

SHM THROUGH FLEXIBLE VIBRATION SENSING TECHNOLOGIES AND ROBUST SAFETY EVALUATION PARADIGM

Theanh Nguyen
M.E., B.E.

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

School of Civil Engineering and Built Environment
Science and Engineering Faculty
Queensland University of Technology

2014

Keywords

Structural Health Monitoring, Large-scale Civil Structures, SHM-oriented Wireless Sensor Network, Semi-complete Data Synchronization, Data Synchronization Error, Ethernet distributed Data Acquisition, Data-based Safety Evaluation, Output-only Modal Analysis, Level-1 Damage Identification, Statistical Unsupervised Learning, Mahalanobis Squared Distance, Multivariate Normal Distribution, Controlled Monte Carlo Data Generation, Synthetic Health Monitoring System

Abstract

Long-term and frequent vibration monitoring has been seen one of the most suitable approaches to achieve a timely decision-making process concerning the health status of civil infrastructure including the emergency cases such as possible structural failures. The fact that the tragic collapse of the I-35W Bridge in United States in 2007 occurred only three months after its most recent inspection has strongly supported this argument. However, implementing such a system in practice is not always feasible due to many obstacles such as the large scale of civil structures, tight budget, uncertainties of new sensing technologies and ineffective applications. To tackle these problems altogether, this research program aims to develop a practical and reliable synthetic Structural Health Monitoring (SHM) system with two core subsystems namely vibration sensing system and data-based safety evaluation system for use in large-scale civil infrastructure. Two popular vibration sensing platforms considered for the first subsystem are SHM-oriented Wireless Sensor Network (SHM-oriented WSN) and Ethernet distributed Data Acquisition (Ethernet distributed DAQ) system as they can provide flexibility in selection for different (e.g. continuous/non-continuous) measurement purposes. To enable a frequent safety evaluation basis, the feature extraction function is built upon the convenient Output-only Modal Analysis (OMA) techniques particularly those of the powerful data-driven Stochastic Subspace Identification (SSI-data) family whilst the damage identification function is constructed from the computationally efficient Mahalanobis Squared Distance (MSD) based unsupervised learning algorithm. A number of evaluations, improvements and new developments are then made in all the components in both subsystems to select the most suitable candidates as well as to enhance their robustness and practicalities. First, a flexible semi-complete data synchronization scheme is explored for SHM-oriented WSNs as a relaxing synchronization solution for more efficient usage in large-scale civil structures. Realizing its desired cost-effective and flexible features, an extension of this scheme is then made for use with the TCP/IP communication medium before being used, along with other novel sensor and peripheral DAQ hardware solutions, to realize an actual Ethernet distributed DAQ system. On the side of safety evaluation components, intensive investigations are first carried out to assess the possible

impact of remaining synchronization uncertainties as well as the robustness of the SSI-data techniques and their auxiliary processing tools. To overcome the inherent impact of variable Environmental and Operational (E&O) factors when evaluating random Data Synchronization Error (DSE) of the Ethernet distributed DAQ system, a novel daisy chain data selection scheme is derived in order to effectively assist statistical assessment tasks. For damage identification function, what is uncovered is that the weakness of the MSD-based method is associated to the difficulty in satisfying this method requirement of data distribution at an early monitoring stage or during short-lived SHM programs. A so-called Controlled Monte Carlo Data Generation (CMCDG) scheme is then proposed as a prescription for such an illness. The results from intensive evaluations in this research program show that the semi-complete data synchronization approach is feasible not only for SHM-oriented WSNs but also for the Ethernet distributed DAQ system platform. In fact, the efficacy of the data synchronization scheme derived for the latter platform is experimentally verified thereby not only successfully establishing a real-world SHM testbed but also providing a promising alternative for use in the other SHM projects with tight budget and/or sparse measurement coverage where conventional cable-based synchronization solution may be too costly. Besides, the evaluation results also confirm the robustness of the primary SSI-data technique and the post-processing data merging method in coping with DSE; and highlight the efficacy of the channel projection option in assisting the SSI-data techniques in such adverse circumstances. Finally, it is shown that the ultimate assistance from the proposed CMCDG scheme is well able to assist the MSD-based damage identification method to overcome the problems of experimental data shortage and distributional insufficiency, attain sufficient computational stability and achieve satisfactory damage identification outcome. Applications onto both laboratory and full-scale vibration data highlight the advantages of CMCDG not only in helping to provide optimal synthetic data for adequate MSD-based learning processes but also at the dynamic structure of this scheme making it well adaptive to any input data with any primary distributional condition. With these validated developments and enhancements, it is evident that the developed synthetic SHM system will function reliably from very early operational stages thereby enabling prompt protection for both valuable civil infrastructure and the user community involved.

Table of Contents

Keywords	i
Abstract	iii
Table of Contents	v
List of Figures	ix
List of Tables.....	xi
List of Acronyms and Abbreviations.....	xiii
Statement of Original Authorship.....	xv
Acknowledgments	xvii
CHAPTER 1: INTRODUCTION.....	1
1.1 Overview and Motivations	1
1.2 Aim and Objectives.....	4
1.3 Scope and Research Significance.....	5
1.4 Account of Scientific Progress Linking the Research Papers.....	7
1.5 Research Innovation.....	12
1.6 List of Research Papers Included in Thesis.....	13
CHAPTER 2: LITERATURE REVIEW.....	15
2.1 General Concepts	15
2.1.1 SHM	15
2.1.2 Vibration-based SHM.....	16
2.1.3 Hierarchy and Approaches of Damage Identification.....	17
2.1.4 Coping with Impact of Variable E&O Factors.....	18
2.2 Vibration Sensing Technologies	20
2.2.1 WSN	20
2.2.2 Wired Sensing System.....	24
2.3 Data-based Safety Evaluation	29
2.3.1 Features and Feature Extraction Techniques	29
2.3.2 Level-1 Damage Identification	32
2.4 Concluding Remarks.....	34
CHAPTER 3: EFFECTS OF WSN UNCERTAINTIES ON OUTPUT-ONLY MODAL-BASED DAMAGE IDENTIFICATION.....	39
Statement of Joint Authorship.....	41
Abstract	42
3.1 Introduction.....	43
3.2 Major Uncertainties of Generic and SHM-oriented WSNs.....	44
3.3 OMA and level 1 OMDI approach under investigation	47
3.4 Research Methodology.....	50
3.5 Description of Datasets and Analysis.....	51
3.5.1 The Benchmark Structure and Datasets	51
3.5.2 Simulation of DSE and Analyses of DSE Impact	52

3.6	Result and Discussions.....	53
3.6.1	Common Results of FDD and SSI-data	53
3.6.2	Effect of DSE on Level 1 of OMDI.....	54
3.7	Conclusions.....	58
CHAPTER 4: EFFECTS OF WSN UNCERTAINTIES ON OUTPUT-ONLY MODAL ANALYSIS EMPLOYING MERGED DATA OF MULTIPLE TESTS.....		61
	Statement of Joint Authorship.....	63
	Abstract	64
4.1	Introduction.....	65
4.2	Major Uncertainties of SHM-oriented WSNs.....	66
4.3	OMA and Data Merging Methods	68
4.4	Research Methodology	71
4.5	Brief Description of Tests and Analysis	72
4.5.1	The Bridge Model and Wired Sensing System.....	72
4.5.2	Simulation of Noise and Initial DSE	74
4.5.3	OMA and Analyses of Effects of DSE	75
4.6	Results and Discussions.....	76
4.6.1	Common Results of OMA for DSE-free Data	76
4.6.2	The Use of Channel Projection.....	77
4.6.3	Effects of DSE on Outcomes of Three OMA Techniques	78
4.7	Conclusions.....	80
CHAPTER 5: DEVELOPMENT OF A COST-EFFECTIVE AND FLEXIBLE SENSING SYSTEM FOR LONG-TERM CONTINUOUS VIBRATION MONITORING.....		83
	Statement of Joint Authorship.....	85
	Abstract	86
5.1	Introduction.....	87
5.2	Overview of The Instrumented Complex and Monitoring Systems	89
5.3	Vibration Sensor and Data Synchronization Solutions	92
5.3.1	Vibration Sensor Solution.....	92
5.3.2	Data Synchronization Solution	94
5.4	Experimental Evaluations	96
5.4.1	General Evaluation of The Sensing System.....	96
5.4.2	Statistical Evaluation of The Data Synchronization Solution	101
5.4.3	Continuous Structural Safety Evaluation	102
5.5	Conclusions.....	104
	Appendix 5.A MSD-based Learning for Level-1 Damage Identification.....	106
	Appendix 5.B AANN-based Learning for Level-1 Damage Identification	106
CHAPTER 6: CONTROLLED MONTE CARLO DATA GENERATION FOR STATISTICAL DAMAGE IDENTIFICATION EMPLOYING MSD.....		109
	Statement of Joint Authorship.....	110
	Abstract	111
6.1	Introduction.....	111
6.2	Damage Identification and Data Generation Methods	113
6.2.1	MSD-based Damage Identification.....	113
6.2.2	CMCDG	114
6.3	Description of the benchmark structures and data	116
6.4	Analyses and Discussion.....	118
6.4.1	MSD-based Damage Identification Performance on Pure Experimental Data	120

6.4.2	Performance of Two Condition Assessment Methods on Pure Experimental Data	121
6.4.3	Performance of CMCDG on Premature Data	123
6.5	Conclusions	129
CHAPTER 7: FIELD VALIDATION OF CONTROLLED MONTE CARLO DATA		
GENERATION SCHEME FOR STATISTICAL DAMAGE IDENTIFICATION EMPLOYING		
MSD		
	Statement of Joint Authorship	132
	Abstract	133
7.1	Introduction	133
7.2	Theory	136
7.2.1	MSD-based Damage Identification	136
7.2.2	CMCDG	137
7.3	Description of benchmark Structures and Their Data Status	138
7.3.1	SMC Benchmark Structure and Data Status	138
7.3.2	QUT-SHM Benchmark Structure and Data Status	139
7.4	Analyses and Discussion	140
7.4.1	SMC Vibration Data	141
7.4.2	QUT-SHM Vibration Data	151
7.5	Conclusions	156
CHAPTER 8: CONCLUSIONS AND FUTURE WORK		
159		
8.1	Summary and Conclusions	159
8.2	Future Work	163
BIBLIOGRAPHY		
167		
APPENDICES		
179		
	Appendix A: Flowchart of MSD-based Damage Identification Employing CMCDG	179
	Appendix B: Fundamentals of Monte Carlo Simulation and Monte Carlo Data	
	Generation	180

List of Figures

Figure 1-1 Main connections amongst the papers included in the thesis	10
Figure 2-1 Hierarchy of damage identification process [adapted from (Farrar and Worden, 2013)]	17
Figure 2-2 Temperature impact on the first-mode shape of one bridge span during two distinct times of the day: (a) in the morning (7.75 Hz); (b) in the afternoon (7.42 Hz) (Figueiredo <i>et al.</i> , 2011).....	19
Figure 2-3 Typical diagram of a star WSN topology (Libelium, 2010)	21
Figure 2-4 Typical diagram of a peer-to-peer WSN topology (Libelium, 2010).....	21
Figure 2-5 A SHM-oriented wireless sensor node based on Imote2 control and communication platform [adapted from (Rice and Spencer, 2009)]	23
Figure 2-6 Tsing Ma bridge with 1377 m main span (Chan <i>et al.</i> , 2011).....	26
Figure 2-7 Schematic diagram of a typical Ethernet distributed DAQ system used in long-span bridges [adapted from (Xu and Xia, 2012)]	27
Figure 2-8 Example of CRIO peripheral DAQ model (National Instruments, 2012a)	28
Figure 3-1 Position of the accelerometers and the wired DAQ system on GNTVT	51
Figure 3-2 Singular value plot for FDD (left) and stabilization diagram for SSI-data (right) from DSE-free data	53
Figure 3-3 Four typical modes in the short axis estimated by FDD and SSI-data from DSE-free data	54
Figure 3-4 Box-plots of MAC deviation under DSE of generic WSNs.....	56
Figure 3-5 Box-plots of MAC deviation under DSE of SHM-oriented WSNs	56
Figure 3-6 Box-plots of MAC deviation under DSE of SHM-oriented WSNs (zoomed scale)	57
Figure 4-1 Flowchart of the investigation approach	72
Figure 4-2 Physical bridge model and its wired sensing system.....	73
Figure 4-3 Two examples of the sensor setups	74
Figure 4-4 Typical mode shapes estimated from DSE-free data	76
Figure 4-5 Stabilization diagram of UPC-PreGER with projection: (a) disabled and (b) enabled	77
Figure 4-6 Stabilization diagram of UPC-PoSER with projection: (a) disabled & (b) enabled	77
Figure 4-7 Box-plots of MAC deviations (from unity) of mode shapes for four cases	79
Figure 5-1 Science and Engineering Centre (SEC) at QUT Gardens Point campus.....	90
Figure 5-2 General diagram of the SEC monitoring systems	91
Figure 5-3 Sensor position on the footbridge (left) and the building (right)	93
Figure 5-4 Typical low-level acceleration time history	97
Figure 5-5 A typical SSI-data stabilization diagram for OMA of the building.....	100
Figure 5-6 Building model and typical animation views of seven estimated modes	100
Figure 5-7 Statistics of mode shape agreement across multiple datasets.....	102
Figure 5-8 Level-1 damage identification results: MSD-based method (left) and AANN-based method (right).....	103
Figure 5-9 Box-plots of training data (left) and testing data (right)	104

Figure 6-1 The test structure (left) and damage simulation mechanism (right) at LANL (Figueiredo <i>et al.</i> , 2009).....	117
Figure 6-2 RMSE of AR models of increasing order for each undamaged state.....	120
Figure 6-3 Type I error of increasing observation size	121
Figure 6-4 COND in linear (left) and log (right) scales	122
Figure 6-5 Beta Q-Q plot of (a) AR10-05 tests, (b) AR10-13 tests, (c) AR30-08 tests and (d) AR30-16 tests.....	123
Figure 6-6 COND and mean error rate of increasing noise level: (a and c) AR10 and (b and d) AR30	125
Figure 6-7 COND, Q-Q RMSE and Type I error rate of increasing replication size: (a, c, and e) AR10 and (b, d, and f) AR30	127
Figure 6-8 Overlay of one typical seed observation and its 15 variants: AR10 (left) and AR30 (right)	128
Figure 6-9 Post-CMCDG beta Q-Q plots: AR10 (left) and AR30 (right)	128
Figure 7-1 SMC benchmark structure (Li <i>et al.</i> , 2014).....	138
Figure 7-2 QUT-SHM benchmark structure	139
Figure 7-3 SMC bridge model in ARTeMIS Extractor software	141
Figure 7-4 Detected modes of SMC benchmark structure: (a) State 1 and (b) State 2.....	143
Figure 7-5 Representative correlation between (SMC) mode shapes of two states.....	144
Figure 7-6 Characteristics of SMC data and original testing results: (a, b) datasets 1 and 2 (State 1); (c) dataset 3 (State 2); (d) beta Q-Q plot of dataset 1; and (e) original testing results.....	146
Figure 7-7 Results of two simulation rounds in CMCDG for SMC data: (a, b) round 1 with COND of 9 and 18 replication blocks; (c, d) round 2 with COND and beta Q-Q RMSE at noise level of 2%	148
Figure 7-8 (a) beta Q-Q plot and (b) testing results of a typical enhanced (SMC) learning dataset	149
Figure 7-9 Impact of input perturbation on (SMC) MSD computation: (a) original learning dataset and (b) typical enhanced learning dataset.....	151
Figure 7-10 Detected modes of QUT-SHM benchmark structure.....	152
Figure 7-11 Mode shapes of six usable modes of QUT-SHM benchmark structure	152
Figure 7-12 Characteristics of QUT-SHM data and original testing results: (a) original learning dataset; (b) testing dataset; (c) beta Q-Q plot of original learning dataset; and (d) original testing results	154
Figure 7-13 Results of two simulation rounds in CMCDG for QUT-SHM data: (a, b) round 1 with COND for 7 and 14 replication block cases; (c, d) round 2 with COND and beta Q-Q RMSE at noise level of 0.6%	154
Figure 7-14 (a) beta Q-Q plot and (b) testing results of a typical enhanced (QUT-SHM) learning dataset.....	156
Figure 7-15 Impact of input perturbation on (QUT-SHM) MSD computation: (a) original learning dataset and (b) typical enhanced learning dataset.....	156

List of Tables

Table 3-1 Effects of DSE of 15dt on frequency estimates by SSI-data	55
Table 4-1 Effects of DSE on frequency estimates by UPC-PoSER and UPC-PreGER.....	78
Table 5-1 Features of seven estimated modes	100
Table 6-1 Data labels of the structural state conditions [adapted from (Figueiredo <i>et al.</i> , 2009)]	118
Table 7-1 Selected testing days and usable SMC datasets	142

List of Acronyms and Abbreviations

AANN	Auto-Associative Neural Network
AE	Acoustic Emission
ANSHM	Australian Network of Structural Health Monitoring
AR	Auto-Regressive
ARMA	Auto-Regressive with Moving Average
ARX	Auto-Regressive with Exogenous Inputs
ARTeMIS	Ambient Response Testing and Modal Identification Software
cDAQ	compact Data Acquisition
CLT	Central Limit Theorem
CMCDG	Controlled Monte Carlo Data Generation
COMAC	Coordinate Modal Assurance Criterion
COND	Condition number
CRIO	Compact Reconfigurable Input/Output
DAQ	Data Acquisition
DSE	Data Synchronization Error
E&O	Environmental and Operational
FDD	Frequency Domain Decomposition
FE	Finite Element
FPGA	Field-Programmable Gate Arrays
GNTVT	Guangzhou New TV Tower
GPS	Global Positioning System
HTC	Heritage Court Tower
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
IQR	Inter-Quartile Range
LANL	Los Alamos National Laboratory
LLN	Law of Large Numbers
MAC	Modal Assurance Criterion
MACEC	Modal Analysis on Civil Engineering Constructions
MSD	Mahalanobis Squared Distance

multinormal	multivariate normal
multinormality	multivariate normality
NI	National Instruments
OMA	Output-only Modal Analysis
OMDI	Output-only Modal-based Damage Identification
PoSER	Post Separate Estimation Re-scaling
PreGER	Pre Global Estimation Re-scaling
Q-Q	Quantile-Quantile
QUT	Queensland University of Technology
RMS	Root Mean Square
RMSE	Root Mean Square Error
SHM	Structural Health Monitoring
SMC	Structural Monitoring and Control
SSI-cov	covariance-driven Stochastic Subspace Identification
SSI-data	data-driven Stochastic Subspace Identification
SVD	Singular Value Decomposition
TCP/IP	Transmission Control Protocol/Internet Protocol
TSE	Time Synchronization Error
UPC	Unweighted Principal Component
URL	Uniform Resource Locator
WSN	Wireless Sensor Network

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Author's Signature: [QUT Verified Signature](#)

Date: 12/11/2014

Acknowledgments

Pursuing a PhD overseas has been a long and challenging journey for me. Getting to this point, I must say that I have been so fortunate to be supported by a caring family; a professional supervisory team; and lovely friends and colleagues.

I thank my parents for giving me the birth, for their love and encouragement while I studied abroad. The love and support from my wife (Haiyen) and my children (Ruby & Beam) have always been non-stop though during the time of this study they had to spend many weekends and holidays on their own.

I wish to pass my sincere appreciation to my two supervisors:

Prof. Tommy Chan, my principal supervisor. During my candidature, I have indefinitely benefited from his enthusiasm, high sense of responsibility, and invaluable advice to my research. His ongoing support has kept me passionate on tough research topics. I also thank him for being always available for me to keep in touch and for timely editing for my papers no matter where he is (in Australia or overseas). Considering him a role model for my future career, I have greatly learnt from his expertise in teaching; and research cooperation and coordination.

Prof. David Thambiratnam, my associate supervisor. Throughout my candidature, his care, enlightenment, and incentive on my study and life were always there; without these, I could not achieve much from the research project. In addition, I appreciate his careful editing for my papers making them more readable as well as his willingness to share academic experience with me. I also love listening to his stories about his experience in Vietnam.

I also greatly acknowledge Les King for his professional working attitude and valuable cooperation in the sensing system development work presented in Chapter 5 of this thesis. Also in relation to this work, Mark Barry, Lance Wilson and other technicians in the HPC and Research Support Group must be thanked for their timely assistance for data storage and access particularly for the initial phase of this

instrumentation project. I also thank Mr. Jonathan James for his assistance for the project.

My thanks also spread to Frank Wang, Buddhi Wahalathantri, Craig Cowled, and other fellows in the QUT-SHM research team for their help, debates and feedbacks on my work. It was my honour to be a member of such a friendly, knowledgeable, and dynamic group.

My acknowledgements also go to Dr. Charles Farrar (LANL Engineering Institute, USA), Prof. Hui Li and Dr. Shunlong Li (Harbin Institute of Technology, China), Prof. Yi-Qing Ni and Dr. Yong Xia (Hong Kong Polytechnic University, China) for generous sharing of benchmark data and other research resources; Prof. Keith Worden (University of Sheffield, UK) for kind instructions on machine learning issues; Prof. Peter Avitabile (University of Massachusetts Lowell, USA) for his help with some modal analysis problems; and Dr. Kirill Mechitov (University of Illinois at Urbana-Champaign, USA) and Dr. Tomonori Nagayama (University of Tokyo, Japan) for useful discussions on wireless sensor topics. I thank Dr. Ying Wang, Dr. Xinqun Zhu, Dr. Ching-Tai Ng and the other ANSHM executive members; and Prof. Douglas Adams (Journal of SHM) for their assistance for my publications. I also thank Dr. Tuan Ngo, A/Prof. Nicholas Haritos and Prof. Priyan Mendis (University of Melbourne) for their help for my first days with SHM topics.

Last but not least, I wish to thank the two external panel members of my final seminar (Prof. Mahen Mahendran and A/Prof. Cheng Yan) as well as the anonymous thesis examiners for their constructive comments and suggestions in making this thesis better. I would also like to express my appreciation to Vietnam Government, QUT, QUT School of Civil Engineering and Built Environment for the scholarships that supported me in the past years. I sincerely thank QUT Research Student Centre and QUT Science and Engineering Faculty HDR Student Support Team for their direct and indirect support during my PhD candidature.

(Andy) Theanh Nguyen

Brisbane, November 2014

Chapter 1: Introduction

1.1 OVERVIEW AND MOTIVATIONS

Structural Health Monitoring (SHM) is emerging as a very active research area beginning from the fact that civil infrastructure around the world is experiencing degradation due to their aging and improper usage. Further, introduction of innovative structures with bold concepts, use of new materials, excessively increased maintenance costs and the need for effective post disaster condition evaluations have also caused infrastructure to become more reliant on SHM technologies (Ansari, 2005). In short, SHM can be defined as the use of on-structure, non-destructive sensing systems to collect data in order to evaluate the health state of a structure and monitor its performance for decision making and advancing the current practice of structural design, maintenance and rehabilitation (Chan *et al.*, 2011). Therefore, a practical SHM system should essentially consist of two main subsystems namely (i) sensing system, and (ii) structural (performance and/or safety) evaluation system. Over the past two decades, there have been numerous advances in development in these two subsystems in order to make them more applicable to actual civil structures with high degrees of scale and complexity. Various types of sensors such as accelerometers, optical fibers, and global positioning systems have been developed and successfully integrated into sophisticated wired sensing systems with hundreds to thousands channels for the purpose of long-term and frequent monitoring of long-span bridges and high-rise building structures (Karbhari and Ansari, 2009; Ni *et al.*, 2009). Ethernet has been employed to replace conventional communication buses (such as RS-232 or GPIB) to spread out measurement to remote locations and form distributed Data Acquisition (DAQ) systems (Eren, 2011). As an emerging opponent of the Ethernet distributed DAQ platform, wireless sensor networks (WSNs) have recently drawn significant attention from the SHM community. Significantly enhanced from generic versions, SHM-oriented WSNs have been transitioned from structural laboratories to full-scale structures with the expectation that they would be an inexpensive and a more flexible substitute to the wired sensing system. On the

side of structural evaluation aspect, the most studied problems are unsurprisingly concerning structural damage identification (Doebbling *et al.*, 1996; Sohn *et al.*, 2003; Farrar and Worden, 2013). Numerous types of damage-sensitive (or health-representative) features and the corresponding techniques to extract them from vibration data have been investigated. Due to their applicability to in-service civil structures, there has recently been an increasing trend of using dynamic properties obtained by means of output-only modal analysis (OMA) techniques (Brincker *et al.*, 2001; Brincker *et al.*, 2003; Karbhari and Ansari, 2009). Popular OMA techniques such as those of Frequency Domain Decomposition (FDD) and data-driven Stochastic Subspace Identification (SSI-data) families have seen their applications in many SHM projects and have been incorporated in several commercial software packages such as ARTeMIS (www.svibs.com) and MACEC (<http://bwk.kuleuven.be/bwm/macec>). To cope with inherent impact of changing Environmental and Operational (E&O) conditions on the feature, a number of studies have been conducted towards the use of machine learning algorithms and their pattern recognition aspects in assisting the structural damage identification process (Sohn *et al.*, 2003; Gul and Catbas, 2009; Figueiredo *et al.*, 2011; Farrar and Worden, 2013). Through these studies, several promising learning algorithms have been shown and demonstrated by numerical and experimental data. As it is mainly based on the measured data, this evaluation approach is often known as the data-based approach as opposed to the model-based evaluation counterpart which is generally based on a physical model of the investigated structure such as the Finite Element (FE) model.

Even though prior studies have accomplished significant achievements, there have still been limited deployments of SHM systems onto real civil infrastructure due to a number of key limitations in both aforementioned subsystems that could hinder the widespread practice. With regards to structural safety evaluation, the first limitation is that most feature extraction and damage identification methods have been experimentally validated in well-controlled environment such as laboratories. The robustness of most feature extraction techniques has also not been thoroughly assessed against realistic measurement uncertainties particularly those recently emerged from the use of new sensing technologies or measurement strategies to suit

the needs in SHM of actual civil infrastructure. In a similar fashion, most damage identification methodologies proposed earlier have not thoroughly considered the impact of variable E&O factors and practical circumstances for long-term and frequent monitoring purposes. Although machine learning based damage identification has been shown to be a promising approach, most prior studies in this trend have often overlooked the prerequisite knowledge concerning validity of data condition and learning system architecture. If data condition is unstable or system architecture is invalid, it can induce ill-conditioned or invalid learning system realizations and subsequently erroneous results of structural evaluation. Such overlooking may be the cause for the false indications of structural states in prior studies (Worden *et al.*, 2003; Figueiredo *et al.*, 2011). In the sensing system aspect, most wired sensing systems for global vibration monitoring in general and those employing the Ethernet distributed DAQ architecture in particular though reliable and durable have still been too costly for most structures especially from the return on investment viewpoint (Ansari, 2005; Karbhari and Ansari, 2009). SHM-oriented WSNs though potentially being an inexpensive sensing substitute are still not efficient for use in full global vibration monitoring programs for actual civil structures due to inherent technical difficulties such as those related to data synchronization and system latency (Chintalapudi *et al.*, 2006; Karbhari and Ansari, 2009; Xu and Xia, 2012; Farrar and Worden, 2013). Such problems are partially due to the implementation of very precise synchronization solutions which are either expensive in the Ethernet distributed DAQ systems or computationally costly as well as causing additional system latency in the SHM-oriented WSNs. Besides, the large scale of critical civil structures also further complicates system deployment work and increases the cost. More cost-effective and flexible sensing solutions; as well as more robust structural evaluation paradigms are therefore in need to enhance the efficiency, flexibility and reliability of SHM systems. Naturally, these problems and the corresponding needs have become the main motivations for the research presented herein.

1.2 AIM AND OBJECTIVES

The overall aim of this thesis is to develop a practical synthetic SHM system that is cost-effective and flexible in sensing and DAQ; as well as robust in the structural safety evaluation aspect for the purpose of long-term and frequent monitoring of large-scale civil infrastructure during their service lives. The two types of sensing systems to be considered are the SHM-oriented WSNs and the Ethernet distributed DAQ systems due to their own merits and applicability. While the former sensing platform is more applicable for the case when continuous sensing is not required and/or a mobile motoring system is sought, the latter sensing platform is more suited to serve as a continuous sensing system that is often desirable for critical civil structures. Based upon the data-based approach, the structural safety evaluation system should be robust in both feature extraction and damage identification phases while usable from very early monitoring stages in order to safeguard the unfortunate short-lived structures.

This research can be achieved upon completing the following research tasks

- **Task 1**: Evaluate the feasibility of using the semi-complete data synchronization approach in global vibration sensing of large-scale civil structures. The main target in this task will be a relaxing data synchronization scheme that is computationally effective for SHM-oriented WSNs when being used in large-scale structures. Included in the research task will also be a thorough understanding of the impact of such a synchronization scheme as well as the (corresponding) robustness of the most popular global feature extraction techniques to facilitate reliable applications in subsequent stages of this research.
- **Task 2**: Develop a cost-effective and flexible wired sensing system employing the Ethernet distributed DAQ architecture and the semi-complete data synchronization approach to overcome the difficulties of having large or sparse measurement coverage while the budget is tight. Budget constraint is often encountered in practice and tends to hinder the use of high-end signal acquisition devices to cope with such coverage problems.

- **Task 3:** (a) Develop a robust data-based damage identification method for the purpose of frequent or continuous structural safety evaluation with time. The method should be well usable even with limited observations such as at an early monitoring stage; (b) Validate the developed damage identification method by well-established laboratory experimental datasets.
- **Task 4:** Validate the damage identification method by measured data collected by the wired sensing system (developed in Task 2) as well as by other real-world vibration measurement systems.

1.3 SCOPE AND RESEARCH SIGNIFICANCE

As discussed earlier, this research addresses two main (i.e. sensing and structural safety evaluation) problems associated with the development of a practical SHM system for civil infrastructure. In the aspect of sensing issues, the sensing systems herein are only intended for the global vibration-based SHM approach with the main measurand being acceleration. The merit for this type of system is first that it often allows long-term and frequent (or even continuous) monitoring by utilizing ambient and operational loadings as excitation sources and output-only vibration analysis (such as OMA) for feature extraction while the tested structure remains in normal operation. This is crucial for achieving a timely decision-making process concerning the structural health with time including the emergency cases such as possible structural failures. Compared to most other types of measurements, acceleration measurement is more reliable while the sensor (i.e. accelerometer) is more durable over a long period of time. Further, as a global monitoring approach, acceleration measurement allows the monitoring of the entire structure, rather than just each structural component, by a relatively small set of sensors and equipment. However, as a global monitoring approach, vibration measurement has been often believed to be associated with very precise data synchronization. This tends to lead to the dependence on either precise data synchronization methods (causing additional burden of computation and latency for SHM-oriented WSNs); or turnkey (but costly) synchronization hardware solutions in Ethernet distributed DAQ systems particularly when these two sensing platforms are used in large-scale structures. Besides, the (large) scale of actual civil infrastructure often requires large or sparse system

coverage making wired system deployment in general and synchronization hardware solutions in particular very costly. More cost-effective and flexible solutions for synchronization and other DAQ issues are therefore in need to provide more effective operational schemes for SHM-oriented WSNs as well as more affordable Ethernet distributed DAQ systems.

On the other hand, the structural safety evaluation problem considered herein is mainly restricted to following the model-based approach and within level 1 of the damage identification process that is to identify the presence of structural change such as damage or degradation. The data-based approach is adopted because this approach first can avoid challenging tasks in the counterpart (i.e. model-based) approach such as modelling accurately material (e.g. concrete or composite) and structural joints; and updating satisfactorily the model (Farrar and Worden, 2013). As it could bypass the time-consuming model updating process, the data-based approach also tends to be more capable of real-time or near real-time diagnostics which are desirable for the aforementioned timely decision-making goal. For the level-1 scope, although a great deal of studies have been carried out in this level, most of them were based on numerical or laboratory data and therefore have not taken into account the inherent influence of E&O factors in practical SHM programs (Doebbling *et al.*, 1996; Sohn *et al.*, 2003; Farrar and Worden, 2013). Further, in the limited research that has coped with such influence, the problem has not been thoroughly addressed and the false detection rate has still been significant (Gul and Catbas, 2009; Figueiredo *et al.*, 2011). Finally, although OMA is among the most popular feature extraction approaches, commonly-used powerful parametric OMA techniques such as SSI-data techniques have not been thoroughly assessed against distributed measurement uncertainties such as Data Synchronization Error (DSE). Hence, it is believed by the author of this thesis that enhancing the reliability and efficacy in this phase is still very crucial besides addressing the damage identification problems at higher levels. Such enhancements are truly necessary to ensure that the data collected from any sensing system (including those considered herein) will be used in a reliable and effective manner and the whole SHM system will be able to safeguard civil infrastructure for society.

1.4 ACCOUNT OF SCIENTIFIC PROGRESS LINKING THE RESEARCH PAPERS

This research is presented in the format of thesis by published papers in accordance with the PhD regulations at QUT. As the primary contribution of the research project, five thesis chapters (Chapters 3–7) are made of five research papers that have been published or accepted or under review in peer-reviewed journals. For reference purposes, a list of these papers in the same order as numbered hereafter is provided in section 1.6. Three of the papers which are published (Papers No. 1 & 2) or in-press (Paper No. 4) are listed with their URLs included. For the other two papers, the proof of manuscript submission of Paper No. 3 and the proof of full acceptance (for publication) of Paper No. 5 are provided in Appendices A and B, respectively. In addition, the statement of joint authorship for each paper can be viewed in each chapter right before the paper content. The remaining three chapters of this thesis are the introduction (Chapter 1), the literature review (Chapter 2) and the conclusion (Chapter 8). The following subsections are to provide the linkages between the five research papers themselves (with illustrations in Figure 1-1) and with the other chapters. More detailed contributions and linkages of each paper with the others are provided in the opening section of each chapter.

This thesis starts with the introduction chapter for the provision of the background, rationale and outline of the research. As the ultimate target of this study is a practical synthetic SHM system in the data-based approach, Chapter 2 is to provide critical reviews of prior research related to the two main subsystems viz. vibration sensing and data-based structural safety evaluation. From these reviews, promising approaches as well as remaining problems associated are identified for both subsystems. First of all, it was found that synchronization has still been a grand problem particularly in ambient vibration monitoring of large-scale civil structures. For the SHM-oriented WSN platform, some software-based enhancement solutions such as the resampling-based DSE correction method tend to cause costly computation effort and additional system latency when this sensing platform is used for large-scale structures. On the side of Ethernet distributed DAQ systems, the traditional dependence to turnkey synchronization hardware solution can make this type of sensing system very costly for use in large-scale infrastructure which is often

associated with large or sparse measurement coverage. Besides, such coverage also often leads to the use of more sophisticated DAQ equipment thereby further increasing the system development cost while general sensor solutions may not work effectively in demanding ambient excitation conditions. Moreover, the robustness of such popular OMA techniques as those of the SSI-data family with respect to DSE and uncertainties from the data merging process has not been thoroughly investigated. In these regards, Papers No. 1 and 2 (Chapters 3 and 4, respectively) are intended to cope with the main problems of SHM-oriented WSNs and SSI-data techniques (i.e. Research Task 1, as listed in section 1.2) whilst Paper No. 3 (Chapter 5) is to deal with those associated mostly with the Ethernet distributed DAQ platform (Research Task 2). In more detail, Paper No. 1 is to provide (i) a feasibility study towards the use of the semi-complete data synchronization approach in SHM-oriented WSNs for application onto large-scale civil structures; and (ii) more thorough understandings towards the effects of different DSE levels on the widely-used OMA (or equivalently level-1 output-only modal-based damage identification) techniques in general and on the popular primary SSI-data technique (i.e. employing the UPC estimator) in particular. As a continuation study in this trend, Paper No. 2 provides similar evaluations to those presented in Paper No. 1, that is, towards the feasibility of the targeted synchronization scheme and the robustness of the primary SSI-data but for other types of sensor setup and test structure. Specifically, this paper focusses on the problem related to the use of the multi-test setup scheme which is also popular in practice for measuring large-scale structures with a limited number of sensors; as well as the employment of a test structure with a higher range of vibration frequencies to facilitate comparisons amongst different frequency ranges. This assessment task undoubtedly facilitates the recognition of the most robust data merging method and the efficacy of the channel projection scheme in assisting the primary SSI-data to cope with synchronization and data merging uncertainties. Based upon evaluation results of the semi-complete data synchronization approach and the primary SSI-data technique from Papers No. 1 and 2, Paper No. 3 (Chapter 5) proceeds with the development of a wired sensing system solution based on the Ethernet distributed DAQ architecture for the purpose of continuous monitoring of large-scale civil structures. While common measurement and system development issues such as low-level ambient vibration, sparse sensing coverage and budget

constraint as previously mentioned have called for innovations in this research work, the feasibility confirmation for the semi-complete data synchronization approach has provided the basis for the synchronization solution derived in this development. Utilizing the TCP/IP communication technology, this data synchronization solution can not only help to reduce the total cost but also provide greater flexibility for system development work in demanding circumstances such as large or sparse measurement coverage. Using this synchronization solution as well as other effective sensor and peripheral DAQ solutions, a cost-effective and flexible Ethernet distributed DAQ system is developed and implemented onto an actual building at QUT which has a (rather low) frequency range of interest similar to the majority of real-world civil structures. By means of the primary SSI-data technique coupled with the channel projection scheme (as suggested in Paper No. 2), general assessment as well as statistical analysis incorporated daisy data selection, the reliability of this sensing system is thoroughly verified. The vibration data acquired by the system is then used for the purpose of validating the damage identification method and related data generation scheme presented in Paper No. 5.

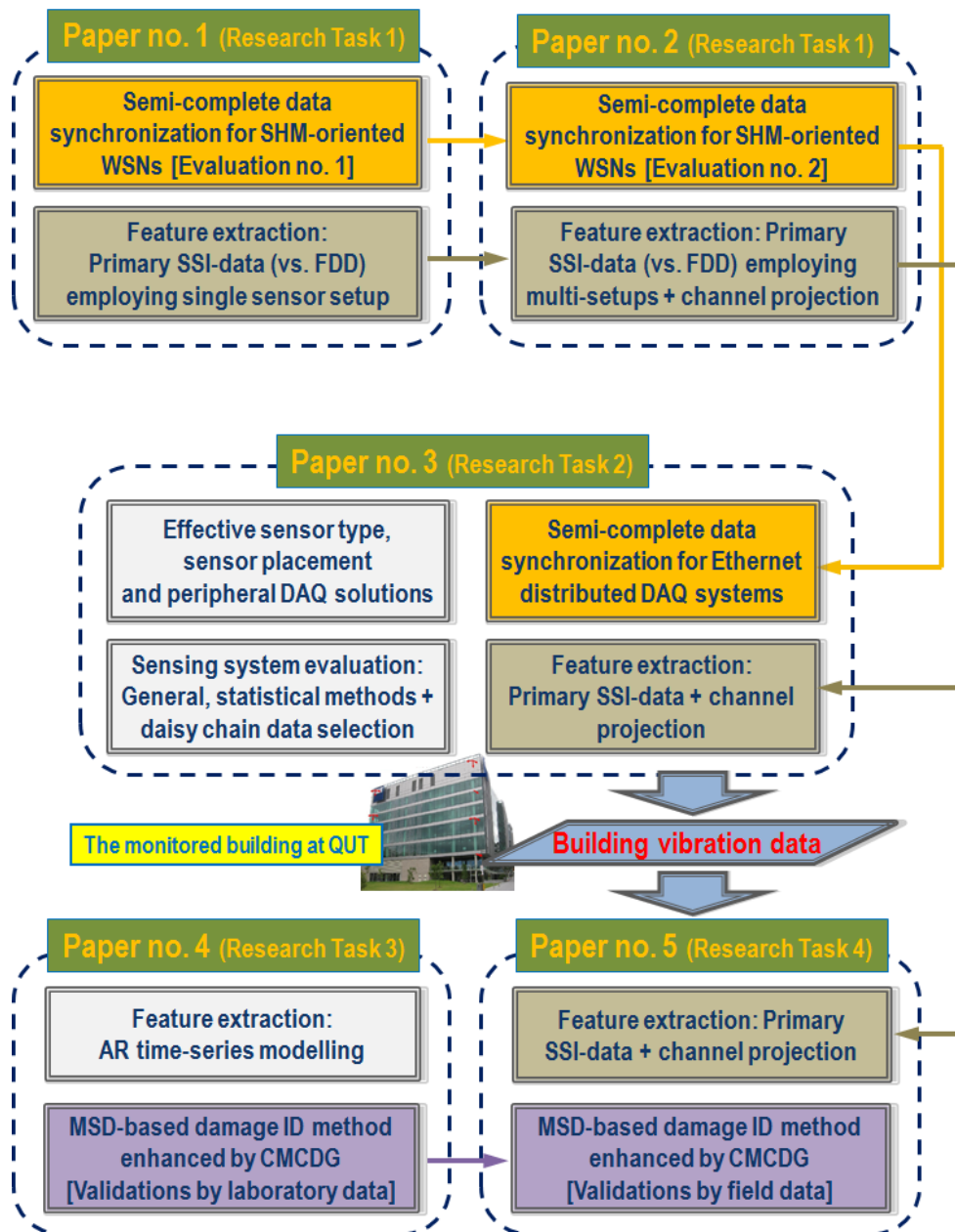


Figure 1-1 Main connections amongst the papers included in the thesis

Besides coping with the issues in sensing systems and the main output-only feature extraction (i.e. OMA) approach, it is also important to develop a robust level-1 data-based damage identification method to overcome the impact of E&O factors on actual vibration monitoring data. In this regard, the outcome of Chapter 2 shows that Mahalanobis Squared Distance (MSD) based method in the statistical damage identification approach is among the most promising candidates for long-term and frequent monitoring purposes due to its architectural simplicity and computational efficiency. Such features make the MSD-based method advantageous when dealing

with large volume of data resulted from frequent or continuous monitoring manners. However, it is also pointed out that this method has an “Achilles heel” that is the requirement of data under a multivariate normal (multinormal) distribution. This tends to become problematic in an early monitoring stage or during short-term SHM programs when only limited actual observations are available. To tackle this problem in a systematic manner, Paper No. 4 (Chapter 6) proposes a so-called Controlled Monte Carlo Data Generation (CMCDG) scheme to assist the MSD-based method in such adverse circumstances and initially validates this scheme with some Auto-Regressive (AR) time-series feature data extracted from a sophisticated SHM benchmark dataset of a laboratory structural model with multiple damage scenarios (Research Task 3, see section 1.2). The theoretical bases of this scheme and its key components are uncovered to consolidate their scientific validity. Two premature datasets with limited observation are extracted to represent an early monitoring stage for two different cases of the feature dimension. To achieve the most desirable goal towards the field validation (Research Task 4), Paper No. 5 (Chapter 7) further applies the CMCDG scheme onto real ambient vibration data of the building studied in Paper No. 3 as well as from an actual bridge with a naturally damaged state. While the shortage of the building vibration data truly represents the problem at an early monitoring stage, the data shortage occurred with the bridge is due to that this bridge was damaged shortly after the introduction of the SHM program making the monitoring (time) span too short.

In correlation with the aim of this research (in section 1.2), the cost-effectiveness and flexibility of the two targeted vibration sensing platforms will be enhanced through novel inexpensive and flexible data synchronization schemes and effective solutions of sensor type, sensor placement and peripheral DAQ presented in Papers No. 1, 2 and 3. On the other hand, the robustness of the safety evaluation paradigm developed herein will be improved by reliable feature extraction by means of robust primary SSI-data configurations (derived in Papers No. 1 and 2) and enhanced MSD-based damage identification (by means of the CMCDG scheme as presented in Papers No. 4 and 5) in every monitoring stage in general and in the challenging early monitoring stages in particular. These enhancements and new developments will be progressively detailed throughout Chapters 3–7 whereas the most significant findings

of the five papers as well as the other achievements of this research program will be summarized in Chapter 8 before recommendations for future work are made.

1.5 RESEARCH INNOVATION

This research includes several main innovations. First of all, it provides two effective semi-complete data synchronization schemes for two most popular vibration sensing platforms (SHM-oriented WSNs and Ethernet distributed DAQ systems) when they are used to monitor large-scale civil structures. Specifically, the first scheme assists SHM-oriented WSNs in more efficient and energy-saving operation through bypassing a computationally costly processing step at leaf nodes whilst retaining sufficiently accurate vibration analysis outcome. On the side of Ethernet distributed DAQ systems, the second scheme provides a cost-effective and flexible data synchronization solution which can be used to replace the costly traditional hardware-based synchronization methods in cases of large or sparse measurement coverage as often encountered with large-scale infrastructure. Besides providing optimal solutions for sensor selection and sensor placement for ambient vibration monitoring, this research is also innovative in the way it evaluates the data synchronization scheme directly on an in-service structure by means of robust statistical assessment and particularly a novel daisy chain data selection scheme so as to overcome the inherent impact of variable E&O. In addition to effectively addressing vibration sensing issues, this research also focuses on the improvement of applicability and reliability of the safety evaluation system so that it can make effective use of vibration data acquired by the investigated sensing platforms. Besides thoroughly evaluating the primary SSI-data technique for the purpose of rapid and reliable OMA, this research also critically selects the computationally efficient MSD-based damage identification method for effectively coping with large volume of data commonly encountered in practical long-term and frequent SHM programs. Realizing the weakness of the MSD-based method, a novel data generation scheme termed CMCDG is then derived so as to truly assist the MSD-based method in overcoming its most vulnerable phases such as those occurring at the early monitoring stages. Not only is CMCDG able to detect such a computational weakness and to provide optimal synthetic data for quality MSD-based learning

processes but the dynamic structure of this scheme also makes it well adaptive to any input data with any primary distributional condition. The strengths and versatilities of both the CMCDG scheme and the MSD-based damage identification method ensure that the safety evaluation system will function reliably from very early operational stages thereby enabling prompt protection for both valuable civil infrastructure and the user community involved.

1.6 LIST OF RESEARCH PAPERS INCLUDED IN THESIS

- (1) Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014. Effects of wireless sensor network uncertainties on output-only modal-based damage identification. *Australian Journal of Structural Engineering* 15 (1):15-25. URL: <http://dx.doi.org/10.7158/S12-041.2014.15.1>
- (2) Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014. Effects of wireless sensor network uncertainties on output-only modal analysis employing merged data of multiple tests. *Advances in Structural Engineering* 17 (3):219-230. URL: <http://dx.doi.org/10.1260/1369-4332.17.3.319>
- (3) Nguyen, T., Chan, T. H. T., Thambiratnam, D. P. and King, L. Development of a cost-effective and flexible sensing system for long-term continuous vibration monitoring. (Under review)
- (4) Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. Controlled Monte Carlo data generation for statistical damage identification employing Mahalanobis squared distance. *Structural Health Monitoring* 13 (4):461-472; URL: <http://dx.doi.org/10.1177/1475921714521270>
- (5) Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. Field validation of controlled Monte Carlo data generation for statistical damage identification employing Mahalanobis squared distance. *Structural Health Monitoring* 13 (4):473-488; URL: <http://dx.doi.org/10.1177/1475921714542892>

Chapter 2: Literature Review

This chapter begins with the review (section 2.1) of the most critical concepts of SHM, vibration-based SHM, damage identification and issues that should be noted in applications of vibration-based SHM methods onto actual civil structures. Then, detailed reviews of vibration sensing and data-based structural safety evaluation (two main vibration-based SHM problems to be considered in this research) are presented in sections 2.2 and 2.3*. Finally, section 2.4 summarizes the core literature review results as well as the research gaps that are identified.

