Sensor Technologies for Civil Infrastructures

Volume 2: Applications in Structural Health Monitoring

Edited by M. L. Wang, J. P. Lynch and H. Sohn

WP
Woodhead
Publishing

Sensor Technologies for Civil Infrastructures

Related titles:

Handbook of terahertz technology for imaging, sensing and communications
(ISBN 978-0-85709-235-9)

Ultrasonic transducers
(ISBN 978-1-84569-989-5)

MEMS for automotive and aerospace applications
(ISBN 978-0-85709-118-5)

Sensor Technologies for Civil Infrastructures

Volume 2: Applications in Structural Health Monitoring

Edited by
M. L. Wang, J. P. Lynch
and H. Sohn

AMSTERDAM • BOSTON • CAMBRIDGE • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Woodhead Publishing is an imprint of Elsevier

ELSEVIER

WP
WOODHEAD
PUBLISHING

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

For information on all Woodhead Publishing publications
visit our website at http://store.elsevier.com/

Typeset by Newgen Knowledge Works Pvt Ltd, India

Printed and bound in the United Kingdom

Contents

R. Birken, J. Zhang and G. Schirner,
Northeastern University, USA

J. M. W. Brownjohn, K-Y. Koo and N. de Battista,
University of Exeter, UK

11   Sensing solutions for assessing and monitoring tunnels   309

N. A. HOULT, Queen's University, Canada and K. SOGA, University of Cambridge, UK

12   Mapping subsurface utilities with mobile electromagnetic geophysical sensor arrays   347

R. BIRKEN, Northeastern University, USA and M. ORISTAGLIO, Yale University, USA

13   Sensing solutions for assessing the stability of levees, sinkholes and landslides   396

D. INAUDI, SMARTEC SA, Switzerland

Contributor contact details

(* = main contact)

Editors

M. Wang
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA

E-mail: Mi.Wang@neu.edu

J. P. Lynch
Department of Civil and
    Environmental Engineering
University of Michigan
Ann Arbor, MI 48109-2125, USA

E-mail: jerlynch@umich.edu

H. Sohn
Department of Civil and
    Environmental Engineering
Korea Advanced Institute of
    Science and Technology (KAIST)
291 Daehak-ro
Yuseong-gu
Daejeon, 305-701, South Korea

E-mail: hoonsohn@kaist.ac.kr

Chapter 1

K. H. Law*
Department of Civil and
    Environmental Engineering
Stanford University
Stanford, CA 94305, USA

E-mail: law@stanford.edu

K. Smarsly
Department of Civil Engineering
Bauhaus University Weimar
99423 Weimar, Germany

E-mail: kay.smarsly@uni-weimar.de

Y. Wang
School of Civil and Environmental
    Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA

E-mail: yang.wang@ce.gatech.edu

Chapter 2

D. Zonta
Department of Civil,
    Environmental and Mechanical
    Engineering
University of Trento
Via Mesiano 77
38123, Trento, Italy

E-mail: daniele.zonta@unitn.it

Chapter 3

D. Bernal
Department of Civil and
    Environmental Engineering
405 Cushing Hall

102 The Fenway
Northeastern University
Boston, MA 02115, USA

E-mail: bernal@neu.edu

Chapter 4

S. Nagarajaiah*
Department of Civil &
    Environmental Engineering
    and Department of Mechanical
    Engineering
Rice University
Houston, Texas, USA

E-mail: Satish.Nagarajaiah@rice.
    edu

B. Basu
Department of Civil, Structural &
    Environmental Engineering
Trinity College Dublin,
Dublin 2, Ireland

E-mail: basub@tcd.ie

Y. Yang
Department of Civil &
    Environmental Engineering
Rice University
Houston, Texas, USA

E-mail: yy13@rice.edu

Chapter 5

D. Frangopol*
Lehigh University
Department of Civil and
    Environmental Engineering,
Engineering Research Center for
    Advanced Technology for Large
    Structural System (ATLSS)
117 ATLSS Drive
Imbt Labs

Bethlehem, PA 18015-4729, USA

E-mail: dan.frangopol@lehigh.edu

S. Kim
Korea Hydro & Nuclear Power Co.,
    Ltd.
Central Research Institute
1312-gil Yusong-daero
Yuseong-gu
Daejeon, 305-343, South Korea

E-mail: suk206@alum.lehigh.edu

Chapter 6

R. Birken,* J. Zhang and
    G. Schirner
Northeastern University
360 Huntington Avenue
514ST
Boston, MA 02115, USA

E-mail: r.birken@neu.edu;
    jxzhang@ece.neu.edu;
    schirner@ece.neu.edu

Chapter 7

J. M. W. Brownjohn,* K-Y. Koo
    and N. de Battista
College of Engineering,
    Mathematics and Physical
    Sciences
University of Exeter
North Park Road
Exeter, EX4 4QF, UK

E-mail: j.brownjohn@exeter.ac.uk

Chapter 8

A. Mita
Department of System Design
    Engineering
Keio University

3-14-1 Hiyoshi
Yokohama 223-8522, Japan

E-mail: mita@sd.keio.ac.jp

Chapter 9

Y. Q. Ni
Department of Civil and
    Environmental Engineering
The Hong Kong Polytechnic
    University
Hung Hom, Kowloon, Hong Kong

E-mail: ceyqni@polyu.edu.hk

Chapter 10

C-H. Loh
Department of Civil Engineering
National Taiwan University
Taipei, 10617, Taiwan

E-mail: lohc0220@ntu.edu.tw

Chapter 11

N. A. Hoult*
Department of Civil Engineering
Queen's University
58 University Ave.
Kingston
Ontario, K7L 3N6, Canada

E-mail: neil.hoult@civil.queensu.ca

K. Soga
Department of Engineering
University of Cambridge
Trumpington Street
Cambridge, CB2 1PZ, UK

E-mail: ks207@cam.ac.uk

Chapter 12

R. Birken*
Northeastern University
360 Huntington Avenue
514ST
Boston, MA 02115, USA

E-mail: r.birken@neu.edu

M. Oristaglio
Yale University
210 Whitney Avenue
New Haven, CT 06511,USA

E-mail: michael.oristaglio@yale.edu

Chapter 13

D. Inaudi
SMARTEC SA
Via Pobiette 11
Manno, CH-6928, Switzerland

E-mail: daniele.inaudi@smartec.ch

Chapter 14

Branko Glisic
Department of Civil and
    Environmental Engineering
Princeton University E330 EQuad
Princeton, NJ 08544, USA

E-mail: bglisic@princeton.edu

Chapter 15

M. Wang* and R. Birken
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA

E-mail: Mi.Wang@neu.edu;
    r.birken@neu.edu

Chapters 16 and 17

P. Rizzo
Laboratory for Nondestructive
    Evaluation and Structural Health
    Monitoring Studies
Department of Civil and
    Environmental Engineering
University of Pittsburgh
729 Benedum Hall
3700 O'Hara Street
Pittsburgh, PA 15261, USA

E-mail: pir3@pitt.edu

Chapter 18

M. H. Kim* and J. M. Lee
Department of Naval Architecture
    and Ocean Engineering
Pusan National University
30 Jangjeon-dong
Geumjeong-gu
Busan, 609–735, South Korea

E-mail: kimm@pusan.ac.kr

Chapter 19

R. Rolfes,* S. Tsiapoki and
    M. W. Häckell
Institute of Structural Analysis
Gottfried Wilhelm Leibniz
    Universität Hannover

Appelstraße 9A
30167, Hannover, Germany

E-mail: r.rolfes@isd.uni-hannover.
    de; s.tsiapoki@isd.uni-hannover.
    de; m.haeckell@isd.uni-hannover.
    de

Chapter 20

H. Sohn,* J. Y. Yang, H. S. Lee and
    B. J. Park
Department of Civil and
    Environmental Engineering
Korea Advanced Institute of
    Science and Technology (KAIST)
291 Daehak-ro
Yuseong-gu
Daejeon, 305-701, South Korea

E-mail: hoonsohn@kaist.ac.kr

Chapter 21

I. A. Hiskens
Department of Electrical
    Engineering and Computer
    Science
University of Michigan
1301 Beal Avenue
Ann Arbor, MI 48109-2125, USA

E-mail: hiskens@umich.edu

Woodhead Publishing Series in Electronic and Optical Materials

This page intentionally left blank

The oldest of all of the engineering disciplines, Civil Engineering, enjoys a proud history of providing society with infrastructure systems to ensure economic prosperity and a high quality of life. In recent years, a number of grand challenges have emerged that fundamentally jeopardize the profession's ability to continue designing and maintaining infrastructure with the high level of performance as experienced in the past. For example, structures in urban environments are seeing unprecedented levels of demand from growing populations leading to higher levels of deterioration occurring at faster rates. This is an especially demanding issue for civil engineers in developed nations who are now dealing with a growing number of infrastructure components exceeding their intended design lives (typically half a century). It is costlier to maintain the safe operating condition of aging infrastructure. Another challenge for civil engineers is the need to continue to improve the design of their infrastructure to withstand extreme loadings associated with natural hazard events including earthquakes and tropical storms. An especially pressing challenge is the frequent occurrence of failures at points of interconnection between two or more interdependent infrastructure systems. Unfortunately, functional interdependencies create the potential for cascading failures that can result in systemic disasters. A case in point was the progression of infrastructure system failures that occurred during both Hurricane Katrina (2005, New Orleans) and the Tohoku Earthquake (2011, Japan). The breaching of the levees in New Orleans led to cascading failures that ultimately crippled an entire city; to this day, New Orleans has not fully recuperated from this natural disaster. The Tohoku Earthquake offered another case of cascading failures with the failure of a protective sea wall setting off a chain reaction of failures leading to the meltdown of the Fukushima nuclear reactors; Fukushima remains an ongoing disaster with major public health implications for the coming century.

The civil engineering community is facing yet another grand challenge in how to innovatively design and construct more durable and safer infrastructure systems while being responsible stewards of the natural environment. The field is now exploring new ways of designing resilient infrastructure

with lower levels of embodied energy, while emitting lower levels of greenhouse gasses during construction.

The magnitude and scale of these challenges are necessitating new approaches to learning how infrastructure systems behave under both normal and extreme load conditions. Sensors are a key technology that allows civil engineers unprecedented levels of observation of operational infrastructure systems. Surveillance and observation are the main elements that aid professions to appropriately model their systems, improve system designs, and optimize future operations and maintenance processes. Concurrent to the growing demand for sensing, there has been a proliferation of new enabling technologies to improve capabilities of sensing systems while lowering their costs and simplifying their deployments. The creation of new, micro-scale sensors through the adoption of micro-electro-mechanical systems (MEMS) has led to a new generation of sensors with compact forms and low costs. In addition, the adoption of low power microprocessors that benefit from Moore's Law has helped create a class of 'intelligent and smart' sensors with computing autonomy. Similar to Moore's Law, Eldholm's Law is projecting continued improvement in wireless communications. The result has been scalable, high rate wireless communication standards that have been optimized for wireless communications within sensor networks. Undoubtedly, these trends have improved the sensing arsenal available to civil engineers to observe their systems while lowering traditional barriers to adoption including cost and ease of use.

At the outset, the editors felt there was a pressing need to create an authoritative reference of established and emerging sensing technologies impacting the field's ability to monitor its infrastructure systems. With that goal in mind, the editors invited the leaders of the field to provide detailed overviews of sensing hardware in Volume 1 of the two volume set. Specifically, Volume 1 concentrates on describing explicit sensor types and data acquisition methods relevant to the civil infrastructure domain. Sensors have the potential to generate unprecedented amounts of data; this can create a bottleneck in the decision-making process of engineers. In order to aid decision makers responsible for the operation and upkeep of infrastructure, data analysis systems are direly needed to process growing amounts of sensor data. The experts in the domain of data interrogation and decision support systems were invited to author chapters in Volume 2 of the book. Furthermore, Volume 2 showcases how sensors and sensing systems are already being used to observe specific infrastructure system types. The editors home these case studies provided the basis for others to implement impactful sensing systems of their own.

At the beginning of this book project, the editors felt there was an equally pressing need for a comprehensive reference that could be used to educate students. Specifically, graduate programs in civil engineering have begun to

add new disciplinary concentrations that are intended to train students to be the thought leaders that will usher in the solutions required to solve some of the aforementioned grand challenge problems. For example, many universities across the globe are now offering a specialization in 'infrastructure systems' with the goal of training students to approach infrastructure system problems with a 'systems' perspective. Sensors are a key enabling technology required to observe operational infrastructure systems. For this reason, these new graduate programs now offer courses in sensors, sensing systems, and data processing using sensor data. This book has been designed to offer educators a reference book that offers students a complete review of the fundamental operational principles of current and future sensor technologies (Volume 1) and showcasing the use of these sensing technologies to observe real-world systems (Volume 2). The book is written assuming students are upper-level undergraduates or at the post-graduate level.

When undertaking a project of this scale, it should be acknowledged that the project was only possible due to the support and encouragement of the editors' peers in the international research community. The pace of technological development that has occurred over the past quarter century has been fueled by the intellectual vitality and camaraderie of the research community engaged in sensing technologies for infrastructure monitoring. The editors are grateful that this community was willing and eager to contribute to this book. The result is a reference that accurately represents the state-of-the-art of the field and the impact sensing technologies has in solving emerging problems in the field of civil infrastructure systems. Enjoy reading!

Ming L. Wang, PhD
Boston, MA

Jerome Lynch, PhD
Ann Arbor, MI

Hoon Sohn, PhD
Daejeon, Korea

This page intentionally left blank

Part I
Sensor data interrogation and decision making

This page intentionally left blank

1

Sensor data management technologies for infrastructure asset management

K. H. LAW, Stanford University, USA, K. SMARSLY,
Bauhaus University Weimar, Germany, and Y. WANG,
Georgia Institute of Technology, USA

Abstract: This chapter discusses selected data management issues that are commonly shared by structural health monitoring (SHM) systems. First, data processing and management issues at the wireless sensor node are discussed, with special attention on reducing energy consumption inherent to wireless data communication. Second, issues such as communication constraints, protocol design and dynamic code migration, that are related to in-network data and software management are discussed. Third, persistent data storage and remote data access schemes, including a multi-agent-based framework for data processing and sensor self-monitoring, are presented. Technologies will continue to evolve to address these issues and to support data-rich SHM systems.

Key words: sensor network, data compression, communication protocol, code migration, multi-agent system.

1.1   Introduction

Structural health monitoring (SHM) systems are being deployed to collect measurements of structural responses originating from ambient and/or external disturbances, and to draw conclusions about the state of health of a structure based on the measurement data. Typically, sensors are strategically placed in a structure to measure and record environmental and response data. The collected data, either in raw form or being filtered or pre-processed, are then transmitted to another sensor node or to a data server for further processing and archival. Data management thus plays a very critical role for successful deployment of an SHM system.

An overview of data management issues has been discussed previously by McNeill (2009). As noted, data management issues include data collection and management at the site, data communication and transfer to off-site facilities, and data storage. Depending on the type of structure, the extent of information to be measured, and the purpose for monitoring, different

3

system configurations, measurement equipment, data communication and storage apparatus may be used. To provide a comprehensive treatment of all the issues related to data processing, data communication and data management of SHM systems is a non-trivial endeavor. In this chapter, the discussion will focus on selected data issues for the deployment of wireless sensors and sensor networks, and persistent backend data storage, management, and access.

Sensors are the most basic and primitive entities of a SHM system. They are instrumented in a structure and its vicinity to measure response data (e.g. strain, acceleration, and displacement) as well as environmental data, such as temperature, wind speed, and wind direction. The past couple of decades have seen tremendous development of wireless sensors and communication technology. With embedded microcontrollers, wireless sensors and wireless sensor networks have opened many new and exciting opportunities for their deployment in SHM systems. Wireless sensor technology not only eradicates cables and the associated material and labor cost, but also allows flexible system configurations. As many wireless sensors are designed to be powered by batteries, special considerations should be paid to minimize energy usage as much as possible. Furthermore, wireless sensors and wireless systems have their limitations, among many other issues, on communication range and possible data loss. Communication protocols must be designed to ensure that data can be transmitted in the monitoring environment and that data are being transmitted and received properly.

Measurement data are a valuable asset, not only for SHM but also for life-cycle assessment and management of the structure and the system as a whole. Persistent data storage that can effectively support data access is an important consideration of the overall data management system design. With the proliferation of the Internet and ubiquitous computing, advanced software technologies can be deployed to facilitate data access, support maintenance and operation of the monitored structure, and self-monitor the SHM system itself. The intriguing integration of data management and system monitoring is a subject worth consideration as a system solution to the monitoring of a civil infrastructure system.

This chapter is organized as follows. Section 1.2 discusses the data processing and management issues at the wireless sensor node, with special attention on reducing energy consumption inherent to wireless data communication. Section 1.3 discusses the data issues related to in-network communication. Specifically, the section examines the issues related to wireless communication range in a monitoring environment and robust communication protocol design. Furthermore, a dynamic code migration paradigm, taking advantage of wireless in-network communication, is introduced as a flexible approach to data processing. Section 1.4 discusses persistent data management and retrieval, and presents software modules developed to facilitate data access

and support system monitoring of a SHM system. Section 1.5 concludes the chapter with a brief summary and discussion.

1.2   Sensor level data processing and management

When using wireless sensors for acquiring and transmitting measurements, issues worth considering are the constraints regarding battery life and capacity, as well as communication range. These constraints could affect how data are being processed and managed at the sensor nodes. Generally speaking, wireless sensors consume less battery power to perform onboard data processing, thereby reducing the amount of data to be transmitted, than to transmit lengthy raw time-histories of sensor data. In other words, in terms of saving battery power, onboard processing is 'cheaper' than wireless communication. This is particularly relevant for applications in civil structures, which typically involve long-range communication requiring signal boosting and/or multi-hopping, and consumption of significant power by the wireless transceiver. Onboard processing measures can be adopted to preserve the life of portable batteries coupled with the wireless sensing node. These measures include embedding engineering analyses and performing data compression locally on the sensor node prior to data transmission. Using a wireless sensing node designed by Lynch (Lynch and Law, 2002) as an example, this section briefly reviews a power-efficiency study illustrating the demands for local data processing and data communication.

1.2.1   Power efficiency of an example
wireless sensing node

As an early research prototype effort, the wireless sensing node designed by Lynch (Lynch and Law, 2002) features two onboard microcontrollers, a low-power 8-bit Atmel AVR AT90S8515 model and a high-performance 32-bit Motorola MPC555 PowerPC model. Normally on for maintaining the overall operation of the wireless sensing node, the AVR microcontroller draws 8 mA of current when powered by a 5 V source. On the other hand, the MPC555 microcontroller, which has a floating point unit and ample internal program memory, draws 110 mA of current when powered at 3.3 V. The MPC555 microcontroller is normally off to save power. When needed, the AVR microcontroller wakes up the MPC555 to perform computationally demanding tasks. For wireless communication, a Proxim RangeLAN2 7911 wireless modem, operating on the 2.4 GHz Federal Communications Commission (FCC) unlicensed band, is incorporated. Using a 1 dBi omnidirectional antenna, open space ranges of 1000 ft can be attained. To attain such a large communication range, the wireless radio consumes a significant

Table 1.1 Duration of battery sources for various operational states

Operational state	Circuit Current (mA)	Internal Voltage (V)	Energizer L91 7.5 V Li/FeS$_2$ (h)
AVR On/MPC555 Off	8	5	500
AVR On/MPC555 On	160	5/3.3	15
RangeLAN Active	190	5	13
RangeLAN Sleep	60	5	40

amount of power. When internally powered by 5 V, the wireless modem draws 190 mA of current during data transmission and reception, and draws 60 mA of current when idling.

To quantify power saving, the energy consumed by the wireless sensing node is experimentally measured using 7.5 V battery sources. Table 1.1 summarizes the expected operational life of a battery pack with Energizer AA L91 lithium-based battery cells when continuously drained (Lynch *et al.*, 2004). As shown in Table 1.1, the wireless modem consumes a relatively large amount of battery energy. To preserve battery life, the use of the modem should be minimized by limiting the amount of data wirelessly transmitted. When executing an embedded analysis task, the MPC555 consumes 363 mW by drawing 110 mA at 3.3 V, while for data transmission, the RangeLAN2 radio consumes 950 mW of power running 190 mA at 5 V. Clearly, the MPC555 is 2.6 times more power efficient than the wireless radio. To determine the total amount of energy saved, the time required to perform an embedded computational task needs to be calculated. The time for data transmission can be computed on the basis of the wireless radio baud rate (19 200 bit per second).

1.2.2   Power saving measure I – embedded engineering analyses

With the MPC555 computational core, engineering analyses can be encoded and embedded in the wireless sensing node (Lynch *et al.*, 2004). To assess the energy saved by the sensing node by locally processing data, fast Fourier transform (FFT) and autoregressive (AR) time series analysis, which are two algorithms commonly used for system identification and damage detection in SHM, are used as illustrative examples. The first algorithm, FFT, can transform time series data into frequency domain for determining structural modal properties. In this experimental study, the Cooley-Tukey version of the FFT is embedded in the wireless sensing node to locally calculate the Fourier coefficients (Press, 1995). The second algorithm, which is based on AR analysis and has been widely used in damage detection studies (Sohn

Table 1.2 Energy analysis of data interrogation versus transmission of time series record

Analysis	Length of record N	Time of MPC555 calculation (s)	Energy consumed MPC555 (J)	Time for wireless transmission (s)	Energy consumed ratio (J)	Energy saved (%)
FFT	1024	0.0418	0.0152	1.7067	1.6213	99.1
FFT	2048	0.0903	0.0328	3.4133	3.2426	99.0
FFT	4096	0.1935	0.0702	6.8267	6.4854	98.9
AR (10 coefficients)	2000	1.3859	0.5031	3.3333	3.1666	84.1
AR (20 coefficients)	2000	2.8164	1.0224	3.3333	3.1666	67.7
AR (30 coefficients)	2000	4.2420	1.5398	3.3333	3.1666	51.4
AR (10 coefficients)	4000	2.7746	1.0072	6.6667	6.3333	84.1
AR (20 coefficients)	4000	5.6431	2.0484	6.6667	6.3333	67.7
AR (30 coefficients)	4000	8.5068	3.0879	6.6667	6.3333	51.2

and Farrar, 2001), fits discrete measurement data to a set of linear AR coefficients weighing past time-history observations:

$$y_k = \sum_{i=1}^{p} b_i y_{k-i} + r_k$$

[1.1]

Here y_k denotes the response of the structure at the k-th discrete point, which is expressed as a function of p previous observations of the system response, plus a residual error term, r_k. Weights on the previous observations of y_{k-i} are denoted by the b_i coefficients. Burg's approach to solving the Yule-Walker equations can be used for calculating the weighting coefficients (Press, 1995).

Table 1.2 presents the time associated with each analysis and the energy saved. The time series data were obtained during a field test conducted at the Alamosa Canyon Bridge in New Mexico (Lynch, *et al.*, 2003). As compared to the transmission of time-history record, the computational efficiency of the embedded FFT and transmission of (a few) modal frequencies can achieve major energy savings of over 98%. Calculation of AR coefficients is more complex and requires external memory for temporary data storage, resulting in longer execution times. Nevertheless, energy savings of over 50% can be achieved. Clearly, end-users of wireless sensing nodes should be cognizant of the execution times of their analyses, but on an average, significant energy can be saved by local data interrogation on the sensor nodes.

1.2.3 Power saving measure II – data compression

Time series data can be compressed by exploiting natural internal structures of data prior to transmission, so that wireless transmission can be reduced
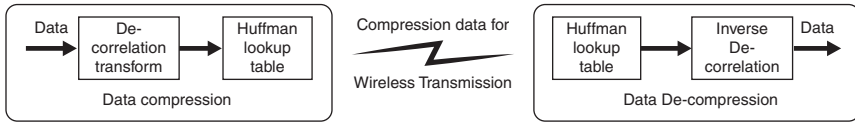
1.1 Huffman compression of sensor data using wireless sensing nodes.

and power saved. Compression algorithms can be broadly categorized into two classes: lossless and lossy compression. Lossless compression guarantees the integrity of the data without distortion. In contrast, lossy compression reduces data with reasonable distortions but can achieve higher compression rates. There are many lossless and lossy compression techniques (Salomon and Motta, 2010). A simple and computationally inexpensive compression technique, known as Huffman coding, is selected for illustration (Sayood, 2000).

Lossless Huffman coding exploits statistical relationships in the data, and pairs short symbols to data values with high probability of occurrence and long symbols to those with low probability. For example, if the 16-bit integer value '0x2342' was the most commonly occurring data sample, a short 1-bit symbol can be given to it, such as '0'. Next, if '0x2455' was the second most occurring data sample, it might be given the 2-bit symbol '10'. Hence, depending on the probability mass density of the data, a compact binary representation of variable length can be used for compressed coding. Prior to generation of a Huffman lookup table, inherent structures in data can be exploited for achieving greater compression rates by using a de-correlation transform, such as (Daubechies) wavelet transform. The data compression and de-compression process is shown in Fig. 1.1.

In the illustrated example, Huffman coding is performed for data compression prior to wireless transmission. The compression results reported in Table 1.3 are obtained from the acceleration response acquired from a shake table test on a 5 degree-of-freedom laboratory structure subjected to sweeping sinusoidal and white noise inputs (Lynch and Law, 2002). The acceleration data are recorded by the wireless sensing node using an effective 12-bit A/D converter. Table 1.3 shows the performance of lossless compression and the amount of energy saved through compressing data with the MPC555 and wirelessly transmitting the compressed record. Huffman coding compression is performed with and without wavelet transform for de-correlation. For the case of sweep excitation input, a compression rate of 61% was achieved after the initial record is de-correlated using wavelet transform. If the record is not de-correlated and internal statistical structures are not exploited in the creation of the Huffman coding lookup table, a compression rate of approximately 71% can still be attained. However,

Table 1.3 Compression of structural response data using Huffman coding

Excitation Type	De-correlation	A/D resolution (bits)	Total record size (bytes)	Compressed record size (bytes)	Compression rate (%)	Energy saved (%)
Sweep	None	12	1024	733	71.6	28.4
Sweep	Wavelets	12	1024	626	61.2	38.8
White	None	12	1024	795	77.6	22.4
White	Wavelets	12	1024	791	77.3	22.7

Table 1.4 Lossy compression of measurement data using uniform quantization

Compression scheme	Compress rate (%)	Data (MSE/ mean)	AR coefficients (30) (MSE)
Lossless	83.0	0.0	0.0
Lossy (2)	65.6	10^{-8}	10^{-6}
Lossy (4)	58.8	10^{-7}	10^{-4}

for the white noise excitation, the response lacks an inherent structure that the de-correlation transform can leverage for compression. Since the time required to compress data is practically negligible, the compression rate essentially determines the energy saved by the wireless radio in transmitting the compressed record.

Lossy compression techniques are also widely available to further compress the data with 'reasonable' data distortion within certain tolerance for a particular application. The time series data can be compressed by adding signal quantization between the de-correlation (or inverse de-correlation) filter and the Huffman coder. Table 1.4 shows the data compression results on a 16-bit time series data set collected from a high-speed boat (Sohn, *et al.*, 2001), using a simple uniform quantizer, $x_q = \text{round}(x/q)$, with quantization (sampling) factors of 2 and 4. The results demonstrate that lossy compression can significantly reduce the size of the time series data. However, the distortion errors can propagate to subsequent analyses performed, in this example shown as the mean-square error (MSE) on the AR coefficients computed with the compressed data series.

In summary, with the availability of a microcontroller onboard with the sensor nodes, significant energy can be saved by pre-processing the measurement data (i.e. directly embedding engineering analyses and/or compressing the raw sensor measurement data), prior to wireless transmission.

1.3   In-network data communication and management

Compared to cable-based systems, wireless structural monitoring systems have a unique set of advantages and technical challenges. Besides the desire for portable long-lasting energy sources, such as batteries, reliable data communication is another key data management issue to be considered in the design and implementation of wireless structural monitoring systems. Section 1.3.1 discusses the important communication issues and metrics, such as transmission latency and communication range, for adopting wireless sensors in an SHM application. Furthermore, communication constraints need to be considered carefully in the selection of hardware technologies and the design of software/algorithmic strategies. Section 1.3.2 presents the utilization of a state machine concept to design reliable communication protocols in a wireless sensor network. In addition to data transmission, software code/programs can be wirelessly transmitted and dynamically migrated to sensor nodes on an as-needed basis. This dynamic code migration strategy can greatly ease the onboard memory limitations on the sensor nodes, and save effort in the field for programming and incorporating new algorithmic developments on the sensor nodes. Section 1.3.3 describes the concept of dynamic code migration, illustrated with an example prototype implementation.

1.3.1   Communication constraints in wireless sensor network

To quantify transmission latency in wireless sensor networks, it is necessary to first review the overall architecture of a wireless sensor node. A wireless sensor node usually consists of three functional modules, including sensor interface, computational core, and wireless communication (Lynch and Loh, 2006). The sensor interface converts analog sensor signals into digital data, which are then transferred to the computational core. Besides a microcontroller/processor executing embedded program, external static random access memory (SRAM) is often integrated with the computational core to facilitate local data storage and analysis. A complete wireless communication process includes not only wireless data exchanging among transceivers of different wireless nodes, but also onboard data exchange between the computational core and the transceiver inside each wireless node.

The time required for a single transmission is defined as the wireless transmission latency between the sender and the receiver ends of the communication. A single transmission time refers to the period starting when the sender's processor begins pushing data to its transceiver, and ending when the receiver's processor obtains all the data from its transceiver. As shown in Fig. 1.2, T_{Onboard} represents the time required to transfer the data packet

Data packet sent from
processor to transceiver

Sender
side

T_{Onboard}

T_{Latency}

T_{Onboard}

Time

Receiver
side

Time

Data packet coming out of
transceiver and going into processor
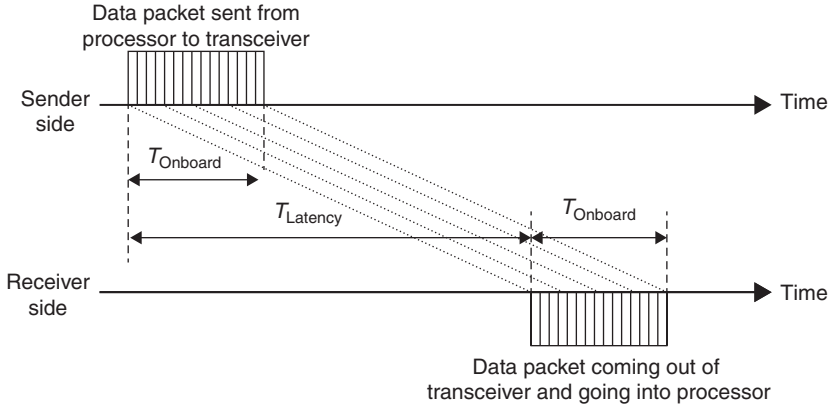
1.2 Illustration of time consumption by a single wireless transmission.

onboard, between the processor and wireless transceiver. Typical interfaces for the onboard transfer are either serial peripheral interface (SPI) or universal asynchronous receiver/transmitter (UART) interface. Once the sender's wireless transceiver obtains the first bit of data from the its processor, the transceiver starts preparing wireless transmission of the data packet. The communication latency, T_{Latency}, refers to the period starting when the sender side's processor begins pushing the first bit of data to its transceiver, and ending when the receiver side's transceiver is able to push out the first bit of the data to its processor. Assuming both the sender and the receiver have the same onboard interfaces between their processor and wireless transceiver, T_{Onboard} is usually the same for both sides.

Wireless transmission latency, T_{Latency}, is among the basic characteristics of a wireless transceiver. For example, latency of the 24XStream and 9XCite transceiver from Digi International Inc. are measured to be about 15 ms and 5 ms, respectively (Wang and Law, 2007), while the Chipcon CC2420 wireless transceiver offers a lower latency of less than 2 ms (Swartz and Lynch, 2009; Wang and Law, 2011). The onboard data transfer time, T_{Onboard}, is determined by the transfer rate of the onboard interface. For example, a UART interface can be used with a data rate of R_{UART} at 38 400 bps (bits per second). UART is usually set to transmit 10 bits for every one byte (8 bits) of sensor data, including one start bit and one stop bit. Therefore, the data rate is equivalent to R_{UART} /10 bytes per second, or R_{UART} /10 000 bytes per ms. If a data packet to be transmitted contains N bytes, the single transmission time of the data packet can be estimated as:

$$T_{\text{SingleTransm}} = T_{\text{Latency}} + \frac{10\,000N}{R_{\text{UART}}}(\text{ms}) \qquad [1.2]$$

In the analysis herein, it is assumed that the bandwidth of the wireless transceiver is higher than or equal to the onboard transfer rate R_{UART}. Otherwise, wireless bandwidth becomes a bottleneck in the communication, and the wireless baud rate should be used to replace R_{UART} in Equation [1.2]. This amount of single transmission time typically has minimal effect in most monitoring applications, but can have a noticeable effect when in-network analysis is performed. An in-network analysis may require frequent exchange of data among wireless nodes, while harnessing the embedded computing power of microcontrollers for collaborative data processing.

The other constraint, the achievable wireless communication range, is related to the attenuation of the wireless signal traveling along the transmission path. The path loss PL[dB] of a wireless signal is measured as the ratio between the transmitted power, P_{TX} (mW), and the received power, P_{RX} (mW) (Molisch, 2005):

$$\text{PL}(\text{dB}) = 10 \log_{10} \frac{P_{TX}(\text{mW})}{P_{RX}(\text{mW})} \qquad [1.3]$$

Path loss generally increases with the distance, d, between the transmitter and the receiver. However, the loss of signal strength varies with the environment along the transmission path and is difficult to quantify precisely. Experiments have shown that a simple empirical model may serve as a good estimate to the mean path loss (Rappaport and Sandhu, 1994):

$$\overline{PL}(d)(\text{dB}) = PL(d_0)(\text{dB}) + 10n \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma(\text{dB}) \qquad [1.4]$$

Here $PL(d_0)$ is the free-space path loss at a reference point close to the signal source (d_0 is usually selected as 1 meter for ease of calculation). Parameter X_σ represents the variance of the path loss, which is a zero-mean log-normally-distributed random variable with a standard deviation of σ. Parameter n is the path loss exponent that describes how fast the wireless signal attenuates over distance. In essence, Equation [1.4] indicates an exponential decay of signal power:

$$P_{RX}(\text{mW}) = P_0(\text{mW})\left(\frac{d}{d_0}\right)^{-n} \qquad [1.5]$$

where P_0 is the received power at the reference distance d_0. Typical measured values of n are reported to be between 2 and 6 (Rappaport and Sandhu, 1994).

Table 1.5 Link budget analysis to two types of wireless transceivers (1 W transmitted power)

	900 MHz	2.4 GHz
P_{TX}(dBm)	30.00	30.00
AG (dBi)	4.00	4.00
RS (dBm)	−104.00	−105.00
FM (dB)	22.00	22.00
$\overline{\mathrm{PL}} = P_{\mathrm{TX}} + \mathrm{AG} - \mathrm{RS} - \mathrm{FM}$ (dB)	116.00	117.00
$\mathrm{PL}(d_0)$ (dB), $d_0 = 1$ m	31.53	40.05
$\overline{\mathrm{PL}} - \mathrm{PL}(d_0)$ (dB)	84.47	76.95
n	2.80	2.80
$\overline{d}$ (m)	1039.72	560.23

A link budget analysis can be used to estimate the range of wireless communication (Molisch, 2005). To achieve a reliable communication link, it is required that following inequality holds:

$$P_{TX}(\mathrm{dBm}) + \mathrm{AG}(\mathrm{dBi}) \geq \mathrm{PL}(d)(\mathrm{dB}) + \mathrm{RS}(\mathrm{dBm}) + \mathrm{FM}(\mathrm{dB}) \qquad [1.6]$$

where AG denotes the total antenna gain for the transmitter and the receiver, RS the receiver sensitivity, FM the fading margin to ensure the quality of service, and $\mathrm{PL}(d)$ the realized path loss at some distance d within an operating environment. Table 1.5 summarizes the link budget analysis for a 900 MHz and a 2.4 GHz transceiver (both operating in a free unlicensed frequency band in the United States), and their estimated indoor ranges. The transmitted power for both transceivers, P_{TX}, is set at 1 W (30 dBm), the maximum power allowed by the Federal Communications Committee.

In Table 1.5, a total antenna gain AG of 4 dBi is employed by assuming that low-cost 2 dBi whip antennas are used at both the transceiver and receiver. The receiver sensitivity RS and fading margin FM of the two wireless transceivers are chosen as typical values available from commercial transceivers. The free-space path loss at d_0 is computed using the Friis transmission equation (Molisch, 2005):

$$\mathrm{PL}(d_0)(\mathrm{dB}) = 20\log_{10}\left(\frac{4\pi d_0}{\lambda}\right) \qquad [1.7]$$

where λ is the wavelength of the corresponding wireless signal. Rappaport and Sandhu (1994) reported values of path loss exponent, n, that are measured

for a number of different building types. The value for soft-partitioned office building ($n = 2.8$) is listed in Table 1.5 for both transceivers. Finally, assuming that the variance X_σ is zero, the mean communication range $\bar{d}$ can be derived from Equation [1.4] as shown at the bottom row of Table 1.5:

$$\bar{d} = d_0 10^{(\overline{PL} - PL(d_0))/(10n)} \qquad [1.8]$$

It is important to note the sensitivity of the communication range with respect to the path loss exponent n in Equation [1.8]. For instance, if the exponent of 3.3 for indoor traveling (through brick walls, as reported by Janssen & Prasad (1992) for 2.4 GHz signals) is used for the 2.4 GHz transceiver, the mean communication range reduces to 214.7 m. Our experience in the design of wireless sensors and their placement has demonstrated the importance of considering the path loss estimation in the physical implementation of a SHM system with reliable data collection (Wang and Law, 2007).

1.3.2 Communication protocol development through state machine concept

To ensure reliable communication between a wireless server and wireless nodes, or among multiple wireless nodes, the data communication protocol needs to be carefully designed and implemented. A commonly used network communication protocol is the transmission control protocol (TCP). Establishing a full duplex virtual connection between two endpoints, TCP is a sliding window protocol that handles both timeouts and retransmissions. Although TCP is highly reliable, it is usually too general and cumbersome to be employed by low-power and low data-rate wireless nodes. If the complete TCP suite is adopted in a low-power wireless sensor network, protocol overhead can cause long latency while transmitting each wireless packet, and significantly slow down the communication throughput. While general purpose operating systems for wireless sensors are available to conveniently implement TCP, for practical and efficient application in a wireless structural sensor network, a simpler communication protocol is quite desirable to minimize the transmission overhead without overwhelming the limited resources of the sensor nodes. Furthermore, the protocol needs to be designed to ensure reliable wireless transmission by properly addressing possible data loss. Nevertheless, a light-overhead communication protocol designed for a wireless structural sensor network can still inherit many useful features of TCP, such as data packetizing, sequence numbering, timeout checking, and retransmission.

Finite state machine concepts can greatly facilitate the designing and programming of an efficient communication protocol. A finite state machine consists of a set of discrete states and specified transitions between the states (Tweed, 1994). Taking the protocol between a wireless server and multiple wireless sensing nodes as an example, the state machine of either the server or a wireless node can only be in one of the possible states at any point in time. In response to different events, each machine transits between states. Figure 1.3 illustrates part of a communication protocol for one round of sensor data collection, as described by state machine diagrams for the server and for the wireless nodes; note that the server and the nodes have separate sets of state definitions. During each round of data collection, the server collects sensor data from all wireless nodes. The detailed communication protocol for initialization and synchronization has been discussed by Wang, *et al*. (2007).

As shown in Fig. 1.3, at the beginning of data collection, the server and all the wireless nodes are all set in their own State 1. Starting with the first wireless node in the network, the server queries each node for the availability of data by sending the '01Inquiry' command. If the data is not ready, the node replies '02NotReady'; otherwise, the node replies '03DataReady' and transits to State 2. After the server ensures that the data from this wireless node is ready for collection, the server transits to State 3. To request a data segment from a node, the server sends a '04PlsSend' command that contains a packet sequence number. One round of data collection from one wireless node is ended with a two-way handshake, where the server and the node exchange '05EndTransm' and '06AckEndTransm' commands. The server then moves on to the next node and continuously collects sensor data round-by-round.

Since the typical objective in structural monitoring is to transmit sensor data or analysis results to the server, in this set of state machine designs, the server is assigned the responsibility for ensuring reliable wireless communication. If the server sends a command to a wireless sensing node and does not receive an expected response from the node within a certain time limit, the server will resend the last command again until the expected response is received. However, after a wireless sensing node has responded to the server, the node does not check if the message has successfully arrived at the server, because the server is assigned the overall responsibility. The wireless sensing node becomes aware of data loss only when the server queries the node for the same data again.

1.3.3   Dynamic wireless code migration

In addition to transmitting data sets, algorithms and software programs can also be transmitted over the wireless network. Specifically, the concept of dynamic wireless code migration is a powerful approach toward resource-efficient and reliable data processing and management at the

(a)

(b)
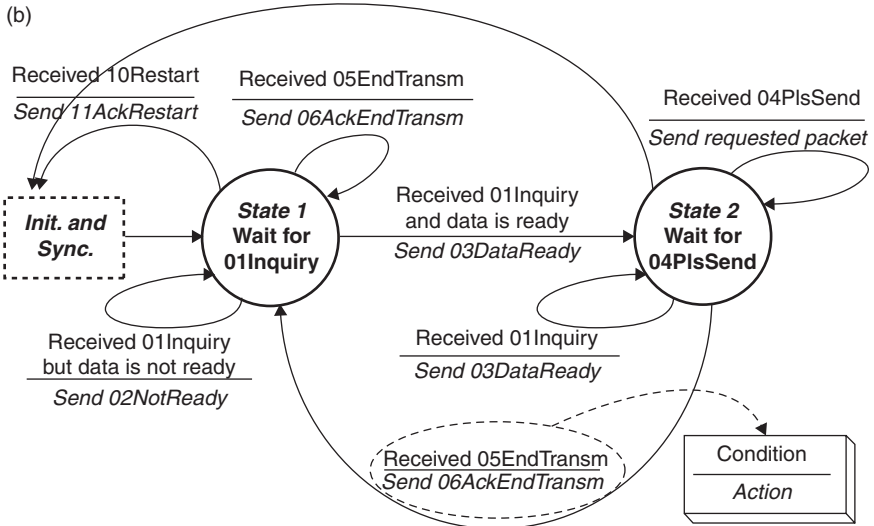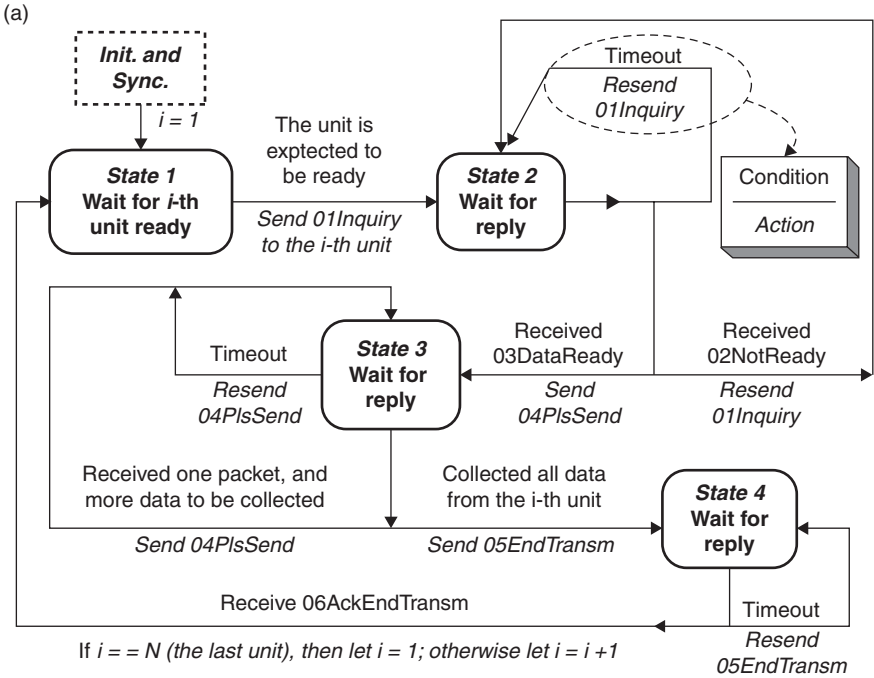
1.3 Communication state machine diagrams for wireless SHM. (a) State diagram of the server. (b) State diagram of a wireless sensing node.

sensor node. Wireless code migration, i.e. software programs (including dynamic behavior, actual state, and specific knowledge) physically migrating from one sensor node to another, has been used in a number of areas, such as mobile commerce, medical applications, and distributed traffic detection (Chen *et al.,* 2009; Herbert *et al.,* 2006; Mihailescu *et al.,* 2002). The concept can be beneficial for wireless SHM as well.

Relatively sophisticated wireless sensing nodes are now available to support dynamic code migration. For example, Smarsly *et al.* (2011a) have employed a Java-based wireless sensing node, the SunSPOT node developed by Sun Microsystems, for a prototype implementation. Unlike common embedded applications for wireless sensor networks, which are usually written in low-level native languages (such as C/C++ and assembly language), the SunSPOT node contains a fully capable Java ME, which is widely used, for example, on advanced mobile phones. The computational core is an Atmel AT91RM9200 system-on-a-chip (SoC) incorporating a 32-bit ARM920T ARM processor with 16 kB instruction and 16 kB data cache memories and executing at 180 MHz maximum internal clock speed. The SunSPOT node also incorporates a Spansion S71PL032J40 chip that consists of 4 MB flash memory and 512 kB RAM.

