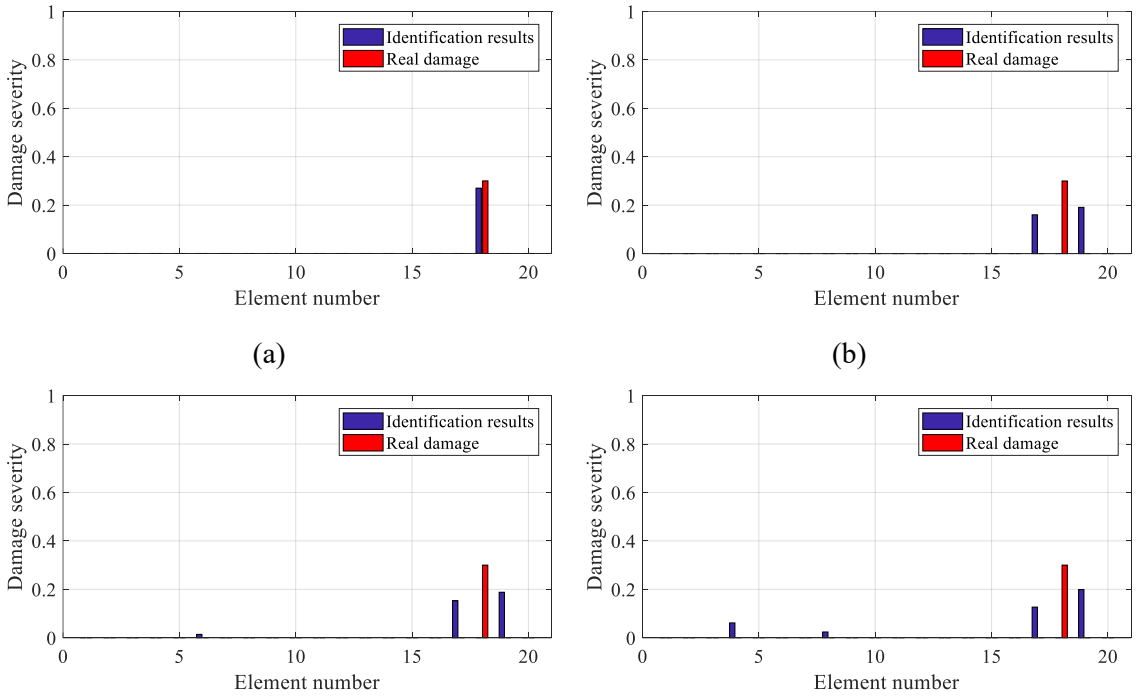
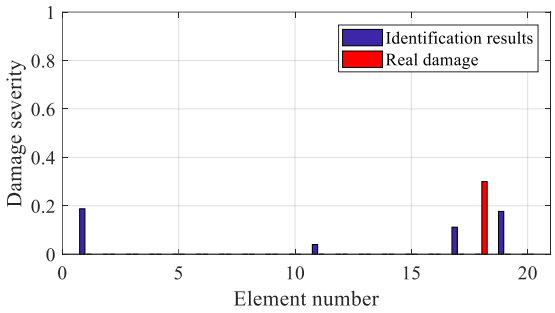



Figure 6.7 Residual values of DS1

As can be seen, the residual value is the largest when assuming no damage occurs, and it decreases to a low value and has a small fluctuation when the assumed number of damages exceeds 1. Therefore, the true damage number is 1. The optimal solution under each assumption is given in Figure 6.8, and Figure 6.8(a) gives the final result. Obviously, the damage location and severity are correctly recognised.






(e)

Figure 6.8 Damage identification results (DS1) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

In damage scenario DS2, the variation of residuals is shown in Figure 6.9. Intuitively, we can recognise that the true damage is 1. Then, the optimal solution is displayed in Figure 6.10(a). The damage is successfully identified.

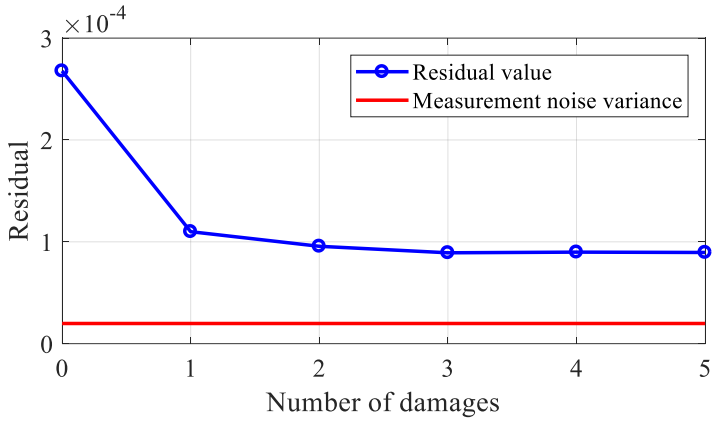
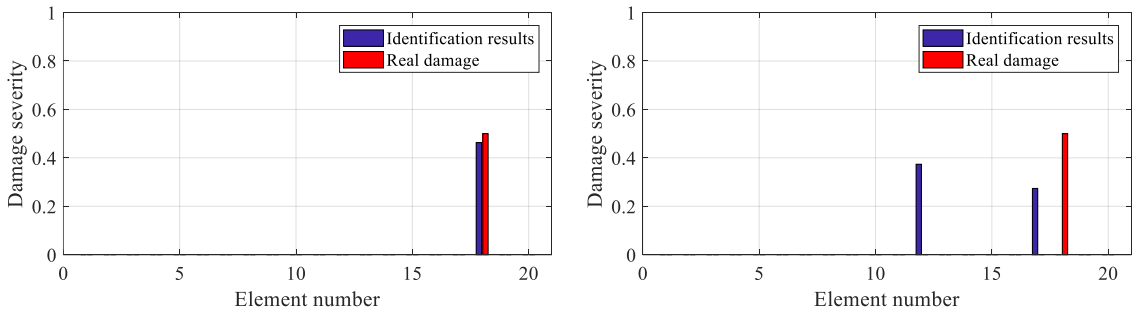
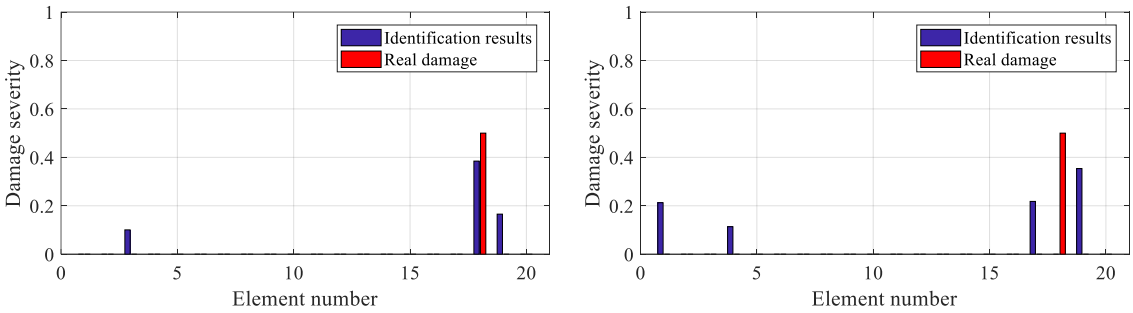


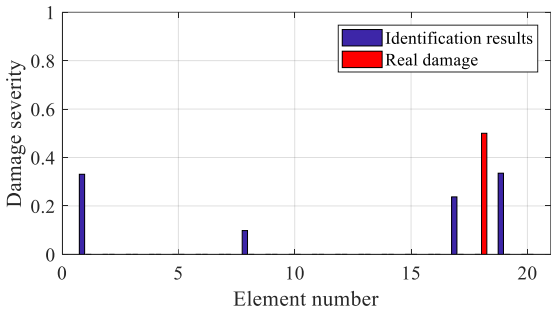

Figure 6.9 Residual values of DS2




(a) (b)




(c) (d)




(e)

Figure 6.10 Damage identification results (DS2) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

The results of damage scenario DS3 are portrayed in Figure 6.11 and Figure 6.12. As displayed in Figure 6.11, it is clear that there are two damages in the structure. As shown in Figure 6.12(b), the damage locations and severities are correctly quantified.