2.1 GENERAL CONCEPTS

2.1.1 SHM

Most SHM experts agree that SHM can be defined as the use of on-structure, non-destructive sensing systems to collect data for the purpose of safety and performance evaluations (Chan *et al.*, 2011; Karbhari and Ansari, 2009). Compared to performance evaluation, safety evaluation through damage assessment has drawn much more attention from SHM researchers as evidenced by the large amount of published literature on the topic as summarized in several comprehensive reviews (Doebbling *et al.*, 1996; Sohn *et al.*, 2003; Yan *et al.*, 2007). One possible reason for this interest symptom is that safety evaluation in general and damage identification in particular are often considered as more immediate problems than performance evaluation. This is supported by the fact that many real civil structures are now approaching or even exceeding their designated service life; and/or being classified as structurally deficient but are still being used due to economic reasons (Farrar and Worden, 2013; Karbhari and Ansari, 2009).

The benefits of SHM are widely envisioned in many textbooks and can be summarized as follows (Chan *et al.*, 2011; Karbhari and Ansari, 2009)

* Parts of these contents are also presented in the “Introduction” sections of Chapters 3–7

- Enhancing the current practice of structural inspection and maintenance from local and subjective condition to global and objective condition.
- Enabling the cost-effective, timely and proactive processes of safety assurance and structural intervention.
- Enabling timely evaluation of structural integrity immediately after extreme events such as typhoons, earthquakes and vehicle/vessel collisions.
- Enabling cumulative and more comprehensive understanding of performance, reliability and risk associated with individual structures thereby enabling more cost-effective and refined methods of structural design as well as linking better design, construction, maintenance and rehabilitation processes together.

2.1.2 VIBRATION-BASED SHM

In order to achieve the above benefits, it is apparent that the structure needs to be monitored and evaluated in a long-term and frequent, or more desirably continuous, basis. The fact that the tragic collapse of the I-35W Bridge in United States in 2007 occurred only three months after its most recent visual inspection has strongly supported this argument (Reid, 2008; Hilkevitch, 2010). In this regard, vibration monitoring offers one of the most effective SHM approaches through utilization of ambient and operational loadings (e.g. traffic, wind and other human related activities) as excitation sources for many civil structures (Karbhari and Ansari, 2009). The merits of this approach is that ambient and operational excitations are essentially at no cost thus are very effective particularly in the long-term and/or frequent monitoring basis while there is no requirement for closure of the structure (the structure remains in normal operation during the testing time). Also, corresponding vibration data reflects actual behavior of the monitored structure. Finally, utilization of global vibration signatures facilitates the monitoring of the entire structure, instead of a structural component, by a relatively small set of sensors and equipment (Karbhari and Ansari, 2009). However, it should be noted that global vibration monitoring requires data streams from different DAQ units to be synchronized with each other for the purpose of accurate mode shape estimation (Aktan *et al.*, 2003; Chintalapudi *et al.*, 2006). Besides, ambient vibration monitoring often requires more

thorough selection of sensor type (than forced vibration testing) for overcoming inherent problems of low-frequency and low-level response measurement (Jo *et al.*, 2012; Karbhari and Ansari, 2009; Xu and Xia, 2012).

The basic premise of vibration-based SHM is that changes in structural properties such as stiffness and boundary conditions will result in changes in vibration characteristics such as modal parameters. When these changes become significant such that they could adversely affect the performance of the structure, the changes can be considered as damage (Karbhari and Ansari, 2009; Farrar and Worden, 2013). By studying the changes in measured vibration features, the presence and the other information (such as location) of damage can be identified.

2.1.3 HIERARCHY AND APPROACHES OF DAMAGE IDENTIFICATION

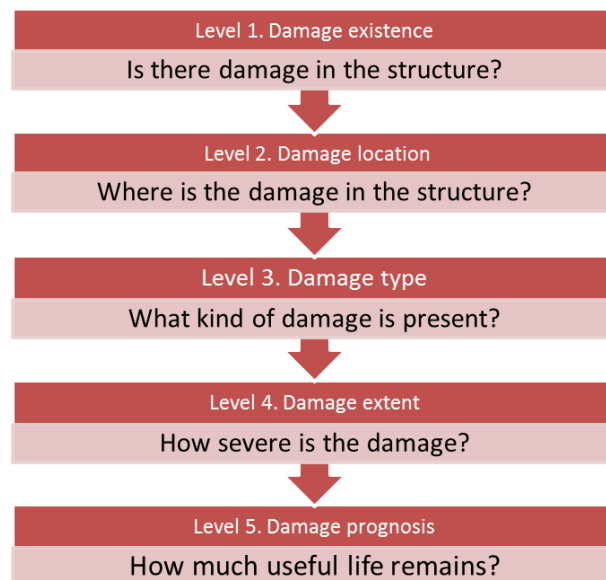


Figure 2-1 Hierarchy of damage identification process [adapted from (Farrar and Worden, 2013)]

The damage identification process is often considered as consisting of five steps which can be formulated as five questions as illustrated in Figure 2-1 (Farrar and Worden, 2013). Answering these questions in the presented steps represents increasing knowledge about structural damage and one can do so by means of one (or combination) of two main approaches commonly known as model-based and data-based approaches. In the former approach, a physical model of the structure of interest is used such as its FE model formulated from detailed physical description of

the structure and updated on the basis of measured data recorded in the normal condition to create a “baseline” model. When data from a subsequent monitoring phase become available with significant deviations (e.g. in modal parameters) compared to the normal condition, a further update of the model will be employed to indicate the location and extent of the structural changes or damage. Theoretically, the model-based approach can be used to provide solutions for problems at levels 2–5. However, as pointed out in Farrar and Worden (Farrar and Worden, 2013), there are several challenging tasks in the model-based approach such as accurate modelling of the material (e.g. concrete or composite), bonds and joints; and updating the model satisfactorily.

On the other hand, the data-based approach does not directly proceed from a physical model. In this approach, one constructs training data from all the possible normal and damaged states and then utilizes pattern recognition (via machine learning) to assign a diagnostic class label to measured data each time it is available from the testing phase (Farrar and Worden, 2013). Theoretically, the data-based approach can be used to answer questions corresponding to levels 1–4. Finally, even though not directly based on the physical model of the structure, data-based methods can still make use of this type of models as a means of investigating effective features or augmenting data of a specific structural state when necessary.

2.1.4 COPING WITH IMPACT OF VARIABLE E&O FACTORS

Even though many prior damage assessment studies have been successful with numerical and laboratory data, implementation of damage identification in actual civil engineering structures is not a trivial task. Vibration characteristics can be sensitive to a certain level of damage but they are also sensitive to changing E&O conditions. This is not a problem of vibration-based SHM alone but tends to be a grand problem of other SHM approaches leading to one of the fundamental axioms of SHM (Farrar and Worden, 2013). In vibration-based SHM (as well as SHM in general), the most notable E&O factors are temperature and operational loading which can cause up to 5-10% deviation on the magnitude of fundamental modal frequencies (Kim *et al.*, 2003; Soyoz and Feng, 2009). Temperature impact was also believed to be the cause of (1) why two lightly damaged cases introduced in the I-40

Bridge did not induce decrease in measured frequencies; and (2) asymmetrical variation, as illustrated in Figure 2-2, in the first-mode shape that changed throughout the day (Farrar and Worden, 2013). It can therefore be seen that such changes in the vibration characteristics can, if not properly accounted for, potentially result in false indications of damage.

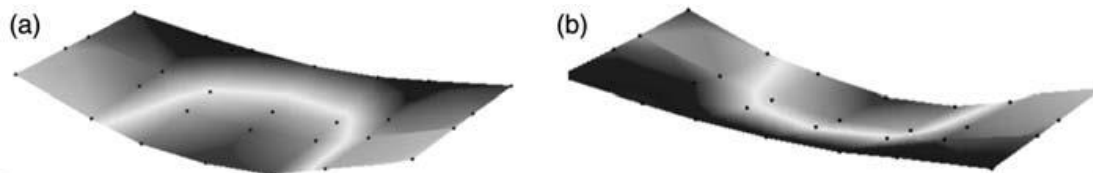


Figure 2-2 Temperature impact on the first-mode shape of one bridge span during two distinct times of the day: (a) in the morning (7.75 Hz); (b) in the afternoon (7.42 Hz) (Figueiredo *et al.*, 2011).

There are two main approaches for coping with the impact of variable E&O factors. Based upon strict measurement of these factors, the first approach tracks the influence of E&O factors on the feature of interest typically by developing predictive model by means of regression techniques (Farrar and Worden, 2013). In the testing phase, this model will be used to predict the feature for a particular E&O factor; and the predicted value and the measured one can be compared to infer if a damaged state has occurred. It is apparent that, the first approach is more applicable to the case when as few as possible number of E&O factors is prevailing (and completely recorded as discussed earlier) which will be difficult to be satisfied in many practical monitoring circumstances of civil structures. Also, it is often impossible to accurately measure such E&O factors as occupant-induced loading in buildings or traffic-induced loading in major bridges. In contrast, the second approach does not require measurement of E&O factors but utilizes (unsupervised) machine learning algorithms (or more simply look-up tables) to “learn” the underlying trends (caused by E&O factors) present in the training data (Farrar and Worden, 2013). This way, these trends can be separated from the influence of any damage present in the testing data. Compared to the look-up tables, machine learning algorithms are generally more reliable for high-dimensional features (as often encountered in SHM) and more applicable for being operated in an autonomous manner. Finally, since they are also frequently used to implement pattern recognition in the data-based approach (as

reviewed in section 2.1.3), the popularity of employing machine learning algorithms in recent damage identification studies can be naturally explained (Sohn *et al.*, 2003; Gul and Catbas, 2009; Figueiredo *et al.*, 2011; Farrar and Worden, 2013).

2.2 VIBRATION SENSING TECHNOLOGIES

From the perspective of data transmission and system communication, there are two main types of vibration sensing systems namely wired and wireless DAQ systems. In spite of being the conventional approach for vibration measurement for many years, the former is still very popular due to its ruggedness and reliability particularly over long periods of time. On the other hand, the wireless DAQ system commonly known as Wireless Sensor Network (WSN) has been recently employed for SHM purposes due to several promising factors such as cost-effectiveness; and rapid and flexible deployment. This section reviews critical global characteristics and remaining issues for each type of systems mainly from global SHM application viewpoints.

2.2.1 WSN

- *System topologies*

A sensor network is considered to be wireless if the communication between its nodes is based on radio frequency (RF) transmission. Most of WSNs for SHM have been operated on the frequency band of 2.4 GHz specified by several IEEE standards such as 802.11b (Wi-Fi), 802.15.1 (Bluetooth) and 802.15.4 (Zig Bee). The main differences between these protocols such as bandwidth, transmission range and power consumption, are needed to be considered depending on specific applications. Among these standards, 802.15.4 has been mainly used on recent WSNs for SHM due to its distinct features including lower energy consumption and longer communication range.

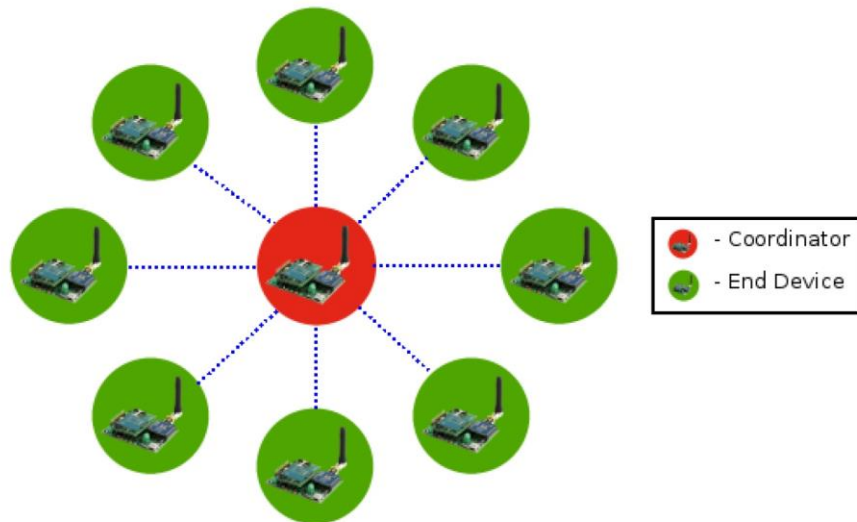


Figure 2-3 Typical diagram of a star WSN topology (Libelium, 2010)

Generally, there are two main basic WSN topologies, namely star and peer-to-peer, the latter is only available in the 802.15.4 standard. In star topology (Figure 1-1), wireless communication is merely established between leaf nodes (or end devices) and the cluster head node (or the coordinator) for one cluster. This topology may be expanded to a three-layer tree topology by employing several clusters and a coordination node at a higher administration level and so on. In contrast, the peer-to-peer protocol (Figure 2-4) supports communication between leaf nodes resulting in more sophisticated architectures such as mesh and multi-hop. For actual WSN deployments so far, the star topology has been preferred due to its simplicity and convenience for the purpose of system debugging and uncertainty assessment.

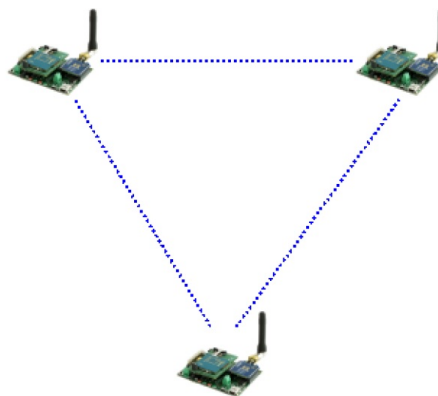


Figure 2-4 Typical diagram of a peer-to-peer WSN topology (Libelium, 2010)

- ***Data aggregation***

There are two primary approaches for data aggregation in WSN, namely centralized and decentralized approaches. In the first one, time histories from each sensor node are transmitted to the base station for data processing and aggregation (Nagayama and Spencer Jr., 2007). This approach has advantages of using simpler algorithms in leaf nodes, providing more intensive vibration analysis, and storing time series data for various post-processing, data mining and structural control purposes (Whelan, 2009; Linderman *et al.*, 2013). On the other hand, the decentralized approach is based on the distributed computing strategy initially proposed by Gao and Spencer (2005). Using this approach, local sensor communities can be established with overlapping nodes and only the data collected by sensors within each community needs to be shared and processed. This way, the amount of data to be transmitted can be reduced and power preservation or system scalability may be better achieved. However, there are several limitations in this approach. First, since feature data is processed directly on the leaf node with limited processing capacity, only limited types of features can be obtained and this limits the number of features as well the methods that can be implemented. Second, since the accuracy of some important decentralized SHM techniques (such as mode shape estimation) depends significantly on (limited) overlapping nodes and the quality of their signals (Sim, 2011), failure or malfunction at any of these nodes can potentially make the feature data useless while the original time histories may be no longer available (Linderman *et al.*, 2013). These limitations could help explain for the popularity of centrally logging and storing data in the format of time histories in most of full-scale WSN deployments so far.

- ***Other advances in WSNs for SHM and remaining issues***

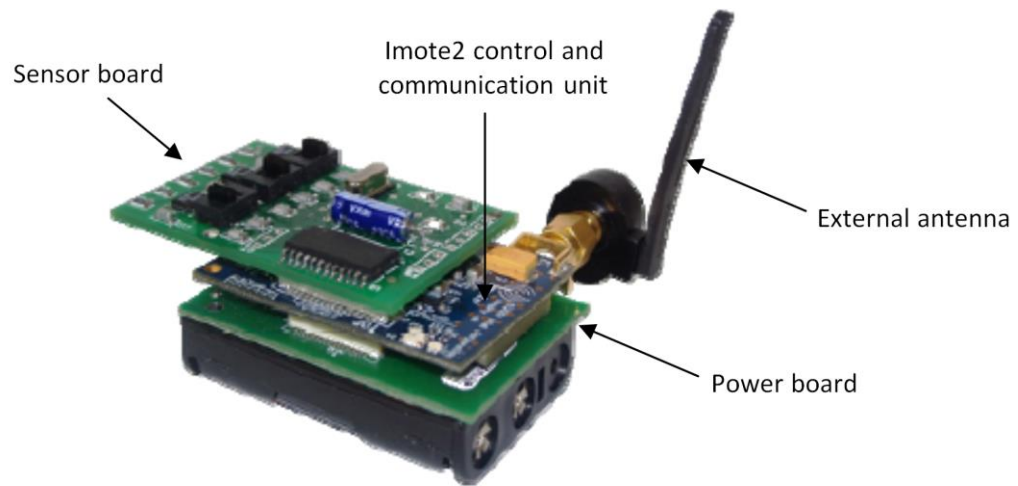


Figure 2-5 A SHM-oriented wireless sensor node based on Imote2 control and communication platform [adapted from (Rice and Spencer, 2009)]

The use of WSNs for SHM has posed a large number of technical challenges (Spencer *et al.*, 2004; Lynch and Loh, 2006). Most commercial WSNs have been initially designed for generic purposes rather than SHM leading to such immediate limitations as low-sensitivity sensors with high noise floor, poor resolution of analog-digital converters, large Data Synchronization Error (DSE) and data loss. Realizing such problems, SHM researchers have begun enhancing capacity of selective WSN models in order to align them with requirements of SHM applications. This can be well illustrated in the case of SHM-oriented WSNs employing MicaZ or particularly Imote2 (Figure 2-5) control and communication platforms (Pakzad *et al.*, 2008; Rice and Spencer, 2009). High-fidelity sensor boards have been customized to enhance the sensing sensitivity and ADCs. External antennas have been applied and reliable communication protocols have been written to achieve longer and more reliable wireless communication (Pakzad *et al.*, 2008; Nagayama *et al.*, 2009; Rice and Spencer, 2009). With such efforts, wireless data transmission without loss is currently achievable though it has not been available in a real-time manner. In the data synchronization aspect, DSE varies significantly from model to model in the generic WSN platform (Lynch and Loh, 2006; Rice and Spencer, 2009). In the SHM-oriented WSN platform, there are several solutions in both hardware and software customization efforts to cope with DSE. Along with customized sensor boards, clock drift estimation and compensation features have also been added into the middleware service in order to combat the timing drift and fluctuation in each node and to

mitigate the timing difference across multiple nodes (Rice and Spencer, 2009; Nagayama *et al.*, 2009). Nagayama *et al.* (2009) proposed a resampling algorithm to provide tight DSE correction and claimed that it can achieve synchronized sensing with accuracy of 30 μ s. However, as it is directly executed at the leaf node, this algorithm can cost more computation effort and increase the data transmission latency. In the regard of DSE impact evaluations, one notable point is that most prior evaluations of DSE employed large DSE values and/or rather high vibration frequencies (of laboratory structural models) to highlight the impact of DSE. Krishnamurthy *et al.* (2008) showed large DSE effects at the modes of 56 and 93 Hz, respectively. Nagayama *et al.* (2007) showed large impact at the DSE of several tens of millisecond. Since the remaining DSE in SHM-oriented WSNs operating without resampling-based DSE correction has been significantly reduced and the measurable vibration frequency range of actual civil structures is much lower (in many cases less than 10 Hz), it is necessary to reassess the DSE impact with these updated/realistic circumstances.

Besides the above data synchronization problems, some general concerns towards effective long-term application of WSNs to actual civil structures still persist. First, continuous sensing is still impossible in SHM-oriented WSNs due to the nature of time-division multiple access protocol in wireless sensors that permits only one node to send data at a time (Linderman *et al.*, 2013). Power supply by normal batteries requires frequent replacement while alternative energy harvesting solutions such as solar rechargeable batteries may still be expensive when considering the possible replacement period such as one year (Jang *et al.*, 2010a). The low reliability of each node and of the whole system particularly operated on the bases of demanding sleeping/waking cycles and shared transmission bandwidth makes the long-term autonomous operation of SHM-oriented WSNs still very challenging.

2.2.2 WIRED SENSING SYSTEM

- *System topologies*

From the signal acquisition perspective, there are two main types of wired sensing system topologies viz. centralized and distributed DAQ systems. While the former has been the traditional measurement approach and mostly suited to small-scale

structures, the latter has been more recently employed for SHM to deal with large-scale structures (Aktan *et al.*, 2003; Van Der Auweraer and Peeters, 2003). Typically, each centralized system consists of a single internal DAQ device or external DAQ chassis housing a number of signal acquisition modules. The sensors are then cabled to these modules via their connectors in a parallel manner which can cause problems for analog signals and installation works if the cable is too long. In contrast, the distributed DAQ is transformed into a network of multiple peripheral DAQ units each of which only needs to be in charge of a group of adjacent sensors. The DAQ units are linked with each other and to the host device by Ethernet cables for the purpose of delivering the digital data (Eren, 2011). As the main signal acquisition tasks take place immediately at each DAQ device, this type of system greatly mitigates the burden of wiring directly from every sensor to the host device and avoids the problem of signal noise induced by lengthy cabling.

- ***Data aggregation***

As wired sensing systems are not subject to limited bandwidth of communication and data transmission as seen in the case of WSNs, data aggregation is often implemented in the centralized scheme. As previously mentioned, this type of data aggregation allows higher versatility of data mining and better efficacy for SHM methods.

- ***Other advances in wired sensing systems for SHM and remaining issues***

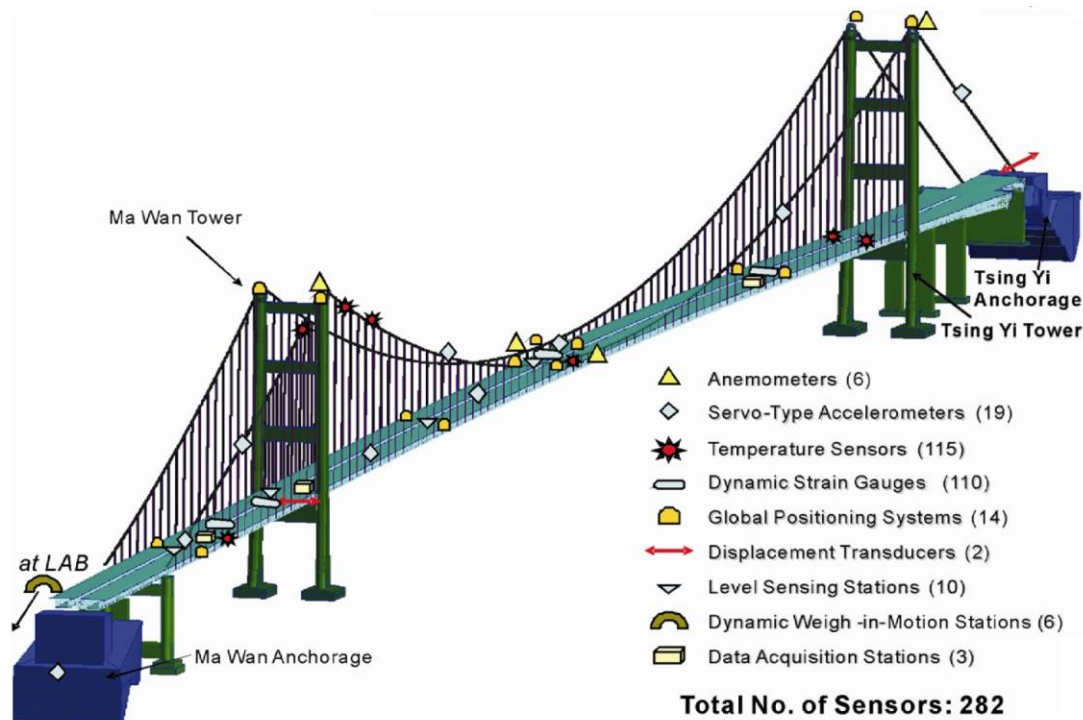


Figure 2-6 Tsing Ma bridge with 1377 m main span (Chan *et al.*, 2011)

Recent deployments of long-term sensing systems for critical civil structures have shown the increasing popularity of the Ethernet distributed DAQ platform. Due to their scale (see Figure 2-6 for such an example), most long-span bridges have employed this type of systems for their permanent SHM systems in the past two decades (Ko and Ni, 2005; Li *et al.*, 2008; Karbhari and Ansari, 2009; Yang and Kranjc, 2010). Multiple sensory and DAQ units are placed along the bridge to ensure the fidelity of the acquired data (Figure 2-7). With similar problems, tall building-type structures have also been equipped with such systems in order to have adequate quality coverage along their height (Ni *et al.*, 2009; Su *et al.*, 2013).

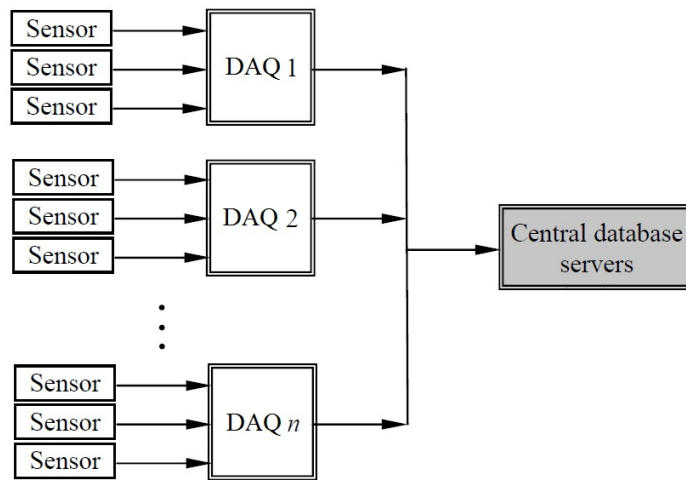


Figure 2-7 Schematic diagram of a typical Ethernet distributed DAQ system used in long-span bridges [adapted from (Xu and Xia, 2012)]

In recent years, the use of off-the-shelf programmable automation controllers such as the popular CRIO (also known as CompactRIO, see Figure 2-8 for an example) as autonomous and rugged Ethernet distributed DAQ systems has become increasingly popular. Initially, these types of systems were mainly designed for high-performance control purposes (Hristu-Varsakelis and Levine, 2005). However, their versatility; and ruggedness and reliability in harsh environments as well as in autonomous continuous acquisition make them suitable for SHM particularly for long-term and permanent deployment purposes. In practice, the CRIO-based systems have seen their applications in a number of SHM projects including both of the main 2008 Summer Olympic venues in Beijing, China namely the National Stadium and the National Aquatics Center (Cigada *et al.*, 2010; Li *et al.*, 2010; McDonald, 2012; Moser and Moaveni, 2013).

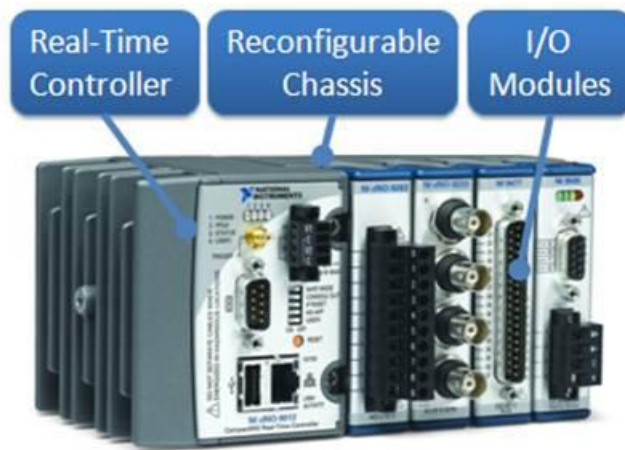


Figure 2-8 Example of CRIO peripheral DAQ model (National Instruments, 2012a)

Although the Ethernet distributed DAQ architecture has been seen applicable to real-world civil infrastructure, actual deployments of such systems have still been rather limited. One main problem for this is the development cost has still been expensive for most of SHM projects especially from the return on investment viewpoint (Ansari, 2005; Karbhari and Ansari, 2009). This is particularly true for vibration sensing systems as this type of system is often believed to require, besides normal expenses for individual sensors and DAQ units, very precise synchronization which often becomes costly particularly in applications onto large-scale structures. Typical examples for this are the adoption of a digital I/O module in each DAQ unit in order to form a dedicated synchronization bus; or the use of Global Positioning System (GPS) in bridges or large-scale building-type structures (Li *et al.*, 2010; Xu and Xia, 2012; McDonald, 2012; National Instruments, 2012c). Besides these prime (hardware-based) methods, there has been a software-based synchronization approach based on TCP/IP command communication (national instrument, 2009). Analogous to what are often used in WSNs, utilizing this approach can provide initial synchronization as well as control the DSE with time through starts and stops processes. However, DSE encountered in this approach if not strictly controlled may be large. In practice, some researchers have attempted to apply this software-based synchronization approach (Cigada *et al.*, 2010; Devriendt *et al.*, 2013). However, since implementation and/or evaluation of the derived synchronization solution were not thoroughly reported, the feature and efficacy of the solutions have not been clear. Overall, it can be concluded that software-based synchronization approach is rather

promising for use with the Ethernet connection since quality timing coordination is feasible with this type of connection (Semancik, 2004). However, more intensive realization and evaluations are in need to make it truly compatible with Ethernet-based DAQ systems in general and CRIO-based systems in particular.

2.3 DATA-BASED SAFETY EVALUATION

As discussed earlier, the structural evaluation paradigm developed herein belongs to the data-based approach with the main scope being at level 1 as clearly defined and explained in section 1.3. As feature extraction and level-1 damage identification are two main consecutive steps in this paradigm, the features, extraction techniques and damage identification methods commonly used in each step will be briefly reviewed in the following.

2.3.1 FEATURES AND FEATURE EXTRACTION TECHNIQUES

In this section, much attention will be given for the features that can be extracted by the output-only extraction approach as these features are more suited for the purpose of long-term and frequent SHM utilizing ambient excitation conditions. Nevertheless, any input-output extraction counterpart related to such a feature will also be reviewed for the purpose of comparison. Besides, robustness of the features and extraction techniques as well as remaining issues with respect to the main sensing uncertainty of interest herein (i.e. DSE) will also be discussed.

- ***Modal parameters***

Modal parameters such as modal frequencies and mode shapes are amongst the most popular vibration features. The basis for this is that damage is assumed to alter stiffness, mass or energy dissipation characteristics (of the structure) which have some relationship with these parameters (Farrar and Worden, 2013). Of the two types of modal parameters, modal frequencies can be estimated more rapidly and more accurately than mode shapes (Salawu, 1997; Dilena and Morassi, 2009). This is particularly true in long-term ambient vibration monitoring where mode shape estimation is often more challenging. Comparing the results from the Z24 highway bridge, Brincker et al. (2001) believed that modal frequencies can be an effective

damage index if the impact from temperature can be tackled. However, if a higher damage identification level (e.g. level 2 to identify damage location) is desired, mode shapes or modal derivatives are generally more effective. It might be worth noting in this regard that modal derivatives should be used with care as some numerical processes (such as differentiation or polynomial fitting) may either amplify high-frequency noise or smooth out the small local impact caused by damage (Farrar and Worden, 2013).

There are two main modal parameter extraction approaches of which OMA is newer than its traditional counterpart named input-output modal analysis. For long-term SHM of civil structures, OMA has often been preferred since this approach does not require expensive excitation resources and can be applied to the structures when they are in operation (Brincker *et al.*, 2003; Karbhari and Ansari, 2009). Amongst numerous OMA techniques, two most popular families are FDD and SSI-data. FDD techniques are well-known for their simplicity and therefore more suited to general users. On the other hand, SSI-data techniques in general and the primary SSI-data technique (i.e. employing UPC estimator) in particular are among the most powerful OMA techniques and therefore often preferred by experienced modal analysts to deal with complicated vibration tests such as those of actual large-scale civil infrastructure (Brincker *et al.*, 2001; Peeters and Ventura, 2003). Besides their original advantages of automated modal identification and coping with closely-spaced or repeated modes, the strength of SSI-data techniques has been significantly improved by the recent employment of the fast multi-order least squares algorithm as a replacement for the traditional least squares solution (Dohler *et al.*, 2012).

Of vibration-based SHM applications, impact of DSE has often been studied on OMA and Output-only Modal-based Damage Identification (OMDI) though the total number of actual studies is still rather limited. In such limited studies, the main finding is that DSE tends to affect modal phase the most and the induced phase shift tends to be proportional to the DSE magnitude and the modal frequency value (Krishnamurthy *et al.*, 2008; Nagayama *et al.*, 2009; Yan and Dyke, 2010). However, the impact of DSE has mainly been examined (i) against the OMA techniques in the non-parametric approach (e.g. FDD) and two-stage parametric approach (e.g. Natural Excitation Technique–Eigensystem Realization Algorithm or NExT–ERA); (ii)

against excessively large DSE values and for simple structural models; and (iii) employing the single test setup scheme. Hence, equivalent investigation problems should be formulated for (1) one-stage parametric OMA techniques (i.e. the SSI-data techniques) due to their popularity as previously discussed; (2) updated DSE values (after recent enhancements in SHM-oriented WSNs) and large-scale civil structures; and (3) sophisticated testing setup schemes such as the multiple test setup strategy that are used for achieving a large number of measurement points with limited sensors.

- ***Time series model properties***

Properties of time series models (such as model coefficients or model residuals) are another popular type of features that has been used more recently for damage identification purposes (Sohn *et al.*, 2001; Carden and Brownjohn, 2008; Gul and Catbas, 2009; Figueiredo *et al.*, 2011). Commonly-used models include AR, ARX (Auto-Regressive with Exogenous Inputs) and ARMA (Auto-Regressive with Moving Average) models. Of these three types of models, ARX model requires both input and output data to be measured therefore tends to be more applicable for controlled force vibration tests such as those using shakers. In contrast, AR and ARMA models are advantageous for the case when only output vibration data is available. With a simpler architecture, AR model has a more straightforward estimation process and more frequently witnesses its applications in the SHM context. However, the direct use of AR model fitted to measured data tends to be more suited to stationary data such as those collected during Gaussian-type forced vibration tests (Farrar and Worden, 2013). As ambient vibration data is often non-stationary, some forms of data normalization such as random decrement (Gul and Catbas, 2009) tend to be in need for this type of data prior to the process of AR model fitting. However, effects of using different normalization input parameters on AR modelling and subsequent damage identification processes need to be thoroughly assessed in order to quantify the associated uncertainties and achieve the most satisfactory normalization outcome.

Depending on the type of the models, the model coefficients can be estimated by means of several popular techniques such as Burg and Yule-Walker (Ljung, 2011).

The rationale of using time series model properties for structural health and damage evaluation is that AR coefficients are directly related to modal frequencies and damping ratios (Carden and Brownjohn, 2008). Also, this employment has several advantages. First, these properties can be used as a signal-based feature and therefore can provide certain local information in a similar manner as from mode shapes. Further, as the feature is computed directly from the time series sequence of each sensing channel, data synchronization across channels is not required (Sohn and Farrar, 2001). Finally, AR coefficients tend to achieve high sensitivity towards nonlinear-type damage such as fatigue cracks that open and close under dynamic loads (Figueiredo *et al.*, 2009).

- ***Other potential features***

Besides, modal parameters and time series model coefficients, properties of wavelet transforms may also be used as damage-sensitive features in general and signal-based features in particular (Reda Taha *et al.*, 2006). However, the fact that there are many types of wavelets, wavelet transforms and the uncertainties involved in the selection processes tends to hinder systematic application of this type of feature. More in-depth comparative studies are in need in order to find the most effective types of wavelets and transforms for safety evaluation purposes.

2.3.2 LEVEL-1 DAMAGE IDENTIFICATION

As discussed earlier, the level-1 damage identification approach adopted herein is constructed in the pattern recognition framework by means of unsupervised learning algorithms. As mentioned in section 2.1.4, the rationale of this adoption is related to the adverse impact of variable E&O factors such as temperature and operational loading; as well as the difficulty in tracking and measuring accurately each of these factors. As they are equivalent to novelty detection methods, methods of damage identification employing unsupervised learning can be classified in the same fashion as for novelty detection methods, that is, conforming to either statistical approach or neural network based approach (Markou and Singh, 2003a, 2003b). Hence, the following section will review the most commonly-used damage identification methods in these two approaches.

- ***MSD-based method***

MSD-based method is one of the most popular methods in the statistical approach of level-1 damage identification. MSD is actually a well-known distance measure in statistics for the purposes of cluster analysis and classification. In damage identification context, MSD is directly employed as the damage (or novelty) index and the underlying trends of E&O factors are taken into account mainly through the sample covariance matrix (see Appendix 5.A or in Chapters 6 and 7). MSD-based method is well-known for its architectural simplicity making it very robust (if not immune) to the uncertainties of architectural assumptions. Further, such system simplicity results in the feature of computational efficiency which make it a potential candidate for embedded sensing systems such as wireless sensors (Worden *et al.*, 2000b; Figueiredo *et al.*, 2011). Also known as outlier analysis based damage detection, MSD-based method has often seen its success in a large number of experimental studies (Worden *et al.*, 2000a; Sohn *et al.*, 2003; Gul and Catbas, 2009; Figueiredo *et al.*, 2011; Farrar and Worden, 2013). However, one notable difficulty in implementing MSD-based method is that it assumes the learning data under a multinormal distribution. In the SHM context, the shortage of measured data due to testing constraints has often been mitigated by means of a basic data generation scheme based on the Monte Carlo simulation methodology (Worden *et al.*, 2000a; Worden *et al.*, 2002; Farrar and Worden, 2013). However, the fact that this scheme is implemented in an uncontrolled simulation manner can be seen as a significant limitation in this data generation approach. More systematic schemes are therefore in need to cope with such realistic problems.

- ***Auto-Associative Neural Network (AANN) based method***

AANN-based method is probably the most popular method in the neural network based approach of damage identification. AANN is actually a multilayer feed-forward perceptron network which is trained to produce, at the output layer, the patterns that are presented at the input layer (Chan *et al.*, 2011; Farrar and Worden, 2013). The network contains three hidden layers: the mapping layer, the bottleneck layer and the demapping layer. By employing fewer nodes than the other two layers, the bottleneck layer can discard trivial variations and extract the predominant trends

such as those induced by E&O factors. Similar to MSD-based method, AANN-based method has often had high rate of success in level-1 damage identification (Worden *et al.*, 2000a; Sohn *et al.*, 2003; Figueiredo *et al.*, 2011; Farrar and Worden, 2013). Advantages of AANN-based method include the capacities of dealing with data that may not have a multinormal distribution and recognizing nonlinear underlying trends. However, AANN-based method has significantly more complicated analyzing architecture than the MSD-based method therefore costs more computational effort in training and testing processes and often requires certain amount of user judgment in setting up the layers (e.g. determining number of nodes) for the network. More details of AANN-based damage identification method are presented in Appendix 5.B.

- ***Other potential methods and other issues***

Besides the above two methods, the other unsupervised damage identification methods include those based on kernel density estimation, factor analysis and principal component analysis (Worden *et al.*, 2000b; Figueiredo *et al.*, 2009; Yan *et al.*, 2005). However, the first method may suffer from the problem of absolute density estimation particularly when the training data is sparse while the other two methods may encounter difficulty in determining number of common factors or principal components (with a similar role as the bottleneck layer in AANN). Finally, it appears that the impact of observation size on all the damage identification methods reviewed herein has not been thoroughly investigated.

2.4 CONCLUDING REMARKS

From the above literature review, the following remarks can be made

- ***On vibration sensing technologies***
 - ✓ SHM-oriented WSNs and Ethernet distributed DAQ systems are currently two increasingly popular sensing system platforms for vibration monitoring of real civil engineering structures. Although it is advantageous in the aspects of easier installation and lower cost, the former platform is still not capable of continuous sensing whereas its long-term stability and reliability are still

uncertain and require more intensive full-scale evaluations. On the other hand, the Ethernet distributed DAQ platform tends to still be more applicable for long-term monitoring whilst it is the sole option (of these two platforms) for continuous monitoring which is often desirable for critical infrastructure.

- ✓ Most of the recent enhancements at the sensor-board and middleware levels have worked well towards realization of more effective WSNs for SHM purposes. However, some software-based solutions such as resampling-based DSE correction tend to cause costly computation effort and additional system latency particularly for large-scale civil structures. This is because these types of structures generally require more lengthy signal sequences (than small-scale structures) in ambient vibration tests for quality implementation of OMA while it is known that the longer the signal sequence is, the higher computational effort the resampling task costs. Under these circumstances, more flexible and effective solutions are in need in order to avoid unnecessarily excessive computation and latency burden.

- ✓ As the main impact of DSE tends to be proportional to the DSE magnitude and the modal frequency value, it may be possible to have a relaxing (semi-complete) data synchronization scheme for use in ambient vibration monitoring of large-scale structures by means of SHM-oriented WSNs. This is because the large-scale structures often have relatively low-value frequencies for the measurable modes while SHM-oriented WSNs can currently attain such rather small DSE as mostly within a single sampling period without the use of the resampling-based DSE correction algorithm (Linderman *et al.*, 2011; Linderman *et al.*, 2013). In addition, ambient monitoring often uses rather high sampling rate (such as around 100 Hz) to achieve better quality of data (e.g. via data decimation process for noise reduction). Hence, it is anticipated that the correspondingly induced DSE impact would be reasonable such that it might be accepted for demanding vibration analysis such as OMA. This can be evaluated by means of vibration data of large-scale civil structures and simulated DSE.

- ✓ Ethernet distributed DAQ systems though rugged and reliable are still expensive as there is yet a lack of cost-effective and flexible solutions at both sensor and DAQ levels particularly for coping with large or sparse measurement coverage problems in large-scale structures. In this regard, using CRIO models especially from the cost-optimized series can provide budget solutions for peripheral DAQ device. Vibration sensor solution though possibly adequate with a relatively small quantity should be carefully chosen to overcome the adversities related to ambient vibration of large-scale structures such as low-frequency and low-level response measurement. Further, using software-based synchronization approach is cost-effective as the corresponding solution itself is essentially at no cost. In this regard, it might be possible to extend the semi-complete data synchronization approach (used earlier in the SHM-oriented WSN platform) for applications to the Ethernet connection since quality timing coordination is feasible with this type of network connection.
- ***On structural safety evaluation***
 - ✓ Modal parameters in general and modal frequencies in particular are among the most popular and robust damage-sensitive features for level-1 damage identification. OMA techniques are more applicable than input-output techniques for the purpose of long-term and frequent monitoring of on-operation civil structures such as existing buildings and bridges. However, the impact of common DSE levels on commonly-used one-stage parametric OMA techniques such as the popular primary SSI-data technique as well as associated practical sensing setup schemes need to be thoroughly investigated to facilitate derivation of appropriate actions. It is necessary because even though one-stage parametric OMA techniques are amongst the most powerful OMA techniques, they can be more susceptible to direct data disturbances such as DSE or from the data merging process when employing the multi-test setup strategy.
 - ✓ Features based on AR coefficients are interesting for provision of spatial information and immunity to DSE impact. However, the direct employment

of this type of features is more applicable to stationary vibration data. Often contaminated with certain amount of non-stationarity and high level of measurement noise, ambient vibration data tends to require some forms of data normalization or signal transformation prior the AR modelling process. This should also be taken into account when considering other types of signal-based features besides AR coefficients.

- ✓ Due to their popularity and high rate of success in level-1 damage identification, MSD-based and AANN-based can be considered as representative methods in statistical and neural approaches, respectively. AANN-based method tends to be more applicable to non-Gaussian multivariate data but generally requires more computational effort and user experience for setting up the network. With a simple computational architecture, MSD-based is among the most computationally efficient methods and therefore more suited to frequent or continuous SHM programs which are apparently data-intensive. However, MSD-based has one “Achilles heel” that is the requirement of the learning data to be multinormal distributed which tends to be problematic at an early monitoring stage or during short SHM programs when only limited experimental observations are available. To enable the effective and reliable use of MSD-based method in such circumstances, the problem of multinormality insufficiency must be properly addressed. With its applicability to multidimensional problems, the Monte Carlo data generation approach if properly systematized can be seen as very suitable tool for assisting the MSD-based method to cope with the aforementioned adversities.

The desire for addressing the above problems altogether has become the motivation for this research program aiming for the development of a practical and reliable synthetic SHM system of which vibration sensing and safety evaluation will naturally become the two subsystems. Due to their own merits that can complement each other in catering for different deployment purposes, SHM-oriented WSN and Ethernet distributed DAQ system are adopted as the two sensing platform subjects of study. The synchronization problems in both sensing platforms are intended to be addressed by the semi-complete data synchronization approach for more effective applications

onto large-scale structures. The adversities from the (large) scale of actual infrastructure and nature of ambient vibration monitoring will also be addressed by means of optimal sensor selection and placement schemes and cost-optimized peripheral DAQ solution with the convenient Ethernet connection. For the purpose of ultimate validation of these solutions altogether at one place, the development for a wired sensing system will be carried out in an actual building structure right at the institutional campus where this research program is carried out. On the side of safety evaluation system, multiple tasks need to be conducted. With their robustness and applicability in tackling large volume of data generally encountered in long-term ambient vibration monitoring programs, the primary SSI-data technique and the MSD-based method are selected as the main feature extraction and damage identification functions, respectively, in the safety evaluation system. As it works directly with time history vibration data, it will be necessary that the primary SSI-data be thoroughly evaluated against dominant uncertainties from both sensing platforms such as those related to synchronization issues. To systematically cope with the problem of multinormality insufficiency associated with the MSD-based method, a fully controlled data generation scheme will be developed by adding on top of the basic Monte Carlo scheme the most necessary advanced tools for determining appropriate simulation configurations so that optimal synthetic data can be obtained. Experimental data not only from laboratories but also from field full-scale vibration tests are always desirable for validation purposes. All these issues will be detailed in subsequent chapters as well as in the two appendices[†] of the thesis.

[†] Appendix A provides connections between the main safety evaluation system components (MSD-based method, CMCDG, etc...) whereas appendix B provides fundamentals of the base Monte Carlo methods

Chapter 3: Effects of WSN Uncertainties on Output-only Modal-based Damage Identification

This chapter is made of the following published journal paper

- ✚ Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014. Effects of wireless sensor network uncertainties on output-only modal-based damage identification. *Australian Journal of Structural Engineering* 15 (1):15-25. URL: <http://dx.doi.org/10.7158/S12-041.2014.15.1>

The main contributions of this chapter to the overall research program are (1) a feasibility study towards the use of the semi-complete data synchronization approach in SHM-oriented WSNs for application to large-scale civil structures; and (2) more thorough understandings towards the impact of different DSE levels on the commonly-used OMA (or equivalently level-1 OMDI) techniques in general and on the popular primary SSI-data technique (i.e. employing the UPC estimator) in particular. In this chapter, a realization of this (semi-complete) data synchronization approach, the relaxing data synchronization scheme, is made by simply deactivating the resampling process in every leaf node and permitting a controllably relaxed DSE level. The feasibility of the data synchronization scheme is then confirmed by the remarkable robustness of modal frequencies as well as the very small and predictable impact of the permitted DSE on MAC values from the mode shapes estimated by both FDD and primary SSI-data techniques employing the single sensor setup strategy. One experimental dataset recorded by a wired DAQ system on a frequently-studied benchmark structure (Ni *et al.*, 2012; Niu *et al.*, 2012) is first selected to act as uncertainty-free data before being numerically contaminated with different levels of DSE in random manners. Since the chosen structure is very flexible, the impact predictability is well evaluated via a fairly large number of vibration modes conveniently available from this type of structure. It should be noted that, even though only synthesized (rather than real) wireless signal data is used, the data

synthesis process has been thoroughly designed and implemented by taking into account the frequently reported DSE levels for both generic as well as SHM-oriented WSN platforms and using one of the most accurate DSE simulation techniques. The random nature of DSE across the sensor nodes has also been taken into account as well. It can therefore be seen that the synthesized wireless signal data used herein can well represent the actual WSN data particularly in terms of network synchronization uncertainties which is the main focus of this research task. Readers interested in the DSE simulation process can refer to Section 4.5.2 for details of this simulation.