Smarsly *et al.* (2011a) have coupled the migration-based approach with mobile multi-agent technology for wireless SHM. The wireless monitoring system consists of a base node and clusters of sensor network systems, each cluster being composed of one head node that manages the cluster, and a number of sensor nodes that collect structural sensor data (Fig. 1.4). The base node is connected to a local computer that includes a database and an information pool providing global information on the monitoring system and on the structure. Example information includes modal properties of the structure under monitoring, sensor nodes installed, and a catalog of data analysis algorithms.

Two basic types of mobile software programs are designed, namely 'onboard agents' permanently residing at the head and sensor nodes, and 'migrating agents' located at the head nodes to be sent to the sensor nodes
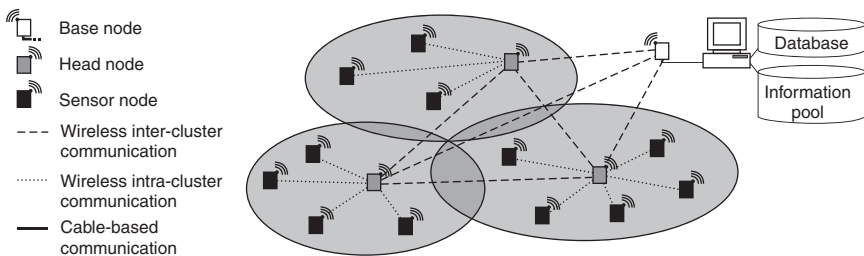
1.4 Hierarchical architecture of the monitoring system.

upon request. The onboard agents installed on the sensor nodes are self-contained, interacting software programs capable of making their own decisions and acting in the wireless sensor network with a high degree of autonomy. The onboard agents are designed to continuously record sensor data from the monitored structure, to perform simple routines for detecting suspected structural abnormalities, and to aggregate the sensor data (for example, to extract representative features and values from sets of measurements). The aggregated data is then transmitted to the database installed on the connected local computer for persistent storage.

As opposed to onboard agents that are permanently residing at the nodes, the migrating agents are capable of physically migrating from one node to another in real-time. While the onboard agents at the sensor nodes are continuously executing relatively simple yet resource-efficient routines, the migrating agents are designed to carry out more comprehensive data analysis algorithms directly on a sensor node. Acting upon a request by an onboard agent in the case of detected or suspected abnormal changes of the monitored structure ('anomalies'), a migrating agent is dynamically composed of the most appropriate algorithm selected from the information pool for analyzing the detected anomaly (here a Cooley-Tukey FFT algorithm as introduced in the previous section). In the prototype implementation of dynamic code migration, a 96.4% reduction of wirelessly transferred data has been achieved compared to transferring the collected raw sensor data to a central server (Smarsly *et al*., 2011b). Furthermore, the memory consumption on a sensor node utilizing the code migration strategy can potentially be reduced compared to the conventional execution of embedded algorithms.

1.4   Persistent data management and retrieval

This section provides an overview on data management and retrieval in SHM systems by means of a wind turbine monitoring system as an illustrative example (Hartmann *et al.,* 2011; Smarsly *et al.,* 2012a). Section 1.4.1 gives a brief description of the monitoring system. The basic steps of data collection, processing, and archiving are discussed in Section 1.4.2, followed by an example of remote communication with the monitoring system in Section 1.4.3. Finally, in Section 1.4.4, specific features of the monitoring system are illuminated, such as the autonomous detection of sensor malfunctions.

1.4.1   A wind turbine monitoring system

A wind turbine monitoring system is prototypically implemented on a 500 kW wind turbine located in Germany. The system is designed to

On-site hardware system   Decentralized software system
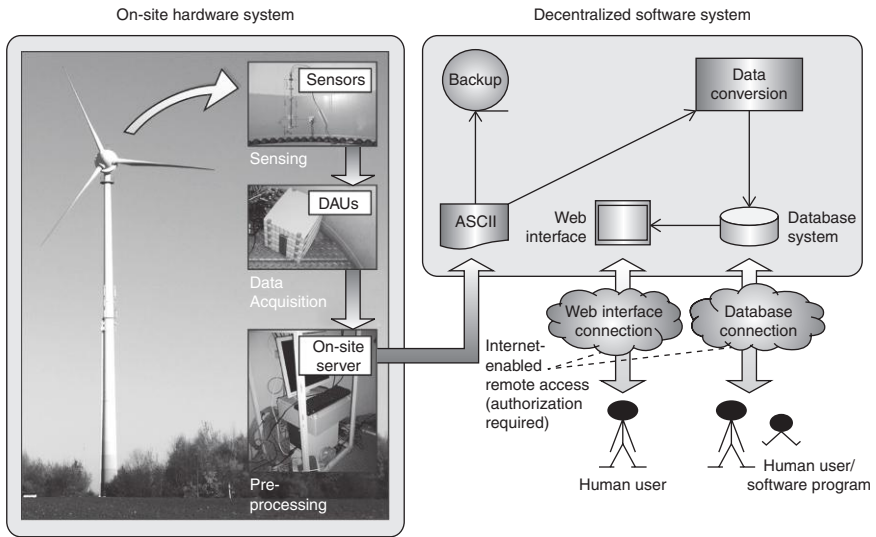
1.5 Architecture of the wind turbine monitoring system.

systematically assess the condition of the wind turbine, detect damage and deterioration, and estimate its service lifespan. The monitoring system consists of an on-site hardware system and a software system for supporting distributed and remote access (Fig. 1.5). Installed in the wind turbine, the on-site hardware system includes sensors, data acquisition units (DAUs), and a local computer (on-site server). Remotely connected to the hardware system, the software system is composed of software modules designed to continuously execute relevant monitoring tasks, such as storing and converting the sensor data collected from the wind turbine.

The on-site hardware system is implemented to collect structural, environmental, and operational data for assessing the condition of the wind turbine. For that purpose, the wind turbine is instrumented with sensors installed both inside and outside the tower as well as on the foundation of the wind turbine. Six inductive displacement transducers, type HBM-W5K, are mounted at two different levels inside the tower. The displacement transducers are complemented by Pt100 resistance temperature detectors to capture temperature influences on the displacement measurements. Additional temperature sensors are placed at two other levels inside and outside the tower to measure temperature gradients. In addition, six three-dimensional 3713D1FD3G piezoelectric accelerometers, manufactured by PCB Piezotronics, are placed at five levels in the tower. On the foundation of the wind turbine, three single-axis PCB-393B12 piezoelectric seismic accelerometers are installed. For acquiring

(a)

Displacement transducer
(HBM-W5K)

RTD surface sensor
(Pt100)

Accelerometer
(PCB-3713D1FD3G)

(b)

3D ultrasonic anemometer
(METEK USA-1)
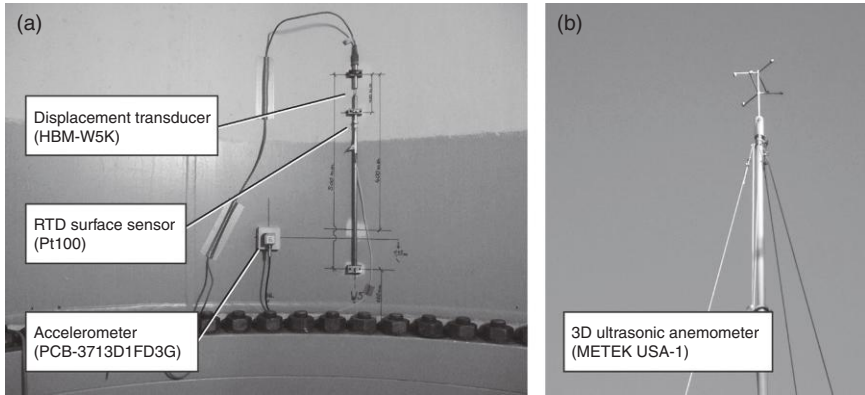
1.6 Sensors inside the wind turbine tower (a) and three-dimensional anemometer (b) being part of the on-site hardware system.

wind information, a Metek USA-1 ultrasonic anemometer is mounted on a telescopic mast next to the wind turbine. Figure 1.6 shows components of the on-site hardware system installed in the wind turbine tower as well as the anemometer.

The sensors are controlled by the DAUs, which are connected to the on-site server located in the maintenance room of the wind turbine. For the acquisition of temperature data, three 4-channel Picotech resistance thermometer data logger (RTD) input modules PT-104 are deployed; for the acquisition of acceleration and displacement data, four Spider8 measuring units are used. Each Spider8 unit has separate A/D converters, ensuring simultaneous measurements at sampling rates between 1 and 9600 Hz. All data sets, being sampled and digitized, are continuously forwarded from the DAUs to the on-site server for temporary storage.

The software system is installed on different computers at the Institute for Computational Engineering (ICE) in Bochum (Germany). The data collected by the on-site hardware system is forwarded to the software system using a permanently installed digital subscriber line (DSL) connection. As shown in Fig. 1.5, the software system is designed to provide a persistent data storage, and to support remote access to the data sets and the monitoring system. For data storage, the software system comprises (i) server systems for online data synchronization, data conversion, and data transmission, (ii) RAID-based storage systems for periodic backup, and (iii) a MySQL database for persistent data storage. A web interface connection is designed to facilitate interactions by the human users to remotely monitor the wind turbine. In addition, a database connection can be utilized by both human users and software programs to remotely access the monitoring system and to download and analyze the data.
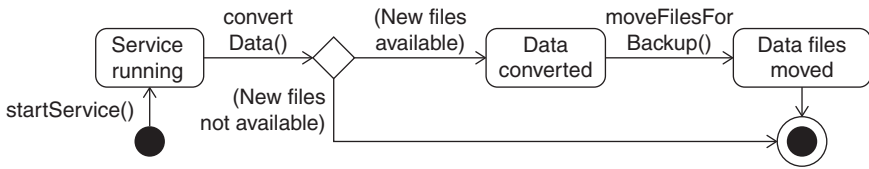
1.7 Abridged illustration of the automated data conversion expressed in terms of a state machine protocol.

1.4.2   Data processing and management

The data sets collected on the wind turbine are stored temporarily on the on-site server. The on-site server automatically creates local backups of all recorded data sets and, through a permanently installed DSL connection, transmits the raw data to the MySQL database installed on a main server at ICE. The data transmission is automatically executed by a 'Cron' job scheduler, which is a time-based Unix utility running on the on-site server to ensure the periodic execution of tasks according to specified time intervals. When uploading the collected raw data to the main server for persistent archival, metadata are added to provide definitions of installed sensors, DAU IDs, output specification details, date and time formats, etc. This data conversion process is implemented using a commercial, open-source tool 'Pentaho Data Integration' (PDI), which offers metadata-driven conversion and data extraction capability (Castors, 2008; Roldan, 2009). Figure 1.7 shows the basic tasks defined for executing the data conversion process, which includes (i) starting the conversion service, (ii) data processing, and (iii) moving the input data files for storage on a database.

Once the data are successfully converted and stored in the MySQL database, an acknowledgement is sent from the main server at ICE to the on-site server in the wind turbine, whereupon the corresponding data set on the on-site server is deleted. During the conversion process, all data sets involved are automatically 'locked' and cannot be accessed by software programs or by human users, in order to avoid inconsistencies. Performance tests validating the automated data conversion process have been documented (Smarsly and Hartmann, 2009a, b, 2010). After being stored in the database at ICE, the data are available for remote access.

Figure 1.8 shows the basic structure of the database. Database tables (such as 'pt104', 'spider8' and 'usa') are defined for the DAUs (PT-104 modules, Spider8 measuring units, USA-1 anemometer) installed in the wind turbine. Each field in a database table represents one sensor connected to the DAU. For example, the database table 'pt104', shown in Fig. 1.8, comprises ten fields, which correspond to the data recorded by the PT-104 modules through the temperature sensors T1 to T10.

Monitoring database

Table	Records	Type	Size
spider8	1,724,319,519	MyISAM	432.6 GiB
usa	877,378,974	MyISAM	57.2 GiB
pt104	17,574,367	MyISAM	1.9 GiB

pt104

Field	Type	Null	Default	Comments
time	bigint(20)	No		
t1	float	No		
t2	float	No		
t3	float	No		
t4	float	No		
t5	float	No		
t6	float	No		
t7	float	No		
t8	float	No		
t9	float	No		
t10	float	No		

usa

Field	Type	Null	Default	Comments
time	bigint(20)	No		
v	int(11)	No		
d	int(11)	No		
z	int(11)	No		
t	int(11)	No		

spider8

Field	Type	Null	Default	Comments
time	bigint(20)	No		
B1_x	float	No		
B1_y	float	No		
B1_z	float	No		
B2_x	float	No		
B2_y	float	No		
B2_z	float	No		
B3_x	float	No		
B3_y	float	No		
B3_z	float	No		
B4_x	float	No		
B4_y	float	No		
B4_z	float	No		
B5_x	float	No		
B5_y	float	No		
B5_z	float	No		
B6_x	float	No		
B6_y	float	No		
B6_z	float	No		
W1	float	No		
W2	float	No		
W3	float	No		
W4	float	No		
W5	float	No		
W6	float	No		
B7	float	No		
B8	float	No		
B9	float	No		

1.8 Structure of the monitoring database (extract).

During the automated conversion process, the basic statistics of the data sets, such as quartiles, medians and means, are computed at different time intervals and stored in the MySQL database (Table 1.6). As an example, Fig. 1.9 shows the database table 'usa_3' which summarizes the statistics of the data sets collected by the USA-1 anemometer over a period of 3 s, as indicated by the suffix '_3'. The data sets and statistics are made available for analyzing the physical and operational states of the wind turbine structure.

1.4.3   Remote access to the monitoring system

Remote access to the monitoring data is provided through the web interface and through the direct database connection. The web interface offers

Table 1.6 Characteristic values describing the secondary monitoring data

Description	Suffix
Least value not considered an outlier	_least_non_outlier
First (lower) quartile Q1	_lower_quartile
Second quartile Q2 (median)	_median
Mean value	_mean
Third (upper) quartile Q3	_upper_quartile
Largest value not considered an outlier	_largest_non_outlier

Monitoring database (aggregated)

Table	Records	Type	Size
pt104_3	17,806,670	MyISAM	4.4 GiB
pt104_60	946,176	MyISAM	237.7 MiB
pt104_360	148,397	MyISAM	37.3 MiB
pt104_600	94,617	MyISAM	23.8 MiB
pt104_1800	29,686	MyISAM	7.5 MiB
pt104_21600	2,477	MyISAM	638.3 KiB
spider8_3	17,475,701	MyISAM	12.9 GiB
spider8_60	944,449	MyISAM	712.8 MiB
spider8_360	145,640	MyISAM	109.9 MiB
spider8_600	94,446	MyISAM	71.3 MiB
spider8_1800	29,135	MyISAM	22.0 MiB
spider8_21600	2,434	MyISAM	1.8 MiB
usa_3	17,740,632	MyISAM	2.0 GiB
usa_60	946,001	MyISAM	107.7 MiB
usa_360	147,847	MyISAM	16.8 MiB
usa_600	94,600	MyISAM	10.8 MiB
usa_1800	29,575	MyISAM	3.4 MiB
usa_21600	2,467	MyISAM	289.0 KiB

usa_3

Field	Type	Null	Default	Comments
time	bigint(20)	No		
v_least_non_outlier	float	No		
v_lower_quartile	float	No		
v_median	float	No		
v_mean	float	No		
v_upper_quartile	float	No		
v_largest_non_outlier	float	No		
d_least_non_outlier	float	No		
d_lower_quartile	float	No		
d_median	float	No		
d_mean	float	No		
d_upper_quartile	float	No		
d_largest_non_outlier	float	No		
z_least_non_outlier	float	No		
z_lower_quartile	float	No		
z_median	float	No		
z_mean	float	No		
z_upper_quartile	float	No		
z_largest_non_outlier	float	No		
t_least_non_outlier	float	No		
t_lower_quartile	float	No		
t_median	float	No		
t_mean	float	No		
t_upper_quartile	float	No		
t_largest_non_outlier	float	No		

1.9 Tertiary monitoring data composed of characteristic values (database structure).

graphical user interfaces (GUIs) for remotely visualizing, exporting, and analyzing the monitoring data, while the database connection allows accessing the data directly, for example by software programs and other tools. Figure 1.10 shows the web interface displaying the monitoring data collected during 24 h on 24 April 2011. As shown in the left pane of the web interface, the monitoring data are graphically displayed as selected by the user through the control panel on the right pane. Using the control panel, the user can select the data collected by specific sensors, specify the time intervals to be plotted, conduct online data analyses, and export data sets. In Fig. 1.10, a time
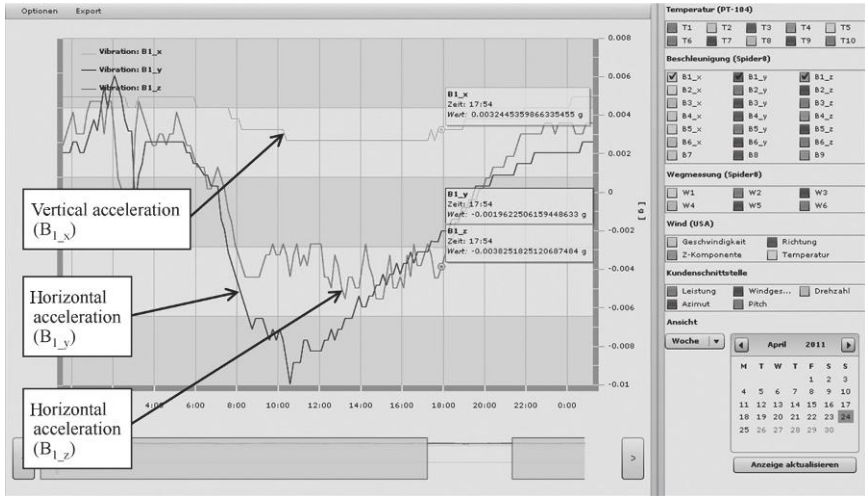
1.10 Data sets remotely accessed via the web interface provided by the monitoring system (collected during 24 h on 24 April 2011).

history of acceleration data is displayed, collected by a Spider8 measuring unit through a 3D accelerometer installed at the height of 63 m in the wind turbine tower (sensor B1). The web interface and the database connection provide easy access to the users and software modules for interacting with the monitoring system and conducting monitoring tasks.

1.4.4   Detection of data anomalies and sensor malfunctions

One of the key issues in an SHM system is to ensure that the collected sensor data are reliable and, furthermore, the sensors or the DAUs are in good condition. Typical malfunction may be caused by communication problems due to long-distance lines, breakdowns of sensors, or temporary power blackouts that affect the computer systems. A flexible and extensible multi-agent system is designed and connected to the existing monitoring system through the database connection (Smarsly *et al.,* 2011c, d, 2012b). The purpose of the multi-agent system is to self-detect malfunctions and enable corrective actions. Once a malfunction of a DAU is observed, the system informs the responsible individuals about the detected defects through email alerts. The affected DAUs can be restarted remotely or replaced in a timely manner.

A multi-agent system consists of multiple interacting software components or 'agents.' Software agents are characterized by two basic capabilities: *autonomy* and *flexibility*, which make multi-agent technology well suited for implementing distributed, real-time applications. An 'autonomous' software

agent is able to operate without any direct intervention by humans or other software systems, to control its actions, and to decide independently which actions are appropriate for achieving prescribed goals (Russell and Norvig, 1995; Wooldridge, 2009). In this implementation, the multi-agent system is developed on the basis of the widely used Java Agent Development Framework JADE (Bellifemine *et al.,* 2003, 2004, 2007). To ensure extensibility and interoperability, the multi-agent system is implemented in compliance with the specifications issued by the Foundation for Intelligent Physical Agents (FIPA). FIPA, the IEEE Computer Society standards organization for agents and multi-agent systems, promotes agent-based technology, interoperability of agent applications, and interoperability with other technologies (FIPA, 2004, 2002a, b, c, d). Besides ensuring extensibility and interoperability, adhering to the FIPA specifications provides considerable advantages with respect to performance and robustness of the implemented multi-agent system. To illustrate the integration of data management system with self-monitoring functions, the following briefly describes two specific agents, namely an 'interrogator agent' for sensor malfunction detection, and a 'mail agent' for sending email alerts to responsible personnel.

One typical potential DAU malfunction that causes an interruption of the data acquisition process is often implicitly indicated by anomalies in the data sets, such as a long consecutive sequence of unusual, identical measurements. To detect such a data anomaly, the interrogator agent at certain time intervals extracts and analyzes the data sets stored in the MySQL monitoring database (Smarsly *et al*., 2012a). The interrogator agent is connected to the monitoring database through the Java Database Connectivity (JDBC). Security of database requests and data transmissions is ensured by security mechanisms provided by the MySQL database, which requires password and username as well as secure drivers to be specified by the interrogator agent when trying to access the database. A set of configuration files defining interrogation parameters, database URL, database driver, sensor specifications, scheduled interrogation intervals, etc., is predefined and stored in the multi-agent system (Smarsly and Law, 2012, Smarsly *et al.,* 2012a). Figure 1.11 shows an abridged example of how a connection is established using the required specifications of URL, database driver, username, and password (connect method). Similarly, other complex queries can be executed by the agent (interrogate method). As shown in the figure, SQL command queries, issued to request for the sensor data or for performing data analyses, can be dynamically executed.

Upon detecting possible data anomalies, the responsible individuals are immediately notified by the mail agent. On behalf of the interrogator agent the notification is issued by the mail agent via an agent-based email messaging service in a two-step process. In the first step, the mail agent collects all relevant information from the interrogator agent about the observed

```
private Connection connect( String DRIVER,
                           String URL,
                           String USERNAME,
                           String PASSWORD )){

    Connection connection = null;
    ...
    Class.forName(DRIVER).newInstance();
    connection = DriverManager.getConnection(URL, USERNAME, PASSWORD);
    ...
    return connection;
}


private void interrogate( Connection connection,
                          SaveData saveData, ... ){

    ...
    Statement statement = connection.createStatement();
    statement.execute(query)
    ResultSet resultSet = statement.getResultSet();
    ...
}
```

Driver and URL information: available in the multi-agent system

Username and password: individually given to authorized agents

SQL command gets
(i) Data sets to be used for agent-based analyses or
(ii) Computed outliers if the data analyses have been executed by the database system

1.11 Abridged example of establishing and utilizing a database connection.

anomaly. Furthermore, metadata stored in the configuration files, such as addresses of the recipients or the email server to be used, are acquired to compose the email, which is then automatically sent to the email clients using the Simple Mail Transfer Protocol (SMTP). For that purpose, the mail agent communicates with an SMTP server that is specified in the configuration files (and can be changed any time by the users, as approved by and registered with the multi-agent system). Secure email messages are ensured by username- and password-based authentications that the mail agent, like a human user, needs to specify when trying to access the SMTP server.

Since its initial deployment in 2009, the multi-agent system has reliably detected all the malfunctions that have occurred and has notified individuals via email alerts. To illustrate the remote access to the monitoring system using the web interface, the detection of a malfunction of a temperature DAU, as occurred on 17 April 2011, is shown in Fig. 1.12. Identical temperature measurements have been repeatedly stored by the DAU for a long period of time, while other DAUs show variations on the temperature measurements. Figure 1.13 shows a printed excerpt of the corresponding email, assembled and sent by the mail agent, which includes detailed information on the revealed anomaly. An internal system malfunction in the DAU (one of the PT-104 input modules installed in the wind turbine) was identified as the cause and the engineer, after having received the email alert, remotely restarted the DAU in a timely manner.

One distinct advantage of the agent-based approach is that the multi-agent system can be easily extended to accommodate additional functions. A variety of self-sensing and self-detection mechanisms can be built and
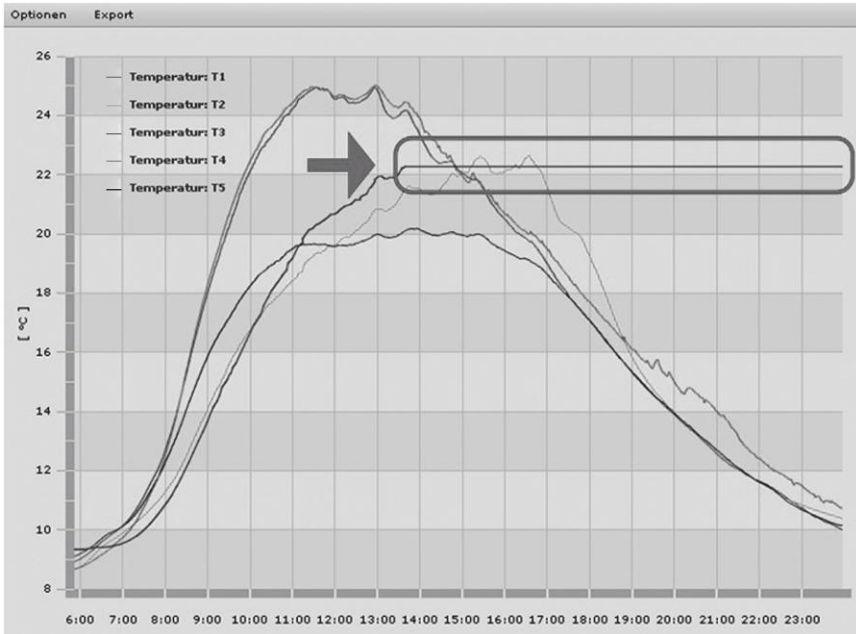
1.12 Malfunction detected by the multi-agent system on 17 April 2011.

1.13 Excerpt of the email alert, assembled and sent by the multi-agent system.

data interrogation methods, such as FFTs and autoregressive models, can be implemented as specialized agents readily integrated into the multi-agent system.

1.5   Conclusion and future trends

Measurement data collection, data transmission, and persistent storage for data retrieval and access are among the key issues for successful deployment of an SHM system. With a few selected examples from prior wireless sensor research and from system deployment efforts, this chapter has discussed a number of issues faced by sensor data management of SHM systems. Issues have been discussed and demonstrated at three different operational levels: sensor level, in-network level, and system and database level.

1. At the sensor level, more specifically for wireless sensor nodes, energy constraints are an important issue when handling collected measurement data. To reduce the energy-demanding wireless transmission of data, onboard data processing and compression are viable approaches to optimize the performance at the sensor level. A dual-microcontroller architecture has been shown, allowing the power-consuming and more sophisticated microcontroller to be switched on only when complicated computing tasks are needed. Algorithms can be embedded to pre-process the data prior to transmission.
2. With respect to data management occurring at the sensor network level, communication bandwidth and range are important issues to consider. Robust and light-overhead end-to-end communication protocols are necessary to achieve reliable data exchange. For flexible onboard data processing, wireless exchange of analysis code, i.e. dynamic wireless code migration, offers new opportunities for conveniently providing wireless sensor nodes with a wide variety of engineering analysis algorithms on demand.
3. At the system and database level, persistent data storage and convenient data retrieval from potentially off-site database is fundamentally important for the maintenance and operation of the monitored structure. JDBC has been illustrated to provide online interfaces for querying a relational database. A multi-agent system allows a flexible and modular approach for implementing external functions for processing the data and for performing operational activities, including self-monitoring and diagnosis.

In today's rapidly evolving technological world, new hardware/software platforms will continue to emerge. Although the technologies used as examples in this chapter are not necessarily the latest, the data management and

processing issues discussed are commonly shared by different generations of SHM systems. Future energy harvesting technologies may alleviate energy constraints at the sensor node level, and empower more sophisticated in-network analysis. New data modeling techniques, such as the hierarchical data format HDF5 (HDF-Group, 2011), provide enhanced capabilities to support complex and large-volume data sets. New technologies and analyses will eventually enable low-cost, pervasive, and ubiquitous sensing that supports data-rich SHM systems. It can be expected that future SHM systems will provide a much more detailed and accurate understanding of structural performance than today's systems.

1.6   Acknowledgements

1.7   References

Bellifemine, F., Caire, G., Poggi, A. and Rimassa, G. (2003). 'JADE – A white paper', *EXP in Search of Innovation*, **3**(3): 6–19.

Bellifemine, F., Caire, G. and Greenwood, D. (2004). *Developing Multi-Agent Systems with JADE*, John Wiley & Sons, West Sussex, UK.

Bellifemine, F., Caire, G., Trucco T. and Rimassa, G. (2007). *JADE Programmer's Guide*, [online]. http://jade.tilab.com/doc/programmersguide.pdf.

Castors, M. (2008). *PDI Performance Tuning Check-List, Technical Report*, Pentaho Corporation, Orlando, FL, USA.

Chen, B., Cheng, H. H. and Palen, J. (2009). 'Integrating mobile agent technology with multi-agent systems for distributed traffic detection and management systems', *Transportation Research Part C-Emerging Technologies*, **17**(1): 1–10.

Foundation for Intelligent Physical Agents – FIPA (2002a). *SC00001:2002 FIPA Abstract Architecture Specification*, FIPA, Alameda, CA, USA.

Foundation for Intelligent Physical Agents – FIPA (2002b). *SC00026:2002 FIPA Request Interaction Protocol Specification*, FIPA, Alameda, CA, USA.

Foundation for Intelligent Physical Agents – FIPA (2002c). *SC00037:2002 FIPA Communicative Act Library Specification*, FIPA, Alameda, CA, USA.

Foundation for Intelligent Physical Agents – FIPA (2002d). *SC00061:2002 FIPA ACL Message Structure Specification*, FIPA, Alameda, CA, USA.

Foundation for Intelligent Physical Agents – FIPA (2004). *SC00023K:2004 FIPA Agent Management Specification*, FIPA, Alameda, CA, USA.

Hartmann, D., Smarsly, K. and Law, K. H. (2011). 'Coupling sensor-based structural health monitoring with finite element model updating for probabilistic lifetime estimation of wind energy converter structures', *Proceedings of the 8th International Workshop on Structural Health Monitoring*. Stanford, CA, USA, 13–16 September 2011.

HDF-Group (2011). *Hierarchical Data Format HDF5* [online]. http://hdfgroup.org/HDF5/.

Herbert, J., O'Donoghue, J., Ling, G., Fei, K. and Fok, C.-L. (2006). 'Mobile agent architecture integration for a wireless sensor medical application', *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Hong Kong, China, 18–22 December 2006.

Janssen, G. J. M. and Prasad, R. (1992). 'Propagation measurements in an indoor radio environment at 2.4 GHz, 4.75 GHz and 11.5 GHz', *Proceedings of IEEE 42nd Vehicular Technology Conference*, 2: 617–620, Denver, CO, 10–13 May 1992.

Lynch, J.P. and Law, K.H. (2002). *Decentralization of Wireless Monitoring and Control Technologies for Smart Civil Structures*. 140, John A. Blume Earthquake Eng. Ctr., Stanford University, Stanford, CA.

Lynch, J.P., Sundararajan, A., Law, K.H., Kiremidjian, A.S., Carryer, E., Sohn, H. and Farrar, C.R. (2003). 'Field validation of a wireless structural monitoring system on the Alamosa Canyon Bridge', *Proceedings of the SPIE 10th Annual International Symposium on Smart Structures and Materials*, San Diego, CA, USA, 2–6 March 2003.

Lynch, J.P., Sundararajan, A., Law, K.H., Kiremidjian, A.S. and Carryer, E. (2004). 'Embedding damage detection algorithms in a wireless sensing unit for operational power efficiency', *Smart Materials and Structures*, **13**(4): 800–810.

Lynch, J.P. and Loh, K.J. (2006). 'A summary review of wireless sensors and sensor networks for structural health monitoring', *The Shock and Vibration Digest*, **38**(2): 91–128.

McNeill, D.K. (2009). 'Data management and signal processing for structural health monitoring of civil infrastructure systems', in *Structural Health Monitoring of Civil Infrastructure Systems*, V.M. Karbhari and F. Ansari (eds.), CRC Press, Boca Raton, FL.

Mihailescu, P., Binder, W. and Kendall, E. (2002). 'MAE: A mobile agent platform for building wireless M-commerce applications', *Proceedings of the 8th ECOOP Workshop on Mobile Object Systems: Agent Applications and New Frontiers*, Malaga, Spain, 10–14 June 2002.

Molisch, A.F. (2005). *Wireless Communications*, John Wiley & Sons, IEEE Press, Chichester, West Sussex, England.

Press, W.H. (1995). *Numerical Recipes in C*, Cambridge University Press, Cambridge, England.

Rappaport, T.S. and Sandhu, S. (1994). 'Radio-wave propagation for emerging wireless personal-communication systems', *Antennas and Propagation Magazine, IEEE*, **36**(5): 14–24.

Roldan, M.C. (2009). *Pentaho Data Integration (Kettle) Tutorial, Technical Report*, Pentaho Corporation, Orlando, FL, USA.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, USA.

Salomon, D. and Motta, G. (2010). *Handbook of Data Compression*, Springer, London, New York.

Sayood, K. (2000). *Introduction to Data Compression*, Morgan Kaufmann Publishers, San Francisco, CA, USA.

Smarsly, K. and Hartmann, D. (2009a). 'Real-time monitoring of wind energy converters based on software agents', *Proceedings of the 18th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering*, Weimar, Germany, 7–9 July 2009.

Smarsly, K. and Hartmann, D. (2009b). 'Multi-scale monitoring of wind energy plants based on agent technology', *Proceedings of the 7th International Workshop on Structural Health Monitoring*, Stanford, CA, USA, 9–11 September 2009.

Smarsly, K. and Hartmann, D. (2010). 'Agent-oriented development of hybrid wind turbine monitoring systems', *Proceedings of ISCCBE International Conference on Computing in Civil and Building Engineering and the EG-ICE Workshop on Intelligent Computing in Engineering*, Nottingham, UK, 30 June–2 July 2010.

Smarsly, K., Law, K.H. and König, M. (2011a). 'Resource-efficient wireless monitoring based on mobile agent migration', *Proceedings of the SPIE: Health Monitoring of Structural and Biological Systems 2011*, San Diego, CA, USA, 7–10 March 2011.

Smarsly, K., Law, K.H. and König, M. (2011b). 'Autonomous structural condition monitoring based on dynamic code migration and cooperative information processing in wireless sensor networks', *Proceedings of the 8th International Workshop on Structural Health Monitoring*, Stanford, CA, USA, 13–15 September 2011.

Smarsly, K., Law, K.H. and Hartmann, D. (2011c). 'Implementing a multiagent-based self-managing structural health monitoring system on a wind turbine', *Proceedings of the 2011 NSF Engineering Research and Innovation Conference*, Atlanta, GA, USA, 4–7 January 2011.

Smarsly, K., Law, K.H. and Hartmann, D. (2011d). 'Implementation of a multiagent-based paradigm for decentralized real-time structural health monitoring', *Proceedings of the 2011 ASCE Structures Congress*, Las Vegas, NV, USA, 14–16 April 2011.

Smarsly, K. and Law, K.H. (2012). 'Advanced structural health monitoring based on multi-agent technology', in *Computation for Humanity: Information Technology to Advance Society*, J. Zander and P. Mostermann (eds.) Taylor & Francis Group, LLC., Boca Raton, FL, USA.

Smarsly, K., Law, K.H. and Hartmann, D. (2012a). 'A multiagent-based collaborative framework for a self-managing structural health monitoring system', *ASCE Journal of Computing in Civil Engineering*, **26**(1), 76–89.

Smarsly, K., Hartmann, D. and Law, K.H. (2012b). 'Structural health monitoring of wind turbines supervised by autonomous software components – 2nd level

monitoring', *Proceedings of the 14th International Conference on Computing in Civil and Building Engineering*, Moscow, Russia, 26–29 June 2012.

Sohn, H., Farrar, C.R., 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*. LA-13761-MS, 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.

Swartz, R.A. and Lynch, J.P. (2009). 'Strategic network utilization in a wireless structural control system for seismically excited structures', *Journal of Structural Engineering*, **135**(5): 597–608.

Tweed, D. (1994). 'Designing real-time embedded software using state-machine concepts', *Circuit Cellar Ink*, **53**: 12–19.

Wang, Y. and Law, K.H. (2007). *Wireless Sensing and Decentralized Control for Civil Structures: Theory and Implementation*. 167, John A. Blume Earthquake Eng. Ctr., Stanford University, Stanford, CA.

Wang, Y., Lynch, J.P. and Law, K.H. (2007). 'A wireless structural health monitoring system with multithreaded sensing devices: design and validation', *Structure and Infrastructure Engineering*, **3**(2): 103–120.

Wang, Y. and Law, K.H. (2011). 'Structural control with multi-subnet wireless sensing feedback: experimental validation of time-delayed decentralized H_∞ control design', *Advances in Structural Engineering*, **14**(1): 25–39.

Wooldridge, M. (2009). *An Introduction to Multiagent Systems*, 2nd Edn., John Wiley & Sons, West Sussex, UK.

2

Sensor data analysis, reduction and fusion for assessing and monitoring civil infrastructures

D. ZONTA, University of Trento, Italy

Abstract: Structural health monitoring data analysis is basically a logical inference problem, wherein we attempt to gain information on the structural state based on sensor responses. In this chapter, we first introduce Bayesian logic as the main instrument to formulate the inference problem in rigorous mathematical terms, properly accounting for data and model uncertainties. Next, an overview of the most popular data reduction techniques is provided, with a special focus on principal component analysis (PCA). The chapter then introduces the concept of data fusion and discusses techniques to handle multi-temporal and multi-sensor data based on Bayesian statistics. Alternative non-probabilistic logical models for handling uncertainties are outlined at the end.

Key words: probabilistic data analysis, Bayesian inference, data fusion, data reduction.

2.1   Introduction

In structural health monitoring we acquire data using sensors to understand the condition state of a structure. Therefore the process of data interpretation is basically a logical inference problem, wherein we attempt to gain information on the structural state based on sensor responses. We have an ideal condition when the state of the structure is bi-univocally and deterministically mapped to the sensor observations. However, in the real world, data and interpretation models are very often uncertain, so we need logic which accounts for uncertainties; in this chapter we will use Bayesian logic as the main instrument to formulate the problem in rigorous mathematical terms. In Section 2.2, we first introduce the concept of Bayesian inference and explain with simple examples how it applies to structural health monitoring problems. Permanent monitoring often results in a large amount of data from different types of sensor, the data usually being highly redundant. Reducing the size of the dataset helps reduce the computational effort and

33

optimize the analysis performance. An overview of the most popular data reduction techniques is provided in Section 2.3, with special focus on principal component analysis (PCA). Data fusion refers to techniques and tools used for combining sensor data, which is a typical problem encountered in structural monitoring. In Section 2.4 we introduce the concept of data fusion and discuss some of the most popular techniques to handle multi-temporal and multi-sensor data, always based on Bayesian statistics. Alternative non-probabilistic logical schemes for handling uncertainties are outlined at the end of the chapter.

2.2   Bayesian inference and monitoring data analysis

In structural health monitoring we acquire data to understand the condition state of a structure. For example, we use strain gauges to measure the deformation of a concrete surface. We may want to use the same information to understand whether the concrete surface is cracked or not. We use load cells to understand the tension of a tendon in a cable-stayed bridge. On a bridge we may use accelerometers to acquire the free or forced vibration modes, and then use the acceleration data to estimate the natural frequencies or mode shapes. Based on changes in model parameters, we can infer whether any components of the bridge are damaged or not.

   In more abstract terms we can say that to monitor is to infer logically the state of a system (the structure) based on observations (the instrumental measurements), on assumptions as to the state of the structure (prior knowledge), and on the relationship with the observations (model).

   This section introduces the reader to this logical process (Section 2.2.1) and states the formal tools for Bayesian inference (Section 2.2.2). It first analyses the case where the data interpretation model is a probabilistic parametric one, generic (Section 2.2.3) or linear Gaussian (Section 2.2.4). To highlight the connection between deterministic and probabilistic parameter identification, the most popular statistical parameter estimators are discussed (Section 2.2.5). Eventually, the approach is further generalized to those problems whereby the interpretation model can change with the structural state (Section 2.2.6).

2.2.1   Inference in health monitoring

The problem of data interpretation is basically a logical problem, so we will start by defining the tools of logical inference. To illustrate the ideal monitoring system, Wenzel (2011) liked to use, as an example, a fuel warning light: when driving, if this warning light is on, we infer that the fuel is low, so we stop to refuel, even if we cannot physically observe the fuel level in the tank. Deterministic logic is based on the assumption that information

from a sensor is deterministically related to the state of the object monitored. Thus, when we see the warning light, we can infer the state of the item monitored (low fuel). When the relationship between observation and state is (or assumed to be) bi-univocal and deterministic ('light on' if and only if the fuel is low), the deductive inference can follow the classical Aristotelian deductive reasoning ('light on' necessarily implies that the fuel is low).

In real life, things are not so simple. For example, the warning light could be burnt out so we will not see any warning even if the fuel is low (false negative case). Or the float mechanism in the tank could be faulty, so the warning light stays on even when the fuel is not low (false positive case). So in general we can see that the relationship between the observation (the sensor response) and the state is not fully deterministic, but has some degree of randomness. In these cases classical deterministic logic fails to apply strictly: 'light on' implies that either the gas is low or the float is faulty, so the observation does not provide any deterministic information on the state of the tank level. That said, when we observe the light on, we can still say that it is plausible that the fuel is low.

To deal formally with such uncertainties we need to recast the problem in a probabilistic framework. The approach followed here is based on Bayesian logic. The Bayesian theory of probability originates from Bayes' well-known essay (Bayes, 1763), and today many modern specialized textbooks can provide the reader with a critical review and applications of this theory to data analysis; see for instance Gregory (2005) and Sivia (2006). Of all the papers dealing with the application of Bayesian theory to engineering problems, I wish to underline Beck's work in Papadimitriou *et al*. (1997), Beck and Katafygiotis (1998), Beck and Au (2002), which by disseminating these concepts have had great impact on the civil engineering community.

2.2.2   Formal probabilistic framework

In the examples above we have introduced indirectly the idea that structural health monitoring entails an inference process to correlate observation and state. Now we want an abstract from the physical system and define the basic component of the probabilistic inference process involved in health monitoring. First, we define the concepts of sensor, state observation and model that we will use in the rest of this chapter.

Sensor: there are many uses of the term sensor; in this chapter we will refer to a sensor in logical terms as suggested by Hall and McMullen (2004): a logical sensor is defined as 'any device which functions as a source of information'.

Observation: in general, we use objective information in our inference problem; in structural health monitoring, the raw data acquired by the sensors are observations.

State: an object which represents the condition of the system. The state of the system can be specified by a discrete variable or class (for example, a concrete beam can be cracked or not cracked), by a set of continuous parameters $\boldsymbol{\theta}$ (for example, Young modulus of concrete, the opening of a crack) or by a combination of the two.

Model: the way we assume the observations are correlated to the state. Typically, in structural engineering problems, the model has a physical background (e.g. a finite element or analytical model). Sometimes the model can be heuristic. Very often it depends on some state parameters; it can be probabilistic, in the sense that the parameters are random and are represented by a distribution.

Going back to the definition introduced above, the basic monitoring problem is to infer the state S of the structure based on a set of instrumental observations, or measurements, $\mathbf{y}$. In general the state S is identified by set of variables which characterize the state. To start with, we assume that the structural system can be in one of N mutually exclusive and exhaustive scenarios, so the structural state can be univocally defined by a variable S which can assume the discrete values $S \in (S_1, S_2, \ldots, S_N)$. Once measurements $\mathbf{y}$ become available from the monitoring system, Bayes' theorem allows calculation of the updated, or *posterior*, probability for each state S_n, from prior probability $\mathrm{prob}(S_n)$, state likelihood $\mathrm{pdf}(\mathbf{y}|S_n)$ and evidence $\mathrm{pdf}(\mathbf{y})$, using the following expression:

$$\mathrm{prob}\left(S_n \,|\, \mathbf{y}\right) = \frac{\mathrm{pdf}\left(\mathbf{y}\,|\,S_n\right)\mathrm{prob}\left(S_n\right)}{\mathrm{pdf}\left(\mathbf{y}\right)} \tag{2.1}$$

where pdf denotes the *probability density function* of a random variable or vector. Equation [2.1] reads: my estimated probability of being in state n after acquiring the sensor data $\mathbf{y}$ is equal to the likelihood of the observed data multiplied by my state probability estimated a priori, normalized by the evidence of the data acquired. The first term in the numerator is normally referred to as the likelihood of the scenario, a function that states how credible are the observations assuming the known state S (how plausible is the strain recorded when my concrete is cracked). The calculation of this likelihood entails definition of a probabilistic model which connects the structural state to the sensor response. As the scenario set is complete and mutually exclusive, evidence of measurement $\mathbf{y}$ is simply obtained by summing over the scenarios:

$$\mathrm{pdf}\left(\mathbf{y}\right) = \sum_{n=1}^{N} \mathrm{pdf}\left(\mathbf{y}\,|\,S_n\right) \cdot \mathrm{prob}\left(S_n\right) \tag{2.2}$$

(a)

(b)
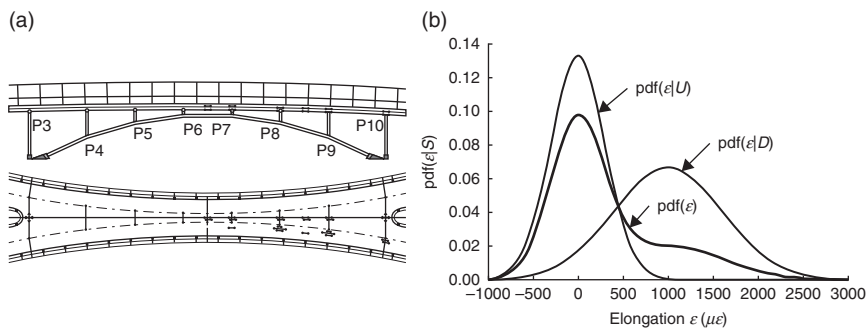
2.1 Location of sensors in the main span of the bridge (a); likelihoods and evidence (b), Glisic *et al*. (2012).