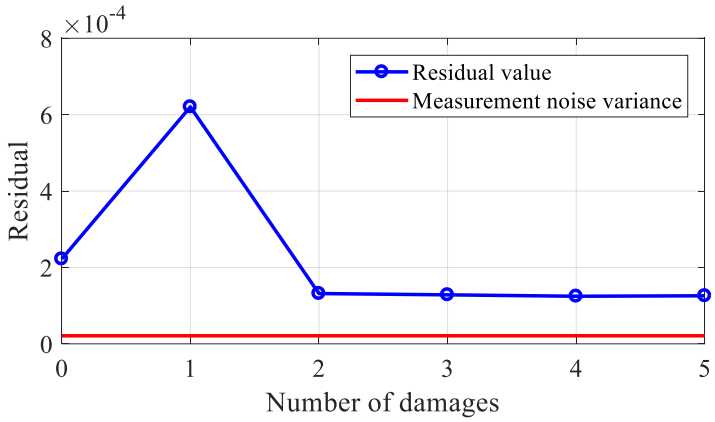



Figure 6.11 Residual values of DS3

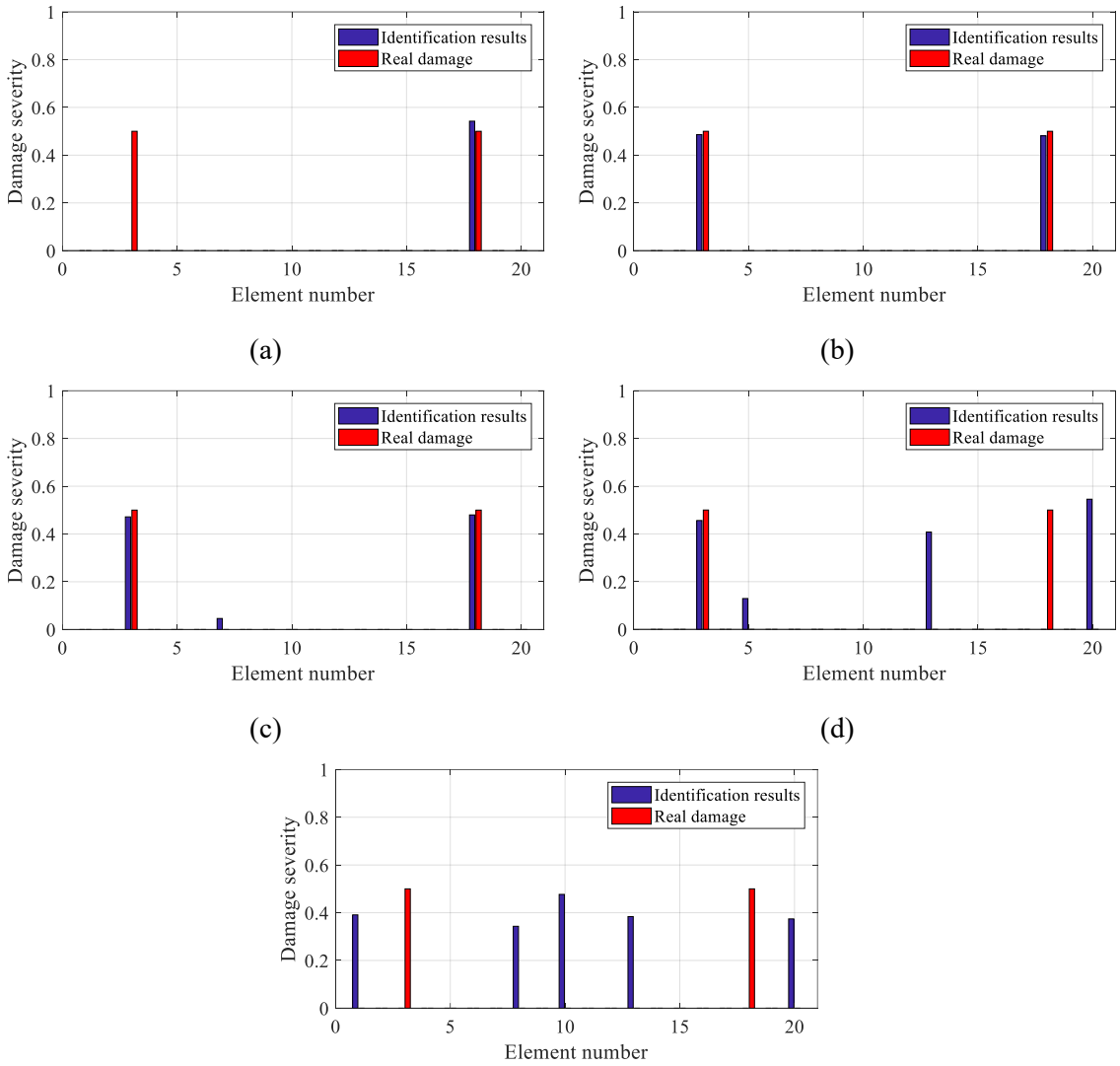



Figure 6.12 Damage identification results (DS3) under assumption of damage number (a) one; (b) two; (c) three; (d) four; (e) five

## 6.4 Summary

A novel sparse damage identification technique for nonlinear system, called approximate $l_0$ sparse damage identification, has been developed in this chapter. Through giving assumptions to the number of damages on the structure, the damage identification issue is transformed into an approximate $l_0$ sparse representation problem. Subsequently, a heuristic searching algorithm, Cuckoo Search Algorithm (CSA), is leveraged to find the optimal solution of this problem under

each assumption. Ultimately, the best solution sparsity (number of damages) could be determined by employing the Discrepancy Principle (DP), and the corresponding damage locations and severities can both be obtained via CSA. Apparently, with the aid of CSA, the proposed approach overcome the challenge in the traditional sparse representation methods that the system should be linear. Meanwhile, choosing proper regularisation parameter value in DP is replaced by giving assumptions to the number of damages to control the solution sparsity. The computational burden of performing heuristic algorithm with DP is largely lightened. However, this strategy is limited to solving problem with small scale. For large scale problem, the computational cost of the heuristic algorithm is still a heavy burden.

The proposed approach is validated on a simulation bridge model and an experimental beam model based on the TL-MFPCE models established in the previous chapter. Results show that the proposed approach has promising performance in identifying single and multiple damages. The variance of measurement noise that estimated in training the TL-MFPCE models contributes to the damage identification in this chapter. Furthermore, in the experimental case, different number of measurements are tested. Results manifest that the quantity of measurements has low impact to the performance of this strategy. However, the estimated noise variances of different frequencies in the experimental case are inconsistent with the assumption in Equation (6.7). The reason might be that the measurements were collected within short time intervals. In practice, it is recommended that the training data should be collected over a certain period to cover sufficient samples to increase the

modelling precision.

# CHAPTER 7

# CONCLUSIONS AND FUTURE WORKS

## 7.1 Conclusions

This thesis aims to apply the Polynomial Chaos Expansion (PCE) method for building surrogate model of a bridge structure and then use the surrogate model to help in damage identification. Due to the sophisticated mechanism of temperature effect on bridge structures, it is normally impractical to simulate the temperature effect correctly in a physical model. Thus, a Transfer Learning (TL) based Multi-Fidelity (MF) modelling technique is proposed to update the surrogate model instead of the physical model for the sake of eliminating the temperature-induced modelling error. As a result, damage identification based on the updated surrogate model can achieve accurate results.