The evaluation results from this chapter and from Chapter 4 are to act as bases for another semi-complete data synchronization scheme developed for use in an actual Ethernet distributed DAQ system (see Chapter 5). Finally, the robustness of the primary SSI-data technique is confirmed to be more or less the same as the robust FDD technique. The robustness confirmation for primary SSI-data from this chapter as well as from Chapter 4 is also very important to facilitate rapid while reliable implementation of feature extraction by means of this powerful OMA technique in Chapters 5 and 7 of this thesis.

STATEMENT OF JOINT AUTHORSHIP

The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- There are no conflicts of interest
- They agree to the use of the publications in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

Contributors:

- Mr. Theanh Nguyen (PhD Student): Conceived the paper ideas; designed and conducted experiments (for those conducted at QUT); analyzed data; wrote the manuscripts; and addressed reviewers comments (for published or in-press papers) to improve the quality of paper.
- Prof. Tommy Chan (Principal Supervisor), Prof. David Thambiratnam (Associate Supervisor): Provided critical comments on the student's formulation of the concepts as well as on his experiment and data analysis programs; and provided editorial comments to enable the student to improve the presentation quality of the manuscripts and the revisions for the published and in-press papers.

Principal Supervisor Confirmation: I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Tommy Chan

Name



Signature

18 June 2014

Date

ABSTRACT

The use of Wireless Sensor Networks (WSNs) for vibration-based Structural Health Monitoring (SHM) has become a promising approach due to many advantages such as low cost, fast and flexible deployment. However, inherent technical issues such as data synchronization error and data loss have prevented these distinct systems from being extensively used. Recently, several SHM-oriented WSNs have been proposed and believed to be able to overcome a large number of technical uncertainties. Nevertheless, there is limited research examining effects of uncertainties of generic WSN platform and verifying the capability of SHM-oriented WSNs, particularly on demanding SHM applications like modal analysis and damage identification of real civil structures. This article first reviews the major technical uncertainties of both generic and SHM-oriented WSN platforms and efforts of SHM research community to cope with them. Then, effects of the most inherent WSN uncertainty on the first level of a common Output-only Modal-based Damage Identification (OMDI) approach are intensively investigated. Experimental accelerations collected by a wired sensory system on a benchmark civil structure are initially used as clean data before being contaminated with different levels of data pollutants to simulate practical uncertainties in both WSN platforms. Statistical analyses are comprehensively employed in order to uncover the distribution pattern of the uncertainty influence on the OMDI approach. The result of this research shows that uncertainties of generic WSNs can cause serious impact for level 1 OMDI methods utilizing mode shapes. It also proves that SHM-WSN can substantially lessen the impact and obtain truly structural information without having used costly computation solutions.

KEYWORDS

Wireless Sensor Networks (WSNs), Structural Health Monitoring (SHM), Uncertainties, Generic, SHM-oriented, Data Synchronization Error (DSE), Output-only Modal Analysis (OMA), Output-only Modal-based Damage Identification (OMDI)

3.1 INTRODUCTION

The use of Wireless Sensor Networks (WSNs) for vibration-based Structural Health Monitoring (SHM) has increasingly become popular due to many features such as low cost, fast and flexible deployment. Moreover, this sensing technology is capable of processing data at individual nodes and therefore enabling each measurement point to be a mini intelligent monitoring station (Lynch and Loh, 2006). As a result, many WSNs have been proposed for SHM applications and their capacity and features can be found in several comprehensive reviews (Lynch and Loh, 2006; Rice and Spencer, 2009). In more recent time, SHM research community has paid more attention on commercial WSN platforms as they offer modular hardware and open software which can be further customized with ease to meet requirements of SHM applications.

However, the use of WSNs for SHM poses a number of technical challenges. Most commercial WSNs have been initially designed for generic purposes rather than SHM (Ruiz-Sandoval et al., 2006). As a result, there are many limitations of such a generic platform such as low-sensitivity sensors, high noise, poor resolution of analog-digital converters, inaccurate synchronization and unreliable data transmission (Spencer et al., 2004). Some typical examples can be seen in the cases of the generic version of the Mica or Imote2 WSNs, i.e. using their basic sensors and sensor boards (Ruiz-Sandoval et al., 2006; Rice and Spencer, 2009). Realizing such limitations, a number of research centers have begun enhancing capacity of selective WSN models in order to align them with requirements of SHM applications. High-fidelity hardware components for SHM have been customized and specific middleware algorithms have been written to achieve tighter network synchronization and more reliable wireless communication (Pakzad et al., 2008; Nagayama et al., 2009; Rice and Spencer, 2009). This SHM-oriented WSN platform at present can be best illustrated in the combination of Imote2-based control & communication unit with SHM-A sensor board and middleware developed by the Illinois Structural Health Monitoring Project (ISHMP, see e.g. Rice and Spencer, 2009). With belief of having overcome a large number of WSN uncertainties, these SHM-oriented WSN have been moved from laboratory applications to be deployed in real large-scale infrastructure (Pakzad *et al.*, 2008; Jang *et al.*, 2010b).

Although SHM-oriented WSNs have achieved a number of promising results, there has been very limited validation research examining the effect of improvement of this platform in comparison with its generic counterparts from the SHM application aspect. Impact of uncertainties of both platforms has not been studied in depth, particularly with respect to very popular but demanding global SHM methods such as Output-only Modal Analysis (OMA) and Output-only Modal-based Damage Identification (OMDI). It is worth noting that, OMDI and corresponding OMA techniques, have gained more popularity in comparison to their input-output counterparts in recent years as they are more applicable for monitoring in-service civil structures such as bridges under normal traffic operation (Brincker et al., 2003).

To address this need, this article first presents a review of major uncertainties of both generic and SHM-oriented WSN platforms and their effects on the most popular OMA and OMDI techniques from prior studies. Then, effects of the most inherent uncertainty are investigated with respect to the outcome of a common level 1 OMDI approach, i.e. detecting the presence of structural damage based on the deviation of modal parameters estimated by two most popular OMA techniques. The OMA techniques adopted herein are Frequency Domain Decomposition (FDD) and data-driven Stochastic Subspace Identification (SSI-data) based on the fact that they have been considered as the most robust technique in either frequency domain or time domain and they can well complement each other. For the sake of completeness, FDD, SSI-data and their corresponding level 1 OMDI approach are also described in brief in one of the following sections. Effect of WSN uncertainties on higher levels of the OMDI approach will be addressed in future work. As being the most advanced WSN, Imote2 and its customized hardware and software as previously mentioned are selected as the representative for the SHM-oriented WSN platform in this study.

3.2 MAJOR UNCERTAINTIES OF GENERIC AND SHM-ORIENTED WSNS

There are a number of technical uncertainties or challenges that have been identified by prior studies (Spencer *et al.*, 2004; Lynch and Loh, 2006). However, from a perspective of the most popular SHM methods, two major and distinct WSN uncertainties that can directly degrade data quality are data synchronization error and data loss (Nagayama *et al.*, 2007). A brief review regarding these two factors in both

generic and SHM-oriented WSN platforms and the effort of the SHM research community to address the associated issues are presented below.

Data loss is one intrinsic uncertainty in the generic WSN platform due to two main factors, viz. poor radio signal and packet collision. Sources of the first factor include excessive range of communication (i.e. too far distance between communicating nodes) with limited on-board antenna capacity and interference of environmental factors that can obstruct or degrade the radio signal. Examples for the latter case are the presence of other wireless communication systems or certain building materials like steel (Rice and Spencer, 2009). Data loss due to packet collision occurs when multiple nodes attempt to send data at the same time leading to inference between packets. Prior studies have shown that data loss in generic WSNs can be as large as 20 to 30 percent. In SHM-oriented WSNs, there are both hardware and software solutions to mitigate effects of this uncertainty. External antennas have helped SHM-oriented Imote2 sensors to increase the communication range three times compared to its generic model (Rice and Spencer, 2009). The use of external antennas has also proved to exhibit more consistent behavior with different communication distances. In the software aspect, several reliable communication protocols have been developed in middleware services so that lost data packets can be resent (Mechitov *et al.*, 2004; Nagayama *et al.*, 2009). With such efforts, wireless data transmission without loss is currently achievable though it has not been available in a real-time manner.

Data Synchronization Error (DSE) is probably the most well-known uncertainty in WSNs which consists of two main components, namely initial DSE and jitter-induced DSE. Major sources of initial DSE include the timing offset among local clocks of nodes and the random delay in start time of sensing in each sensor node (Nagayama *et al.*, 2009). Jitter-induced DSE is mainly due to (1) clock drift, (2) fluctuation in sampling frequency of each sensor node and (3) difference in sampling rate among sensor nodes. The combination of the timing offset and clock-drift-induced DSE has been well known as Time Synchronization Error (TSE) which only reflects part of DSE. In the generic WSN platform, DSE varies significantly from model to model and previous reviews (Lynch and Loh, 2006; Rice and Spencer, 2009) have reported fairly large initial DSE values in order of tens to a hundred of

milliseconds for relatively limited communication ranges. Lynch et al. (2005) commented that initial DSE might become larger when a longer transmission range is in use. Clock drift rate difference among nodes can be as large as fifty microseconds per second (Nagayama *et al.*, 2009). It might be worth noting that the total DSE of one sensing segment can be seen as the initial DSE of the next segment (Yan and Dyke, 2010). These mean that DSE could become much larger in practical data acquisition for SHM which can be as long as tens of minutes or more. In the SHM-oriented WSN platform, there are a number of solutions in both hardware and software customization efforts to cope with DSE. Rice and Spencer (2009) have customized a multi-metric sensor board named SHM-A in order to effectively mitigate the second and third source of incremental DSE. The first source of incremental DSE, clock drift, can be effectively dealt with using clock drift compensation algorithm in the time-stamping process during sensing (Nagayama *et al.*, 2009). As a result, the remaining synchronization error for SHM-oriented Imote2 platform is mainly initial DSE which is random in range of a single sampling period (Linderman *et al.*, 2011). Even though a lower initial DSE can be further achieved with resampling algorithm (Nagayama *et al.*, 2009), this algorithm can cost more computation effort at leaf nodes and increase the data transmission latency. Tolerance capacity of SHM applications with respect to relatively small DSE in SHM-oriented WSNs needs to be assessed in order to avoid unnecessarily excessive computational and latent burden.

There are limited studies that have investigated effects of DSE on SHM applications and almost all of them have focused on effects of DSE on limited aspects of outcomes of several OMA and OMDI techniques. The rationale for such studies is, as global SHM methods, OMA and OMDI generally require data from different measurement points to be well-synchronized with each other (Nagayama *et al.*, 2007). It is worth noting that this requirement can be easily met in the traditional wired sensing system but not in case of WSNs with inherent synchronization errors. Nagayama et al. (2007) noted substantial effects of initial DSE on modal phases estimated from one time domain on correlation functions of responses. On the other hand, Krishnamurthy et al. (2008) observed considerable influence of initial DSE on mode shape magnitudes estimated by FDD from experimental data with artificial

introduction of DSE. Later, Yan and Dyke (2010) confirmed effects of DSE on both components of mode shapes and one OMDI method employing one flexibility-based index. All these three investigations have concluded that, DSE has no impact on modal frequencies and damping ratios estimated in the adopted OMA techniques. It is also notable in the latter investigation (Yan and Dyke, 2010) that DSE was randomly contaminated into different sensing nodes within a pre-determined range. This type of simulation can be seen to partially reflect the nature of DSE in WSNs, i.e. randomly different for different measurement points. However, since no multiple-round simulations have been made in this work, statistical properties of impact of DSE randomness on modal parameters and modal-based damage indices have not been derived. In addition to investigations that have been made for nonparametric and correlation-driven OMA, influence of DSE on commonly-used data-driven techniques such as SSI-data needs to be assessed. Finally, although DSE has been greatly reduced in SHM-oriented WSN platform, there appears no comparative study which has been made to evaluate effects of this improvement from a perspective of one OMDI approach. These issues will be addressed later in this study.

3.3 OMA AND LEVEL 1 OMDI APPROACH UNDER INVESTIGATION

Representing for non-parametric OMA is FDD, proposed by Brincker et al. (2000). This technique starts with estimation of output power spectral density matrices each of which (G_{yy}) at a discrete frequency ω_i is then decomposed by the Singular Value Decomposition (SVD) algorithm as below.

$$G_{yy}(j\omega_i) = U_i S_i H_i^H \quad (3-1)$$

Here, U_i is a unitary matrix containing singular vectors u_{ij} as columns and S_i is a diagonal matrix containing singular values (s_{ij}). Next, singular value lines are formed by assembling s_{ij} for all discrete frequencies of interest and plotted for implementing peak-picking technique (see Figure 3-2 for illustration). A mode is generally estimated as close as possible to the corresponding resonance peak of the first singular value line where the influence of the other modes is as small as possible. In the case of two orthogonally coupled modes at one frequency, the

previous step is for the stronger mode whereas a peak on the second singular value line will be “picked” for the weaker mode (Structural Vibration Solutions A/S, 2011). Mode shapes are finally derived from singular vectors corresponding to selected frequencies.

There are two variants of this technique, i.e. Enhanced FDD and Curve-fit FDD but these techniques work similarly except for the fact that estimation of damping ratios is only implemented in the two variants. Similar to traditional input-output non-parametric techniques, FDD family is said to be fast, simple and user-friendly as well immune to computational modes (Zhang *et al.*, 2005). However, difficulties may be arisen in the case that dense and close modes are simultaneously present.

On the other hand, SSI-data has been considered as one of the most robust techniques in time domain since it can take into account furious modes from measurement noise; cope well with dense and closely spaced modes and avoid spectrum leakage (Brincker *et al.*, 2001; Zhang *et al.*, 2005). This method relies on directly fitting parametric state space models to the measured responses of a linear and time invariant physical system (Overschee and Moor, 1996; Structural Vibration Solutions A/S, 2011).

$$x_{t+1} = Ax_t + w_t ; y_t = Cx_t + v_t \quad (3-2)$$

Here, x_t and y_t are the state vector and the response vector at time t , respectively. A is the system state matrix whereas C is the observation matrix. Amongst two stochastic processes, w_t is the process noise (i.e. the input) that drives the system dynamics whilst v_t is measurement noise of the system response.

In later phase, subspace models are first established for different dimensions up to the user-defined maximum value. Among three subspace estimation algorithms, Un-weighted Principal Component (UPC) has been used most for SSI-data of civil structures. Estimates of matrices A and C (i.e. \hat{A} and \hat{C} respectively) are then obtained by the least square solution. By performing the eigenvalue decomposition of the system matrix (\hat{A}), its discrete poles (μ_i) and eigenvectors (Ψ) can be found as described in (Brincker and Andersen, 2006).

$$\hat{A} = \Psi[\mu_i]\Psi^{-1} \quad (3-3)$$

The continuous time poles and subsequently modal frequencies and damping ratios are then obtained.

$$\lambda_i = \frac{\ln(\mu_i)}{\Delta t} \quad (3-4)$$

$$\omega_i = |\lambda_i|; f_i = \frac{|\lambda_i|}{2\pi} \quad (3-5)$$

$$\zeta_i = \frac{\text{Re}(\lambda_i)}{|\lambda_i|} \quad (3-6)$$

Where Δt is the sampling period, the subscript “i” is the index of modes. Mode shape matrix is finally derived from the observation matrix and eigenvectors

$$\Phi = \hat{C}\Psi \quad (3-7)$$

By using increasing subspace model orders, multiple sets of modal parameters for each mode are obtained and their deviation can be used to examine whether that mode is sufficient stable to be from a genuine structural pole. This leads to the extensive use of the stabilization diagram not only in SSI-data (see Figure 3-2 for illustration) but also in most parametric modal analysis methods. It might be worth noting that there is another SSI technique that is based on covariance of data and therefore named covariance-driven SSI but this technique is likely to confront higher computational errors due to the issue of matrix squared up in its calculation process (Zhang *et al.*, 2005).

In practice, FDD and SSI-data have been often used together to complement each other and they are often used for OMA of real civil structures including those being sensed by WSNs (see e.g. Weng *et al.*, 2008; Cho *et al.*, 2010). As for other damage identification approaches, level 1 of OMDI employing these two OMA techniques address the simple but most critical question (i.e. whether the structural damage takes place) by examining changes in modal parameters (Brincker *et al.*, 2001). In this regard, the frequency change and the deviation (from unity) of Modal Assurance

Criterion (MAC, see e.g. Allemang, 2003) are frequently-used damage indices among others (Doebling *et al.*, 1998). The increasing use of this approach for SHM employing WSNs in recent time has proved that it deserves more thorough investigations especially those related to WSN uncertainties as previously reviewed.

3.4 RESEARCH METHODOLOGY

To facilitate a comparative study of effects of DSE of generic and SHM-oriented WSNs on level 1 OMDI employing FDD and SSI-data, a dataset from a benchmark structure was first selected and acts as the DSE-free dataset. This dataset was subsequently polluted with random DSE within a specified range to simulate this uncertainty of both generic and SHM-oriented WSNs based on the review results of each platform. At each DSE range, this pollution process was randomly repeated fifty times to generate fifty datasets for subsequent analyses. In the simulations for SHM-oriented WSNs, DSE range was set to be within one sampling period as previously reviewed and as to relax the option of having to use the costly resampling algorithm. In those for generic WSNs which generally suffer from much larger errors, three DSE ranges selected are five, ten and fifteen times of the sampling period to enable the trend of effects to be quantitatively investigated. For the sake of simplicity, impact of jitter-induced DSE is excluded in this study but one can conclude that it would cause additional adverse influence in SHM applications using generic WSNs. The DSE-free data and all DSE-corrupted datasets are used as the input, for FDD and SSI-data techniques, to identify modal frequencies, mode shapes and corresponding comparative indicators. Damping ratios are not considered in this study based on the fact that their estimation with acceptable accuracy may still be uncertain in OMA and they are not among commonly-used indices for SHM (Brincker *et al.*, 2001). In a similar fashion as for damage detection process, relative frequency change and the deviation of MAC from unity are selected as two indicators for assessing effects of different DSE ranges on level 1 OMDI herein. Details of the original dataset, simulation approach for DSE and the analysis procedure are given the following sections.

3.5 DESCRIPTION OF DATASETS AND ANALYSIS

3.5.1 THE BENCHMARK STRUCTURE AND DATASETS

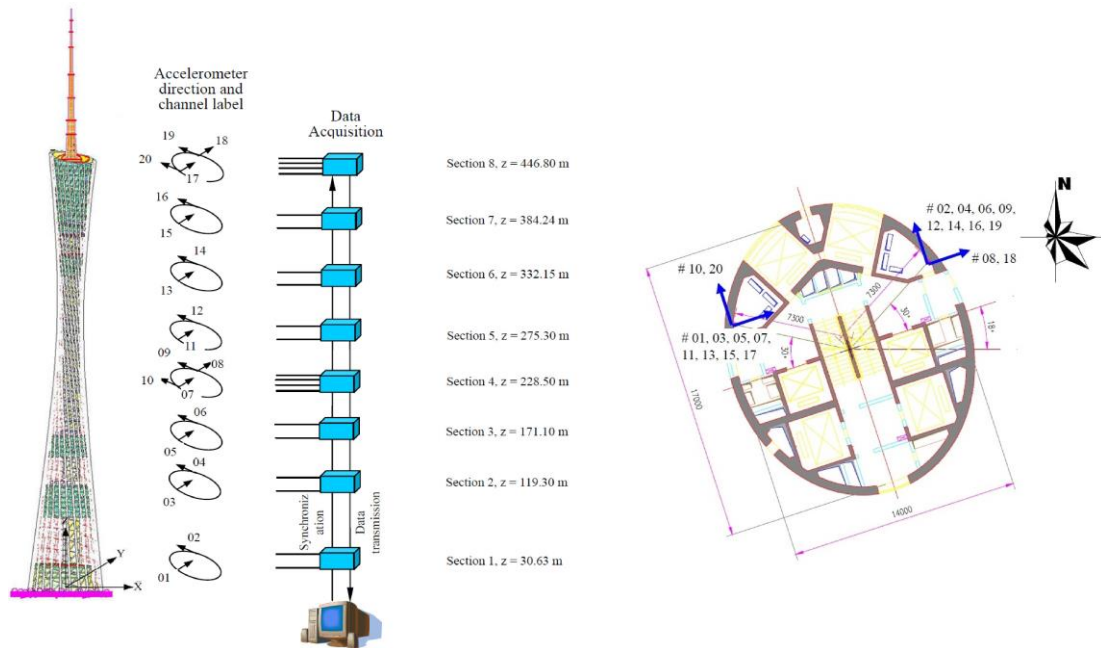


Figure 3-1 Position of the accelerometers and the wired DAQ system on GNTVT

Even though simpler types of data can be generated through computer simulations or laboratory experiments, it is the intention of the authors to use real monitoring data from real civil structures in this study. The rationale for this is the pattern of real data is likely to be different from that of data generated in numerical simulations or laboratory experiments since real civil structures are subjected to influence of different operational or environmental factors such as wind and measurement noise. The dataset selected to use for this case study is from Guangzhou New TV Tower (GNTVT). This 610m super-tall tower has been considered as a benchmark structure for SHM and its one-day data as well as the full description of the SHM system are freely provided in a website for SHM research community (<http://www.cse.polyu.edu.hk/benchmark/>). Figure 3-1, taken from this website, shows the arrangement of 20 uni-axial accelerometers installed at eight levels along the height of this tower. Sensors were placed along short-axis and long-axis of the inner structure. The sampling rate was set at 50 samples per second which can be seen to belong to a common range for SHM of real structures. The provided data were split into 24 sets of one hour length and the 7th dataset (i.e. named

accdata_2010-01-20-00) was chosen as benchmark (or DSE-free) dataset in this study.

3.5.2 SIMULATION OF DSE AND ANALYSES OF DSE IMPACT

As previously discussed, four DSE ranges were chosen which are within one, five, ten and fifteen times of the sampling period (i.e. dt) in which the first range represents DSE of SHM-oriented WSNs and the others represent those of generic WSNs. For a given DSE range, fifty sets of the time delay vector were randomly generated and each set was used to interpolate the corresponding DSE-corrupted dataset from the benchmark dataset. It is worth noting that, among various one-dimensional interpolation techniques, the linear interpolation technique has already been utilized in the resampling algorithm for SHM-oriented WSN middleware (Nagayama *et al.*, 2009) due to the fact that it requires less computational effort from sensor resources. Since the simulations herein are not subjected to such a computational constraint, the cubic spline interpolation technique (MathWorks, 2011) was adopted to achieve more accurate simulation results.

The DSE-free and DSE-corrupted datasets were used as the input for FDD and SSI-data techniques. The analyses were conducted using ARTeMIS Extractor software ver. 5.3 (Structural Vibration Solutions A/S, 2011). Since number of the sensors was rather large, the channel projection was adopted which can help to reduce effects of noise and avoid too much redundant cross information. The minimum number of the projection channels is generally three. The basis behind this is that, in case of two close modes, at least two projection channels are needed to separate the modes plus one additional channel to account for the measurement noise (Structural Vibration Solutions A/S, 2011). After several trials, the number of projection channels selected was four as they provided the most stable stabilization diagram with the least noise modes. Also, the dimension for the state space model was set 160 as it was found to be sufficient for performing SSI-data. For each DSE range, fifty sets of modal parameters (i.e. frequencies and mode shapes) were estimated at each mode, compared with the benchmark modal parameter set (i.e. from the DSE-free or original dataset) to calculate fifty corresponding sets of relative frequency changes and MAC deviations from unity.

To evaluate narrow-range changes of modal parameters like frequencies, basic statistic figures are sufficient such as Root Mean Square Error (RMSE) of DSE-corrupted frequencies with respect to the DSE-free frequency and relative difference of maximum and minimum DSE-corrupted frequencies with respect also to the DSE-free frequency. In order to visualize largely different deviations of different variables in one plot, box-plot function (MathWorks, 2011) was adopted to visualize some useful statistical properties (such as median, quartiles and extremes) of MAC deviations under impact of different DSE ranges.

3.6 RESULT AND DISCUSSIONS

3.6.1 COMMON RESULTS OF FDD AND SSI-DATA

The first twelve modes investigated lie on the frequency range of between 0.09 Hz and 1.3 Hz. The results of the mode estimation are in excellent agreement between both OMA techniques (i.e. FDD and SSI-data) for both DSE-free and DSE-corrupted data. MAC values of twelve mode shape vectors estimated by two techniques are approximately unity. All these twelve modes vibrate mostly along with either of two sensor-placement directions, except modes 6 and 12 which are coupled between the two directions. Figure 3-2 shows the singular value plot employed in FDD technique and stabilization diagram utilized in SSI-data whereas Figure 3-3 illustrates some typical modes vibrating mostly along with the short-axis direction (of the inner structure, see also Figure 3-1) for the DSE-free dataset. The results are also in good agreement with prior studies in regards to OMA of this benchmark structure (Chen et al., 2011).

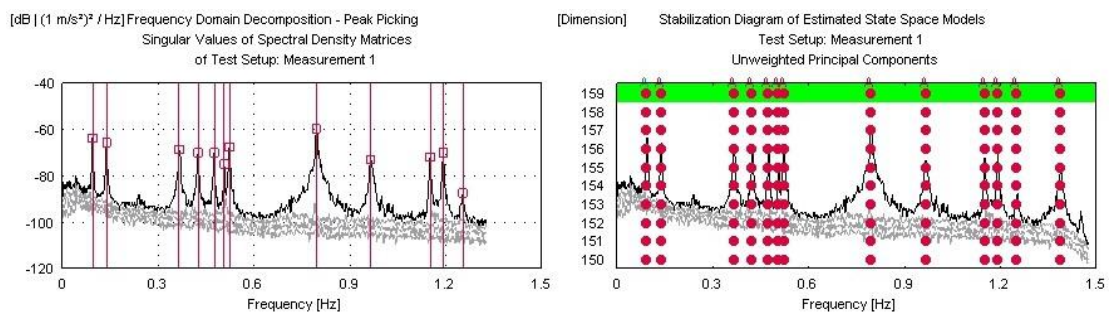


Figure 3-2 Singular value plot for FDD (left) and stabilization diagram for SSI-data (right) from DSE-free data

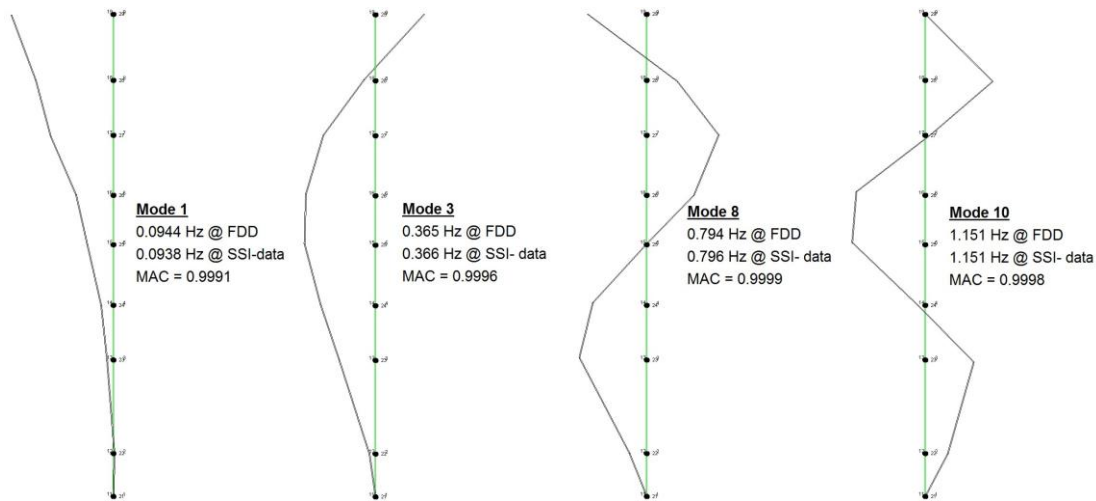


Figure 3-3 Four typical modes in the short axis estimated by FDD and SSI-data from DSE-free data

3.6.2 EFFECT OF DSE ON LEVEL 1 OF OMDI

- **Frequency change**

There is no change in frequencies estimated by FDD for both DSE-free and DSE-corrupted data. This once again reinforces prior findings that DSE does not affect frequencies estimated by FDD (Krishnamurthy *et al.*, 2008; Yan and Dyke, 2010) and highlights the robustness of this technique with respect to DSE impact.

Frequency estimates by SSI-data are subjected to certain influence from DSE but the impact is very small even for the case of the largest DSE considered such as 15dt as illustrated in Table 3-1. The maximum RMSE, occurred at the highest mode of interest (mode 12) is only $0.392e-3$ Hz whereas the maximum relative difference is less than 0.5 percent. Therefore, effects of DSE on frequency estimates by SSI-data can apparently be considered to be negligible.

Table 3-1 Effects of DSE of 15dt on frequency estimates by SSI-data

<i>Mode</i>	<i>DSE-free (Hz)</i>	<i>RMSE (mHz)</i>	<i>Min (Hz)</i>	<i>Max (Hz)</i>	<i>RD (%)</i>
1	0.0938	0.123	0.0938	0.0935	0.367
2	0.1382	0.079	0.1382	0.1380	0.121
3	0.3661	0.051	0.3661	0.3660	0.032
4	0.4241	0.030	0.4241	0.4240	0.023
5	0.4748	0.036	0.4749	0.4748	0.017
6	0.5060	0.089	0.5061	0.5058	0.050
7	0.5228	0.017	0.5228	0.5227	0.015
8	0.7957	0.266	0.7958	0.7951	0.082
9	0.9663	0.022	0.9664	0.9663	0.010
10	1.1509	0.044	1.1510	1.1508	0.013
11	1.1916	0.034	1.1917	1.1916	0.012
12	1.2520	0.392	1.2525	1.2509	0.127

[Note: Relative difference, RD = (Max-Min)/DSE-free]

• ***Mode shape change - MAC deviation (from unity)***

At each DSE range for one OMA technique, a 50×12 (number of simulations/observations × number of modes) matrix of MAC deviations from unity was established and visualized by the box-plot function. To facilitate a complete comparison across different DSE ranges and with respect to two OMA techniques, a total of 8 box-plots are illustrated with the same scale in Figure 3-4 (for generic WSNs) and Figure 3-5 (for SHM-oriented WSNs). However, MAC deviations for the latter (i.e. SHM-oriented WSNs) case were re-plotted in a zoomed scale in Figure 3-6 to have a clearer view of their statistical distribution.

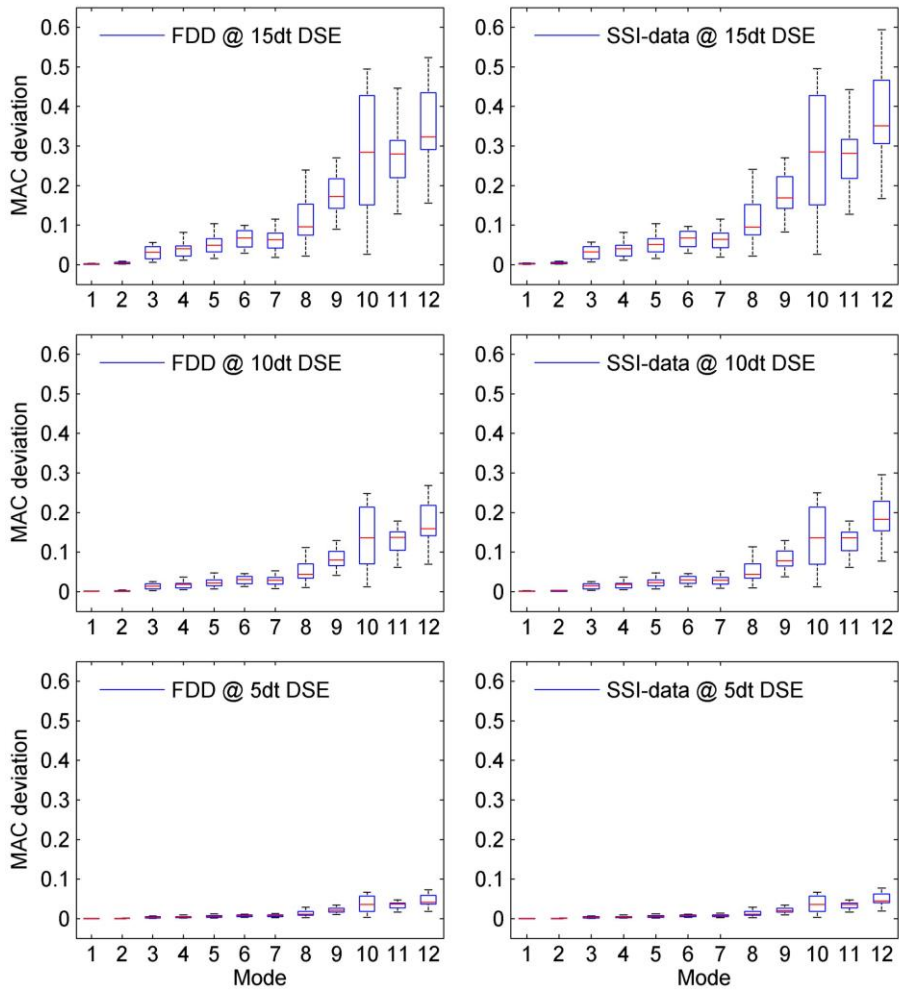


Figure 3-4 Box-plots of MAC deviation under DSE of generic WSNs

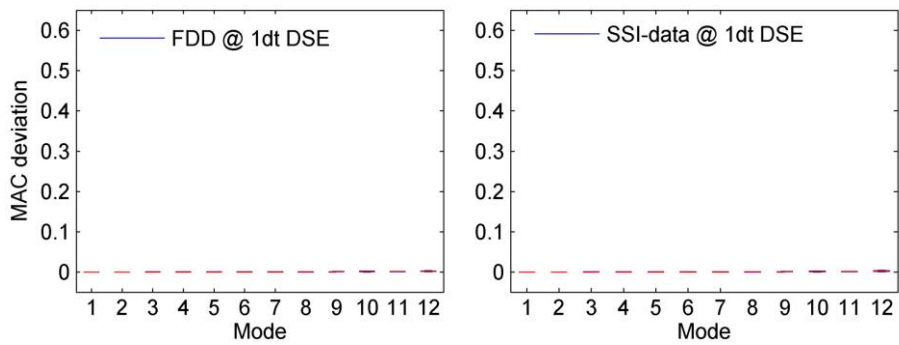


Figure 3-5 Box-plots of MAC deviation under DSE of SHM-oriented WSNs

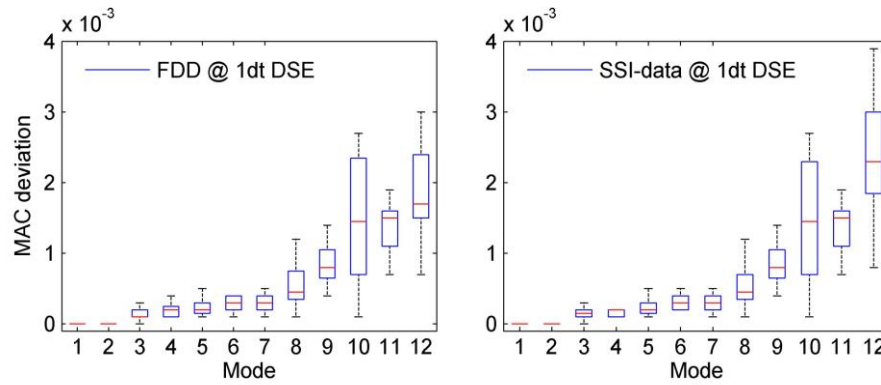


Figure 3-6 Box-plots of MAC deviation under DSE of SHM-oriented WSNs (zoomed scale)

Figure 3-4 and Figure 3-5 clearly show the negative impact of DSE of generic WSNs which increases rapidly for higher DSE ranges especially for higher modes. For instance of modes 10 and 11, whilst the median of MAC deviations under DSE range of 5dt is only about 0.05, those under DSE of 10dt and 15dt are around 0.13 and 0.28, respectively. Also, the variation of MAC deviations at higher DSE ranges is much larger than that at the lower range, again particularly at higher modes. This can be obviously seen through the total range (i.e. distance between the lower and upper extremes) as well as the Inter-Quartile Range (IQR) which is the difference between the 75th and 25th percentile of the presented data. For instance of modes 8 and 9, IQR of MAC deviations under DSE range of 5dt is approximately 0.15, those under DSE ranges of 10dt and 15dt are around 0.35 and 0.8, respectively. MAC deviations as large as from 0.2 to 0.35 or even larger for extreme cases (i.e. almost up to 0.6) would cause problems for level-1 OMDI relying on mode shape changes since 0.2 is also the MAC deviation (of the highest detectable mode) for the most severe damage cases in a real bridge (Brincker *et al.*, 2001).

In general, DSE impact on statistical features of MAC deviations such as their median value and their variation is higher for higher modes. This trend can be seen not only on plots for cases of generic-WSN DSE but also on the zoomed plots (i.e. Figure 3-6) for the case of SHM-oriented DSE for both OMA techniques. However, for the latter case, the actual impact magnitude is drastically reduced with even the highest extreme value of MAC deviations being less than 0.005 (i.e. at mode 12). Obviously, this impact level can be considered to be marginal in comparison with

those levels which have been discussed above and it also shows that OMDI is likely to tolerate the DSE level of SHM-oriented WSN without having to use costly computational algorithm like resampling approach.

Under adverse influence of DSE, outcomes of FDD and SSI-data are generally rather similar though one could see higher impact on SSI-data at the highest mode (i.e. mode 12). The robustness of SSI-data herein once again evidences why this technique has been believed to be the best choice for accurate OMA in both off-line and automate manner (Brincker *et al.*, 2001).

3.7 CONCLUSIONS

This article has presented a comprehensive investigation of uncertainties of both generic and SHM-oriented WSN platform and their effects on a common level 1 OMDI approach. Based on an intensive review, this study has first revealed that whilst data loss can be effectively treated using reliable communication protocols, DSE is still unavoidable and can be considered as the most inherent uncertainty. The review has also shown that the DSE magnitude has been considerably alleviated in the SHM-oriented WSN platform by advanced combination of hardware and middleware solutions, and will possibly help avoiding the use of costly computational methods for compensation of DSE impact. To evaluate such improvements in SHM-oriented WSNs as well as highlight the limitation of the generic WSN platform, a comparative study was carried out with focus on applications on real civil structures. One experimental dataset from a benchmark structure was first selected to act as uncertainty-free data before being contaminated with different levels of DSE in random manners to practically simulate this uncertainty in both WSN platforms. In order to gain a more thorough understanding of DSE impact, statistical analyses, for the first time, were employed to derive critical distributions and variation patterns of two common level-1 OMDI indices i.e. frequency changes and MAC deviations from unity. The results have first shown that, the robustness of SSI-data with respect to DSE impact can be more or less the same as that of FDD. In terms of damage indices, the frequency-based index is the most robust one since DSE causes no (or almost no) change for frequency estimates. However, the second index (i.e. MAC deviation which is commonly used for

assessing mode shape change) has subjected to rather significant influence from DSE, particularly with large DSE ranges of generic WSNs. Likewise; this impact has been shown to be increased with the order of modes, proving that higher modes are more sensitive to DSE. In the same regard, capacity of SHM-oriented WSN platform has been assessed and shown that its improvement has greatly lessen the adverse impact of DSE and that OMDI is likely to perform well with data from SHM-oriented WSNs. It is also worth noting that although the effects of uncertainties like DSE on OMDI have just been investigated for level 1, the outcomes of this study can act as basis for further investigations on higher levels of OMDI. As the final finding from this study, statistical approach is ultimately recommended for investigations of WSN uncertainties particularly for data synchronization errors of generic WSNs. Beside basic statistical feature such as RMSE or relative difference, box-plots have been proved to be useful in presenting, in one plot, different variables with variations in rather different scales.

Chapter 4: Effects of WSN Uncertainties on Output-only Modal Analysis Employing Merged Data of Multiple Tests

This chapter is made up of the following published journal paper

- ✚ Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014. Effects of wireless sensor network uncertainties on output-only modal analysis employing merged data of multiple tests. *Advances in Structural Engineering* 17 (3):219-230. URL: <http://dx.doi.org/10.1260/1369-4332.17.3.319>

As a continuation of the study in Chapter 3, this chapter has similar contributions to those presented in Chapter 3 but for other types of sensor setups and test structures. Specifically, this chapter focuses on the problem related to the use of the multiple test setup strategy which is also popular in practice for measuring large-scale structures with a limited number of sensors. In this regard, there are two methods for merging data from multiple tests leading to another SSI-data variant (termed UPC-PreGER in the chapter) besides the primary SSI-data technique (termed UPC-PoSER herein for this particular sensor setup scheme). On the other hand, the feasibility of the relaxing data synchronization scheme is assessed by means of vibration data from a structure with a lower degree of flexibility than the one used in Chapter 3. The rationale behind this is to examine the impact at a frequency range that is higher than the range investigated in Chapter 3. In a similar fashion as those used in Chapter 3, synthesized wireless signal datasets are thoroughly created so that they can represent the actual WSN data particularly in the aspect of network synchronization uncertainties. Once again, the feasibility of the (semi-complete) data synchronization scheme is reconfirmed by the undeniable robustness of modal frequencies as well as the relatively small and predictable impact of the permitted DSE level on MAC values particularly for those modes estimated by primary SSI-data or in the low vibration frequency range (e.g. 0–10 Hz) of civil structures. The DSE impact is also reconfirmed to be analogous to the influence of measurement noise, that is, generally

being more severe for modes of higher order. Besides facilitating such confirmations, the data synchronization evaluation results from this chapter and from Chapter 3 are to act as bases for the cost-effective and flexible data synchronization solution presented in Chapter 5. Finally, besides the anticipated robustness of FDD against noise-type uncertainties, the chapter also highlights the necessity of using channel projection in helping in general the SSI-data family to cope with such adversities as well as in assisting in particular the primary SSI-data technique to attain a robustness level analogous to that of FDD. Along with related results from Chapter 3, these findings contribute further to the reliable implementation of this SSI-data technique on subsequent applications in Chapters 5 and 7.

STATEMENT OF JOINT AUTHORSHIP

The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- There are no conflicts of interest
- They agree to the use of the publications in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.


Contributors:

- Mr. Theanh Nguyen (PhD Student): Conceived the paper ideas; designed and conducted experiments (for those conducted at QUT); analyzed data; wrote the manuscripts; and addressed reviewers comments (for published or in-press papers) to improve the quality of paper.
- Prof. Tommy Chan (Principal Supervisor), Prof. David Thambiratnam (Associate Supervisor): Provided critical comments on the student's formulation of the concepts as well as on his experiment and data analysis programs; and provided editorial comments to enable the student to improve the presentation quality of the manuscripts and the revisions for the published and in-press papers.

Principal Supervisor Confirmation: I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Tommy Chan

Name



Signature

18 June 2014

Date

ABSTRACT

The use of Wireless Sensor Networks (WSNs) for vibration-based Structural Health Monitoring (SHM) has become a promising approach due to many advantages such as low cost, fast and flexible deployment. However, inherent technical issues such as data asynchronicity and data loss have prevented these distinct systems from being extensively used. Recently, several SHM-oriented WSNs have been proposed and believed to be able to overcome a large number of technical uncertainties. Nevertheless, there is limited research verifying the applicability of those WSNs with respect to demanding SHM applications like modal analysis and damage identification. Based on a brief review, this paper first reveals that Data Synchronization Error (DSE) is the most inherent factor amongst uncertainties of SHM-oriented WSNs. Effects of this factor are then investigated on outcomes and performance of the most robust Output-only Modal Analysis (OMA) techniques when merging data from multiple sensor setups. The two OMA families selected for this investigation are Frequency Domain Decomposition (FDD) and data-driven Stochastic Subspace Identification (SSI-data) due to the fact that they both have been widely applied in the past decade. Accelerations collected by a wired sensory system on a large-scale laboratory bridge model are initially used as benchmark data after being added with a certain level of noise to account for the higher presence of this factor in SHM-oriented WSNs. From this source, a large number of simulations have been made to generate multiple DSE-corrupted datasets to facilitate statistical analyses. The results of this study show the robustness of FDD and the precautions needed for SSI-data family when dealing with DSE at a relaxed level. Finally, the combination of preferred OMA techniques and the use of the channel projection for the time-domain OMA technique to cope with DSE are recommended.

KEYWORDS

Wireless Sensor Networks (WSNs), Data Synchronization Error (DSE), Output-only Modal Analysis (OMA), Multi-setup, Frequency Domain Decomposition (FDD), data-driven Stochastic Subspace Identification (SSI-data)

4.1 INTRODUCTION

The use of Wireless Sensor Networks (WSNs) for vibration-based Structural Health Monitoring (SHM) has increasingly become popular due to many features such as low cost, fast and flexible deployment. Moreover, this sensing technology is capable of processing data at individual nodes and therefore enabling each measurement point to be a mini intelligent monitoring station (Lynch and Loh, 2006). As a result, many WSNs have been proposed for SHM applications and their capacity and features can be found in several comprehensive reviews (Lynch and Loh, 2006; Rice and Spencer, 2009). In more recent time, SHM research community has paid more attention on commercial WSN platforms as they offer modular hardware and open software which can be further customized with ease to meet requirements of SHM applications.

However, the use of WSNs for SHM poses a number of technical challenges. Most WSNs have been initially designed for generic purposes rather than SHM (Ruiz-Sandoval *et al.*, 2006). As a result, there are many limitations of such a generic platform such as low-sensitivity sensors, high noise, poor resolution of analog-digital converters, inaccurate synchronization and unreliable data transmission (Spencer *et al.*, 2004). Typical example can be seen in the case of the generic version of the Imote2 WSN, i.e. using basic sensors and sensor board ITS400 (Rice and Spencer, 2009). Realizing such limitations, a number of research centers have begun enhancing capacity of selective WSN models in order to align them with requirements of SHM applications. High-fidelity sensor boards for SHM have been customized and specific middleware algorithms have been written to achieve tighter network synchronization and reliable wireless communication (Rice and Spencer, 2009; Pakzad *et al.*, 2008; Nagayama *et al.*, 2009). This SHM-oriented WSN platform can be best illustrated in the combination of Imote2-based control & communication unit with SHM-A sensor board and middleware developed in the Illinois Structural Health Monitoring Project (Rice and Spencer, 2009). Since they are the most popular WSNs which have been used for SHM applications, the generic and SHM-oriented platforms of Imote2 are selected as representatives for this study hereafter.

Although SHM-oriented WSNs have achieved initial promising results, uncertainties of this platform have not been completely removed. Effects of SHM-oriented WSN uncertainties have not been studied in depth, particularly with respect to popular but demanding global SHM applications such as output-only modal analysis (OMA) and output-only modal-based damage identification (OMDI). It is worth noting that, OMA and OMDI have gained more popularity in comparison to their input-output counterparts in recent years as they are more applicable for monitoring in-service civil structures such as bridges under normal traffic operation (Brincker *et al.*, 2003).