Equation [2.1], known as Bayes' rule, is technically the result of manipulating a conditional probability, but its meaning is fundamental to understanding the cognitive essence of the inference process. Bayes' rule shows that the posterior probability is always the result of the update of prior knowledge. The equation does not determine the source of the prior; it simply states that the data interpretation process always depends on prior knowledge. Prior probabilities assigned to each scenario reflect the initial judgement of the evaluator, independently of the outcome of monitoring. In summary, the interpretation of the monitoring data always presumes (i) the existence of a model that correlates the observation to the state, and (ii) a prior idea about the state of the structure.

Example 1. State classification. In Glisic *et al*. (2012) is described the case of 'Tom', a fictitious maintenance manager, who wants to understand the damage state of the steel-concrete arch bridge depicted in Fig. 2.1a, based on a single strain observation ε, recorded at midspan by an optical strain gauge. The details of the bridge and its monitoring system are reported in (Glisic, 2011). In this problem, S is formally a variable which can assume the two mutually exclusive and exhaustive values undamaged (U) and damaged (D).

Tom expects that, if the bridge is virtually undamaged, the change in strain will be close to zero. He is also aware of the natural fluctuation of the midspan curvature, mainly due to thermal effects; his monitoring system provider told him that this fluctuation might be of the order of ±300 $\mu\varepsilon$. Formally, we can encode this knowledge in a theoretical *likelihood of no-damage* pdf($\varepsilon|U$), normally distributed, with zero mean value and standard deviation $\sigma = 300$ $\mu\varepsilon$ which represents Tom's expectation of the system response in the undamaged state. Conversely, on the assumption that the bridge is heavily damaged but still standing, Tom expects a significant change in strain; we can model the

evidence of damage PDF($\varepsilon|D$) as a Gaussian distribution with mean value of 1000 $\mu\varepsilon$, and standard deviation of $\sigma = 600 \mu\varepsilon$, which reflects Tom's uncertainty of expectation. In addition, he guesses that after a damage event (a road accident), U is more likely than scenario D and he estimates prob(D) = 30% and prob(U) = 70% as prior probabilities. Using his prior judgement Tom can also predict the distribution of ε, before this data is available, by marginalizing the system states through the following formulation:

$$\text{pdf}(\varepsilon) = \text{pdf}(\varepsilon|D)\text{prob}(D) + \text{pdf}(\varepsilon|U)\text{prob}(U) \qquad [2.3]$$

When the monitoring observation, ε, is available to the manager, he can update his estimate of the probability of damage using Bayes' formula:

$$\text{prob}(D|\varepsilon) = \frac{\text{pdf}(\varepsilon|D)\,\text{prob}(D)}{\text{pdf}(\varepsilon)} \qquad [2.4]$$

2.2.3 Probabilistic parametric model

In the previous example the likelihood functions are defined heuristically, based on the engineer's judgement, and basically reflect her/his knowledge of the mechanical behaviour of the bridge. In general, an evaluation of likelihood requires knowledge of a probabilistic model which correlates the measurements and the parameters which define the system state. Very often the model includes a response model r, which encodes the expected physical behaviour, and an epistemic component, which reproduces the epistemic uncertainty. The mechanical part r reflects our physical interpretation of the behaviour, and may depend on a set of parameters θ_r. Civil engineers are normally very confident with mechanical models, as these are the typical tools used to predict the structural response at the design stage. In the real world, we do not expect the model to fit the observation perfectly, even after identification of the optimal parameters, because the response prediction is usually affected by errors. The sources of discrepancy can include (Bevington and Robinson, 2003):

- systematic errors in the measurement or bias errors (for example: calibration errors);
- random fluctuation of the observation due to limited precision of the measurement system or to the physical nature of the quantity observed;
- incompleteness of the model: the model does not fully reproduce the physics of the problem (for example, we did not account for thermal expansion, while local temperature changes are significant).

Whatever the source of the discrepancy, the scatter between model prediction and observation can be modelled with an error, e, or epistemic component, which in turn may depend on a number of additional parameters, $\boldsymbol{\theta}_e$. Formally, we can write:

$$y = r(\boldsymbol{\theta}_r) + e(\boldsymbol{\theta}_e) = \hat{y}(\boldsymbol{\theta}) \qquad [2.5]$$

Therefore, the observations $\mathbf{y}$ can be seen as the outcome of a probabilistic model $\hat{\mathbf{y}}(\boldsymbol{\theta})$ which depends on a set of parameters $\boldsymbol{\theta}$ that includes both physical and epistemic parameters. It is important to observe that, having defined the probabilistic model in the form of Equation [2.5], it is irrelevant whether the nature of the parameter is mechanical or epistemic for its identification.

When the state is defined by a set of continuous parameters $\boldsymbol{\theta}$, we can rewrite Equation [2.1] in the form:

$$\text{pdf}(\boldsymbol{\theta}|\mathbf{y}) = \frac{\text{pdf}(\mathbf{y}|\boldsymbol{\theta}) \cdot \text{pdf}(\boldsymbol{\theta})}{\text{pdf}(\mathbf{y})} \qquad [2.6]$$

which shows that the posterior distribution $\text{pdf}(\boldsymbol{\theta}|\mathbf{y})$ of the parameters is proportional to the likelihood $\text{pdf}(\mathbf{y}|\boldsymbol{\theta})$ and to the prior distribution $\text{pdf}(\boldsymbol{\theta})$ of the parameters. The evidence can be seen as a normalization constant, which must be calculated by integrating over the parameter space:

$$\text{pdf}(\mathbf{y}) = \int_{D\boldsymbol{\theta}} \text{pdf}(\mathbf{y}|\boldsymbol{\theta}) \cdot \text{pdf}(\boldsymbol{\theta}) \cdot d\boldsymbol{\theta} \qquad [2.7]$$

where $D\boldsymbol{\theta}$ means the domain of the parameters $\boldsymbol{\theta}$. Assume that $\mathbf{y}$ is a set of m uncorrelated observations $\{y_1, y_2, ..., y_m\}$; in this case the application of the joint probability rule yields:

$$\text{pdf}(\mathbf{y}|\boldsymbol{\theta}) = \text{pdf}\left(\bigcap_{i=1}^{m} y_i \,|\, \boldsymbol{\theta}\right) = \prod_{i=1}^{m} \text{pdf}(y_i\,|\,\boldsymbol{\theta}) \qquad [2.8]$$

which is to say that the likelihood of a set of uncorrelated measurements is simply the product of the likelihood $\text{pdf}(y_i|\boldsymbol{\theta})$ of the individual measurements.

Example 2. Probabilistic model of a concrete column. The expected axial contraction $\hat{\varepsilon}_r(t)$ at time t of a concrete column under a constant load N can

be seen as a combination of an elastic term, a creep term ε_{cc} and a shrinkage term ε_{cs}. We assume that the structural response depends on three random mechanical parameters: Young modulus of concrete E_c, the creep coefficient ϕ, and the ultimate shrinkage $\varepsilon_{cs,\infty}$; while the other quantities in the physical model are deterministically known. In formula:

$$\hat{\varepsilon}_r\left(t; E_c, \phi, \varepsilon_{cs,\infty}\right) = \frac{N}{E_c A} + \varepsilon_{cc}\left(t; \phi\right) + \varepsilon_{cs}\left(t; \varepsilon_{cs,\infty}\right) \qquad [2.9]$$

where $\varepsilon_{cc}\left(t; \phi\right)$ and $\varepsilon_{cs}\left(t; \varepsilon_{cs,\infty}\right)$ denote the creep and shrinkage models assumed, respectively. The column is instrumented with a long gauge strain sensor, which records m measurements of the contraction of the column $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, ..., \varepsilon_m\}$ at times $t_1, t_1, ..., t_m$. Our goal is to identify the posterior distribution of the three parameters based on the observed strain. We can model the discrepancy between the observed response and the model prediction with a random Gaussian noise $g\left(\sigma\right)$ with unknown standard deviation σ; in this case the probabilistic model which predicts an observation will read:

$$\hat{\varepsilon}\left(t; E_c, \phi, \varepsilon_{cs,\infty}, \sigma\right) = \hat{\varepsilon}_r\left(t; E_c, \phi, \varepsilon_{cs,\infty}\right) + g\left(\sigma\right) = \frac{N}{E_c A} + \varepsilon_{cc}\left(t; \phi\right) + \varepsilon_{cs}\left(t; \varepsilon_{cs,\infty}\right) + g\left(t; \sigma\right)$$

$$[2.10]$$

so in turn any observation ε_i is interpreted through a probabilistic model $\hat{\varepsilon}(t_i; \boldsymbol{\theta})$ which depends on parameters $\boldsymbol{\theta} = \{E_c, \phi, \varepsilon_{cs,\infty}, \sigma\}$. Given a specific set of parameters and times, the distribution of the prediction is normal distributed with mean value $\hat{\varepsilon}_r\left(t; E_c, \phi, \varepsilon_{cs,\infty}\right)$ and standard deviation σ:

$$\mathrm{pdf}\left(\hat{\varepsilon}\right) = \mathrm{Norm}\left\{\hat{\varepsilon}; \hat{\varepsilon}_r\left(t; E_c, \phi, \varepsilon_{cs,\infty}\right), \sigma\right\} \qquad [2.11]$$

where in general the notation $\mathrm{Norm}\{x; \mu, \sigma\}$ indicates normal distribution of variable x with mean value μ and standard deviation σ. When at time t_i we observe measurement ε_i, the likelihood of the parameters is:

$$\mathrm{pdf}\left(\varepsilon_i \mid \boldsymbol{\theta}\right) = \mathrm{pdf}\left(\varepsilon_i = \hat{\varepsilon}(t_i; \boldsymbol{\theta})\right) = \mathrm{Norm}\left\{\varepsilon_i; \hat{\varepsilon}_r\left(t_i; E_c, \phi, \varepsilon_{cs,\infty}\right), \sigma\right\} \quad [2.12]$$

or in other words it is the normal distribution with mean value equal to the response model prediction $\hat{\varepsilon}_r\left(t_i; E_c, \phi, \varepsilon_{cs,\infty}\right)$ and standard deviation equal to σ, calculated for value ε_i. Last, the likelihood of the dataset can be calculated using Equation [2.8]:

$$\text{pdf}\left(\boldsymbol{\varepsilon}\,|\,\boldsymbol{\theta}\right) = \prod_{i=1}^{m} \text{Norm}\left\{\varepsilon_i; \hat{\varepsilon}_r\left(t_i; E_c, \phi, \varepsilon_{cs,\infty}\right), \sigma\right\} \tag{2.13}$$

2.2.4   Linear Gaussian model

A special case is that of a linear Gaussian model. Here we assume a linear relationship between the observations **y** and the parameters **θ**:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{g} \tag{2.14}$$

where **A** is a linear transformation, deterministic, and **g** is a zero mean Gaussian noise with known covariance. This model sometimes reflects the actual physics of the problem, but may be just an approximation driven by computational convenience. Very often engineers prefer to work with normal distributions for such convenience, because any linear combination of normal distributions is still normal. Unfortunately, probabilistic models are generally non-linear as to parameters, even assuming a linear relationship between physical model and mechanical parameters.

Example 3. Temperature compensation. We want to estimate the baseline deformation ε_0 at reference temperature T_0 of a steel cable subject to temperature changes. The cable has a strain gauge and a thermometer, and the recordings at instants t_1, t_1, …, t_m are denoted with $\boldsymbol{\varepsilon} = \left\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_m\right\}$ and $\mathbf{T} = \left\{T_1, T_2, ..., T_m\right\}$, respectively. Our physical knowledge of the problem lets us state the following relationship:

$$\hat{\varepsilon}(t) = \varepsilon_0 + \alpha\left(\hat{T}(t) - T_0\right) \tag{2.15}$$

where α is the (unknown) apparent thermal expansion coefficient, and $\hat{\varepsilon}$ and $\hat{T}$ denote the supposed physical strain and temperature, respectively. In turn the values of the physical quantities are related to the observations by:

$$\begin{aligned} \varepsilon_i &= \hat{\varepsilon}(t_i) + g(\sigma_\varepsilon) \\ T_i &= \hat{T}(t_i) + g(\sigma_T) \end{aligned} \tag{2.16a,b}$$

where σ_ε and σ_T are the standard deviations of strain and temperature noise and where as usual we indicate with $g(\sigma)$ a zero mean uncorrelated Gaussian noise. Combining Equations [2.15] with [2.16] yields the probabilistic relationship between observations and parameters:

$$\varepsilon_i = \varepsilon_0 + \alpha(T_i - T_0) + g\left(\sqrt{\sigma_\varepsilon^2 + (\alpha\sigma_T^2)}\right) \tag{2.17}$$

The physical quantities in Equations [2.15] and [2.16] are sometimes improperly referred to as the 'true' values in some literature; however, it is important to note that these quantities have meaning only within the model assumed by the engineer in charge of the data interpretation, while the only 'true' values involved in the inference process are the observations. Equation [2.17] shows that the relationship between parameters and observations is non-linear, even assuming as deterministic the standard deviation of the measurements, because the standard deviation of the noise depends on parameter α. For computational reasons, therefore, it is convenient to assume that the standard deviation is independent of α.

$$\varepsilon_i = \varepsilon_0 + \alpha(T_i - T_0) + g(\sigma) \tag{2.18}$$

We can easily verify that a probabilistic model underlying Equation [2.18] is a linear Gaussian model where temperature T is seen as a dependent variable of $\hat{\varepsilon}$ (similarly to t in Equation [2.15]), and where the standard deviation of noise incorporates all the uncertainties of the problem. In formula:

$$\hat{\varepsilon}(T; \varepsilon_0, \alpha) = \varepsilon_0 + \alpha(T - T_0) + g(\sigma) \tag{2.19}$$

or:

$$\hat{\varepsilon}(T; \varepsilon_0, \alpha) = \left[1\ (T - T_0)\right]\begin{bmatrix}\varepsilon_0 \\ \alpha\end{bmatrix} + g = \mathbf{A}\boldsymbol{\theta} + \mathbf{g} \tag{2.20}$$

which highlights the canonical form of Equation [2.14]. With this approximation, the temperature compensation problem reduces to a linear fit in the plane ε–T, where ε_0 is the intercept and α is the slope of the fitting line.

2.2.5 Statistical estimators

Many civil engineers dealing with structural monitoring are familiar with structural identification techniques, but are not very familiar with probability. In general, the problem of structural identification is to find the optimal set of parameters, where by optimal we mean those which somehow minimize the discrepancy between prediction and observation. To see the

connection between deterministic and Bayesian parameter identification, it is useful to analyse some of the more popular statistical estimators, and in particular the maximum a posteriori (MAP), the maximum likelihood (ML) and the least square (LS). The objective is to highlight the underlying metric of the methods and its meaning in Bayesian terms.

MAP. This set of parameters is the set that maximizes the posterior probability of the observations:

$$\boldsymbol{\theta}_{MAP} = \arg\max_{\boldsymbol{\theta}} \mathrm{pdf}(\boldsymbol{\theta}|\mathbf{y}) \qquad [2.21]$$

From a logical point of view, the MAP parameters are those that identify the most likely state of the structure, given the prior and the observations. Noting that Bayes' principle allows not only calculation of the 'optimal' parameters values, but also their posterior distribution, knowledge of the posterior distribution allows evaluation of the degree of confidence and determinacy of the solution.

ML. We define ML solution as the set of parameters that maximizes the likelihood of the observations:

$$\boldsymbol{\theta}_{ML} = \arg\max_{\boldsymbol{\theta}} \mathrm{pdf}(\mathbf{y}|\boldsymbol{\theta}) \qquad [2.22]$$

As suggested by Equation [2.6] the posterior probability is proportional to prior and likelihood:

$$\mathrm{pdf}(\boldsymbol{\theta}|\mathbf{y}) \propto \mathrm{pdf}(\mathbf{y}|\boldsymbol{\theta}) \cdot \mathrm{pdf}(\boldsymbol{\theta}) \qquad [2.23]$$

thus we recognize immediately that the ML solution coincides with the MAP solution when the prior is uniform. Uniform prior assumption means that, without knowledge of any observation, any value of parameters $\boldsymbol{\theta}$ (or any possible state of the structure) is equally likely. The $\boldsymbol{\theta}_{ML}$ can be seen as the most likely set of parameters given only the information from the monitoring system. More correctly, we must say that the uniform assumption is per se strong prior information, while a deeper discussion of the meaning of non-informative prior can be found in Kass and Wasserman (1994).

Nevertheless, the uniform prior assumption is very often used in physical data analysis, mainly because it allows easy calculation of the posterior parameters. In fact, if the probabilistic model is such that the likelihood is differentiable, the $\boldsymbol{\theta}_{ML}$ can be calculated by solving the system:

$$\theta_{ML}: \left[\frac{\partial \mathrm{pdf}(\mathbf{y}|\theta)}{\partial \theta}\right]_{\theta=\theta_{ML}} = \mathbf{0} \qquad\qquad [2.24]$$

Using ML means disregarding the prior information and this could produce unacceptable errors or even make the inference problem indeterminate.

LS. Starting from Equation [2.6], it is easily demonstrated that the LS solution of an identification problem is also the ML, under the assumption that the likelihood is normally distributed over the parameters; see for example van der Heijden *et al.* (2004) for a formal demonstration. Assuming uniform prior and normal likelihood it follows that $\theta_{MAP} = \theta_{ML} = \theta_{LS}$, which is to say that the traditional LS method provides the most likely solution, also in Bayesian terms, when the likelihood is normal and any value of the parameters is equally likely a priori. The normal likelihood assumption depends on the nature of the probabilistic model adopted. A special case is that of a linear Gaussian model, already discussed in Section 2.2.4.

2.2.6   Bayesian model selection

We discussed earlier the problem of parameter estimation under the assumption that the model is determinate, which is to say that we can use the same parametric model to reproduce the structural response independently of the structural state. However, there may be problems where the mechanical model changes on the basis of the state of the structure. For example, the moment–curvature relationship for a concrete beam changes depending on whether the beam is cracked or not. Thus we need different models for the original state and for the cracked case, and the interpretation of the observations changes based on the state of the structure.

In this case, it is convenient to divide the domain of the possible structural responses into a mutually exclusive and exhaustive set of scenarios $(S_1, S_2, ..., S_N)$, each defining the structural behaviour in a specific condition state. The structural response ${}^n r\,(t;{}^n\theta)$ for time t and sensor s_j in scenario n is controlled by a certain number of parameters ${}^n\theta$ (e.g. damage position, activation time, corrosion rate). The structural response is completely defined by specifying a scenario and a value for the correlated parameter set. Here, as in Bayesian model selection theory (Bretthorst, 1996; Beck and Katafygiotis, 1998; Yuen, 2002; Sivia, 2006), the discrete meta-parameter scenario identifies the response function, which in turn is specified by a parameter set. Bayes' rule assumes in this case the same expression of Equation [2.1]:

$$\mathrm{prob}\left(S_n|\mathbf{y}\right)=\frac{\mathrm{pdf}\left(\mathbf{y}|S_n\right)\mathrm{prob}\left(S_n\right)}{\mathrm{pdf}\left(\mathbf{y}\right)}\qquad\text{(2.1, repeated)}$$

and similarly the evidence assumes the same form of Equation [2.2]. However, since here the scenario response is controlled by a number of unknown parameters ${}^{n}\boldsymbol{\theta}$, the *likelihood* of scenario S_n, requires integration over the whole parameter domain $D^n\boldsymbol{\theta}$, using:

$$\mathrm{pdf}\left(\mathbf{y}|S_n\right)=\int_{D^n\boldsymbol{\theta}}\mathrm{pdf}\left(\mathbf{y}|{}^{n}\boldsymbol{\theta},S_n\right)\cdot\mathrm{pdf}\left({}^{n}\boldsymbol{\theta}|S_n\right)\cdot d^n\boldsymbol{\theta}\qquad[2.25]$$

where the first term in the integral is the likelihood of the S_n-parameters given scenario S_n, while the second is their prior distribution. In the assumption of uncorrelated observations, the likelihood can be calculated as:

$$\mathrm{pdf}\left(\mathbf{y}|{}^{n}\boldsymbol{\theta},S_n\right)=\prod_{i=1}^{m}\mathrm{pdf}\left(y_i|{}^{n}\boldsymbol{\theta},S_n\right)\qquad[2.26]$$

Once having calculated the posterior scenario probability, we can calculate the pdf of the parameters governing a specific scenario, again using Bayes' principle, which takes a form similar to Equation [2.6]:

$$\mathrm{pdf}\left({}^{n}\boldsymbol{\theta}|\mathbf{y},S_n\right)=\frac{\mathrm{pdf}\left(\mathbf{y}|{}^{n}\boldsymbol{\theta},S_n\right)\cdot\mathrm{pdf}\left({}^{n}\boldsymbol{\theta}|S_n\right)}{\mathrm{pdf}\left(\mathbf{y}|S_n\right)}\qquad[2.27]$$

When many parameters are involved, the exact integration of Equation [2.25] might require an exceptional computational effort and need to be circumvented with numerical techniques.

One of the simplest ways is to apply a Laplace asymptotic expansion, a deterministic method which approximates the posterior distribution of the parameters with an un-normalized multivariate Gaussian function (Beck and Katafygiotis, 1998). The method is applied independently for each scenario and, more in detail, requires the following two steps: (i) using an optimization technique, we find the MAP set of parameters ${}^{n}\boldsymbol{\theta}_{\mathrm{MAP}}$; (ii) we compute, at ${}^{n}\boldsymbol{\theta}_{\mathrm{MAP}}$, the Hessian $\mathbf{H}_p$ of the logarithm of the posterior distribution $\mathrm{pdf}\left(\mathbf{y}|{}^{n}\boldsymbol{\theta}\right)$: this quantity can be regarded as a measure of the fitting sensitivity. It can be shown (MacKay, 2003) that the likelihood of the scenario S_n can be derived by the following formula:

$$\mathrm{pdf}\left(\mathbf{y}\mid S_n\right) = \mathrm{pdf}\left({}^n\theta_{\mathrm{MAP}}\right)\left[2\pi\right]^{M/2}\det\left(\mathbf{H}_p\right)^{-1/2} \qquad\qquad [2.28]$$

where M is the number of parameters in the scenario.

Monte Carlo algorithms are alternative methods: their implementation may be computationally more demanding but their validity is more general, as they also work well with non-Gaussian distributions. They are classified as classical Monte Carlo methods and Markov chain Monte Carlo methods (MacKay, 2003): both rely on the possibility of drawing samples from a target distribution and of computing an integral, averaging along the sample. While the former family draws samples independently from a fixed distribution, the latter produces a step-by-step random walk in the parameter domain. The Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Beck and Au 2002) is possibly the most popular in this family. Parallel tempering (Gregory, 2005) is a robust evolution of Metropolis-Hastings, where several pseudo-random samples are produced in parallel by using different distributions: the algorithm allows information exchange between the simulations running in parallel. The reader is referred to the technical literature for more details of this topic; suggested readings are the textbooks by Brandt (1999), MacKay (2003), Bolstad (2010), Yuen (2010).

Example 4. Monitoring the tilt of a historic tower. Zonta *et al.* (2008) report the case of the leaning tower shown in Fig. 2.2, continuously monitored from October 2003 to October 2007. The main instrument installed is an inclinometer, based on a pendulum hanging from level $H = 29.90$ m, for which the position is returned in the form of two coordinates, x and y, representing the shift in directions West-East and North-South, respectively. In addition to the pendulum inclinometer, the system records the temperature field $\mathbf{T}(t)$ at a number of thermocouples. Figure 2.3a and 2.3b report a typical temperature record and the pendulum position in the two directions, x and y, during the whole monitoring period. The monitoring objective is to determine as soon as possible whether or not the tower tilt progresses with time. Using the formal approach introduced in this section, we can define two scenarios: in the first scenario, S_1, the tower inclination does not change with time, any pendulum shift from the mean position being due to daily and annual temperature changes; according to the second, S_2, the tower tilt is increasing, with a trend we can assume to be linear. In scenario S_1 the compensated inclination $\varphi_x(t)$ of the tower in direction x (i.e. east–west) is modelled as constant and equal to parameter ${}^1\varphi_{0,x}$. Conversely, in scenario S_2, the compensated inclination is a linear function trend w_x with an offset ${}^2\varphi_{0,x}$. We can see the misalignment measured by the pendulum as the sum of a term depending on tower inclination, a term linearly correlated with the temperatures recorded and a Gaussian noise. Particularly, for direction x:
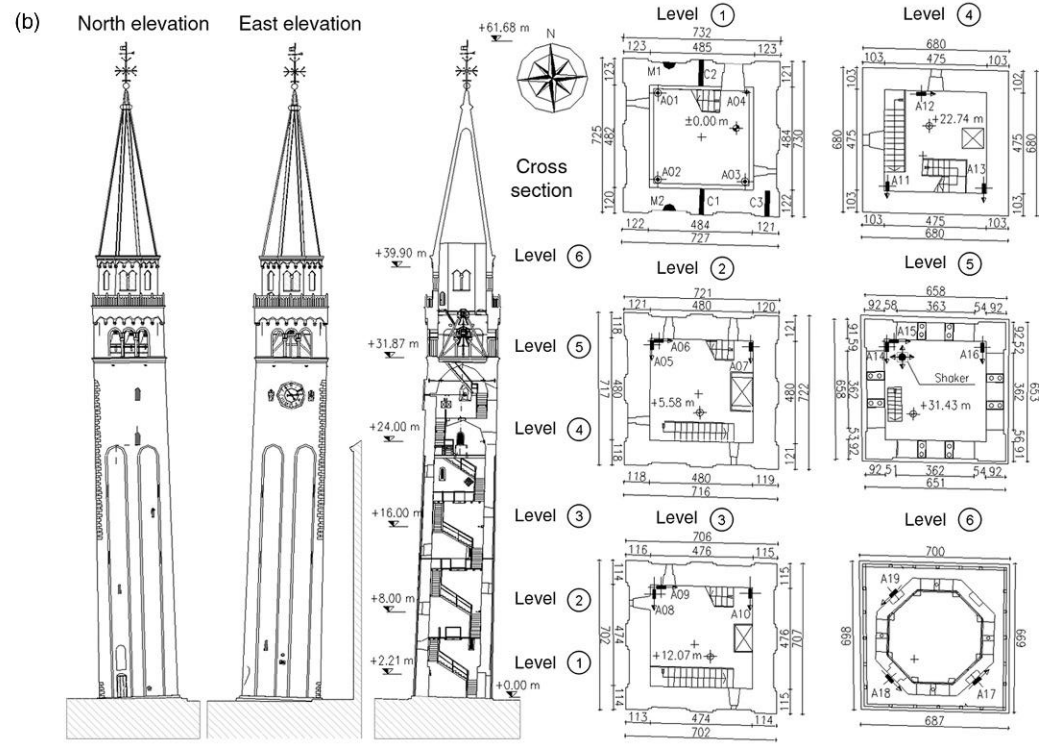
2.2 Overview of the tower (a); north and east elevation, cross-section and plan views at different levels (b), Zonta *et al.* (2008).
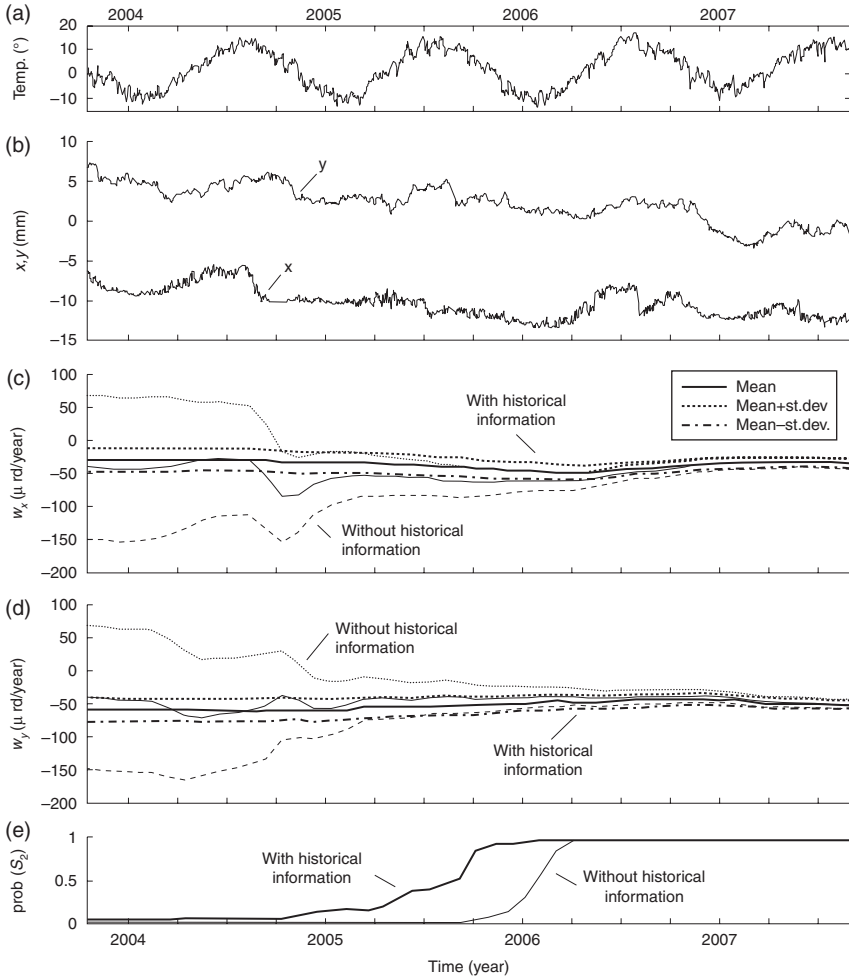
2.3 Temperature measurements at thermocouple (a); out of plumb measurements (b); posterior distribution of angular trend w_x (c); posterior distribution of angular trend w_y (d); posterior probability of scenario S_2 (e), Zonta *et al.* (2008).

$$
\begin{aligned}
x(t) &= \left({}^1\varphi_{0,x}\right)\cdot H + {}^1\mathbf{a}_x \cdot \mathbf{T}(t) + g(t) && \text{for } S_1 \\
x(t) &= \left({}^2\varphi_{0,x} + w_x \cdot t\right)\cdot H + {}^2\mathbf{a}_x \cdot \mathbf{T}(t) + g(t) && \text{for } S_2
\end{aligned}
\tag{2.29}
$$

where $\mathbf{a}_x$ is the linear transformation that correlates the temperatures to the misalignment, while indices 1 or 2 indicate that this vector generally assumes different values in different scenarios. Similar equations can be written for direction y. Following the notation outlined in this chapter, we can group the

parameters into a single vector: in scenario S_1, this vector is $^1\boldsymbol{\theta} = [^1\boldsymbol{\varphi}_{0,x} \ ^1\mathbf{a}_x \ ^1\boldsymbol{\varphi}_{0,y}$ $^1\mathbf{a}_y]^T$, while in scenario S_2 it is $^2 \boldsymbol{\theta} = [^2\boldsymbol{\varphi}_{0,x} \ w_x \ ^2\mathbf{a}_x \ ^2\boldsymbol{\varphi}_{0,y} \ w_y \ ^2\mathbf{a}_y]^T$.

The posterior probability of tilting is calculated assuming two different prior knowledge conditions: (i) a low prior credibility of tilting, equal to prob(S_2) = 1/1000, which reflects the actual perception of the owner at the beginning of the monitoring; (ii) a relatively higher credibility of tilting prob(S_2) = 3.7% based on technical information found in historical documents; the reader is referred to Zonta *et al.* (2008) for a detailed discussion on choice of the prior information. The light plot in Fig. 2.3e shows how the monitoring data modifies the system perception of an existing trend, assuming a low prior credibility of tilting. We can see that during the first 2 years of monitoring the probability of scenario S_2 is always close to zero, while only during the third year the monitoring information starts to alter the initial perception, to the point where in April 2006 the data are sufficient to convince the system that the tower is tilting. Similarly, the light plots of Fig. 2.3c and 2.3d show the evolution of the trend distributions w_x and w_y: we see that the trend estimates, which are very uncertain during the first 2 years, rapidly converge to more reliable values. The bold plots in the same diagrams reproduce the outcome of the inference process assuming the higher prior credibility of tilting: in this case, the system would theoretically become aware of the trend as early as September 2005; therefore, we can conclude that the historical information is roughly equivalent to 6 months of instrumental monitoring.

2.3   Data reduction

Permanent monitoring often results in a large amount of often highly redundant data from different types of sensor. In many cases, the dimensions M of a dataset can be very large, and many elements of the dataset may be redundant or irrelevant to the classification or identification process (van der Heijden *et al.*, 2004). Bayes' rule is a very powerful principle which lets us handle consistently any type of problem involving uncertainties: however, in practice the difficulty of its application grows quickly with the size and complexity of the dataset.

This section introduces the process of data reduction (Section 2.3.1), with special focus on Principal Component Analysis (Section 2.3.2). An overview of other popular feature reduction techniques is provided at the end of the Section 2.3.3.

2.3.1   Definition and scope

Reducing the size of the dataset is advisable for a number of reasons. The most obvious is that less data reduces the computational complexity and effort. However, we may believe that selecting or compressing data always

deteriorates the final result, while adding data in principle increases information but also increases noise. Van der Heijden *et al*. (2004, pages 183–185) demonstrate that increasing the dimensions of the dataset can cause an ultimate loss of monitoring performance; this mechanism is at the basis of the so-called *catastrophic data fusion* highlighted by Movellan and Mineiro (1998).

In general, most data reduction techniques are based on the concept of feature, where by feature we mean those components of the dataset that carry the information relevant to the monitoring inference problem. For example, in a damage detection problem, the features are the components of the dataset that are damage-sensitive. Data reduction occurs through two different steps: feature extraction and feature selection. Feature extraction is basically a transformation of the raw measurement space into the feature space (for example, from the acceleration records to the modal parameters). Feature selection basically consists in discarding those elements of the features vector that carry little or no information, and using the remainder.

Feature extraction and selection can be seen as processes that attempt to distil the knowledge hidden into the raw dataset; thus, data reduction is strictly related to the process of learning in Artificial Intelligence (Kantardzic, 2003). From this point of view, data reduction techniques are classified as: *supervised*, when the transformation from data space to the feature space is chosen based on the knowledge of the physical nature of the problem; and *unsupervised*, when the transformation is merely data driven. Based on the mathematical nature of the transformation, data reduction techniques are also classified as *linear* or *non-linear*. Because supervised techniques are, due to their nature, strictly application dependent, in this section we will focus principally on unsupervised techniques.

2.3.2   Principal component analysis (PCA)

PCA is possibly the most popular technique for reducing the dimensions of a dataset. Especially over recent years, PCA and its variants have found widespread application in structural health monitoring data analysis (see among many: Sohn *et al*., 2000, Zang and Imregun 2001; Giraldo *et al*., 2006; Ni *et al*., 2006; Li and Zhang 2006; Loh *et al*., 2011), as well as in data compression algorithms which are needed for optimizing data transmission in wireless sensor networks (see for example Park *et al*., 2007). The concept of principal component (PC) was first introduced by Hotelling (1933) although the method has its background in Pearson (1901). However, widespread interest in the method began only after the works of Anderson (1963), Rao (1964), Gower (1966) and Jeffers (1967), as correctly pointed out by Jolliffe (1986). Both Anderson's paper and Jolliffe's book are good references for approaching the method, but see also Kantardzic (2003).

Assume we have a dataset $\mathbf{y}$ resulting from monitoring a structure, instrumented with M sensors, labelled $(s_1, s_2, \ldots, s_M)$, each providing measurements at each of N_T time instants $(t_1, t_2, \ldots, t_{Nt})$. We denote as $^{(i)}y_k$ the sample acquired by sensor s_i at time t_k, and as $\mathbf{y}_k$ the acquisition of the whole M-dimensional sensor set at time t_k. The objective of PCA is to extract an m-dimensional linear projection of the M-dimensional dataset in the LS sense. Generally, data are correlated and the degree of correlation is well represented by the covariance matrix Σ. Because the covariance is not known a priori, we can use the sample covariance as its estimator:

$$\Sigma = \frac{1}{N_T} \sum_{k=1}^{N_T} (\mathbf{y}_k - \mu)(\mathbf{y}_k - \mu)^T \qquad [2.30]$$

where the mean value vector is defined as:

$$\mu = \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{y}_k \qquad [2.31]$$

We want first to recast the dataset $\mathbf{y}$ in a new coordinate system where the new variables $\mathbf{x}$, the data features or PCs, are uncorrelated. We call $\mathbf{A}$ the linear orthonormal transformation from the original data system to the feature system:

$$\mathbf{y}_k - \mu = \mathbf{A}\mathbf{x}_k \qquad [2.32]$$

Transformation $\mathbf{A}$ is found by solving the classical eigenvalue problem:

$$\Sigma \mathbf{A} = \mathbf{A} \Lambda \qquad [2.33]$$

where Λ is the diagonal matrix whose i-th diagonal element is the i-th eigenvalue λ_i of Σ. From Equation [2.33] it is clear that $\mathbf{A}$ collects the eigenvector $\mathbf{a}_i$ of covariance matrix Σ. We can sort the eigenvalues and label so that:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \qquad [2.34]$$

Because λ represents the variance of each PC of the dataset, the larger is λ, the higher is the quantity of information that the component carries. To reduce the initial M-dimensional dataset to an m-dimensional set, we can choose to represent it by the best subset of m PCs. An issue is how many PCs

do we need to obtain a meaningful representation of the original data. The most straightforward criterion is to select a subset of components whose cumulative variance is greater than a desired percentage $\bar{R}$ (for example: 90%). In formula:

$$R_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{M} \lambda_i} = \frac{\sum_{i=1}^{m} \lambda_i}{\mathrm{tr}(\Sigma)} \geq \bar{R} \tag{2.35}$$

where we note that here the cumulative variance is the trace of the covariance matrix.

When the dataset is heterogeneous, it is useful to carry out PCA using the correlation matrix, instead of the covariance, whose elements are defined as:

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sqrt{\sigma_{ii}^2 \sigma_{jj}^2}} \tag{2.36}$$

being σ_{ij}^2 the generic element of the covariance matrix. An interesting property of the correlation matrix eigenvalues is that they are all equal to one when the data records are mutually uncorrelated (and in this case the PCs coincide with the original variables); this suggests that a PC with variance less than one carries less information than any of the original records. In this case another popular method for selecting the subset dimension is to pick those PCs associated to eigenvalues greater than one; this criterion is sometime referred to as Kaiser's rule (Kaiser, 1960). Further criteria for selecting PCs are, for example, the broken stick method (Legendre and Legendre, 1983) and the 'scree' graph method (Cattel, 1966).
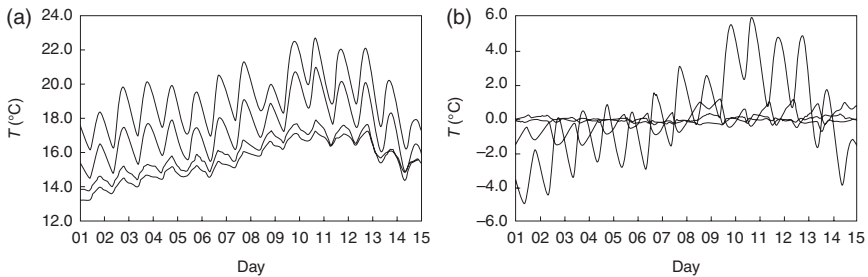
2.4 14-day temperature records from Zonta *et al.* (2008): sensor data (a) and projection onto the principal component space (b).

Table 2.1 Outcome of PCA on a 4-dimensional temperature dataset

Principal component i	Eigenvalue λ_i (°C^2)	Cumulative variance R_m	Eigenvector $\mathbf{a}_i^T$			
1	6.357	92.8%	0.595	0.348	0.612	0.387
2	0.434	99.2%	−0.147	0.714	−0.533	0.387
3	0.039	99.7%	0.734	−0.339	−0.580	0.428
4	0.017	100%	0.291	0.504	−0.057	−0.811

Example 5. PCA of a temperature dataset. Figure 2.4a represents the dataset $\mathbf{T}=[^{(1)}\mathbf{T}\ ^{(2)}\mathbf{T}\ ^{(3)}\mathbf{T}\ ^{(4)}\mathbf{T}]$ acquired in 14 days by four thermocouples in the application discussed in Zonta *et al.* (2008). The PCA of the dataset produces the results reported in Table 2.1. In particular, we observe that the first eigenvalue λ_1 is dominant with respect to the other three, having a variance corresponding to 92.8% of the total cumulative variance. This suggests we reduce the 4-dimensional dataset to a 1-dimensional record $^{(1)}\mathbf{x}$ which is the projection of the dataset on the first PC, according to:

$$^{(1)}\mathbf{x} = \mathbf{a}_1\mathbf{T} = 0.595\ ^{(1)}\mathbf{T} + 0.348\ ^{(2)}\mathbf{T} + 0.612\ ^{(3)}\mathbf{T} + 0.387^{(4)}\mathbf{T} \qquad [2.37]$$

where $\mathbf{a}_1$ is the first eigenvector. Figure 2.4b represents the dataset projected in the PC space and highlights the dominance of the first component. Using two PCs, the cumulative variance is 99.2%, meaning that using two PCs allows data reduction with virtually no loss of information.

2.3.3   Other feature reduction techniques

PCA is based on the information provided by the data only, and is performed by a simple linear transformation. However, the simplicity of the method is sometimes inappropriate given the nature of the dataset or the objective of the analysis. Many methods have been proposed to overcome PCA limits: we discuss in this section the most important of these methods, while an exhaustive review can be found in Skocaj's dissertation (2003) and Mitchell's textbook (2005).

   Probabilistic PCA. To properly account for the prior information, which is neglected in PCA, a probabilistic version of PCA was suggested in Tipping and Bishop (1999). The assumption is that the observed data $\mathbf{y}$ can be seen as the outcome of a linear Gaussian model controlled by a number m of latent parameters $\mathbf{x}$, of the type:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{\mu} + \mathbf{g} \qquad\qquad\qquad [2.38]$$

where $\mathbf{\mu}$ is as usual the mean value vector and $\mathbf{g}$ a zero mean Gaussian noise with variance σ^2 independent of the parameters. In practice, probabilistic PCAs reduce the problem of data reduction to a Bayesian parameter identification problem with a linear Gaussian model, which is discussed in Section 2.2.4. The reader can find the details of the mathematical manipulation in Tipping and Bishop (1999).

Multidimensional Scaling (MDS). MDS is a non-linear mapping technique proposed by Kruskal and Wish (1997), although it has its background in Sammon (1969) and is also known as Sammon mapping. The objective of MDS is to reduce the dimensions of the dataset while preserving, as far as possible, the Euclidean distance between the objects $\mathbf{y}_k$; this is done by minimizing a target function which cumulates the error of all possible mutual distances between objects. The classic application of MDS has the problem of representing a 3-dimensional space in 2 dimensions, while only recently has it been applied to mechanical system identification problems (see for instance Lopez and Sarigul-Klijn, 2009).

Kernel PCA (KPCA). Kernel PCA refers to a group of techniques that extend PCA to non-linear feature mapping functions, taking advantage of the properties of kernel functions. The general approach has its background in the work of Aizerman *et al*. (1964), while an extensive review of its applications can be found in the technical report by Schöelkopf *et al*. (1996). Recent applications of KPCA to damage detection and identification of mechanical systems are found in the work of He *et al*. (2007), Žvokelj *et al* (2011), Reynders *et al*. (2012). See also in Yan *et al*. (2005) an interesting way to override the complexity of KPCA in non-linear feature extraction using local PCA.

Linear Discriminant Analysis (LDA). This is a supervised linear mapping technique that has been popular mainly in pattern recognition problems. It assumes that each measurement $\mathbf{y}_k$ is associated with a class, or system state. While PCA seeks to maximize the variance of projections, LDA aims at maximizing the distance between the projected class means, while minimizing the variance within classes. Skocaj (2003) provides a broad overview on LDA and its variants, but see also Mitchell (2005).

2.4   Data fusion

Data fusion refers to techniques and tools used for combining sensor data, which is a typical problem encountered in structural monitoring. In this section, we introduce the concept of data fusion (Section 2.4.1) and we discuss some of the most popular techniques to handle multi-temporal (Section 2.4.2) and multi-sensor (Section 2.4.3) data, always based on Bayesian statistics. Alternative non-probabilistic logical schemes for handling uncertainties are outlined at the end of Section 2.4.4.