First of all, to enable the PCE method available for modelling problems with high input dimension or high complexity, a novel sequential sampling strategy and an adaptive basis selection strategy are proposed. The PCE model is thus established adaptively by determining the sample set and truncation degree automatically, and the computational burden in the training process is reduced to a large extent. Then, PCE models as surrogate to a bridge structure are established and trained depending on the finite element model. Next, a TL-based MFPCE (TL-MFPCE) approach is

developed to update the modelling error caused by temperature in the PCE models, where only the measurement data from structure in healthy condition is necessary. Ultimately, a sparse damage identification method for nonlinear system is presented, which is able to identify the damage locations and severities based on the updated PCE models. A numerical bridge case and an experimental beam structure are utilised to validate the effectiveness of the proposed TL-MFPCE approach and the sparse damage identification method. The primary contributions of this study are summarised and highlighted as follows:

**(1) A novel hybrid sequential sampling strategy is developed for PCE modelling.**

The quantity and distribution of samples will greatly affect the training precision of a PCE model. Generally, the traditional sampling strategies for PCE are input-dependent only or output-dependent only, by which the modelling accuracy and results stability cannot be pursued simultaneously. Therefore, a novel sampling strategy for PCE modelling is developed, which intends to leverage both the input information and model feature to instruct a sequential sampling process. The samples can be collected with high quality and in relatively small quantity, and meanwhile, the best sample number can be automatically determined. Besides, sparse representation is employed in this strategy for regression calculation, thereby further reducing the computational cost. Benchmark tests on several functions and simulation cases have demonstrated that the proposed strategy has higher convergence speed and modelling stability compared with those traditional sampling approaches.

**(2) An adaptive basis selection strategy is proposed for PCE modelling.**

It is always a challenge to determine the most appropriate truncation degree of PCE model in advance when little information about the target problem is available. In this study, a new adaptive basis selection strategy is proposed, which can automatically determine the best truncation degree by identifying significant basis terms and removing insignificant basis terms. Moreover, the regression calculation cost for PCE training is also reduced benefitting from removing the insignificant basis terms.

**(3) A stability evaluation strategy is proposed to reconcile the sequential sampling process and the adaptive basis selection process in an adaptive PCE modelling framework.**

In practice, the sampling issue and the adaptive modelling problem co-exist in an PCE modelling task, and they should be concurrently addressed. In this study, a two-loop framework is constructed, in which the sequential sampling process is nested as the inner loop and the adaptive basis selection process is introduced as the outer loop. A novel stability evaluation approach is presented to determine which loop should be performed in each iteration. More importantly, this framework enables the sparse representation to be integrated in the adaptive modelling for the sake of cost saving. By validating on several benchmark functions, it is demonstrated that the proposed adaptive modelling framework is more practical than the traditional adaptive modelling strategies. Furthermore, this framework is also developed for addressing the multi-output modelling problem.

**(4) A TL-MFPCE approach is developed to eliminate the temperature-induced modelling error.**

Due to the complex mechanism of temperature in affecting bridge structures, the physical model of a bridge structure generally contains modelling errors that caused by the incorrect consideration of temperature. Based on the proposed adaptive PCE modelling framework, the PCE models as surrogate to the physical model is formulated. After that, a TL-MFPCE approach is proposed to update the PCE models under the circumstance that only the measurements from structure in healthy condition are available. The coefficients of the basis terms in the PCE models associated with the structural parameters are frozen, and the remaining coefficients related to the temperature variables are updated. Finally, the updated PCE models are able to predict the structural responses correctly. Through validating on a numerical bridge case and an experimental beam case by comparing with the traditional MFPCE (tMFPCE) method, it is demonstrated that the TL-MFPCE models outperform the tMFPCE models and have the ability to predict the structural responses accurately even though the structure is damaged.

**(5) A sparse damage identification method for nonlinear system is proposed.**

As most of the traditional sparse representation methods were presented for linear system, a novel approximate $l_0$ sparse damage identification method is developed in this study to work with nonlinear PCE model. By formulating the damage identification problem into an approximate $l_0$

regularised optimisation problem, Discrepancy Principle (DP) is employed to help determine the solution sparsity, and Cuckoo Search Algorithm (CSA) is introduced to obtain the final solution of this problem. The damage locations and severities are both received. Studies on a numerical case and an experimental case demonstrate the effectiveness of the proposed method in identifying single and multiple damages.

## 7.2 Recommendations for Future Works

Although the improvements on the PCE surrogate modelling technique and damage identification methods are developed and demonstrated, there are still some issues that deserve further research:

(1) The "curse of dimensionality" issue is an inherent challenge in the PCE method. Even though the proposed adaptive PCE modelling framework could help reduce the computational cost, it is still difficult for PCE to cope with problems with extremely large input dimension. For those large-scale bridges that have thousands of structural parts, thousands of input variables should be defined for building the corresponding surrogate model. Thus, more effective modelling approaches or strategies are required to overcome this challenge.

(2) In this research, only frequency data is interpreted for damage identification, so further studies can investigate more other structural characteristics, such as modal shape, strain or even image.

(3) Apart from the temperature effect, many other factors may also lead to modelling errors in a physical model, such as the model simplification and incorrect parameter values, etc. Considering the situation of real structure in a comprehensive way still need more in-depth studies.

(4) The proposed TL-MFPCE approach and sparse damage identification method have been validated in a simulation case and an experimental case. Further studies on their application to practical bridge structures are still required.

# APPENDIX A

The EI-ELF criterion is a sequential sampling method, which leverages the information from the PCE modelling result in each iteration to give sampling instruction of next point when Bayesian regression approaches are employed. A candidate samples pool $\boldsymbol{\zeta}_{cand}$ is defined in advance, and samples drawn during the sampling process are all from this pool. With the calculated coefficients by Equations (3.22) and (3.23), the PCE model could make predictions at any point $\boldsymbol{\zeta}$ with a distribution:

$$Y_{pre}(\boldsymbol{\zeta}) \sim N\left(\hat{y}(\boldsymbol{\zeta}), \sigma^2(\boldsymbol{\zeta})\right) \tag{A.1}$$

in which $\hat{y}(\boldsymbol{\zeta}) = \boldsymbol{\Psi}_{new}(\boldsymbol{\zeta})\mu_c$ and $\sigma^2(\boldsymbol{\zeta}) = \boldsymbol{\Psi}_{new}(\boldsymbol{\zeta})\Sigma_c\boldsymbol{\Psi}_{new}(\boldsymbol{\zeta})^T + \beta^{-1}$. By assuming that the observation of the target model at a point $\boldsymbol{\zeta}$ is $Y(\boldsymbol{\zeta})$, the quadratic loss function is defined as:

$$L(\boldsymbol{\zeta}) = \left(Y(\boldsymbol{\zeta}) - Y_{pre}(\boldsymbol{\zeta})\right)^2 \tag{A.2}$$

The expectation of the loss function is defined as the Expected Loss Function (ELF) which can be decomposed as (Zhou et al., 2019a):

$$E\left(L(\boldsymbol{\zeta})\right) = E\left(\left(Y(\boldsymbol{\zeta}) - \hat{y}(\boldsymbol{\zeta}) + \hat{y}(\boldsymbol{\zeta}) - Y_{pre}(\boldsymbol{\zeta})\right)^2\right)$$
$$= E\left(\left(Y(\boldsymbol{\zeta}) - \hat{y}(\boldsymbol{\zeta})\right)^2\right) + E\left(\left(\hat{y}(\boldsymbol{\zeta}) - Y_{pre}(\boldsymbol{\zeta})\right)^2\right) \quad \text{(A.3)}$$
$$+ E\left(\left(Y(\boldsymbol{\zeta}) - E\left(Y(\boldsymbol{\zeta})\right)\right)^2\right)$$

The first term $E\left(\left(Y(\boldsymbol{\zeta}) - \hat{y}(\boldsymbol{\zeta})\right)^2\right)$ is the bias that represents the average difference between the predicted response and the target model output; the second term $E\left(\left(\hat{y}(\boldsymbol{\zeta}) - Y_{pre}(\boldsymbol{\zeta})\right)^2\right)$ denotes the prediction variance of the PCE model, which equals to $\sigma^2(\boldsymbol{\zeta})$; the third term $E\left(\left(Y(\boldsymbol{\zeta}) - E\left(Y(\boldsymbol{\zeta})\right)\right)^2\right)$ represents the intrinsic noise of the observations, which can be ignored since the prediction accuracy of PCE does not depend on this term. Then, ELF can be rewritten as:

$$ELF(\boldsymbol{\zeta}) \approx \left(Y(\boldsymbol{\zeta}) - \hat{y}(\boldsymbol{\zeta})\right)^2 + \sigma^2(\boldsymbol{\zeta}) \quad \text{(A.4)}$$