To address this need, this paper first presents a brief review of major uncertainties of the SHM-oriented WSN platform and their effects on OMA techniques from prior studies. Then, effects of the most inherent uncertainty are investigated with respect to one of the frequent OMA applications, i.e. OMA employing merged data from multiple tests (Dohler *et al.*, 2010). Frequency Domain Decomposition (FDD) and data-driven Stochastic Subspace Identification (SSI-data) are selected for this investigation as each of them has been considered as the most robust technique for either frequency domain or time domain.

4.2 MAJOR UNCERTAINTIES OF SHM-ORIENTED WSNS

There are a number of technical uncertainties or challenges that have been identified by prior studies (Lynch and Loh, 2006; Spencer *et al.*, 2004). However, from a perspective of SHM applications, two major and distinct WSN uncertainties that can directly degrade data quality are data loss and data synchronization error (Nagayama *et al.*, 2007). Brief review and discussion regarding these two factors are presented below.

Data loss has been seen as a serious problem for the generic WSN platform and resulted from unreliable wireless communications between sensor nodes (Nagayama *et al.*, 2007). In SHM-oriented WSNs, reliable communication protocol based on acknowledgement approach have been developed in middleware services so that lost data packets can be resent (Nagayama *et al.*, 2009). Wireless data transmission without loss is currently achievable though it has not been available in a real-time manner.

Data Synchronization Error (DSE) is another well-known uncertainty in WSNs which consists of two main components, namely initial DSE and jitter-induced DSE. Major sources of initial DSE include the timing offset among local clocks and the random delay in start time of sensing in sensor nodes (Nagayama *et al.*, 2009). Jitter-induced DSE is mainly due to (1) clock drift, (2) fluctuation in sampling frequency of each sensor node and (3) difference in sampling rate among sensor nodes. In the SHM-oriented WSN platform, there are a number of solutions in both hardware and software customization efforts to cope with DSE. Rice and Spencer (2009) customized a multi-metric sensor board named SHM-A in order to effectively mitigate the second and third sources of jitter-induced DSE. The first source of jitter-induced DSE, clock drift, can be effectively dealt with using clock drift compensation algorithm (Nagayama *et al.*, 2009). As a result, the remaining synchronization error for SHM-oriented Imote2 platform is mainly initial DSE which is random in range of a single sampling period (Linderman *et al.*, 2011). Even though a lower initial DSE can be further achieved with resampling algorithm (Nagayama *et al.*, 2009), this algorithm costs more computation effort at leaf nodes. Tolerance capacity of SHM applications with respect to relatively small DSE in SHM-oriented WSNs needs to be assessed in order to avoid unnecessarily computational burden.

There are limited studies that have investigated effects of DSE on SHM applications and almost all of them focused on effect of DSE on OMA techniques. The rationale for that is, as a global SHM approach, OMA generally requires data from different measurement points to be well-synchronized with each other (Nagayama *et al.*, 2007). It is worth noting that this requirement can be easily met in the traditional wired sensing system but not in case of WSNs with inherent synchronization errors. Nagayama *et al.* (2007) noted substantial effects of initial DSE on modal phases detected from simulation model by one parametric OMA method, whereas Krishnamurthy *et al.* (2008) observed considerable influence of initial DSE on mode shape magnitudes estimated by FDD in an experiment. Yan and Dyke (2010) confirmed effects of DSE on mode shapes in both simulation and experimental studies. Nguyen *et al.* (2014b) compared effects of different DSE levels on data collected from one real tower structure using the single sensor setup. Since previous research has mostly focused on simple structures such as cantilever and simply

supported beams or on the use of the single sensor setup, effects of initial DSE on OMA of civil structures in larger scales, which in many cases need to employ multi-setup tests, need to be further studied. Such effects on the most popular (but in different domains) OMA techniques (i.e. FDD and SSI-data) definitely deserve a comparative investigation in order to uncover their strength and weakness. For sake of completeness, FDD, SSI-data and associated strategies of data merging are briefly described in the next section.

4.3 OMA AND DATA MERGING METHODS

Representing non-parametric OMA is FDD, proposed by Brincker *et al.* (2000). This technique starts with estimation of output power spectral density matrices each of which (G_{yy}) corresponds to one of the discrete frequencies (ω_i) in the frequency range of interest. These matrices are then decomposed by the Singular Value Decomposition (SVD) algorithm as follows

$$G_{yy}(j\omega_i) = U_i S_i H_i^H \quad (4-1)$$

Where $U_i = [u_{i1}, u_{i2}, \dots, u_{im}]$ is a unitary matrix containing the singular vectors u_{ij} ; S_i is a diagonal matrix containing singular values s_i ; j and m are the index and total number of measured responses, respectively. Next, singular value lines are formed by assembling s_i for all discrete frequencies (ω_i) and plotted for implementing the peak-picking of modes. A mode is generally estimated as close as possible to the corresponding resonance peak of the first singular value line where the influence of the other modes is as small as possible. In the case of two orthogonally coupled modes occurring at one frequency, the previous step is carried out for the stronger mode whereas the peak on the second singular value line will be “picked” for the weaker mode (Structural Vibration Solutions A/S, 2011). Mode shapes are finally derived from singular vectors (u_{ij}) corresponding to selected frequencies. Besides FDD, there are two variants of this technique, i.e. Enhanced FDD and Curve-fit FDD but three techniques work similarly except the fact that estimation of damping ratios is only implemented in the two later ones. Similar to traditional input-output non-parametric techniques, FDD family is said to be fast, simple and user-friendly as well

as immune to computational modes (Zhang *et al.*, 2005). However, difficulties may arise in the case that dense and close modes are simultaneously present.

On the other hand, SSI-data has been considered as one of the most robust families of OMA time domain techniques since it can take into account furious modes from measurement noise; cope well with dense and closely spaced modes and avoid spectrum leakage (Zhang *et al.*, 2005; Brincker *et al.*, 2001). This OMA family relies on directly fitting parametric state space models to the measured responses of a linear and time invariant physical system (Overschee and Moor, 1996; Structural Vibration Solutions A/S, 2011) as follows.

$$x_{t+1} = Ax_t + w_t ; y_t = Cx_t + v_t \quad (4-2)$$

Here, x_t and y_t are the state vector and the response vector at time t , respectively. A is the system state matrix whereas C is the observation matrix. Amongst two stochastic processes, w_t is the process noise (i.e. the input) that drives the system dynamics whilst v_t is measurement noise of the system response.

In later phase, subspace models are first established for different dimensions up to the user-defined maximum value. Estimates of matrices A and C (i.e. \hat{A} and \hat{C} , respectively) are then obtained by the least square solution. By performing the eigenvalue decomposition of the system matrix estimate (\hat{A}), its discrete poles (μ_i) and eigenvectors (Ψ) can be found as follows (Brincker and Andersen, 2006):

$$\hat{A} = \Psi[\mu_i]\Psi^{-1} \quad (4-3)$$

The continuous time poles and subsequently modal frequencies and damping ratios are then obtained.

$$\lambda_i = \frac{\ln(\mu_i)}{\Delta t} \quad (4-4)$$

$$f_i = \frac{|\lambda_i|}{2\pi} \quad (4-5)$$

$$\zeta_i = \frac{\text{Re}(\lambda_i)}{|\lambda_i|} \quad (4-6)$$

Where Δt is the sampling period. Mode shape matrix is finally derived from the observation matrix and eigenvectors.

$$\Phi = \widehat{C}\Psi \quad (4-7)$$

By using increasing subspace model orders, multiple sets of modal parameters for each pole are obtained and their deviation can be used to examine whether the pole is as stable as a genuine structural mode. This leads to the extensive use of the stabilization diagram not only in SSI-data (see Figure 4-5 or Figure 4-6 for illustration) but also in most parametric modal analysis methods. It might be worth noting that there is another SSI family that is based on covariance of data and therefore named covariance-driven SSI (SSI-cov) but this OMA approach is likely to confront higher computational errors due to the issue of matrix squared up in its calculation process (Zhang *et al.*, 2005). Among different estimation algorithms for SSI-data (Structural Vibration Solutions A/S, 2011), Un-weighted Principal Component (UPC), has been used most for OMA of civil structures. Another advantage of the SSI-data techniques over the FDD family is that they have potential to be operated in the automate manner.

Besides the use of a single dataset, it is not unusual in practice, to merge data from multiple setups in both input-output and output-only modal analysis (Reynders *et al.*, 2009). Such a usage is able to cover a large number of measurement points using a limited number of sensors for the denser measurement which is always desirable in modal analysis, particularly for mode shape estimation. Multiple successive test setups are employed with a few sensors (known as reference sensors) being kept fixed while the others are being roved along the structure. A common problem with this usage is the inconsistency and non-stationary amongst different datasets (for instance, due to different operational and environmental conditions) which may cause estimation errors in the OMA process (Reynders *et al.*, 2009). Since DSE is a newer type of measurement uncertainty as previously mentioned, it is necessary that its impact be thoroughly investigated.

There are two most common ways of merging data in both SSI approaches in general and SSI-data family in particular from multiple tests, namely Post Separate Estimation Re-scaling (PoSER) and Pre Global Estimation Re-scaling (PreGER). By means of data of reference sensors, the former merges secondary data (i.e. mode shapes) estimated by SSI of all individual tests whilst the latter relies on merging the correlation of all primary sub-datasets (i.e. time series) into a unified set before performing SSI techniques (Dohler *et al.*, 2010). Compared to PoSER, the advantage of PreGER is that only one stabilization diagram needs to be dealt with in the identification phase regardless of number of the setups while the user of PoSER may need to work with every single diagram of each setup. However, PreGER is likely to be less robust with respect to small non-stationarities (Reynders *et al.*, 2009) which may be the case of DSE. The robustness of both methods and particularly PreGER with respect to DSE obviously deserves further investigation. For the sake of simplicity, SSI-data-UPC-PoSER and SSI-data-UPC-PreGER are hereafter shortened as UPC-PoSER and UPC-PreGER, respectively.

4.4 RESEARCH METHODOLOGY

As previously discussed, effects of common initial DSE on OMA approach especially on two most popular OMA techniques (i.e. FDD and SSI-data-UPC) need to be investigated more thoroughly on more complex structures with another realistic sensor arrangement strategy (i.e. employing multiple sensor setups). To realize this aim, a sophisticated and large-scale laboratory bridge model is selected for data acquisition with multiple successive tests using limited number of sensors. In order to have DSE-free data, the original data herein was collected by a precisely synchronized wired sensing system, before being contaminated with an additional amount of measurement noise to account for the higher presence of this factor on WSNs in comparison with the wired sensing system employed herein. Serving as benchmark (or DSE-free) data, the noise-added accelerations are then introduced with a relaxed level of initial DSE for SHM-oriented WSN platform in a random manner. To investigate impact of DSE randomness, this pollution process was repeated fifty times to generate fifty sets of DSE-corrupted data for subsequent analyses. Both DSE-free and DSE-corrupted data are used for OMA utilizing FDD

and two variants of SSI-data-UPC techniques, to identify modal frequencies, mode shapes and their changes with respect to DSE. Damping ratios are not under consideration of this study based as their estimation can be inaccurate in OMA approach and they are not among commonly-used damage indices for SHM (Brincker *et al.*, 2001). The use of projection channels is also explored to see whether it can mitigate DSE impact. The basis for this is that the impact of DSE is generally higher for higher modes (see section 6.3) which is similar to the impact of conventional measurement uncertainties such as measurement noise which can be handled by the use of the projection option. Figure 4-1 presents the flowchart of the investigation approach and further details can be found in section 5.

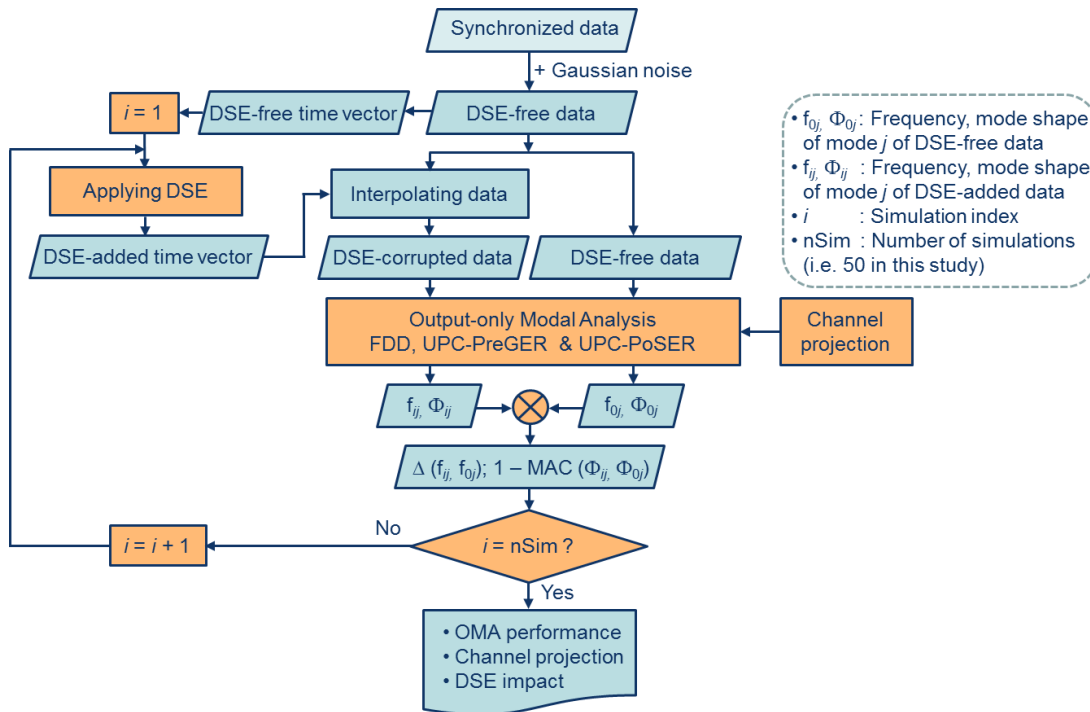


Figure 4-1 Flowchart of the investigation approach

4.5 BRIEF DESCRIPTION OF TESTS AND ANALYSIS

4.5.1 THE BRIDGE MODEL AND WIRED SENSING SYSTEM

Object for data acquisition is the through-truss bridge model at the Queensland University of Technology (Figure 4-2). With almost 600 degrees of freedom and dimensions of 8550mm by 900mm for its foot print and the height of 1800mm at the two towers, this bridge model can be one of the largest laboratory through-truss

bridge models for SHM purposes. To simulate ambient excitation, three large industrial fans were used at three different positions along the structure. Fan speed and direction were altered from one test to another to take into account changes of wind speed and wind direction in reality.

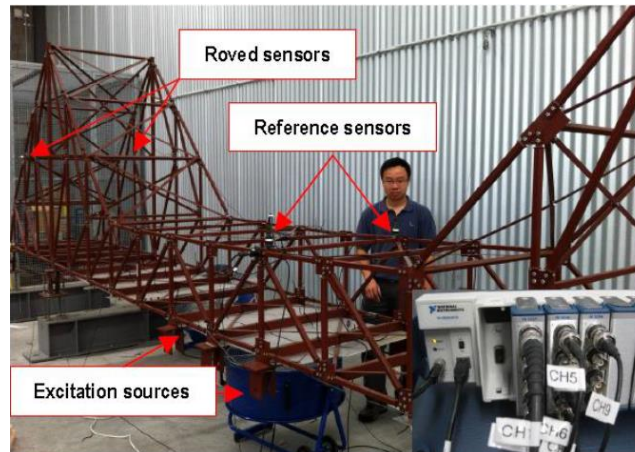


Figure 4-2 Physical bridge model and its wired sensing system

The bridge model was instrumented with nine high-quality uni-axial seismic ICP® accelerometers (www.pcb.com) with the sensitivity of 10 V/g. In each test, the sensors were divided into three groups each of which covers one cross section. This is based on the assumption of the cross section moving as a rigid body, the movement of one rectangular cross section can be described by three uni-axial accelerometers (Structural Vibration Solutions A/S, 2011). In each group, two accelerometers were used for vertical measurement and the other was to measure the lateral response. Of three groups, one was kept as the reference (i.e. near mid-span) and the other two were roved along the bridge model. Figure 4-3 illustrates two examples of the sensor setups. The total number of successive sensor setups was set at seven.

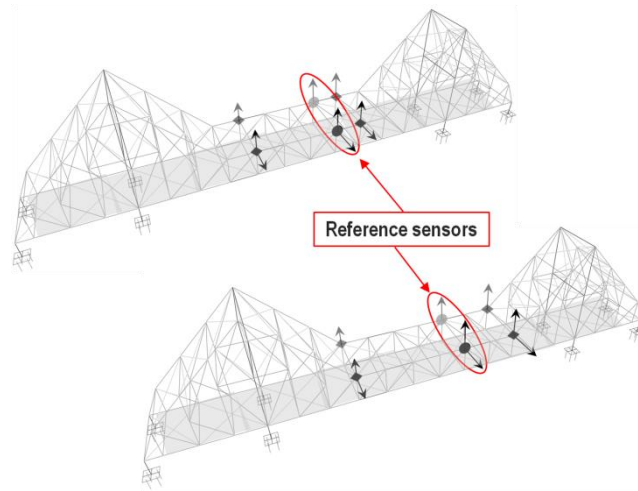


Figure 4-3 Two examples of the sensor setups

The sensing system was controlled by a National Instruments (NI) data acquisition system including NI cDAQ 9172 chassis, NI 9234 dynamic signal acquisition modules and LabVIEW Signal Express software (www.ni.com). To achieve precise synchronization, the internal timebase of one module is selected to be shared with the other modules so that all modules can use the same timebase in the sampling process. Sampling rate was set at relatively high value, i.e. 1766 Hz which allows the use of different decimation factors to achieve different lower sampling rates. For illustration purpose, the data used hereafter was obtained by decimating ten times the original data, therefore resulting in 176.6 Hz as the effective sampling rate. This effective rate can be considered belonging to a common range for practical SHM applications.

4.5.2 SIMULATION OF NOISE AND INITIAL DSE

All seven-subset data were added with relatively high level of Gaussian noise (i.e. 20 percent in root-mean-square sense) to account for the presence of higher noise in WSNs in comparison with the wired sensing system used herein. In this step, the MATLAB function named “randn” was utilized to create sequences of Gaussian distributed numbers with the specified root-mean-square values (MathWorks, 2011). Acting as the DSE-free source, each noise-added acceleration sequence is then contaminated with an initial DSE which was randomly assigned between zero and the effective sampling period in the simulation process. To do so, the DSE value was first added to the initial time vector of each time series to obtain the (DSE-induced) delayed time vector and based on these two time vectors, the DSE-corrupted data was

then derived from the DSE-free acceleration sequence by means of the MATLAB one-dimensional interpolation function named “interp1”. The “interp1” function has a number of options which are actually the methods of interpolation including popular ones such as linear or cubic spline interpolation methods. It is worth noting that the linear interpolation method has already been utilized in the re-sampling algorithm for SHM-oriented WSN middleware (Nagayama *et al.*, 2009) due to the fact that it requires less computational effort from sensor resources. Since the simulations herein are not subjected to such a computational constraint, the cubic spline interpolation method in the “interp1” function was adopted to achieve more accurate simulation results (MathWorks, 2011). This process was run fifty times to generate fifty DSE-corrupted datasets to facilitate statistical analyses.

4.5.3 OMA AND ANALYSES OF EFFECTS OF DSE

The DSE-free and fifty DSE-corrupted datasets were used as the input for FDD, UPC-PoSER and UPC-PreGER techniques. The analysis was conducted using ARTeMIS Extractor software (Structural Vibration Solutions A/S, 2011) with two options for channel projection as previously mentioned (i.e. enable and disable). It is worth noting that the use of channel projection is mainly recommended to the case that has many sensors. After several trials, the number of projection channels selected was four as they provided the best results. Also, the dimension for the state space model (i.e. the maximum model order) was set 180 as it was found to be sufficient for both UPC-PoSER and UPC-PreGER. In ARTeMIS Extractor software, UPC-PoSER is simply called UPC or Unweighted Principal Component whilst UPC-PreGER is known as UPC Merged Test Setups.

For each OMA technique, fifty sets of modal parameters (i.e. frequencies and mode shapes) were estimated at each mode and can be used to compare with the benchmark modal parameter set (i.e. from the DSE-free data). As this direct comparison is the same as level 1 of modal-based damage identification process (Brincker *et al.*, 2001), popular damage indices such as frequency changes and the deviation from unity of Modal Assurance Criterion (MAC) of mode shape pairs can be used as primary indicators for assessment of DSE impact. Interested readers could refer to Allemang (2003) for more details of the MAC index.

To evaluate changes of modal parameters with different bases like frequencies under DSE impact, basic statistical measures are employed including root-mean-square error (RMSE) of DSE-corrupted frequencies (with respect to the DSE-free frequency) and relative difference of DSE-corrupted frequency estimates. With MAC deviations which share the same base (i.e. zero), box-plot function (MathWorks, 2011) was adopted to visualize some useful statistical properties (such as median, quartiles and extremes) of MAC deviations at a number of first modes.

4.6 RESULTS AND DISCUSSIONS

4.6.1 COMMON RESULTS OF OMA FOR DSE-FREE DATA

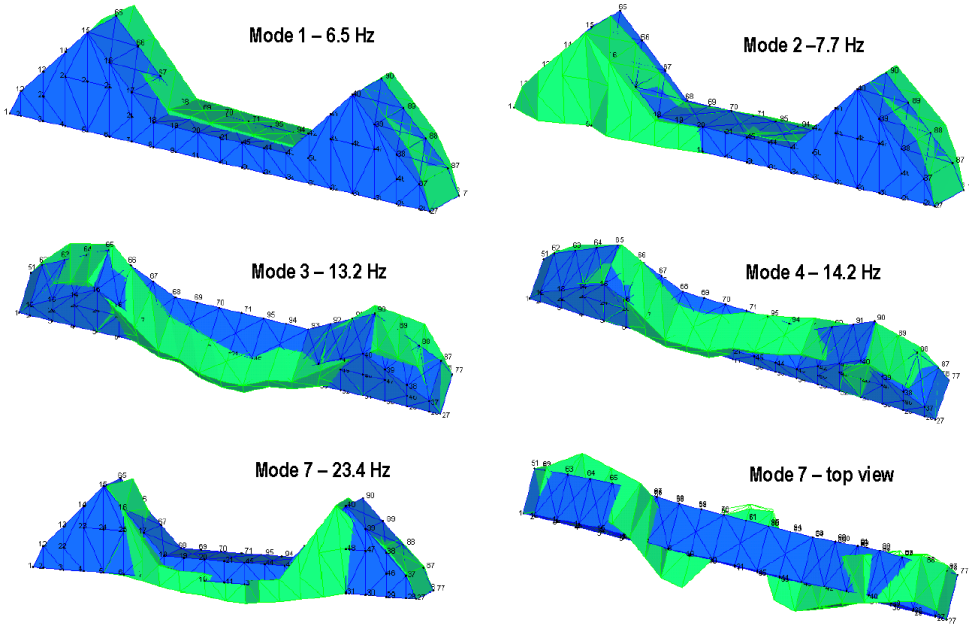


Figure 4-4 Typical mode shapes estimated from DSE-free data

The first four modes detected are purely (or almost purely) lateral modes, at around 6.5, 7.7, 13.2 and 14.2 Hz, respectively (Figure 4-4) whilst three higher modes detected (at around 18.2, 22.2 and 23.4 Hz) are mostly coupled ones between lateral and vertical responses. Figure 4-4 shows such a coupled mode (mode 7).

4.6.2 THE USE OF CHANNEL PROJECTION

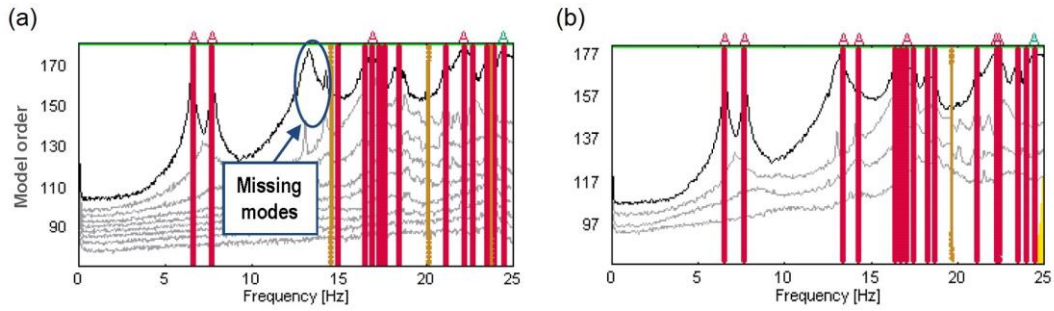


Figure 4-5 Stabilization diagram of UPC-PreGER with projection:
(a) disabled and (b) enabled

Of the three techniques, the channel projection has the most substantial influence on robustness of UPC-PreGER with respect to DSE presence. While the projection-disabled version of UPC-PreGER works properly with DSE-free data estimating all aforementioned modes, it completely fails detecting modes 3 and 4 from most of the fifty sets of DSE-corrupted data [Figure 4-5(a)] even though higher dimensions of the state space model were tried. However, the use of channel projection has enhanced UPC-PreGER so that these two modes can be estimated again in the projection-enabled version [Figure 4-5(b)]. Besides, the channel projection has also certain effect on the way UPC-PoSER copes with DSE. Some noise modes are mistakenly detected at locations of true modes [see Figure 4-6(a) for the case of mode 2] if the channel projection is not used. This problem is also resolved once the projection option is enabled [Figure 4-6(b)]. Obviously, channel projection is needed for two SSI-data-UPC techniques in order to effectively detect genuine modes under the presence of DSE. Impact of the projection option on performance of FDD technique is presented in the next section.

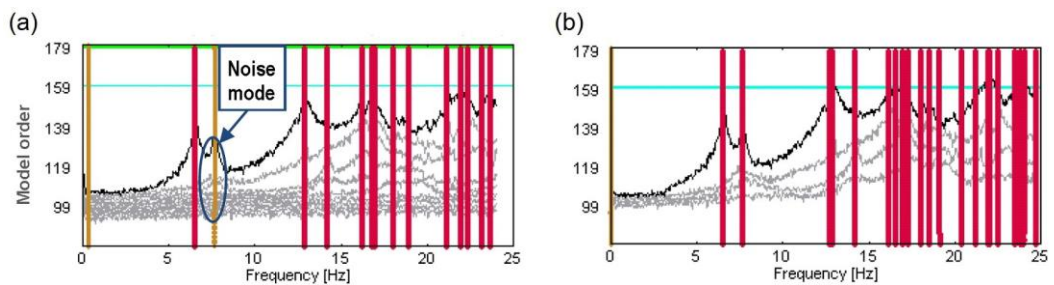


Figure 4-6 Stabilization diagram of UPC-PoSER with projection:
(a) disabled & (b) enabled

4.6.3 EFFECTS OF DSE ON OUTCOMES OF THREE OMA TECHNIQUES

The previous section has proven the necessity of applying the channel projection for UPC-PreGER and UPC-PoSER when DSE is present in the sensing system. Besides, it is necessary to examine whether FDD is under the same impact of the projection. Therefore, in each OMA round, the projection method was applied for two UPC-PreGER and UPC-PoSER whilst the robustness of FDD was also examined for both cases i.e. with the channel projection being enabled and disabled. The remaining of this section will present and discuss the results of DSE impact on estimates of frequencies and mode shapes.

There is no change in frequencies estimated by FDD for both DSE-free and DSE-corrupted data. This once again reinforces the prior findings that DSE does not affect frequencies estimated by FDD (Krishnamurthy *et al.*, 2008; Yan and Dyke, 2010) and highlights the robustness of this technique with respect to DSE impact on frequency estimation.

Table 4-1 Effects of DSE on frequency estimates by UPC-PoSER and UPC-PreGER

<i>Technique</i>	<i>Mode</i>	<i>DSE-free (Hz)</i>	<i>RMSE (mHz)</i>	<i>Min (Hz)</i>	<i>Max (Hz)</i>	<i>RD (%)</i>
UPC-PoSER	1	6.550	0.04	6.550	6.550	0.00
	2	7.700	0.15	7.700	7.701	0.01
	3	13.187	6.36	13.167	13.183	0.13
	4	14.271	2.32	14.272	14.278	0.04
	5	18.171	7.41	18.158	18.180	0.12
	6	22.207	1.29	22.203	22.208	0.02
	7	23.425	6.23	23.426	23.447	0.09
UPC-PreGER	1	6.547	19.33	6.573	6.616	0.66
	2	7.705	3.05	7.706	7.718	0.15
	3	13.215	89.66	13.290	13.643	2.67
	4	14.288	65.04	14.348	14.540	1.35
	5	18.210	70.63	18.304	18.601	1.63
	6	22.094	109.89	21.913	22.273	1.63
	7	23.433	101.45	23.339	23.676	1.44

[Note: Relative difference, RD = (Max-Min)/DSE-free]

Frequency estimates by SSI family are subjected to certain influence from DSE but the effects are fairly different for two SSI-data-UPC sub-techniques as illustrated in Table 4-1. Whilst UPC-PoSER experiences the maximum frequency RMSE of less than 0.01 Hz, that figure of UPC-PreGER can be as large as 0.1 Hz. Similarly, the

upper bound for relative frequency difference of the former technique is only 0.13 percent whereas that of the latter is up to 20 times larger.

Figure 4-7 shows the distribution of MAC deviations (from unity) of mode shapes estimated by FDD (with two options for the channel projection), UPC-PoSER and UPC-PreGER. Obviously, the results of these four cases can be seen to be classified into two groups. MAC indices of mode shapes estimated by FDD with two cases and UPC-PoSER mainly experience drops of less than 0.1 and their trend clearly show that DSE impact increases along with the increase in the mode order. However, those from UPC-PreGER can be as large as 0.3 or even higher for extreme cases and their trend is somewhat non-stationary at transitions between certain modes. These trends are reflected not only via median value of MAC deviations but also in general through the dispersion statistics such as the inter-quartile range (i.e. the height of the box in Figure 4-7).

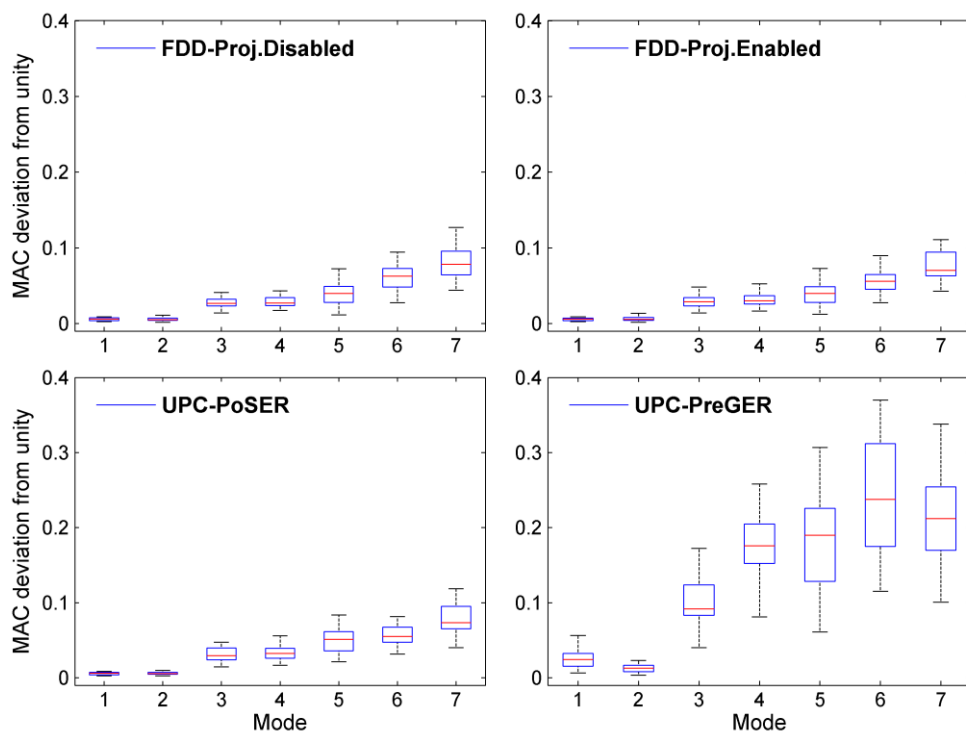


Figure 4-7 Box-plots of MAC deviations (from unity) of mode shapes for four cases

The results above show that FDD is the most robust technique among those studied herein with respect to DSE effects. Its frequency estimates stay unchanged under the impact of DSE regardless of whether the channel projection is applied or not. The

mode shape magnitudes estimated by this technique have also changed the least. It appears, with reasonable number of sensors like those used in this research, that FDD does not necessarily require the assistance from projection method even though a slight improvement in MAC values can be seen if the projection option is enabled. Amongst three techniques, UPC-PreGER is the worst possibly due to the fact that this technique merges the correlation of data before performing SSI and errors may be exaggerated during this merging phase. With the help of the channel projection, UPC-PoSER also overcomes negative impact of DSE on local sets of data and achieves considerable robustness to cope with this uncertainty.

It can also be seen from the above results that, impact of DSE on estimates of mode shapes generally increases with the order of modes which is similar to effects of measurement noise. One simple way to combat this negative influence is to limit number of modes of interest and this fact has become a fundamental axiom to achieve a feasible modal-based SHM solution in practice. MAC deviation (from unity) of around 0.05 at the sixth mode estimated by FDD or UPC-PoSER might be considered as an acceptable fluctuation threshold for monitoring of structural damage in real civil structures, see for instance (Brincker *et al.*, 2001).

4.7 CONCLUSIONS

This paper has presented an intensive investigation of effects of uncertainties of SHM-oriented WSNs on performance and outcome of several popular OMA techniques considering a frequent realistic application. Based on a brief review, the paper has first revealed that whilst data loss can be effectively treated using reliable communication protocols, DSE is still unavoidable and can be considered as the most inherent uncertainty. Since OMA has been identified as one of the SHM approaches possibly suffering the most from negative impact of DSE, effects of the updated DSE level on three most frequently-used OMA techniques have been investigated with respect to one of the common usages i.e. merging data from multiple tests. A combination of precisely synchronized experimental data of a large-scale laboratory structure, simulation of SHM-oriented WSN uncertainties including random noise and random DSE and commonly-used statistical tools such as the box-plot has been adopted to facilitate the assessment process. The results have first shown that the

impact of DSE on modal parameters (except frequencies estimated by FDD) tends to be more severe for higher-order modes and this trend is similar to conventional measurement uncertainties such as measurement noise. Of the three OMA techniques, FDD is the most robust technique possibly because it avoids working directly with time-domain data like the other two and impact of DSE at spectral peaks is the least. Without using channel projection, both variants of SSI-data (i.e. UPC-PoSER and UPC-PreGER) have been found to suffer from unreliable estimation of modal characteristics under disturbance of DSE. In this regard, the use of the channel projection has been proven to be able to enhance the performance of the two SSI-data variants to some extent. Nevertheless, the remaining impact of DSE on the outcome of UPC-PreGER is still considerable while that of UPC-PoSER is reduced to be more or less the same as the impact on the outcome of FDD. Since parametric and non-parametric OMA approaches have always been recommended to be used together to complement each other, the combination of both FDD and UPC-PoSER with the channel projection option has been shown to be effective and highly recommended for OMA of multi-setup datasets subjected to DSE such as those collected by WSN.

Chapter 5: Development of a Cost-effective and Flexible Sensing System for Long-term Continuous Vibration Monitoring

This chapter consists of the following manuscript submitted for review

- ✚ Nguyen, T., Chan, T. H. T., Thambiratnam, D. P. and King, L. Development of a cost-effective and flexible sensing system for long-term continuous vibration monitoring. (Under review)

The ultimate contribution of this chapter to the overall research program is a novel cost-effective and flexible realization of the Ethernet distributed DAQ platform for the purpose of continuous monitoring of large-scale civil structures. While common measurement and system development issues such as low-level ambient vibration, sparse sensing coverage and budget constraint as previously mentioned have called for innovations in this piece of research, the feasibility confirmation for the semi-complete data synchronization approach has provided the basis for the synchronization solution derived in this development. Utilizing the TCP/IP communication technology, this data synchronization solution can not only help to reduce the total cost but also provide greater flexibility for system development work in demanding circumstances such as large or sparse measurement coverage. Using this synchronization solution as well as other effective sensor and peripheral DAQ solutions, a cost-effective and flexible Ethernet distributed DAQ system is developed and implemented onto an actual building at QUT which has a (rather low) frequency range of interest similar to the majority of real-world civil structures. By means of the primary SSI-data technique coupled with the channel projection scheme (as suggested from Chapter 4), general assessment as well as statistical evaluation incorporated daisy data selection, the reliability of this sensing system is thoroughly verified. The daisy data selection scheme derived herein is to assist the statistical evaluation overcome the inherent impact of common E&O factors so that the effect

of random DSE on mode shapes can be more precisely evaluated. The vibration data acquired by the system is then used for the purpose of validating the damage identification method and related data generation scheme presented in Chapter 7.

STATEMENT OF JOINT AUTHORSHIP

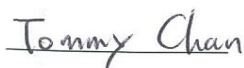
The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- There are no conflicts of interest;
- They agree to the use of the publication in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

Contributors:

- Mr. Theanh Nguyen (PhD Student): Conceived the paper idea; designed part of the data synchronization scheme; designed the sensor placement scheme and data analysis programs; implemented Matlab programming; and wrote the manuscript.
- Mr. Les King (Senior Instrumentation Engineer): Designed part of the data synchronization scheme; implemented LabVIEW programming; and provided critical comments for part of the manuscript.
- Prof. Tommy Chan (Principal Supervisor), Prof. David Thambiratnam (Associate Supervisor): Provided critical comments on the student's formulation of the concepts as well as on his experiment and data analysis programs; and provided editorial comments to enable the student to improve the presentation quality of the manuscript.

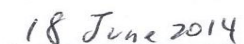
Principal Supervisor Confirmation: I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.



Name



Signature



Date

ABSTRACT

In the Structural Health Monitoring (SHM) field, long-term continuous monitoring is becoming increasingly popular as this could keep track of the health status of structures during their service lives. However, implementing such a system is not always feasible due to numerous constraints. Of the most common ones, the compromise between the cost and the reliability of a system is a significant challenge especially when the measurement coverage is large or sparse. To address this issue, this paper presents the development of a cost-effective and flexible sensing system for continuous vibration monitoring of a newly-built institutional complex with a focus on the deployment on its main building. At first, the most appropriate accelerometers and their optimal positions are selected in order to overcome adversities such as low-frequency and low-level vibration measurement. In order to accommodate sparse measurement points, a cost-optimized distributed data acquisition model is adopted to provide the skeleton for the sensing system. Instead of using rather costly turnkey hardware-based synchronization modules, the synchronization task is left open for possible software-based solutions. As one of such solutions, a combination of a high-resolution timing coordination method and a periodic system resynchronization strategy both based on the TCP/IP communication technology is proposed to synchronize data from multiple peripheral data acquisition units. The results of both general and statistical evaluations show that the proposed sensing and data synchronization solutions work truly well and can provide a promising cost-effective and flexible alternative for use in the SHM projects with tight budget and/or sparse system coverage in which the conventional synchronization method might be unreasonable. Using these solutions, the sensing system developed herein is shown to be able to provide stable and useful feature databank which can be used to construct unbiased safety evaluation processes for civil structures in practice. With such an effective and flexible sensing system, the instrumented building can act as a multi-purpose benchmark structure for vibration-based SHM problems in general as well as for addressing system synchronization issues in particular.

KEYWORDS

SHM, Continuous Vibration Monitoring, Cost-effective, Sensor Solution, Distributed Data Acquisition, Data Synchronization

5.1 INTRODUCTION

In the Structural Health Monitoring (SHM) field, long-term continuous monitoring of structural vibration is becoming more popular in recent years. This is owing to the fact that this type of systems will enable to keep track of the genuine health status of real structures under the disturbance of Environmental and Operational (E&O) factors (Ko and Ni, 2005; Farrar and Worden, 2013). As a result, the SHM community has seen an increasing number of bridges and buildings around the world equipped with such monitoring systems (Ansari, 2005; Ko and Ni, 2005; Karbhari and Ansari, 2009; Chan *et al.*, 2011). Nevertheless, implementing such a monitoring system is not always feasible due to numerous difficulties of which the budget constraints and the requirement for a reliable system are the most common ones among others (Karbhari and Ansari, 2009; Farrar and Worden, 2013). These problems often become worse when the measurement coverage is large or sparse. This is particularly true for the conventional Data Acquisition (DAQ) system where every sensor needs to be cabled to one single centralized DAQ station leading to massive cost of the cable itself and the installation labor. To overcome these issues, the recent trend in SHM is to use the distributed DAQ system platform in which each system consists of a network of peripheral DAQ units and their sensors (Aktan *et al.*, 2003; Van Der Auweraer and Peeters, 2003). The DAQ units are linked to each other and to the base station via either radio signal (wireless system) or Ethernet cable (Ethernet-based system). Since each local DAQ unit only needs to be in charge of a small group of adjacent sensors, this type of system greatly mitigates the burden of wiring directly from every sensor to the host station and avoids the problem of signal noise induced by lengthy cabling (Aktan *et al.*, 2003). In this regard, wireless sensor networks especially those belonging to the SHM-oriented platform could be considered as the first promising candidate. This type of sensors can completely eliminate the demand for cabling among units and therefore be more affordable while suffice for even demanding global SHM applications (to real civil structures) such as

modal analysis (Rice and Spencer, 2009; Nguyen *et al.*, 2014b). However to date, successful applications reported for this type of systems have mostly been for short-term monitoring projects (Rice and Spencer, 2009; Cho *et al.*, 2010). The usage of wireless sensor systems for long-term monitoring is still uncertain due to inherent difficulties such as constraints of power supply and bandwidth; and lower reliability of each node and the whole system particularly over long periods of time (Karbhari and Ansari, 2009; Farrar and Worden, 2013; Xu and Xia, 2012). On the other hand, the Ethernet-based sensing systems, though well-known for their reliability, durability and therefore being more suited for long-term monitoring purposes, may still be expensive for most of SHM projects especially from the return on investment viewpoint (Ansari, 2005; Karbhari and Ansari, 2009). More reasonable solutions at sensor and DAQ levels would make this type of systems more appealing so that they can be more widely applied on civil infrastructure particularly in permanent monitoring basis to provide higher safety for society.

To address this issue, this paper presents the development of a cost-effective and flexible Ethernet-based sensing system for long-term continuous vibration monitoring of a newly constructed complex with a focus on the deployment on the main building. The greatest challenges in developing this system are that the budget is tight while vibration sensors are distantly located as well as operated in challenging ambient vibration conditions. To overcome such difficulties, multi-layer solutions are derived and implemented. At the lowest system level, the most appropriate accelerometers and their optimal positions are first selected so that they could overcome common adversities in ambient vibration monitoring of civil infrastructure such as low-frequency and low-level vibration measurement. A cost-optimized distributed DAQ model is then selected to form the system skeleton so that each group of nearby measurement channels can be handled by one local DAQ unit within a network. Instead of using costly hardware-based synchronization methods as normally seen in the traditional approach, the synchronization task for the building vibration sensors is left open for possible inexpensive software-based solutions. As one of the first solutions in the proposed direction, a combination of a customized timing coordination method and a system resynchronization strategy both based on the TCP/IP communication technology is proposed to synchronize data from multiple

distributed DAQ units. To facilitate accurate assessments of this initial (as well as any future synchronization) solution, both robust general and statistical evaluation methods are derived and intensively applied to assess the quality of the data acquired by the realized system. The outcomes of these evaluations show that the proposed sensing and data synchronization solutions work very well and the corresponding system is able to provide stable and useful feature databank to enable unbiased pattern recognition processes for evaluating the health status of civil infrastructure.

The layout of this paper is as follows. The second section first provides an overview of the instrumented complex and its monitoring systems before detailing the distributed DAQ system solution. The third section presents two solutions - one for selection and placement of vibration sensors and the other for synchronization of data from multiple DAQ units on the main building. Intensive general and statistical validations as well as applications employing the pattern recognition framework are provided in section 4 before section 5 summarizes and concludes the research work.

5.2 OVERVIEW OF THE INSTRUMENTED COMPLEX AND MONITORING SYSTEMS

The instrumented complex is the newly-constructed Science and Engineering Centre (SEC) at the Gardens Point campus of Queensland University of Technology (QUT), Australia. Costing around AUD 230 million, this complex has achieved 5-star Green Star rating from the Green Building Council of Australia making it one of the highest rated green buildings in Brisbane City. Besides being a main teaching and research facility, SEC is also notable for its giant digital lab named “the Cube” (www.thecube.qut.edu.au). With a two-story interactive digital learning and display screen and many other digital science spaces, the Cube is dedicated to be a dynamic hub of hands-on scientific exploration for the QUT community and the wider public. More interestingly, the real-time data from the monitoring systems will be viewed on the Cube screens once all the monitoring systems are fully deployed.



Figure 5-1 Science and Engineering Centre (SEC) at QUT Gardens Point campus

In structural details, the SEC complex comprises two 10-level buildings (named Y and P, respectively from left to right in Figure 5-1) and various functional spaces underneath the atrium connecting these two buildings. As the main building of the complex, P block houses the Cube and is the main site for deployment of three major monitoring systems namely (1) vibration system; (2) structural system; and (3) subsurface system. Besides P block, part of the vibration system is also deployed on the reinforced concrete footbridge located in the front corridor linking the main entrances of the two buildings. This is also the deployment location of two Acoustic Emission (AE) sensors which, due to their proximity, are able to share the DAQ unit with all the footbridge vibration sensors. Whilst accelerometers are the sole sensor type in the vibration system, the structural and subsurface systems have mostly employed dynamic strain transducers, vibrating wire and soil pressure sensors. Beside the main vibration system for the global vibration monitoring purpose, there are also two minor vibration monitoring points also within the P block. Of these two points, the first one targets the local monitoring of a slab underneath a transmission electron microscope at level 6 whereas the other is for ground acceleration (seismic) monitoring and referencing purpose at the ground floor. Further usage descriptions of these two accelerometers as well as the other sensors of the AE, structural and

subsurface systems are beyond the scope of the present paper and may be reported elsewhere in future.