2.4.1   Definition and scope

Today there is still no univocal definition of data fusion. Historically, according Goodman *et al.* (1997), data fusion is the name given to a variety of problems arising primarily in military applications, and particularly in air-sea surveillance systems, against possible threats. Goodman defines data fusion as 'locate and identify an unknown number of unknown objects of many different types on the basis of different kind of evidence'. Therefore, apart from the specific application, the general problem in data fusion is to combine information from different sources to identify a physical quantity, and this has led over the years to the development of a theory that is independent of the application. Mitchell (2007) defines data fusion as 'the theory, techniques and tools which are used for combining sensor data, or data derived from sensory data, into a common representational format'. Regardless of the definition, data fusion entails combination of uncertain information of differing nature, and therefore it is clear to the reader that Bayesian logic as introduced in Section 2.2 can serve as a natural instrument in tackling the problem. Application of data fusion to structural health monitoring of structural mechanical systems has seen rapid growth in recent years, including applications in temperature compensation in civil structures (Balmès *et al.*, 2008); fault detection in mechanical systems (Chen *et al.*, 1995; Worden and Dulieu-Barton 2004; Subrahmanya *et al.*, 2010); monitoring adaptive structures (Garg *et al.*, 2002); vibration control (Tjepkema *et al.*, 2012); structural damage evaluation in CFRP structures (Su *et al.*, 2009); and damage location (Guo, 2006).

A classification of data fusion techniques is based on the object of fusion:

> In *multi-sensor data fusion*, fusion occurs across different sensors;
> In *multi-temporal data fusion*, fusion occurs across time at the same sensor;
> In *multi-temporal multi-sensor data fusion*, we combine data from different sensors and times.

Based on the nature of the information fused, the fusion may occur at the *data level* or at the *feature level*.

2.4.2   Multi-sensor data fusion

In the case of multi-sensor data fusion we distinguish:

> a *homogeneous case*, in which the dataset originates from sensors which nominally measure the same quantity and the objective is to determine the state $\boldsymbol{\theta}$;
> a *heterogeneous case*, in which the sensors involved measure different quantities which are used to estimate a further physical quantity x

(for example: we use a strain gauge and a thermometer to estimate the tension of a steel cable).

In both cases the problem is easily handled with Bayesian inference. Assume the case of two homogeneous sensors, which provide measurements ^1y and ^2y. The problem can be tackled using Bayes' rule in the form of Equation [2.6]:

$$\text{pdf}\left(\boldsymbol{\theta}|\mathbf{y}\right)=\frac{\text{pdf}\left(\mathbf{y}|\boldsymbol{\theta}\right)\cdot\text{pdf}\left(\boldsymbol{\theta}\right)}{\text{pdf}\left(\mathbf{y}\right)} \tag{2.6. repeated}$$

where the dataset $\mathbf{y}$ includes the two measurements $\{^1y, {}^2y\}$. Thus the likelihood is calculated by applying the joint probability relationship:

$$\text{pdf}\left(\mathbf{y}|\boldsymbol{\theta}\right)=\text{pdf}\left({}^1y\cap{}^2y|\boldsymbol{\theta}\right)= \text{pdf}\left({}^1y|\boldsymbol{\theta}\right)\text{pdf}\left({}^1y|{}^1y,\boldsymbol{\theta}\right) \tag{2.39}$$

and when the measurements are not correlated, Equation [2.39] simply reduces to Equation [2.8]. Equations [2.6] and [2.8] preserve their form independently of the number of observations.

2.4.3   Multi-temporal data fusion

When dealing with a dynamic system (i.e. a system which changes in time), the efficiency of data fusion can be improved by applying sequentially the inference process: this recursive application of Bayes' formula is commonly known as *Bayesian filtering*. Assume that we want to estimate in real time the state at each measurement time $(t_1, t_2, \ldots, t_k \ldots)$. We label $\mathbf{y}_k$ the measurement taken at time t_k and $\mathbf{y}_{1:k}$ the whole dataset from the first time interval (i.e. from the start of monitoring) to time t_k. Let $\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)$ be a prediction of parameters $\boldsymbol{\theta}_k$ at time t_k, based on the information $\mathbf{y}_{1:k-1}$ acquired to time t_{k-1}: Bayes' theorem allows us to update this distribution using the fresh data $\mathbf{y}_k$:

$$\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k}\right)=\frac{\text{pdf}\left(\mathbf{y}_k|\mathbf{y}_{1:k-1},\boldsymbol{\theta}_k\right)\cdot\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)}{\text{pdf}\left(\mathbf{y}_k|\mathbf{y}_{1:k-1}\right)} \tag{2.40}$$

Equation [2.40] is equivalent to the general formulation of Equation [2.6], with the difference that here the data are formally processed sequentially, and not all at once. Although very elegant, this sequential scheme does not

necessarily provide any advantage in computational terms over the classical scheme, because its implementation requires, at every step, recalculation of the prediction $\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)$ and the likelihood $\text{pdf}\left(\mathbf{y}_k|\mathbf{y}_{1:k-1},\boldsymbol{\theta}_k\right)$, which in turn depend on all the past data. The problem gets simplified when we make the assumption that $\boldsymbol{\theta}_k$, the state of system at time t_k, depends only on the state $\boldsymbol{\theta}_{k-1}$ at time t_{k-1}. In statistical terms, this is to say that the system state evolution follows a Markov process of order one. In this case, the prediction in Bayesian terms follows the calculation of the integral:

$$\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)= \int_{D\boldsymbol{\theta}_{k-1}} \text{pdf}\left(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}\right)\cdot\text{pdf}\left(\boldsymbol{\theta}_{k-1}|\mathbf{y}_{1:k-1}\right)\cdot d\boldsymbol{\theta}_{k-1} \qquad [2.41]$$

which depends on the distribution $\text{pdf}\left(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}\right)$ of the transition from state $k-1$ to k, and the posterior state estimate $\text{pdf}\left(\boldsymbol{\theta}_{k-1}|\mathbf{y}_{1:k-1}\right)$ which is known at time t_{k-1}. Then, once the sample $\mathbf{y}_{k+1}$ is acquired, the state prediction of Equation [2.41] can be updated (or corrected) with:

$$\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k}\right)=\frac{\text{pdf}\left(\mathbf{y}_k|\boldsymbol{\theta}_k\right)\cdot\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)}{\text{pdf}\left(\mathbf{y}_k|\mathbf{y}_{1:k-1}\right)} \qquad [2.42]$$

where now, contrary to Equation [2.40], the likelihood $\text{pdf}\left(\mathbf{y}_k|\boldsymbol{\theta}_k\right)$ is independent of past observations. Similarly, the evidence is calculated by integrating:

$$\text{pdf}\left(\mathbf{y}_k|\mathbf{y}_{1:k-1}\right)=\int_{D_*} \text{pdf}\left(\mathbf{y}_k|\boldsymbol{\theta}_k\right)\cdot\text{pdf}\left(\boldsymbol{\theta}_k|\mathbf{y}_{1:k-1}\right)\cdot d\boldsymbol{\theta}_k \qquad [2.43]$$

which does not involve quantities differing from the likelihood and the prediction. In summary, the estimation of the structural state is obtained by recursive application of Equation [2.42], known as *prediction*, and Equation [2.43], known as *correction* (or *update*).

The practical implementation of this recursive scheme is very burdensome, because as usual the problem is to calculate the integrals of Equations [2.42] and [2.43]. A useful simplification of the problem is possible when the model is linear Gaussian. In this case Equation [2.36] allows a closed form solution, which is usually referred to as the *Kalman filter* (Kalman, 1960; Kalman and Bucy, 1961). The linear Gaussian assumption can be formally written as:

$$\mathbf{y}_k = \mathbf{A}_k\boldsymbol{\theta}_k+\mathbf{g}_k \qquad [2.44]$$

where $\mathbf{g}_k$ is a zero mean Gaussian noise with covariance $\Sigma_{g,k-1}$. Also, assuming that the states at time k are linearly related to the parameters at time $k-1$:

$$\boldsymbol{\theta}_k = \mathbf{F}_k \boldsymbol{\theta}_{k-1} + \mathbf{v}_k \qquad [2.45]$$

where $\mathbf{v}_k$ is a zero mean Gaussian noise with covariance $\Sigma_{v,k-1}$ (linear Gaussian process assumption). Further, it is assumed that the prior distribution of the parameter vector is also Gaussian:

$$p(\boldsymbol{\theta}_0 | \mathbf{y}_0) = \mathrm{Norm}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \qquad [2.46]$$

Under these conditions, it is shown that the posterior pdfs are Gaussian, too:

$$\begin{aligned} p(\boldsymbol{\theta}_k | \mathbf{y}_{1:k-1}) &= \mathrm{Norm}(\boldsymbol{\mu}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1}) \\ p(\boldsymbol{\theta}_k | \mathbf{y}_{1:k}) &= \mathrm{Norm}(\boldsymbol{\mu}_{k|k}, \boldsymbol{\Sigma}_{k|k}) \end{aligned} \qquad [2.47]$$

and their mean value and covariance can be recursively calculated in closed form through the following prediction-correction scheme:

$$\boldsymbol{\mu} \begin{cases} \text{prediction} & \boldsymbol{\mu}_{k|k-1} = \mathbf{F}_k \boldsymbol{\mu}_{k-1|k-1} \\ \text{correction} & \boldsymbol{\mu}_{k|k} = \boldsymbol{\mu}_{k|k-1} + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}_{k|k-1} \right) \end{cases} \qquad [2.48]$$

$$\boldsymbol{\Sigma} \begin{cases} \text{prediction} & \boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{\Sigma}_{v,k-1} + \mathbf{F}_k \boldsymbol{\Sigma}_{k-1|k-1} \mathbf{F}_k^{T} \\ \text{correction} & \boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k-1} - \mathbf{K}_k \mathbf{A}_k \boldsymbol{\Sigma}_{k|k-1} \end{cases} \qquad [2.49]$$

where $\mathbf{K}_k$ is the so-called Kalman gain, defined as:

$$\mathbf{K}_k = \boldsymbol{\Sigma}_{k|k-1} \mathbf{A}_k^T \left(\mathbf{A}_k \boldsymbol{\Sigma}_{k|k-1} \mathbf{A}_k^T + \boldsymbol{\Sigma}_{g,k} \right)^{-1} \qquad [2.50]$$

Among the many applications of the Kalman filter to mechanical system identification, we wish to mention the work by the research groups Bernal (Bernal, 2005; Hernandez, 2011; Hernandez and Bernal, 2012), Smyth (Wu and Smyth, 2007; Wu and Smyth, 2008) and Brownjohn (Omenzetter and Brownjohn, 2006). Among other books, the reader can find further information on the Kalman filter and its application in Grewal and Andrews (2001) and in Zarchan and Mussov (2009).

2.4.4   Alternative non-probabilistic models

Although Bayes' theorem lets us state a rigorous mathematical formulation for any inference problem encountered in structural monitoring, it is a fact that civil engineers are not usually very familiar with probability theory, and are often discouraged by the complexity of its application. Indeed, implementations of probabilistic methods usually involve great computational effort to integrate the likelihood and evidence functions; it also requires a formal definition of all the uncertainties in the variables used in the inference process, which is not an easy task if one is unfamiliar with statistics.

These limits of Bayesian probability theory have justified in the past the redevelopment of alternative methods to handle uncertainties. Among these methods, possibly the most popular approaches are *fuzzy logic* and *interval algebra*. Below we provide a brief outline of these approaches.

Fuzzy logic. Fuzzy set theory was first introduced by Zadeh (1965) as a way to handle semantically imprecise concepts. Fuzzy logic is based on the concept of membership function. While in deterministic logic an element x can belong or not to set A, in fuzzy logic this is defined through a membership function $\mu_A(x)$; the metric of the membership is a value ranging between 1 and 0, known as grade of truth, or truth value. In fuzzy logic the classical Boolean logic operators (NOT, AND, OR) are redefined and often referred to as Zadhe operators. For example NOT is obviously defined as:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \tag{2.51}$$

Operators AND and OR are not univocally defined but are usually defined as follows: the degree of truth of x being simultaneously in A AND B (which is to say x is in the interception of A and B, $A \cap B$), is the minimum grade of truth of A and B

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \tag{2.52}$$

Similarly, operator OR is usually defined as the maximum grade of truth of the two

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \tag{2.53}$$

It is almost natural to seek some analogy between truth value and Bayesian probability; specifically, the grade of truth, varying between 0 and 1, strictly recalls the concept of probability. However, comparison

between the expression of the joint probability and Equation [2.52] shows that fuzzy logic coincides with probability only under special conditions, and in any case does not have the flexibility to handle correlation. Critics of fuzzy logic note, often on the basis of de Finetti's theory (1993), the lack of a need for a method for handling uncertainties, alternative to a probabilistic method.

Interval-based techniques. Interval algebra, first disseminated by the work of Moore (1966), is an even rougher way to handle uncertainties. While in deterministic logic the value of a variable x is determinate, and in probabilistic analysis is defined by a distribution pdf(x), in interval analysis the uncertainty of a variable is defined through an interval $\mathbf{x}$:

$$\mathbf{x} = \left[\underline{x}, \overline{x}\right] \tag{2.54}$$

which means that the variable $\mathbf{x}$ is expected to range from a minimum value of $\underline{x}$ to a maximum of $\overline{x}$. Interval algebra provides a formal extension of classic algebraic operators; for example, sum and difference are redefined as:

$$\mathbf{x} + \mathbf{y} = \left[\underline{x}, \overline{x}\right] + \left[\underline{y}, \overline{y}\right] = \left[\underline{x} + \underline{y}, \overline{x} + \overline{y}\right] \tag{2.55}$$

$$\mathbf{x} - \mathbf{y} = \left[\underline{x}, \overline{x}\right] - \left[\underline{y}, \overline{y}\right] = \left[\underline{x} - \overline{y}, \overline{x} - \underline{y}\right] \tag{2.56}$$

Interval analysis has gained some popularity in the past, and is sometimes still used today for its simplicity and easy interpretation of the results. From a statistical perspective, it can be seen as a very rough approximation of the probability theory, where distributions are conservatively forced to be uniform, and correlation is ignored. These simplifications often cause too conservative results: for example, direct application of Equation [2.6] produces the following apparent paradox:

$$\mathbf{x} - \mathbf{x} = \left[\underline{x}, \overline{x}\right] - \left[\underline{x}, \overline{x}\right] = \left[\underline{x} - \overline{x}, \overline{x} - \underline{x}\right] \tag{2.57}$$

for which an interval minus itself is not equal to zero, but to an interval centred on zero. That said, using intervals can be useful when scarce information makes it difficult to define the distributions necessary to state the inference problem in rigorous probabilistic terms. The reader is also referred to the textbook by Alfred and Herzberger (1983) for an exhaustive explanation of interval logic.

2.5   Further trends

The process of fusing data from different sources into a single piece of information is often conceptually represented with a *fusion cell or fusion node* (Mitchell, 2005). For example, we can conceive the temperature-compensated strain as the output of a fusion node where the rough strain observation is fused with the temperature observation. Rather than being a physical entity, the fusion node is an object in a logical network, which represents the logical interconnection of the different components in the fusion process. The output of a fusion node can be seen as a sensor in logical terms, even if this has no connection with the physical architecture of the sensor network. Imagine for example a monitoring system including a network of strain gauges distributed over a bridge, all wired to a single interrogation unit, and a second network of thermocouples paired with the strain gauges but independently wired to a second interrogation unit; the two interrogation units are in turn connected to a computer, which analyses and stores the data. Temperature compensation of an individual strain gauge occurs by combining its data with that of the corresponding thermocouple. Therefore, from a logical point of view, data from each strain gauge–thermocouple pair are locally fused in a compensated measurement, through a fusion node. However, from the physical point of view, all data processing occurs at the computer CPU, and therefore there is no direct relationship between the logical architecture (software) and physical architecture (hardware) of the system.

An alternative solution is to physically combine the parent sensors in a single device with a CPU to carry out data acquisition and processing at the local level. In this case a computer system designed for the specific application is *embedded* in the device, in contrast to the computer system needed to logically process the data at the network level. This device, sometimes referred to as sensor node, embeds not only hardware but also the logic of the fusion node. The output of the node is the result of data fusion (the compensated strain in our example) and the device itself can be seen as a sensor from the point of view of network logic. Embedding a fusion block in a single device scales the level of abstraction of the network and reduces its complexity and cost while facilitating deployment, scalability and operation. Wireless sensor networks are a typical application of embedded systems: in this case the sensor node is embedded. Wireless sensor networks are discussed in Section 10.4 of this book, while key references on the topics are Lynch and Loh (2006) and Lynch (2007).

Apart from providing information on the state, the utility of the monitoring system is to drive decisions by the owner as to infrastructure maintenance, repair, retrofit or replacement. There are cases when knowledge of the state automatically drives the decision (going back to the introductory example:

the fuel gauge is on, so we stop to refuel). In other cases the knowledge of the state does not automatically entail a specific action, because the decision requires prediction of the effect of the action. When the effect can be quantified in economic terms, the decision problem can be formally addressed by the utility theory (von Neumann and Morganstern, 1944; Jensen, 1967), which is discussed in its general terms in many modern textbooks (see for example Parmigiani and Lurdes, 2009). An application of decision theory to a civil engineering monitoring problem is discussed in the tutorial paper by Glisic *et al*. (2012), which recognizes the utility of monitoring using the concept of value of information (Russell and 1991), but see also Pozzi *et al*. (2010) and van der Hijden *et al*. (2004).

2.6   Sources of further information and advice

Bayesian theory of probability is clearly stated in many modern textbooks including Jaynes (2003), Gregory (2005) and Sivia (2006). For a historical background, good readings are Weber (1973), Stigler (1986) and of course Bayes' essay (Bayes, 1763), where you might be surprised to see that he was not Bayesian in the modern sense. Bayesian numerical techniques are the subject of a number of specialized textbooks, including Brandt (1999), McKay (2003), Bolstad (2010) and Yuen (2010). The definitive work on PCA is Jolliffe (1986), while for alternative data reduction techniques suggested readings are Bevington and Robinson (2003), Kantardzic (2003), van der Hijden *et al*. (2004), Mitchell (2005) and Skocaj (2003). Mitchell (2005) is also one of the most comprehensive works approaching data fusion, but see also Clark and Yuille (1990), Goodman *et al*. (1997) and again van der Hijden *et al*. (2004). Among many, a suggested key reference for fuzzy set theory is Duboi and Prade (1980), while for interval analysis see Moore (1966) and Alfred and Herzberger (1983).

2.7   Acknowledgements

The author wishes to thank Matteo Pozzi (Carnagie Mellon University) and Federico Bruschetta (University of Trento) for their help in reviewing the final version of this chapter.

2.8   References

Aizerman MA, Braverman EM and Rozonoer LI (1964), 'Theoretical foundations of the potential function method in pattern recognition learning', *Automation and Remote Control*, **25**, 821–837.

Balmès E, Basseville M, Bourquin F, Mevel L, Nasser H and Treyssède F (2008), 'Merging sensor data from multiple temperature scenarios for vibration monitoring of civil structures', *Structural Health Monitoring*, **7**(2), 129–142.

Bayes T (1763), 'An essay toward solving a problem in the doctrine of chances', *Philosophical Transactions of the Royal Society of London*, **53**, 370–418.

Beck JL and Au SK (2002), 'Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation', *Journal of Engineering Mechanics*, **128**(4), 380–391.

Beck JL and Katafygiotis LS (1998), 'Updating models and their uncertainties, I: Bayesian statistical framework', *Journal of Engineering Mechanics*, **124**(2), 455–461.

Bernal D. (2005), 'Closely spaced roots and defectiveness in second-order systems', *Journal of Engineering Mechanics*, **131**(3), 276–281.

Bevington PR and Robinson DK (2003), *Data Reduction and Error Analysis for the Physical Sciences* (3rd Edn.), New York, McGraw-Hill.

Bolstad WM (2010), *Understanding Computational Bayesian Statistics*, New York, Wiley.

Brandt S (1999), *Data Analysis: Statistical Computational Methods for Scientists and Engineers*, New York, Wiley.

Bretthorst GL, 'An introduction to model selection using probability theory as logic', In Heidbreder GR (Ed), *Maximum Entropy and Bayesian Methods*, New York, Springer.

Cattell RB (1966), 'The scree test for the number of factors', *Multivariate Behavioral Research*, **1**, 245–276.

Chen YD, Du R and Qu LS (1995), 'Fault features of large rotating machinery and diagnosis using sensor fusion', *Journal of Sound and Vibration*, **188**(2), 227–242.

Clark JJ and Yuille AL (1990), *Data Fusion for Sensory Information Processing Systems*, Dordrecht (NL), Kluwer.

de Finetti B (1993), *Probabilità e induzione*, Bologna, Clueb.

Dubois D and Prade H (1980), *Fuzzy Sets and Systems: Theory and Applications*, New York, Academic Press.

Garg DP, Zikry MA, Anderson GL and Stepp D (2002), 'Health monitoring and reliability of adaptive heterogeneous structures', *Structural Health Monitoring*, **1**(1), 23–39.

Giraldo FD, Dyke SJ and Caicedo JM (2006), 'Damage detection accommodating varying environmental conditions', *Structural Health Monitoring*, **5**(2), 155–172.

Glisic B (2011), 'Streicker bridge: An on-site SHM laboratory at Princeton University campus', *Proc. First Middle East Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures, Paper No. 306*, Dubai, 8–10 February.

Glisic B, Adriaenssens S and Zonta D (2012), 'The impact of monitoring on decision making', *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2012, Proceedings of SPIE Vol. 8345*, Bellingham, (WA), SPIE.

Goodman IR, Mahler RPS and Nguyen HT (1997), *Mathematics of Data Fusion*, Dordrecht (NL), Kluwer.

Gregory P (2005), *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge (UK) Cambridge University Press.

Grewal MS and Andrews AP (2001), *Kalman Filtering* (2nd Edn.), New York, Wiley.

Guo HY (2006), 'Structural damage detection using information fusion technique', *Mechanical Systems and Signal Processing*, **20**(5), 1173–1188.

Hall DL and McMullen SAH (2004), *Matematical Techniques in Multisensor Data Fusion* (2nd Edn.), Norwood, MA, Artech House Inc.

He Q, Kong F and Yan R (2007), 'Subspace-based gearbox condition monitoring by kernel principal component analysis', *Mechanical Systems and Signal Processing*, **21**(4), 1755–1772.

Hernandez EM (2011), 'Optimal model-based state estimation in mechanical and structural systems', *Structural Control & Health Monitoring*, **20**(4), 532–543. doi: 10.1002/stc.513.

Hernandez EM and Bernal D (2012), 'State estimation in structural systems with model uncertainties', *Journal of Engineering Mechanics*, **134**(3), 252–257.

Hotelling H (1933), 'Analysis of a complex of statistical variables into principal components', *Journal of Educational Psychology*, **24**, 417–441.

Jaynes ET (2003), *Probability Theory: The Logic of Science*, Cambridge (UK), Cambridge University Press.

Jensen NE (1967), 'An introduction to Bernoullian utility theory: I Utility functions,' *Swedish Journal of Economics*, **69**, 163–183.

Jolliffe IT (1986), *Principal Component Analysis*, New York, Springer.

Kaiser HF (1960), 'The application of electronic computers to factor analysis,' *Educational and Psychological Measurement*, **20**, 141–151.

Kalman RE (1960), 'A new approach to linear filtering and prediction problem', *Transactions of the ASME, Journal of Basic Engineering,* **82D**, 35–45.

Kalman RE and Bucy R (1961), 'New results in linear filtering and prediction', *Transactions of the ASME, Journal of Basic Engineering*, **83D**, 95–108.

Kantardzic M (2003), *Data Mining: Concepts, Models, Methods, and Algorithms*, New York, Wiley.

Kass RE and Wasserman L (1994), *Formal Rules for Selecting Prior Distributions. A Review and Annotated Bibliography, Technical Report #583*, *Department of statistics,* Pittsburgh (PA), Carnegie Mallon University.

Legendre L and Legendre P (1983), *Numerical Ecology*. New York, Elsevier.

Li J. and Zhang Y. (2006), 'Interactive sensor network data retrieval and management using principal components analysis transform', *Smart Materials and Structures*, **15**, 1747–1757.

Loh CH, Chen CH and Hsu TY (2011), 'Application of advanced statistical methods for extracting long term trends in static monitoring data from an arch dam', *Structural Health Monitoring*, **10**(6), 587–601.

Lopez I and Sarigul-Klijn N (2009), 'Distance similarity matrix using ensemble of dimensional data reduction techniques: Vibration and aerocoustic case studies', *Mechanical Systems and Signal Processing*, **23**(7), 2287–2300.

Lynch JP (2007), 'An overview of wireless structural health monitoring for civil structures', *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **365**(1851), 345–372.

Lynch JP and Loh K (2006), 'A summary review of wireless sensors and sensor networks for structural health monitoring', *Shock and Vibration Digest*, **38**(2), 91–128.

MacKay DJC (2003), *Information Theory, Inference and Learning Algorithms*, Cambridge (UK), Cambridge University Press.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A and Teller E (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics*, **21**(6), 1087–1092.

Mitchell HB (2007), *Multi-Sensor Data Fusion*, Heidelberg, Springer.

Moore RE (1966), *Interval Analysis*, Englewood Cliffs (NJ), Prentice Hall.

Movellan JR and Mineiro P (1998), 'Robust sensor fusion: Analysis and application to audio visual speech recognition', *Machine Learning*, **32**, 85–100.

Ni YQ, Zhou XT and Ko JM (2006), 'Experimental investigation of seismic damage identification using PCA-compressed frequency response functions and neural networks', *Journal of Sound and Vibration*, **290** (1–2), 242–263.

Papadimitriou C, Beck JL and Katafyogiotis LS (1997), 'Asymptotic expansion for reliability and moments of uncertain systems', *Journal of Engineering Mechanics*, **123**(12), 380–391.

Park S, Lee JJ, Yun CB and Inman DJ (2007), 'Electro-mechanical impedance-based wireless structural health monitoring using PCA-data compression and K-means clustering algorithms', *Journal of Intelligent Material System and Structures*, **19**(4), 509–520.

Parmigiani and Lourdes (2009), *Decision Theory: Principles and Approaches*, Chichester (UK), Wiley.

Pearson K (1901), 'On lines and planes of closest fit to systems of points in space', *Philosophical Magazine*, **2**(6), 559–572.

Pozzi M, Zonta D, Wang W and Chen G (2010), 'A framework for evaluating the impact of structural health monitoring on bridge management,' *Proc. 5th International Conf. on Bridge Maintenance, Safety and Management (IABMAS2010)*, Philadelphia, 11–15 July.

Russell SJ and Wefald E (1991), *Do the Right Thing*, Cambridge (MA), MIT Press.

Sivia DS (2006), *Data Analysis: A Bayesian Tutorial*, Oxford, Oxford University Press.

Schölkopf B, Smola A and Müller KR (1996), *Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Technical Report No. 44*, Tübingen, Max-Plank-Institute für biologische Kybernetik.

Skocaj D (2003), *Robust Subspace Approaches to Visual Learning and Recognition,* PhD Thesis, University of Lubijana.

Sohn H, Czarneki JA and Farrar CR (2000), 'Structural health monitoring using statistical process control', *Journal of Structural Engineering*, **126**(1), 1356–1363.

Stigler SM (1986), *The History of Statistics*, Cambridge (MA), Harvard University Press

Su Z, Wang X, Cheng L., Yu L and Chen Z (2009), 'On selection of data fusion schemes for structural damage evaluation', *Structural Health Monitoring,* **8**(3), 223–241.

Subrahmanya N, Shin YC, Meckl PH (2010), 'A Bayesian machine learning method for sensor selection and fusion with application to on-board fault diagnostics', *Mechanical Systems and Signal Processing*, **24**(1), 182–192.

Tipping ME and Bishop CM (1999), 'Probabilistic principal component analysis', *Journal Royal Statistical Society*, **61B**, 611–622.

Tjepkema D, van Dijk J and Soemers HMJR (2012), 'Sensor fusion for active vibration isolation in precision equipment', *Journal of Sound and Vibration*, **331**(4), 735–749.

van der Heijden F, Duin RPW, Ridder D and Tax DMJ (2004), *Classification, Parameter Estimation and State Estimation*, New York, Wiley.

von Neumann J and Morgenstern O (1944), *Theory of Games and Economic Behavior*, Princeton (NJ), Princeton University Press.

Weber JD (1973), *Historical Aspects of the Bayesian Controversy*, Tucson (AZ), The University of Arizona.

Wenzel H (2011), verbal communication, panel discussion at the 9th International Workshop on Structural Health Monitoring, Stanford (CA), Stanford University, September 15.

Worden K and Dulieu-Barton JM (2004), 'An overview of intelligent fault detection in systems and structures', *Structural Health Monitoring*, **3** (1), 85–98.

Wu M and Smyth A (2007), 'Application of the unscented Kalman filter for real-time, nonlinear structural system identification', *Structural Control and Health Monitoring*, **14**, 971–990.

Wu M and Smyth A (2008), 'Real-time parameter estimation for degrading and pinching hysteretic models', *International Journal of Non-Linear Mechanics*, **43**, 822–833.

Yan AM, Kerchen G, De Boe P and Golinval JC (2005), 'Structural damage diagnosis under changing environmental conditions-part II: Local PCA for non-linear cases.' *Mechanical Systems and Signal Processing*, **19**, 865–880.

Yuen KV (2002), *Model Selection, Identification and Robust Control for Dynamical Systems, Ph.D. thesis, Report EERL 2002–03*, Pasadena (CA), California Institute of Technology.

Yuen KV (2010), *Bayesian Methods for Structural Dynamics and Civil Engineering*, New York, Wiley.

Zadeh LA (1965), 'Fuzzy sets', *Information and Control*, **8**(3), 338–353.

Zarchan P and Musoff H (2009), *Fundamentals of Kalman Filtering* (3rd Edn.), Reston (VA), American Institute of Aeronautics and Astronautics.

Zang C and Imregun M (2001), 'Structural damage detection using artificial neural networks and measured FRF data reduced via principal component projection', *Journal of Sound and Vibration*, **242**(5), 813–827.

Zonta D, Pozzi M, Zanon P, Anese GA and Busetto A (2008), 'Real-time probabilistic health monitoring of the Portogruaro Civic Tower', in D'Ayala D and Fodde E, *Structural Analysis of Historic Construction: Preserving Safety and Significance,* London, CRC Press, 723–731.

Žvokelj M, Zupan S and Prebil I (2011) 'Non-linear multivariate and multiscale monitoring and signal denoising strategy using kernel principal component analysis combined with ensemble empirical mode decomposition method', *Mechanical Systems and Signal Processing*, **25**(7), 2631–2653.

3

Analytical techniques for damage detection and localization for assessing and monitoring civil infrastructures

D. BERNAL, Northeastern University, USA

Abstract: The chapter contains a review of some aspects of structural dynamics and includes a section on damage detection and damage localization from vibration signals. Novelty detection is treated in detail, with special attention given to the situation where the environmental variables are not measured. Two techniques that can detect damage without the need for repeated identifications are presented (one based on a subspace approach and one on the Kalman filter) and a discussion of change detection using a cumulative sum chart from process control is included. The techniques presented are covered in sufficient detail to allow implementation by an interested reader.

Key words: damage detection, damage localization, environmental effects, change detection.

3.1   Introduction

The chapter contains a review of some aspects of structural dynamics and includes a section on damage detection and damage localization from vibration signals. The material is not a survey but a description of some specific techniques in sufficient detail to allow implementation by an interested reader.

3.2   Linear time invariant systems

The equations of dynamic equilibrium for a linear time invariant finite dimensional system with viscous dissipation can be written as

$$M\ddot{q}(t) + C_{dam}\dot{q}(t) + Kq(t) = b_2 f(t) \qquad [3.1]$$

where: $q \in R^{n \times 1}$ is the displacement vector; n is number of degrees of freedom (DOF); M, C_{dam}, and $K \in R^{n \times n}$ are the symmetric mass, damping, and

67

stiffness matrices, respectively; $b_2 \in R^{n \times r}$ is a vector describing the spatial distribution of the forces $f(t) \in R^{r \times 1}$; and dots indicate differentiation.

First order formulation

Augmenting Equation [3.1] with the equality $\dot{q}(t) = \dot{q}(t)$ gives

$$\begin{Bmatrix} \dot{q}(t) \\ \ddot{q}(t) \end{Bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C_{dam} \end{bmatrix} \begin{Bmatrix} q(t) \\ \dot{q}(t) \end{Bmatrix} + \begin{bmatrix} 0 \\ M^{-1}b_2 \end{bmatrix} f(t) \qquad [3.2]$$

taking

$$z = \{q(t) \quad \dot{q}(t)\}^T \qquad [3.3]$$

and defining

$$A_z = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C_{dam} \end{bmatrix} \qquad [3.4]$$

and

$$B_z = \begin{bmatrix} 0 \\ M^{-1}b_2 \end{bmatrix} \qquad [3.5]$$

gives

$$\dot{z}(t) = A_z z(t) + B_z f(t) \qquad [3.6]$$

Measurements are typically a linear combination of the state variables with the possibility of a direct transmission from the input, namely

$$y(t) = C_z z(t) + D_z f(t) \qquad [3.7]$$

where $y(t) \in R^{m \times 1}$ and D_z is known as the direct transmission term. Examination shows that the map from the inputs to the state, and from the state to the outputs, is not unique; indeed, letting

$$x(t) = Tz(t) \qquad [3.8]$$

where T is invertible but otherwise arbitrary, one has

$$\dot{x}(t) = A_c x(t) + B_c f(t) \qquad [3.9]$$

and

$$y(t) = C_c x(t) + D_c f(t) \tag{3.10}$$

where

$$A_c = T^{-1} A_z T \qquad B_c = T^{-1} B_z \qquad C_c = C_z T \qquad D_c = D_z \tag{3.11a-d}$$

Two points to note are: (1) that the A_c matrices are related by a similarity transformation, and thus have the same eigenvalues; and (2) that D_c is independent of T.

3.3   Modal form

Assuming that A_c is not defective, a spectral decomposition gives

$$A_c = \Psi \Lambda \Psi^{-1} \tag{3.12}$$

where Ψ are the eigenvectors of A_c and $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_{2n})$ is the diagonal matrix of the complex eigenvalues. Introducing the coordinate transformation

$$x(t) = \Psi Z(t) \tag{3.13}$$

gives

$$\dot{Z}(t) = \Lambda Z(t) + \Gamma^T f(t) \tag{3.14}$$

and

$$y(t) = \Theta Z(t) + D_c f(t) \tag{3.15}$$

where

$$\Gamma^T = \Psi^{-1} B_c \tag{3.16}$$

and

$$\Theta = C_c \Psi \tag{3.17}$$

where Θ are the complex modes of the physical system at the measured coordinates. Taking a Laplace transform of Equations [3.14] and [3.15], assuming zero initial conditions, and eliminating the internal state, one gets

$$y(s) = G(s) \times f(s) \tag{3.18}$$

where the transfer matrix is

$$G(s) = \Theta [I \times s - \Lambda]^{-1} \Gamma^T + D_c \qquad\qquad [3.19]$$

Taking $\Theta = [\varphi_1\ \varphi_2, ..., \varphi_{2n}]$ and $\Gamma = [t_1, t_2, ..., t_{2n}]$ where t_j is a column vector, Equation [3.19] can be written as

$$G(s) = \sum_{j=1}^{2n} \frac{\varphi_j t_j^T}{s - \lambda_j} + D_c \qquad\qquad [3.20]$$

For under-damped systems the eigenvalues appear in complex conjugate pairs and it is customary to write Equation [3.20] as

$$G(s) = D_c + \sum_{j=1}^{n} \frac{\gamma_j}{s - \lambda_j} + \frac{\gamma_j^*}{s - \lambda_j^*} \qquad\qquad [3.21]$$

where $\gamma_j = \varphi_j t_j^T$ is the residue of mode j and the asterisk stands for conjugation. The direct transmission term is non-zero only in the case where $G(s)$ is Accelerance and there is collocation. In the common case where one is interested in the relation between forces and displacements or velocities, D_c is zero and one has

$$G(s) = \sum_{j=1}^{n} \frac{\gamma_j}{s - \lambda_j} + \frac{\gamma_j^*}{s - \lambda_j^*} \qquad\qquad [3.22]$$

which is the transfer matrix in pole residue form.

3.4   Relation between the complex and the normal mode models

The eigenvalues of the first order formulation are complex and carry information on the homogeneous solution; namely, the real part gives the rate of decay and the imaginary the damped vibration frequency. If the system is classically damped, one has

$$\lambda_j = -\omega_j \xi_j + i\omega_j \sqrt{1 - \xi_j^2} \qquad\qquad [3.23]$$

where ω = undamped frequency and ξ = the modal damping ratio. In practice, Equation [3.23] is used to estimate ω and ξ from the complex eigenvalues, even when the classical damping premise is not satisfied. The simplest

approximation of normal modes, given the complex modes, is obtained by normalizing the vectors so that the entry with the largest amplitude is real, and discarding the imaginary components.

3.5   Damage detection

Detection is the first level in a damage assessment scheme. The objective in detection is to announce, with a predetermined level of confidence, whether a structure is in its reference state or not. When the damage detection algorithm incurs an error by announcing damage, while the system is healthy, there is a false positive or Type I error. On the contrary, if it announces that the system is in the reference condition when some damage has taken place, then one has a false negative or Type II error.

The resolution of the damage detection is characterized by the power of the test (POT), which is the probability that the test classifies data from the damaged system correctly given a pre-selected Type I error rate α, e.g. $\alpha = 5\%$. The POT is particular to each damage and is the complement of the Type II error β. In reference to Fig. 3.1a, it is the area to the right of the threshold on the damaged state distribution. An alternative graphical depiction of the effectiveness of an algorithm to detect a particular damage is the receiver operating characteristic (ROC) curve. The ROC curve plots the Type I error on the horizontal axis vs the POT on the vertical, each point corresponding to a different discriminating threshold. The ROC curve for the distributions depicted in Fig. 3.1a is illustrated in Fig. 3.1b.

Detection can be based on model updating or be data-driven. In the model updating approach, the measured data are used to estimate the parameters of a (physics-based) model and damage or no-damage is inferred by inspection of the parameter values. Data-driven algorithms decide on damage by contrasting identification models obtained at different times, or by inspecting residuals that reflect how well the recorded data fit an identification
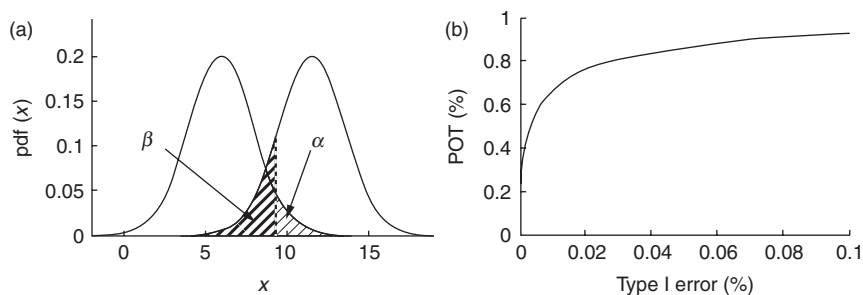
3.1 (a) Distributions of discriminating metric in reference and damaged states and (b) ROC curve for the distributions in (a).

model of the reference state. This chapter considers the multiple identification approach and the residual-based scheme, but does not discuss the model updating alternative.

We note, for clarity, that the term point is used here to refer to an element of a vector space. A point, for example, can be a collection of frequencies and mode shapes listed on a vector. 'Input space' is used to indicate the support of the points (e.g. all the real positive numbers if the points are vectors collecting frequencies) while 'feature space' is the scalar valued space of the metric used to decide on the system's state. In some settings the input and the feature spaces are the same. We define the term pre-damaged image (PDI) as the location in input space that a damaged point 'would have occupied' (given the present environmental conditions) if the structure were undamaged. Damage, therefore, can be viewed as the translation from the PDI to the position where the point is actually observed with time frozen. We designate the vector from the PDI to the damaged location as δx.

Novelty detection

Let $t_1, t_2, \ldots, t_L$ be times indicating the initiation of measurements, and $\Delta t_1, \Delta t_2, \ldots, \Delta t_L$ the duration of the records. We assume that the durations Δt_j $\forall j$ are short compared to the time scale in which the system properties change due to environmental fluctuations. Let the recorded data be used to obtain a vector $x \in \mathbb{R}^{m \times 1}$ that depends on system properties. One has, therefore, $X = \{x_1 \quad x_2 \quad \cdots \quad x_L\} \in \chi$. Assume it is known that the structure, for $j = 1, \ldots, L$, is in the undamaged state. The issue in damage detection is to decide, informed by the data in X, on the subset of the input space, S, for which data points are to be classified as normal.

The decision on S is often realized by specifying a function, or algorithm, that operates on x and produces a scalar $\rho = g(x)$ that is in some way correlated with the probability density of the points in input space. S is then taken as the collection of all points for which $\rho \geq \kappa$, where κ is a threshold. The assumption that the points in X are samples from a single distribution is justified when these points are compensated for environmental fluctuations, but otherwise they should be thought of as coming from a collection of distributions that depend on the environmental state. This chapter does not discuss the formulation of environmental models, but interested readers can consult the work presented in Sohn *et al.* (2001), who used Artificial Neural Networks and a spatial description of the temperature field and the one by Peeters and De Roeck (2000), who presented an autoregressive model and a temporal representation of the temperature state.

If the points in X belong to multiple distributions the most favorable scenario for detection is that where the distributions are well separated. Indeed, in this instance the data points form well-defined clusters and the densities

for each cluster can be identified separately. In many cases, however, the environment produces continuous changes in the identification results and well-defined clusters do not appear. In these instances the topology of the data cloud in input space limits the performance that can be realized in damage detection. If the intrinsic dimensionality of the fluctuations in x due to the environment equals m (where we recall m is the dimension of the points in input space) then good damage detection resolution is difficult. On the contrary, if the data cloud has 'narrow dimensions', as illustrated in Fig. 3.2, good performance can be attained if the damage has important projections in these directions.

3.5.1   Selection of the decision boundary Γ_s

Let Γ_s be the boundary of the S domain. A simple approach to select Γ_s is to take the Euclidian distance of the points to the center of X as a measure of novelty and decide on a threshold (assuming the histogram of the data points has a single mode). Another (also assuming a single mode) is to use the covariance of the data to normalize the distance, in which case one obtains the widely used Mahalanobis metric (Mahalanobis 1936). Another approach, often claimed as a way to 'eliminate environmental effects', is to project the data into a subspace where the variance in the reference state is small, and measure novelty there. Projection procedures include principal component
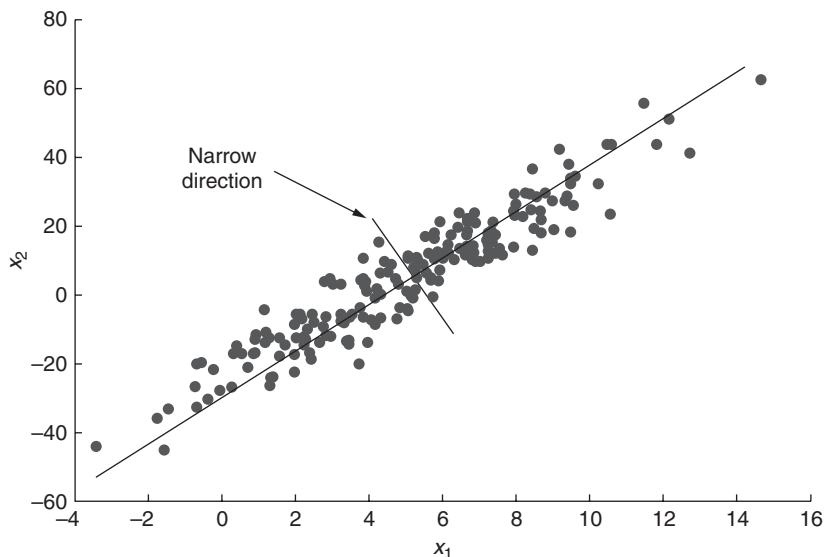
3.2 Bivariate data with narrow dimension.

analysis, factor analysis, and missing data analysis (Kullaa 2003, 2005; Yan *et al.*, 2005a), for 'linear data', and nonlinear principal component analysis (PCA) or kernel PCA (among others) for data whose principal directions are not linear in input space (Hoffman, 2007; Hsu and Loh, 2010; Yan *et al.*, 2005b). Commentary on the results of applying these techniques is presented next.

Consider the bivariate data plotted in Fig. 3.3a. Accepting a hard boundary the line labeled 'a' is a reasonable Γ_S. In PCA the data are projected on to the 2–2 axis, and Γ_S is changed from the *a*-boundary to the two straight lines shown. Assume p_1 is observed. If one were confident that the data of the reference condition cover all the environmental fluctuations, it would be reasonable to consider p_1 as a novel point. Processing the point using PCA would thus lead to a false negative. If, on the contrary, there is significant uncertainty on the adequacy of the boundary in the 1–1 direction, it appears preferable to treat p_1 as a previously unseen healthy point, especially if the prior probability that the structure is undamaged is large. The usefulness of PCA, therefore, is strongly related to the uncertainty that there is on the distribution of the data projected on the principal directions. As Fig. 3.3b shows, PCA increases damage detection sensitivity in some directions and decreases it in others, the directions depending on the location of each point in input space. The steps in the implementation of PCA, when the distance metric is the Mahalanobis distance, are presented next.