Clearly, the true observation $Y(\boldsymbol{\zeta})$ is unknown at an arbitrary point $\boldsymbol{\zeta}$ in the input space, so an approximation term is used to replace the first term in Equation (A.4), which is called expected improvement criterion (Beck and Katafygiotis, 1998; Zhou et al., 2019a):

$$EI(\boldsymbol{\zeta}) = \left(Y(\boldsymbol{\zeta_m}) - \hat{y}(\boldsymbol{\zeta})\right)^2 T(N, P_s) \quad \text{(A.5)}$$

where $Y(\boldsymbol{\zeta_m})$ represents the observed response of the target model at the known point $\boldsymbol{\zeta_m}$ which has the least Euclidean distance to the candidate point $\boldsymbol{\zeta} \in \boldsymbol{\zeta}_{cand}$; $T(N, P_s) = \frac{N}{N-P_s}\left(1 + tr\left(\left(\boldsymbol{\Psi}^{\boldsymbol{T}}\boldsymbol{\Psi}\right)^{-1}\right)\right)$ is a correction factor for the regression with a small number of samples, which can reduce the sensitivity of the bias estimate to overfitting or underfitting (Chapelle et al., 2002);

$N$ is the sample number for training the PCE model and $P_s$ denotes the cardinality of the trained PCE model. As a result, the EI-ELF criterion can be written as:

$$ELF_{EI}(\boldsymbol{\zeta}) = \left(Y(\boldsymbol{\zeta_m}) - \hat{y}(\boldsymbol{\zeta})\right)^2 T(N, P_s) + \sigma^2(\boldsymbol{\zeta}) \tag{A.6}$$

Finally, the sequential sampling process based on the EI-ELF criterion is to select the new sample in each iteration to maximise the following equation:

$$\boldsymbol{\zeta}_{new} = \underset{\boldsymbol{\zeta} \in \boldsymbol{\xi}_{cand}}{\operatorname{argmax}}\, ELF_{EI}(\boldsymbol{\zeta}) \tag{A.7}$$

It can be seen that the first term in Equation (A.6) would like to improve the estimation accuracy at the neighbour of observed points and the second term prefers to collect more information in the regions with large uncertainty. So, this criterion acts to achieve trade-off between the global exploration and local exploitation. This criterion has been demonstrated to be an effective sequential sampling method that the trained PCE model will have a high precision while the samples can be controlled to a small quantity (Zhou et al., 2019a).

# APPENDIX B

The ANOVA method is a variance-based global sensitivity analysis method, which works on the entire input domain and analyses the influence of each independent input variable on the model output. Suppose a model can be decomposed as the following summands (Crestaux et al., 2009):

$$f(\xi_1, \xi_2, \cdots, \xi_d) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{1 \le i < j \le d} f_{ij}(\xi_i, \xi_j) + \cdots + f_{12\cdots d}(\xi_1, \xi_2, \cdots, \xi_d) \tag{B.1}$$

and

$$\int_D f_{i_1 i_2 \cdots i_k}(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k})\, p(\xi_{i_r}) d\xi_{i_r} = 0 \quad r \in [1, k] \quad k \le d \tag{B.2}$$

where $D$ is the definition domain of input variables; $p(\xi_{i_r})$ is the marginal probabilistic density function of $\boldsymbol{\xi}_{i_r}$. The total variance of the model $f(\xi_1, \xi_2, \cdots, \xi_d)$ is decomposed into the summation of each summand variance which can be simply calculated by integration. The total variance is obtained by:

$$V_{total} = \sum_{1 \le i \le d} V_i + \sum_{1 \le i < j \le d} V_{i,j} + \cdots + V_{1,2,\cdots d} \tag{B.3}$$

where

$$V_{i_1 i_2 \cdots i_k} = \int_D \cdots \int_D f_{i_1 i_2 \cdots i_k}(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k})^2 \, p(\xi_{i_1} \cdots \xi_{i_k}) d\xi_{i_1} \cdots d\xi_{i_k} \quad k \leq d \tag{B.4}$$

Subsequently, the variance of each summand can be normalised by the total variance $V_{total}$, and the variance contribution of each summand can be obtained by

$$S_{i_1 i_2 \cdots i_k} = \frac{V_{i_1 i_2 \cdots i_k}}{V_{total}} \tag{B.5}$$

To quantify the variance contribution of each variable, the first-order index is defined as the $S$ value that only contains variance value from one variable in numerator, which represents the independent variance contribution from this variable to the model output. The total index is the sum of all $S$ values that involve the target variable in numerator, including interaction terms with other variables. By comparing the first-order index and the total order index, the interaction effect of input variables on the output can be identified. When applying ANOVA based on the PCE model, the calculation of indices becomes convenient since the multivariate orthogonal polynomials satisfy the condition in Equation (B.2). The ANOVA indices can be easily calculated from the PCE coefficients.

# REFERENCES

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., and Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. Journal of Sound and Vibration, 388, 154–170.

Alemazkoor, N., and Meidani, H. (2018). A near-optimal sampling strategy for sparse recovery of polynomial chaos expansions. Journal of Computational Physics, 371, 137–151.

Alış, Ö. F., and Rabitz, H. (2001). Efficient implementation of high dimensional model representations. Journal of Mathematical Chemistry, 29(2), 127–142.

Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2010). Bayesian compressive sensing using Laplace priors. IEEE Transactions on Image Processing, 19(1), 53–63.

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

Beck, J. L., and Katafygiotis, L. S. (1998). Updating models and their uncertainties. I: Bayesian statistical framework. Journal of Engineering Mechanics, 124(4), 455–461.

Behmanesh, I., and Moaveni, B. (2016). Accounting for environmental variability, modeling errors, and parameter estimation uncertainties in structural identification. Journal of Sound and

Vibration, 374, 92–110.

Berchier, M. (2016). Multi-fidelity surrogate modelling with polynomial chaos expansions. Master thesis, Swiss Federal Institute of Technology.

Bigoni, C., and Hesthaven, J. S. (2020). Simulation-based anomaly detection and damage localization: An application to structural health monitoring. Computer Methods in Applied Mechanics and Engineering, 363, 112896.

Blatman, G. (2009). Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis. Doctoral dissertation, Blaise Pascal University.

Blatman, G., and Sudret, B. (2008). Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. Comptes Rendus Mécanique, 336(6), 518–523.

Blatman, G., and Sudret, B. (2010). An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. Probabilistic Engineering Mechanics, 25(2), 183–197.

Blatman, G., and Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. Journal of Computational Physics, 230(6), 2345–2367.

Bruckstein, A. M., Donoho, D. L., and Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review, 51(1), 34–81.

Burnaev, E., Panin, I., and Sudret, B. (2017). Efficient design of experiments for sensitivity analysis based on polynomial chaos expansions. Annals of Mathematics and Artificial Intelligence, 81(1–2), 187–207.

Cameron, R. H., and Martin, W. T. (1947). The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. The Annals of Mathematics, 48(2), 385.

Candes, E. J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2), 489–509.

Chakraborty, S. (2021). Transfer learning based multi-fidelity physics informed deep neural network. Journal of Computational Physics, 426, 109942.

Chaloner, K., and Verdinelli, I. (1995). Bayesian experimental design: A review. Statistical Science, 10(3), 273–304.

Chapelle, O., Vapnik, V., and Bengio, Y. (2002). Model selection for small sample regression. Machine Learning, 48(1–3), 9–23.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing, 20(1), 33–61.

Chen, S. X., Zhou, L., Ni, Y. Q., and Liu, X. Z. (2021). An acoustic-homologous transfer learning

approach for acoustic emission–based rail condition evaluation. Structural Health Monitoring, 20(4), 2161–2181.

Chen, Z., Zhang, R., Zheng, J., and Sun, H. (2020). Sparse Bayesian learning for structural damage identification. Mechanical Systems and Signal Processing, 140, 106689.

Cheng, K., and Lu, Z. (2020). Active learning polynomial chaos expansion for reliability analysis by maximizing expected indicator function prediction error. International Journal for Numerical Methods in Engineering, 121(14), 3159–3177.