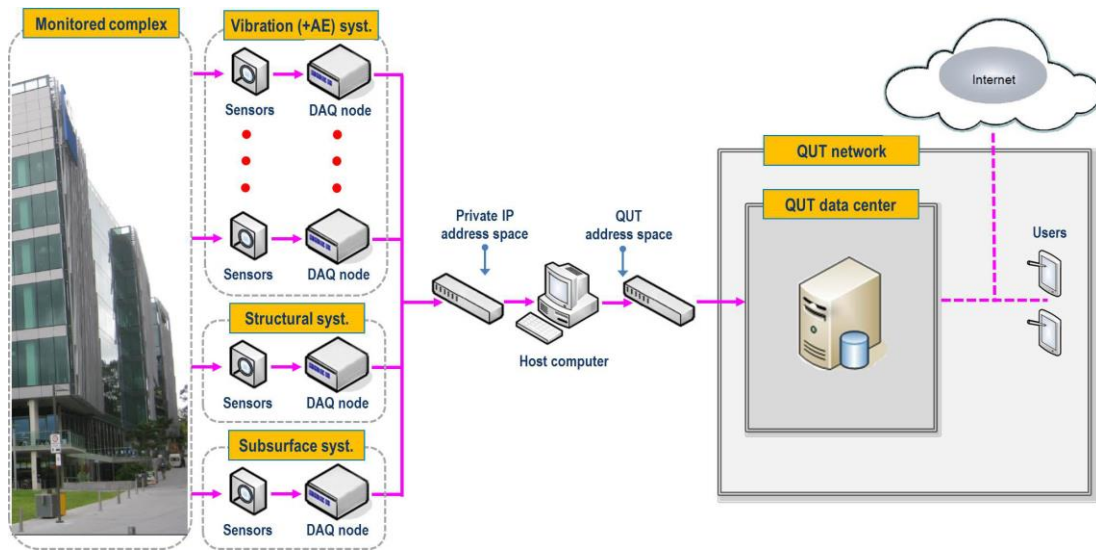


Figure 5-2 General diagram of the SEC monitoring systems

With all the sensors in general and the vibration sensors in particular broadly and distantly located across structures and levels, the distributed DAQ architecture, as illustrated on the left of Figure 5-2, is found the most applicable. Well-known for their compact-size, stability and cost-effectiveness, the reconfigurable embedded control and acquisition system platform CRIO provided National Instruments (www.ni.com/crio) in general and the cost-optimized CRIO-907x series in particular is specially considered as a potential candidate for the DAQ skeleton for all the sensing systems herein. The finally selected model (CRIO-9074) is one of three models (besides CRIO-9072 and CRIO-9073) of the CRIO-907x series available during the design phase of this project. The CRIO-9074 is more advantageous than the other two models due to its capacity of chassis expansion as it has another Ethernet port (beside the one exclusively for network connectivity) that can be used for such a purpose. Last but not least, the adoption of the same DAQ model for all systems in this instrumentation project ensures that the tasks of system deployment, programming, operational management and future maintenance can be all simplified.

5.3 VIBRATION SENSOR AND DATA SYNCHRONIZATION SOLUTIONS

5.3.1 VIBRATION SENSOR SOLUTION

For the main vibration system, 8 tri-axial and 4 single-axis analog accelerometers has been pre-allocated for use to monitor global vibration characteristics of the P block and the footbridge. To connect these sensors to the CRIO-9074 chassis, the four-channel analog input module (NI-9239) is used to provide high-resolution (24-bit) analog-digital converter, anti-aliasing filter and peripheral sampling control. As the main target for the system is to be serviceable in ambient vibration context, the main characteristics of the accelerometers such as sensor type, measurement range and sensitivity are the most critical factors among others. Manufactured by Silicon Designs, Inc., the accelerometers finalized are all capacitive-type with a reasonable measurement range ($\pm 2g$) and high sensitivity for this sensor type (2000mV/g). The capacitive type is selected as this type of sensors is more applicable for low-frequency (as small as around 1 Hz) and low-level (i.e. as small as 1–2 mg) vibration measurement than other types such as piezoelectric sensors (Jo *et al.*, 2012; Karbhari and Ansari, 2009; Xu and Xia, 2012). With the same concern as the latter, selection of high-sensitivity accelerometers is also prioritized in order to enhance the quality of low-amplitude vibration response data (Jo *et al.*, 2012; Karbhari and Ansari, 2009; Xu and Xia, 2012). This low-level vibration measurement issue is very likely to occur in the case of the two instrumented structures herein. This is because the P block, with a limited height of around 40 m against its typical (upper) floor dimensions of 45 m by 63.5 m, is no doubt a non-slender building. Spanning merely over a distance of 8.4 m with the slab thickness of 375 mm can make the concrete-type footbridge herein to be categorized as a relatively inflexible planar structure. Such building and footbridge structures tend to be marginally excited by ambient excitation and therefore need higher attention in selection of the sensors in order to enable their small motions to be properly recorded.

With such a limited number of sensors available for use, sensor positioning must be carefully designed in order to obtain sufficient modal information (of each target structure) while being able to keep number of sensors used on each component/portion at minimum. First of all, both of these two structures should have

tri-axial sensors so that more types of modes (e.g. in different directions or coupled ones) can be measured. Besides, in planar-type structures such as the footbridge herein, the tri-axial sensors should be placed in the positions that are the most sensitive to the modes that may be undetected by single-axis sensors such as lateral or torsional modes. Towards this goal, the footbridge is first allocated two tri-axial accelerometers to be positioned on middle of the two unsupported edges as shown in Figure 5-3 (left). Additionally, two single-axis sensors are placed to measure the vertical motion at a quarters and three-quarters of the span along the longitudinally central line of this footbridge. The motion of the middle point of this central line can be interpolated from motions captured by the two tri-axial sensors based on the assumption of the cross-section moving as a rigid body.

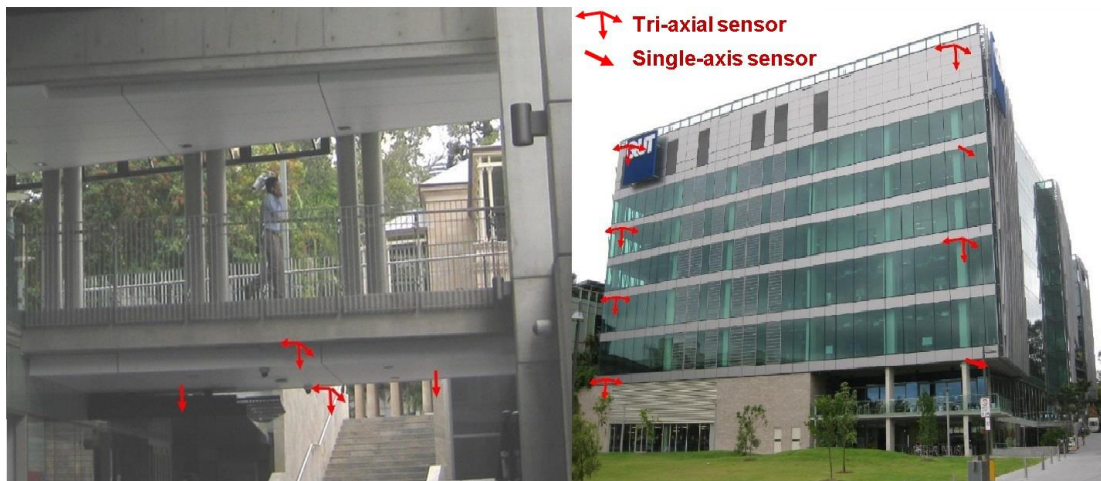


Figure 5-3 Sensor position on the footbridge (left) and the building (right)

On the other hand, the P block has a rather common level configuration for the building structures of its type with a rather wide semi-underground base consisting of the lowest four levels (see Figure 5-6 for illustration of the base). It is therefore sensible to anticipate this part of the building would not be well excited in the horizontal plane under ambient excitation conditions. In this regard, it is apparent that the remaining upper six levels are horizontally more sensitive to ambient excitation and should be the locations for most of the vibration sensors where the signal-to-noise ratio of acceleration data will be enhanced. Note that in global vibration monitoring of buildings the vertical vibration measurement is often of little interest due to the excitation problem. Further, since each stiff (reinforced concrete)

floor of the building can be considered as a rigid body, the horizontal movement of each floor level can be described via two horizontal displacements and one deformation angle (Structural Vibration Solutions A/S, 2011). Hence, a minimum number of two sensors can be used at two adjacent corners of the floor and one of these sensors can be in sing-axis type. Consequently, four floor levels can be covered with eight remaining accelerometers leading to the building sensor placement solution as illustrated in Figure 5-3 (right). Unmeasured horizontal degrees of freedom for mode shape animation purposes will be interpolated from measured ones.

5.3.2 DATA SYNCHRONIZATION SOLUTION

Being located within proximity of each other, all footbridge sensors only required one DAQ unit to be shared and consequently, these sensors can be precisely synchronized without requiring any additional hardware. This synchronization is realized by sharing the master timebase of any input module with the others in the same chassis. In contrast, as the building vibration sensors are distant from each other, one DAQ node needs to be allocated for each of these sensors to effectively reduce the cable length and enhance the data quality as previously discussed. The remaining issue is how to properly synchronize multiple peripheral DAQ nodes deployed on the building. Large DSE has been shown to negatively affect global vibration-based SHM applications such as modal analysis and associated damage identification (Nguyen *et al.*, 2014b; Krishnamurthy *et al.*, 2008). The main hardware-based multi-chassis synchronization option during the system design phase was the use of a digital I/O module in every local DAQ unit in order to form a dedicated synchronization bus. A synchronization trigger pulse can be then generated from the master device and passed to each of the slave devices (Semancik, 2004; National Instruments, 2012c). However, as this option was prohibitively expensive for the system coverage required herein, the desire for inexpensive and flexible alternatives has triggered the promotion of alternative software-based solutions. The main bases for this direction are first that the main impact of DSE tends to be proportional to the DSE magnitude and the modal frequency value (Krishnamurthy *et al.*, 2008; Nguyen *et al.*, 2014b, 2014c). Second, the frequencies of interest for civil structures where the synchronization task often becomes more

problematic are rather low and, for many cases (including the case considered herein), less than 10 Hz. Third, high-resolution timing coordination is feasible with Ethernet connection (Semancik, 2004). It is therefore feasible in sensing system development for such structures to use software-based semi-complete data synchronization solutions to achieve a reasonable DSE level so that the impact of this level can be negligible (Nguyen *et al.*, 2014b). As one of the first solutions in this direction, the combination of a customized timing coordination method and a periodic system resynchronization strategy will be presented below.

The manner in which a distributed DAQ system operates without a conventional cable-based synchronization bus is as follows. The peripheral DAQ units start their own data sampling process asynchronously based on the sampling command each unit receives from the host computer via TCP/IP communication. Upon receiving this command, the Field-Programmable Gate Arrays (FPGA) chip in each unit immediately initiates its analog input module to acquire data based on the internal master timebase of the module. These result in inherent DSE with two main components, namely initial DSE and accumulated DSE, similar to those from wireless sensors (Nguyen *et al.*, 2014b). Whilst initial DSE is induced from the difference in the start (sampling) times of multiple DAQ units, accumulated DSE is caused by the inherent difference between multiple local timebases of multiple input modules. Besides controlling the starting times of the individual sampling processes, it is therefore also necessary to resynchronize the system after certain duration of time before the accumulated DSE could become significant to spoil the recorded data.

In order to combat both DSE components, two sub-solutions are derived and implemented in the system programming environment (LabVIEW). First, data acquired by each DAQ node is timestamped based on the time the sampling command is supposed to be received at the FPGA chip of the unit. This command delivery time is calculated by the time the sampling command is sent from the host computer plus the duration for the command message to travel from the host computer to the DAQ node. The latter component is actually half of the ping time (obtained in each ping test) which measures the round-trip travel time of the command message and can be well estimated by utilizing statistical measures (such

as the mean and standard deviation) from a sufficiently large number of ping tests (Wikipedia contributors, 2001). To achieve high accuracy, the advanced hrPing utility with the time resolution of microseconds is used instead of the basic windows ping utility with limited resolution of milliseconds (cFos Software GmbH, 2013). This way, the initial DSE has been significantly reduced and has found to be less than 0.1 millisecond (which is equivalent to one fifth of the sampling period); and now mostly due to small fluctuations of the message delivery times at different DAQ nodes. In order to cope with DSE accumulation, the system is programmed to reinitialize the sampling process after a predetermined duration to cut off the growth of DSE. This duration should at least equate the uninterrupted data length required by intended applications such as modal analysis (see the next section for details). Finally, the data streams from different DAQ nodes are aligned with each other by matching their timestamps before other pre-processing tasks such as data decimation can take place.

5.4 EXPERIMENTAL EVALUATIONS

As the main focus of this study, only experimental evaluations of the sensor and data synchronization solutions for the P block are reported in this section. Analysis results of the footbridge are currently underway to be presented in a separate paper. Besides, as the building has experienced some major alterations during the construction phase, the finite element model used earlier in the design phase requires an intensive revision before it can represent the actual structure. This revision task is expected to start shortly and the results of analytical-experimental correlations and model updating will be published in a future work.

5.4.1 GENERAL EVALUATION OF THE SENSING SYSTEM

Owing to the delta-sigma modulation of the input module, possible sampling rates can be derived through the formula of $50/n$ kilo samples per second (kS/sec) where the denominator (n) can be any integer between 1 and 31 (National Instruments, 2012b). To cater for convenient choices of decimation factors (in post-processing phases), the primary sampling rate is selected to be 2 kS/sec (i.e. $n = 25$) which is close to the minimum sampling rate value. Next, the data length for an uninterrupted

acquisition process between two consecutive synchronization points needs to be decided. The commonly-used rule of thumb for this length in ambient modal testing is that it should be around 1000–2000 times the fundamental vibration period (Structural Vibration Solutions A/S, 2011; Cantieni, 2005). Based on preliminary spectral analyses, the first natural vibration period of P block is shown up to be around 0.9 sec and the uninterrupted data acquisition length should therefore be 900–1800 sec. Hence, the resynchronization span is set at 30 minutes and the data streams acquired by multiple DAQ nodes are timestamped and synchronized as previously described. Besides, such a low value of the fundamental frequency (of around 1.1 Hz) has confirmed that the assumption of low-frequency measurement in the instrumentation design phase has become true. A similar confirmation can also be made for the low-level measurement assumption as one thirds of horizontal measurement channels have been found to experience the (peak-to peak) amplitudes of around 2 mg in normal excitation conditions. For the illustration purpose, Figure 5-4 shows a 10-minute time history of such a typical channel in level 8. Overall, the confirmations for both earlier assumptions mean that the adopted accelerometers are appropriate to the vibration characteristics of the intended structure.

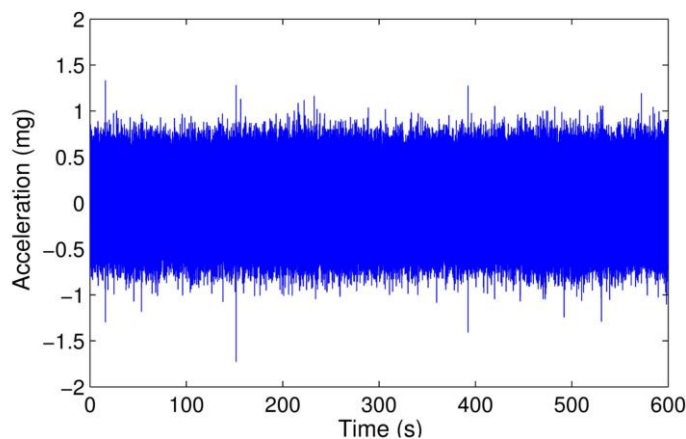


Figure 5-4 Typical low-level acceleration time history

To obtain global vibration characteristics of the P block under ambient excitation conditions, the Output-only Modal Analysis (OMA) approach is used (Structural Vibration Solutions A/S, 2011). For the purpose of modal animation and validation, the building is modelled, as illustrated in Figure 5-6, at measured levels plus its lowest (ground floor) level that is deemed to have negligible displacement. Amongst

different OMA algorithms, the primary technique of the data-driven Stochastic Subspace Identification (SSI-data) family [i.e. SSI-data employing Unweighted Principal Component (UPC) estimator] is selected as the main OMA technique (Structural Vibration Solutions A/S, 2011). This is because this technique has been shown to be as robust against the impact of initial DSE as the well-known Frequency Domain Decomposition (FDD) technique (Nguyen *et al.*, 2014b). Further, in comparison with FDD, SSI-data techniques are more advantageous in coping with closely spaced or repeated modes as well as in the implementation of automated modal identification. These can facilitate rapid and accurate OMA which is particularly meaningful for continuous monitoring generally with large quantity of datasets. Nevertheless, FDD is still utilized as the secondary OMA technique for cross-check purposes particularly with the initial configuration setup stage for primary SSI-data.

Theoretically, SSI-data relies on directly fitting parametric state space models to the measured responses of the structure with model orders being varied up to a user-defined maximum dimension (Overschee and Moor, 1996; Structural Vibration Solutions A/S, 2011). To reduce computational effort and noise impact associated with the use of a large number of measurement channels, channel projection is often applied (Herlufsen *et al.*, 2005; Structural Vibration Solutions A/S, 2011). By varying the model order, multiple sets of each modal parameter at a pole are obtained and their deviations are used to examine whether the pole is as stable as to represent a genuine structural mode. This leads to the extensive use of the stabilization diagram in SSI-data to facilitate the task of automated modal identification (see Figure 5-5 for illustration). It is also worth noting from this figure that, the background wallpaper of the diagram consists of the spectral density singular value plots the peaks of which can be used to correlate with the pole locations determined by SSI-data for cross-checking purposes.

Back to this study, via comparing the results of SSI-data of incremental dimensions and projection channels, the most stable range of maximum dimensions is found to be between 120 and 200 whereas that of the projection is from 8 to 10 channels. Hence, the maximum state space dimension of 160 and the projection of 9 channels are selected as the common SSI-data configuration for all vibration datasets of the P

block used in this paper. Using this SSI-data configuration, a total of first seven modes can be estimated via the stabilization diagram and validated using mode shape animation as typically illustrated in Figure 5-5 and Figure 5-6, respectively. Further details of the estimated modes are provided in Table 5-1. Note that the values of frequencies and damping ratios in this table are the mean values obtained from a number of datasets which are mainly for use for statistical assessment of synchronization solutions in the next section. Of the seven modes, most are reasonably well excited (i.e. corresponding well to the spectral peaks) with two of them (i.e. modes 2 and 3) being rather closely spaced. In addition, the most weakly-excited one (mode 6) though not always present can still be identified in a certain number of datasets. Such capacities and the achievement of clear mode shape animation views (Figure 5-6) for all the seven modes can be attributed for the suitability of not only the sensors and their positions but also the combined synchronization solution in general and the system resynchronization strategy in particular. For correlation purposes, it might be worth comparing the OMA results of the P block herein with similar cases in literature. Towards this point, the two 15-story buildings to be cited are an office building in Japan and the well-known Heritage Court Tower (HTC) in Canada which has been considered as a benchmark structure for ambient vibration testing (Horyna and Venture, 2000; Tamura *et al.*, 2002). By using popular OMA techniques such as FDD or SSI-data, 9 modes from the first building were identified whilst this figure for the HTC building was from 9 to 11 depending on the OMA technique in use. Since the P block has only 10 stories with a high percentage of base levels (35% total height vs. around 25% of the HTC building case), the number of estimated modes (i.e. 6–7) for this building can be seen as to be well reasonable against both quoted reference results.

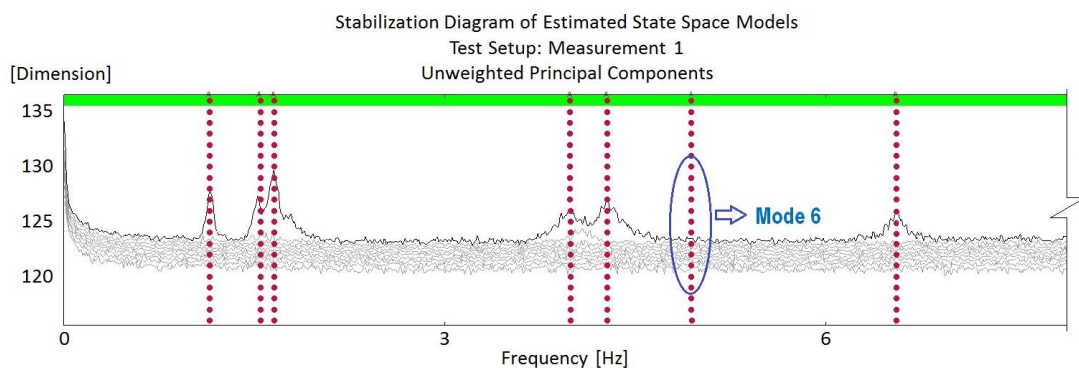


Figure 5-5 A typical SSI-data stabilization diagram for OMA of the building

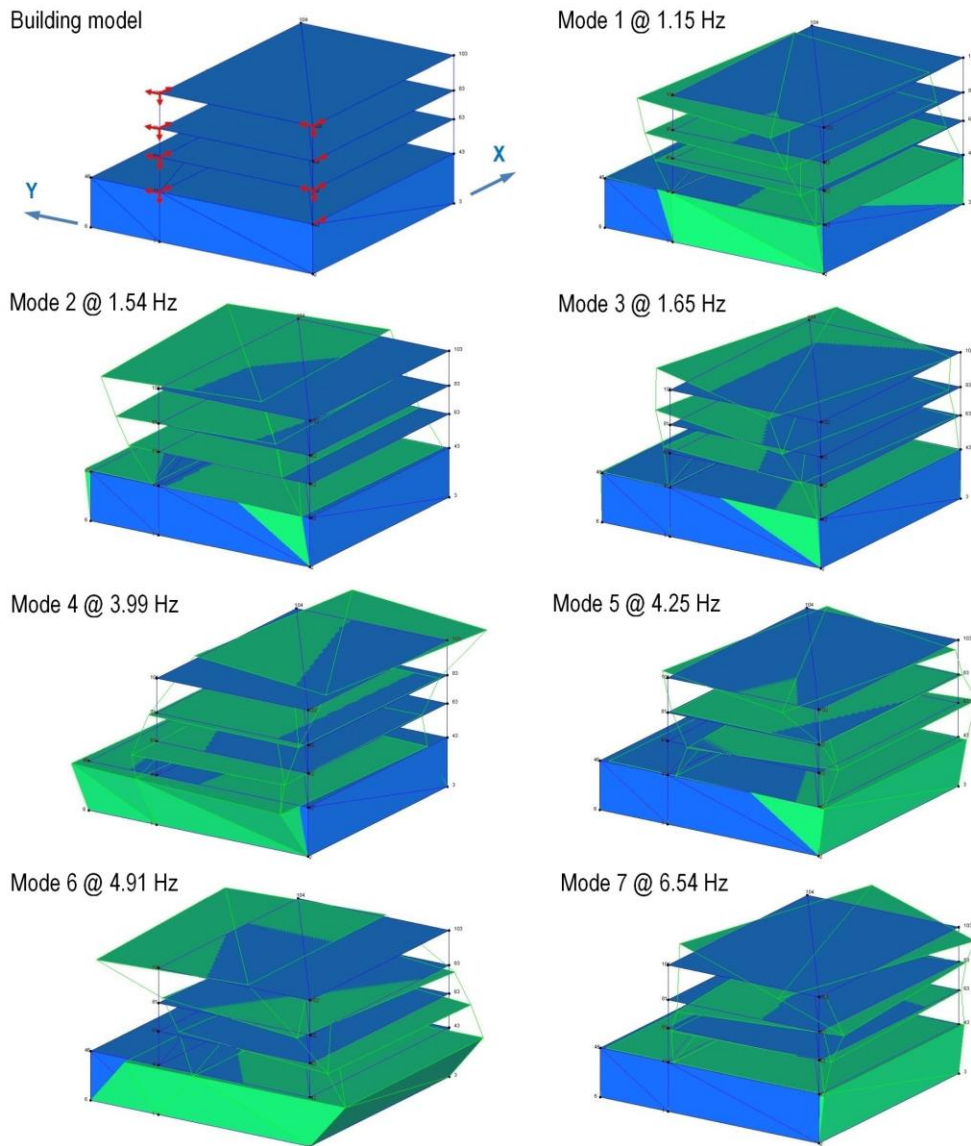


Figure 5-6 Building model and typical animation views of seven estimated modes

Table 5-1 Features of seven estimated modes

Mode	Description of modes	Frequency (Hz)	Damping ratio
1	1 st translational – X direction	1.15	2.03
2	1 st translational – Y direction	1.54	3.59
3	1 st torsional	1.65	1.88
4	2 nd translational – X direction	3.99	1.98
5	2 nd torsional	4.25	2.08
6	2 nd translational – Y direction	4.91	5.47
7	3 rd torsional	6.54	2.85

5.4.2 STATISTICAL EVALUATION OF THE DATA SYNCHRONIZATION SOLUTION

To evaluate the actual efficacy of the synchronization solution for long-term purposes, repetition checks need to be carried out on multiple datasets to examine the impact of remaining initial DSE that is random in its nature. Since mode shapes have been shown to be significantly more sensitive than frequencies and damping ratios with respect to initial DSE in prior studies (Nguyen *et al.*, 2014b; Nagayama *et al.*, 2007), it is necessary to assess the impact of initial DSE on the former parameter. The most common method for this is to track changes of Modal Assurance Criterion [MAC, (Allemang, 2003)] of pairs of mode shape datasets recorded upon different synchronization spans. The challenge for this type of assessment on real structures is that the modal parameters in general and mode shapes in particular are influenced not only by special DAQ uncertainties such as DSE herein but also from common E&O factors such as temperature, wind conditions or human-induced activities. The impact of the latter factors especially temperature has been considered as one of the most significant obstacles against the success of SHM in practice for the purpose of detecting critical changes such as structural damage (Farrar and Worden, 2013). To overcome this problem, a daisy chain data selection scheme is derived in this study for computation of MAC data for the aforementioned tracking purpose. To do so, MAC values are strictly calculated from any (two) consecutive datasets in rather uniform E&O conditions. Such selections of data and E&O conditions are to ensure that the time and meteorological spans are so short that the paired datasets are deemed to be subjected to the similar E&O influences. Any unusual shift in MAC values could therefore be attributed to the initial DSE.

Applying the above data selection scheme, 50 MAC vectors can be calculated from a total number of 64 reasonable datasets obtained in various days during the system development phase in late 2013. As it is occasionally under-excited during the testing days, mode 6 is excluded from this assessment. Figure 5-7 shows the distribution of MAC data for the remaining six (regularly-excited) modes in box-plot, a useful graphical tool for presenting robust statistics (Nguyen *et al.*, 2014b). There are a few outliers (maximum 3 out of 50 observations) at some modes but this has no

significant impact on the overall result. In fact, one can clearly see that most of the mode-shape sets are in excellent agreement (with each other) with their MAC values above 0.95 and narrow (statistical) dispersion of around 0.02 or less. The lower figures of MAC at modes 2 or 7 can be attributed to the nature of these two modes i.e. being more damped than the other modes. In contrast, the highest agreement at mode 3 is mainly due to the fact that this is the least damped mode in the detected range. These are reflected via the magnitudes of the damping ratios in Table 5-1. Nevertheless, such overall high and stable MAC values have proved that the remaining initial DSE is insignificant and the proposed synchronization solution has worked very well.

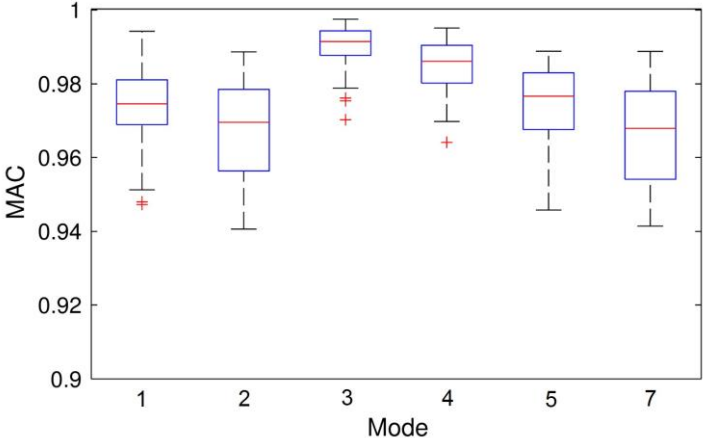


Figure 5-7 Statistics of mode shape agreement across multiple datasets

5.4.3 CONTINUOUS STRUCTURAL SAFETY EVALUATION

Previous statistical evaluation has shown that the vibration monitoring system can operate stably and provide useful and reliable vibration data for the long-term purpose. This would lead to the establishment of representative databases for every definite period of time (e.g. annually) during the structural service life to enable continuous (or at least frequent) evaluation of structural safety. In order to deal with the inherent impact of E&O factors, the evaluation problem should be formulated in the pattern recognition framework by means of machine learning algorithms (Farrar and Worden, 2013). The use of such an algorithm is to “learn” the underlying trends (caused by E&O factors) present in the training data so that these trends can be separated from the symptom of any potential structural anomaly or novelty (such as

damage) present in the testing data. The consequence of this is that the testing result can accurately inform whether the structure has still been in its normal state or not. For illustration purposes, two commonly-used unsupervised learning algorithms, that is, based on Mahalanobis Squared Distance (MSD) and Auto-Associative Neural Network (AANN) are employed for the purpose of damage identification. While MSD-based algorithm obtains the knowledge of E&O impact trends mainly through the inclusion of the sample covariance matrix in its computational structure (Appendix5.A), AANN mainly does this at the bottleneck layer where the trends induced by E&O factors are forced to be prevailing (Appendix5.B). Note that since only the unsupervised learning algorithm are used, the damage identification problem herein is restricted to level 1 of the identification hierarchy that is to identify the presence of a possible damaged state. To implement the damage identification process, the training data used in this section is the frequency data corresponding to the six frequently-excited modes of the 64 aforementioned datasets. Testing data is from 36 datasets collected more recently (in early 2014). The confidence level used for both learning algorithms is set at a commonly-used level of 99 % while a simple AANN architecture is used with the number of nodes at both outer hidden (i.e. mapping and demapping) layers equating to the number of feature variables. At the bottleneck layer, two hidden nodes are used as they are assumed to represent two most dominant E&O factors namely temperature and structural mass (Farrar and Worden, 2013). As P block is a main institutional building for teaching and working purposes, the change of the latter factor is supposed to be due to the occupants of this block entering or vacating the building.

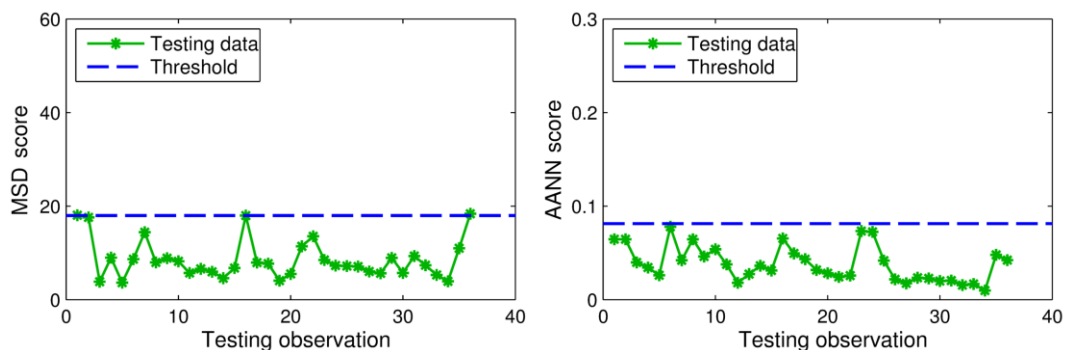


Figure 5-8 Level-1 damage identification results: MSD-based method (left) and AANN-based method (right)

After computation, all 64 training distance scores are ranked and the distance threshold is chosen among the largest scores using the selected 99 % confidence level. The threshold can then be used to compare against the 36 testing distance scores to infer the status of the structure. Figure 5-8 shows the level-1 damage identification results by MSD-based and AANN-based methods. All AANN scores and almost all of MSD ones lie below the threshold meaning that all the testing observations are likely to be under normal conditions. This can be confirmed by the good agreement between the magnitudes of training and testing data as reflected by their box-plots in Figure 5-9. A few extreme MSD scores (slightly larger than the threshold) might be related to the problem of insufficient multinormality in the training dataset as this set has had rather limited (i.e. 64) observations at the current stage. Such a problem has been shown to be able to cause unstable MSD computational outcome (Nguyen *et al.*, 2014a). While this problem tends to be overcome in later monitoring stages when more measured data is provided by the sensing system, one possible immediate solution is the use of the controlled Monte Carlo data generation scheme to enhance the training data multinormality degree and therefore robust MSD computation (Nguyen *et al.*, 2014a). This will be further addressed in the future work of the present authors.

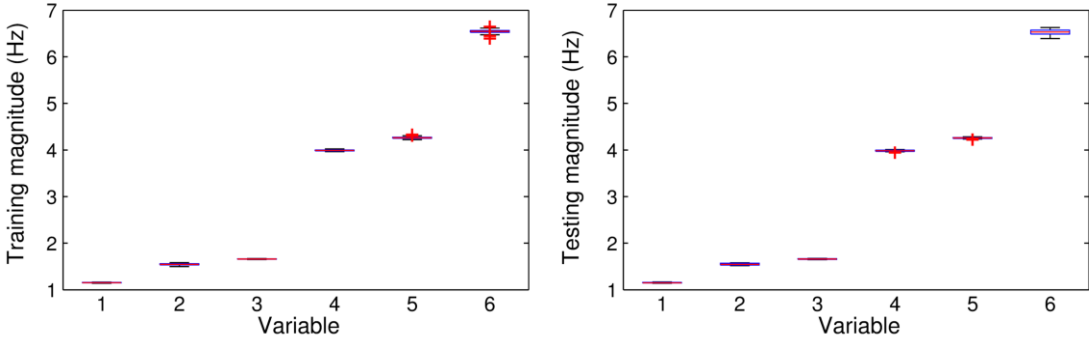


Figure 5-9 Box-plots of training data (left) and testing data (right)

5.5 CONCLUSIONS

This paper has presented the detailed development of a cost-effective and flexible sensing system for long-term continuous vibration monitoring of a newly constructed complex in general and its main building in particular. Under the challenging

characteristics of the monitored structures and the monitoring conditions, accelerometers of capacitive type and high sensitivity have first been selected so that they could function well under practical difficulties such as low-frequency and low-level vibration measurement. Sensor positioning has then been optimized to overcome the problem of the sensor quantity constraint and maximize the structural information acquired. To tackle the challenge of sparse measurement coverage, the distributed DAQ architecture has been adopted and a cost-optimized distributed DAQ model has been selected to establish the skeleton for the monitoring system. In search of inexpensive and flexible alternatives, the synchronization task for the building vibration sensors has been left open for possible software-based solutions. To prove the feasibility of this direction, a combined solution has been derived by means of high-resolution timing coordination and periodic system resynchronization both enforced from the host level via TCP/IP communication medium. By applying this solution, the initial DSE of the system has been well reduced from uncontrolled levels to be very marginal whereas the accumulated DSE can be effectively kept at minimum. By means of a robust OMA technique and a novel daisy chain data selection scheme, general and statistical evaluation methods have been derived to rigorously assess the sensing and data synchronization solutions while continuous health evaluation processes have been constructed in light of the pattern recognition framework. The evaluation results have showed that the sensing and data synchronization solutions work truly well and can provide a promising alternative for use in the SHM projects with tight budget and/or sparse measurement coverage where conventional hardware-based synchronization may be too costly. Using these solutions, the developed sensing system has been shown to be able to provide quality feature databases which can be used to combat the impact of practical E&O factors and establish unbiased pattern recognition processes for health evaluation of civil infrastructure. With such an effective and flexible sensing system, the instrumented building herein can be used as a flexible benchmark structure for vibration-based SHM problems in general and for addressing system synchronization issues in particular. For upcoming future work, building vibration data is being continuously collected and analyzed under different E&O conditions to construct representative databases for tracking the health status of the building or its deterioration process. Other potential synchronization methods can be rapidly applied to examine their

efficacy while multiple classes of data with different levels of DSE can be generated (by relaxing the synchronization process) for related uncertainty assessment studies.

APPENDIX 5.A MSD-BASED LEARNING FOR LEVEL-1 DAMAGE IDENTIFICATION

Suppose that the training dataset has p variables and a sufficient multivariate normal (multinormal) distribution. This dataset can therefore be represented by its sample mean vector (\bar{x}) and sample covariance matrix (S). By means of MSD technique, each p -variate observation (x_i) in either training or testing phases can be transformed into a scalar in the form of distance (or novelty) measure as follows.

$$d_i = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x}) \quad (5.A)$$

After computing all training distances, the assumption of a multinormal distribution again allows the estimation of the novelty threshold from the basis of chi-square distribution for the training distances (Farrar and Worden, 2013). It is because under such an assumption, one can specify a statistical threshold based on a distribution quantile or a confidence level (Farrar and Worden, 2013; Nguyen *et al.*, 2014a). In the testing phase, whenever a new feature observation of the structure is recorded, its corresponding distance can then be used to compare against the threshold to determine whether it corresponds to a normal condition or a novelty (such as a damaged state in the SHM context). In spite of being selective in multinormality degree of the input data, MSD-based learning algorithm is well-known for its architectural simplicity and computational efficiency which are advantageous for dealing with large volume of data (Nguyen *et al.*, 2014a).

APPENDIX 5.B AANN-BASED LEARNING FOR LEVEL-1 DAMAGE IDENTIFICATION

AANN is a multilayer feed-forward perceptron network which is trained to produce, at the output layer, the patterns that are presented at the input layer (Chan *et al.*, 2011; Farrar and Worden, 2013). The network contains three hidden layers: the mapping layer, the bottleneck layer and the demapping layer. The mapping and demapping layers consist of neurons with hyperbolic tangent sigmoid transfer functions while the bottleneck and output layers are formed by linear neurons. Typically, both mapping and demapping layers often use the same number of nodes while the bottleneck layer uses fewer nodes so that this special layer can

discard trivial variations and extract the predominant trends such as those induced by E&O factors. Also because of this, the number of hidden nodes at the bottleneck layer is often selected in order to represent such prevalent trends in the data.

Again, suppose the training dataset has p variables, after the network is trained, each p -variate input observation (x_i) in either training or testing phases is passed into the trained network to yield an (p -variate) observation (\hat{x}_i) at the network output layer. The corresponding observation of the novelty index in form of the Euclidean distance can therefore be as follows (Farrar and Worden, 2013).

$$d_i = \|x_i - \hat{x}_i\| \quad (5.B)$$

The tasks of threshold determination (based on confidence level) and novelty detection of AANN-based algorithm are then similar to those of MSD-based algorithm (Appendix5.A). Advantages of AANN-based algorithm include the capacities of dealing with data that may not have a multinormal distribution and recognizing nonlinear underlying trends. However, AANN-based algorithm has more complicated architecture than the MSD-based counterpart and therefore generally requires some user judgment and costs more computational effort in training and testing processes.

Chapter 6: Controlled Monte Carlo Data Generation for Statistical Damage Identification Employing MSD

This chapter is made up of the following published journal paper

- ✚ Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014. Controlled Monte Carlo data generation for statistical damage identification employing Mahalanobis squared distance. *Structural Health Monitoring* 13 (4):461-472; URL: <http://dx.doi.org/10.1177/1475921714521270>

While the three previous chapters (Chapters 3, 4 and 5) focus on enhancement and new development of the vibration sensing and feature extraction issues, the contribution of this chapter is associated with the improvement of the reliability of the damage identification function in the second subsystem of the targeted synthetic SHM system. Specifically, this chapter is to develop a data generation scheme to assist the MSD-based damage identification method in such adverse circumstances as at an early monitoring stage or during short SHM programs. Realizing the disadvantages of the basic Monte Carlo data generation method, an enhanced scheme named CMCDG is derived to achieve optimal data generation configurations in a systematic way. Not only are theoretical bases of CMCDG uncovered but the efficacy of this scheme is also intensively validated with a sophisticated SHM benchmark dataset collected from a laboratory structural model with multiple levels of damage. Extension of this work for the ultimate purpose of field application and validation is presented in Chapter 7.

Readers interested in implementing MSD-based damage identification employing CMCDG in the context of the safety evaluation system proposed in this thesis may refer to Appendix A for detailed implementation procedure and Appendix B for fundamentals of Monte Carlo simulation and basic data generation methods.

STATEMENT OF JOINT AUTHORSHIP

The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- There are no conflicts of interest
- They agree to the use of the publications in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.


Contributors:

- Mr. Theanh Nguyen (PhD Student): Conceived the paper ideas; designed and conducted experiments (for those conducted at QUT); analyzed data; wrote the manuscripts; and addressed reviewers comments (for published or in-press papers) to improve the quality of paper.
- Prof. Tommy Chan (Principal Supervisor), Prof. David Thambiratnam (Associate Supervisor): Provided critical comments on the student's formulation of the concepts as well as on his experiment and data analysis programs; and provided editorial comments to enable the student to improve the presentation quality of the manuscripts and the revisions for the published and in-press papers.

Principal Supervisor Confirmation: I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Tommy Chan

Name



Signature

18 June 2014

Date

ABSTRACT

The use of Mahalanobis Squared Distance (MSD) based novelty detection in statistical damage identification has become increasingly popular in recent years. The merit of the MSD-based method is that it is simple and requires low computational effort to enable the use of a higher-dimensional damage sensitive feature which is generally more sensitive to structural changes. MSD-based damage identification is also believed to be one of the most suitable methods for modern sensing systems such as wireless sensors. Although possessing such advantages, this method is rather strict with the input requirement as it assumes the training data to be multivariate normal which is not always available particularly at an early monitoring stage. As a consequence it may result in an ill-conditioned training model with erroneous novelty detection and damage identification outcomes. To date, there appears to be no study on how to systematically cope with such practical issues especially in the context of a statistical damage identification problem. To address this need, this paper proposes a controlled data generation scheme which is based upon the Monte Carlo simulation methodology with the addition of several controlling and evaluation tools to assess the condition of output data. By evaluating the convergence of the data condition indices, the proposed scheme is able to determine the optimal setups for the data generation process and subsequently avoid unnecessarily excessive data. The efficacy of this scheme is demonstrated via applications to a benchmark structure data in the field.

KEYWORDS

Statistical Damage Identification, Mahalanobis Squared Distance (MSD), Novelty Detection, Multivariate Normal (Multinormal), Data Generation, Monte Carlo, Data Condition Assessment

6.1 INTRODUCTION

It is well-known that Environmental and Operational (E&O) variations can prevent genuine structural damage in real civil structures from being identified since their effects can be larger than those from the genuine structural damage (Farrar *et al.*, 2001b; Sohn *et al.*, 2003). One of the most popular approaches to deal with this,

especially when measures of E&O variations are not fully available, is based on statistical pattern recognition. In this case, machine learning algorithms are oftentimes used to learn the underlying trend induced by E&O variations and create a robust damage index which can be considered to be invariant under the E&O variation presence. Amongst different methods in this approach, Mahalanobis Squared Distance (MSD) based damage identification is believed to be one of the best in unsupervised learning mode i.e. only using data from undamaged structures (Manson *et al.*, 2003; Figueiredo *et al.*, 2011). In this regard, one will simply turn MSD-based (multivariate) outlier analysis into a novelty detection method and attempt to identify a potentially damaged observation as an outlier (Farrar and Worden, 2013; Worden *et al.*, 2000a). Well-known for its simplicity and computational efficiency, MSD-based method has good potential to be cooperated on embedded modern sensing systems such as wireless sensors (Figueiredo *et al.*, 2009; Figueiredo *et al.*, 2011). However, the proper use of the standard MSD for the novelty detection purpose theoretically requires the training data needs to be multivariate normal (short as multinormal) or also known as multi-Gaussian (Farrar and Worden, 2013; Filzmoser *et al.*, 2005). Due to the unavailability of complete multinormal data in many practical applications, one can obtain an approximation by increasing the observation-to-variable ratio (Johnson and Wichern, 2002; Rencher, 2002). In practical structural monitoring, however, this is not always experimentally available particularly at an early monitoring stage. To systematically cope with such an adverse situation, this paper present a controlled data generation scheme which is based upon the Monte Carlo simulation methodology cooperated with several controlling and evaluation tools to assess the output data condition. By evaluating the convergence of the data condition indices, the proposed data generation scheme is able to determine the optimal simulation input parameters that need to be used and subsequently avoid improper simulation setups or unnecessarily excessive data. The efficacy of this scheme is demonstrated via applications to benchmark experimental data in the field. The layout of this paper is as follows. The next section provides descriptions of MSD-based damage identification and the Controlled Monte Carlo Data Generation (CMCDG) scheme. The benchmarks and their dataset used in this study are then briefly described. In the two last sections, detailed analyses and

discussions are first provided before the key findings are summarised in the conclusion.

6.2 DAMAGE IDENTIFICATION AND DATA GENERATION METHODS

6.2.1 MSD-BASED DAMAGE IDENTIFICATION

There are two main types of data used in statistical damage identification process. In general, the primary (or raw) data acquired by sensors is not directly used but is transformed into a damage-sensitive feature which then become input data for the statistical training model. This secondary data is oftentimes in a much lower dimension compared to the primary one so as to alleviate the computational effort and to extract the most meaningful structural information. Typical examples for this can be found in the case of common features such as modal parameters and Auto-Regressive (AR) vectors (Farrar *et al.*, 2001a; Worden and Manson, 2007; Zhang, 2007; Figueiredo *et al.*, 2009; Gul and Catbas, 2009; Worden *et al.*, 2002; Sohn *et al.*, 2001).

Suppose that a training dataset consists of p (i.e. feature dimension) variables and n observations. If its shape approximates a multinormal distribution, this dataset can be represented by the sample mean vector (\bar{x}) and the sample covariance matrix (S). In this case, these two parameters are often referred as “sufficient statistics”. By using the standard MSD technique as a multivariate outlier analysis (Worden *et al.*, 2000a), each feature vector (x_i) for either the training or testing purposes will be converted into a damage index in terms of distance measure (d_i) as follows.

$$d_i = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x}) \quad (6-1)$$

In damage identification context, the mean and covariance should be formulated as an exclusive measure, or in other words, consisting of no potential outlier from the testing phase (Worden *et al.*, 2000a). After computing all training distances, the assumption of a multinormal distribution again allows the estimation of the threshold from the basis of chi-square distribution for the training distances (Farrar and Worden, 2013). It is because under such an assumption, one can specify a statistical threshold for the distances based on a distribution quantile or equivalently a

confidence level (Filzmoser *et al.*, 2005; Farrar and Worden, 2013). There might be a trade-off in choosing the confidence level: using very high level of confidence level might not be able to detect a lightly damaged case that is known as one class of Type II errors but the least critical. However, such confidence level can assist in avoiding as many as possible false-positive indication of damage (i.e. Type I errors) (Farrar and Worden, 2013).

In the testing phase, whenever a new feature observation comes, its corresponding distance can be used to compare against the threshold to determine whether it corresponds to a normal or damaged state. In this sense, the anticipation is that the more severe a damaged state is, the more significant the difference between its actual distance and the threshold becomes. This has been observed in prior studies in this area (Worden *et al.*, 2000a; Gul and Catbas, 2009; Figueiredo *et al.*, 2011).

As seen earlier, even though the MSD-based damage identification possesses a simple computational structure, the success of this method depends on whether its assumption of data distribution (i.e. multinormal) can be adequately satisfied. Since complete multinormal data is seldom available in practice, the overall remedy, stemming from the Central Limit Theorem (CLT) and the Law of Large Numbers (LLN), is to increase the observation size (n) relative to number of variables (p) (Johnson and Wichern, 2002; Rencher, 2002). One simple and inexpensive approach to realize this remedy in the context of measured data shortage is using the CMCDG scheme.

6.2.2 CMCDG

As previously mentioned, the controlled data generation scheme developed in this paper originates from the Monte Carlo simulation methodology. In a broad sense, a Monte Carlo method today refers to any simulation method that involves the use of random numbers and was termed by Neumann and Ulam in the 1940's (Martinez and Martinez, 2002; Wikipedia contributors, 2002). Being easy and inexpensive, this approach is particularly applicable for evaluation of highly multidimensional and complex problems (Dunn and Shultis, 2011). To conduct a Monte Carlo simulation, one just needs to define a model that represents the population or phenomenon of interest and a criterion to generate random numbers for the model. The latter

commonly involves the use of a user-selected probability distribution. Once completed, the data generated from the model can then be used as though they were actual observations.