Principal component analysis

Let $\mu \in \mathbb{R}^{m \times 1}$ be the mean of the columns in X, define

$$E = \frac{1}{L} \sum_{k=1}^{L} x_k x_k^T \qquad [3.24]$$

Perform an SVD of E and partition into the subspaces ϑ_1 and ϑ_2, namely

$$E = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \qquad [3.25]$$

where S_2 is selected such that the largest entry in S_2 is small compared to the smallest entry in S_1. The model of the reference is $\{\mu, S_2 \text{ and } U_2\}$. Let the data matrix in the validation stage be X_v. Each vector of the matrix X_v is centered using the mean obtained in the model to obtain

$$\overline{x}_{v,k} = x_{v,k} - \mu \qquad [3.26]$$
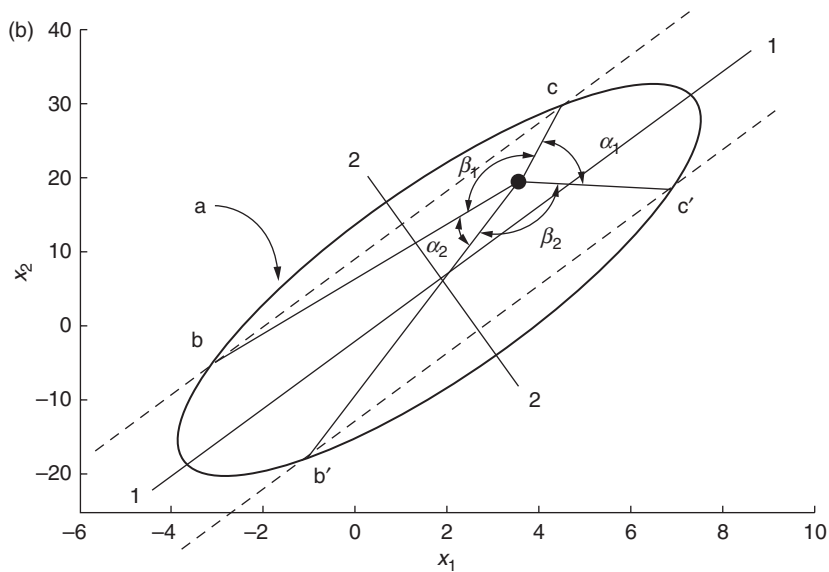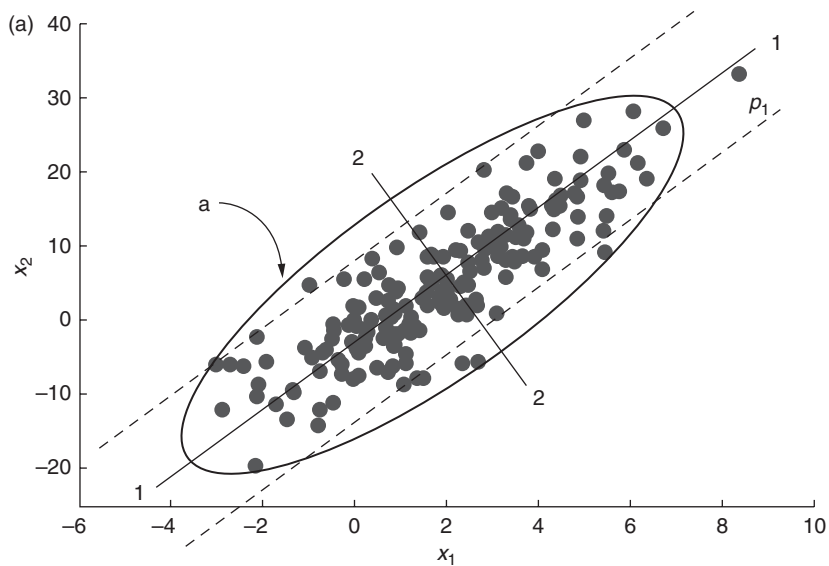
3.3 (a) PCA on bivariate data and (b) change of damage detection sensitivity by PCA.

The projection in ϑ_2 (sometimes referred to as the reconstruction error) is

$$pr_k = U_2^T \bar{x}_{v,k} \qquad [3.27]$$

and the squared Mahalanobis distance is

$$M(pr_k) = pr_k^T S_2^{-1} pr_k \qquad [3.28]$$

The results from Equation [3.28] are computed for all the L_v data. Sorting these values from high to low, the threshold for a 5% Type I error is then taken as the pth term, where p is the closest integer to $0.05 \times L_v$. The vectors in the test phase are treated as explained in the validation phase except that in this case their Mahalanobis is contrasted with threshold to decide on novelty.

Kernel principal component analysis

If the data points in input space do not have principal directions that are linear in input space, the covariance is a poor indicator of the structure of the data, and the square Mahalanobis distance is poorly correlated with density. One approach that can provide a decision boundary for arbitrary distributions, by performing linear algebra, is a generalization of PCA known as kernel PCA (KPCA) (Schölkopf *et al.*, 1998). There is an important difference between KPCA and PCA that is worth noting from the outset. Namely, in PCA the input space is of dimension m and, if there are pc principal components, the effect is to specify an S that is 'open' in pc directions. In contrast, in KPCA the analysis is carried out in a high dimensional feature space and although pc directions are 'opened' in this space, S is generally closed in the input space.

A key feature in KPCA is that the nonlinear function used to expand the input space to the feature space is not explicitly defined, but only the dot product between the vectors is needed (Hoffmann, 2007). KPCA is based on the 'Kernel Trick', a scheme that can extend any algorithm that can be cast in terms of dot products to a higher dimensional space. The advantage of the higher dimensional space is that linear operations carried out there can correspond to nonlinear operations in the original one. To illustrate how the approach works, assume that the L vectors in the input space are mapped to an L-dimensional feature space by an undetermined function. The data matrix in the feature space is thus $B \in R^{L \times L}$. Since B is square BB^T and B^TB are both symmetric with identical eigenvalues, one can write

$$Q_1 = BB^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \qquad [3.29]$$

$$Q_2 = B^TB = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} A_1^T \\ A_2^T \end{bmatrix} \qquad [3.30]$$

and it is not difficult to show that

$$U_j = BA_j S_j^{-0.5} \text{ with } j = 1 \text{ or } 2 \qquad [3.31]$$

The projection of the vectors in B is

$$pr = U_j^T B = S_j^{-0.5} A_j^T B^T B = S_j^{-0.5} A_j^T Q_2 \qquad [3.32]$$

or

$$pr = S_j^{0.5} A_j^T \qquad [3.33]$$

So a projection in the direction of the eigenvectors of the covariance (the U directions) can be realized without having the covariance explicitly. The square of the distance to the origin in the non-principal subspace follows from Equation [3.33] as

$$\gamma = \text{diag}(A_2 S_2 A_2^T) \qquad [3.34]$$

The kernel most widely used to specify Q_2 is the Gaussian Kernel, given by

$$\tilde{Q}_2(i,j) = b_i^T b_j = \exp\left(-\frac{\|f_i - f_j\|_2^2}{2\sigma^2}\right) \qquad [3.35]$$

where the tilde indicates that the matrix is not Q_2, because Q_2 must be based on features that are centered (zero mean). It is not difficult to show that the centered Q_2 can be obtained from the non-centered one as

$$Q_2 = \tilde{Q}_2 - \Gamma\tilde{Q}_2 - \tilde{Q}_2\Gamma + \Gamma\tilde{Q}_2\Gamma \qquad [3.36]$$

where $\Gamma \in \mathbb{R}^{L \times L}$ is a matrix of ones divided by the number of data points, L. Computation of the novelty measure in the reference state is thus straight-forward, namely,

1. $\tilde{Q}_2$ is formulated from Equation [3.35].
2. The matrix is centered using Equation [3.36].
3. The SVD of $\tilde{Q}_2$ is carried out and a dimension for S_2 is selected.
4. The squares of the norms from Equation [3.34] are ordered from large to small and the threshold is selected from the sequence, so a given Type I error is realized.

The testing state in KPCA is somewhat more complex because the new vectors must be centered using the mean implicit in the reference state and this requires some manipulations. The original derivation of the procedure can be found in Hoffman (2007), where the summation notation is used. Here we present the result using matrix products, namely, taking a test vector in the input space as z one has

$$q_1(z) = \left(d^T T_1 - T_2 + T_3\right) A_1 S_1^{-0.5} \qquad [3.37]$$

$$q_2(z) = 1 - \frac{2}{L} d^T r + \frac{1}{L^2} r^T Q_2 r \qquad [3.38]$$

$$\gamma(z) = q_1(z) - \left\| q_2(z) \right\|_2^2 \qquad [3.39]$$

where

$$T_1 = (I - \Gamma) \qquad [3.40]$$

$$T_2 = \frac{1}{L} r^T Q_2 \qquad [3.41]$$

$$T_3 = r^T Q_2 \Gamma \qquad [3.42]$$

where r is a column vector of 1s with L entries and $d \in \mathbf{R}^{L \times 1}$ is such that

$$d(j) = \exp\left(-\frac{\left\| z - f_j \right\|_2^2}{2\sigma^2}\right) \qquad [3.43]$$

User selected parameters in KPCA are the width of the kernel, σ, and the number of components to retain, pc i.e., $S_1 \in \mathbf{R}^{pc \times pc}$. As it turns out, these parameters are related. Namely, as σ increases the columns of Q_1 become increasing collinear and pc decreases. While automated procedures for optimal selection of σ appear feasible, at present these have not yet been developed and some trial and error by the user is typically required to select a reasonable value.

3.5.2   Techniques that operate with a single identification

Techniques that use results from a single identification judge novelty based on the statistical properties of residuals. An attractive property of

these techniques is that they avoid the difficulties that can arise in the multiple identifications setting, where the need to ensure that the quantities that are compared are in fact the same, e.g. the frequencies of the same modes. Two techniques that operate in this manner are described in the next sections.

Robust statistical subspace-based damage detection

Let the system considered in Equations [3.9]–[3.10] be sampled and given at discrete time instants $t = k\tau$, where the inputs are unknown and are replaced by noise terms, leading to the model

$$\begin{aligned} x_{k+1} &= A_d x_k + v_k \\ y_k &= C_d x_k + w_k \end{aligned}$$
[3.44]

with $A_d = \exp(A_c \tau)$ and $C_d = C_c$. A Hankel matrix of output covariances of this system is defined as

$$H_{p+1,q} = \begin{pmatrix} R_1 & R_2 & \cdots & R_q \\ R_2 & R_3 & \cdots & R_{1+q} \\ \vdots & \vdots & \ddots & \vdots \\ R_{p+1} & \cdots & \cdots & R_{p+q} \end{pmatrix}$$
[3.45]

where p and q are user defined parameters and $R_j = E(y_{k+j} y_k^T)$. Since $R_j = C_d A_d^{j-1} G$ where $G = E(x_{k+1} y_k^T)$ (Van Overschee and De Moor, 1996), the Hankel matrix can be factored into an observability and controllability product as

$$H_{p+1,q} = O_{p+1} C_q \text{ where } O_{p+1} = \begin{pmatrix} C_d \\ C_d A_d \\ \vdots \\ C_d A_d^p \end{pmatrix} \text{ and } C_q = \begin{pmatrix} G & A_d & \cdots & A_d^{q-1} G \end{pmatrix}. \quad [3.46]$$

A necessary condition for the information from all the modes to be contained in the matrix of Equation [3.46] is that the number of rows and the number of columns be no less than the system order. Since the true system is infinitely dimensional, the order selected is always that of a truncated model on which results are based. With the number of output sensors as m and the selected system order as n it follows that $p \geq n/m - 1$ and $q \geq n/m$; therefore, one can take $p = \text{round}(n/m)$ and $q = p + 1$.

Subspace-based damage detection

Damage introduces a change in the modes and thus a change in the sub-space defined by the observability matrix O_{p+1} in Equation [3.46]. Defining the left null space S of O_{p+1} in the reference state, the characteristic property of the system being in the reference state can be written as

$$S^T O_{p+1} = 0 \text{ and thus } S^T H_{p+1,q} = 0. \qquad [3.47]$$

Note that the left null space S, given a selected order n, is the matrix given by columns $(n + 1)$, to the last column, from the left side singular vectors of the SVD of $H_{p+1,q}$ in the reference state. The damage detection technique exploits the orthogonality condition in Equation [3.47] using the residual vector (Basseville *et al.*, 2000)

$$\zeta_N = \sqrt{N} \, \text{vec}\left(S^T \widehat{H}_{p+1,q}\right) \qquad [3.48]$$

where $\widehat{H}_{p+1,q}$ is the Hankel matrix of Equation [3.45] computed from cova-riance estimates $\widehat{R}_j = \dfrac{1}{N}\sum_{k=1}^{N} y_{k+j} y_k^T$ on test data containing N data points, and $\text{vec}(\cdot)$ denotes column stacking vectorization.

The statistical evaluation of the residual vector of Equation [3.48] in order to test if $\zeta_N = 0$ (reference state) or $\zeta_N \neq 0$ (damaged state) requires the knowledge of its covariance, which can be estimated from a collection of data sets obtained while the structure is in the reference state. Since the approach is formulated on the premise that the response is stationary, a collection of data sets can be obtained by dividing a long record into segments, e.g. if the total number of points is N_1 one can divide the data into n_b blocks with length $N = N_1/n_b$ and use each set as an independent test. Namely, if $\zeta_N^{(1)}$ is the residual obtained using Equation [3.48] for the lth data set, one estimates the covariance as

$$\Sigma_\varsigma \frac{1}{n_b} \sum_{l=1}^{n_b} \zeta_N^{(l)} \zeta_N^{(l)T} \qquad [3.49]$$

Finally, to test if a newly collected data set corresponds to the reference state or not, one computes the metric

$$\chi_N^2 = \zeta_N^T \, \Sigma_\varsigma^{-1} \, \zeta_N \qquad [3.50]$$

which is chi-squared distributed with as many DOFs as the dimension of the residual vector. In theory, the threshold of the metric in Equation

[3.50] can be determined from the cumulative χ^2 distribution by selecting an acceptable Type I error, but in practice, due to inevitable deviations between the assumptions and the real situation, the threshold is best established by inspecting results for a sequence of data sets from the reference state. Specifically, one orders the values of the metric from high to low and selects the threshold as the αth quantile of the sequence, where α is the Type I error rate.

Changes in the excitation statistics

A limitation of the residual in Equation [3.48] is that the Hankel matrix at the time of inquiry depends on the intensity of the excitation, and thus the amplitude of the residual depends, not only on changes in the system, but also in fluctuations in the disturbances. A modification that sharply improves the situation, recently introduced by Döhler and Mevel (2011), replaces the Hankel matrix at the time of enquiry by the left side singular vectors that define its span, namely, with

$$\widehat{H}_{p+1,q} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Delta_1 & \\ & \Delta_2 \end{bmatrix} V^T \qquad [3.51]$$

One obtains the residual as

$$\zeta_N = \sqrt{N}\ \mathrm{Vec}\left(S^T U_1\right) \qquad [3.52]$$

where the number of columns in U_1 is the selected order n.

Environmental changes (temperature)

One way to mitigate the effect of temperature in the subspace detection approach is to compute the matrix S from a composite Hankel matrix formed by concatenating results for all the environmental conditions of interest (Balmès *et al.*, 2008). Specifically, with $H^1, H^2, \ldots, H^t$ as Hankel matrices from Equation [3.45] computed for different environmental conditions, one has

$$H = \begin{bmatrix} H^1 & H^2 & H^t \end{bmatrix} \qquad [3.53]$$

The matrix S is then defined as the left null space of this matrix. Needless to say, the premise is that the temperature effects produce changes that are limited to particular directions so that the matrix in Equation [3.53] retains a left null space. Once S and the covariance are established, the approach

proceeds as described previously. As is the case with PCA, this approach reduces the sensitivity of the residual to damage since the dimension of the null space of H decreases.

Kalman filter-based damage detection

The idea of using the Kalman filter as a fault detector was introduced by Mehra and Peschon (1971), soon after the filter was developed (Kalman, 1960). The strategy is based on the fact that when the filter is operating optimally (i.e., when the prevailing conditions satisfy the assumptions), differences between measurements and filter predictions, known as innovations, are realizations of a white noise process. Detection, therefore, is carried out by formulating the filter for some reference state, which is treated as the healthy condition, and subsequently carrying a hypothesis test (Melsa and Cohn, 1978) on the innovation's whiteness. For damage detection purposes, the physical meaning of the state vector is not relevant and this allows, as shall be shown, formulation of the filter from data measured in the reference state (Bodeux and Golinval, 2001; Di Ruscio 1996). This section illustrates how the filter can be used as a damage detector; a discussion on how the filter can be extracted from reference state data can be found in Bodeux and Golinval, (2001).

The innovations form of the state space expressions are

$$x_{k+1}^- = Ax_k^+ + AKe_k \qquad\qquad [3.54]$$

$$y_k = Cx_k^- + e_k \qquad\qquad [3.55]$$

where e is the innovation, K is the Kalman gain, and the superscripts – and + indicate before and after information from the measurement is accounted for in the estimation. Note that in the literature the matrix AK is also referred to as the Kalman gain. The algorithm to compute the innovations sequence can be summarized as follows:

- Let x_k^- be known or assumed
- The innovation at k is computed using Equation [3.55]
- The state at k is updated using

$$x_k^+ = x_k^- + Ke_k \qquad\qquad [3.56]$$

- and it is advanced as

$$x_{k+1}^- = Ax_k^+ \qquad\qquad [3.57]$$

The next step is to determine if it is reasonable to assume that the innovation sequence obtained is a realization from a white noise process.

Whiteness test

Consider any one of the innovation signals (there is one for each output channel). Under the null hypothesis, the auto-correlation for the *jth* lag, $q_j = E(e_i e_{i+j})$ has an empirical estimate given by

$$q_j = \frac{1}{L} \sum_{i=1}^{L} e_i e_{i+j} \qquad [3.58]$$

where L is the number of points selected for the summation. Under the null hypothesis, provided the innovations are normalized to unit variance, q_j has zero mean and a variance equal to $1/L$. A χ^2 test for innovations whiteness can thus be carried out as follows:

- Normalize $e(k)$ to unit variance
- Compute the correlation function q_j for lags n_1 to n_2

  Compute the χ^2 statistic, Γ, as

$$\gamma = L \sum_{j=n_1}^{n_2} (q_j)^2 \qquad [3.59]$$

where n_1 is the first lag considered and n_2 the last. The null hypothesis, i.e. that the innovation sequence is a realization of a white noise process, is not rejected if Γ is no larger than the *(1−α)*th quantile of the χ^2 with $\bar{n} = n_2 - n_1 + 1$ DOF.

Aggregation

Since the whiteness test (as shown) is carried out one sensor at a time, the question of aggregation in the multi-sensor case arises. A simple option is to decide on damage or no-damage using an OR strategy. Namely, the null is rejected if Γ exceeds the associated threshold at any sensor. For an OR strategy the rate at which the null is rejected at each sensor $\bar{a}_j$ differs from the value of the global Type I error. An expression for $\bar{a}_j$ as a function of α can be derived by noting that the distributions of Γ at each sensor are independent under the null hypothesis; on this premise one has

$$\alpha = 1 - \prod_{j=1}^{m} (1 - \bar{\alpha}_j) \qquad [3.60]$$

So, taking all the values of $\bar{\alpha}_j$ as equal one gets

$$\bar{\alpha} = 1 - (1 - \alpha)^{1/m} \qquad\qquad [3.61]$$

which, for small values of α, is closely approximated by α/m.

Selection of n_1 and n_2

Analytical optimization of n_1 and $\bar{n} = n_2 - n_1 + 1$ is not practical because the result depends on the specific damage. An alternative that has been shown to provide reasonable results is to use $\bar{n} = \text{round}(f_s/f_1)$ where f_s = sampling frequency and f_1 = fundamental frequency (from the identification) and to select n_1 from inspection of a plot of the threshold vs n_1. Namely, the metric Γ for each channel is computed for a number of n_1 values using the training data (for fixed $\bar{n}$) and the result is organized from largest to smallest. The threshold for each channel, ϑ, at each n_1 value is the $\bar{\alpha} \cdot P$ entry, where P is the number of data points used for training. A plot of ϑ vs n_1 can then be made and one selects n_1 as the lowest value for which the rate of change of the threshold is small. This approach, which entails judgment from the user, is analogous to the selection of a regularization parameter in the solution of a poorly conditioned system of equations. The whiteness test with the n_1 and $\bar{n}$ selection described has been designated as the lag shifted whiteness test (LSWT).

3.5.3   Change detection

The advantage that the system state does not switch back-and-forth from healthy to damage can be realized by framing the problem as one of change detection. In this case, the question posed is not whether a particular data point is an outlier, but rather whether the aggregate of a sequence of points (typically assumed independent under the null hypothesis) is such that a change should be announced. A convenient approach to processing the data for change detection is by means of control charts devised in the process control field (Montgomery, 2009). A control chart is a depiction in the y-axis of some metric whose probability depends on the state of the process vs the sample number. Control charts were introduced by Shewhart, (1931) and since then many types have been proposed. A recent literature review can be found in Schmid (2007). The CUSUM chart is designed to perform a hypothesis test regarding the shift in any parameter of a distribution that is assumed to prevail when the null hypothesis is true. In this section we present general expressions for the exponential distribution family and particular results for the Gaussian and the Gamma density functions.

With y as the metric and θ as the parameter that is allowed to vary in the distribution (parameters vary one at a time) one has, for the exponential family

$$f(y|\theta) = e^{\{a(y)b(\theta)+c(y)+d(\theta)\}}$$ [3.62]

Assuming independence, the probability of the sequence of values $Y = y_i$ and $i = 1,\ldots, n$, follows from Equation [3.62] as

$$f(Y|\theta) = e^{\left\{b(\theta)\sum\limits_{j=1}^{n}a(y_j)+\sum\limits_{j=1}^{n}c(y)+nd(\theta)\right\}}$$ [3.63]

Consider two postulated parameter θ_0 and θ_1; we want to determine which of the two is more likely, given the data. The answer is given by Equation [3.63] evaluated for $\theta = \theta_1$ divided by the same expression evaluated for $\theta= \theta_0$. If the ratio is >1 then θ_1 is more likely, and the opposite is true if the value is <1. The ratio is

$$\Upsilon = \frac{f(Y|\theta_1)}{f(Y|\theta_0)} = e^{\left\{(b(\theta_1)-b(\theta_0)\sum\limits_{j=1}^{n}a(y_j)+n(d(\theta_1)-d(\theta_0))\right\}}$$ [3.64]

or, designating $\Im = \ln(\Upsilon)$

$$\Im_n = (b(\theta_1)-b(\theta_0))\sum\limits_{j=1}^{n}a(y_j)+n(d(\theta_1)-d(\theta_0))$$ [3.65]

From inspection of Equation [3.65] it follows that

$$\Im_n = \Im_{n-1} +(b(\theta_1)-b(\theta_0))a(y_n)+(d(\theta_1)-d(\theta_0))$$ [3.66]

or

$$\Im_n = \Im_{n-1} +(b(\theta_1)-b(\theta_0))\{a(y_n)+k\}$$ [3.67]

where one has

$$k = \frac{d(\theta_1)-d(\theta_0)}{b(\theta_1)-b(\theta_0)}$$ [3.68]

Negative values of $\Im_n$ indicate that $\theta = \theta_0$ is most likely. A single positive value of $\Im_n$ is not strong evidence that a switch to the alternative hypothesis is appropriate because $\Im_n$ is a random variable, but a trend in the upward direction indicates that the alternative hypothesis is the more likely one. While one can focus on the slope of $\Im_n$, it is more convenient to inspect the attainment of a critical upper movement once the slope turns positive. This is easily done by reinitializing $\Im_n$ whenever it goes negative and setting some limiting positive value, h^+, as a level which, when crossed, damage is announced. Needless to say, there is always a finite probability that any level will be crossed, even when the structure is not damaged, and this is known as ARL_0 (average run length under the null hypothesis). Once damage takes place there is also an average number of data points needed before the threshold is crossed, and this depends on the amount of damage and on the specific value of h^+ selected, and is known as ARL_1. Optimal values of θ_1 and h^+ depend, in principle, on the desired ARL_0 and on the magnitude of $\Delta\theta = \theta_1 - \theta_0$ for which one wishes to attain the minimum ARL_1. In practice, however, at least in damage detection of civil structures, our experience is that refinement in the selection of these levels is unwarranted. Specifically, it generally suffices to take $\theta_1 = 1.5\theta_0$ (or 1/1.5 if damage decreases the parameter) and select h^+ as the hard threshold using data for the training phase. Although the shift suggested is large, numerical results indicate that the results are insensitive to the specific choice in a wide range. Expressions for $a(y)$, $b(\theta)$ and $d(\theta)$ for specific parameters and distributions are presented in Table 3.1.

Step by step procedure

- Compute L values of the metric in the reference state
- Fit a distribution to the metric (if neither Gaussian nor gamma work, then results for some others can be found in Hawkins and Olwell (1997).

Table 3.1 Parameters for implementation of the CUSUM

Distribution	θ	$a(y)$	$b(\theta)$	$d(\theta)$
Gaussian $N(\mu, \sigma)$	μ	Y	$\dfrac{\mu}{\sigma^2}$	$-\dfrac{\mu^2}{2\sigma^2}$
	σ	$(y-\mu)^2$	$-\dfrac{1}{2\sigma^2}$	$-\ln(\sigma)$
Gamma $\Gamma(\alpha, \beta)$	α	$\ln(y)$	$\alpha\text{-}1$	$\ln(\Gamma(\alpha)) - \alpha\ln(\beta)$
	β	y	$-\dfrac{1}{\beta}$	$-\alpha\ln(\beta)$

- Determine if θ increases or decreases due to damage, and select θ_1 accordingly
- Starting with $\Im_0 = 0$, compute k and evaluate $\Im_n$ using Equation [3.67] for the healthy state. Recall that if $\Im$ goes negative at any step it is replaced by zero before executing the next step.
- From inspection of the results select h^+
- Process the data of the test phase exactly as before and announce damage when the threshold h^+ is crossed.

In closing the damage detection discussion, we note that:

1. The difficulties in detecting damage under environmental changes increase as the structure gets larger, because the environment affects the complete system, while damage is typically local in nature.
2. Damage is easier to detect in lightly damped structures than in structures with heavier damping, because the variability of the discriminating metric resulting from environmental changes is not significantly affected by damping, while sensitivity to damage decreases with damping.

An example illustrating novelty detection using KPCA in conjunction with the CUMSUM chart is presented next.

### 3.5.4 Example: detection of novel patterns using KPCA in conjunction with the CUSUM chart

This example illustrates the detection of novel patterns by combining KPCA with the CUSUM chart. The data considered, as shown in Fig. 3.4, are strongly nonlinear in the input space. We note that the points marked as stars are not included in the formulation of the KPCA model, but are used subsequently to test novelty. The reconstruction error over the input space is depicted in Plate I in the color section between pages 374 and 375 for $\sigma = 1.5$ and $pc = 180$. As can be seen, the approach captures the structure of the data accurately.

CUSUM

An inspection of the histogram of the reconstruction error of the model, for the training data, shows that the distribution is approximately gamma, and a distribution fit leads to $\alpha = 2.14$ and $\beta = 0.02$. The results of the CUSUM (implemented as described) for the data of the model, with the test data appended at the end, are depicted in Figs [3.5a] and [3.5b], where thresholds based on hard boundaries are illustrated. As can be seen, the novelty of the outliers is clearly detected.
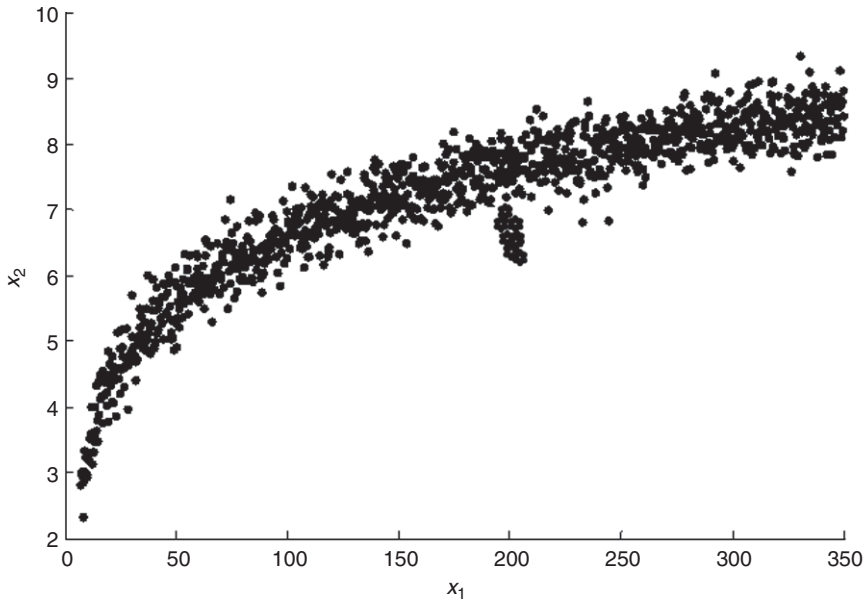
3.4 Input reference and test data.
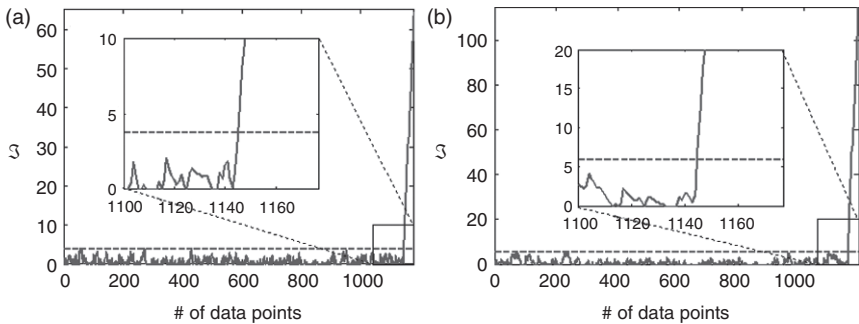
3.5 CUSUM metric while (a) $\theta = \alpha$ (b) $\theta = \beta$.

3.6   Damage localization

Once damage is identified it is sometimes of interest to obtain information on its location. In contrast to detection, localization (with sparse sensors) requires use of a model.

3.6.1   Stochastic damage locating vector approach (SDLV)

We present a damage localization approach that applies when the identification is carried out from output signals. The approach operates using a

static model of the reference condition (i.e., no mass or damping properties are needed) and is described in detail in Bernal (2006). Related procedures that apply when the input can be measured, as well as variants that use the system dynamics, are described in Bernal (2002, 2007, 2009, 2010).

Let $\{A_c$ and $C_c\}$ be the system matrix and the state to output matrix in continuous time, and let there be two sets of these matrices obtained from identification, one for the reference condition and one for the damaged state. For each state one computes

$$R(s) = C_c A_c^{-1} H^{-*} L \qquad [3.69]$$

where

$$H = \begin{bmatrix} C_c A_c^{-1} \\ C_c \end{bmatrix} \quad \text{and} \quad L = \begin{bmatrix} I \\ 0 \end{bmatrix} \qquad [3.70a,b]$$

The difference between the damaged state and the reference state matrices, R, is

$$\Delta R = R_D - R_R \qquad [3.71]$$

Performing an SVD of the transpose of ΔR gives

$$\Delta R^T = U \times \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \times [V_1 \quad V_2]^T \qquad [3.72]$$

where the singular values s_2 are assumed to be small. Leaving aside temporarily the issue of how to decide when a singular value is sufficiently small, let the s_2 set have p entries so $V_2 \in R^{m \times p}$ where m = number of sensors. The SDLV approach localizes the damage by inspecting the stress fields associated with the vectors in V_2. Specifically, let $V_2 = \begin{bmatrix} v_1 & v_2 & L & v_p \end{bmatrix}$ and the stress field on a model of the reference state due to v_j be σ_j. Taking the normalized stress index for the jth vector as nsi_j, namely

$$nsi_j = \frac{|\sigma_j|}{\max |\sigma_j|} \qquad [3.73]$$

one computes an aggregate index as

$$\text{WSI} = \frac{\sum_{j=p}^{p-z+1} nsi_j}{\max(\sum_{j=p}^{p-z+1} nsi_j)} \qquad [3.74]$$

with the set of potentially damaged elements announced as those for which the weighted stress index, WSI, is less than some threshold. In Bernal (2006) it is recommended that the number of terms (z) in Equation [3.74] be computed as follows: (1) normalize the singular values to a maximum of unity; (2) take the square root; (3) count how many are less than 0.1 (say there are q), and; (4) take z equal to $q/2$, rounding downward.

We conclude by noting that the SDLV approach does not state that the damage is in elements where the stress field is small but rather that the elements that have significant stresses are not damaged. The theoretical resolution of the SDLV approach depends on the location of the sensors and is clarified in the original reference. Improvements in resolution can be realized by calling on the dynamic system properties using the SDDLV approach (Bernal, 2010).

3.7   Future trends

The literature in vibration-based structural health monitoring is vast and continues to expand at a fast rate. Progress in the detection side is likely to come from improvements in the formulation of refined models for environmental effects. In localization, methods that avoid a model altogether and make up for it with a large sensor density may prove feasible and, if so, could lead to improved robustness and practicality. Although not treated here due to space constraints, damage quantification is a relatively simple step once the correct locations are identified.

3.8   Sources of further information and advice

There are various excellent sources of information on analytical techniques for damage characterization in the literature. Chapter 2 of the book Structural Health Monitoring, edited by Balageas, Fritzen and Guemes (2006), for example, is a good reference for vibration-based techniques, and the texts by Patton and Frank (2000) on fault diagnosis provide a summary of the state of the art from the systems theory perspective.

3.9   References

Balageas D, Fritzen C P and Guemes A (2006), *Structural Health Monitoring*, London-Newport Beach: ISTE.

Balmès É, Basseville M, Bourquin F, Mevel L, Nasser H and Treyssède F (2008), 'Merging sensor data from multiple temperature scenarios for vibration monitoring of civil structures', *Structural Health Monitoring*, **7**(2), 129–142.

Basseville M, Abdelghani M and Benveniste A (2000), 'Subspace-based fault detection algorithms for vibration monitoring', *Automatica*, **36**(1), 101–109.

Bernal D (2002), 'Load vectors for damage localization', *Engineering Mechanics, ASCE,* **128**(1), 7–14.

Bernal D (2006), 'Flexibility-based damage localization from stochastic realization results', *Engineering Mechanics, ASCE,* **132**(6), 651–658.

Bernal D (2007), 'Damage localization from the null space of changes in the transfer matrix', *American Institute of Aeronautics and Astronautics (AIAA),* **45**(2), 374–381.

Bernal D (2009), 'Damage localization from transmission zeros of delta systems', *Engineering Mechanics*, *ASCE,* **135**(2), 93–99.

Bernal D (2010), 'Load vectors for damage location in systems identified from operational loads', *Engineering Mechanics, ASCE,* **136**(1), 31–39.

Bodeux J B and Golinval J C (2001), 'Application of ARMAV models to the identification and damage detection of mechanical and civil engineering structures', *Smart Materials and Structures*, **10**(3), 479.

Di Ruscio D (1996), 'Combined deterministic and stochastic system identification and realization: DSR – a subspace approach based on observations', *Journal of Modeling, Identification, and Control*, **17**(3), 193–230.

Döhler M and Mevel L (2011), 'Robust subspace based fault detection', In *Proceedings of 18th IFAC World Congress*, Milan.

Hawkins D M and Olwell D H (1997), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.

Hoffmann H (2007), 'Kernel PCA for novelty detection', *Pattern Recognition*, **40**(3), 863–874.

Hsu T Y and Loh C H (2010), 'Damage detection accommodating nonlinear environmental effects by nonlinear principal component analysis', *Structural Control and Health Monitoring*, **17**(3), 338–354.

Kalman R E (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering*, **82**(1), 35–45.

Kullaa J (2003), 'Is temperature measurement essential in structural health monitoring', *4th International Workshop on Structural Health Montoring*, Stanford, CA, 717–724.

Kullaa J (2005), 'Damage detection under a varying environment using the missing data concept', *Structural Health Monitoring*, Stanford, CA, 565–573.

Mahalanobis P C (1936), 'On the generalised distance in statistics', In *Proceedings of the National Institute of Sciences of India*, **2**(1), 49–55.

Mehra R K and Peschon J (1971), 'An innovations approach to fault detection and diagnosis in dynamic systems', *Automatica*, **7**(5), 637–640.

Montgomery D C (2009), *Introduction to Statistical Quality Control*, John Wiley and Sons.

Patton R J and Frank P M (2000), *Issues of Fault Diagnosis for Dynamic Systems.* s.l.:Springer, Berlin, Heidelberg, New York.

Peeters B and De Roeck G (2000), 'One year monitoring of the Z24 bridge: environmental influences versus damage effects', In *Proceedings of IMAC-XVIII,* San Antonio, TX, 1570–1576.

Schmid W (2007), 'Eighty years of control charts', *Sequential Analysis: Design Methods and Applications,* **26**(2), 117–122.

Schölkopf B, Smola A J and Müller K R (1998), 'Nonlinear component analysis as a kernel eigenvalue problem', *Neural Computation*, **10**, 1299–1319.

Sohn H, Worden K and Farrar C R (2001), 'Novelty detection under changing environmental conditions', In *Proceedings of the 8th Annual SPIE International Symposium on Smart Structures and Materials*, San Diego, CA, 108–118.

Van Overschee P and De Moor B (1996), S*ubspace Identification for Linear Systems: Theory, Implementation, Applications*, Boston: Kluwer Academic.

Yan A M, Kerschen G, De Boe P and Golinval J C (2005a), 'Structural damage diagnosis under varying environmental conditions-Part I: a linear analysis', *Mechanical Systems and Signal Processing*, **19**, 847–864.

Yan A M, Kerschen G, De Boe P and Golinval J C (2005b), 'Structural damage diagnosis under varying environmental conditions – Part II: local PCA for nonlinear cases', *Mechanical Systems and Signal Processing*, **19**, 865–880.

4

Output only modal identification and structural damage detection using time-frequency and wavelet techniques for assessing and monitoring civil infrastructures

S. NAGARAJAIAH, Rice University, USA, B. BASU, Trinity College Dublin, Ireland and Y. YANG, Rice University, USA

Abstract: The primary objective of this chapter is to develop output only modal identification and structural damage detection. Identification of multidegree of freedom (MDOF) linear time invariant (LTI) and linear time variant (LTV – due to damage) systems based on time-frequency (TF) techniques – such as short-time Fourier transform (STFT), empirical mode decomposition (EMD), and wavelets – and also a newly-merging blind source separation (BSS) technique is discussed. STFT, EMD, and wavelet methods developed to date are reviewed in sufficient detail. In addition, a Hilbert transform (HT) approach to determine frequency and damping is also presented. In this chapter STFT, EMD, HT, and wavelet techniques are developed for decomposition of free-vibration response of MDOF systems into their modal components. Once the modal components are obtained, each one is processed using HT to obtain the modal frequency and damping ratios. In addition, the ratio of modal components at different degrees of freedom facilitate determination of mode shape. In cases with output only modal identification using ambient/random response the random decrement technique is used to obtain free-vibration response.

The advantage of TF techniques is that they are signal-based, hence, they can be used for output only modal identification. A three degree of freedom 1:10 scale model test structure is used to validate the proposed output only modal identification techniques based on STFT, EMD, HT, and wavelets. Both measured free-vibration and forced-vibration (white noise) response are considered. The second objective of this chapter is to show the relative ease with which the TF techniques can be used for modal identification and their potential for real-world applications wherein output only identification is essential. Recorded ambient vibration data processed using techniques such as random decrement technique can be used to obtain the free-vibration response, so that further processing using TF-based modal identification can be performed.

Key words: output only modal identification, time-frequency techniques, short-time Fourier transform (STFT), empirical mode decomposition (EMD), Hilbert transform (HT), wavelets, blind source separation (BSS), independent component analysis (ICA).

4.1   Introduction

Modal identification of structural systems is a key step in the process of structural identification, structural health monitoring, and damage detection. It essentially requires an inverse problem to be solved from measured or recorded response of the structures under ambient or dynamic loading such as earthquakes, wind and waves. The aim is to estimate properties of the structures such as natural frequencies, mode shapes, energy dissipation characteristics, and strength and stiffness deterioration due to damage. Such estimates are required to be performed only from output signals. This is especially the case for large civil structures where input excitation cannot be directly measured in most practical situations.

System identification of structures has classically been performed in two different paradigms: (i) time-domain analysis; and (ii) frequency domain analysis. Several approaches to time-domain system identification have been developed, such as state estimation using Kalman filter, stochastic analysis and modeling, recursive modeling, and least squares method. System identification and fault detection techniques are also currently being developed. The work of Nagarajaiah and co-workers has led to the development new interaction matrix formulation and input error formulation (Koh *et al*., 2005a, b, 2008), based on the concept of analytical redundancy, to detect and isolate the damage/faults in structural members, sensors, and actuators in a structural system (Li *et al*., 2007; Chen *et al*., 2007, 2008a, b). The new techniques can detect the presence of fault/damage in a structure (Level 1), locate the member/sensor/actuator in which fault/damage is existent (Level 2), and determine the time instants of occurrence (Level 3). The resulting error function would indicate real-time failure/damage of a member, sensor, or actuator. The interaction matrix technique allows the development of input–output equations that are influenced only by one target input. These input–output equations serve as an effective tool to monitor the integrity of each member, sensor, or actuator, regardless of the status of the others. The procedure requires knowledge of the analytical model of the healthy system being tested, and the analytical redundancy can be experimentally predetermined through input–output-based system identification. Additionally, the authors have developed an ARMarkov observer bank algorithm to detect the extent of damage – Level 4 (Dharap *et al*., 2006). They have also shown experimentally that the proposed algorithms successfully identify failures of actuators or sensors that are attached to the truss structure in tests on the NASA 8-bay 4 m long truss (Koh *et al*., 2005a; Li *et al*., 2007). Considering the limited number of measurements and the complexity of the structure, test results ensure the capability of the proposed procedure in detecting and isolating the simultaneously and arbitrarily occurred multiple failures. In addition, new time-segmented

system identification techniques have been proposed (Nagarajaiah *et al.*, 2003, 2004). More recently, Yang and Nagarajaiah established a series of output only algorithms based on a new unsupervised learning technique blind source separation (BSS) or independent component analysis (ICA). Specifically, the developed time-frequency (TF) STFT-ICA can perform effective modal identification of highly-damped structures (Yang and Nagarajaiah, 2013a); also, ICA has been discovered to bias toward sparse components that typically signify damage, thus capable of blindly identifying even multiple damage instants (Level 3) and damage locations (Level 2), simultaneously (Yang and Nagarajaiah, 2012). In addition, the proposed sparse component analysis (SCA) modal identification method (Yang and Nagarajaiah, 2013b) in the framework of underdetermined BSS is able to identify all the present modes using dramatically few sensors, in which situation traditional modal identification methods may not work. These algorithms are validated using numerical and experimental models, as well as real-world seismic-excited structures. The advantages of these BSS methods lie in their elegant formulations, straightforward implementations, and efficient computations.

Signal-based identification techniques, i.e. those based on analysis of response signals of structures, have also been developed. A large volume of work has been carried out based on time-series analysis in the context of output only modal identification and damage detection. Spiridonakos *et al.* (2010) carried out time-domain identification of time varying (TV) mechanical structures using time-dependent auto-regressive moving average (ARMA) models. Laboratory tests on TV systems simulated by a beam with a moving mass were performed and the results compared with 'frozen-configuration' baseline model. It was concluded that functional series-based models performed best. Worden *et al.* (2000) considered damage detection using statistical process control charts using mean and variance of the residual of the auto-regressive (AR) model of the signals. Sohn and Farrar (2001) presented damage detection using statistical pattern recognition methodology on AR time series of dynamics signals recorded. Change in AR coefficients from recorded signals embedded into a wireless sensor has shown the prospect of implementation of damage detection algorithms in the context of a wireless sensor network (Lynch *et al.*, 2004).

A significant amount of research is currently being focused on output only modal identification and damage detection techniques for various types of applications. Deraemaeker *et al.* (2008) studied the output only damage detection under environmental changes using an automated subspace identification methodology. Damping ratios of bridges were identified by Gonzalez *et al.* (2012) using acceleration records from vehicle axles using simulated signals from a quarter car model. The vehicle–bridge interaction (VBI) was simulated, and the effect of noise and road surface roughness was

investigated. Fan and Qiao (2011) carried out a review of modal parameter-based damage identification methods in beams and plates. Milanese *et al.* (2008) detected joint loosening using output only strain data from broadband vibration excitation. Zheng and Mita (2012) proposed distance measures of AR models to identify damage from the output signals and applied the technique to a five-story building model. Gul and Catbas (2008) used a complex mode indicator function with random decrement on output only data from ambient vibration to identify modal parameters. Laboratory test on a steel grid and field test on a long span bridge were conducted. The methodology was used to identify damage on the steel grid.

Parallel to the advances in sensing and data acquisition techniques, there has been a tremendous development of signal processing techniques, which allows extraction of information from the available data sensed in the form of either signals or images. Several identification techniques have been proposed for structural dynamic systems in the recent past (Worden and Tomlinson, 2001). Most of the techniques and algorithms proposed are based on the use of different integral transforms. Among the available techniques, those based on the use of Hilbert transform (HT) by Tomlinson (1987) and Feldman (1994a, b) have become popular. The classical method of frequency domain analysis is by means of Fourier transform (FT) and its algorithmic implementation, the discrete Fourier transformation (DFT). Though DFT has been widely used for modal analysis and other system identification tasks, it has several limitations. Fourier analysis is inherently global in nature and provides average information over time, ignoring the TV nature of a non-stationary signal.