Cheng, K., Lu, Z., and Zhen, Y. (2019). Multi-level multi-fidelity sparse polynomial chaos expansion based on Gaussian process regression. Computer Methods in Applied Mechanics and Engineering, 349, 360–377.

Choi, S. K., Grandhi, R. V., Canfield, R. A., and Pettit, C. L. (2004). Polynomial chaos expansion with Latin hypercube sampling for estimating response variability. AIAA Journal, 42(6), 1191–1198.

Crestaux, T., Le Maıˆtre, O., and Martinez, J.-M. (2009). Polynomial chaos expansion for sensitivity analysis. Reliability Engineering & System Safety, 94(7), 1161–1172.

Crombecq, K., Laermans, E., and Dhaene, T. (2011). Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. European Journal of Operational Research, 214(3), 683–696.

Das, A., and Debnath, N. (2018). A Bayesian finite element model updating with combined normal and lognormal probability distributions using modal measurements. Applied Mathematical Modelling, 61, 457–483.

De'Ath, G., and Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology, 81(11), 3178–3192.

Debusschere, B. J., Najm, H. N., Pébay, P. P., Knio, O. M., Ghanem, R. G., and Le Maı̂tre, O. P. (2004). Numerical challenges in the use of polynomial chaos representations for stochastic processes. SIAM Journal on Scientific Computing, 26(2), 698–719.

Deng, Y., Ding, Y., and Li, A. (2010). Structural condition assessment of long-span suspension bridges using long-term monitoring data. Earthquake Engineering and Engineering Vibration, 9(1), 123-131.

Deraemaeker, A., Reynders, E., De Roeck, G., and Kullaa, J. (2008). Vibration-based structural health monitoring using output-only measurements under changing environment. Mechanical Systems and Signal Processing, 22(1), 34–56.

Diaz, P., Doostan, A., and Hampton, J. (2018). Sparse polynomial chaos expansions via compressed sensing and D-optimal design. Computer Methods in Applied Mechanics and Engineering, 336, 640–666.

Diez-Olivan, A., Del Ser, J., Galar, D., and Sierra, B. (2019). Data fusion and machine learning for

industrial prognosis: Trends and perspectives towards Industry 4.0. Information Fusion, 50(October 2018), 92–111.

Ding, Z., Li, J., and Hao, H. (2019). Structural damage identification using improved Jaya algorithm based on sparse regularization and Bayesian inference. Mechanical Systems and Signal Processing, 132, 211–231.

Dong, J., Zhang, G., Zhang, Z., Geng, Y., and Wang, J. (2018). Inverse problem solution and regularization parameter selection for current distribution reconstruction in switching arcs by inverting magnetic fields. Mathematical Problems in Engineering, 2018, 1–11.

Donoho, D. L. (2006). Compressed sensing. IEEE Transactions on Information Theory, 52(4), 1289–1306.

Eldred, M. S., Ng, L. W. T., Barone, M. F., and Domino, S. P. (2017). Multifidelity uncertainty quantification using spectral stochastic discrepancy models. In Handbook of Uncertainty Quantification (pp. 991–1036). Springer International Publishing.

Entezami, A., Shariatmadar, H., and Sarmadi, H. (2017). Structural damage detection by a new iterative regularization method and an improved sensitivity function. Journal of Sound and Vibration, 399, 285–307.

Erazo, K., Sen, D., Nagarajaiah, S., and Sun, L. (2019). Vibration-based structural health monitoring under changing environmental conditions using Kalman filtering. Mechanical Systems and

Signal Processing, 117, 1–15.

Ernst, O. G., Mugler, A., Starkloff, H.-J., and Ullmann, E. (2012). On the convergence of generalized polynomial chaos expansions. ESAIM: Mathematical Modelling and Numerical Analysis, 46(2), 317–339.

Fajraoui, N., Marelli, S., and Sudret, B. (2017). Sequential design of experiment for sparse polynomial chaos expansions. SIAM/ASA Journal on Uncertainty Quantification, 5(1), 1061–1085.

Fan, W., and Qiao, P. (2011). Vibration-based damage identification methods: A review and comparative study. Structural Health Monitoring, 10(1), 83–111.

Fang, S. E., and Perera, R. (2011). Damage identification by response surface based model updating using D-optimal design. Mechanical Systems and Signal Processing, 25(2), 717–733.

Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2007). Multi-fidelity optimization via surrogate modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 463(2088), 3251–3269.

Fricker, T. E., Oakley, J. E., Sims, N. D., and Worden, K. (2011). Probabilistic uncertainty analysis of an FRF of a structure using a Gaussian process emulator. Mechanical Systems and Signal Processing, 25(8), 2962–2975.

Funahashi, K., and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. Neural Networks, 6(6), 801–806.

Ghanem R. G., and Ghiocel D. M. (1998a). A new implementation of the spectral stochastic finite element method for stochastic constitutive relations. In: Proceedings of the 12th ASCE engineering mechanics conference, La Jolla, California, USA.

Ghanem R. G., and Ghiocel D. M. (1998b). Stochastic seismic soil–structure interaction using the homogeneous chaos expansion. In: Proceedings of the 12th ASCE engineering mechanics conference, La Jolla, California, USA.

Ghiocel, D. M., and Ghanem, R. G. (2002). Stochastic finite-element analysis of seismic soil–structure interaction. Journal of Engineering Mechanics, 128(1), 66–77.

Gomes, G. F., Mendéz, Y. A. D., da Silva Lopes Alexandrino, P., da Cunha, S. S., and Ancelotti, A. C. (2018). The use of intelligent computational tools for damage detection and identification with an emphasis on composites – A review. Composite Structures, 196(September 2017), 44–54.

Gregory, A., Lau, F. D. H., Girolami, M., Butler, L. J., and Elshafie, M. Z. E. B. (2019). The synthesis of data from instrumented structures and physics-based models via Gaussian processes. Journal of Computational Physics, 392, 248–265.

Guo, J., Jiao, J., Fujita, K., and Takewaki, I. (2020). A spectrum-driven damage identification by

minimum constitutive relation error and sparse regularization. Mechanical Systems and Signal Processing, 136, 106496.

Guo, X., Chen, L., and Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement, 93, 490–502.

Haftka, R. T. (1991). Combining global and local approximations. AIAA Journal, 29(9), 1523–1525.

Hämarik, U., and Raus, T. (2006). On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. Journal of Inverse and Ill-Posed Problems, 14(3), 251–266.

Hampton, J., and Doostan, A. (2015). Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. Journal of Computational Physics, 280, 363–386.

Han, Q., Ma, Q., Xu, J., and Liu, M. (2021). Structural health monitoring research under varying temperature condition: a review. Journal of Civil Structural Health Monitoring, 11(1), 149–173.

Hernandez, E. M. (2014). Identification of isolated structural damage from incomplete spectrum information using l1-norm minimization. Mechanical Systems and Signal Processing, 46(1), 59–69.

Hosder, S., Walters, R., and Balch, M. (2007). Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables. 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 3, 2946–2961.

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

Hou, R., and Xia, Y. (2021). Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019. Journal of Sound and Vibration, 491, 115741.

Hou, R., Xia, Y., Bao, Y., and Zhou, X. (2018a). Selection of regularization parameter for l1-regularized damage detection. Journal of Sound and Vibration, 423, 141–160.

Hou, R., Xia, Y., and Zhou, X. (2018b). Structural damage detection based on l 1 regularization using natural frequencies and mode shapes. Structural Control and Health Monitoring, 25(3), e2107.

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

Huan, X., Safta, C., Sargsyan, K., Vane, Z. P., Lacaze, G., Oefelein, J. C., and Najm, H. N. (2018). Compressive sensing with cross-validation and stop-sampling for sparse polynomial chaos expansions. SIAM/ASA Journal on Uncertainty Quantification, 6(2), 907–936.

Huang, H. B., Yi, T. H., Li, H. N., and Liu, H. (2020). Strain-based performance warning method for bridge main girders under variable operating conditions. Journal of Bridge Engineering, 25(4), 04020013.

Huang, M., Lei, Y., and Cheng, S. (2019). Damage identification of bridge structure considering temperature variations based on particle swarm optimization - cuckoo search algorithm. Advances in Structural Engineering, 22(15), 3262–3276.