In the damage identification context, Monte Carlo data generation has also seen its applicability since the features are often in high dimension. However, prior studies in the field have mainly applied the Monte Carlo simulation methodology in an ad hoc manner. The conventional trend in such studies was to generate large number of observations from the data seed of a single or few feature(s) by applying certain amount of random Gaussian noise onto each copy (Worden *et al.*, 2000a; Worden *et al.*, 2002). Even though the noise was constructed from a Gaussian distribution, its magnitude and the sample quantity were generally set in a rather uncontrolled manner. Another general suggestion from prior research is that using lower levels of noise allows more lightly damaged cases to be detected (Worden *et al.*, 2007). However, a possible problem for applying a too low level of noise in data generation is that subsequently generated observations might not be sufficiently random with respect to initial observations to improve the data condition (and this issue will be examined in the application section). Obviously, a more systematic data generation scheme is in need particularly when considering real structural monitoring circumstances with a certain number of observations initially available to form the seed. Such a type of seed apparently reflects more accurately the training conditions of structures but also requires a more thorough data generation scheme to be cooperated.

To cater to this need, the present paper proposes an enhanced data generation scheme termed as CMCDG. This is realized by adding into the conventional scheme two controlling tools that are in fact two data condition assessment methods and a robust probability-based evaluation procedure to assist these methods. Of the two condition assessment methods, the first one is based on evaluating the condition of the generated data through the condition of its sample covariance matrix which is represented by a well-known and robust index, i.e. (2-norm) condition number (COND) in linear algebra (Golub and Van Loan, 1996; Strang, 2006). On the other hand, the second method is based on one of the most popular graphical tools for evaluating multinormality of data i.e. the Quantile-Quantile (Q-Q) plot of a beta

distribution or, in certain cases, a chi-square distribution (Johnson and Wichern, 2002; Rencher, 2002; Sharma, 1995). In this study, the beta Q-Q plot is employed since it is generally more accurate than the chi-square counterpart (Rencher, 2002). To evaluate multinormality of a dataset, the actual plot of data is compared with the theoretical one and a significant discrepancy in the plot would indicate that the data no longer belongs to a multinormal distribution. Since the number of datasets generated by CMCDG for statistical evaluations is large, the Root-Mean-Square Error (RMSE), one of the most commonly-used discrepancy measures, between the theoretical and actual Q-Q plots will be used as another condition index. The mathematical expression of this measure will be included in the application section. The rationale of employing these two methods to evaluate CMCDG process is as follows. First, under the regulation of CLT and LLN, the sample covariance matrix (S) converges in probability to the actual population covariance matrix (Σ) as number of random observations (n) increases (Johnson and Wichern, 2002). It is therefore sensible to anticipate that, as n increases, $COND(S)$ also converges in probability to $COND(\Sigma)$. Similarity can be seen for the second method. As n increases, the Q-Q plot is expected to converge in probability to the theoretical line and its RMSE is therefore anticipated to converge in probability to zero.

Inherent in the way that the two data condition assessment methods is implemented in CMCDG is a robust probability-based evaluation procedure with two robust measures i.e. the median and Inter-Quartile Range (IQR) (Martinez and Martinez, 2005) to examine the central tendency and dispersion of $COND$ and beta Q-Q RMSE. By tracking the convergence of these measures, CMCDG is able to determine the optimal noise level and possibly minimum number of data replications that need to be set in the simulation process. Details of CMCDG and its controlling and evaluation components are illustrated in the application section.

6.3 DESCRIPTION OF THE BENCHMARK STRUCTURES AND DATA

The benchmark dataset used in this study is from Los Alamos National Laboratory (LANL), USA and has been intensively used in recent statistical damage identification studies (Figueiredo *et al.*, 2009; Figueiredo *et al.*, 2011). This data was collected by four accelerometers from a benchmark building model (Figure 6-1) with

varied practical conditions (Table 6-1) including stiffness deviation due to temperature change and mass difference (e.g. caused by traffic). Nonlinear damage was generated by contacting a suspended column with a bumper mounted on the floor below to simulating fatigue crack that can open and close under loading conditions, or loose connections in structures. Different levels of damage were created by adjusting the gap between the column and the bumper. In total, there were 9 undamaged states and 8 damaged states each of which consists of a number of tests performed to take into account excitation variability. In this study, the largest dataset available for public use with 50 tests for each state is used (SHMTTools Development Team, 2010). According to the test description (Figueiredo *et al.*, 2009), state 14 can be considered as the most severe one since it corresponds to the smallest gap case which induces the highest impact of contact. State 10 is the least severe damaged scenario whereas state 11, 12 and 13 can represent mid-level damage scenarios. Other states (i.e. 15, 16 and 17) are the variant states of either state 10 or 13 with mass added effect.

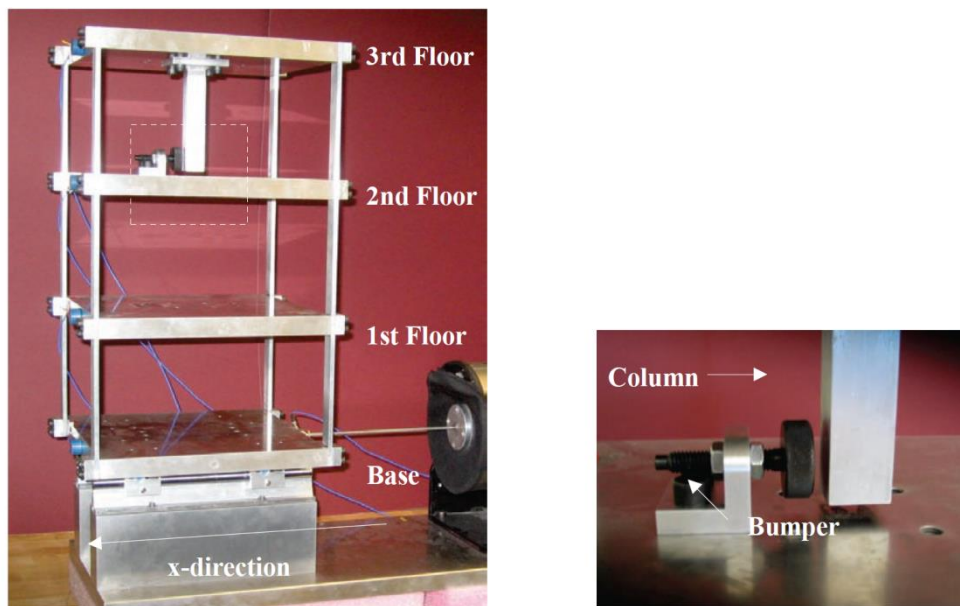


Figure 6-1 The test structure (left) and damage simulation mechanism (right) at LANL (Figueiredo *et al.*, 2009)

Table 6-1 Data labels of the structural state conditions
 [adapted from (Figueiredo *et al.*, 2009)]

<i>Label</i>	<i>Feature</i>	<i>Description</i>
State 1	Undamaged	Baseline condition
State 2	Undamaged	Mass = 1.2 kg added at the base
State 3	Undamaged	Mass = 1.2 kg added on the 1 st floor
State 4	Undamaged	
State 5	Undamaged	
State 6	Undamaged	State 4-9: 87.5% stiffness reduction at various positions to simulate temperature impact [see (Figueiredo <i>et al.</i> , 2009) for details]
State 7	Undamaged	
State 8	Undamaged	
State 9	Undamaged	
State 10	Damaged	Gap = 0.20 mm
State 11	Damaged	Gap = 0.15 mm
State 12	Damaged	Gap = 0.13 mm
State 13	Damaged	Gap = 0.10 mm
State 14	Damaged	Gap = 0.05 mm
State 15	Damaged	Gap = 0.20 mm & mass = 1.2 kg added at the base
State 16	Damaged	Gap = 0.20 mm & mass = 1.2 kg added on the 1 st floor
State 17	Damaged	Gap = 0.10 mm & mass = 1.2 kg added on the 1 st floor

6.4 ANALYSES AND DISCUSSION

The data used in this study is from the second floor sensor which is close to the damage location to guarantee the sensitivity of the method when classifying different-level damage cases. The testing data, established by taking 20 first tests in each of 9 undamaged states and all tests of damaged structure, therefore has 580 (i.e. $20 \times 9 + 50 \times 8$) observations. With 30 remaining tests in each undamaged state for the training purpose, differently sized learning data can be formed by varying number of training tests (i.e. from as low as 1 up to 30) taken in each learning state. This is to illustrate the impact of the observation size reflected through the two data condition assessment methods (as previously mentioned) by means of pure experimental data. For the sake of simplicity, this number of tests per learning state will be referred below as “state observation size”.

The feature used in this investigation is the Auto-Regressive (AR) vector which has also been used in recent studies using this dataset (Figueiredo *et al.*, 2009; Figueiredo *et al.*, 2011). Each raw data time series is first standardized to zero mean and unit

variance before being transformed into an AR vector with a user-selected model order. Even though there are a number of order estimation techniques (Figueiredo *et al.*, 2009; Figueiredo *et al.*, 2011), in this study, the heuristic technique, based on directly observing RMSE of the AR model, is adopted. The basis for this adoption is that it reflects the actual impact of order change on prediction capacity of AR model which, in the opinion of the present authors, is the most crucial. For the sake of completeness, the following part will present a brief description of AR model and the order estimation method based on RMSE.

The AR (p) model, for a regularly sampled time series process Y with n observations can be described by the following formulae

$$y_i = \hat{y}_i + \varepsilon_i \quad (6-2)$$

$$\hat{y}_i = \sum_{j=1}^p \phi_j y_{i-j} \quad (6-3)$$

where y_i , \hat{y}_i and ε_i are the measured signal the predicted signal and the residual error, respectively at the discrete time index i while ϕ_j is the j th AR variable which can be estimated by one of a number of techniques such as Burge, least squares and Yule-Walker (Ljung, 2011). RMSE of the time series predicted by an AR (p) model with respect to the measured signal is therefore as follows

$$\text{RMSE}(p) = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6-4)$$

To find an appropriate model order, RMSE is plotted as a function of the model order which in turn can be estimated by minimizing the RMSE value. Figure 6-2 shows the average RMSE of AR models with the orders ranging from 1 to 40 for each of the 9 undamaged states. One can see that, RMSE becomes significantly steady for all 9 states from the order of 10 which suggests that one should choose the order at least from this value. In the following sections, this suggested starting order (i.e. 10) and one rather high (i.e. 30), along with one medium (i.e. 15) at some points when necessary will be used in the succeeding sections.

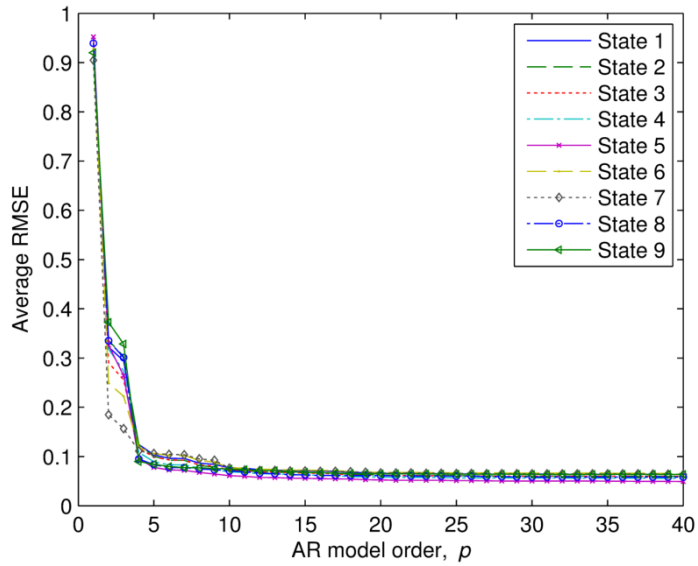


Figure 6-2 RMSE of AR models of increasing order for each undamaged state

6.4.1 MSD-BASED DAMAGE IDENTIFICATION PERFORMANCE ON PURE EXPERIMENTAL DATA

At the model order p , one feature (i.e. AR vector) for each observation in either training or testing data is computed by least squares technique. This leads to 270 by p training data and 580 by p testing data. The threshold distance which is used to differentiate between the undamaged and damaged states is established based on the highest confidence level (i.e. 100%). This can avoid as many as possible the Type I error which, in the opinion of the present authors, is more crucial than the ability of detecting lightly damaged cases which one might achieve by using a lower confidence level. Using this confidence level, the MSD training model is able to correctly detect almost all damage cases – only 1 out of 400 Type II error tests is occasionally found across the lower-dimensional feature (i.e. AR10 and AR15). The high-dimensional feature (AR30) herein has seen no Type II error indicating that it is slightly more sensitive to damage than AR10 and AR15. Overall, the results have confirmed that the previously selected confidence level is appropriate.

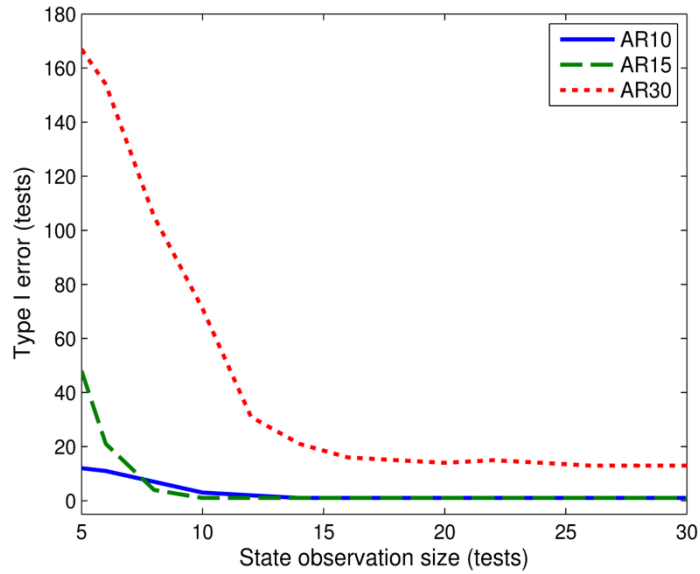


Figure 6-3 Type I error of increasing observation size

In spite of using the highest confidence level, the result of Type I errors significantly differs from that of the Type II errors especially for the smaller range of observations. Figure 6-3 plots the number of false positive errors (out of total 180 tests) against the state observation size in the range between 5 to the maximum (i.e. 30) as previously described. It can be seen that, the Type I error becomes significant for most of the feature dimensions when less than a quarter of the maximum training data is available and is generally higher for higher dimensions. This is most likely due to the fact that, with higher number of variables, higher-order AR models require more observations to be as sufficiently trained as lower-order models. It is worth noting that this problem is well known as “curse of dimensionality” (Farrar and Worden, 2013) and the use of CMCDG herein should be seen to mitigate this problem.

6.4.2 PERFORMANCE OF TWO CONDITION ASSESSMENT METHODS ON PURE EXPERIMENTAL DATA

In Figure 6-4, the condition number of the MSD model is plotted against the state observation size across three feature dimensions in normal linear scale as well as logarithmic (log) scale to facilitate the comparison at different ranges.

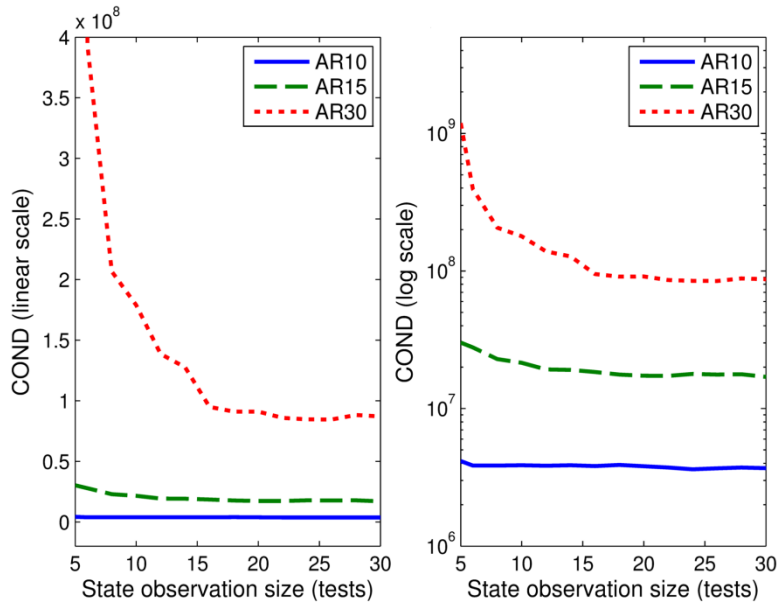


Figure 6-4 COND in linear (left) and log (right) scales

From this figure, one could see that the condition number tends to converge after certain number of observations which is larger for higher feature dimensions. Overall, it can be seen that the convergence trend of this condition number is in fairly good agreement with the performance result presented in Figure 6-3.

To construct a beta Q-Q plot, the training distance in formula (6-1) first needs to be scaled by a factor related to the sample size (n) as follows

$$d_i^* = \frac{n d_i}{(n-1)^2} \quad (6-5)$$

If the training data is multinormal, this scaled distance would follow a beta distribution. The scaled distance is then ranked in ascending order and plotted with the corresponding beta quantiles (Rencher, 2002). For illustration purpose, Figure 6-5 shows the beta Q-Q plots of AR10 (at the state observation of 5 and 13 tests) and AR30 (at the state observation of 8 and 16 tests). These two (one small and one medium) datasets are selected to represent two (one unstable and one improved) conditions of the data, respectively.

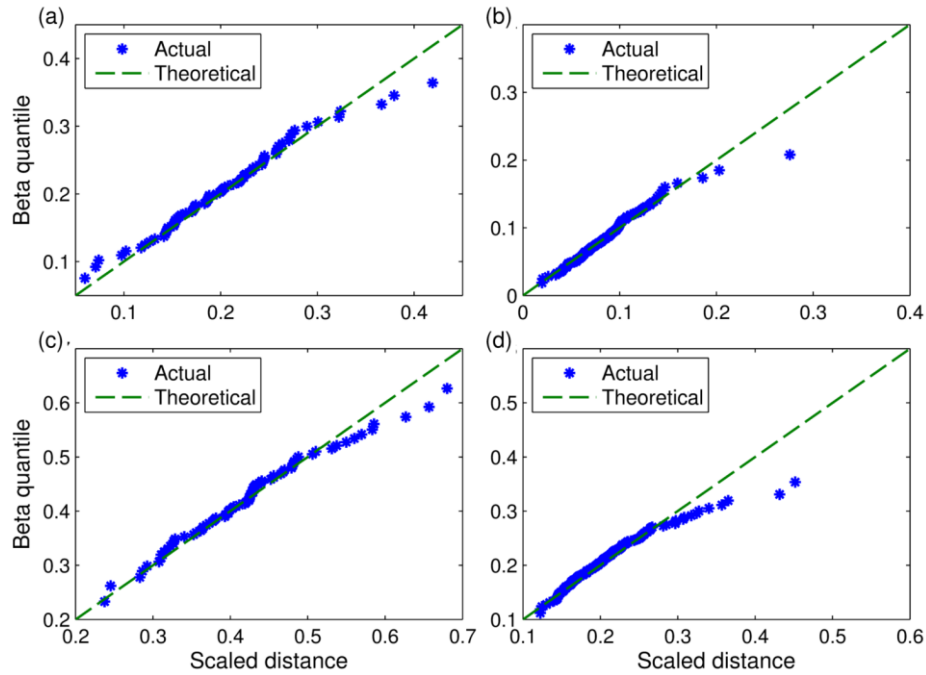


Figure 6-5 Beta Q-Q plot of (a) AR10-05 tests, (b) AR10-13 tests, (c) AR30-08 tests and (d) AR30-16 tests

From Figure 6-5, one can see that increasing number of observations generally improves the agreement between the actual and theoretical plots for most of the data points. This reveals that it is feasible to use a good-of-fitness measure between the two plots as another data condition index (besides COND) to evaluate a huge number of datasets generated from CMCDG process. As previously mentioned, the measure adopted is RMSE which is one of the most commonly-used measures for this type of purpose.

6.4.3 PERFORMANCE OF CMCDG ON PREMATURE DATA

Previous results have shown that the condition of the experimental data will require certain numbers of observations to reach a stable point. Before that, data can be considered as premature and will therefore need a compensation solution such as from CMCDG to improve its condition. In this section, CMCDG will be applied on two premature training datasets each of which is for each feature type, i.e. at the state observation size of 5 tests (for AR10) and 8 tests (for AR30) as preliminarily checked by COND and beta Q-Q plot as shown in Figure 6-4 and Figure 6-5. With such limited observations, the main problem for these two premature datasets is the Type I

errors as previously discussed and presented in Figure 6-3. Out of a total of 180 tests, the original Type I errors of these two (AR10 and AR30) training datasets are 13 and 105 tests (or 7.2% and 58.3% in terms of the error rate), respectively. Under the CMCDG scheme, each premature dataset is first employed as the seed to generate a (user-specified) number of additional datasets of the same size as the seed (by means of random noise) and all the datasets are then tiled one after another to obtain the final data. The random noise herein is generated based on its optimal level in Root-Mean-Square (RMS) sense with respect to the largest deviation of the training features. In this study, the optimal level of noise is determined by the convergence basis of median and IQR of COND. As an example, Figure 6-6(a) and (b) shows the probability distribution of COND values at different noise levels (from 0.05 to 5%) when running 10,000 simulations to evaluate the case of using CMCDG generating 19 additional data replications. Note that the presented noise levels on Figure 6-6 are unequally distributed to accommodate different ranges of noise. From Figure 6-6(a) and (b), one can clearly see that the median and IQR of COND are very large if very low level of noise is employed such as at 0.05 or 0.1%. This is because when noise levels that are too low are applied to the data generation process, subsequently generated observations will have inadequate randomness with respect to the initial observations in the seed as previously discussed. In this case, the covariance matrix becomes more computationally unstable [reflected by larger and more widely variable COND values (Statistics Toolbox Development Team, 2011)] than those formulated by later ranges of noise levels.

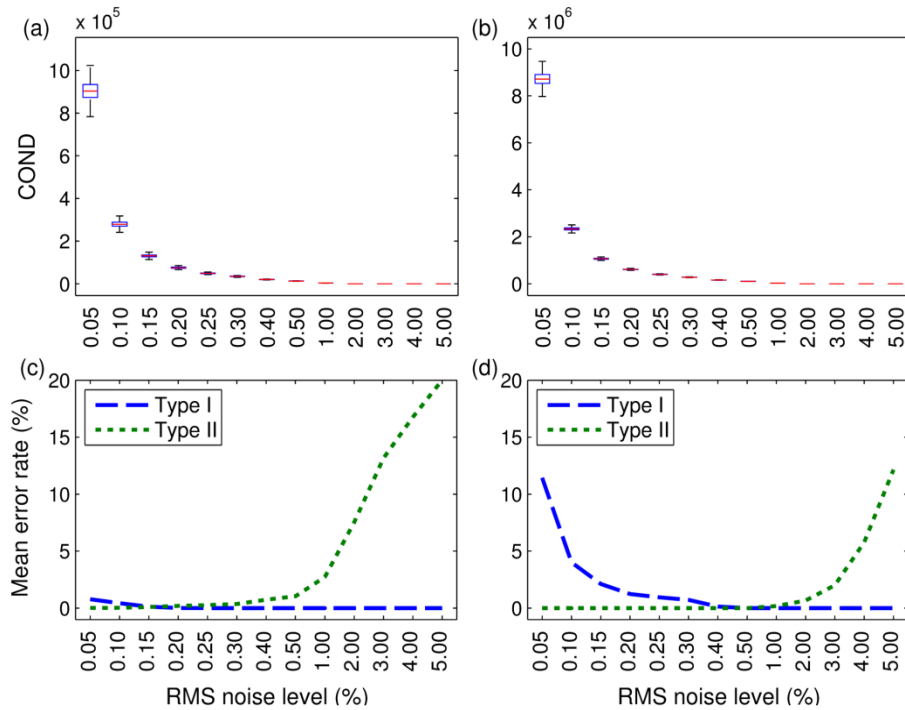


Figure 6-6 COND and mean error rate of increasing noise level:
(a and c) AR10 and (b and d) AR30

However, when the noise level increases, COND rapidly decreases in both median and IQR values. This results in unnoticeable difference in these values from the noise level of around 0.4% onward even though the noise increment later is set at 1%. For correlation purposes, the corresponding mean Type I and Type II errors are also shown in Figure 6-6(c) and (d) in relative sense with respect to a total of 180 Type I and 400 Type II tests. One can first see that the Type I error result is generally in good agreement with the convergence trend of COND. Note that higher Type I error rate for AR30 (in comparison with AR10) at low noise levels should not be seen as abnormal since the initial rate of the premature AR30 data is 58.3% (while that of AR10 is only 7.2%) as previously mentioned. On the other hand, Figure 6-6(c) and (d) appear to show certain impact for the Type II error at high noise levels. However, checking the details across multiple noise levels from 0.5% (for AR10) or 1% (for AR30) up to 5% has revealed that all the Type II errors for both AR10 and AR30 merely belong to the most lightly damaged states (i.e. state 10 and its two variants, state 15 and 16 as illustrated in Table 6-1). Detecting such a damage state may be desirable but not always in the highest priority of damage identification as previously discussed in the regard to choosing the confidence level. Nevertheless, using a

higher-dimensional feature (such as AR30 that has lower Type II error rate) and/or a correct noise level (close to such an optimal level as 0.4% herein) will enhance the damage identification outcome. This also reaffirms the need to determine of an optimal noise level such as being considered in the CMCDG scheme herein since this can lead to a more satisfactory solution.

To find a possibly minimum number of data replications to be used in CMCDG, the same approach used to produce Figure 6-6 will be implemented with a minor swap. The noise level is fixed (at 0.3% for AR10 and 0.5% for AR30) while number of data replications is varied. Figure 6-7 shows the probability distribution of COND and beta Q-Q RMSE along with the mean rate of the Type I error. Again, one can see that both COND and RMSE tend to rapidly converge in both median and IQR values after a certain number of data replications. The figure also shows that the convergence trends of these two indices are in excellent agreement with each other and with that of the Type I error. On the other side, the Type II error results can be retained as more or less the same as those from pure experimental data as previously presented. Once again, there is no single error for AR30 while AR10 only fails to detect one or two most lightly damaged cases out of total 400 tests. This is probably mostly due to the nature of lower-dimensional features such as AR10 which is less sensitive to damage than high-dimensional features like AR30 as previously remarked. This also highlights the feasibility of CMCDG in assisting the use of the high-dimensional feature that may result in higher capability of detecting structural damage.

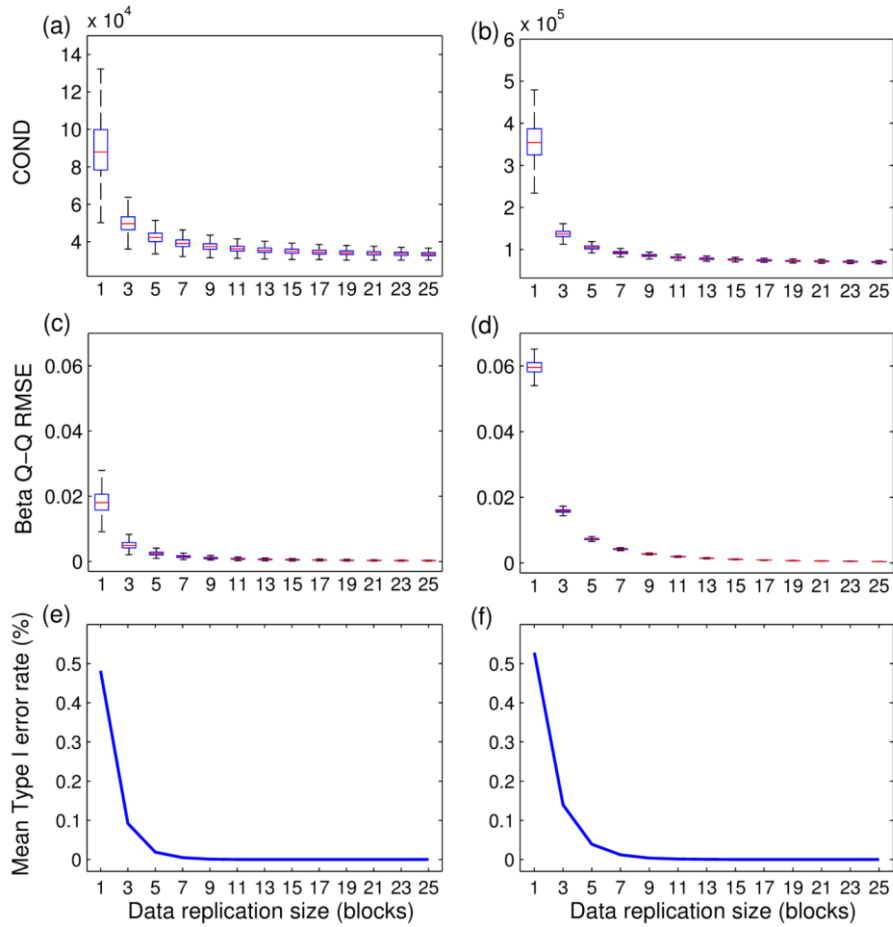


Figure 6-7 COND, Q-Q RMSE and Type I error rate of increasing replication size: (a, c, and e) AR10 and (b, d, and f) AR30[‡]

Based on the convergence of these two condition indices, one can adopt 15 as a possibly minimum number of additional data replications that need to be generated in CMCDG for both feature types of this demonstration example. At this replication size, both post-CMCDG datasets (of both feature types) face no single Type I error across 180 total tests. Compared to aforementioned initial error rates (7.2% and 58.3%) of original datasets, this obviously reflects excellent improvements for the Type I testing performance for both feature types in general and for high-dimensional feature (AR30) in particular.

[‡] As they are (nearly) zero, Type II error rates have been omitted for a better display of Type I errors

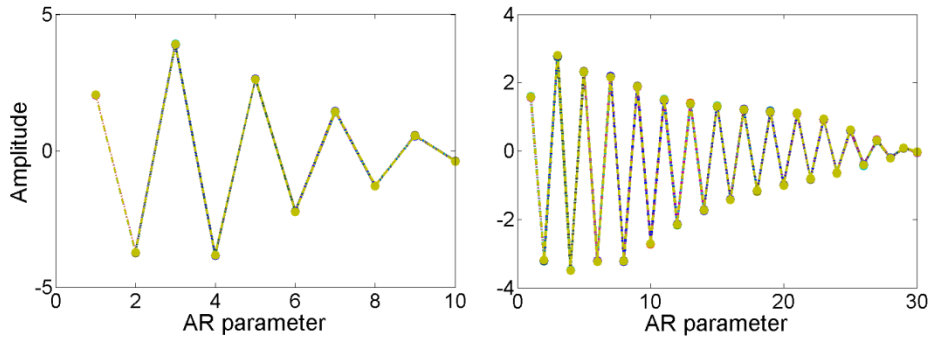


Figure 6-8 Overlay of one typical seed observation and its 15 variants:
AR10 (left) and AR30 (right)

In Figure 6-8 for each feature type, one typical seed (initial) observation and its 15 variants generated by CMCDG are overlaid together and one can see that they are almost identical. This means that the noise addition process in CMCDG does not induce significant variations on the amplitude of the observation. Instead, the efficacy of CMCDG is mainly from the generation of multiple additional random observations to provide a sufficiently large random dataset as directed by CLT and LLN. Finally, to illustrate detailed effectiveness of CMCDG on the training data multinormality, the beta Q-Q plots of two typical datasets generated by CMCDG using aforementioned selected noise levels (0.3% and 0.5%) and replication sizes (15 blocks for both feature types) are shown in Figure 6-9. Compared to those of original (pre-CMCDG) datasets as in Figure 6-5(a) and (c), there are inarguable improvements in terms of the agreement between actual and theoretical lines of the post-CMCDG datasets of both feature types. This once again confirms the effectiveness of the CMCDG scheme.

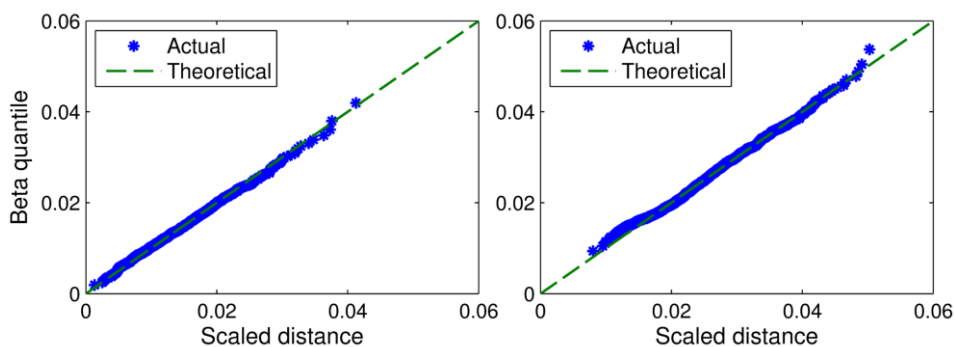


Figure 6-9 Post-CMCDG beta Q-Q plots: AR10 (left) and AR30 (right)

From results presented, it has become apparent that one can conquer the data shortage by employing CMCDG without having to suffer from data burden that is more likely to be confronted during the application of the uncontrolled data generation approach.

6.5 CONCLUSIONS

This paper has proposed an enhanced data generation scheme named CMCDG which can be used to compensate for the shortage of data such as at an early monitoring stage. Targeting a more systematic approach, CMCDG is constructed by adding into the conventional data generation approach two condition assessment methods cooperated with a robust probability-based evaluation procedure. Stemming from a computationally efficient method in linear algebra, COND has been shown to be a simple but useful condition index. This indicator can be used for not only assessing the data condition but also statistically evaluating the effect of random disturbance at different levels such as random noise. Based on the latter usage, the optimal noise level and the possibly minimum number of data replications to be used in CMCDG can be derived so that the generated data can be used for reliable damage identification while being kept reasonable in size. As a different approach, the second assessment method can first act as a convenient tool for graphically examining the status of any single dataset. To use in CMCDG besides COND to work with huge number of simulated datasets, the previous graphical evaluation method is transformed into a single condition indicator which is actually one of the most common good-of-fitness measures, RMSE, to track the discrepancy between actual data and theoretical data. The rationale of utilising the convergence basis of all of data condition indices for determining optimal input for CMCDG has been proved under the regulation of two well-known theorems i.e. CLT and LLN. These two theorems have also been found to be the theoretical bases not only for the CMCDG scheme developed herein but also for the traditional data generation approach. The implementation and application of CMCDG to a benchmark data have shown that CMCDG and its added components can compensate well for the data shortage, improve computational stability and therefore the reliability of MSD-based damage identification. This has also highlighted an important role of CMCDG in assisting the

high-dimensional features such as AR30 that is likely to have higher sensitivity towards a lightly damaged case. Finally, as been shown to be able to improve multinormality of data, CMCDG can be seen as a promising scheme not only for novelty detection based damage identification but also for statistically-based structural analysis in a broader field.

Chapter 7: Field Validation of Controlled Monte Carlo Data Generation Scheme for Statistical Damage Identification Employing MSD

This chapter is made up of the following published journal paper

- ✚ Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. Field validation of controlled Monte Carlo data generation for statistical damage identification employing Mahalanobis squared distance. *Structural Health Monitoring* 13 (4):473-488; URL: <http://dx.doi.org/10.1177/1475921714542892>

As it is an extension of Chapter 6, the main contribution of this chapter is undoubtedly field application and validation for the data generation scheme named CMCDG proposed and initially validated by laboratory data in the previous chapter. CMCDG is previously derived to assist the MSD-based damage identification method in adverse data shortage circumstances such as at an early monitoring stage or during short-lived SHM programs. To illustrate these circumstances in reality, two practical case studies are used in which the first one is established by means of the building vibration datasets recorded by the sensing system developed in Chapter 5 while the second case study comes from an actual bridge with a naturally damaged state. By using these field vibration databanks, not only the efficacy of CMCDG is thoroughly validated but the dynamic structure of this scheme is also highlighted in the aspect of making CMCDG itself well adaptive to any input data with any primary distributional condition. The results show that the assistance from CMCDG is valuable truly enhancing the computational robustness of the MSD-based damage identification method and consequently the reliability of the safety evaluation system.

STATEMENT OF JOINT AUTHORSHIP

The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- There are no conflicts of interest
- They agree to the use of the publications in the student's thesis and its publication on the Australasian Research Online database consistent with any limitations set by publisher requirements.

Contributors:

- Mr. Theanh Nguyen (PhD Student): Conceived the paper ideas; designed and conducted experiments (for those conducted at QUT); analyzed data; wrote the manuscripts; and addressed reviewers comments (for published or in-press papers) to improve the quality of paper.
- Prof. Tommy Chan (Principal Supervisor), Prof. David Thambiratnam (Associate Supervisor): Provided critical comments on the student's formulation of the concepts as well as on his experiment and data analysis programs; and provided editorial comments to enable the student to improve the presentation quality of the manuscripts and the revisions for the published and in-press papers.

Principal Supervisor Confirmation: I have sighted email or other correspondence from all Co-authors confirming their certifying authorship.

Tommy Chan

Name



Signature

18 June 2014

Date

ABSTRACT

This paper presents the field applications and validations for the Controlled Monte Carlo Data Generation (CMCDG) scheme. This scheme was previously derived to assist the Mahalanobis Squared Distance (MSD) based damage identification method to cope with data shortage problems which often cause inadequate data multinormality and unreliable identification outcome. To do so, real vibration datasets from two actual civil engineering structures with such data (and identification) problems are selected as the test objects which are then shown to be in need of enhancement to consolidate their conditions. By utilizing the robust probability measures of the data condition indices in CMCDG and statistical sensitivity analysis of the MSD computational system, well-conditioned synthetic data generated by an optimal CMCDG configurations can be unbiasedly evaluated against those generated by other setups and against the original data. The analysis results reconfirm that CMCDG is able to overcome the shortage of observations, improve the data multinormality and enhance the reliability of the MSD-based damage identification method particularly with respect to false positive errors. The results also highlight the dynamic structure of CMCDG that makes this scheme well adaptive to any type of input data with any (original) distributional condition.

KEYWORDS

Statistical Damage Identification, Mahalanobis Squared Distance (MSD), Controlled Monte Carlo Data Generation (CMCDG), Field Validation, Multinormal, Sensitivity Analysis

7.1 INTRODUCTION

The use of machine learning algorithms for practical Structural Health Monitoring (SHM) in general and structural damage identification in particular has become increasingly popular in recent years. This is due to the fact that this approach could help overcome the adverse impact from inherent Environmental and Operational (E&O) factors that otherwise can prevent the intended objective such as structural

damage from being detected (Sohn *et al.*, 2003; Farrar and Worden, 2013). To do so, a broad range of measured data collected under different E&O conditions of the structure is first used to train the learning algorithm. Once completed, the trained algorithm is supposed to understand the internal relationships of the data within each class (e.g. undamaged or at a specific level of damage) as well as to account for the underlying trend induced by E&O factors. Misjudgement induced from E&O impact can therefore be greatly mitigated and the algorithm can be used to identify genuine structural damage. In this context, one of the most promising methods particularly in the unsupervised learning category is the use of statistical damage identification by means of the Mahalanobis Squared Distance (MSD) based learning algorithm. In the more general disciplines such as novelty detection, the use of MSD-based learning algorithm is also very popular especially in the parametric statistical approach (as opposed to the non-parametric statistical approach) (Markou and Singh, 2003a). Compared to other popular damage identification methods such as those based on neural network, MSD-based method is generally more advantageous towards practical SHM systems which are often associated with the long-term and/or frequent Data Acquisition (DAQ) strategies. This is due to the architectural simplicity and computational efficiency of the MSD-based learning algorithm (Figueiredo *et al.*, 2011) making it more suited for dealing with large volume of data often encountered in such SHM systems in later monitoring stages. In recent experimental evaluations, MSD-based damage identification has also been seen among the most effective methods (Worden *et al.*, 2000a; Sohn *et al.*, 2003; Figueiredo *et al.*, 2011; Worden *et al.*, 2003; Worden *et al.*, 2000b). Besides its own application, MSD is also closely related to the popular Hotelling's T^2 control chart and indeed equivalent to the T^2 statistic when the subgroup size is set at unity for the latter method (Wang and Ong, 2008; Farrar and Worden, 2013). In spite of having such wide connection and merits, MSD-based damage identification method has however had one "Achilles heel", that is, the requirement of the learning data to be multivariate normal (multinormal) distributed. This tends to be more problematic for the cases of employing the infrequent DAQ mode or at an early monitoring stage when not much measured data is available. To cope with this problem, a so-called Controlled Monte Carlo Data Generation (CMCDG) scheme has been derived and reported in one of recent publications of the present authors (Nguyen *et al.*, 2014a). Using this scheme,

additional data can be produced from a limited number of original observations by means of an optimised Monte Carlo simulation process. Such an optimised simulation is useful not only to estimate an optimal noise level (which is to provide optimal randomness for the outcome data) but also to retain the (outcome) data at a reasonable size. Even though this scheme has been intensively tested against a sophisticated laboratory dataset, one may still be concerned that the success of using CMCDG has only been experimentally proved in a well-controlled testing environment. Additional applications towards real infrastructure vibration data are therefore in need in order to further evaluate and demonstrate the efficacy of this scheme.

To address this need and further extend the study on CMCDG, this paper presents applications of this scheme onto real vibration monitoring data from two actual civil engineering (one bridge and one building) structures each of which has been considered as an SHM benchmark structure. Of these two structures, the bridge can represent for the case of having inadequate quality data and/or infrequent measurements while the building represents the case where the data shortage issue occurs at an early monitoring stage. To overcome such a data shortage problem in either case, the CMCDG scheme is applied to the original learning data in order to generate well-conditioned synthetic data and therefore numerically stable (computational) system realizations. Besides utilizing two existing assessment indices in CMCDG, this study also employs statistical sensitivity analysis of the testing MSD computation using representative generated datasets to further validate the efficacy of CMCDG. The outcome of these applications reconfirms that the CMCDG scheme is able to help overcome the data shortage problem and enhance the reliability of the MSD-based damage identification method. The layout of this paper is as follows. The next section provides concise theoretical descriptions of the MSD-based damage identification method and the CMCDG scheme. The benchmark structures and their datasets used in this study are then briefly described. In the last two sections, detailed analyses and discussions are first provided before the key issues and findings are summarised in the conclusion. It might be worth noting that the scope of this research is currently restricted to level 1 of the damage identification hierarchy, that is, to identify the presence of damage. However, as the problem of

false indications has persisted fairly significantly at this level in the prior studies (Figueiredo *et al.*, 2011; Gul and Catbas, 2009), the present authors believe that enhancing the accuracy of this phase is still very crucial besides addressing problems of the higher damage identification levels.

7.2 THEORY

7.2.1 MSD-BASED DAMAGE IDENTIFICATION

There are two main types of data used in a statistical damage identification process. In general, the primary (or raw) data acquired by sensors is not directly used but is transformed into a (damage sensitive) feature which then become the input data for the learning algorithm. Since this transformation process is often conducted by means of data compression methods such as modal analysis or time series modelling, feature data is often in a much lower dimension. The most popular features in SHM include the vectors of modal parameters or auto-regressive coefficients amongst others.

Suppose that a training feature dataset consists of p variables and n observations. If it approximates a multinormal distribution, this dataset can be represented by the sample mean vector (\bar{x}) and the sample covariance matrix (S). Next, each feature observation (x_i) for either training or testing purposes will be converted into a damage index in the form of distance (i.e. MSD) measure (d_i) as follows

$$d_i = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x}) \quad (7-1)$$

Here, the mean and covariance are also the two representatives for the realization (of the MSD computational system) by the given dataset. This point is emphasized as there will be a large number of synthetic datasets (and therefore system realizations as well as their representatives) generated in the CMCDG process. In the damage identification context, the mean and covariance should be formulated as an exclusive measure, or in other words, consisting of no potential outlier from the testing phase (Worden *et al.*, 2000a). After computing all training distances, the assumption of a multinormal distribution again allows the estimation of the threshold from the basis of chi-square distribution for the training distances (Farrar and Worden, 2013). It is because under such an assumption, one can specify a statistical threshold for the

distances based on a distribution quantile or equivalently a confidence level (Filzmoser *et al.*, 2005; Farrar and Worden, 2013). In the testing phase, whenever a new observation comes, its corresponding distance can then be used to compare against the threshold to determine whether it corresponds to a normal or damaged state. There might be a trade-off in choosing the confidence level: using a high level of confidence might not be able to detect a lightly damaged case that is known as one class of Type II errors but the least critical. However, such confidence levels can assist in avoiding as many as possible false-positive indication of damage (i.e. Type I errors) (Farrar and Worden, 2013). In this study, one of such high levels (i.e. 99%) will be used in the application.

7.2.2 CMCDG

The CMCDG scheme proposed is an enhanced version of the conventional Monte Carlo data generation scheme which has been frequently used in the MSD-based damage identification context (Gul and Catbas, 2009; Worden *et al.*, 2002; Worden *et al.*, 2000a). In both schemes, the shortage of the data is compensated by the provision of statistical replications of each initial observation by means of Gaussian noise (Nguyen *et al.*, 2014a). However, the core components of the CMCDG scheme that make it more advanced than the conventional scheme are two data condition indices and a robust probability based evaluation procedure used to obtain robust statistical measures for either index. Of the two indices, the (two-norm) condition number (COND) of the covariance matrix is intended to monitor potential computational instability associated with the use of the inverse of the matrix component in the equation (7-1). On the other hand, the second index is the Root Mean Square Error (RMSE) between the theoretical and actual beta Quantile-Quantile (Q-Q) plots of each dataset generated during CMCDG process. By running a sufficiently large number of data generation simulations, the relationships between the commonly-used robust probability measures (of either index) such as median and Inter-Quartile Range (IQR) and the variable such as the noise level or the replication size can be constructed. The user is then able to use the convergence of these statistical measures to determine the optimal value for each of the two variables. The theoretical bases of the CMCDG scheme and the probability convergences of COND and beta Q-Q RMSE have been proved under the regulation of two well-known

theorems, i.e. Central Limit Theorem (CLT) and the Law of Large Numbers (LLN). Details of these can be found in the first paper of the CMCDG scheme (Nguyen *et al.*, 2014a). Since its target is the enhancement of learning data multinormality, CMCDG can also be considered as a (multivariate) data normalization scheme with the focus on the Gaussian-type prerequisite for the learning process. Although it has such desirable features, it should be noted that CMCDG may only be in need for novelty detection methods (as well as associated damage identification methods) in the parametric statistical approach which are often formulated from the multinormal data assumption (Markou and Singh, 2003a). Methods from other approaches such as multivariate exponentially weighted moving average have been shown to have higher tolerance to non-multinormality and then utilize simpler normalization schemes such as data shuffling to overcome the related impact (Wang and Ong, 2010).