TF methods (Cohen 1995; Huang *et al.*, 1998), such as short-time Fourier transform (STFT) and wavelets, are used extensively for signal processing. New techniques such as empirical mode decomposition (EMD) (Huang *et al.*, 1998) have been developed for signal processing of non-stationary signals. STFT and EMD techniques, with HT, have played a key role in the development of new TF-based controllers for semiactive, smart tuned mass dampers (Nagarajaiah *et al.*, 1999; Nagarajaiah and Varadarajan, 2001; Nagarajaiah and Varadarajan, 2005; Nagarajaiah and Sonmez, 2007; Nagarajaiah, 2009; Narasimhan and Nagarajaiah, 2005; Varadarajan and Nagarajaiah, 2004). Modal identification using EMD and HT has been developed (Nagarajaiah and Varadarajan, 2001; Nagarajaiah, 2009; Yang *et al.*, 2003, 2004). Wavelets have played a key role in the development of new linear quadratic TV controllers (Basu and Nagarajaiah, 2008) and modal identification of TV systems (Basu *et al.*, 2008) by the authors. Wavelets, with HT, have also been used to estimate frequency and damping (Staszewski, 1997), modal and damage identification (Staszewski *et al.*, 1998, 2007; Basu, 2007; Chakraborty *et al.*, 2006; Basu, 2008; Pakrashi *et al.*, 2007; Goggins *et al.*, 2007).

Recently, several TF analysis tools, particularly wavelet analysis technique, have proved to be powerful for system assessment, structural health and fault monitoring (Staszewski and Tomlinson, 1994; Wang and McFadden, 1996; Al-Khalidy *et al.*, 1997; Ghanem and Romeo, 2000; Addison *et al.*, 2002), system identification (Staszewski, 1997, 1998; Ruzzene *et al.*, 1997; Gurley and Kareem, 1999; Kitada, 1998; Kypriyanou and Staszewski, 1999; Robertson *et al.*, 1998; Lardies and Gouttebroze, 2000; Piombo *et al.*, 2000; Ghanem and Romeo, 2001; Kijewski and Kareem, 2002, 2006, 2007; Jiang and Mahadevan, 2008; Chen *et al.*, 2009; Le and Paultre, 2012), and damage detection (Naldi and Venini, 1997; Staszewski *et al.*, 1998; Liew and Wang, 1998; Okafor and Dutta, 2000; Wang and Deng, 1999; Hou *et al.*, 2000; Chatterjee and Basu, 2001, 2006; Patsias and Staszewski, 2002; Patsias *et al.*, 2002; Melhem and Kim, 2003; Chang and Chen, 2003; Gentile and Messina, 2003; Loutridis *et al.*, 2004; Rucka and Wilde, 2006; Spanos *et al.*, 2006; Gangadharan *et al.*, 2009) by analyzing vibration signals. Some of the early researchers on analysis of vibration signals using wavelets include Newland (1993, 1994a, b), Zeldin and Spanos (1998) and Basu and Gupta (1997, 1998, 1999a, b). The above mentioned references are representative of a vast amount of literature available as result of the research in the past decade and a half. An overview of wavelet analysis with several different applications has been provided by Robertson and Basu (2008) and Staszewski and Robertson (2007).

Lately, the unsupervised BSS learning technique (Comon and Jutten, 2010) has attracted considerable attention as a new signal processing tool in structural dynamics and structural identification (Antoni, 2005). BSS is able to recover the hidden source signals and the unknown mixing process using only the measured mixture signals; it is thus particularly suitable for output only identification. Several BSS learning algorithms – such as ICA (Hyvärinen *et al.*, 2000), second order blind identification (SOBI) (Belouchrani *et al.*, 1997), and SCA (Gribonval and Lesage, 2006) – have been explored to perform output only structural identification, enjoying intuitive physical interpretation, straightforward implementation, and computational efficiency. As the most popular BSS technique, ICA uses a higher order statistics of signals to estimate the BSS model, while SOBI is based on the second order statistics of signals and uses the joint approximate diagonalization (JAD) algorithm to perform BSS estimation. Kerschen *et al.* (2007) proposed that the vibration modes can be viewed as the virtual ICA modes and used the time-domain ICA method to identify the modal parameters of undamped or lightly damped structures. SOBI was shown to have potential in identification of higher-damped structures (Poncelet *et al.*, 2007; Zhou and Chelidze, 2007), complex modes (McNeill and Zimmerman, 2008), and structures subjected to non-stationary excitation (Hazra *et al.*, 2010a, b). To tackle the damping effect in the ICA method, Yang and Nagarajaiah (2013a) proposed a TF BSS algorithm STFT-ICA, and showed that it is

effective for identifying even highly-damped structures in free or random vibration (stationary or non-stationary) subject to heavy measurement noise. Furthermore, Yang and Nagarajaiah (2012) discovered that the ICA learning rule naturally extracts sparse components, which typically indicate damage features; combining the wavelet transform, they proposed a new output only method WT-ICA that is able to simultaneously identify single or multiple damage instants and damage locations and is validated by detailed simulations and practical examples. An outstanding advantage of the ICA method is that it has a robust FastICA algorithm that enjoys cubic convergence rate, and is thus superior for identification of large-scale structures and promising in wireless platform.

In addition, the modal identification issue with fewer sensors than modes is also addressed in the framework of underdetermined BSS. Hazra *et al.* (2012) combined the EMD and SOBI to solve the underdetermined modal identification problem. They assumed the *a priori* knowledge of modeshapes to proceed with the identification. Sadhu *et al.* (2011) conducted the modal identification in the wavelet domain using SOBI, and combined several identified partial modeshapes to yield a complete modeshape matrix. Abazarsa *et al.* (2012) used a new underdetermined BSS technique called PARAFAC to handle this problem. In particular, they provided the identifiability criterion of the minimum sensor number of their method. Yang and Nagarajaiah (2013b) used the SCA method to perform output only identification. They showed that the SCA method can accurately identify the modal parameters in either the determined or underdetermined case where sensors may be dramatically few. In the numerical and experimental examples, they illustrated that it can be run automatically without any prior knowledge of the structure or expert interference.

The primary objective of this chapter is to present output only modal identification and structural damage detection. Identification of multidegree of freedom (MDOF) linear time invariant (LTI) and linear time variant (LTV – due to damage) systems based on TF techniques – such as STFT, EMD, and wavelets – is discussed. STFT, EMD, and wavelet methods developed to date are reviewed in sufficient detail. In addition, an HT approach to determine frequency and damping is also presented. In this chapter STFT, EMD, HT, and wavelet techniques are developed for decomposition of free-vibration response of MDOF systems into modal components. Once the modal components are obtained, each one is processed using HT to obtain the modal frequency and damping ratios. In addition, the ratio of modal components at different degrees of freedom facilitates determination of mode shape. In cases with output only modal identification using ambient/random response, the random decrement technique is used to obtain free-vibration response.

The advantage of TF techniques is that they are signal-based; hence, they can be used for output only modal identification. A three degree of freedom 1:10 scale model test structure is used to validate the output only

modal identification techniques based on STFT, EMD, HT, and wavelets. Both measure free-vibration and forced-vibration (white noise) response are considered. The second objective of this chapter is to show the relative ease with which the TF techniques can be used for modal identification, and their potential for real-world applications wherein output only identification is essential. Recorded ambient vibration data processed using techniques such as random decrement technique can be used obtain the free-vibration response, so that further processing using TF-based modal identification can be performed.

4.2 Time-frequency (TF) methods: STFT, EMD and HT

In this section an introduction to TF methods STFT, EMD and HT is provided. The concept of analytical signal is described and the means to obtain it using HT is presented.

4.2.1 Analytical signal and Hilbert transform

Signals in nature are real valued, but for analysis it is often more convenient to deal with complex signals. One wants the real part, $s(t)$, of the complex signal, $s_a(t)$, to be the actual signal under consideration. How does one fix the imaginary part, $\overline{s(t)}$, to form the complex signal? In particular, if one wants to write a complex signal, how does one choose $\overline{s(t)}$? The standard method is to form the 'analytic' signal, $s_a(t)$,

$$s_a(t) = s(t) + j\overline{s(t)} \qquad [4.1]$$

where $j = \sqrt{-1}$. This can be achieved by taking the spectrum of the actual signal, $s(\omega)$, deleting the negative part of the spectrum, retaining only the positive part of the spectrum, multiplying it by a factor of 2, and then forming the new (complex) signal by Fourier inversion. More specifically, if one has a real signal, $s(t)$, calculate $s(\omega)$. Form the complex signal with the positive part of $s(\omega)$ only,

$$s_a(t) = 2\frac{1}{\sqrt{2\pi}} \int_0^\infty s(\omega)e^{j\omega t} d\omega \qquad [4.2]$$

The factor of 2 is inserted so that the real part of the complex signal will be equal to the real signal that one started out with. Therefore substituting for $s(\omega)$:

$$s_a(t) = \frac{1}{\pi} \int_0^\infty \int s(t')e^{-j\omega t'} e^{-j\omega t} dt' d\omega \qquad [4.3]$$

Using the fact that

$$\int_0^\infty e^{j\omega x} d\omega = \pi\delta(x) + \frac{j}{x} \qquad\qquad [4.4]$$

one has

$$\int_0^\infty e^{j\omega(t-t')} d\omega = \pi\delta(t-t') + \frac{j}{t-t'} \qquad\qquad [4.5]$$

hence

$$s_a(t) = \frac{1}{\pi}\int s(t')\left(\pi\delta(t-t') + \frac{j}{t-t'}\right) dt' \qquad\qquad [4.6]$$

or

$$s_a(t) = s(t) + \frac{j}{\pi}\int_{-\infty}^\infty \frac{s(t')}{t-t'} dt' \qquad\qquad [4.7]$$

The imaginary part turns out to be the HT:

$$\bar{s} = H[s(t)] = \frac{1}{\pi}\int_{-\infty}^\infty \frac{s(t')}{t-t'} dt' \qquad\qquad [4.8]$$

Hence,

$$s_a(t) = s(t) + jH[s(t)] = s(t) + j\bar{s} \qquad\qquad [4.9]$$

The complex signal thus formed, $s_a(t)$, is called the analytic signal. Note that, by definition, analytic signals are signals whose spectra consist only of positive frequencies. That is, the spectrum is zero for negative frequencies.

As per Equations [4.1]–[4.9], the analytic signal can be obtained by: (1) taking the FT of $s(t)$; (2) zeroing the amplitude for negative frequencies and doubling the amplitude for positive frequencies; and (3) taking the inverse FT. The analytic signal $s_a(t)$ can also be expressed as

$$s_a(t) = A(t)e^{j\varphi(t)} \qquad\qquad [4.10]$$

where $A(t)$ = instantaneous amplitude and $\varphi(t)$ = instantaneous phase.

4.2.2 Instantaneous frequency

In the analytic signal given by Equations [4.1] and [4.10], $A(t) = \sqrt{(s^2(t) + \bar{s}^2(t))}$ and $\varphi(t) = \arctan(\overline{s(t)}/s(t))$, the instantaneous frequency $\omega_i(t)$ is given by

$$\omega_i(t) = \dot{\varphi}(t) = \frac{d}{dt}\left(\arctan\left(\frac{\overline{s(t)}}{s(t)}\right)\right) \tag{4.11}$$

where

$$\frac{d\varphi}{dt} = \frac{1}{1+\left(\dfrac{\bar{s}^2(t)}{s^2(t)}\right)^2}\frac{d\,\overline{s(t)}/s(t)}{dt} \tag{4.12}$$

and

$$\frac{d}{dt}\frac{\overline{s(t)}}{s(t)} = \frac{s(t)\overline{\dot{s}(t)} - \overline{s(t)}\dot{s}(t)}{s^2(t)} \tag{4.13}$$

From Equations [4.12] and [4.13] one obtains

$$\omega_i(t) = \frac{d\varphi(t)}{dt} = \frac{\left(s(t)\overline{\dot{s}(t)}\right) - \overline{s(t)}\dot{s}(t)}{s^2(t) + \bar{s}^2(t)} \tag{4.14}$$

4.2.3 Short-time Fourier transform (STFT) and spectrogram

The FT of a signal $s(t)$ is given by $s(\omega) = 1/\sqrt{2\pi}\int s(t)e^{-j\omega t}dt$. The STFT, the first tool devised for analyzing a signal in both time and frequency, is based on the FT of a short portion of signal $s_h(\tau)$ sampled by a moving window $h(\tau - t)$. The running time is τ and the fixed time is t. Since the time interval is short compared to the whole signal, this process is called taking the STFT.

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}s_h(\tau)e^{-j\omega\tau}d\tau \tag{4.15}$$

where $s_h(\tau)$ is defined as follows:

$$s_h(\tau) = s(\tau)h(\tau - t) \tag{4.16}$$

in which $h(\tau - t)$ is an appropriately chosen window function that emphasizes the signal around the time t, and is a function $\tau - t$ i.e., $s_h(\tau) = s(\tau)$ for τ near t and $s_h(\tau) = 0$ for τ far away from t. Considering this signal as a function of τ, one can ask for its spectrum. Since the window has been chosen to emphasize signal at t, the spectrum will emphasize, and hence give an indication of, the frequencies at that time. In particular the spectrum is

$$S_t(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(\tau)h(\tau - t)e^{-j\omega\tau}d\tau \tag{4.17}$$

which is the STFT.

Summarizing, the basic idea is that if one wants to know the frequency content of the signal at a particular time, t, then take a small piece $s_h(\tau)$ of the signal around that time and Fourier analyze it, neglecting the rest of the signal, obtaining a spectrum at that time. Next take another small piece, of equal length of the signal, at the next time instant and obtain another spectrum. Proceed in this way till the entire signal is sampled. The collection of all these spectra (or slices at every time instant) gives a TF spectrogram, which covers the entire signal, and captures the localized TV frequency content of the signal. If one performs an FT then the localized variations of frequency content are lost, since FT is performed on the whole signal; one gets an average spectrum of all those obtained by STFT.

The energy density of the modified signal and the spectrogram are given by:

$$P(t, \omega) = \left| S_t(\omega)^2 \right| \text{ or} \tag{4.18}$$

$$P_{SP}(t, \omega) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(\tau)h(\tau - t)e^{-j\omega\tau}d\tau \right|^2 \tag{4.19}$$

By analogy with the previous discussion it can be used to study the behavior of the signal around the frequency point ω. This is done by choosing a window function whose transform is weighted relatively higher at the frequency ω.

$$H(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(t)e^{-j\omega\tau}d\tau \tag{4.20}$$

$$s_\omega(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S(\omega') H(\omega - \omega') e^{-j\omega't} \, d\omega' \qquad [4.21]$$

$$s_h(\tau) = s(\tau) h(\tau - t) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} S_t(\omega) e^{j\omega\tau} \, d\omega \right| \qquad [4.22]$$

where ω' is running frequency and ω is fixed frequency. The spectrogram is given by

$$P_{SP}(t, \omega) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(\omega') H(\omega - \omega') e^{-j\omega't} \, d\omega' \right|^2 \qquad [4.23]$$

The limitation of STFT is its fixed resolution (we will discuss this more in a later section on wavelets), which can be overcome with multiresolution analysis (MRA) using wavelets. In STFT, the length of the signal segment chosen, or the length of the windowing function $h(t)$, determines the resolution: a broad window results in better frequency resolution but poor time resolution, and a narrow window results in good time resolution but poor frequency resolution, due to the time-bandwidth relation (uncertainty principle; Cohen, 1995). Note $h(t)$ and $H(\omega)$ are FT pairs (Equation [4.20]) i.e., if $h(t)$ is narrow, more time resolution is obtained; however, $H(\omega)$ becomes broad, resulting in poor frequency resolution and vice versa.

4.2.4   STFT implementation procedure

The implementation procedure for the STFT in the discrete domain is carried out by extracting time windows of the original non-stationary signal $s(t)$. After zero padding and convolving the signal with the Hamming window, the DFT is computed for each windowed signal to obtain STFT, $s_t(\omega)$, of signal $s_h(\tau)$. If the window width is $n.\Delta t$ (where n is the number of points in the window, and Δt is the sampling rate of the signal), the kth element in $s_t(\omega)$ is the Fourier coefficient that corresponds to the frequency,

$$\omega_k = \frac{2\pi k}{n.\Delta t} \text{(for window width } n.\Delta t) \qquad [4.24]$$

4.2.5   Empirical mode decomposition (EMD)

For a multicomponent signal – as in a multimodal or MDOF response – the procedure described in the previous section to obtain the analytic signal

and instantaneous frequency cannot be applied directly, as described earlier. The EMD technique, developed by Huang (Huang, 1998), adaptively decomposes a signal into 'intrinsic mode functions (IMF),' which can then be converted to an analytic signal using HT. The TF representation and instantaneous frequency can be obtained from the intrinsic modes extracted from the decomposition, using HT. The principal technique is to decompose a signal into a sum of functions that (1) have the same numbers of zero crossings and extrema, and (2) are symmetric with respect to the local mean. The first condition is similar to the narrow-band requirement for a stationary Gaussian process. The second condition modifies a global requirement to a local one, and is necessary to ensure that the instantaneous frequency will not have unwanted fluctuations, as induced by asymmetric waveforms. These functions are called intrinsic mode functions (IMF denoted by imf_i) obtained iteratively (Huang $et\ al.$, 1998). The signal, $x_j(t)$, for example, j^{th} degree of freedom displacement of a MDOF system, can be decomposed as follows

$$x_j(t) = \sum_{i=1}^{n} imf_i(t) + r_n(t) \hspace{2cm} [4.25]$$

where $imf_i(t)$ are the 'IMF' (note: dominant IMFs are equivalent to individual modal contributions to $x_j(t)$ – which will be demonstrated in a later section) and $r_n(t)$ is the residue of the decomposition. The IMFs are obtained using the following algorithm:

1. Initialize; $r_0 = x_j(t)$, $i = 1$
2. Extract the imf_i as follows:
   a. Initialize: $h_0(t) = r_{i-j}(t)$, $j = 1$
   b. Extract the local minima and maxima of $h_{j-t}(t)$
   c. Interpolate the local maxima and the local minima by a spline to form upper and lower envelopes of $h_{j-t}(t)$, $e_{max}(t)$ and $e_{min}(t)$, respectively.
   d. Calculate the mean m_{j-1} of the upper and lower envelopes
   $$= \frac{e_{max}(t) + e_{min}(t)}{2}$$
   e. $h_j(t) = h_{j-1}(t) - m_{j-1}(t)$.
   f. If stopping criterion is satisfied then set $imf_i(t) = h_j(t)$, else go to (b) with $j = j + 1$
3. $r_i(t) = r_{i-1}(t) - imf_i(t)$
4. If $r_i(t)$ still has at least two extrema, then go to 2 with $i = i + 1$; otherwise the decomposition is finished and $r_i(t)$ is the residue.

The analytic signal, $s_a(t)$, and the instantaneous frequencies, $\omega_i(t)$, associated with each $imf_i(t)$ component can be obtained using Equations

[4.1]–[4.14] by letting $s(t) = imf_i(t)$ and $s_a(t) = s(t) + jH\left(\overline{s(t)}\right)$ for each IMF component.

To ensure that the IMF components retain the amplitude and frequency modulations of the actual signal, a satisfactory stopping criterion for the sifting process is defined (Rilling *et al.*, 2003). A criterion for stopping is accomplished by limiting the standard deviation, SD (Huang *et al.*, 1998), of $h(t)$, obtained from consecutive sifting results as

$$SD = \sum_{k=0}^{l} \left[\frac{\left|h_{j-1}(k\Delta t) - h_j(k\Delta t)\right|^2}{h_{j-1}^2(k\Delta t)} \right] \qquad [4.26]$$

where $l = T / \Delta t$ and T = total time. A typical value for SD is set between 0.2 and 0.3 (Rilling *et al.*, 2003). An improvement on this criterion is based on two thresholds θ_1 and θ_2, aimed at globally small fluctuations in the mean while taking into account locally large excursions. This amounts to introducing a mode amplitude $a(t)$ and an evaluation function $\sigma(t)$:

$$a(t) = \left(\frac{e_{max}(t) - e_{min}(t)}{2} \right) \qquad [4.27]$$

$$\sigma(t) = \left| \frac{m(t)}{a(t)} \right| \qquad [4.28]$$

Sifting is iterated until $\sigma(t) < \theta_1$ for a fraction of the total duration while $\sigma(t) < \theta_2$ for the remaining fraction. Typically $\theta_1 \approx 0.05$ and $\theta_2 \approx 10\theta_1$ (Rilling *et al.*, 2003).

4.3 Modal identification of linear time invariant (LTI) and linear time variant (LTV) systems using EMD/HT and STFT

EMD can be used to decompose a signal into its multimodal components (+ residual IMF components + residue). In a lightly damped system with distinct modes, EMD can extract the multicomponent modal contributions (or IMFs) from the *j*th degree of freedom (DOF) displacement response of a MDOF system. Each of these IMF components can then be analyzed separately to obtain the instantaneous frequency and damping ratios. If the displacement of the MDOF LTI system is represented by vector $\mathbf{x} = \Phi\mathbf{q}$,

where Φ = modal matrix, $\mathbf{q}$ = modal displacement vector, then combining it with Equation [4.25] leads to the following equation for $x_j(t)$, the jth degree of freedom displacement of a MDOF LTI system:

$$x_j(t) = \sum_{i=1}^{n} imf_i(t) + r_n(t) = \sum_{i=1}^{m} \Phi_{ji} q_i(t) + \sum_{i=m}^{n} imf_i(t) + r_n(t) \qquad [4.29]$$

where m = number of modes of a MDOF system, and IMFs from m to n are treated as residual terms along with the actual residuals and discarded.

The equation of motion of an MDOF is given by

$$M\ddot{x} + C\dot{x} + Kx + MRf \qquad [4.30]$$

substituting $\mathbf{x} = \Phi\mathbf{q}$,

$$\Phi^T M\Phi\ddot{q} + \Phi^T C\Phi\dot{q} + \Phi^T K\Phi q = \Phi^T MRf \qquad [4.31]$$

The proportional damped system with orthonormal Φ leads to m uncoupled equations of motion with

$$\mathrm{diag}[\Lambda_c] = \begin{bmatrix} 2\xi_1\omega_1 & 0 & . & . & . & 0 \\ 0 & 2\xi_2\omega_2 & . & . & . & 0 \\ . & & . & . & . & . \\ . & & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 2\xi_n\omega_n \end{bmatrix} \quad \mathrm{diag}[\Lambda_k] = \begin{bmatrix} \omega_1^2 & 0 & . & . & . & 0 \\ 0 & \omega_2^2 & . & . & . & 0 \\ . & & . & . & . & . \\ . & & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & \omega_n^2 \end{bmatrix}$$

$$\ddot{q}_k + 2\xi_k\omega_k\dot{q}_k + \omega_k^2 q_k = \Gamma_k f \qquad [4.32]$$

where $\Gamma_k = \Phi_k^T MR$, with f as input and q_k as output, taking the Laplace transform

$$(s^2 + 2\xi_k\omega_k s + \omega_k^2)q_k(s) = \Gamma_k f(s) \qquad [4.33]$$

Dropping Γ_k for generality, the transfer function is then given by

$$H_k(s) = \frac{1}{s^2 + 2\xi_k\omega_k s + \omega_k^2} \qquad [4.34]$$

and the frequency response function (FRF) is given by

$$H_k(j\omega) = \frac{1}{-\omega^2 + j2\xi_k\omega_k\omega + \omega_k^2}$$ [4.35]

Noting $x_k = \Phi_k q_k$ and x_{jk} as the jth component of the displacement contributed by the kth mode, and with f as input and x_{jk} as output, the transfer function

$$H_{jk}(j\omega) = \frac{1}{(\omega_k^2 - \omega^2) + j2\xi_k\omega_k\omega}\phi_{jk}$$ [4.36]

If the structure is lightly damped, the peak transfer function occurs at $\omega = \omega_k$ with amplitude

$$\left|H_{jk}(j\omega)\right| = \frac{\sqrt{1 + 4\xi_k^2}}{2\xi_k}\phi_{jk}$$ [4.37]

From Equation [4.37] it is seen that the magnitudes of the FRF peaks at $\omega = \omega_k$ are proportional to the components of the kth modal vector. The sign of these components can be determined by phases associated with the FRFs: if two modal components are in phase they are of the same sign, and if the two modal components are out-of-phase they are of opposite sign. With knowledge of the magnitude of the peaks, the damping factor, ξ_k can be solved from Equation [4.37]. From Equation [4.36] summing over all modes gives

$$H_{ij}(j\omega) = \sum_{k=1}^{n} \frac{\phi_{ik}\phi_{jk}}{(\omega_k^2 - \omega^2) + j2\xi_k\omega_k\omega}$$ [4.38]

which can be written as

$$H_{ij}(j\omega) = \sum_{k=1}^{n} \frac{{}_kA_{ij}}{(\omega_k^2 - \omega^2) + j2\xi_k\omega_k\omega}$$ [4.39]

with ${}_kA_{ij} = \phi_{ik}\phi_{jk}$ being the residues or modal components. Taking the inverse transform of Equation [4.39] gives the general form of the impulse response function (IRF):

$$h_{ij}(t) = \sum_{k=1}^{n} \frac{{}_kA_{ij}}{\omega_{dk}} e^{-\xi_k\omega_k t} \sin(\omega_{dk}t) \qquad [4.40]$$

where $\omega_{dk} = \omega_k\sqrt{1-\xi_k^2}$ = damped frequency of the kth mode. It follows from Equation [4.39] that the MDOF LTI system frequency responses are the sum of n single degree of freedom frequency responses, provided that well separated modes and light proportional damping are valid, and the residues and the modes are real. For non-proportional damped systems the residues and modes become complex.

Consider the function $e^{-\sigma_k+j\omega_k t}$ with $\sigma_k = \xi_k\omega_{nk}$ and $\omega_k = \omega_{dk}$. For a damped asymptotically stable system with $\sigma > 0$, we can rewrite the Equation [4.36] for mode k by taking inverse FT

$$h_{jk}(t) = A_{jk}e^{-\sigma_k t}\sin(\omega_{dk}t) \qquad [4.41]$$

$$\bar{h}_{jk}(t) = A_{jk}e^{-\sigma_k t}\cos(\omega_{dk}t) \qquad [4.42]$$

where $A_{jk} = \Phi_{jk}/\omega_{dk}$, leading to the analytic signal

$$h_{jk_a}(t) = h_{jk}(t) + j\bar{h}_{jk}(t) \qquad [4.43]$$

that can be written as

$$h_{jk_a}(t) = \bar{A}_{jk}e^{j\varphi t} \qquad [4.44]$$

The magnitude of this analytic signal is given by:

$$\left|h_{jk_a}(t)\right| = \left|\bar{A}_{jk}\right| = \sqrt{h_{jk_a}^2(t) + h_{jk_a}^{-2}(t)} \qquad [4.45]$$

Substituting Equation [4.3] and simplifying results:

$$\left|\bar{A}_{jk}\right| = A_{jk}e^{-\sigma_k t} \qquad [4.46]$$

Taking the natural logarithm of this expression yields:

$$\log\left|\bar{A}_{jk}\right| = -\sigma_k t + \log(A_{jk}) = -\xi_k\omega_n t + \log(A_{jk}) \qquad [4.47]$$

4.3.1   Modal identification based on empirical mode decomposition

Nagarajaiah and co-workers originally developed the EMD/HT modal identification approach for tuning STMD in 2001 (Nagarajaiah and Varadarajan, 2001), based on their earlier work (Nagarajaiah *et al.*, 1999) on variable stiffness systems. The advantage of this approach is that it is signal-based and output only; hence, measured response at any one DOF can be used to make useful estimates of the instantaneous frequency and damping ratio. However, to estimate modeshapes, response signals at more degrees of freedom will be needed. Each significant IMF component represents one modal component with unique instantaneous frequency and damping ratio.

Individual mode FRF and corresponding IRF can be extracted when band-pass filters (Thrane, 1984) are applied to the system FRF. Equation [4.46] can be used to estimate damping in the kth mode, as suggested originally by Thrane in 1984 and adopted by Agneni in 1989. In 2003 Yang and co-workers (Yang *et al.*, 2003, 2004) extended this approach by using EMD/HT to decompose and obtain IMFs and perform modal identification. In the case when the inputs are white noise excitation and the output accelerations at a certain floor are available, one can obtain the free-vibration response from the stationary response to white noise using the random decrement technique followed by instantaneous frequency and damping calculations.

The EMD/HT outlined below was developed independently by Nagarajaiah and co-workers (Nagarajaiah and Varadarajan, 2001):

1. Obtain signal $x_j(t)$, jth degree of freedom displacement of a MDOF system, from the feedback response.
2. Decompose the signal $x_j(t)$ into its IMF components as described in Equations [4.29] and [4.25].
3. Construct the analytical signal for each IMF/modal component using HT method described in Equation [4.9].
4. Obtain the phase angle of the analytic signal and then obtain the instantaneous frequency from Equation [4.14].
5. Obtain the log amplitude function of the analytic signal; perform least squares line fit to the function (which will be a decreasing function fluctuating about a line and not necessarily linear at all times). Then using Equation [4.47] compute the slope and damping ratio.
6. The ratio of modal components at different degrees of freedom facilitate determination of modeshape.
7. In the case of output only modal identification with ambient/random excitation, use the random decrement technique (Ibrahim, 1977) to obtain the free-vibration response.

4.3.2   Modal identification based on STFT

After obtaining a spectrogram, an FRF at a given time can be extracted, and the individual mode FRF and corresponding IRF can be extracted when band-pass filters (Thrane, 1984) are applied. The frequencies can be identified by applying HT to the IRF as per Equation [4.44]. Equation [4.46] can be used to estimate damping in the kth mode, as suggested originally by Thrane in 1984 and adopted by Agneni in 1989. The ratios of modal components at different degrees of freedom facilitate determination of modeshape. In the case of output only modal identification with ambient/random excitation, the random decrement technique can be used to obtain free-vibration response.

4.4   Modal identification of LTI and LTV systems using wavelets

The wavelet function can be defined as

$$(W_\psi x)(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \qquad [4.48]$$

where b is a translation indicating the locality, a is a dilation or scale parameter, $\psi(t)$ is an analyzing (basic) wavelet and $\psi^*(\cdot)$ is the complex conjugate of $\psi(\cdot)$. Each value of the wavelet transform $(W_\psi x)(a,b)$ is normalized by the factor $1/\sqrt{a}$. This normalization ensures that the integral energy given by each wavelet $\psi_{a,b}(t)$ is independent of the dilation a. The function $\psi(t)$ qualifies as an analyzing wavelet when it satisfies the admissibility condition

$$C_\psi = \int_{-\infty}^{\infty} \left(\frac{|\psi(\omega)|^2}{|\omega|} \right) d\omega < \infty \qquad [4.49]$$

where $\psi(\omega)$ is the FT of $\psi(t)$. This is necessary for obtaining the inverse of the wavelet transform given by

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(W_\psi x)(a,b) 1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right) \frac{da\,db}{a^2} \qquad [4.50]$$

The possibility of TF localization arises from the $\psi(t)$ being a window function, which means that additionally

$$\int_{-\infty}^{\infty} |\psi(t)| dt < \infty \qquad [4.51]$$

which follows from Equation [4.45]. One of the most widely used functions in wavelet analysis is the Morlet wavelet defined by:

$$\psi(t) = e^{j2\pi f_o |t|} e^{-|t|^2/2} \qquad [4.52]$$

The Fourier spectrum of Morlet wavelet is a shifted Gaussian function

$$\widehat{\psi}(f) = \sqrt{2}\pi e^{-2\pi^2(f-f_o)^2} \qquad [4.53]$$

In practice the value of $f_0 > 5$ is used. The wavelet transform is a linear representation of a signal. Thus it follows that for a given N functions x_i and N complex values $\alpha_i \, (i = 1, 2, \ldots, N)$

$$\left(W_\psi \sum_{i=1}^{N} \alpha_i x_i \right)(a,b) = \sum_{i=1}^{N} \alpha_i \left(W_g x_i \right)(a,b) \qquad [4.54]$$

The frequency localization is clearly seen when the wavelet transform is expressed in terms of the FT,

$$\left(W_\psi x \right)(a,b) = \sqrt{a} \int_{-\infty}^{+\infty} X(f) \psi_{a,b}^*(af) e^{j2\pi fb} df \qquad [4.55]$$

where $\psi^*(\cdot)$ is the complex conjugate of $\psi(\cdot)$. This localization depends on the dilation parameter a. The local resolution of the wavelet transform in time and frequency is determined by duration and bandwidth of the analyzing functions given by

$$\Delta t = a\Delta t_\psi, \quad \Delta f = \frac{\Delta f_\psi}{a} \qquad [4.56]$$

where Δt_ψ and Δf_ψ are duration and bandwidth of the basic wavelet function respectively. For the Morlet analyzing wavelet function, the relationship between the dilation parameter a_f and the signal frequency f_x at which the analyzing wavelet function is focused, can be given as

$$a_f = f_0 \left(\frac{f_s}{f_w} \right) \left(\frac{1}{f_x} \right) \qquad [4.57]$$

where f_s and f_w are the sampling frequencies of the signal and the analyzing wavelet, respectively. The frequency bandwidth of the wavelet function for the given dilation a can be obtained using a frequency representation of the Morlet wavelet and expressed as

$$\Delta f_x = \left(\frac{1}{\pi a}\right)\left(\frac{f_s}{f_w}\right) \qquad [4.58]$$

which allows one to obtain a single element of the wavelet decomposition of the function for a given value of frequency (dilation) and frequency bandwidth.

The wavelets are scaled to obtain a range of frequencies. They are also translated to provide the time information in the transform. The wavelet transform works as a filter, allowing only a certain time and frequency content through. Any given atom in the TF map of the wavelet transform (see Fig. 4.1) represents the correlation between the wavelet basis function at that frequency dilation and the signal in that time segment. The frequency content of the wavelet transform is represented in terms of scales, which are inversely related to frequencies. The squared amplitude of the continuous wavelet transform (CWT) is therefore called the scalogram. The relationship between scales and frequencies can be used to form a TF map from the scalogram.

Since the wavelet works in a manner similar to the STFT, by convolving the signal with a function that varies in both time and frequency, it suffers from similar limitations in the resolution of the TF map. Both transforms are confined by the uncertainty principle, which limits the area of a TF atom in the overall TF map (see Fig. 4.1). The biggest difference between the two transforms is that the atoms in the WT map are not a constant shape. In the lower frequencies, the atoms are fatter, providing a better resolution in frequency and worse resolution in time, whereas in the upper frequencies the atoms are taller, providing better time resolution and worse frequency resolution. This variable resolution can be advantageous in the analysis of structural time response data.

The CWT gets its name from the fact that the mother wavelet is continuously shifted across the length of the data being analyzed. This smooth shifting means that the time/frequency atoms shown in Fig. 4.1 will overlap one another, providing redundant information.

The variable windowing feature of wavelet analysis leads to an important property exhibiting constant Q-factor (defined as the ratio of the center frequency to bandwidth) analysis. For STFT, at an analyzing frequency ω_0, changing the window width will increase or decrease the number of cycles of ω_0 inside the window. In the case of wavelet transforms, with the change
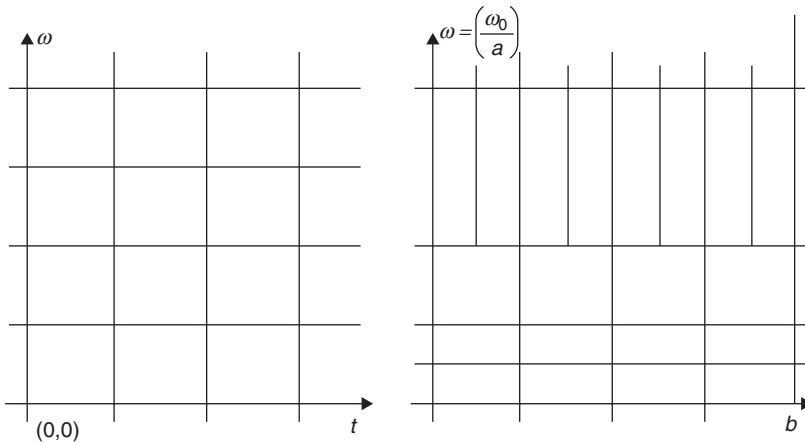
4.1 Comparing STFT and wavelet transform resolution in time and frequency domain.

in window width, mean dilation or compression of the wavelet function changes. Hence, the carrier frequency becomes ω_0 / a, for a window width changing from T to aT. However, the number of cycles inside the window remains constant.

The frequency resolution is proportional to the window width, in the case of both STFT and wavelet transform. However, for wavelet transform, a center frequency shift necessarily accompanies a window width change (time scaling). Thus, Q-factor is invariant with respect to wavelet dilation, and these dilated wavelets may be considered as constant Q band-pass filters, giving rise to the frequency selectivity of the CWT.

Since the wavelet transform is an alternative representation of a signal, it should retain the characteristics of the signal including the energy content in the signal. Thus, there should exist a similar relation to the Parseval's theorem, which provides the energy relationship in Fourier domain. The total energy of a signal in wavelet domain representation is:

$$E_f = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |WT(u,s)|^2 \frac{ds\,du}{s^2} \qquad [4.59]$$

where C_ψ is a scalar constant related to the FT of the wavelet basis (called 'admissibility constant'). The wavelet basis functions can be normalized in a way such that it can attain a value of unity. The differential energy of the signal in the differential tile of scale-translation plane in the wavelet domain is $|WT(u,s)|^2 \, ds\,du/s^2$, which leads to the construction of the scalogram.

4.4.1   Estimates of modal parameters in MDOF systems

Since the analyzing wavelet function has compact support in the time and frequency domains, multicomponent signals can be written as

$$\left(W_\psi x \sum_{i=1}^{N} x_i\right)(a,b) = \frac{1}{\sqrt{a}} \sum_{i=1}^{N} \int_{t-a\Delta t_\psi}^{t+a\Delta t_\psi} x(t) \psi^* \left(\frac{t-b}{a}\right) dt \qquad [4.60]$$

The response of an underdamped SDOF system can be expressed in the form

$$x(t) = A(t) e^{\pm j\omega_n \sqrt{1-\xi^2} t} \qquad [4.61]$$

Assuming the envelope $A(t)$ is slowly varying it follows (Staszewski, 1997; Chakraborty *et al.*, 2006):

$$\ln \left|(W_\psi x)(a,b)\right| \approx -\xi\omega_n b + \ln\left(A_0 \left|\psi^* \left(\pm a_0 j\omega_n \sqrt{1-\xi^2}\right)\right|\right) \qquad [4.62]$$

Subsequently, the response of the MDOF system can be obtained as

$$\left|\left(W_\psi x \sum_{i=1}^{N} x_i\right)(a,b)\right| \approx \sum_{i=1}^{N} A_i e^{-\xi\omega_n b} + \left|\psi^* \left(\pm ja_i \omega_{n_i} \sqrt{1-\xi_i^2}\right)\right| \qquad [4.63]$$

Due to the compact support of the analyzing wavelet functions in time and frequency, the wavelet transform of each separate mode $i = 1, 2, \ldots, N$ becomes (Staszewski, 1997; Chakraborty *et al.*, 2006)

$$\left|(W_\psi x_i)(a,b)\right| \approx A(b) e^{-\xi_i \omega_{n_i} b} \left|\psi^* \left(\pm ja_i \omega_{n_i} \sqrt{1-\xi_i^2}\right)\right| \qquad [4.64]$$

For the given value of dilation a_i related to the natural frequency f_{n_i} of the system, the modulus of the wavelet transform plotted in a semi-logarithmic scale leads to

$$\ln \left|(W_\psi x_i)(a_i,b)\right| \approx -\xi_i \omega_{n_i} b + \ln\left(A_i \left|\psi^* \left(\pm a_i j\omega_{n_i} \sqrt{1-\xi_i^2}\right)\right|\right) \qquad [4.65]$$

and forms the basis of identifying the damping.

The derivations so far are general, and are applicable to any continuous wavelet basis with desired or suitable TF characteristics. Section 4.4.2 provides the detail of a wavelet basis used for identifying the modal parameters of an MDOF system.

4.4.2   Modified Littlewood-Paley (L-P) basis

An equivalent of the harmonic wavelet when the basis function is real is the Littlewood-Paley (L-P) wavelet. This wavelet basis function is defined by

$$\psi(t) = \frac{1}{2\pi} \frac{\sin(4\pi t) - \sin(2\pi t)}{t} \qquad [4.66]$$

A possible variation of the wavelet is one that retains the characteristic of the basis function (close to transient vibration signals, i.e. oscillatory and decaying) but could reduce the frequency bandwidth of the mother wavelet. Hence, the derived modified wavelet is called the modified L-P wavelet, and has been proposed and used by Basu and Gupta (Gupta, 1999a, b). The shifted and scaled version of this is called the baby modified L-P wavelet. This wavelet basis has also been used by Basu (2005, 2007) for damage detection in structures.

The modified L-P basis function is defined by

$$\psi(t) = \frac{1}{\pi\sqrt{\sigma - 1}} \frac{\sin(\sigma\pi t) - \sin(\pi t)}{t} \qquad [4.67]$$

where σ (is a scalar) > 1. In the frequency domain the wavelet basis can be represented by

$$\hat{\psi}(t) = \left\{ \left(\frac{1}{(2\pi(\sigma - 1))} \right) \right\} \ \& \ \text{for} \ \pi \le |\omega| \le \sigma\pi \ @ 0 \ \& \ \text{elsewhere}$$

By choosing appropriate values for the bandwidth, the frequency content of the mother wavelet can be adjusted. If for numerical computation the scaling parameter is discretized as $a_j = \sigma^j$ (in an exponential scale), then the scaled version of the mother basis function has mutually non-overlapping frequency bands and is also orthogonal. This property can be conveniently utilized to detect natural frequencies and modal properties for the dynamical systems, as can be seen in the following sections.

4.4.3   Wavelet packets

While the constant Q-factor and coarser frequency resolution at high frequencies make the wavelet analysis computationally efficient, this may be a disadvantage for analysis of some signals for system/modal identification and structural health monitoring. Better resolution at high frequencies can be obtained by wavelet packet construction.

The discrete wavelet transform based on MRA splits the signal into two bands, a higher band (by using a high pass filter) and a lower band (by using a low pass filter). The lower band is again subsequently split in two bands. This concept can be generalized by splitting the signal into several bands each time. In addition, there could further splitting of higher bands too, not just the lower band. This generalization of MRA produces outputs called wavelet packets. This is a deviation from constant Q analysis and achieves the desired frequency resolution at high frequency bands. Wavelet packets through arbitrary band splitting can choose the most suitable resolution to represent a signal.

The resolution of signals with wavelet packets is not only possible using MRA-based frequency filters in the time domain (starting with Haar wavelets) but also in the frequency domain. For the arbitrary resolution using frequency domain-based filters, the construction for wavelet packets should be based on a modified L-P wavelet basis. The application of wavelet packets is particularly useful in system identification and damage detection for SHM where finer resolution at higher frequency is desired.

4.4.4   Identification of modal parameters

To detect the frequency bands in which the natural frequencies lie, the energy corresponding to each band is calculated for a particular state of response using Equation [4.59]. The bands, which do not contain the natural frequencies, lead to insignificant energy contribution. Hence, the first 'n' bands with significant energy content are the bands where the natural frequencies are located. These bands are in increasing order corresponding to the first 'n' natural frequencies, i.e. the lowest frequency band has the first natural frequency, and so on.

However, the chosen bands may lead to bands with a relatively broad interval in which the natural frequencies lie. To refine the estimates into finer intervals, so that natural frequencies could be determined to a better precision, wavelet packets are used. This is an extension of wavelet transform to provide level-by-level TF description and is easily adaptable for the modified L-P basis. The wavelet packet enables extraction of information from signals with an arbitrary TF resolution satisfying the product constraint in the TF window. In this technique, to refine the estimation of the

kth natural frequency, ω_{n_k}, located in the j_kth band, i.e. with frequency band $[\pi/a_{jk}, \sigma\pi/a_{jk}]$, further re-division is carried out. If it is required to further subdivide the band into 'M' parts, then again an exponential scale is used to divide the band so that the corresponding time-domain function forms a wavelet basis function. In this approach (also sometimes, termed as sub-band coding), for the j_kth band, the mother basis for the packet, $\psi^s(t)$ is formed with the frequency domain description:

$$\psi^{S(\omega)} = \left\{ \left(\left(\frac{1}{\sqrt{2\pi(\delta-1)}} \right) \& \text{ for } \pi \le |\omega| \le \delta\pi \quad @0 \& \text{elsewhere} \right) \right\}$$

where $\delta^M = \sigma$ [with δ (a scalar) > 1]. The corresponding time-domain description is given by:

$$\psi^s(t) = \frac{1}{\pi\sqrt{(\delta-1)}} \frac{\sin(\delta\pi t) - \sin(\pi t)}{t} \qquad [4.68]$$

The frequency band for the pth sub-band within the original j_kth band is the interval $[\delta^{p-1}\pi/a_{j_k}, \delta^p\pi/a_{j_k}]$. The basis function for this is denoted by $\psi^{sp}_{a_{jk},b}(t)$. The wavelet coefficient in this sub-band is denoted by $W_{\psi_{sp}}x_m(a_{j_k},b)$. Using the wavelet coefficients in these sub-bands and then applying an expression similar to that in Equation [4.59] to estimate the relative energies in the sub-bands, the natural frequencies can be obtained more precisely.