Huang, M. S., Gül, M., and Zhu, H. P. (2018). Vibration-based structural damage identification under varying temperature effects. Journal of Aerospace Engineering, 31(3), 04018014.

Huang, Y., and Beck, J. L. (2015). Hierarchical sparse Bayesian learning for structural health monitoring with incomplete modal data. International Journal for Uncertainty Quantification, 5(2), 139–169.

Huang, Y., and Beck, J. L. (2013). Novel sparse Bayesian learning for structural health monitoring using incomplete modal data. Computing in Civil Engineering, 1, 121–128.

Huang, Y., Beck, J. L., and Li, H. (2017a). Hierarchical sparse Bayesian learning for structural damage detection: Theory, computation and application. Structural Safety, 64, 37–53.

Huang, Y., Beck, J. L., and Li, H. (2017b). 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, 318, 382–411.

Jakeman, J. D., Eldred, M. S., and Sargsyan, K. (2015). Enhancing l1-minimization estimates of polynomial chaos expansions using basis selection. Journal of Computational Physics, 289, 18–34.

Jayasundara, N., Thambiratnam, D. P., Chan, T. H. T., and Nguyen, A. (2020). Damage detection and quantification in deck type arch bridges using vibration based methods and artificial neural networks. Engineering Failure Analysis, 109(March 2019), 104265.

Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. IEEE Transactions on Signal Processing, 56(6), 2346–2356.

Jiang, T., Zhang, X., and Li, Y. (2018). Bayesian compressive sensing using reweighted Laplace priors. AEU - International Journal of Electronics and Communications, 97, 178–184.

Jiao, Y., Liu, H., Wang, X., Zhang, Y., Luo, G., and Gong, Y. (2014). Temperature Effect on Mechanical Properties and Damage Identification of Concrete Structure. Advances in Materials Science and Engineering, 2014, 191360.

Jin, S. S., Kim, S. T., and Park, Y. H. (2021). Combining point and distributed strain sensor for complementary data-fusion: A multi-fidelity approach. Mechanical Systems and Signal

Processing, 157, 107725.

Kareem, A. (1988). Effect of parametric uncertainties on wind excited structural response. Journal of Wind Engineering and Industrial Aerodynamics, 30(1–3), 233–241.

Kennedy, M., and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. Biometrika, 87(1), 1–13.

Kesavan, V., Kamalakannan, R., Sudhakarapandian, R., and Sivakumar, P. (2020). Heuristic and meta-heuristic algorithms for solving medium and large scale sized cellular manufacturing system NP-hard problems: A comprehensive review. Materials Today: Proceedings, 21, 66–72.

Kita, A., Cavalagli, N., and Ubertini, F. (2019). Temperature effects on static and dynamic behavior of Consoli Palace in Gubbio, Italy. Mechanical Systems and Signal Processing, 120, 180–202.

Knill, D. L., Giunta, A. A., Baker, C. A., Grossman, B., Mason, W. H., Haftka, R. T., and Watson, L. T. (1999). Response surface models combining linear and euler aerodynamics for supersonic transport design. Journal of Aircraft, 36(1), 75–86.

Krahmer, F., and Ward, R. (2014). Stable and robust sampling strategies for compressive imaging. IEEE Transactions on Image Processing, 23(2), 612–622.

Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. Journal of Machine Learning

Research, 9(2008), 235–284.

Kromanis, R., and Kripakaran, P. (2016). SHM of bridges: characterising thermal response and detecting anomaly events using a temperature-based measurement interpretation approach. Journal of Civil Structural Health Monitoring, 6(2), 237–254.

Lai, Z., and Nagarajaiah, S. (2019). Semi‐supervised structural linear/nonlinear damage detection and characterization using sparse identification. Structural Control and Health Monitoring, 26(3), 1–19.

Lam, H. F., Ko, J. M., and Wong, C. W. (1998). Localization of damaged structural connections based on experimental modal and sensitivity analysis. Journal of Sound and Vibration, 210(1), 91–115.

Lee, S., Park, S., Kim, T., Lieu, Q. X., and Lee, J. (2021). Damage quantification in truss structures by limited sensor-based surrogate model. Applied Acoustics, 172, 107547.

Leng, H., and He, Z. (2006). Computing eigenvalue bounds of structures with uncertain-but-non-random parameters by a method based on perturbation theory. Communications in Numerical Methods in Engineering, 23(11), 973–982.

Liu, F., Li, H., Li, W., and Wang, B. (2014). Experimental study of improved modal strain energy method for damage localisation in jacket-type offshore wind turbines. Renewable Energy, 72, 174–181.

Liu, H., Ong, Y.-S., Cai, J., and Wang, Y. (2018a). Cope with diverse data structures in multi-fidelity modeling: A Gaussian process method. Engineering Applications of Artificial Intelligence, 67(August 2017), 211–225.

Liu, R., Yang, B., Zio, E., and Chen, X. (2018b). Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, 108, 33–47.

Liu, W. K., Belytschko, T., and Mani, A. (1986). Probabilistic finite elements for nonlinear structural dynamics. Computer Methods in Applied Mechanics and Engineering, 56(1), 61–81.

Loukrezis, D., Galetzka, A., and De Gersem, H. (2020). Robust adaptive least squares polynomial chaos expansions in high - frequency applications. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields, 33(6), 1–15.

Lukas, M. A. (1995). On the discrepancy principle and generalised maximum likelihood for regularisation. Bulletin of the Australian Mathematical Society, 52(3), 399–424.

Lüthen, N., Marelli, S., and Sudret, B. (2020). Automatic selection of basis-adaptive sparse polynomial chaos expansions for engineering applications. 1–32.

Lüthen, N., Marelli, S., and Sudret, B. (2021). Sparse polynomial chaos expansions: Literature survey and benchmark. SIAM/ASA Journal on Uncertainty Quantification, 9(2), 593–649.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. Neural

Computation, 4(4), 590–604.

Magalhães, F., Cunha, A., and Caetano, E. (2012). Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. Mechanical Systems and Signal Processing, 28, 212–228.

Mallat, S. G., and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing, 41(12), 3397–3415.

Marelli, S., and Sudret, B. (2018). An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. Structural Safety, 75(August 2017), 67–74.

Matheron, G. (1967). Kriging or polynomial interpolation procedures, The Canadian Mining and Metallurgical Bulletin. 60 (1967), 204-244.

McKeon, P., Yaacoubi, S., Declercq, N. F., Ramadan, S., and Yaacoubi, W. K. (2014). Baseline subtraction technique in the frequency–wavenumber domain for high sensitivity damage detection. Ultrasonics, 54(2), 592–603.

Migliorati, G., Nobile, F., von Schwerin, E., and Tempone, R. (2014). Analysis of discrete L2 projection on polynomial spaces with random evaluations. Foundations of Computational Mathematics, 419–456.

Montgomery, D. C. (2006). Design and analysis of experiments. Technometrics, 48(1), 158–158.

Mosavi, A. A., Dickey, D., Seracino, R., and Rizkalla, S. (2012). Identifying damage locations under ambient vibrations utilizing vector autoregressive models and Mahalanobis distances. Mechanical Systems and Signal Processing, 26(1), 254–267.

Natarajan, B. K. (1995). Sparse approximate solutions to linear systems. SIAM Journal on Computing, 24(2), 227–234.

Needell, D. (2009). Noisy signal recovery via iterative reweighted L1-minimization. 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA.

Neves, A. C., González, I., Leander, J., and Karoumi, R. (2017). Structural health monitoring of bridges: a model-free ANN-based approach to damage detection. Journal of Civil Structural Health Monitoring, 7(5), 689–702.

Ng, L. W. T., and Eldred, M. (2012). Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Honolulu, Hawaii, USA.

Ni, F., Nguyen, P. H., and Cobben, J. F. G. (2017). Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow. IEEE Transactions on Power Systems, 32(1), 694–704.

Ni, P., Xia, Y., Li, J., and Hao, H. (2019). Using polynomial chaos expansion for uncertainty and sensitivity analysis of bridge structures. Mechanical Systems and Signal Processing, 119, 293–311.