7.3 DESCRIPTION OF BENCHMARK STRUCTURES AND THEIR DATA STATUS

7.3.1 SMC BENCHMARK STRUCTURE AND DATA STATUS

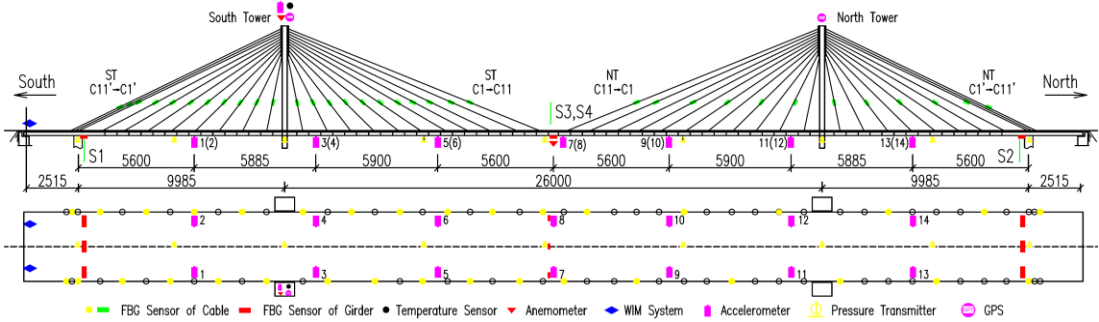


Figure 7-1 SMC benchmark structure (Li *et al.*, 2014)

The first benchmark structure of interest is an actual cable-stayed bridge monitored by the center of Structural Monitoring and Control (SMC) at the Harbin Institute of Technology, China (Li *et al.*, 2014). Opened to traffic in December 1987, this is one of the first cable-stayed bridges in mainland China. This 11-meter-width bridge consists of a main span of 260 m and two side spans of 25.15 + 99.85 m at each end. In 2005 after 19 years of operation, the bridge was found in a rather unsafe condition with a mid-span girder and a number of stay cables being cracked or corroded. Along with major rehabilitation program undertaken to replace the damaged girder segment

and all the stay cables, a sophisticated SHM system (see Figure 7-1) was implemented in order to monitor the bridge from the time of its rebirth in 2007. From monitoring data of this bridge, the SMC research group has been able to develop two SHM benchmark problems: one for stay cable condition assessment and the other for bridge girder damage identification. The context for the second benchmark problem whose data is used in this study is as follows. In August 2008 that is only 8 months after the first complete DAQ after its rehabilitation, the bridge was again found in a new deficient structural condition with several new damage patterns in the girders. Fortunately, this bridge had been frequently monitored during this short period of time and certain distinct difference in modal analysis results could be observed over this monitoring period. Sampled at 100 Hz, acceleration data of 12 days was split into hourly subsets and made available on the SMC website for participants of this benchmark study (Li *et al.*, 2014). In the study herein, only acceleration data recorded from 14 accelerometers installed on the deck are used. Part of this databank (i.e. of several first days) will be employed as the seed data to be input into CMCDG in order to achieve enhanced data for MSD-based learning process. Usable sets of the remaining data will be used for testing purposes. Details of these datasets are presented in the data analysis section.

7.3.2 QUT-SHM BENCHMARK STRUCTURE AND DATA STATUS

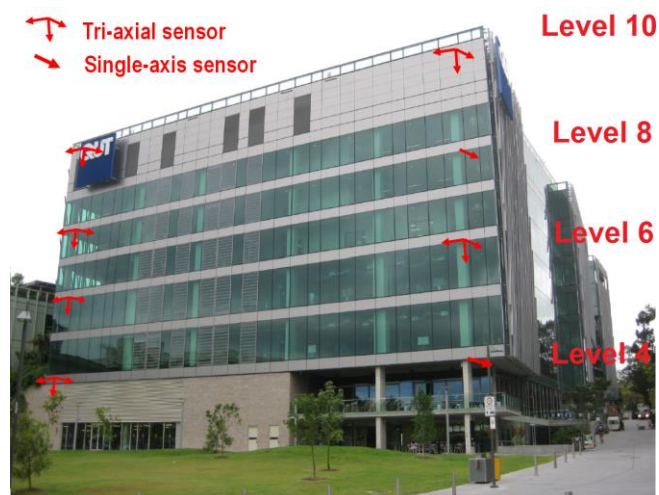


Figure 7-2 QUT-SHM benchmark structure

The second benchmark structure used in this study is the main building with ten stories in the Science and Engineering Centre complex at the Gardens Point campus

of Queensland University of Technology (QUT) in Australia. The most notable feature of this benchmark structure lies at its vibration sensing solution with a software-based synchronization method which can be seen as a promising alternative for use in vibration monitoring of civil infrastructure. At the lowest level of the system, there are only six analog tri-axial and two single-axis accelerometers available for use to capture the vibration responses of this structure. As illustrated in Figure 7-2, the sensors are located on the upper part of the building (i.e. at level 4, 6, 8 and 10) which is globally more sensitive to the ambient excitation sources such as human activities and wind loads. Acceleration data is sampled at the initial rate of 2000 Hz and then split into 30-minute subsets for modal analysis purposes. In spite of using such a limited number of sensors, the sensing system could detect at least six modes with high confidence even under the challenging ambient excitation conditions. However, as the development of this sensing system has recently been completed, its databank is still limited with most of the data being collected during the system implementation phase in late 2013. Such limited data therefore needs the assistance from a data generation scheme like CMCDG to enable the health check process from an early stage. Details of the implementations of CMCDG onto the data of the two benchmark structures are presented in the next section.

7.4 ANALYSES AND DISCUSSION

The feature selected for both benchmark study cases is the vector of modal frequencies estimated by means of the primary technique of the data-driven Stochastic Subspace Identification (SSI-data) family [i.e. SSI-data employing Unweighted Principal Component (UPC) estimator] in Output-only Modal Analysis (OMA) approach. This selection is made due to the following reasons. First, modal frequencies can be more rapidly estimated with higher confidence than other modal parameters such as mode shapes (Salawu, 1997). This is particularly meaningful for SHM in ambient excitation conditions where mode shape estimation is more challenging and time-consuming. Second, primary SSI-data is one of the most robust and advanced OMA techniques which can cope well with large volume of data from long-term SHM processes (Nguyen *et al.*, 2014b, 2014c). Third, online automated frequency estimation is highly possible in practice with the implementation of the

recursive version of SSI-data (Loh *et al.*, 2011). Finally, the modal frequency has been proved to be a main damage index at least for level 1 of damage identification of several large-scale infrastructure such as the well-known Z24 highway bridge (Brincker *et al.*, 2001).

To process vibration data from two benchmark structures, the modal analysis software ARTeMIS Extractor Pro version 5.3 developed by Structural Vibration Solution A/S is used to implement the primary SSI-data technique. Concise descriptions of theory and usage for this technique can be found in several prior papers of the present authors (Nguyen *et al.*, 2014b, 2014c). SSI-data configurations and analysis results for each structure are presented in the following sub-sections.

7.4.1 SMC VIBRATION DATA

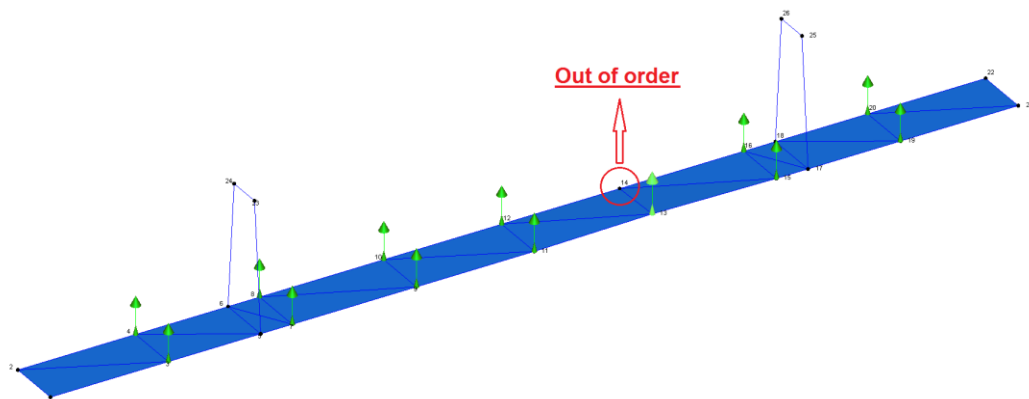


Figure 7-3 SMC bridge model in ARTeMIS Extractor software

For the sake of simplicity, the bridge is modelled, as illustrated in Figure 7-3, mainly with the main span (260 m) and the two larger side spans (99.85 m each) where 14 single-axis accelerometers were deployed. Checking across multiple datasets of this structure has revealed that one of these accelerometers (as circled in Figure 7-3) was out of order but the data from the remaining sensors are still adequate for modal validation (see the analyses later). Also as they are found to be either mostly collected in poor excitation conditions or lacking in the stability along the consecutive sets, data from three days (31 May, 2008; 7 and 16 June, 2008) is excluded from the analyses. Besides these days, the problem of excitation has also

had certain impact on the other days. Table 7-1 lists number of usable datasets from the selected 9 days. Descriptions of data grouping will be detailed later.

Table 7-1 Selected testing days and usable SMC datasets

<i>Selected testing day</i>	<i>Date*</i>	<i>Number of usable subsets</i>	<i>Description of feature dataset</i>
1	01 January	17	Day 1-3: Dataset 1 (State 1, 52 observations)
2	17 January	19	
3	03 February	16	
4	19 March	12	Day 4-8: Dataset 2 (State 1, 46 observations)
5	30 March	13	
6	09 April	7	
7	05 May	7	
8	18 May	7	
9	31 July	24	Dataset 3 (State 2, 24 observations)

[* All within the year of 2008]

The preliminary OMA by SMC group has pointed out certain differences between six frequencies (in the range of 0 to around 1.2 Hz) estimated from data collected in one of the first DAQ days (17 January, 2008) and those from the data acquired in the last DAQ period (31 July, 2008). These differences were assumed to be due to the impact of damage discovered in August, 2008 as mentioned by Li et al (2014). With a similar assumption, the following analyses in this section are to seek the evidence that the usable observations recorded during the first 8 days and the 9th day are likely to belong to two separate states hereafter namely states 1 and 2, respectively. To do so by means of the primary SSI-data technique, a frequency range of interest and a common modal analysis configuration are first required. Owing to small number of sensors and unidirectional measurement which hinder the validation of high-order modes, a decimation factor of 25 times is applied and the frequency range of interest is restricted to between 0 and around 1 Hz to obtain the most accurate modal information. By comparing the results of SSI-data of incremental dimensions and projection channels, the most stable range of the maximum SSI state space dimension is found to be between 120 and 200 whereas that of the projection is from 8 to 11 channels. Hence, the maximum state space dimension of 160 and the option of 9 projection channels are first selected as the main SSI-data configuration for the vibration data subsets used herein.

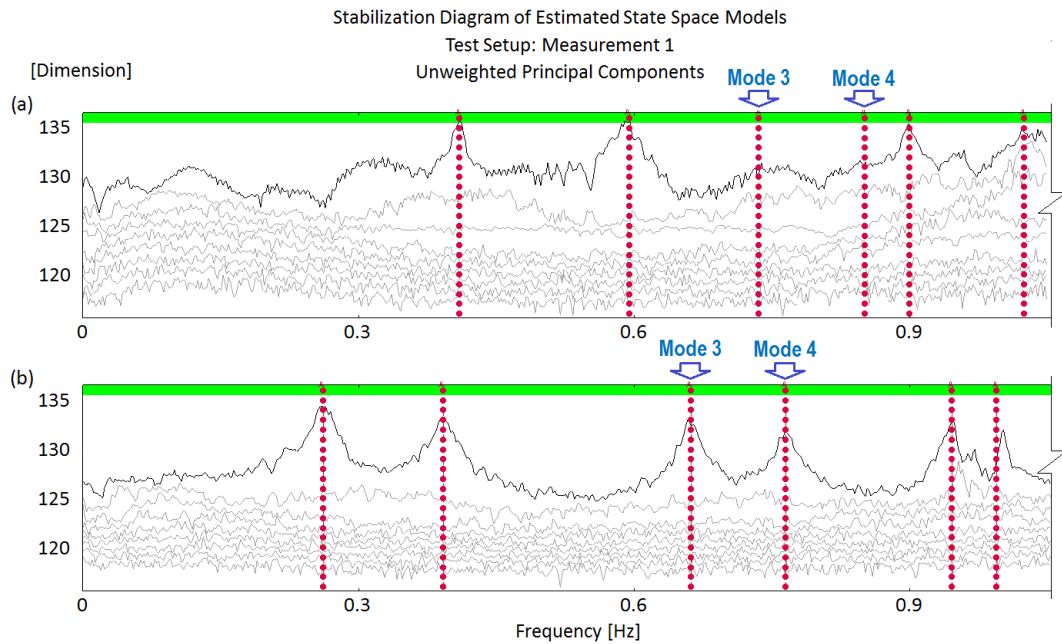


Figure 7-4 Detected modes of SMC benchmark structure: (a) State 1 and (b) State 2

Using the above SSI-data configuration, around six modes may be detected and correlated between the two aforementioned states and these can be illustrated in Figure 7-4 and Figure 7-5 by using two representative datasets for these two states. Of these six modes, four (i.e. modes 1, 2, 5, 6) consistently show up across all datasets of two states. The rather low frequency value of mode 5 in Figure 7-4(a) in comparison with Figure 7-4(b) is mostly due to the former corresponding to an extreme case (see later for detail of frequency comparison). On the other hand, modes 3 and 4, though consistently well-detected in state 2, are only found weakly excited (Figure 7-4) in a limited number of datasets in state 1. This can be seen as the initial evidence for the difference between the two states. The corresponding mode shapes for the two datasets is presented in Figure 7-5 along with the corresponding Modal Assurance Criterion (MAC) for each of the correlated mode shape pairs for the two states. It might be worth noting that all of the first five modes which exhibit a consistently increasing trend in MAC deviation belong to the vertical bending type whilst mode 6 is of a vertical torsion. Compared to the Z24 highway bridge damage identification results (Brincker *et al.*, 2001), low MAC values such as 0.83 and 0.62 (at modes 2 and 5, respectively) can also be seen as truly significant and can therefore serve as the second evidence for the difference between the two states. The last

evidence for such a difference will be inferred from the statistical screening of frequency data in the succeeding paragraph.

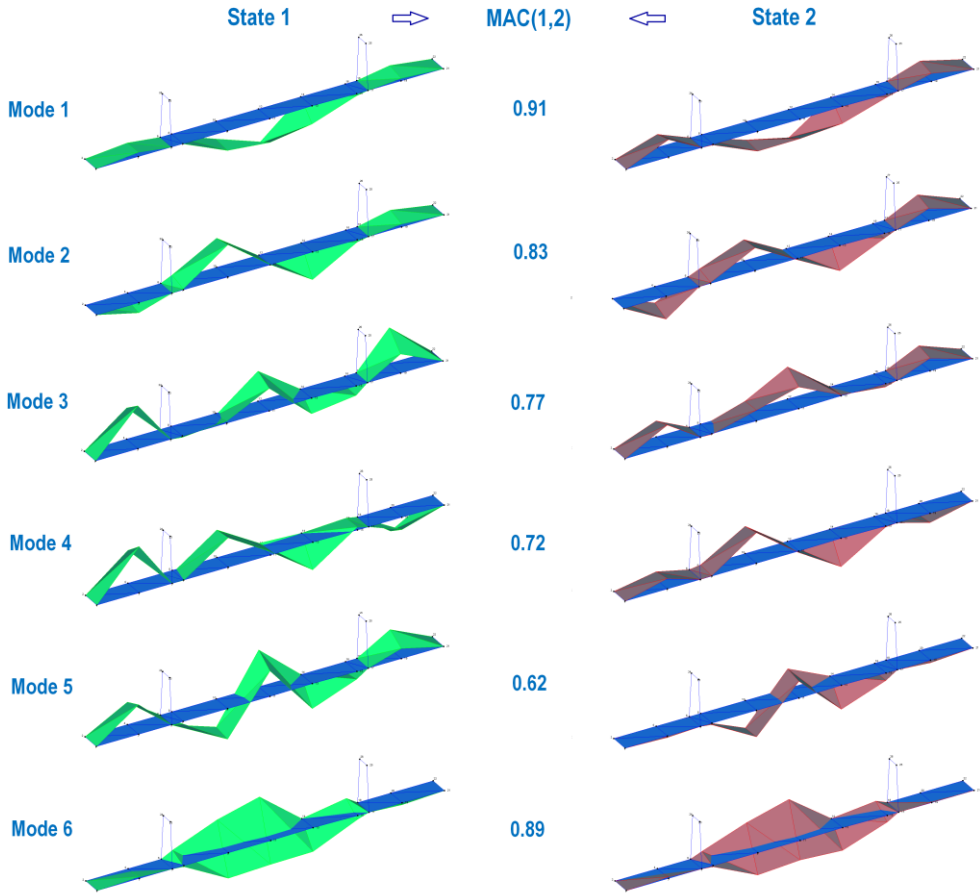


Figure 7-5 Representative correlation between (SMC) mode shapes of two states

Owing to the absence of modes 3 and 4 in the analysis results from most datasets of state 1, the feature data could therefore be established from frequency estimates from the other four modes, or in other words, having four variables. For more detailed comparisons and validation of CMCDG later, feature data of state 1 is split into two sets namely datasets 1 and 2 with 52 and 46 observations, respectively (see Table 7-1 for more details). Figure 7-6(a), (b), and (c) shows box-plots of these two sets along with the third one (of state 2) and one can see that the datasets 1 and 2 are analogous to each other. On the other hand, dataset 3 possesses a distinct difference in the magnitudes of the first two variables. Even though the third variable experiences somewhat opposite change (compared to the other variables), the relative deviation at this variable is rather small (only +1.7%) compared to those at the two first variables

(both almost –30%) in terms of their median values. A possible reason for the former symptom is that the modal frequency of this mode is insensitive to damage but slightly more sensitive to some E&O impact in a similar manner as occurred to the frequencies of the well-known I-40 bridge at its two first damage levels (Farrar and Jauregul, 1998; Farrar and Worden, 2013). Nevertheless, the large reduction in the first two modal frequencies and the two prior evidences can be used as the bases to confirm the discrepancy between the two aforementioned states. Finally, it might be worth noting even though the use of frequencies and MAC values is satisfactory for damage occurrence confirmation herein, this type of damage detection methodologies is only convenient for the case with limited number of datasets. This is because in this approach the analyst would have to check every single feature dataset and compare with the others. For the case having many datasets such as in long-term and/or frequent SHM systems, this type of examination would become extremely time-consuming if not impossible. In this circumstance, the use of MSD-based damage identification is advantageous as it can run autonomously computing the testing distance whenever a new feature observation is available, comparing with threshold and (if larger) giving alarm in a fully automatic manner. Such operation and evaluation capacities of the MSD-based method are critical in order to ensure timely intervention and decision-making towards civil infrastructure and to constantly safeguard the users involved.

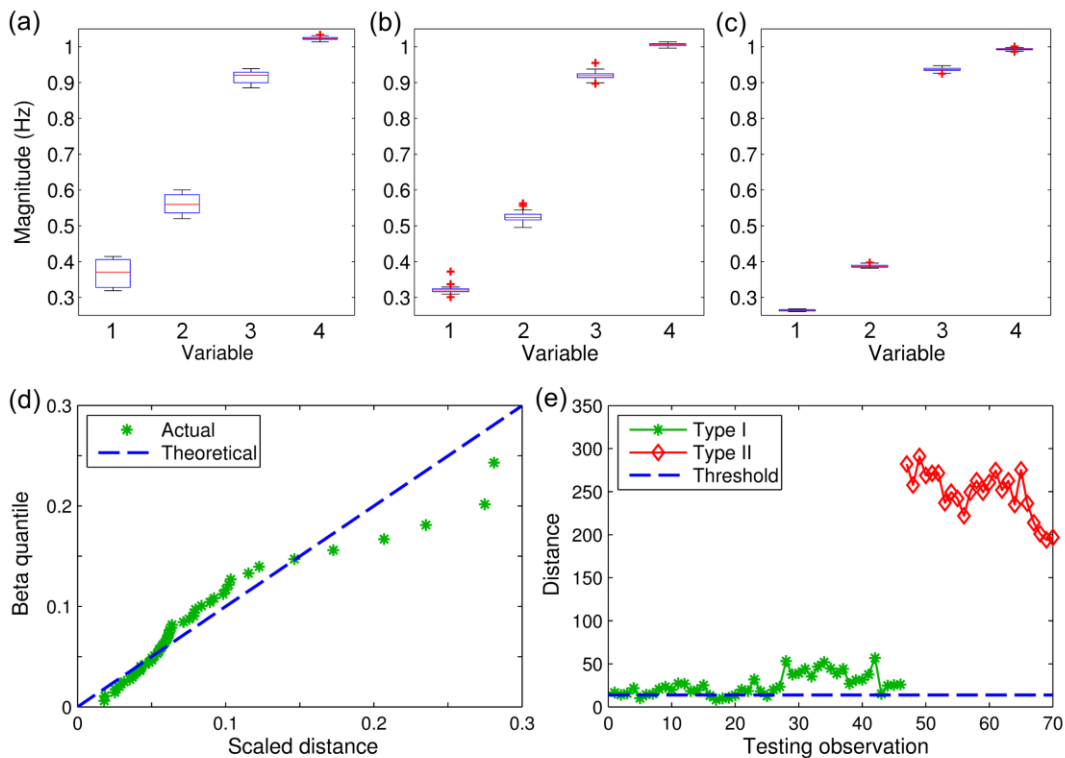


Figure 7-6 Characteristics of SMC data and original testing results: (a, b) datasets 1 and 2 (State 1); (c) dataset 3 (State 2); (d) beta Q-Q plot of dataset 1; and (e) original testing results

In order to rigorously examine the efficacy of a method in distinguishing any two known structural states, the problem should be formulated in the context of hypothesis testing with two hypotheses known as the null hypothesis (H_0) and the alternate hypothesis (H_1). In the damage identification context, the null hypothesis is often assumed for the case when damage is not present while the alternative hypothesis asserts the contrary (Farrar and Worden, 2013). In a probabilistic sense, two kinds of errors may be encountered when testing these hypotheses. If the null hypothesis is rejected even though it is true, then a Type I (false-positive) error has occurred. In contrast, if the null hypothesis is accepted even though it is false, then a Type II (false-negative) error has been committed. In a comprehensive hypothesis testing program, the probabilities of these two error types can then be estimated based on a data distribution under assumption (Montgomery, 2005). However, for the purpose of simplicity, no probability computation will be made and the assessment process herein will be conducted based merely on direct comparison of the error quantities to evaluate the efficacy of CMCDG in assisting the MSD-based damage

identification method. Hence, dataset 1 will be used as the original learning data while datasets 2 and 3 will be employed for the Type I and Type II error testing purposes, respectively.

To check the degree of multinormality of the original learning data, the beta Q-Q plot is employed and the result is shown in Figure 7-6(d). One can see that there is a poor agreement between the theoretical and actual lines and this means that this dataset needs to be enhanced before it can be used for novelty detection or damage identification purposes. As a blind attempt to use this low-quality dataset, the MSD-based damage identification process is implemented onto the 70 (i.e. 46 for Type I and 24 for Type II) testing observations and the testing results are presented in Figure 7-6(e). A closer look for the (selective) Type I distances in conjunction with the threshold can be seen later in Figure 7-9. While no single Type II error is found, Type I errors are extremely severe with more than 80% false indications (as shown in Figure 7-6(e) with most Type I data points lying above the threshold line). To enhance the initial feature data by CMCDG, the optimal Gaussian noise level in the Root Mean Square (RMS) sense is first determined by box-plotting COND of the datasets generated in each data generation setup and tracking the convergence of the median or IQR for multiple setups. Figure 7-7(a) and (b) shows two of such plots of COND at different noise levels (from 0.1 to 25%) when running 10,000 simulations for the first CMCDG round with two illustrative cases, that is, to generate 9 and 18 additional blocks of data replication. Note that three different incremental levels of noise (i.e. 0.3, 1 and 5 %) are used on Figure 7-7(a) and (b) in order to facilitate better displays in different ranges of noise.

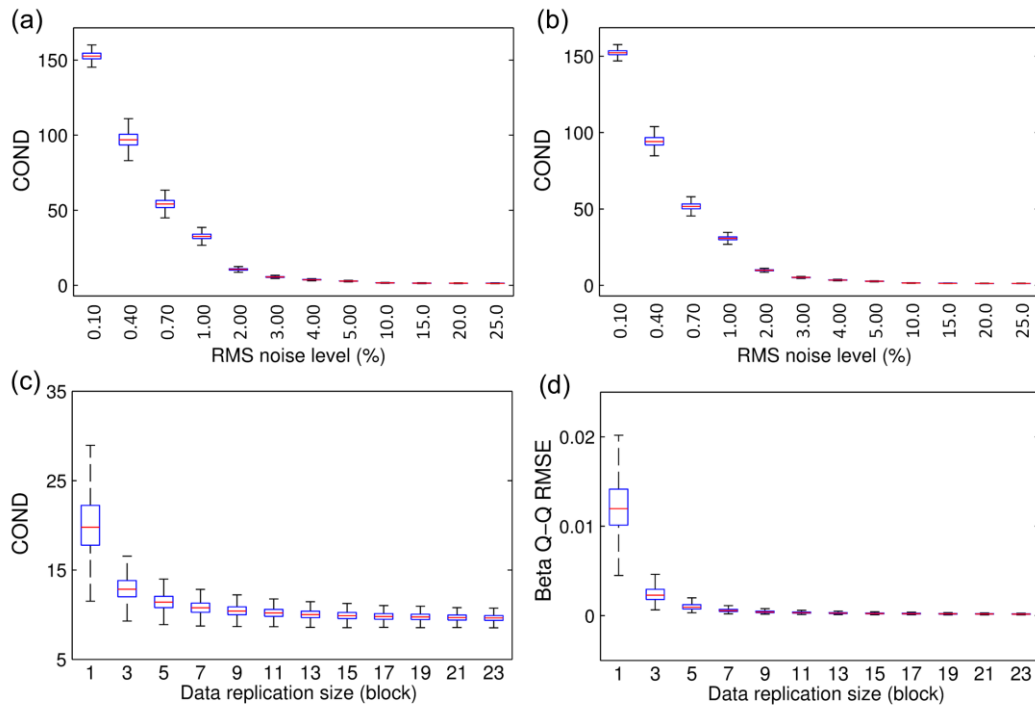


Figure 7-7 Results of two simulation rounds in CMCDG for SMC data: (a, b) round 1 with COND of 9 and 18 replication blocks; (c, d) round 2 with COND and beta Q-Q RMSE at noise level of 2%

As can be seen from Figure 7-7(a) and (b), COND values become significantly small and steady after increasing the noise amount by several small steps and become essentially unchanged at the noise level of 20%. For ease of the selection of an appropriate noise level that corresponds to an essentially small COND (as this level might vary significantly from case to case as to be seen later), a so-called 95% deviation bounds criterion is established as follows. A COND value is considered as essentially small if its deviation from the original COND (i.e. of the original learning dataset) is no less than 95% of the COND span. Here, the COND span is the difference between the original COND and the COND value that has been considered essentially unchanged, that is, corresponding to the noise level of 20% in this case. Applying this criterion upon the medians of COND herein, the appropriate noise levels are found to be from 2% onward. Therefore, the optimal noise level is set at this starting point since the use of higher noise levels tend to reduce the sensitivity in detecting lightly damaged states as noted in the initial investigation with CMCDG (Nguyen *et al.*, 2014a). Employing this noise level, the second round of simulations is operated with the variable being the data replication size and the output being

COND and beta Q-Q RMSE. These two results are graphically shown in Figure 7-7(c) and (d). From this figure, one can find again that COND and beta Q-Q RMSE become significantly small and steady from the replication size of around 9 blocks onward. This figure is therefore considered optimal replication size to provide quality synthetic datasets.

Using this optimal replication size, well-conditioned synthetic data can be generated with the previously selected optimal noise level (2%). Figure 7-8 shows the beta Q-Q plot and the hypothesis testing results for a typical one of such datasets when using it as a replacement for the low-quality original learning data (i.e. dataset 1). Note that the Type I and Type II error testing data are kept the same as earlier (i.e. datasets 2 and 3 with 46 and 24 observations, respectively). Compared to original results reported in Figure 7-6(d, e), substantial improvement in beta multinormality degree is undeniable as reflected in Figure 7-8(a) whilst all the testing observations are accurately identified with no single error in both testing cases as seen in Figure 7-8(b). The enhanced learning data has well improved the reliability of MSD-based method with respect to the Type I error tests while being able to retain sufficient sensitivity to all Type II error testing observations.

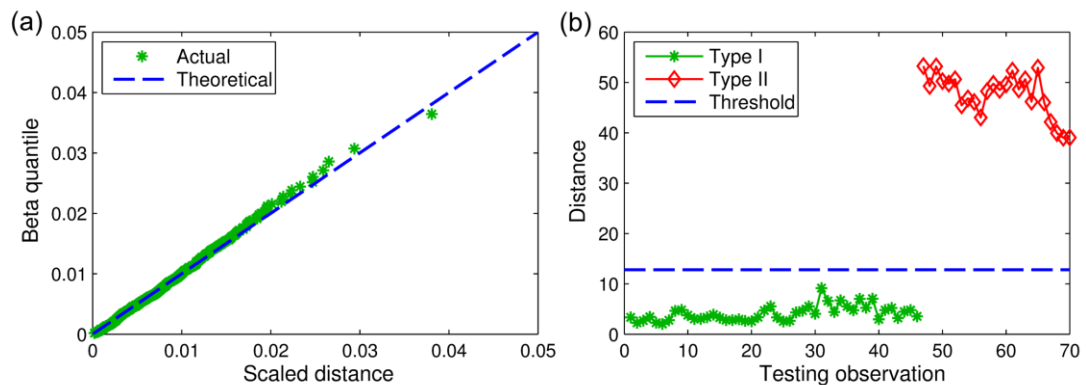


Figure 7-8 (a) beta Q-Q plot and (b) testing results of a typical enhanced (SMC) learning dataset

The earlier problem of having severe Type I errors in the original learning dataset (Figure 7-6) is believed to originate from the instability of the realization (of the MSD computational system) corresponding to this dataset. This has been actually reflected through comparison of COND (in Figure 7-7) since system realizations with

larger COND values tend to suffer from more severe computational instability as previously mentioned. To illustrate this in a more direct manner in MSD-based damage identification process, the robustness of the original computation system realization (i.e. by original learning dataset) with respect to the perturbation of the Type I error testing observation will be assessed against that of the realization by the (typical) enhanced dataset shown in Figure 7-8. Note that this type of assessment is commonly known as sensitivity analysis which is often used to test the robustness of a mathematical model or system in the presence of input uncertainties (Saltelli *et al.*, 2008; Pannell, 1997). Projecting this onto the problem herein, the desired realization (by an appropriate dataset) of the MSD computation system should be as robust as possible against the presence of inherent perturbation (of the testing observation) that may be induced from measurement or data compression phases. Based on this fact, the aforementioned comparative assessments between the original and enhanced datasets are objectively realized by means of the same input (i.e. each of 46 Type I error testing observations), the same magnitude of its statistical perturbation and once again the Monte Carlo simulation in a similar fashion that is used in CMCDG. Specifically, the perturbation level is selected as 2% with respect to the RMS of each investigated observation. Then, 10,000 simulation rounds for the perturbation application and the MSD computation are operated and the corresponding original testing distance and its (10,000) variants are box-plotted in Figure 7-9 for the both original and typical enhanced datasets. Note that due to the paper display limitation, only 12 selective cases (out of a total of 46 testing observations) are reported in this figure for either dataset. Compared to those obtained from the typical enhanced dataset, the fluctuations of the Type I distances computed from the original dataset are significantly (i.e. 10 to 15 times) larger. Further, compared to the magnitude of the threshold, these fluctuations are also truly severe as seen in Figure 7-9(a). Such large fluctuations indicate that it is highly likely that the realization of MSD computational system by the original dataset is in a significantly ill condition and the computational results are unreliable. On the other hand, the marginal fluctuations in Figure 7-9(b) show that the robustness of the computational system has been considerably enhanced through the use of a dataset generated from an optimal CMCDG configuration. Further checks with other datasets generated by succeeding

CMCDG configurations have confirmed the robustness convergence for this configuration but the detailed results are not shown to save space.

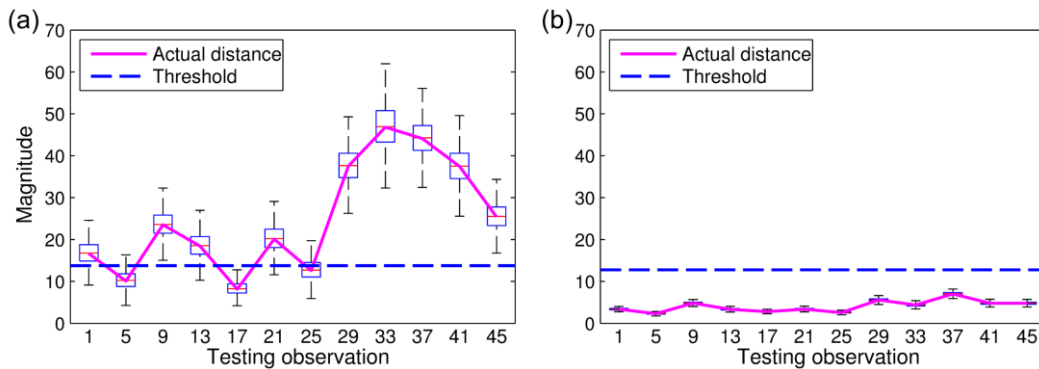


Figure 7-9 Impact of input perturbation on (SMC) MSD computation: (a) original learning dataset and (b) typical enhanced learning dataset

7.4.2 QUT-SHM VIBRATION DATA

As mentioned earlier, as the full monitoring program for this benchmark structure has recently been started, its databank is still limited with 100 subsets at the time of processing data for this paper. Of these subsets, most (64 subsets) were collected during the system development phase in late 2013 and the remaining (36 subsets) were collected in January, 2014. Using an optimal SSI-data configuration similar to the one used for the SMC data, up to seven modes could be estimated as illustrated in Figure 7-10 for one representative data subset. Nevertheless, only six of the modes (i.e. modes 1-5 and 7 as typically animated in Figure 7-11) are usable for the purpose of continuous modal tracking. The exclusion of mode 6 is due to the inconsistency of modal estimation at this particular mode across different datasets recorded under different E&O conditions. As it is a weakly-excited mode (i.e. not corresponding to an obvious peak as seen in Figure 7-10), mode 6 can be only properly estimated when the signal quality is in fairly good condition. To implement the hypothesis testing, the modal frequency data (of the six usable modes) obtained from the two aforementioned portions of the QUT-SHM databank is used to establish the original learning and testing datasets with 64 and 32 observations, respectively. The boxplots of these two datasets, as presented in Figure 7-12(a) and (b), first show that their magnitude distributions are in excellent agreement with each other. Another supporting evidence is that the mode shape agreement across the two sets is very high

with MAC values being frequently higher than 0.9. It is therefore sensible to assume that these two datasets belong to only one structural state. Since no data from another structural state is available with this newly-constructed building, the hypothesis testing is restricted merely to the Type I error tests.

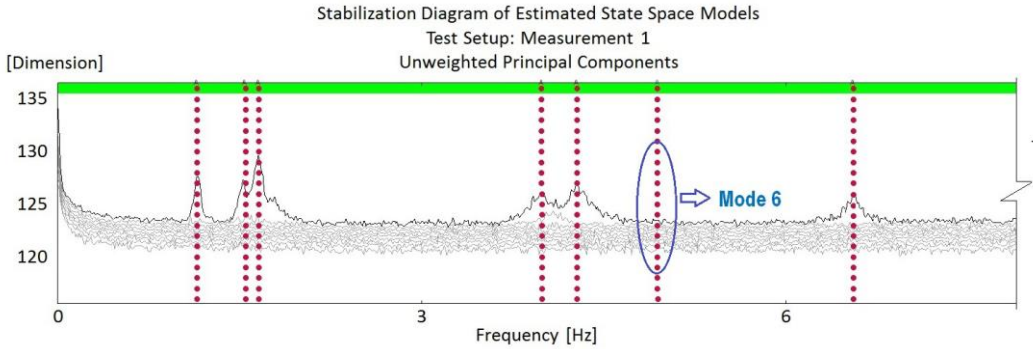


Figure 7-10 Detected modes of QUT-SHM benchmark structure

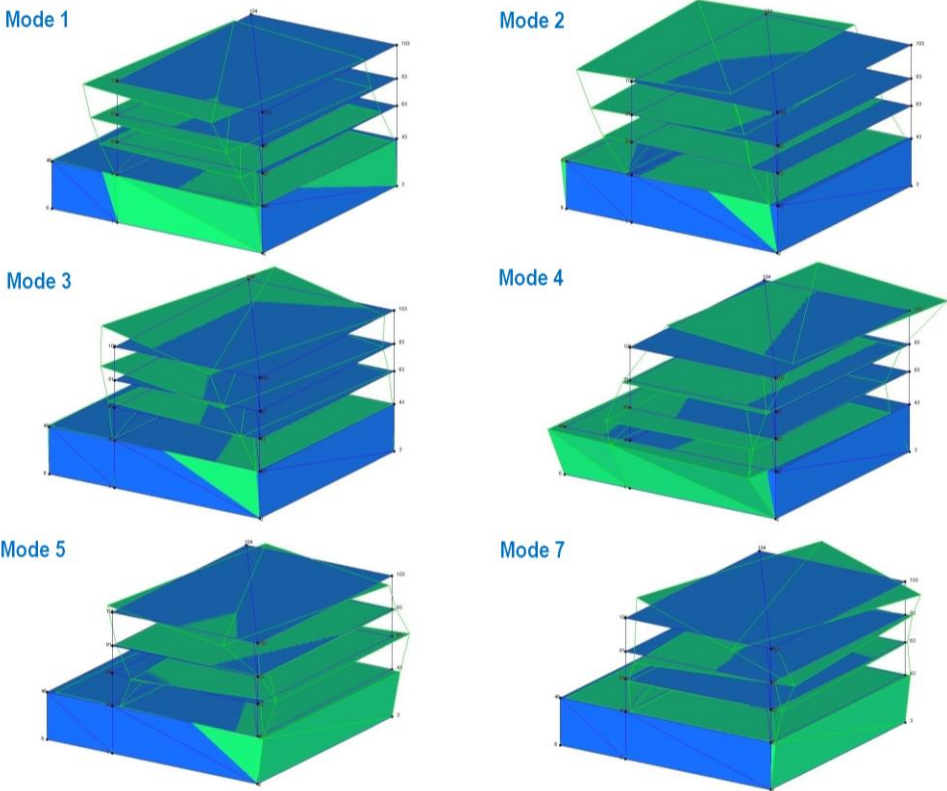


Figure 7-11 Mode shapes of six usable modes of QUT-SHM benchmark structure

Using the same investigation procedure that has been done for the SMC data, the beta Q-Q plot of the original learning data and the Type I testing are conducted for the QUT-SHM data and the results are shown in Figure 7-12(c) and (d). One can see that the agreement between the actual beta Q-Q plot and the theoretical line in this case is slightly better than that of the SMC data. This is reflected by the fact that most of data points in Figure 7-12(c) stay closer to the theoretical line than those data points of the SMC case presented in Figure 7-12(d). The Type I error still comes across but the rate is significantly smaller (than that of the SMC case) with just over 10% false positive detection as illustrated in Figure 7-12(d). To see whether CMCDG could further improve this situation, the same simulation process as for the SMC data is conducted and the results of two simulation rounds in CMCDG are reported in Figure 7-13. Applying again the previous criterion of 95% deviation bounds, the optimal noise level is found at 0.6% and the convergence trends around this level are illustrated in Figure 7-13(a) and (b) for two replication sizes of 7 and 14 blocks, respectively. Employing this noise level and tracking the convergence of both COND and beta Q-Q RMSE from Figure 7-13(c) and (d), one can again select the optimal replication size at 9 blocks. Compared to the optimal noise level (2%) of the SMC data, the optimal level in this case is considerably smaller and a possible reason for this symptom is that the original QUT-SHM learning dataset has better multinormality than that of the SMC bridge structure. This has in fact been reflected through the previous comparison of multinormality degrees (based on the beta Q-Q plots) between two original datasets of the two cases.

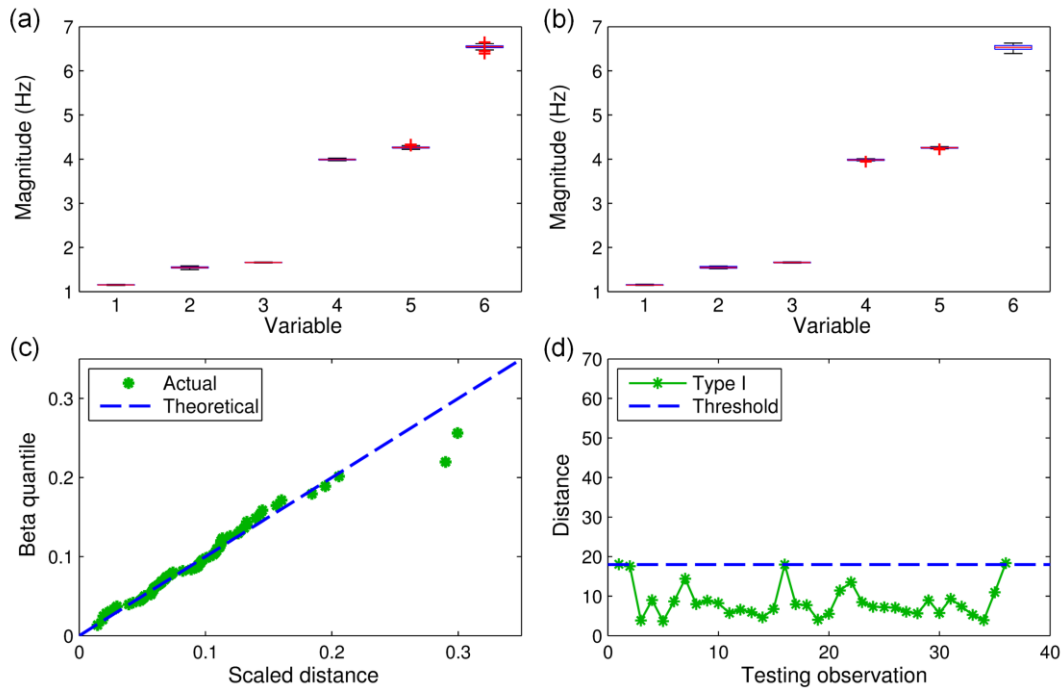


Figure 7-12 Characteristics of QUT-SHM data and original testing results: (a) original learning dataset; (b) testing dataset; (c) beta Q-Q plot of original learning dataset; and (d) original testing results

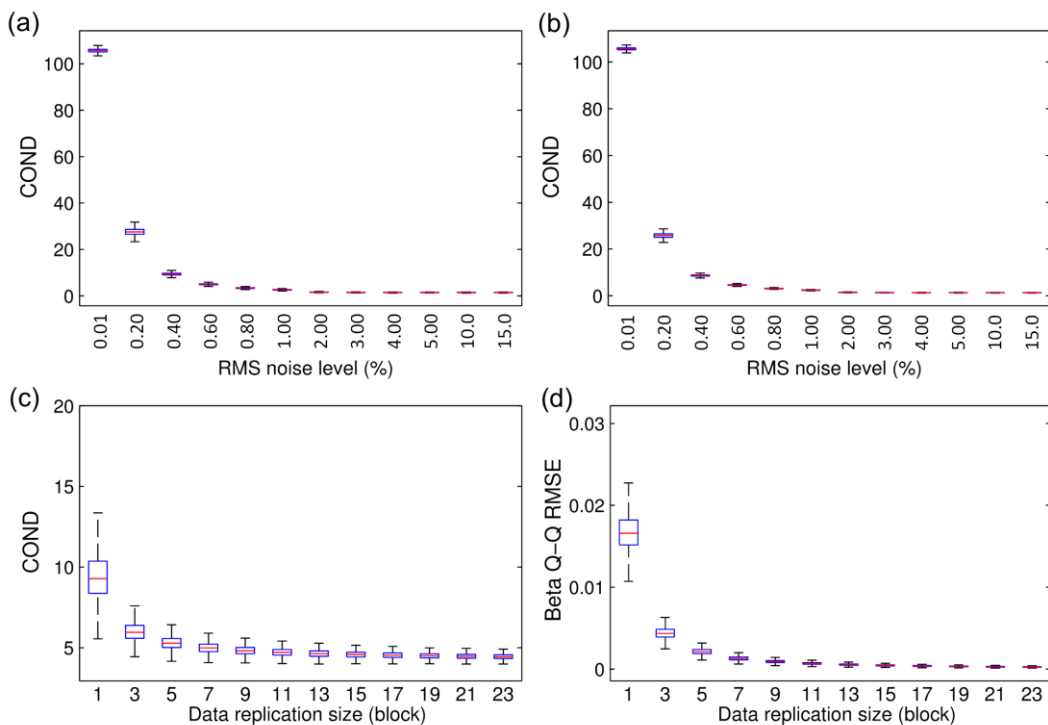


Figure 7-13 Results of two simulation rounds in CMCDG for QUT-SHM data: (a, b) round 1 with COND for 7 and 14 replication block cases; (c, d) round 2 with COND and beta Q-Q RMSE at noise level of 0.6%

For further checking purposes, well-conditioned synthetic datasets are generated by the optimal CMCDG configuration (i.e. noise level of 0.6% and replication size of 9 blocks) previously estimated. Figure 7-14 shows the beta Q-Q plot and the Type I error testing result for a typical one of such datasets whilst Figure 7-15 reports the impact of input perturbation on Type I distance computation based on the same sensitivity analysis procedure as previously conducted for the SMC data. For the latter figure, twelve of the testing observations (i.e. one every three) are selected to fit the paper display space. Once again, improvement can be found for both multinormality and Type I error testing results while the stability of the computational system has been typically improved by 6–10 times by the data optimally enhanced by CMCDG. These results reconfirm the efficacy of the CMCDG scheme in enhancing the condition of learning data and the corresponding computational system realization so that more reliable damage identification outcome can be achieved. Besides, since there is no significant change in the magnitudes of the thresholds between the original learning data and the enhanced data in both SMC (Figure 7-9) and QUT-SHM (Figure 7-15) cases, it can be concluded again that CMCDG does not significantly change the magnitude of feature data as noted in the initial study of this scheme (Nguyen *et al.*, 2014a). Instead, its effectiveness has mainly come from the provision of synthetic observations which are randomly distributed against the original data as led by CLT and LLN theorems and this has been reflected through the irrefutable convergence trends of both COND and beta Q-Q RMSE as previously shown. With the successful applications in two real civil engineering structures herein, the CMCDG scheme can be considered to be successfully validated by field test data.