Once the natural frequencies are obtained and the corresponding bands are identified, the following expression corresponding to the sub-band containing the kth natural frequency with scale parameter j_k and the sub-band parameter 'p' is considered to obtain the kth modeshape.

$$W_\psi x_i(a_j,b) = \sum_{k=1}^{N} \phi_i^k W_\psi(a_{jk},b); i = 1,2,\ldots,N \qquad [4.69]$$

Considering the response or two states or DOF in a MDOF system (with one arbitrarily chosen as $i = 1$, without loss of generality), the ratio of wavelet coefficients of the two considered degrees of freedom at any instant of time $t = b$, corresponding to band j_k

$$\prod_{m}^{j_k} = \frac{W_\psi x_i(a_{jk},b)}{W_\psi x_1(a_{jk},b)} = \frac{\phi_m^k}{\phi_1^k} \qquad [4.70]$$

Thus it is seen that the computed ratios of the wavelet coefficients are invariant with 'b'. These ratios for different states corresponding to different values of 'm,' and assuming $\phi_1^j = 1$ (without loss of generality), the mode-shape for the kth mode (in j_k band with further sub-band division) can be obtained as

$$\{\phi_m^k\} = \left\{ \prod_m^{j_k} \right\}, m = 1, 2, \ldots, N \qquad [4.71]$$

4.4.5 Wavelet-based online monitoring of LTV systems with stiffness changes

Consider a linear TV MDOF system with m degrees of freedom represented by the set of linear TV ordinary differential equations with $[M], [C(t)]$, and $[K(t)]$ as the mass, TV damping, and TV stiffness matrices respectively. The displacement response vector is denoted by $\{X(t)\} = \{x_1(t) x_2(t) \ldots x_m(t)\}^T$, Let us assume that the functions $K_{ij}(t); i, j = 1, \ldots, m$ in the stiffness matrix have discontinuities at a finite number of points. It is then possible to divide the time into several segments with indices arranged as $t_0 < t_1 < t_2 < \ldots t_n$ such that all $K_{ij}(t); i, j = 1, \ldots, m$ are continuous functions in $[t_{i-1}, t_i]$. Further, it is assumed that the variation of all the TV stiffness functions $K_{ij}(t)$ is slower than the fundamental (lowest) frequency of the system (corresponding to the longest period). It subsequently follows that the variation of $\{X(t)\}$ may be represented with a slowly varying amplitude $\{\phi(t)\}_i^k$ and a slowly varying frequency $\omega_{k_i}(t)$ at the kth mode in the time interval $[t_{i-1}, t_i]$.

The modified L-P function has been used as the wavelet basis for analysis for this problem and the basis is characterized by the FT

$$\psi^{(t)} = \left\{ \left(\left(\frac{1}{\sqrt{F_1(\sigma - 1)}} \right) \& \text{ for } F_1 \leq |\omega| \leq \sigma F_1 \ @0 \& \text{ otherwise} \right) \right\}$$

where F_1 is the initial cut-off frequency of the mother wavelet. If this modified L-P basis function is used, then $\hat{\psi}(a_j\omega)$ is supported over $[\sigma F_1/a_j, F_1/a_j]$. It follows that if $\omega_{k_i}(b)$ corresponding to the kth mode is in the j_k^{th} band i.e. $\omega_{k_i}(b) \in [F_1/a_j, F_1/a_{jk}]$ then it can be approximated as

$$\omega_{k_i}(b) \approx \omega_{0jk} = \frac{\sigma + 1}{2} \cdot \frac{\pi}{a_{jk}} \qquad [4.71]$$

for a lightly damped system (with $\eta_k = 1$), where ω_{0jk} is the central frequency of the j_k^{th} band. Let the parameters $\omega_{1_i}(b), \omega_{2_i}(b), \ldots, \omega_{m_i}(b)$ be contained

in the bands with scale parameters identified by indices respectively. Since the response, $z_k(t)$, in the kth mode i.e. in j_k^{th} band, is narrow banded with frequency around $\left[F_1/a_j, F_1/a_{jk}\right]$, it follows that the bands not containing the natural frequency have insignificant energy, which leads to the approximation

$$\left|W_{\psi j} Z_k\left(a_j, b\right)\right| \approx 0 \text{ if } j \neq j_k; k = 1, 2, \ldots, m \qquad [4.73]$$

Thus, the 'm' bands with the 'm' natural frequency parameters $\omega^{k_i}(b); k = 1, \ldots, m$ correspond to m local maxima in the variation of temporal energy, $E_j x_r(b)$ [or its proportional quantity $\left(1/a_j\right)\int_{b-\varepsilon}^{b+\varepsilon}\left|W_\psi x_r(b)\right|^2 db$] (with the integral over $b-\varepsilon$ to $b+\varepsilon$ for the windowed data in case of online identification) with different values of the band parameter 'j.' It may be noted that since the wavelet basis is localized in time, the integral over the window is acceptable with the parameter ε dependent on the frequency scale corresponding to j. If the forcing function is assumed to be described by a broad-banded excitation, then by calculating the relative energies in different bands and comparing, it may be concluded that

$$E_{j-1} x_r(b) < E_j x_r(b) > E_{j+1} x_r(b); \forall j = j_k; k = 1, 2, \ldots, m \qquad [4.74]$$

if the modes are not too closely spaced. Once these bands are detected, the parameters $w_{k_i}(b)$ can be obtained as

$$\omega_{k_i}(b) \approx \frac{\sigma+1}{2} \cdot \frac{F_1}{a_{jk}}; k = 1, 2, \ldots, m \qquad [4.75]$$

over the interval $[b-\varepsilon, b+\varepsilon]$. The sub-band coding with wavelet packets could be applied if the parameters $\omega_{k_i}(b)$ are desired to be obtained with better precision. Once the bands corresponding to the 'm' modes with the parameters $\omega_{k_i}(b)$ are obtained, the TV mode shapes $\{\phi(t)\}_i^k$ could be found by considering the wavelet coefficients of $x_r(t)$ with the scale parameters, j_k, and sub-band parameter p (for wavelet packets). Now, considering two different states of response of the MDOF system, with one considered as $r = 1$ (without loss of generality), the ratio of wavelet coefficients of the considered states at the time instant $t = b$ gives the rth component of the TV kth mode as

$$\pi_r^{j_k}(b) = \frac{W_{\psi_{sp}} x_r\left(a_{jk}, b\right)}{W_{\psi_{sp}} x_1\left(a_{jk}, b\right)} = \frac{\phi_r^k(b)}{\phi_1^k(b)} \qquad [4.76]$$

4.5   Experimental and numerical validation of modal identification of LTI and LTV systems using STFT, EMD, wavelets and HT

This section presents the experimental and numerical validation of the introduced methods for modal identification of LTI and LTV systems.

4.5.1   Modal identification of 1:10 scale three-story model using free-vibration test results and STFT

The 1:10 scale three-story model with a total weight of 1000 lbs shown in Fig. 4.2 is used for a modal identification study based on the proposed STFT and EMD/HT algorithm. The time axis is scaled by $\sqrt{10}$ from the prototype scale for this study. The measured third-floor free-vibration displacement response, shown in Fig. 4.2, is used for output only modal identification. Tests were also performed with white noise excitation, and the FRF was estimated – for further details refer to Nagarajaiah (2009). The identified frequencies of the 3DOF structure, both from free-vibration (output only) as well as forced-vibration test (input–output), are 5.5, 18.7, and 34 Hz for the three modes, respectively. The identified damping ratios are approximately 1.9%, 1.7%, and 1.1% in the three modes, respectively, as shown in Table 4.1. At the prototype scale the three modal frequencies are 1.75, 5.9, and 10.7 Hz, respectively.

4.2 Three-story 1:10 scale building model.

Table 4.1 Frequencies and damping ratios estimated using EMD/HT

Mode	Free-vibration tests		White noise tests	
	Identified frequency (Hz)	Identified damping ratio (%)	Identified frequency (Hz)	Identified damping ratio (%)
1	5.5	1.9	5.5	1.5
2	18.7	1.0	18.7	1.0
3	34	1.1	33.7	1.0

STFT is applied to the free-vibration displacement response of three-story scaled building model. Plate II in the color section between pages 374 and 375 shows the time history (lower right), frequency spectrum (upper left), and the TF spectrogram (upper right). The evolution of the frequency content of the displacement signal as a function of time can be seen in the spectrogram or TF distribution (upper right), shown in Plate II. If one examines the time history alone (lower right), the localized nature of the TV frequency content is not evident. The modal free-vibration response in the three separate modes and the time localization for each mode are clearly evident in the spectrogram, but not in the frequency spectrum or the time history – when examined independently. The three modal frequencies 5.5, 18.7, and 34 Hz (Nagarajaiah, 2009) are clearly evident in the spectrogram and the frequency spectrum (upper left) shown in Plate II. After the STFT spectrogram reveals the modal frequencies, further processing is essential using band-pass filtering to obtain modal components as described in Section 4.3.2. Next we present EMD/HT- and wavelet/HT-based methods that can accomplish output only modal identification without the use of band-pass filters.

4.5.2   Validation of EMD/HT technique using three-story model free-vibration test results

The three-story scaled model, with first mode frequency of 5.5 Hz, is subjected to free-vibration tests. The measured third-floor free-vibration acceleration response signal (we use the acceleration signal since the third mode is dominant, while in the displacement signal it is not dominant) is then analyzed using EMD/HT to extract instantaneous frequency and damping ratios of the three modes as per the procedure described earlier in Section 4.3.2. The free-vibration acceleration response of the third floor is shown in Fig. 4.3. The first three modes are not clearly evident in the time history, as all three modes are present simultaneously and decay at different rates; hence, the need for TF analysis exists to understand localization.
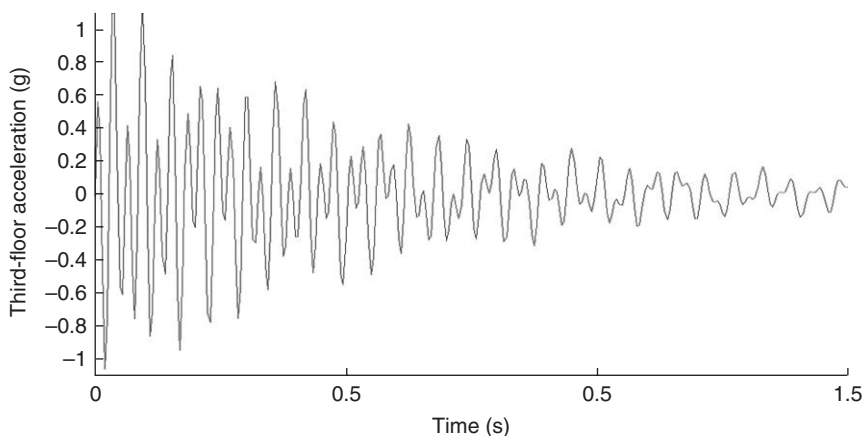
4.3 Measured third-floor acceleration free-vibration response.

The EMD method is capable of extracting all three vibration frequencies and damping ratios from a single measurement of the acceleration response time history based on the procedure outlined in Section 4.3.1. The third-floor acceleration is decomposed into IMFs; the first three are shown in Fig. 4.4 and the rest are discarded, as they are small and below the threshold. Based on the modal identification procedure presented in Section 4.3.1, modal frequencies and damping ratios are identified using linear least squares fit applied to the HT; log amplitude and phase (Equations [4.14], [4.43]–[4.47]) of HT of IMF3 is shown in Fig. 4.5. The modal frequencies and damping ratios obtained are shown in Table 4.1.

IMFs of all three-floor accelerations are obtained. Magnitude/phase information of IMF3 of the three-floor accelerations at a particular time provides the first mode. Similarly second and third modes are obtained. The identified modeshapes (scaled to maximum of 1) are shown in Table 4.2. The analytical modeshapes are shown in Table 4.3. The EMD results are in agreement with the analytical results.

4.5.3   Validation of wavelet/HT technique using three-story model free-vibration test results

The three-story scaled model is subjected to free-vibration tests. The measured third-floor free-vibration displacement response signal is then analyzed using wavelets to extract instantaneous frequency and damping ratios of the three modes, as per the procedure described earlier in Section 4.4. The scalogram of the free-vibration displacement response of the third floor is shown in Plate III in the color section between pages 374 and 375, and relevant wavelet coefficients of the measured free-vibration displacement
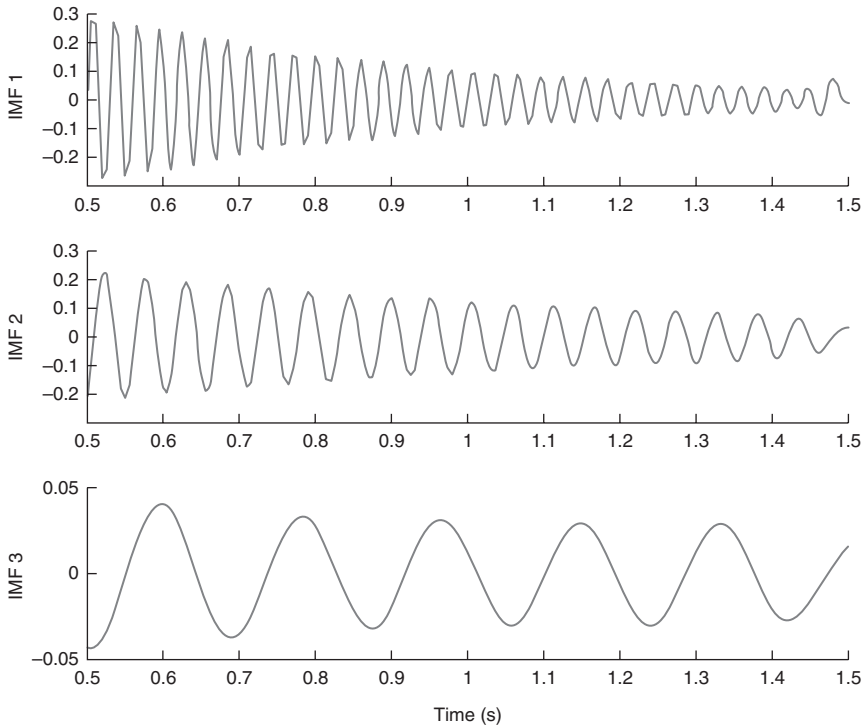
4.4 IMF components of the third-floor acceleration free-vibration response.

Table 4.2 Mode shapes estimated using EMD

	Mode 1	Mode 2	Mode 3
Story-3	1.0000	−0.7025	−0.3787
Story-2	0.6976	0.3265	1.0000
Story-1	0.4696	1.0000	−0.6185

response of all three floors and modes are shown in Fig. 4.6 (the wavelet coefficients have been normalized to have a peak value of 1 in Mode 1). All three modes and their decrement as a function of time are clearly evident in Plate III and Fig. 4.6. The modal frequencies and damping ratios are identified using linear least squares fit applied to the HT; log amplitude and phase (Equations [4.14], [4.43]–[4.47]) of HT of wavelet coefficient corresponding to Mode 1 (Fig. 4.6 top) is shown in Fig. 4.7. The modal frequencies obtained are 5.5, 18.8, and 34 Hz, and the damping ratio of the first mode is estimated to be 1.9%; however, the damping in Modes 2 and 3 are underestimated at 0.08%, as compared to the values shown in Table 4.1. Wavelet coefficients of all three floor displacements, shown in Fig. 4.6, are used to obtain the
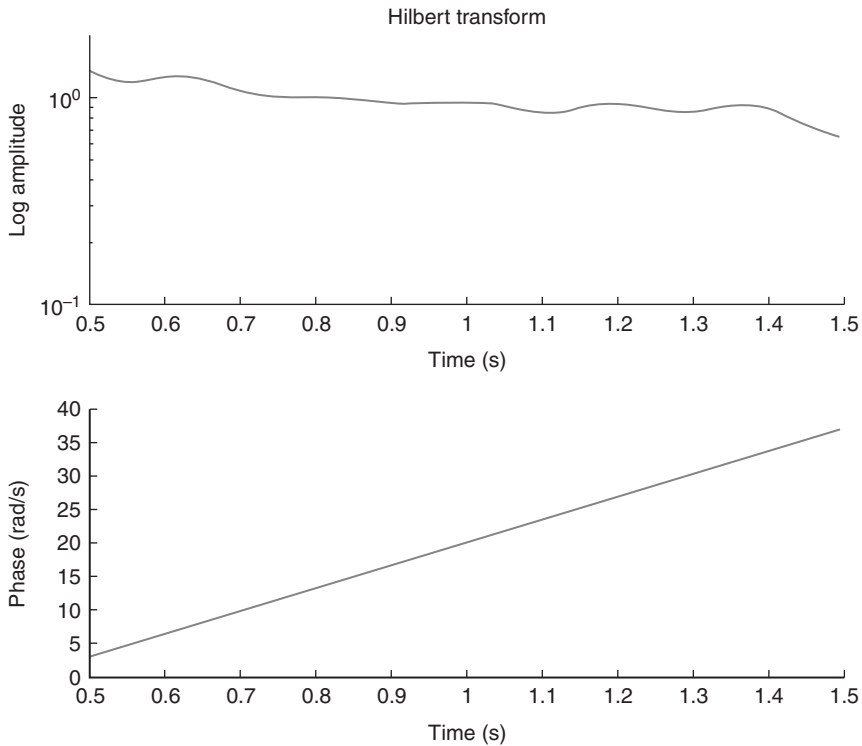
4.5 HT of IMF3 of the third-floor acceleration free-vibration response.

modeshapes. Magnitude/phase information of wavelet coefficients of the three floor displacements at a particular time provides the first mode. The ratio of the wavelet coefficients shown in Fig. 4.6, remain nearly constant as a function of time. Similarly, second and third modes are obtained.

The identified mode shapes (scaled to maximum of 1) are shown in Table 4.4. The analytical modeshapes are shown in Table 4.3. The wavelet results are in agreement with the analytical results.

4.5.4 Validation of wavelet technique using numerical simulation of a 5DOF LTI system

A MDOF model is used to simulate the displacement response and to show the application of the proposed identification methodology. The MDOF system, as shown in Fig. 4.8, is considered. The displacement of the mass relative to the support is denoted by $x_i(t)$. Simulation is carried out for a 5DOF system ($n = 5$). The masses are $m_1 = 300$ kg, $m_2 = 200$ kg, $m_3 = 200$ kg, $m_4 = 250$ kg, and $m_5 = 350$ kg; and the spring stiffnesses are $k_1 = 36\,000$ N/m,
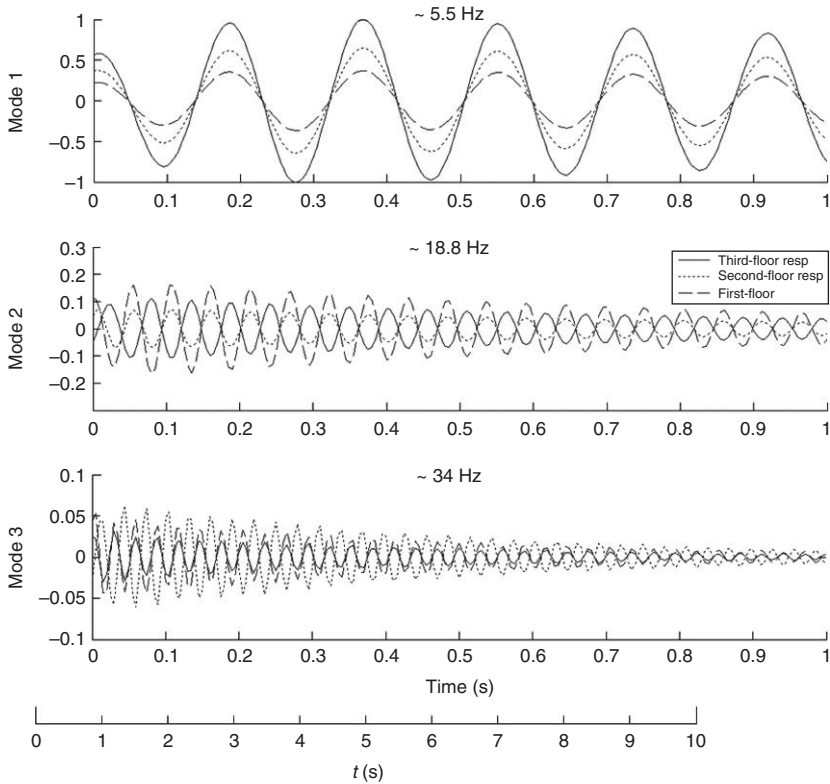
4.6 Wavelet coefficients of the measured free-vibration displacement response of all three floors.

Table 4.3 Analytical mode shapes

	Mode 1	Mode 2	Mode 3
Story-3	1.0000	−0.6416	−0.3946
Story-2	0.6438	0.4299	1.0000
Story-1	0.3648	1.0000	−0.6831

Table 4.4 Mode shapes estimated using wavelets

	Mode 1	Mode 2	Mode 3
Story-3	1.0000	−0.6930	−0.3247
Story-2	0.6437	0.4106	1.0000
Story-1	0.3647	1.0000	−0.7475
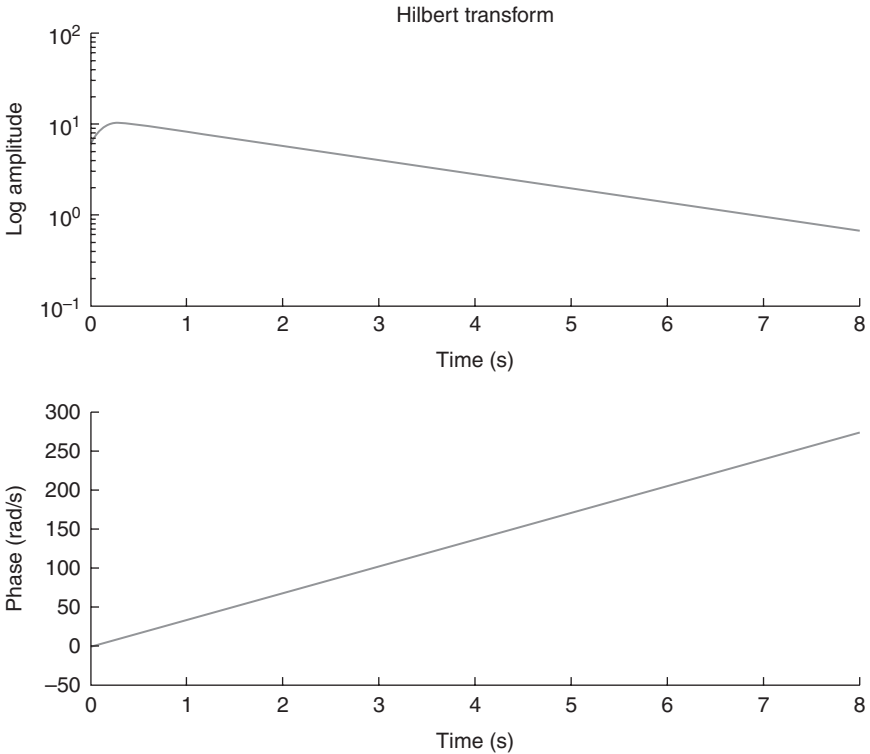
4.7 First mode damping and frequency estimation using wavelet coefficient/HT.
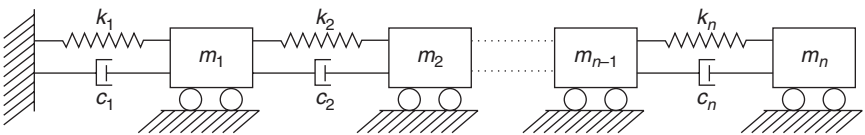
4.8 MDOF system.

k_2 = 24 000 N/m, k_3 = 36 000 N/m, k_4 = 20 000 N/m, and k_5 = 15 000 N/m, respectively. The damping ratio is assumed to be 5% for all modes. The system is subjected to initial displacement of $x_i(0) = 1, i = 1,...,5$ for all the degrees of freedom. Using these, the ambient vibration response is simulated.

The modified L-P wavelet is used to decompose the signals into different frequency levels. Initially, the response energy is calculated for each degree of freedom in frequency bands with $\sigma = 2^{1/4}$ to broadly identify the bands that contain the natural frequencies. These bands are further divided into sub-bands using wavelet packets. Figures 4.6 and 4.10 represent the ratio of

wavelet coefficients of displacements $x_i(t) = 1, i = 2,\ldots,5$ with respect to the wavelet coefficients of displacement $x_1(t)$ over time for the five frequency sub-bands containing the five natural frequencies, respectively. Since the response for different degrees of freedom attain same phase during modal vibration, these ratios are practically constant over time. The natural frequencies are estimated as the central frequency of the corresponding sub-bands, and the corresponding modeshapes are obtained by averaging the ratios. The results for the first two modes are shown in Figs 4.9 and 4.11 using sub-band coding as discussed in Section 4.4. The results are summarized in Table 4.5. Figures 4.10 and 4.12 show the mode shapes estimated using the proposed method and compared with the actual for the first three modes. From Figs 4.10 and 4.12 and Table 4.6, it can be noticed that the modal frequencies, along with other modal parameters, are estimated satisfactorily, which proves the effectiveness of the proposed method.

Although the ratios of wavelet coefficients for higher modes are constant over time, the accuracy in estimation reduces for the higher modes. This is due to the fact that the energy content in bands containing the higher modal frequencies reduces with increase in mode number.
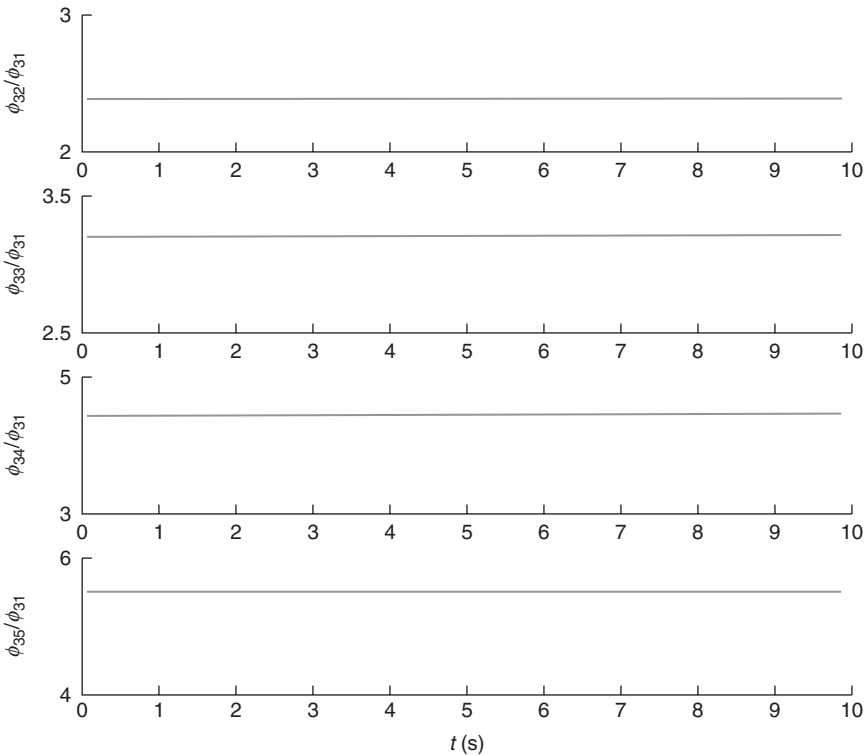
4.9 Modal response at first natural frequency.
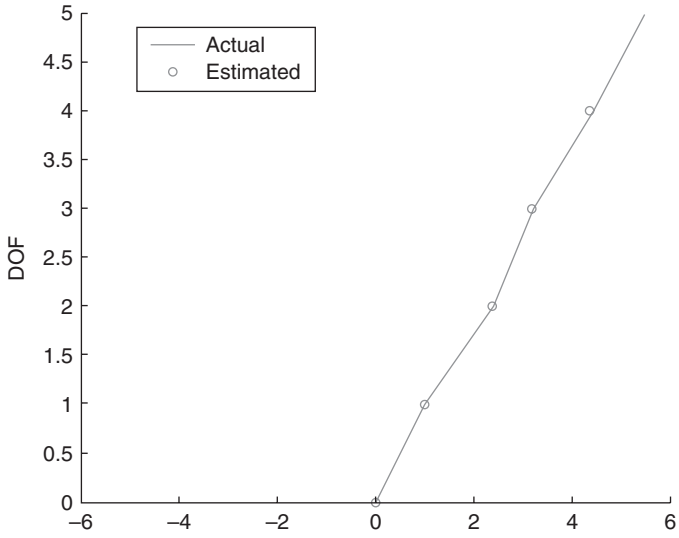
4.10 First mode shape.
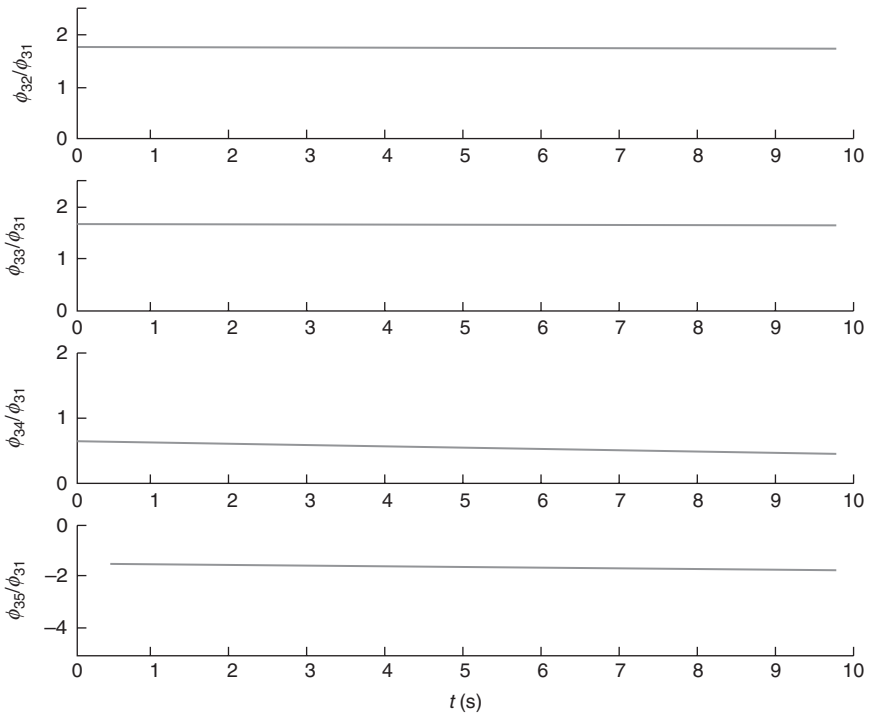
4.11 Modal response at second natural frequency.

Table 4.5 Second mode shape estimated using wavelets

2*Mode	Natural frequency (rad / s)		Damping ratio (%)	
	Actual	Estimated	Actual	Estimated
1	2.84	2.88	0.05	0.04
2	7.69	7.69	0.05	0.03
3	12.35	12.59	0.05	0.02

Table 4.6 Third mode shape estimated using wavelets

3* Mode	Normalized mode shape								
	x_1	x_2	x_2	x_3	x_3	x_4	x_4	x_5	x_5
		Actual	Estimated	Actual	Estimated	Actual	Estimated	Actual	Estimated
1	1.00	2.40	2.37	3.22	3.18	4.45	4.39	5.48	5.39
2	1.00	1.76	1.73	1.69	1.66	0.56	0.69	−1.49	−1.58
3	1.00	0.59	0.89	−0.18	−0.59	−1.30	−1.02	0.51	0.63
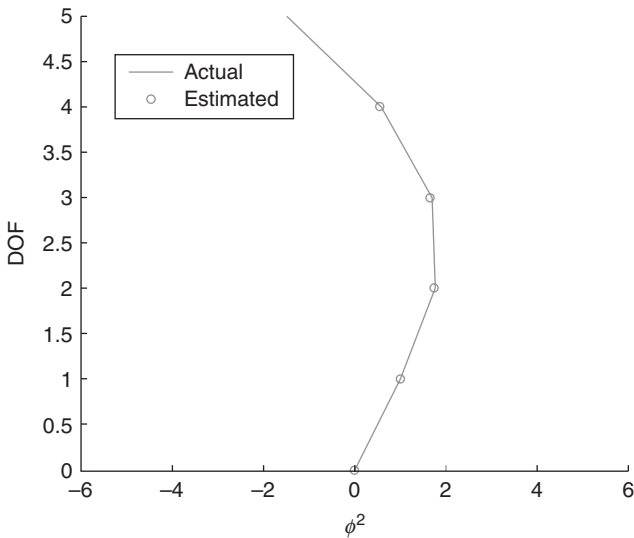
4.12 Second mode shape.

For the 5DOF system, the estimation accuracies start deteriorating from the third mode onwards, and are poorer for the last two modes. This indicates that a higher number of modes and the associated modal properties can be identified with greater accuracy, for systems with a relatively greater number

of degrees of freedom. Also, modal damping ratios can be estimated with reasonable accuracy, with the level of accuracy deteriorating with higher modes. The higher modal damping ratios tend to be underestimated.

4.5.5   Validation of wavelet technique using numerical simulation of a 2DOF LTV system

To demonstrate the application of the tracking methodology, an example of a 2DOF system has been considered. The system considered is a shear-building model. The masses at first and second floors are $m_1 = 10$ unit and $m_2 = 15$ unit, respectively. The floor stiffnesses for the first and second floor are $k_1 = 2500$ unit and $k_2 = 4500$ unit, respectively. These parameters lead to the first and second natural frequencies, of $\omega_1 = 9.04$ rad/s and $\omega_2 = 30.30$ rad/s, respectively. The first and the second mode shapes are $\{\phi_{11}, \phi_{21}\} = \{1, 1.137\}$ and $\{\phi_{12}, \phi_{22}\} = \{1, -0.048\}$, respectively. A band limited white noise excitation has been simulated. The range of frequencies is kept wide enough to cover the frequencies of the system to be identified. The excitation has been digitally simulated at a time step of $\Delta t = 0.0104$ s. The response of the system is simulated with 5% of modal damping. For the frequency-tracking algorithm, a moving window of 400 time steps equal to 4.16 s has been chosen. For the identification of the 2DOF system, the parameters F_1 and σ are taken as 8.25 rad/s and 1.2 respectively. To observe if the proposed method can track a sudden change in the stiffness of an MDOF system and follow the recovery to the original stiffness value(s), the stiffnesses k_1 and k_2 of the 2DOF are changed to 5000 and 5200 unit respectively at an instant of 5.72 s in time. Subsequently, the stiffness are restored to their original value at 12.28 s. During the changed phase the natural frequencies and the mode shapes are changed to $\omega_1 = 11.57$ rad/s; $\omega_2 = 35.11$ rad/s; and $\{\phi_{11}, \phi_{21}\}^t = \{1, 1.157\}$; $\{\phi_{12}, \phi_{22}\} = \{1, -0.048\}$. Figures 4.13 and 4.14 show the tracked first natural frequency and the ratio of first mode shape ϕ_{21}/ϕ_{11}. As expected, there is a time lag in tracking the frequency and mode shape. The change in the frequency is tracked in (three) steps corresponding to the bands of frequencies considered. To investigate if a relatively small change in stiffness can be tracked, a case where the natural frequency of a SDOF representing first mode only changes from 9 rad/s to 9.5 rad/s is considered, and the results for successful tracking are shown in Fig. 4.15 with a window width of 200 sampling points corresponding to a time delay of 2.08 s. For this, the parameters F1 and σ are taken as 8.9 rad/s and 1.02 respectively. This indicates that the minimum change in stiffness that can be tracked is related to the value of σ, and to identify a small change a relatively smaller value will be required.
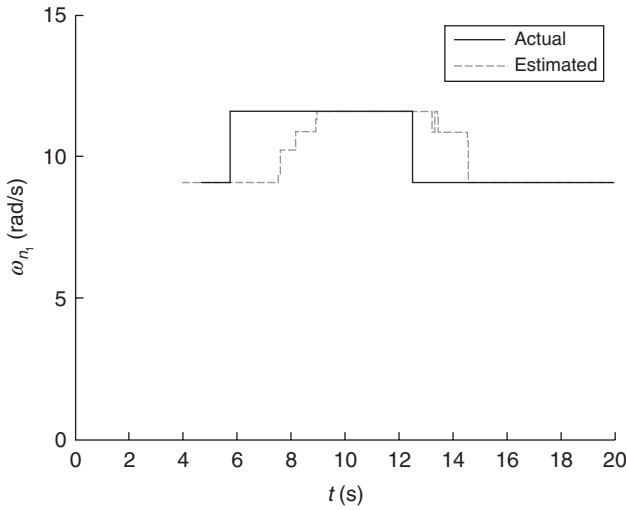
4.13 TV first modal frequency.
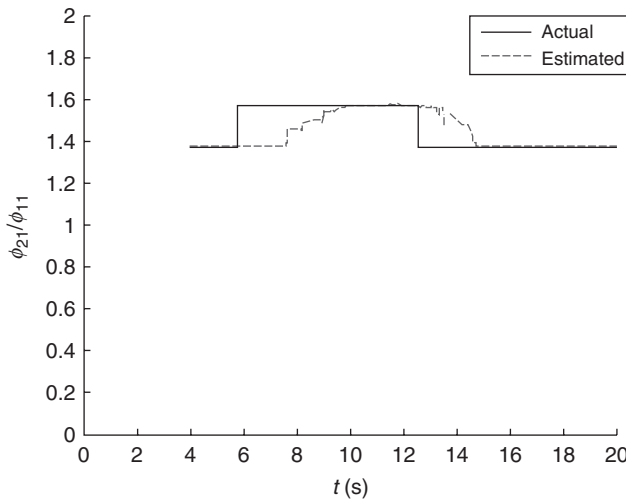
4.14 TV first mode shape

4.5.6 Validation of wavelet and random decrement technique using three-story model test results under white noise excitation: the case of structural damage detection

The three-story scaled model was damaged intentionally to simulate structural deterioration (Nagarajaiah, 2009). The model was subjected to white
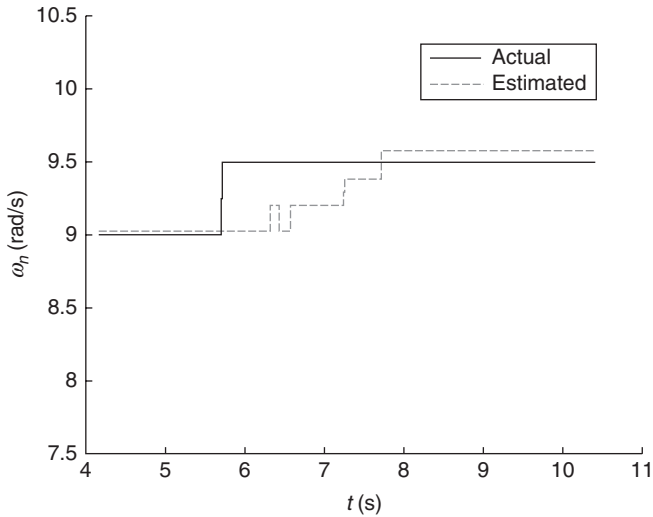
4.15 TV frequency.

noise tests before and after the structural damage. We choose 10 s of the measured acceleration record before damage and 10 s of the measured acceleration after damage. The measured third-floor acceleration response signal is shown in Fig. 4.16, and the corresponding Fourier spectrum is shown in Fig. 4.17. From the Fourier spectrum the first mode frequency evident is ~5.5 Hz, the second mode frequency at ~18.8 Hz, and the third mode frequency at ~34 Hz. The lower first mode frequency after damage is evident. The scalogram of the acceleration response of the third floor is shown in Plate IV in the color section between pages 374 and 375 and relevant scaled wavelet coefficients of the measured third-floor acceleration response are shown in Fig. 4.18. The first two wavelet coefficient time histories in Fig. 4.18 are the most interesting, as they correspond to the first mode frequencies 4.9 Hz (after damage) and 5.5 Hz (before damage). The first two wavelet coefficients in Fig. 4.18 detect the loss of stiffness at 10 s, as evident in the significant change at 10 s in both coefficients. The third and fourth time histories in Fig. 4.18 correspond to the second and third modes, respectively. Even the second mode response reduces at 10 s, although the frequency of the second mode does not change significantly. The third mode response does not indicate any change.

The modal frequency of the second wavelet coefficient in Fig. 4.18b is estimated using linear least squares fit applied to the HT; log amplitude and phase (Equations [4.14], [4.43]–[4.47]) of HT of second wavelet coefficient corresponding to Mode 1 before damage (Fig. 4.18b) is shown in Fig. 4.19. The change in frequency is clearly detected at 10 s: the frequency is ~5.5 Hz prior to damage and ~4.9 Hz after damage.
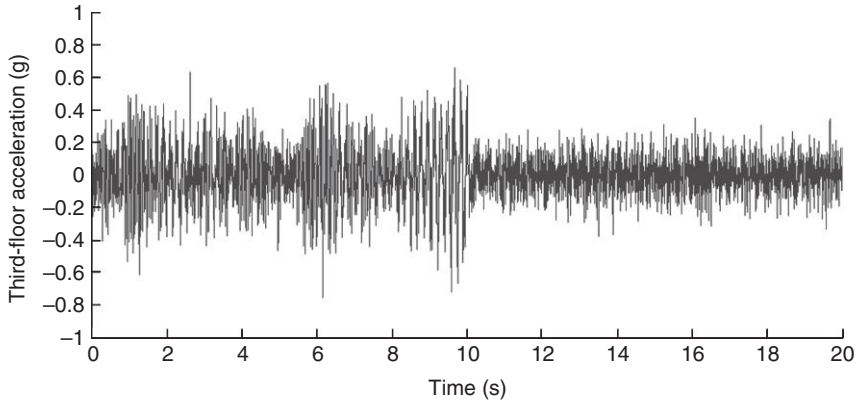
4.16 Measured third-floor acceleration response to white noise excitation.
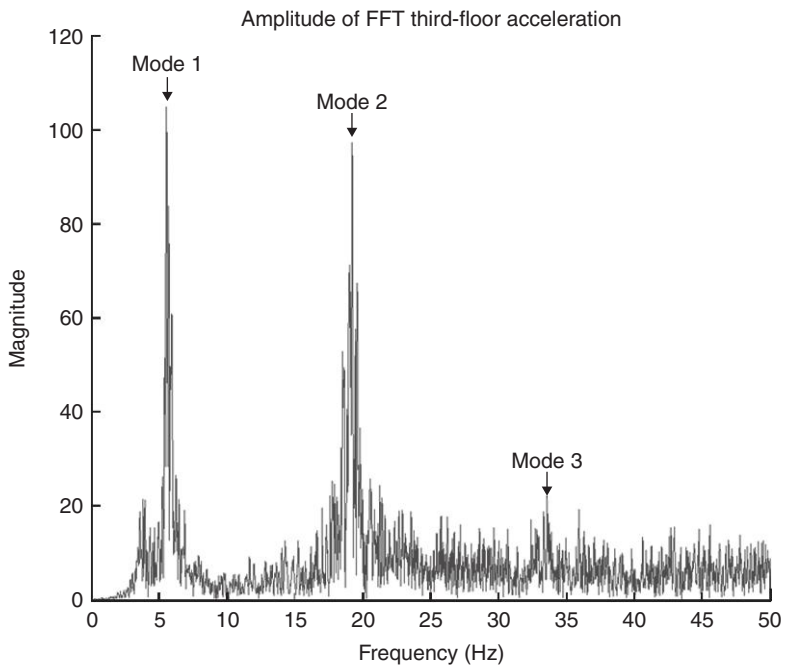
4.17 Fourier spectrum of third-floor acceleration.