Ni, Y. Q., Hua, X. G., Fan, K. Q., and Ko, J. M. (2005). Correlating modal properties with temperature using long-term monitoring data and support vector machine technique. Engineering Structures, 27(12), 1762–1773.

Ni, Y. Q., Zhou, X. T., Ko, J. M., and Wang, B. S. (2000). Vibration-based damage localization in Ting Kau Bridge using probabilistic neural network. Advances in Structural Dynamics, 2, 1069–1076.

Nielsen, F. (2019). On the Jensen–Shannon symmetrization of distances relying on abstract means. Entropy, 21(5), 485.

Olleak, A., and Xi, Z. (2020). Calibration and validation framework for selective laser melting process based on multi-fidelity models and limited experiment data. Journal of Mechanical Design, 142(8), 1–13.

Palar, P. S., Tsuchiya, T., and Parks, G. (2015). Decomposition-based evolutionary aerodynamic robust optimization with multi-fidelity point collocation non-intrusive polynomial chaos. 17th AIAA Non-Deterministic Approaches Conference, Kissimmee, Florida, USA.

Palar, P. S., Tsuchiya, T., and Parks, G. T. (2016). Multi-fidelity non-intrusive polynomial chaos

based on regression. Computer Methods in Applied Mechanics and Engineering, 305, 579–606.

Palar, P. S., Zuhal, L. R., Shimoyama, K., and Tsuchiya, T. (2018). Global sensitivity analysis via multi-fidelity polynomial chaos expansion. Reliability Engineering & System Safety, 170(September 2016), 175–190.

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

Papadimitriou, C., Beck, J. L., and Au, S.-K. (2000). Entropy-based optimal sensor location for structural model updating. Journal of Vibration and Control, 6(5), 781–800.

Park, C., Haftka, R. T., and Kim, N. H. (2017). Remarks on multi-fidelity surrogates. Structural and Multidisciplinary Optimization, 55(3), 1029–1050.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., and Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. Structural Health Monitoring, 18(1), 103–122.

Pearl, J. (1984). Heuristics: intelligent search strategies for computer problem solving. Addison-Wesley Publishing Company, California, USA.

Peeters, B., Maeck, J., and Roeck, G. De. (2001). Vibration-based damage detection in civil

engineering: excitation sources and temperature effects. Smart Materials and Structures, 10(3), 518–527.

Peng, J., Hampton, J., and Doostan, A. (2014). A weighted l1-minimization approach for sparse polynomial chaos expansions. Journal of Computational Physics, 267, 92–111.

Pepper, N., Montomoli, F., and Sharma, S. (2021). Data fusion for uncertainty quantification with non-intrusive polynomial chaos. Computer Methods in Applied Mechanics and Engineering, 374, 113577.

Pérez, M. A., and Serra-López, R. (2019). A frequency domain-based correlation approach for structural assessment and damage identification. Mechanical Systems and Signal Processing, 119, 432–456.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2017). Machine learning of linear differential equations using Gaussian processes. Journal of Computational Physics, 348, 683–693.

Ramos, E. G., and Martínez, F. . (2013). A review of artificial neural networks: How well do they perform in forecasting time series? Analítika: Revista de Análisis Estadístico, 6, 7–18.

Rasmussen, C. E., and Williams, C. K. I. (2006). Gaussian processes for machine learning. the MIT Press, Cambridge, London, England.

Ren, L., Sun, Y., Cui, J., and Zhang, L. (2018). Bearing remaining useful life prediction based on

deep autoencoder and deep neural networks. Journal of Manufacturing Systems, 48(May), 71–77.

Rumpfkeil, M. P., and Beran, P. S. (2020). Multi-fidelity, gradient-enhanced, and locally optimized sparse polynomial chaos and Kriging surrogate models applied to benchmark problems. AIAA Scitech 2020 Forum, Orlando, FL, USA.

Rytter, A. (1993). Vibration based inspection of civil engineering structures, Doctoral dissertation, Aalborg University.

Schlune, H., Plos, M., and Gylltoft, K. (2009). Improved bridge evaluation through finite element model updating using static and dynamic measurements. Engineering Structures, 31(7), 1477–1485.

Seeger, M. W., and Nickisch, H. (2008). Compressed sensing and Bayesian experimental design. Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland.

Sen, D., Erazo, K., Zhang, W., Nagarajaiah, S., and Sun, L. (2019a). On the effectiveness of principal component analysis for decoupling structural damage and environmental effects in bridge structures. Journal of Sound and Vibration, 457, 280–298.

Sen, D., Erazo, K., Zhang, W., Nagarajaiah, S., and Sun, L. (2019b). On the effectiveness of principal component analysis for decoupling structural damage and environmental effects in bridge structures. Journal of Sound and Vibration, 457, 280–298.

Sepahvand, K., Marburg, S., and Hardtke, H.-J. (2010). Uncertainty quantification in stochastic systems using polynomial chaos expansion. International Journal of Applied Mechanics, 02(02), 305–353.

Sheikholeslami, R., and Razavi, S. (2017). Progressive Latin hypercube sampling: An efficient approach for robust sampling-based analysis of environmental models. Environmental Modelling & Software, 93, 109–126.

Shin, Y., and Xiu, D. (2016). Nonadaptive quasi-optimal points selection for least squares linear regression. SIAM Journal on Scientific Computing, 38(1), A385–A411.

Shokrani, Y., Dertimanis, V. K., Chatzi, E. N., and N. Savoia, M. (2018). On the use of mode shape curvatures for damage localization under varying environmental conditions. Structural Control and Health Monitoring, 25(4), e2132.

Sobol′, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1–3), 271–280.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. Technometrics, 29(2), 143.

Stutz, L. T., Rangel, I. C. S. S., Rangel, L. S., Corrêa, R. A. P., and Knupp, D. C. (2018). Structural damage identification built on a response surface model and the flexibility matrix. Journal of Sound and Vibration, 434, 284–297.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety, 93(7), 964–979.

Sun, L. M., Zhang, W., and Nagarajaiah, S. (2019). Bridge real-time damage identification method using inclination and strain measurements in the presence of temperature variation. Journal of Bridge Engineering, 24(2), 04018111.

Sun, L., Zhou, Y., and Min, Z. (2018). Experimental study on the effect of temperature on modal frequencies of bridges. International Journal of Structural Stability and Dynamics, 18(12), 1850155.

Sun, X., Choi, Y. Y., and Choi, J.-I. (2020). Global sensitivity analysis for multivariate outputs using polynomial chaos-based surrogate models. Applied Mathematical Modelling, 82, 867–887.

Surace, C., and Bovsunovsky, A. (2020). The use of frequency ratios to diagnose structural damage in varying environmental conditions. Mechanical Systems and Signal Processing, 136, 106523.

Tan, M. H. Y. (2015). Sequential Bayesian polynomial chaos model selection for estimation of sensitivity indices. SIAM/ASA Journal on Uncertainty Quantification, 3(1), 146–168.

Tang, Z., Chen, Z., Bao, Y., and Li, H. (2019). Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring. Structural Control and Health Monitoring, 26(1), e2296.

Thapa, M., Mulani, S. B., and Walters, R. W. (2018a). A new non-intrusive polynomial chaos using higher order sensitivities. Computer Methods in Applied Mechanics and Engineering, 328, 594–611.

Thapa, M., Mulani, S. B., and Walters, R. W. (2018b). Variance based adaptive-sparse polynomial chaos with adaptive sampling. 2018 AIAA Non-Deterministic Approaches Conference, Kissimmee, Florida, USA.

Thapa, M., Mulani, S. B., and Walters, R. W. (2020). Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling. Computer Methods in Applied Mechanics and Engineering, 360, 112759.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. The Journal of Machine Learning Research, 1, 211–244.

Torre, E., Marelli, S., Embrechts, P., and Sudret, B. (2019). A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas. Probabilistic Engineering Mechanics, 55(2018), 1–16.

Truong, T. T., Dinh-Cong, D., Lee, J., and Nguyen-Thoi, T. (2020). An effective deep feedforward neural networks (DFNN) method for damage identification of truss structures using noisy incomplete modal data. Journal of Building Engineering, 30(January), 101244.