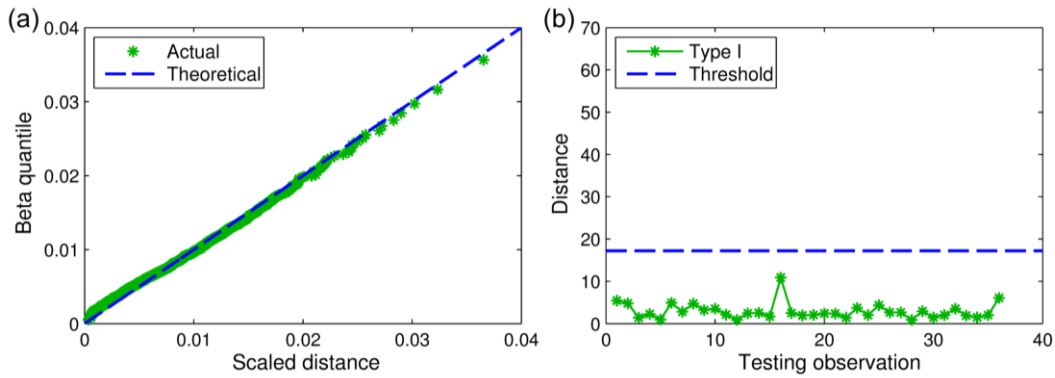


Figure 7-14 (a) beta Q-Q plot and (b) testing results of a typical enhanced (QUT-SHM) learning dataset

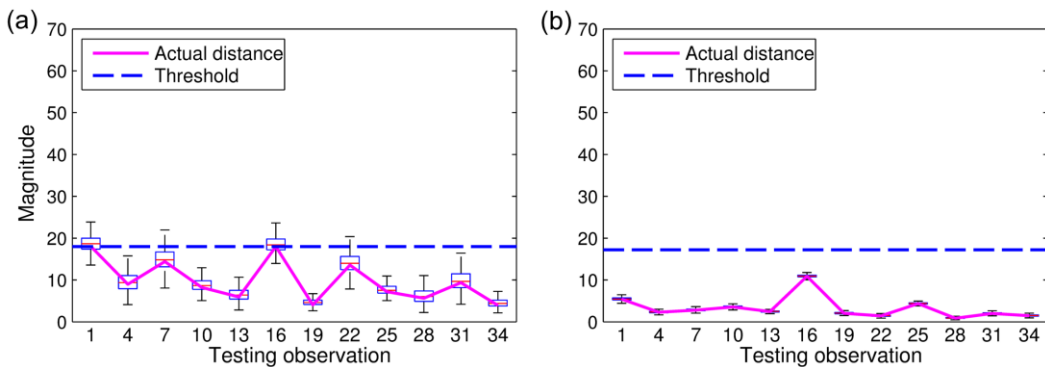


Figure 7-15 Impact of input perturbation on (QUT-SHM) MSD computation: (a) original learning dataset and (b) typical enhanced learning dataset

7.5 CONCLUSIONS

This paper has presented the field applications and validations for the CMCDG scheme recently derived to assist the MSD-based damage identification method to cope with the problem of data shortage which can cause inadequate data multinormality and unstable MSD computation. To do so, two benchmark SHM structures are used in which the bridge represents for the case of having infrequent and/or inadequate quality measurements while the building represents the case where the data shortage problem occurs at an early monitoring stage. Owing to limited availability of actual observations, the original learning dataset of either case has been revealed to be in such poor multinormal distributions that require the data to be enhanced before it can be reliably used for the MSD-based damage identification process. It has also been shown that a blind attempt to use these low-quality data may

result in a significant rate of false positive errors and the severity of this type of errors tends to be proportionate to the poorness of the data multinormality. However, with the enhancement from CMCDG, these problems have been shown to be effectively mitigated. Under optimal data generation configurations derived in CMCDG, well-conditioned synthetic data for the learning process has been generated with remarkable improvements in multinormality degree as well as MSD computational stability. The latter has been critically assessed not only with regards to the initial condition of computationally unstable dataset via COND as in the original work in CMCDG but also with respect to the consequent impact of using such a dataset on the testing results. The ultimate outcome of the applications of CMCDG to the field data herein has reconfirmed that CMCDG is able to overcome the poor data multinormality problem in general and data shortage issues in particular. Under such valuable assistance from CMCDG, the MSD-based damage identification method can deal more effectively and reliably with SHM data recorded from infrequent monitoring mode, right after the completion of the sensing system or even from unfortunately short-lived structures. Finally, since the appropriate noise levels tend to vary from case to case depending on the multinormality degree of the seed data as illustrated with two examples herein, the dynamic structure of CMCDG has apparently made it well adaptive to any data seed with any (original) distributional condition.

Chapter 8: Conclusions and Future Work

8.1 SUMMARY AND CONCLUSIONS

Long-term and frequent vibration monitoring has been seen one of the most suitable approaches that can achieve a timely decision-making process concerning the health status of civil infrastructure including the emergency cases such as possible structural failures. The fact that the tragic collapse of the I-35W Bridge in United States in 2007 occurred in only three months after its most recent inspection has strongly supported this argument. However, implementing such a system in practice is not always feasible due to many obstacles such as the large scale of actual structures, tight budgets, uncertainties of new sensing technologies and ineffective applications. To tackle these problems altogether, this research program has devoted itself towards the development of a practical and reliable synthetic SHM system with two core subsystems namely sensing system and level-1 data-based safety evaluation system. Two vibration sensing technologies namely the SHM-oriented WSN and the Ethernet distributed DAQ platforms have been considered for further improvement before they can be fully accepted into the system. At first, a flexible semi-complete data synchronization scheme to enhance the SHM-oriented WSN efficiency has been derived and then evaluated by means of simulated DSE applied on measured vibration data from two large-scale civil structures. Targeting a cost-effective and flexible sensing solution for continuous vibration monitoring of critical civil infrastructure, the Ethernet-based sensing development work has been realized with multiple cost-effective solutions not only at sensor and DAQ levels but also in the synchronization aspect. Instead of using costly conventional synchronization hardware modules, another inexpensive software-based data synchronization scheme has been derived based on the previous semi-complete data synchronization methodology with adaptation to utilize the TCP/IP communication technology available in the Ethernet connection. To overcome the inherent influence of variable E&O factors, a daisy chain data selection scheme has been derived to assist the statistical assessment work in evaluating the impact of remaining initial DSE in this

wired sensing system. On the side of safety evaluation components, intensive investigations have first been carried out to assess the impact of the remaining DSE on the main modal properties as well as the robustness of the primary SSI-data technique and its auxiliary tools such as data merging or channel projection. Also uncovered was the weakness of the MSD-based method due to its difficulty in satisfying the requirement of data distribution at an early monitoring stage or during short-lived SHM programs. A data generation scheme termed as CMCDG has then been proposed as a prescription for such an illness. Applications not only onto a sophisticated laboratory dataset but also onto ambient structural responses of an actual bridge have been used for validation purposes. Based on the outcomes from extensive evaluations of the above enhancements and new developments, the main findings of this thesis are as follows

- It is feasible to use flexible semi-complete data synchronization schemes for not only SHM-oriented WSNs but also Ethernet-based wired DAQ systems when these measurement systems are used for large-scale civil structures. In the SHM-oriented WSNs, such a scheme has been demonstrated (in Chapters 3 and 4) by simply deactivating the resampling process in the every leaf node and permitting a controllably relaxed DSE level within a sampling period. In the Ethernet distributed DAQ platform, the scheme can be realized by utilizing the TCP/IP communication medium to achieve the quality timing coordination as well as to effectively cut off the accumulation of jitter-related error. In this research, while the feasibility of the former scheme was confirmed by negligible modal results for low frequency ranges such as 0–10 Hz (Chapters 3 and 4), the latter data synchronization scheme has been well developed and fully validated in Chapter 5 thereby providing a promising alternative for use in actual large-scale civil infrastructure.
- Along with the above novel synchronization solution for the Ethernet distributed DAQ platform, an actual cost-effective and flexible sensing system has been developed in Chapter 5 to further illustrate the practicality of the synthetic SHM system targeted in this research. The Ethernet distributed DAQ platform has been found rather ideal for overcoming difficulties associated with the large or sparse measurement coverage problems which are

oftentimes encountered in practice. By carefully selecting the type and optimal positioning for the sensors, number of required vibration sensors can be kept minimum while. The daisy chain data selection scheme proposed in Chapter 5 has also been shown to be a good tool for cooperating with statistical assessment for the purpose of in-situ evaluation of data synchronization methods in the inherent presence of E&O factors.

- For the purpose of cost-effective long-term and frequent SHM, ambient vibration monitoring should be adopted where possible. In order to do a quick-check on whether an ambient loading condition is suitable for such a purpose, a pilot vibration test can be easily and economically conducted using a portable measurement system (such as a SHM-oriented WSN) before a permanent sensing system can be designed. The information obtained from such a preliminary test can then be used for selection of suitable sensor type (as illustrated in Chapter 5) and optimal sensor positions
- OMA techniques in general and the FDD and SSI-data families in particular have become rather mature convenient technologies for vibration analysis. Compared to FDD, SSI-data techniques are more advantageous when dealing with closely spaced or repeated modes and when implementing the automated modal identification process. The FDD technique is previously known for its advantage of working directly with spectral peaks and therefore being very robust against noise-type uncertainties including DSE. In this regard, the robustness of the primary SSI-data technique can, as reflected in this research, be more or less the same as that of FDD when it is assisted by appropriate channel projection options and PoSER data merging method. The PoSER data merging method, though being more time-consuming in the case of using a large number of multiple setups, has been found more robust against DSE than the PreGER counterpart particularly with the modes that are not well excited. For the channel projection problem, it should be noted that using a small number of projected channels may denoise more effectively for some modes but the associated risk is that more spurious computational modes may arrive. In this issue, using multiple adjacent projection setups as mentioned in

Chapters 5 and 7 can assist in finding the most stable projection setup as well as to distinguish the spurious modes from the genuine ones.

- As the outcome of OMA technology, output-only modal parameters are convenient health-representative and damage-sensitive features particularly for long-term and frequent monitoring purposes. Compared to mode shapes, modal frequencies are more robust with respect to DSE as well as to other E&O factors and therefore deserve to be one of the main level-1 damage indices. On the other hand, the impact of initial DSE on mode shapes is in general predictable and can therefore be compensated for the purpose of maintaining the accuracy of further applications to this type of features.
- Owing to its architectural simplicity, MSD-based statistical damage identification can be amongst the most computationally efficient methods and tends to be immune to the uncertainties of learning architectural assumptions. Such advantages make MSD-based method one of the best candidates for the purpose of long-term and/or real-time safety evaluations. MSD-based method has however had a major weakness associated with its requirement of data distribution. If a learning dataset has a low multinormality degree, its computational system realization will suffer from severe computational instability which can, as reflected in Chapters 6 and 7, lead to unreliable damage detection outcome. In this regard, the ultimate assistance from the CMCDG scheme proposed in Chapter 6 has been shown to be able to assist the MSD-based method to overcome the problem and achieve satisfactory damage identification outcome. Not only should COND and beta Q-Q plot/RMSE be employed to examine the multinormality degree of the input data, but statistical sensitivity analysis should also be frequently used as demonstrated in Chapter 7 to evaluate the robustness of the MSD computational solution. Such computational checks ensure the SHM system will yield reliable outcome of structural health.
- The dynamic structure of CMCDG has made it well adaptive to any data seed with any primary distributional condition. Compared to the traditional Monte Carlo data generation approach, the CMCDG scheme has several important

advantages. First, the noise level and replication size are determined in a controlled manner in which the former can be optimally estimated while the latter can be kept at minimum. This way, not only the synthetic database can attain the sufficient multinormality degree but it also can have a modest size. The database may therefore be capable of being implemented along with the MSD-based method in embedded diagnostic systems such as SHM-oriented WSNs which generally have very limited computational resources.

With the above validated developments and enhancements, the developed synthetic SHM system is expected to operate efficiently and reliably in every monitoring stage. The cost-effectiveness and flexibility of the two targeted vibration sensing platforms (SHM-oriented WSNs and Ethernet distributed DAQ systems) have been enhanced by not only effective solutions of sensor type, sensor placement and peripheral DAQ but also by novel inexpensive and flexible data synchronization schemes. Reliability and efficiency of the safety evaluation system have been improved through (i) robust and reliable feature extraction by means of powerful primary SSI-data with valued cooperation of the PoSER data merging method as well as useful channel projection scheme; (ii) computational efficiency of MSD-based damage identification for dealing with large volume of data in later monitoring stages; and (iii) the valuable optimal assistance from CMCDG and its versatility in enhancing the computational reliability of the MSD-based method under adverse circumstances like at an early monitoring stage. Using the Ethernet-based sensing system developed in Chapter 5, the instrumented building at QUT can serve as a flexible benchmark structure not only for vibration-based health monitoring problems but also for more novel synchronization solutions. Building vibration data is being continuously collected and analyzed under different E&O conditions to construct representative databases for tracking the health status of the structure by means of the level-1 safety evaluation paradigm developed herein.

8.2 FUTURE WORK

Although the present research program has carried out intensive investigations, developments and validations in the areas of vibration sensing, feature extraction and damage identification technologies; the proposed subsystems have been restricted to

on-module timebase, output-only modal properties and within the level-1 assessment paradigm employing MSD-based unsupervised damage identification method. In order to build more comprehensive damage diagnostic system as well as more effective sensing measurement systems, additional research may be extended based on the achievements in the present research. Recommendations for future work include the following.

- Employing FPGA timebases for data synchronization in Ethernet distributed DAQ system: Besides the input module, each FPGA has its own timebase which can be used to provide controlling of peripheral sampling process. This option will be investigated in future.
- Real-time or near real-time output-only frequency estimations: The recursive version of SSI-data is capable of conducting online automated frequency estimation. This way, the SHM system can work autonomously without user interaction. Another way is to use the powerful SSI-data family in the batch mode to achieve near real-time manners.
- Level-2 damage identification: After a damage occurrence is confirmed by level-1 methods such as the MSD-based herein, narrowing down the damage location is often desirable. In this regard, using signal-based features such as those based on AR, ARMA and wavelet coefficients or mode shape related features such as the Coordinate Modal Assurance Criterion (COMAC) can do the job to some extent. In the case of the latter, as COMAC tends to be sensitive to E&O impact present in ambient vibration data, a robust statistical program should be developed for the purpose of outlier screening. In the case of the former, robust data normalization or transformation schemes are often in need to be constructed to eliminate the (instantaneous and/or long-term) non-stationarity and measurement noise particularly when they are at severe levels. As they are often in the multivariate data type, signal-based features would still need statistical unsupervised learning techniques such as MSD-based or AANN-based techniques to deal with E&O factors.

- Damage identification at higher levels: Estimating damage extent, though theoretically feasible through the application of supervised learning, is often a very challenging task as the data of damage states is often unavailable beforehand. One common way to overcome this problem is to use the (experimentally) updated FE model of the structure with simulated damage scenarios. Even though it is often said that the simulated damage by no means can be the same as actual damage, the rapid development in FE modelling technologies has continually led to advanced material models that could reflect more closely actual damage mechanisms. As for the QUT-SHM team, the FE model of the P block after being correlated with the as-built drawings will continue to be updated with experimental modal parameters as estimated in Chapter 5 before different damage scenarios can be simulated for the purpose of supervised damage identification validation.
- Efficacy of CMCDG towards other unsupervised (and also supervised) damage identification methods: Although other unsupervised damage identification methods such as AANN and factor analysis do not require multinormal data, they still tend to require a sufficiently large number of random observations (of what the actually measured data may just be a fraction) to reach a stable point in the training process. In this sense, the CMCDG can also be used to generate synthetic random data for such a purpose. It may therefore be desirable to investigate the efficacy of CMCDG or its optimal configuration for each method. One can also see that this issue can be extended to the supervised methods as long as they use multivariate data.

Bibliography

- Aktan, A. E., Catbas, F. N., Grimmelsman, K. A. and Pervizpour, M. 2003. *Development of a Model Health Monitoring Guide for Major Bridges*: Drexel Intelligent Infrastructure and Transportation Safety Institute, Philadelphia, PA, USA. Available online at <http://www.di3.drexel.edu/DI3/Events/Paper-Presentation/FHWAGuideFull-web.pdf> (accessed August, 2013).
- Allemang, R. J. 2003. The modal assurance criterion—twenty years of use and abuse. *Sound and Vibration* 37 (8):14-23.
- Ansari, F. 2005. *Sensing Issues in Civil Structural Health Monitoring*. Dordrecht, The Netherlands: Springer.
- Brincker, R. and Andersen, P. 2006. Understanding stochastic subspace identification. In *Proceedings of the 24th International Modal Analysis Conference, St. Louis, MO, USA, January 30 - February 2*: SEM.
- Brincker, R., Andersen, P. and Cantieni, R. 2001. Identification and level I damage detection of the Z24 highway bridge. *Experimental Techniques* 25 (6):51-57.
- Brincker, R., Ventura, C. and Andersen, P. 2003. Why output-only modal testing is a desirable tool for a wide range of practical applications. In *Proceedings of the 21st International Modal Analysis Conference, Kissimmee, FL, USA, February 3-6*, pp. 1-8: SEM.
- Brincker, R., Zhang, L. and Andersen, P. 2000. Modal identification from ambient responses using frequency domain decomposition. In *Proceedings of the 18th International Modal Analysis Conference, San Antonio, TX, USA*, pp. 625-630: SEM.
- Cantieni, R. 2005. Experimental methods used in system identification of civil engineering structures. In *Proceedings of the 1st International Operational Modal Analysis Conference, Copenhagen, Denmark*, pp. 249–60.
- Carden, P. and Brownjohn, J. M. W. 2008. ARMA modelled time-series classification for structural health monitoring of civil infrastructure. *Mechanical Systems and Signal Processing* 22 (2):295-314.
- cFos Software GmbH. 2013. *Ping Utility hrPING v5.06*. cFos Software GmbH, Bonn, Germany. Available online at <http://www.cfos.de/en/ping/ping.htm> (accessed December, 2013).
- Chan, T. H. T., Wong, K. Y., Li, Z. X. and Ni, Y. Q. 2011. Structural Health Monitoring for Long Span Bridges - Hong Kong Experience & Continuing onto Australia. In *Chapter 1 in Structural Health Monitoring in Australia*, eds.

- T. H. T. Chan and D. P. Thambiratnam. New York: Nova Science Publishers, Inc.
- Chen, W. H., Lu, Z. R., Lin, W., Chen, S. H., Ni, Y. Q., Xia, Y. and Liao, W. Y. 2011. Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures* 33 (12):3628-3646.
- Chintalapudi, K., Fu, T., Paek, J., Kothari, N., Rangwala, S., Caffrey, J., Govindan, R., Johnson, E. and Masri, S. 2006. Monitoring civil structures with a wireless sensor network. *Internet Computing, IEEE* 10 (2):26-34.
- Cho, S., Jo, H., Jang, S., Park, J., Jung, H. J., Yun, C. B., Spencer, B. and Seo, J. W. 2010. Structural health monitoring of a cable-stayed bridge using wireless smart sensor technology: data analyses. *Smart Structures and Systems* 6 (5-6):461-480.
- Cigada, A., Moschioni, G., Vanali, M. and Caprioli, A. 2010. The measurement network of the san siro meazza stadium in milan: Origin and implementation of a new data acquisition strategy for structural health monitoring. *Experimental Techniques* 34 (1):70-81.
- Devriendt, C., Magalhaes, F., El Kafafy, M., De Sitter, G., Cunha, A. and Guillaume, P. 2013. Long-term dynamic monitoring of an offshore wind turbine. In *Proceedings of the 31st International Modal Analysis Conference, Garden Grove, CA, USA, February 11-14*, pp. 253-267: Springer.
- Dilena, M. and Morassi, A. 2009. Structural health monitoring of rods based on natural frequency and antiresonant frequency measurements. *Structural Health Monitoring* 8 (2):149-173.
- Doebling, S. W., Farrar, C. R. and Prime, M. B. 1998. Summary review of vibration-based damage identification methods. *Shock and Vibration Digest* 30 (2):91-105.
- Doebling, S. W., Farrar, C. R., Prime, M. B. and Shevitz, D. W. 1996. *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review*: Los Alamos National Laboratory, Los Alamos, NM, USA.
- Dohler, M., Andersen, P. and Mevel, L. 2010. Data merging for multi-setup operational modal analysis with data-driven SSI. In *Proceedings of the 28th International Modal Analysis Conference, Jacksonville, FL, USA*, pp. 443-452: SEM.
- Dohler, M., Andersen, P. and Mevel, L. 2012. Operational modal analysis using a fast stochastic subspace identification method. In *Proceedings of the 30th International Modal Analysis Conference 2012, , Jacksonville, FL, USA, January 30 - February 2*, pp. 19-24: Springer.

- Dunn, W. L. and Shultis, J. K. 2011. *Exploring Monte Carlo Methods*. Burlington: Elsevier.
- Eren, H. 2011. *Instrument Engineers' Handbook: Vol. 3: Process Software and Digital Networks*. Florence: CRC Press [Imprint].
- Farrar, C. R., Doebling, S. W. and Nix, D. A. 2001a. Vibration-based structural damage identification. *Philosophical Transactions of the Royal Society A* 359 (1778):131-149.
- Farrar, C. R. and Jauregul, D. A. 1998. Comparative study of damage identification algorithms applied to a bridge: I. Experiment. *Smart Materials and Structures* 7 (1998):704-19.
- Farrar, C. R., Sohn, H. and Worden, K. 2001b. Data Normalization: A Key for Structural Health Monitoring. In *Proceedings of the 3rd International Workshop on Structural Health Monitoring, Stanford, CA, USA, September, 17-19*, pp. 1229-1238: Technomic Publishing Company Inc.
- Farrar, C. R. and Worden, K. 2013. *Structural health monitoring: a machine learning perspective*. Chichester, West Sussex: Wiley.
- Figueiredo, E., Park, G., Farrar, C. R., Worden, K. and Figueiras, J. 2011. Machine learning algorithms for damage detection under operational and environmental variability. *Structural Health Monitoring* 10 (6):559-572.
- Figueiredo, E., Park, G., Figueiras, J., Farrar, C. and Worden, K. 2009. *Structural health monitoring algorithm comparisons using standard data sets*: Los Alamos National Laboratory, Los Alamos, NM, USA.
- Filzmoser, P., Garrett, R. G. and Reimann, C. 2005. Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences* 31 (5):579-587.
- Gao, Y. 2005. Structural health monitoring strategies for smart sensor networks. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- Golub, G. H. and Van Loan, C. F. 1996. *Matrix computations*. 3rd ed. Baltimore: Johns Hopkins University Press.
- Gul, M. and Catbas, F. N. 2009. Statistical pattern recognition for structural health monitoring using time series modeling: theory and experimental verifications. *Mechanical Systems and Signal Processing* 23 (7):2192-204.
- Herlufsen, H., Andersen, P., Gade, S. and Møller, N. 2005. Identification Techniques for Operational Modal Analysis—An Overview and Practical Experiences. In *Proceedings of the 1st International Operational Modal Analysis Conference, Copenhagen, Denmark*, pp. 1–13.

- Hilkevitch, J. 2010. *High-tech sensors help spot potentially fatal problems with bridges*: Chicago Tribune. Available online at http://articles.chicagotribune.com/2010-10-17/classified/ct-met-getting-around-1018-20101017_1_interstate-highway-35w-bridge-bridge-piers-bridge-safety-experts (accessed December, 2013).
- Horyna, T. and Venture, C. E. 2000. Summary of HCT building ambient vibration data analyses. In *Proceedings of the 18th International Modal Analysis Conference, San Antonio, TX, USA*, pp. 1095-1098: SEM.
- Hristu-Varsakelis, D. and Levine, W. S. 2005. *Handbook of networked and embedded control systems*. Boston, Mass: Birkhäuser Boston.
- Jang, S., Jo, H., Cho, S., Mechitov, K., Rice, J. A., Sim, S.-H., Jung, H.-J., Yun, C.-B., Spencer Jr, B. F. and Agha, G. 2010a. Structural health monitoring of a cable-stayed bridge using smart sensor technology: deployment and evaluation. *Smart Structures and Systems* 6 (5-6):439-459.
- Jang, S., Sim, S. H., Jo, H. and Spencer, B. F. 2010b. Decentralized bridge health monitoring using wireless smart sensors. In *Proceedings of Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2010, San Diego, CA, USA, March 8-11 SPIE*.
- Jo, H., Sim, S.-H., Nagayama, T. and Spencer, B. F. 2012. Development and Application of High-Sensitivity Wireless Smart Sensors for Decentralized Stochastic Modal Identification. *Journal of Engineering Mechanics* 138 (6):683-694.
- Johnson, R. A. and Wichern, D. W. 2002. *Applied multivariate statistical analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.
- Karbhari, V. M. and Ansari, F. 2009. *Structural health monitoring of civil infrastructure systems*. Cambridge, UK: Woodhead Publishing.
- Kim, C.-Y., Jung, D.-S., Kim, N.-S., Kwon, S.-D. and Feng, M. Q. 2003. Effect of vehicle weight on natural frequencies of bridges measured from traffic-induced vibration. *Earthquake Engineering and Engineering Vibration* 2 (1):109-115.
- Ko, J. M. and Ni, Y. Q. 2005. Technology developments in structural health monitoring of large-scale bridges. *Engineering Structures* 27 (12):1715-1725.
- Krishnamurthy, V., Fowler, K. and Sazonov, E. 2008. The effect of time synchronization of wireless sensors on the modal analysis of structures. *Smart Materials and Structures* 17 (5):1-13 (18).
- Li, N., Feng, L., Zhou, B., Sun, X., Liang, Z. and Xie, H. 2010. A study on a cRIO-based monitoring and safety early warning system for large bridge structures. In *Proceedings of International Conference on Measuring Technology and*

Mechatronics Automation, Changsha, China, March 13 - 14, pp. 361-364: IEEE.

- Li, N., Zhang, X. Y., Zhou, X. T., Leng, J., Liang, Z., Zheng, C. and Sun, X. F. 2008. Introduction of structural health and safety monitoring warning systems for Shenzhen-Hongkong western corridor Shenzhen bay bridge. In *Proceedings of Health Monitoring of Structural and Biological Systems 2008, San Diego, CA, USA, March 10-13*: SPIE.
- Li, S., Li, H., Liu, Y., Lan, C., Zhou, W. and Ou, J. 2014. SMC structural health monitoring benchmark problem using monitored data from an actual cable-stayed bridge. *Structural Control and Health Monitoring* 21 (2):156-172.
- Libelium. 2010. Waspnote technical guide (Document version: v0.9 - 07/2010): Libelium Comunicaciones Distribuidas S.L., Zaragoza, Spain.
- Linderman, L. E., Mechitov, K. A. and Spencer, B. F. 2011. *Real-Time Wireless Data Acquisition for Structural Health Monitoring and Control*. Vol. 29, *NSEL Report*: University of Illinois at Urbana Champaign, Champaign, IL, USA.
- Linderman, L. E., Mechitov, K. A. and Spencer Jr, B. F. 2013. TinyOS-based real-time wireless data acquisition framework for structural health monitoring and control. *Structural Control and Health Monitoring* 20 (6):1007-1020.
- Ljung, L. 2011. System Identification Toolbox™ 7 User's Guide. Natick, MA: MathWorks.
- Loh, C.-H., Weng, J.-H., Liu, Y.-C., Lin, P.-Y. and Huang, S.-K. 2011. Structural damage diagnosis based on on-line recursive stochastic subspace identification. *Smart Materials and Structures* 20 (5):1-10 (055004).
- Lynch, J. P. and Loh, K. J. 2006. A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest* 38 (2):91-128.
- Lynch, J. P., Wang, Y., Law, K. H., Yi, J. H., Lee, C. G. and Yun, C. B. 2005. Validation of a large-scale wireless structural monitoring system on the Geumdang Bridge. In *Proceedings of the 9th International Conference on Safety and Structural Reliability (ICOSSAR), Rome, Italy, June, 19-23*.
- Manson, G., Worden, K. and Allman, D. 2003. Experimental validation of a structural health monitoring methodology. Part II. Novelty detection on a Gnat aircraft. *Journal of Sound and Vibration* 259 (2):345-363.
- Markou, M. and Singh, S. 2003a. Novelty detection: A review - Part 1: Statistical approaches. *Signal Processing* 83 (12):2481-2497.

- Markou, M. and Singh, S. 2003b. Novelty detection: A review - Part 2: Neural network based approaches. *Signal Processing* 83 (12):2499-2521.
- Martinez, W. L. and Martinez, A. R. 2002. *Computational statistics handbook with MATLAB*. Boca Raton: CRC Press.
- Martinez, W. L. and Martinez, A. R. 2005. *Exploratory data analysis with MATLAB*. Boca Raton, Florida: Chapman & Hall/CRC.
- MathWorks. 2011. MATLAB® R2011a Help Browser. Natick, MA, USA: MathWorks
- McDonald, C. 2012. *Performing Structural Health Monitoring of the 2008 Olympic Venues Using NI LabVIEW and CompactRIO*: National Instruments, Austin, TX, USA. Available online at <http://sine.ni.com/cs/app/doc/p/id/cs-11279> (accessed December, 2013).
- Mechitov, K., Kim, W., Agha, G. and Nagayama, T. 2004. High-frequency distributed sensing for structure monitoring. In *Proceedings of the First International Conference on Networked Sensing Systems, Tokyo, Japan, June, 22-23*.
- Montgomery, D. C. 2005. *Introduction to statistical quality control*. Hoboken, N.J: John Wiley.
- Moser, P. and Moaveni, B. 2013. Design and deployment of a continuous monitoring system for the dowling hall footbridge. *Experimental Techniques* 37 (1):15-26.
- Nagayama, T., Sim, S. H., Miyamori, Y. and Spencer, B. F. 2007. Issues in structural health monitoring employing smart sensors. *Smart Structures and Systems* 3 (3):299-320.
- Nagayama, T., Spencer, B. F., Mechitov, K. A. and Agha, G. A. 2009. Middleware services for structural health monitoring using smart sensors. *Smart Structures and Systems* 5 (2):119-37.
- Nagayama, T. and Spencer Jr., B. F. 2007. *Structural Health Monitoring Using Smart Sensors*. Vol. 001, *NSEL Report* University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- National Instruments. 2012a. *Introduction to NI CompactRIO for Structural Health Monitoring*: National Instruments, Austin, TX, USA. Available online at <http://www.ni.com/white-paper/8427/en/> (accessed December, 2013).
- National Instruments. 2012b. *NI 9229/9239 Operating instructions and specifications* National Instruments, Austin, TX, USA. Available online at <http://www.ni.com/pdf/products/us/922939ds.pdf> (accessed December 2013).

- National Instruments. 2012c. *Synchronizing multiple CompactRIO chassis* National Instruments, Austin, TX, USA. Available online at <http://www.ni.com/white-paper/4217/en/> (accessed December, 2013).
- Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014a. Controlled Monte Carlo data generation for statistical damage identification employing Mahalanobis squared distance. *Structural Health Monitoring*: (In press); doi: 10.1177/1475921714521270.
- Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014b. Effects of wireless sensor network uncertainties on output-only modal-based damage identification. *Australian Journal of Structural Engineering* 15 (1):15-25.
- Nguyen, T., Chan, T. H. T. and Thambiratnam, D. P. 2014c. Effects of wireless sensor network uncertainties on output-only modal analysis employing merged data of multiple tests. *Advances in Structural Engineering* 17 (3):319-330.
- Ni, Y. Q., Xia, Y., Liao, W. Y. and Ko, J. M. 2009. Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. *Structural Control and Health Monitoring* 16 (1):73-98.
- Ni, Y. Q., Xia, Y., Lin, W., Chen, W. H. and Ko, J. M. 2012. SHM benchmark for high-rise structures: a reduced-order finite element model and field measurement data. *Smart Structures and Systems* 10 (4):411-426.
- Niu, Y., Kraemer, P. and Fritzen, C. P. 2012. Operational modal analysis for a benchmark high-rise structure. *Smart Structures and Systems* 10 (4):393-410
- Overschee, P. V. and Moor, B. D. 1996. *Subspace Identification for the Linear Systems: Theory–Implementation–Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Pakzad, S. N., Fenves, G. L., Kim, S. and Culler, D. E. 2008. Design and implementation of scalable wireless sensor network for structural monitoring. *Journal of Infrastructure Systems* 14 (1):89-101.
- Pannell, D. J. 1997. Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics* 16 (2):139-152.
- Peeters, B. and Ventura, C. E. 2003. Comparative study of modal analysis techniques for bridge dynamic characteristics. *Mechanical Systems and Signal Processing* 17 (5):965-988.
- Reda Taha, M. M., Noureldin, A., Lucero, J. L. and Baca, T. J. 2006. Wavelet transform for structural health monitoring: A compendium of uses and features. *Structural Health Monitoring* 5 (3):267-295.
- Reid, R. L. 2008. The infrastructure crisis. In *Civil engineering*.

- Rencher, A. C. 2002. *Methods of multivariate analysis*. 2nd ed. New York: John Wiley & Sons.
- Reynders, E., Magalhaes, F., Roeck, G. and Cunha, A. 2009. Merging strategies for multi-setup operational modal analysis: application to the Luiz I steel arch bridge. In *Proceedings of the 27th International Modal Analysis Conference, Orlando, FL, USA, February 9-12*: SEM.
- Rice, J. A. and Spencer, B. F. 2009. *Flexible smart sensor framework for autonomous full-scale structural health monitoring*. Vol. 18, *NSEL Report*. Illinois, USA: University of Illinois at Urbana-Champaign.
- Ruiz-Sandoval, M. E., Nagayama, T. and Spencer, B. F. 2006. Sensor development using Berkeley Mote platform. *Journal of Earthquake Engineering* 10 (2):289-309.
- Salawu, O. S. 1997. Detection of structural damage through changes in frequency: A review. *Engineering Structures* 19 (9):718-723.
- Saltelli, A., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D. and Ratto, M. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, West Sussex: Wiley.
- Semancik, J. 2004. Ethernet-based instrumentation for ATE. In *Proceedings of the IEEE AUTOTESTCON 2004, San Antonio, TX, USA, September, 2004*, pp. 316-320: IEEE.
- Sharma, S. 1995. *Applied multivariate techniques*. New York: John Wiley & Sons.
- SHMTools Development Team. 2010. *SHMTools and mFUSE*: Los Alamos National Laboratory, Los Alamos, NM, USA. Available online at <http://institute.lanl.gov/ei/software-and-data/SHMTools/> (accessed August, 2012).
- Sim, S. H. 2011. Decentralized identification and multimetric monitoring of civil infrastructure using smart sensors. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- Sohn, H., Farrar, C., Hemez, F., Shunk, D., Stinemates, D. and Nadler, B. 2003. *A review of structural health monitoring literature: 1996-2001*: Los Alamos National Laboratory, Los Alamos, NM, USA.
- Sohn, H., Farrar, C., Hunter, N. and Worden, K. 2001. *Applying the LANL statistical pattern recognition paradigm for structural health monitoring to data from a surface-effect fast patrol boat*: Los Alamos National Laboratory, Los Alamos, NM, USA.
- Sohn, H. and Farrar, C. R. 2001. Damage diagnosis using time series analysis of vibration signals. *Smart Materials and Structures* 10 (3):446-451.

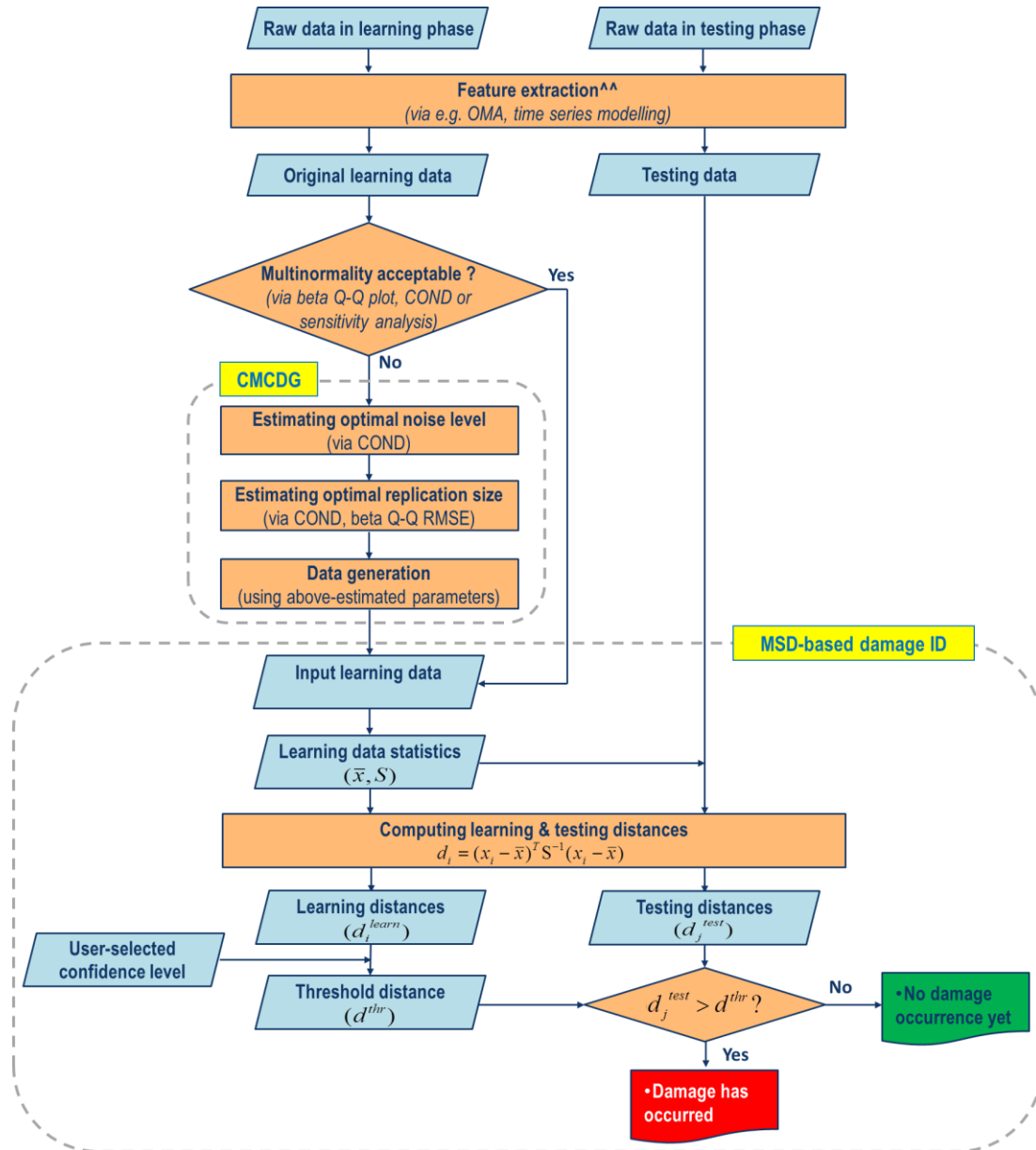
- Soyoz, S. and Feng, M. Q. 2009. Long-term monitoring and identification of bridge structural parameters. *Computer-Aided Civil and Infrastructure Engineering* 24 (2):82-92.
- Spencer, B. F., Ruiz-Sandoval, M. E. and Kurata, N. 2004. Smart sensing technology: Opportunities and challenges. *Structural Control and Health Monitoring* 11 (4):349-368.
- Statistics Toolbox Development Team. 2011. Statistics Toolbox™ 7 User's Guide. Natick, MA: MathWorks.
- Strang, G. 2006. *Linear algebra and its applications*. 4th ed. Belmont, CA: Thomson Brooks/Cole.
- Structural Vibration Solutions A/S. 2011. ARTEMIS Extractor, Release 5.3, User's Manual. Aalborg East, Denmark: Structural Vibration Solutions A/S.
- Su, J.-Z., Xia, Y., Chen, L., Zhao, X., Zhang, Q.-L., Xu, Y.-L., Ding, J.-M., Xiong, H.-B., Ma, R.-J., Lv, X.-L., et al. 2013. Long-term structural performance monitoring system for the Shanghai Tower. *Journal of Civil Structural Health Monitoring* 3 (1):49-61.
- Tamura, Y., Zhang, L., Yoshida, A., Cho, K., Nakata, S. and Naito, S. 2002. Ambient vibration testing and modal identification of an office building. In *Proceedings of the 20th International Modal Analysis Conference, Los Angeles, CA, USA*, pp. 141-146: SEM.
- Thomopoulos, N. T. 2012. *Essentials of Monte Carlo Simulation : Statistical Methods for Building Simulation Models*. Dordrecht: Springer New York.
- Van Der Auweraer, H. and Peeters, B. 2003. Sensors and systems for structural health monitoring. *Journal of Structural Control* 10 (2):117-125.
- Wang, Z. and Ong, K. C. G. 2008. Autoregressive coefficients based Hotelling's T2 control chart for structural health monitoring. *Computers and Structures* 86 (19-20):1918-1935.
- Wang, Z. and Ong, K. C. G. 2010. Multivariate statistical approach to structural damage detection. *Journal of Engineering Mechanics* 136 (1):12-22.
- Weng, J.-H., Loh, C.-H., Lynch, J. P., Lu, K.-C., Lin, P.-Y. and Wang, Y. 2008. Output-only modal identification of a cable-stayed bridge using wireless monitoring systems. *Engineering Structures* 30 (7):1820-1830.
- Whelan, M. 2009. In-service highway bridge condition assessment using high-rate real-time wireless sensor networks. PhD thesis, Clarkson University, Potsdam, NY, United States.

- Wikipedia contributors. 2001. Ping (networking utility). *Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/wiki/Ping_%28networking_utility%29 (accessed December, 2013).
- Wikipedia contributors. 2002. Monte Carlo method. *Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/wiki/Monte_Carlo_method (accessed August, 2012).
- Worden, K., Farrar, C. R., Manson, G. and Gyuhae, P. 2007. The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society of London, Series A (Mathematical, Physical and Engineering Sciences)* 463 (2082):1639-1664.
- Worden, K. and Manson, G. 2007. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A* 365 (1851):515-37.
- Worden, K., Manson, G. and Allman, D. 2003. Experimental validation of a structural health monitoring methodology. Part I. Novelty detection on a laboratory structure. *Journal of Sound and Vibration* 259 (Copyright 2003, IEE):323-43.
- Worden, K., Manson, G. and Fieller, N. R. J. 2000a. Damage detection using outlier analysis. *Journal of Sound and Vibration* 229 (3):647-667.
- Worden, K., Pierce, S. G., Manson, G., Philp, W. R., Staszewski, W. J. and Culshaw, B. 2000b. Detection of defects in composite plates using Lamb waves and novelty detection. *International Journal of Systems Science* 31 (11):1397-1409.
- Worden, K., Sohn, H. and Farrar, C. R. 2002. Novelty detection in a changing environment regression and interpolation approaches. *Journal of Sound and Vibration* 258 (4):741-61.
- Xu, Y.-L. and Xia, Y. 2012. *Structural health monitoring of long span suspension bridges*. New York: Spon Press.
- Yan, A. M., Kerschen, G., De Boe, P. and Golinval, J. C. 2005. Structural damage diagnosis under varying environmental conditions - Part I: A linear analysis. *Mechanical Systems and Signal Processing* 19 (4):847-864.
- Yan, G. and Dyke, S. J. 2010. Structural damage detection robust against time synchronization errors. *Smart Materials and Structures* 19 (6):1-14.
- Yan, Y. J., Cheng, L., Wu, Z. Y. and Yam, L. H. 2007. Development in vibration-based structural damage detection technique. *Mechanical Systems and Signal Processing* 21 (5):2198-2211.

- Yang, N. and Kranjc, M. 2010. Civionics: The Modern Approach to Structural Test and Monitoring. In *Proceedings of the 6th Australasian Congress on Applied Mechanics, Perth, Western Australia, Australia*, pp. 258-266: Engineers Australia.
- Zhang, L., Brincker, R. and Andersen, P. 2005. An overview of operational modal analysis: major development and issues. In *Proceedings of the 1st International Operational Modal Analysis Conference, Copenhagen, Denmark*, pp. 179-190.
- Zhang, Q. W. 2007. Statistical damage identification for bridges using ambient vibration data. *Computers and Structures* 85 (7-8):476-485.

Appendices

APPENDIX A: FLOWCHART OF MSD-BASED DAMAGE IDENTIFICATION EMPLOYING CMCDG



^^ For quality ambient vibration feature extraction, it is suggested that (Structural Vibration Solutions A/S, 2011; Cantieni, 2005):

- (1) minimum history data length be between 1000-2000 times the fundamental period
- (2) sampling rate be at least three times the highest frequency of interest
- (3) data condition be checked in time domain for clipping, spikes and drop-outs; and in frequency domain for excitation quality and/or measurement noise impact. Data with good signal/noise ratio will often result in sufficiently clear spectral peaks and the detected modes being well corresponding to these peaks (see, for instance, Figures 5.5 and 7.4)

APPENDIX B: FUNDAMENTALS OF MONTE CARLO SIMULATION AND MONTE CARLO DATA GENERATION

Monte Carlo simulation

Monte Carlo simulation (or Monte Carlo method) is generally referred to a procedure to solve complex mathematical problems that can be very difficult (or impossible) to be solved using closed-form techniques. The method is based on running the model many times as in random sampling (Thomopoulos, 2012; Dunn and Shultis, 2011). For each sample, random variates are generated on each input variable; computations are run through the model yielding random outcomes on each output variable. Since each input is random, the outcomes are random. In the same way, they generated thousands of such samples and achieved thousands of outcomes for each output variable. Such large number of output data can be used to obtain an expectation value for the outcome or its probability distribution. Although it had been used previously, Monte Carlo simulation was systematically developed by Neumann and Ulam. The method was named after the (Monte Carlo) casino in Monaco where Ulam's uncle often gambled and probability of the success might only be estimated in the same way.

To conduct a Monte Carlo simulation, one will need to define a model that represents the population or phenomenon of interest and a criterion to generate random numbers for the model. In order to obtain satisfactory outcome, a large stream of random numbers is often in need and, due to the expensiveness and sometimes the inconvenience of truly random numbers, pseudo-random numbers are often used instead. In MATLAB software, two popular functions that can be used for this purpose are “rand” and “randn” (MathWorks, 2011). While the outcome of the former will be drawn from a standard uniform distribution on the open (0–1) interval, that of the latter will be taken from a standard normal distribution.

Monte Carlo data generation

Monte Carlo data generation in MSD-based damage identification is actually an adaptation of Monte Carlo simulation to compensate the lack of actual multinomial data. Synthetic data (for the MSD-based training process) is commonly generated by taking multi-copies of the seed data (which can be a single or multiple feature vectors) and applying small amount of random Gaussian noise onto each copy (Worden *et al.*, 2000a; Worden *et al.*, 2002). The data generation process can be illustrated as follows.

Without loss of generality, assume that one has the seed data in the form of a single $1 \times p$ feature vector (x_1, x_2, \dots, x_p) and they wish to apply Gaussian noise of, say, 5% RMS on copies of the feature vector to generate $n \times p$ synthetic (feature) data, the steps shall include:

- (1) Calculate RMS of the feature vector

$$x_{rms} = \sqrt{\frac{1}{p}(x_1^2 + x_2^2 + \dots + x_p^2)}$$

- (2) Copying the feature vector n times and concatenating them consecutively into $n \times p$ matrix X (with all rows now being identical)
- (3) Creating an $n \times p$ (noise) matrix of Gaussian data of 5% RMS of the feature vector. Using MATLAB software, this can be done by using “randn” function

$$N_{RMS0.05} = 0.05 x_{rms} \text{ randn}(n, p)$$

- (4) Adding the noise matrix into the data matrix X to create the synthetic data

$$Y = X + N_{RMS0.05}$$

The outcome data (Y) can then be used as if they are actual observations as seen in a number of prior damage identification studies (Worden *et al.*, 2000a; Worden *et al.*, 2002).