Umar, S., Bakhary, N., and Abidin, A. R. Z. (2018). Response surface methodology for damage

detection using frequency and mode shape. Measurement, 115(November 2015), 258–268.

Vishwanathan, A., and Vio, G. A. (2019). Efficient quantification of material uncertainties in reliability-based topology optimization using random matrices. Computer Methods in Applied Mechanics and Engineering, 351, 548–570.

Vitali, R., Haftka, R. T., and Sankar, B. V. (2002). Multi-fidelity design of stiffened composite panel with a crack. Structural and Multidisciplinary Optimization, 23(5), 347–356.

Wan, H. P., Ren, W. X., and Todd, M. D. (2020). Arbitrary polynomial chaos expansion method for uncertainty quantification and global sensitivity analysis in structural dynamics. Mechanical Systems and Signal Processing, 142, 106732.

Wan, X., and Karniadakis, G. E. (2006). Beyond Wiener–Askey expansions: Handling arbitrary PDFs. Journal of Scientific Computing, 27(1–3), 455–464.

Wang, D., Tan, B., Wang, X., and Zhang, Z. (2021). Experimental study and numerical simulation of temperature gradient effect for steel-concrete composite bridge deck. Measurement and Control, 54(5–6), 681–691.

Wang, F., Xiong, F., Chen, S., and Song, J. (2019). Multi-fidelity uncertainty propagation using polynomial chaos and Gaussian process modeling. Structural and Multidisciplinary Optimization, 60(4), 1583–1604.

Wang, J. Y., and Ni, Y. Q. (2015). Refinement of damage identification capability of neural network techniques in application to a suspension bridge. Structural Monitoring and Maintenance, 2(1), 77–93.

Wang, L., Zhou, J., and Lu, Z. R. (2020a). A fast friction-model-inspired sparse regularization approach for damage identification with modal data. Computers & Structures, 227(November), 106142.

Wang, X., Hou, R., Xia, Y., and Zhou, X. (2020b). Laplace approximation in sparse Bayesian learning for structural damage detection. Mechanical Systems and Signal Processing, 140, 106701.

Wang, Z., Huang, M., and Gu, J. (2020c). Temperature effects on vibration-based damage detection of a reinforced concrete slab. Applied Sciences, 10(8), 2869.

Wiener, N. (1938). The Homogeneous chaos. American Journal of Mathematics, 60(4), 897.

Worden, K., and Lane, A. J. (2001). Damage identification using support vector machines. Smart Materials and Structures, 10(3), 540–547.

Xia, Q., Cheng, Y., Zhang, J., and Zhu, F. (2017). In-service condition assessment of a long-span suspension bridge using temperature-induced strain data. Journal of Bridge Engineering, 22(3), 04016124.

Xia, Q., Zhou, L., and Zhang, J. (2018). Thermal performance analysis of a long-span suspension bridge with long-term monitoring data. Journal of Civil Structural Health Monitoring, 8(4), 543–553.

Xia, Y., Chen, B., Weng, S., Ni, Y. Q., and Xu, Y. L. (2012). Temperature effect on vibration properties of civil structures: A literature review and case studies. Journal of Civil Structural Health Monitoring, 2(1), 29–46.

Xia, Y., Chen, B., Zhou, X., and Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. Structural Control and Health Monitoring, 20(4), 560–575.

Xia, Y., Xu, Y. L., Wei, Z. L., Zhu, H. P., and Zhou, X. Q. (2011). Variation of structural vibration characteristics versus non-uniform temperature distribution. Engineering Structures, 33(1), 146–153.

Xiu, D., and Karniadakis, G. E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM Journal on Scientific Computing, 24(2), 619–644.

Xu, J., and Kong, F. (2018). A cubature collocation based sparse polynomial chaos expansion for efficient structural reliability analysis. Structural Safety, 74(November 2017), 24–31.

Xu, X., Huang, Q., Ren, Y., Zhao, D. Y., Yang, J., and Zhang, D. Y. (2019). Modeling and separation of thermal effects from cable-stayed bridge response. Journal of Bridge Engineering, 24(5),

04019028.

Xu, Y., Choi, J., and Oh, S. (2011). Mobile sensor network navigation using Gaussian processes with truncated observations. IEEE Transactions on Robotics, 27(6), 1118–1131.

Xu, Y. L., Zhang, C. D., Zhan, S., and Spencer, B. F. (2018). Multi-level damage identification of a bridge structure: a combined numerical and experimental investigation. Engineering Structures, 156(November 2017), 53–67.

Yan, A. M., De Boe, P., and Golinval, J. C. (2004). Structural damage diagnosis by Kalman model based on stochastic subspace identification. Structural Health Monitoring, 3(2), 103–119.

Yang, Q., Meng, S., Jin, H., Xie, W., and Zhang, X. (2019). Multi-fidelity uncertainty quantification method with application to nonlinear structural response analysis. Applied Mathematical Modelling, 75, 853–864.

Yang, X. S. (2010). Nature-Inspired Metaheuristic Algorithms Second Edition. Luniver Press, Frome, UK.

Yang, X. S., and Suash Deb. (2009). Cuckoo Search via Levy flights. 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 210–214.

Yarnold, M. T., and Moon, F. L. (2015). Temperature-based structural health monitoring baseline for long-span bridges. Engineering Structures, 86, 157–167.

Zein, S., Colson, B., and Glineur, F. (2013). An efficient sampling method for regression-based polynomial chaos expansion. Communications in Computational Physics, 13(4), 1173–1188.

Zhang, C. D., and Xu, Y. L. (2016). Comparative studies on damage identification with Tikhonov regularization and sparse regularization. Structural Control and Health Monitoring, 23(3), 560–579.

Zhang, G. P., and Berardi, V. L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction. Journal of the Operational Research Society, 52(6), 652–664.

Zhang, J., and Du, X. (2010). A second-order reliability method with first-order efficiency. Journal of Mechanical Design, 132(10), 973–984.

Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D. (2015). A survey of sparse representation: Algorithms and applications. IEEE Access, 3, 490–530.

Zhao, H., Gao, Z., Xu, F., Zhang, Y., and Huang, J. (2019). An efficient adaptive forward–backward selection method for sparse polynomial chaos expansion. Computer Methods in Applied Mechanics and Engineering, 355, 456–491.

Zhou, C., Lu, Z., and Yuan, X. (2013). Use of relevance vector machine in structural reliability analysis. Journal of Aircraft, 50(6), 1726–1733.

Zhou, H. F., Ni, Y. Q., and Ko, J. M. (2011). Eliminating temperature effect in vibration-based structural damage detection. Journal of Engineering Mechanics, 137(12), 785–796.

Zhou, K., and Tang, J. (2021). Uncertainty quantification of mode shape variation utilizing multi-level multi-response Gaussian process. Journal of Vibration and Acoustics, 143(1).

Zhou, S., and Song, W. (2018). Environmental-effects-embedded model updating method considering environmental impacts. Structural Control and Health Monitoring, 25(3), 1–22.

Zhou, X. Q., Xia, Y., and Weng, S. (2015). L1 regularization approach to structural damage detection using frequency data. Structural Health Monitoring, 14(6), 571–582.

Zhou, Y., Lu, Z., Cheng, K., and Ling, C. (2019a). An efficient and robust adaptive sampling method for polynomial chaos expansion in sparse Bayesian learning framework. Computer Methods in Applied Mechanics and Engineering, 352, 654–674.

Zhou, Y., Lu, Z., Cheng, K., and Shi, Y. (2019b). An expanded sparse Bayesian learning method for polynomial chaos expansion. Mechanical Systems and Signal Processing, 128, 153–171.

Zhou, Y., Lu, Z., and Yun, W. (2020). Active sparse polynomial chaos expansion for system reliability analysis. Reliability Engineering & System Safety, 202(May), 107025.

Zhu, Y., Ni, Y., Jesus, A., Liu, J., and Laory, I. (2018). Thermal strain extraction methodologies for bridge structural condition assessment. Smart Materials and Structures, 27(10), 105051.

Zhu, Y., Ni, Y., Jin, H., Inaudi, D., and Laory, I. (2019). A temperature-driven MPCA method for structural anomaly detection. Engineering Structures, 190(September 2018), 447–458.