The first two wavelet coefficient time histories in Fig. 4.18 are processed further to extract the free-vibration response using the random decrement technique. The third-floor acceleration free-vibration time history obtained from the random decrement technique before damage is shown in Fig. 4.20;
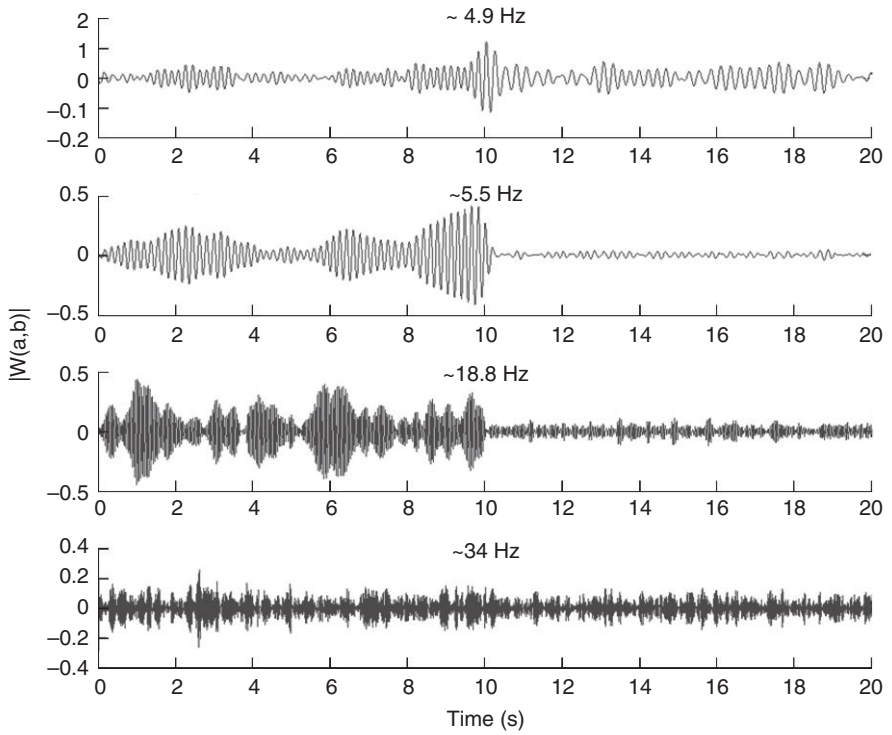
4.18 Wavelet coefficients of the measured third-floor acceleration response.
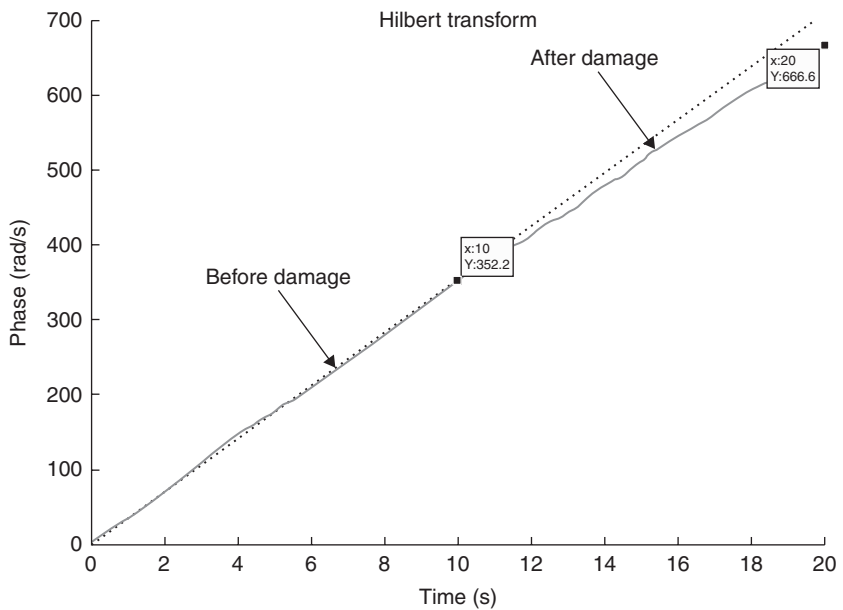
4.19 First mode frequency estimation using wavelet coefficient/HT (note the change in frequency before and after damage at 10 s).
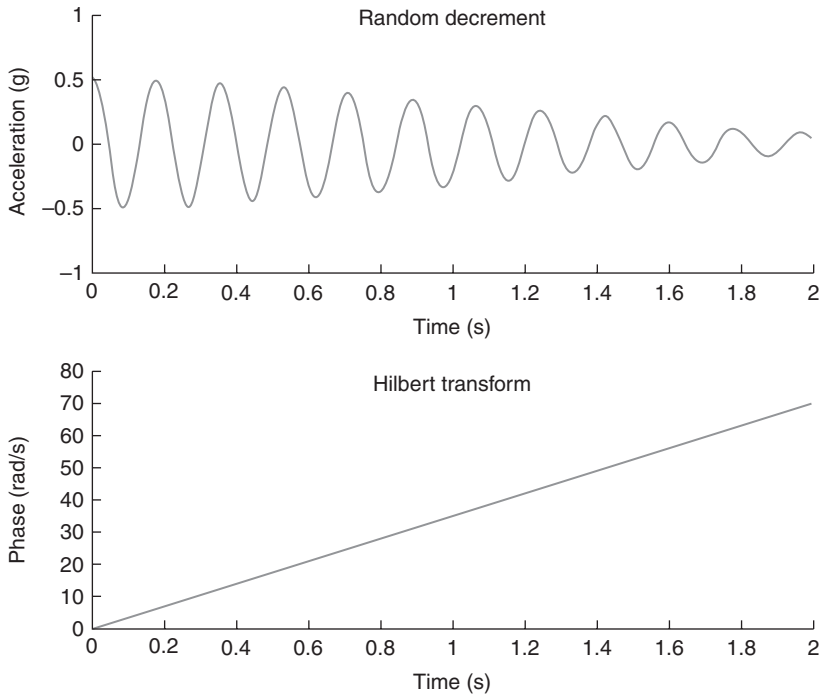
4.20 First mode frequency estimation before damage using random decrement/HT.

also shown is the frequency estimation using HT – the estimated first mode frequency before damage is ~5.5 Hz. The third-floor acceleration free-vibration time history obtained from the random decrement technique after damage is shown in Fig. 4.21; also shown is the frequency estimation using HT – the estimated first mode frequency after damage is ~4.9 Hz. Damping ratios and modeshapes can be obtained as described in Section 4.4 (not shown, due to space limitations).

4.5.7 Three-story model test results under white noise excitation: STFT and EMD for structural damage detection

We process the third-floor acceleration response to white noise excitation using STFT and EMD. The spectrogram is shown in Plate V in the color section between pages 374 and 375. The spectrogram detects the change in frequency from 5.5 to 4.9 Hz at 10 s. However, the fixed TF resolution is a limitation that prevents robust detection when compared to the variable resolution of wavelets that enables more robust detection. In addition, estimation of frequencies,
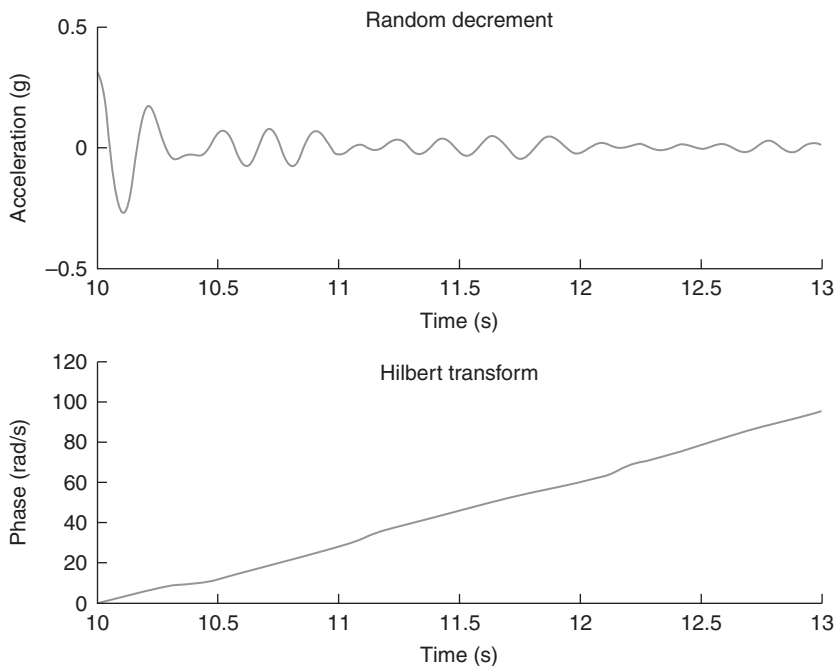
4.21 First mode frequency estimation after damage using random decrement/HT.

damping ratios, and mode shapes would require further processing using band-pass filtering and HT approach, as described earlier in Section 4.3.

We process the third-floor acceleration response to white noise excitation using EMD. The IMFs are shown in Fig. 4.22. The IMFs do detect change at 10 s – particularly the IMF3 for the first mode before damage at 5.5 Hz; however, the detection is not as robust as in the case of the wavelet coefficients shown in Fig. 4.18.

The recently developed output only WT-ICA method also successfully identified damage instant, damage location, and the TV modes of this model due to damage. For details, refer to Yang and Nagarajaiah (2012).

4.6   Conclusion

The effectiveness of the developed TF algorithms for output only modal identification of MDOF LTI and LTV systems has been demonstrated by simulated and experimental results. The algorithms presented demonstrate the powerful capabilities of TF methods for output only modal identification and ease of implementation.

The STFT, EMD, and wavelet HT algorithms applied to MDOF LTI and LTV systems offer different advantages and limitations that can be summarized as follows.
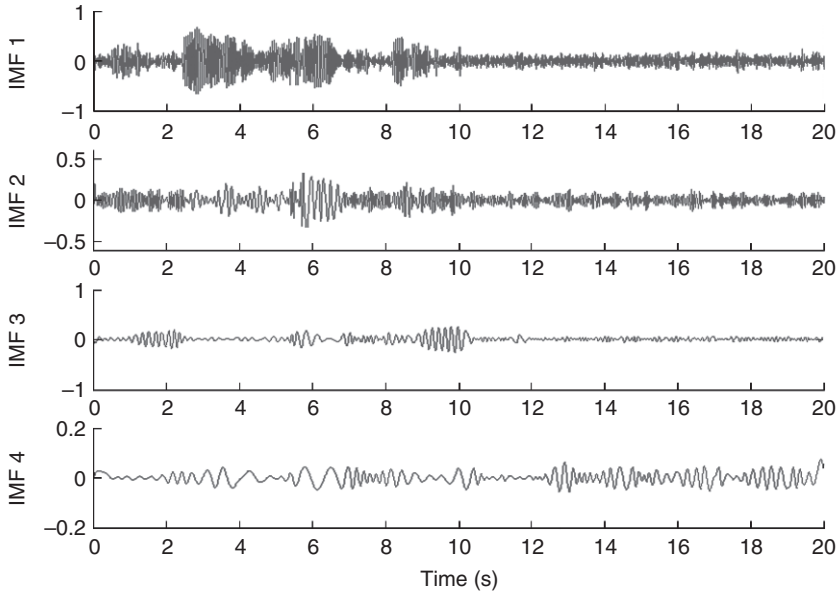
4.22 IMFs of the third-floor acceleration response to white noise excitation.

1. STFT based identification technique presented can detect the modal frequencies of LTI systems and their time localization very well; however, further processing using band-pass filtering is essential to obtain frequencies, damping ratios, and mode shapes. STFT can also detect changes in modal frequency of LTV systems due to structural damage. However, the fixed TF resolution is a limitation that prevents robust detection when compared to the variable resolution of wavelets that enables more robust detection.

2. The EMD-based identification technique presented is capable of decomposing the free-vibration or forced-vibration output signal into its individual modal components – represented by individual IMFs. Frequencies, damping ratios, and mode shapes of LTI systems can be obtained using the IMFs and the HT approach. In case of ambient response, the random decrement technique can be used to obtain the free-vibration response, followed by the application of EMD/HT for modal identification. The EMD technique is capable of detecting changes in frequency of LTV systems due to structural damage; however, the detection may not be as robust as wavelets.

3. The wavelet-based identification technique presented is capable of extracting the modal components represented by wavelet coefficients obtained from the free-vibration or forced-vibration output response signal. Frequencies, damping ratios, and mode shapes of LTI and LTV systems can

be obtained using wavelet coefficients and the HT approach. In the case of ambient response, the random decrement technique can be used to obtain the free-vibration response, followed by the application of wavelet/HT for modal identification. The wavelet technique is very effective in detecting changes in the frequency of LTV systems due to structural damage. The wavelet technique can also detect closely spaced modal frequencies and real-time changes in frequencies and mode shapes of LTV systems.

4.7   Acknowledgments

4.8   References

Abazarsa, F., Ghahari, S.F., Nateghi, F. and Taciroglu, E. (2012) Response-only modal identification of structures using limited sensors. *Structural Control and Health Monitoring*, **20**, 987–1006. DOI: 10.1002/stc.1513 .

Addison, P., Watson, J.N. and Feng, T. (2002) Low-oscillation complex wavelets, *Journal of Sound and Vibration,* **254**(4), 733–762.

Agneni, A. and Balis-Cerma, L. (1989) Damping measurements from truncated signals via the Hilbert transform, *Mechanical Systems and Signal Processing,* **3**(1), 1–3.

Al-Khalidy, A., Noori, M., Hou, Z., Carmona, R., Yamamoto, S., Masuda, A. and Sone, A. (1997) A study of health monitoring systems of linear structures using wavelet analysis, *Approx Methods in the Design and Analysis of Pressure Vessels and Piping Components, ASME PVP,* **347**, 49–58.

Antoni, J. (2005) Blind separation of vibration components: principles and demonstrations, *Mechanical Systems and Signal Processing,* **19**(6), 1166–1180.

Basu, B. and Gupta, V.K. (1997) Non-stationary seismic response of MDOF systems by wavelet modelling of non-stationary processes, *Earthquake Engineering Structural Dynamics*, **26**, 1243–1258.

Basu, B. and Gupta, V.K. (1998) Seismic response of SDOF systems by wavelet modelling of nonstationary processes, *Journal of Engineering Mechaics, ASCE,* **124**(10), 1142–1150.

Basu, B. and Gupta, V.K. (1999a) On equivalent linearization using wavelet transform, *Journal of Vibration and Acoustics, ASME,* **121**(4), 429–432.

Basu, B. and Gupta, V.K. (1999b) Wavelet based analysis of the non-stationary response of a slipping foundation, *Journal of Sound and Vibration,* **222**(4), 547–563.

Basu, B. (2005) Identification of stiffness degradation in structures using wavelet analysis, *Construction and Building Materials*, **19**, 713–721.

Basu, B. (2007) Assessment of structural integrity via wavelet based time-frequency analysis of vibration signals, *International Journal of Materials and Structural Integrity,* **1**, 238–258.

Basu, B. and Nagarajaiah, S. (2008) A wavelet-based time-varying adaptive LQR algorithm for structural control, *Engineering Structures,* **30**(9), 2470–2477.

Basu, B., Nagarajaiah, S. and Chakraborty, A. (2008) 'Online identification of linear time-varying stiffness of structural systems by wavelet analysis,' *International Journal of Structural Health Monitoring,* **7**(1), 21–36.

Belouchrani, A., Abed-Meraim, A.K., Cardoso, J.F. and Moulines, E. (1997) A blind source separation technique using second-order statistics, *IEEE Transactions on Signal Processing,* **45**, 434–444.

Chakraborty, A., Basu, B. and Mitra, M. (2006) Identification of modal parameters of a MDOF system by modified L-P wavelet packets, *Journal of Sound and Vibration,* **295**(3–5), 827–837.

Chang C.C and Chen L.W (2003) Vibration damage detection of a Timoshenko beam by spatial wavelet based approach, *Applied Acoustics*, **64**, 1217–1240.

Chatterjee, P. and Basu, B. (2001) Nonstationary seismic response of tank with soil interaction by wavelets, *Earthquake Engineering Structural Dynamics,* **30**, 1419–1437.

Chatterjee, P. and Basu, B. (2006) Nonstationary seismic response of a tank on a bilinear hysteretic soil using wavelet transform, *Probabilistic Engineering Mechanics,* **1**, 54–63.

Chen, B. and Nagarajaiah, S. (2007) Linear matrix inequality based robust fault detection and isolation using the eigenstructure assignment method, *Journal of Guidance Control and Dynamics, AIAA,* **30**(6), 1831–1835.

Chen, B. and Nagarajaiah, S. (2008a) H_- / H_∞ structural damage detection filter design using iterative LMI approach, *Smart Materials and Structures,* **17**(3), Art. no. 035019.

Chen, B. and Nagarajaiah, S. (2008b) Structural damage detection using decentralized controller design method, *Smart Structures and Systems,* **4**(6), 779–794.

Chen, S-L., Liu, J-J. and Lai, H-C. (2009) Wavelet analysis for identification of damping ratios and natural frequencies, *Journal of Sound and Vibration,* **323**, 130–147.

Cohen, L. (1995) *Time-Frequency Analysis* (1st edn). Prentice Hall: New Jersey.

Comon, P. and Jutten, C. (2010) *Handbook of Blind Source Separation*: *Independent Component Analysis and Applications*. Academic Press, New York.

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

Dharap, P., Koh, B.H. and Nagarajaiah, S. (2006) Structural health monitoring using ARMarkov observers, *Journal of Intelligent Material Systems and Structures,* **17**(6), 469–481.

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

Feldman, M. (1994a) Non-linear system vibration analysis using Hilbert transform – I. Free vibration analysis method 'FREEVIB', *Mechanical Systems and Signal Processing,* **8**(2), 119–127.

Feldman, M. (1994b) Non-linear system vibration analysis using Hilbert transform-II. Forced vibration analysis method 'forcevib', *Mechanical Systems and Signal Processing,* **8**(3), 309–318.

Gangadharan, R., Roy Mahapatra, D., Gopalakrishnan, S., Murthy, C.R.L. and Bhat, M.R. (2009) On the sensitivity of elastic waves due to structural damages: time-frequency based indexing method, *Journal of Sound and Vibration,* **320**, 915–941.

Gentile, A. and Messina, A. (2003) On the continuous wavelet transforms applied to discrete vibrational data for detecting open cracks in damaged beams, *International Journal of Solids and Structures,* **40**, 295–315.

Ghanem, R. and Romeo, F. (2000) A wavelet based approach for the identification of linear time-varying dynamical systems, *Journal of Sound and Vibration*, **4**, 555–576.

Ghanem, R. and Romeo, F. (2001) A wavelet based approach for model and parameter identification of nonlinear systems, *International Journal of Nonlinear Mechanics,* **5**, 835–859.

Goggins, J., Broderck, B.M., Basu, B. and Elghazouli, A.Y. (2007) Investigation of seismic response of braced frames using wavelet analysis, *Structural Control and Health Monitoring,* **14**(4), 627–648.

Gonzalez, A., Obrien, E.J. and McGetrick, P.J. (2012) Identification of damping in a bridge using a moving instrumented vehicle, *Journal of Sound and Vibration,* **331**, 4115–4131.

Gribonval, R. and Lesage, S. (2006) A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges. *Proceedings of European Symposium on Artificial Neural Networks, Bruges*, 323–330.

Gul, M. and Catbas, F.N. (2008) Ambient vibration data analysis for structural identification and global condition assessment, *Journal of Engineering Mechanics ASCE,* **134**(8), 650–662.

Gurley, K. and Kareem, A. (1999) Applications of wavelet transforms in earthquake, wind and ocean engineering, *Engineering Structures,* **21**(2), 149–167.

Hazra B., Roffel A.J., Narasimhan S. and Pandey, M.D. (2010a) Modified cross-correlation method for the blind identification of structures, *ASCE Journal of Engineering Mechanics,* **136**, 889–897.

Hazra, B. and Narasimhan, S. (2010b) Wavelet-based blind identification of the UCLA Factor building using ambient and earthquake responses, *Smart Materials and Structures,* **19**, 025005.

Hazra, B., Sadhu, A., Roffel, A.J., Paquet, P.E. and Narasimhan, S. (2012) Under-determined blind identification of structures using the modified cross-correlation method, *ASCE Journal of Engineering Mechanics,* **138**, 327–337.

Hou Z, Nouri, M and Amand R. St (2000) Wavelet based approach for structural damage detection, *ASCE Journal of Engineering Mechanics,* **126**(7), 677–683.

Huang, E.N., Shen, Z., Long, R.S., Wu, C.M., Shih, H.H., Zheng, Q., Yen, N., Tung, C.C. and Liu, H.H. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Royal Soc. London*, **454**, 903–995.

Hyvärinen, A. and Oja, E. (2000) Independent component analysis: algorithms and applications, *Neural Networks,* **13**, 411–430.

Ibrahim, S.R. (1977) Random decrement technique for modal identification of structures, *Journal of Spacecraft and Rockets,* **14**(11), 696–700.

Jiang, X. and Mahadevan, S. (2008) Bayesian wavelet method for multivariate model assessment of dynamics systems, *Journal of Sound and Vibration,* **312**, 694–712.

Kerschen, G., Poncelet, F. and Golinval J.-C. (2007) Physical interpretation of independent component analysis in structural dynamics, *Mechanical Systems and Signal Processing,* **21**, 1561–1575.

Kijewski-Correa, T. and Kareem, A. (2007) Performance of wavelet transform and empirical mode decomposition in extracting signals embedded in noise, *Journal of Engineering Mechanics, ASCE,* **133**(7), 849–852.

Kijewski-Correa, T. and Kareem, A. (2006) Efficacy of Hilbert and wavelet transforms for time-frequency analysis, *Journal of Engineering Mechanics, ASCE,* **132**(10), 1037–1049.

Kijewski, T. and Kareem, A. (2003) Wavelet transforms for system identification in civil engineering, *Computer-Aided Civil and Infrastructure Engineering,* **18**(5), 339–355.

Kitada, Y. (1998) Identification of nonlinear structural dynamic system using wavelet, *Journal of Engineering Mechanics, ASCE,* **124**(10), 1059–1066.

Koh, B.H., Dharap, P., Nagarajaiah, S. and Phan, M.Q. (2005a) Real time structural damage monitoring by input error function, *AIAA Journal,* **43**(8), 1808–1814.

Koh, B.H., Li, Z., Dharap, P., Nagarajaiah, S. and Phan, M.Q. (2005b) Actuator failure detection through interaction matrix formulation, *Journal of Guidance Control and Dynamics, AIAA,* **28**(5), 895–901.

Koh, B.H., Nagarajaiah, S. and Phan, M.Q. (2008) Direct identification of structural damage through Kronecker product method, *Journal of Mechanical Science and Technology,* **22**(01), 103–112.

Kyprianou, A. and Staszewski, W.J. (1999) On the cross wavelet analysis of the Duffing oscillator, *Journal of Sound and Vibration,* **228**(1), 119–210.

Lardies, J. and Gouttebroze, S. (2000) Identification of modal parameters using the wavelet transform, *International Journal of Mechanical Science,* **44**, 2263–2283.

Le, T.-P and Paultre, P. (2012) Modal identification based on continuous wavelet transform and ambient excitation tests, *Journal of Sound and Vibration*, **331**, 2023–2037.

Li, Z., Koh, B.H. and Nagarajaiah, S. (2007) Detecting sensor failure via decoupled error function and inverse input-output model, *Journal of Engineering Mechanics*, *ASCE*, **133**(11), 1222–1228.

Liew, K.M. and Wang, Q. (1998) Application of wavelet theory for crack identification in structures, *American Society of Civil Engineering, Journal of Engineering Mechanics,* **124**(2), 152–157.

Loutridis, S, Douka, E and Trochidis, A., (2004) Crack identification in double cracked beams using wavelet analysis, *Journal of Sound and Vibration*, **277**, 1025–1039.

Lynch, J.P., Sundararajan, A., Law, K.H., Kiremidjian, A.S. and Carryer, E. (2004) Embedding damage detection algorithms in a wireless sensing unit for attainment of operational power efficiency, *Smart Materials and Structures,* **13**, 800–810.

McNeill, S.I. and Zimmerman D.C. (2008) A framework for blind identification using joint approximate diagonalization, *Mechanical Systems and Signal Processing,* **22**, 1526–1548.

Melhem, H and Kim, H. (2003) Damage detection in concrete by Fourier and Wavelet analysis, *American Society of Civil Engineering, Journal of Engineering Mechanics,* **129**(5), 571–577.

Milanese, A., Marzocca, P., Nichols, J.M., Seaver, M. and Trickey, S.T. (2008) Modeling and detection of joint loosening using output-only broad-band vibration data, *Structural Health Monitoring,* **7**(4), 309–328.

Naldi, G. and Venini, P. (1997) Wavelet analysis of structures: statics, dynamics and damage identification, *Meccanica*, **32**, 223–230.

Nagarajaiah, S., Vardarajan N. and Sahasrabudhe, S. (1999) Variable stiffness and instantaneous frequency. *Proc. Structures Congress, ASCE, New Orleans,* 858–861.

Nagarajaiah, S. and Varadarajan, N. (2001) Semi-active control of smart tuned mass damper using empirical mode decomposition and Hilbert transform algorithm. *Proc. ICOSSAR Newport Beach, CA,* June, CD ROM 2001.

Nagarajaiah, S. and Dharap, P. (2003) Reduced order observer based identification of base isolated buildings, *Earthquake Engineering and Engineering Vibration,* **2**(2), 237–244.

Nagarajaiah, S. and Li, Z. (2004) Time segmented least squares identification of base isolated buildings, *Soil Dynamics and Earthquake Engineering Journal,* **24**, 577–586.

Nagarajaiah, S. and Varadarajan, N. (2005) Semi-active control of wind excited building with variable stiffness TMD using short time Fourier transform, *Engineering Structures,* **27**(3), 431–441.

Nagarajaiah, S. and Sonmez, E. (2007) Structures of semiactive variable stiffness multiple tuned mass dampers under harmonic forces, *Journal of Structural Engineering, ASCE,* **133**(1), 67–77.

Nagarajaiah, S. (2009) Adaptive passive, semiactive, smart tuned mass dampers: identification and control using empirical mode decomposition, Hilbert transform, and short-term Fourier transform, *Structural Control and Health Monitoring,* DOI: 10.1002 stc.349.

Narasimhan, S. and Nagarajaiah, S. (2005) STFT algorithm for semiactive control of base isolated buildings with variable stiffness isolation systems subjected to near fault earthquakes, *Engineering Structures,* **27**(4), 514–523.

Newland, D.E. (1993) *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, Longman, U.K.

Newland, D.E. (1994a) Wavelet analysis of vibration, Part1: Theory, *Transactions of ASME Journal of Vibration and Acoustics*, **116**, 409–416.

Newland, D.E. (1994b) Wavelet analysis of vibration, Part2: Wavelet maps, *Transactions of ASME Journal of Vibration and Acoustics*, **116**, 417–425.

Okafor, A.C. and Dutta, A. (2000) Structural damage detection in beams by wavelet transforms, *Smart Materials and Structures*, **9**, 906–917.

Pakrashi, V., Basu, B. and O'Connor, A. (2007) Structural damage detection and calibration using wavelet-kurtosis technique, *Engineering Structures*, **29**(9), 2097–2108.

Patsias, S., Staszewski, W.J. and Tomlinson, G.R. (2002) Image sequences and wavelets for vibration analysis. Part II – extraction of modal damping and mode-shapes, *Proceedings of the Institution of Mechanical Engineers, Part C, Journal of Mechanical Engineering Science*, **216**(9), 901–912.

Patsias, S. and Staszewski, W.J. (2002) Damage detection using optical measurements and wavelets, *Structural Health Monitoring*, **1**(1), 5–22 .

Piombo, B.A.D., Fasana, A., Marchesiello, S. and Ruzzene, M. (2000) Modelling and identification of the dynamic response of a supported bridge, *Mechanical Systems and Signal Processing*, **14**(1), 75–89.

Poncelet, F., Kerschen, G., Golinval, J.-C. and Verhelst, D. (2007) Output-only modal analysis using blind source separation techniques, *Mechanical Systems and Signal Processing,* **21**, 2335–2358.

Rilling, G., Flandrin, P. and Goncalves, P. (2003) On empirical mode decomposition and its algorithms, *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, NSIP-03, Grado I.

Robertson, A.N., Park, K.C. and Alvin, K.F. (1998) Extraction of impulse response data via wavelet transform for structural system identification, *Journal of Vibration and Acoustics, ASME,* **120**, 252–260.

Robertson, A. and Basu, B. (2008) *Wavelet Analysis*, editors Boller, C., Chang, F.-K. and Fujino, Y., *Encyclopaedia on Structural Health Monitoring*, Wiley, New York.

Rucka, M. and Wilde, K. (2006) Crack identification using wavelets on experimental static deflection profiles, *Engineering Structures,* **28**, 279–288.

Ruzzene, M., Fasana, A., Garibaldi, L. and Piombo, B.A.D. (1997) Natural frequencies and damping identification using wavelet transform: application to real data, *Mechanical Systems and Signal Processing,* **11**, 207–218 .

Sadhu, A., Hazra, B., Narasimhan, S. and Pandey, M.D. (2011) Decentralized modal identification using sparse blind source separation, *Smart Materials and Structures,* **20**, 125009.

Sohn, H. and Farrar, C.R. (2001) Damage diagnosis using time series analysis of vibration signals, *Smart Materials and Structures,* **10**, 446–451.

Spanos, P.D., Failla, G., Santini, A. and Pappatico, M. (2006) Damage detection in Euler-Bernoulli beam via spatial wavelet analysis, *Structural Control and Health Monitoring,* **13**, 472–487.

Spencer, B. and Nagarajaiah, S. (2003) State of the art of structural control, *Journal of Structural Engineering, ASCE,* **129**(7), 845–856.

Spiridonakos, M.D., Poulimenos, A.G. and Fassois, S.D. (2010) Output-only identification and dynamic analysis of time-varying mechanical structures under random excitation: a comparative assessment of parametric methods, *Journal of Sound and Vibration,* **329**, 768–785.

Staszewski, W.J. and Tomlinson, G.R. (1994) Application of the wavelet transform to fault detection in a spur gear, *Mechanical Systems and Signal Processing,* **8**, 289–307.

Staszewski, W.J. (1997) Identification of damping in MDOF systems using time scale decomposition, *Journal of Sound and Vibration,* **203**, 283–305.

Staszewski, W.J. (1998a) Identification of non-linear systems using multi-scale ridges and skeletons of the wavelet transform, *Journal of Sound and Vibration,* **214**(4), 639–658.

Staszewski, W.J., Biemans, C., Boller,C. and Tomlinson, G.R. (1998) Damage detection using wavelet-based statistical analysis, *Proc of ISMA23*, **I**, 59–66.

Staszewski, W.J. (1998b) Structural and mechanical damage detection using wavelets, *The Shock and Vibration Digest,* **30**(4), 457–472.

Staszewski, W.J. and Robertson, A.N. (2007) Time-frequency and time-scale analysis for structural health monitoring, *Philosophical Transactions of the Royal Society, Part A,* **365**(1851), 449–477.

Thrane, N. (1984) *The Hilbert Transform* Bruel & Kjaer Technical Review, **3**.

Tomlinson, G.R. (1987) Developments in the use of the Hilbert transform for detecting and quantifying non-linearity associated with frequency response functions, *Mechanical Systems and Signal Processing,* **1**(2), 151–171.

Varadarajan, N. and Nagarajaiah, S. (2004) Wind response control of building with variable stiffness tuned mass damper using EMD/HT, *Journal of Engineering Mechanics, ASCE,* **130**(4), 451–458.

Wang, W.J. and McFadden, P.D. (1996) Application of wavelets to gearbox vibration signals for fault detection, *Journal of Sound and Vibration,* **192**, 927–939.

Wang, Q. and Deng, X. (1999) Damage detection with spatial wavelets, *International Journal of Solids and Structures,* **36**, 3443–3468.

Worden, K. and Tomlinson, G.R. (2001) *Nonlinearity in Structural Dynamics*, Institute of Physics Publishing, Bristol and Philadelphia.

Worden, K., Manson, G. and Fieller, N.R.J. (2000) Damage detection using outlier analysis, *Journal of Sound and Vibration,* **229**, 647–667.

Yang, J.N., Lei, Y., Pan, S. and Huang, N. (2003) System identification of linear structures based on Hilbert-Haung spectral analysis. Part I: normal modes, *Earthquake Engineering & Structural Dynamics,* **32**, 1443–1467.

Yang, J.N., Lei, Y., Lin, S. and Huang, N. (2004) Hilbert-Haung based apporach for structural damage detection, *Journal of Engineering Mechanics,* **130**(1), 85–95.

Yang, Y. and Nagarajaiah, S. (2013a) Time-frequency blind source separation using independent component analysis for output-only modal identification of highly-damped structures, *ASCE Journal of Structural Engineering*, **139**(10), 1780–1793.

Yang, Y. and Nagarajaiah, S. (2012) Blind identification of damage in time-varying system using independent component analysis with wavelet transform, *Mechanical Systems and Signal Processing*, in press, DOI: 10.1016/j.ymssp.2012.08.029.

Yang, Y. and Nagarajaiah S. (2013b) Output-only modal identification with limited sensors using sparse component analysis, *Journal of Sound and Vibration*, **332**, 4741–4765.

Zeldin, B.A. and Spanos P.D. (1998) Spectral identification of nonlinear structural systems, *Journal of Engineering Mechanics, ASCE,* **124**(7), 728–733.

Zheng, H. and Mita, A. (2012) Localized damage detection of structures subject to multiple ambient excitations using two distance measures for autoregressive models, *Structural Health Monitoring,* **8**(3), 207–222.

Zhou, W. and Chelidze, D. (2007) Blind source separation based vibration mode identification, *Mechanical Systems and Signal Processing,* **21**, 3072–3087.

5

Prognosis and life-cycle assessment based on SHM information

D. M. FRANGOPOL, Lehigh University, USA and S. KIM, Korea Hydro & Nuclear Power Co., Ltd, Republic of Korea

Abstract: Efficient prognosis based on structural health monitoring (SHM) information can improve the accuracy associated with structural performance assessment and prediction, and lead to more rational life-cycle management of civil infrastructure systems. This chapter deals with the statistical and probabilistic aspects for efficient prognosis using SHM data. The concepts of the statistics of extremes and decision analysis are employed for cost-effective monitoring planning considering availability of monitoring data and performance prediction error. In order to quantify this error, the loss function is used. Furthermore, the general concept of life-cycle evaluation, the possible effects of SHM on structural performance and service life prediction, and a practical approach to integrating SHM into life-cycle performance analysis of deteriorating civil infrastructures are presented.

Key words: prognosis, structural health monitoring (SHM), statistics of extremes, decision analysis, life-cycle analysis, structural performance, uncertainty, cost, optimization.

5.1 Introduction

Uncertainty in estimation of the structural safety and prediction of the service life of civil infrastructure systems is unavoidable. Over the last several decades, various types of nondestructive testing and structural health monitoring (SHM) methods have been employed for assisting managers of aging structures with the continuation of safe and economic operation (Brownjohn, 2007). The application of SHM has great potential in rational damage prognosis by reducing the uncertainty (Farrar and Lieven, 2007). This reduction can lead to preventing unexpected failure of a structure, assessing and predicting structural performance more reliably, and applying the appropriate maintenance at the optimum time (Frangopol and Messervey, 2007, 2009). However, if the application of SHM is not cost-effective and the

use of SHM data is not efficient, it will be difficult for structure managers to justify adopting SHM (Frangopol, 2011). Therefore, studies regarding cost-effective monitoring, planning, and efficient prognosis approaches to using SHM data are necessary.

The statistical and probabilistic concepts, along with a probability-based decision-making approach, can provide efficient tools for treating epistemic uncertainty in rational prognosis using SHM information. This chapter presents such concepts and approach. Exceedance probability, availability of monitoring data for performance prediction, and a general concept of the loss function to quantify the error in structural performance assessment and prediction are introduced. Furthermore, a decision-making approach based on availability of monitoring data is provided for cost-effective planning of monitoring. Finally, a general concept of life-cycle evaluation, possible effects of SHM on life-cycle assessment, and a practical approach for the efficient integration of SHM into life-cycle analysis of deteriorating civil infrastructure systems are presented. The concepts and approaches introduced in this chapter can lead to efficient prognosis, and cost-effective and more reliable maintenance management of civil infrastructure.

5.2   Statistical and probabilistic aspects for efficient prognosis

Significant accomplishments focusing on damage diagnosis using advanced SHM technologies have been made over the last several decades (Chang *et al.*, 2003). Despite this progress in damage diagnosis techniques, infrastructure managers are under continuous challenge to accurately predict the remaining service life and structural performance using SHM information (Farrar and Worden, 2007; Glaser *et al.*, 2007). Several studies on damage prognosis using SHM information have been conducted only recently. Farrar and Lieven (2007) provided a comprehensive review of recent advances and future challenges in damage prognosis. Approaches to predict the fatigue life of existing structures using monitoring data were presented by Kulkarni and Achenbach (2008), Kwon and Frangopol (2010), and Zárate *et al.* (2012), among others. Furthermore, Okasha and Frangopol (2012) and Orcesi and Frangopol (2011) developed an integrated life-cycle framework for maintenance-SHM-management of civil infrastructures. However, damage prognosis for more practical applications still requires extensive and multi-disciplinary research efforts (Farrar and Worden, 2007).

SHM for the structural response to external loadings requires a very sizeable storage system if a large-scale continuous SHM is used and all data are recorded (Li and Zhang, 2006). The size of monitored data depends on

the monitoring frequency and number of installed sensors. The statistics of extremes is well suited for efficient data management and structural performance assessment and prediction (Mahmoud *et al.*, 2005).

SHM is generally applied for damage detection and identification, structural performance assessment and prediction, and effective management of structural systems under uncertainty (Doebling *et al.*, 1996; Farhey, 2005; Brownjohn, 2007; Orcesi and Frangopol, 2010). These objectives are related to the reduction of the epistemic uncertainty (Frangopol *et al.*, 2008; Kim and Frangopol, 2011). Therefore, the use of probabilistic concepts and methods is necessary for rational prognosis based on SHM information.

5.2.1 Statistics of extremes

Extreme values from observed data are necessary to design and assess engineering structures. These extreme values may be treated as random variables (Ang and Tang, 1984). The sample size n taken from the population of an initial variate X is considered in order to find the probability density function (PDF) of the extreme value. The largest value $Y_{max,n}$ among n initial variables is $max\{X_1, X_2, ..., X_n\}$. Assuming that the initial variables $X_1, X_2, ..., X_n$ are identically distributed (i.e., $F_{X1}(x) = F_{X2}(x) = ... = F_{Xn}(x) = F_X(x)$) and statistically independent, the cumulative distribution function (CDF) of $Y_{max,n}$ for the n initial variables $X_1, X_2, ..., X_n$ is

$$F_{Y\,max,n}(y) \equiv P(Y_{max,n} \leq y) = P(X_1 \leq y, X_2 \leq y, ..., X_n \leq y) = \left[F_X(y)\right]^n \quad [5.1]$$

In a similar way, the CDF of the smallest value Y_{min} can be obtained

$$F_{Y\,min,n}(y) \equiv 1 - P(Y_{min,n} > y) = 1 - \left[1 - F_X(y)\right]^n \quad [5.2]$$

As the sample size $n \to \infty$, $F_{Ymax,n}(y)$ (see Equation [5.1]) and $F_{Ymin,n}(y)$ (see Equation [5.2]) may converge to a particular distribution type, which depends on how the tail of the initial distribution decays in the direction of the extreme. Gumbel (1958) categorized the asymptotic distributions into three types: (a) Type I asymptotic form (i.e., the double exponential form) (b) Type II asymptotic form (i.e., the exponential form), and (c) Type III asymptotic form (i.e., the exponential form with upper bound). For instance, the largest and smallest values of the normally distributed initial variables having exponentially decaying tails are associated with the Type I asymptotic distribution. The CDF of the Type I asymptotic form is expressed as

$$F_{Y \max, n}(y) = \exp\left[-\exp\left(-\beta_{\max}(y - \alpha_{\max})\right)\right] \quad \text{for the largest value} \quad [5.3a]$$

$$F_{Y \min, n}(y) = 1 - \exp\left[-\exp\left(-\beta_{\min}(y - \alpha_{\min})\right)\right] \quad \text{for the smallest value}$$

$$[5.3b]$$

where $\alpha_{\max}$ and $\alpha_{\min}$ are the location parameters, and $\beta_{\max}$ and $\beta_{\min}$ are the scale parameters. The location parameters $\alpha_{\max}$ and $\alpha_{\min}$ are $F_X^{-1}(1 - 1/n)$ and $F_X^{-1}(1/n)$, respectively, where $F_X^{-1}(\cdot)$ = inverse CDF of the initial variate X, and n = sample size of the initial population X. The scale parameters $\beta_{\max}$ and $\beta_{\min}$ are defined as $n \cdot f_X(\alpha_{\max})$ and $n \cdot f_X(\alpha_{\min})$, respectively, where $f_X(\cdot)$ = PDF of the initial variate X.

5.2.2 Exceedance probability

Exceedance probability is referred to as the probability that a certain value will be exceeded in a predefined future time period. The exceedance probability can be used to predict extreme events such as floods, earthquakes, and hurricanes (Lambert *et al.*, 1994; Kunreuther, 2002). This probability could be dependent on or independent of the PDFs of the existing observations (Ang and Tang, 1984).

Distribution independent exceedance probability: when the number of existing observations is not enough to estimate the PDF of the existing observations, the exceedance probability p_{exd} (i.e., probability that the observed largest value among n existing observations will be exceeded in N future observations) is

$$p_{exd} = \frac{N}{(N+n)} \qquad [5.4]$$

Distribution dependent exceedance probability: considering the PDF of existing observations can lead to more accurate estimation of exceedance probability. If $Y_{\max,n}$ is the largest value among n existing observations (i.e., $Y_{\max,n} = \max\{X_1, X_2, \ldots, X_n\}$) and the initial distribution of X is associated with Type I or II asymptotic distribution, the probability that the largest value among N future observations $Y_{\max,N}$ is larger than $Y_{\max,n}$ is

$$P\left(Y_{\max,N} > Y_{\max,n}\right) = 1 - \exp\left[-\frac{N}{n}\right] \qquad [5.5]$$

5.2.3 Availability of monitoring data for performance prediction

The relation between the predictor variable (i.e., time) and the response variable (i.e., physical quantity) can be formulated by a prediction function $f_p(t)$ approximated to first, second or third order (Frangopol *et al.*, 2008) as

$$f_p(t) = \sum_{j=0}^{m} c_i \cdot t^i \qquad [5.6]$$

where c_i = coefficient, m = order of the prediction function (i.e., $m = 1, 2$, or 3), and t = time. The real physical quantity q_t can be expressed as

$$q_t = f_p(t) + e_t \qquad [5.7]$$

where q_t = monitored data at time t, and $f_p(t)$ and e_t are the regression function and the error between values from the monitored data and values from the prediction function at time t, respectively. The linear performance prediction model based on monitoring data during t_m days is shown in Fig. 5.1a. In general, the errors between the values from the linear regression model and the actual data can be assumed normally distributed with mean value 0, if the data are mutually independent, the number of data is large enough, and the regression model is obtained appropriately (Rosenkrantz, 1997). For this reason, the largest values from the errors as initial variate can be expressed as the Type I asymptotic PDF, as shown in Fig. 5.1b.

The availability of monitoring data for performance prediction is defined as the probability that the performance prediction model based on monitoring data is usable in the future (Kim and Frangopol, 2011). The performance prediction model in a non-usable state can become usable by establishing a prediction model based on new monitoring data. The criterion for availability of monitoring data is as follows: if the largest error between values from performance prediction model and monitoring data in future t_f days exceeds the largest error during the monitoring period t_m days, the monitoring data cannot be used for prediction of the structural performance. Therefore, the average availability $\overline{A}$ of monitoring data for performance prediction can be defined as (Ang and Tang, 1984)

$$\overline{A} = \frac{T_L}{t_f} \cdot p_{\text{exd}}(t_f) + \left(1 - p_{\text{exd}}(t_f)\right) \qquad 0 \le T_L \le t_f \qquad [5.8]$$

(a)

(b)
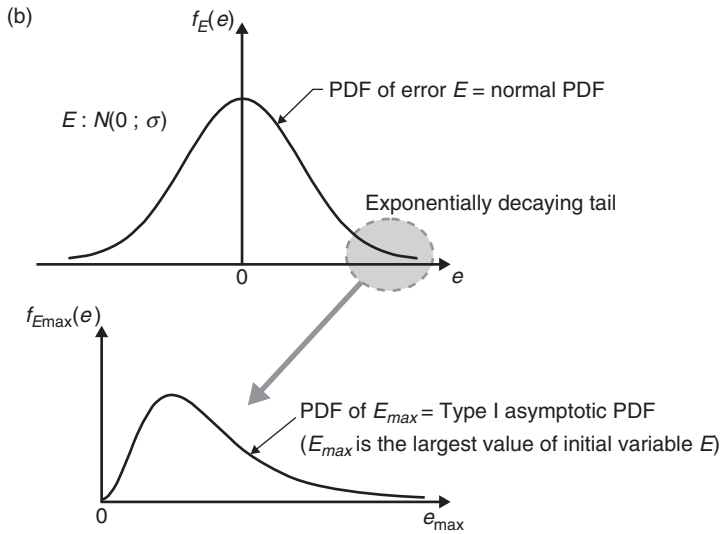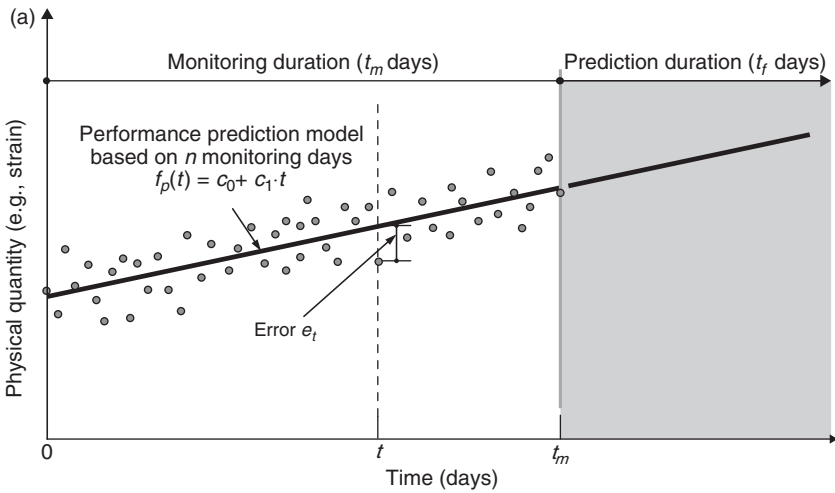
5.1 (a) Performance prediction model based on monitoring data; and (b) PDF of the error between the values from the prediction model and the actual data (i.e., normal PDF), and the associated largest value PDF (i.e., Type I asymptotic PDF).

where $p_{exd}\left(t_f\right) = 1 - \exp\left(-t_f / t_m\right)$ (see Equation [5.5]), and T_L is the time to lose the usability of the prediction model. The expected average availability of monitoring data for prediction is (Ang and Tang, 1984)

$$E\left(\overline{A}\right) = 1 - \frac{1}{t_f} \cdot \int_0^{t_f} p_{exd}\left(x\right)dx = \frac{t_m}{t_f}\left[1 - \exp\left(\frac{t_f}{t_m}\right)\right] \qquad [5.9]$$

The expected average availability of monitoring data $E(\bar{A})$ can be used for optimum monitoring planning (Kim and Frangopol, 2011), where the formulation of the bi-objective optimization problem is associated with maximization of $E(\bar{A})$ and minimization of monitoring cost.

5.2.4 Bayesian updating

The limited monitoring data can be used to systematically update existing information or judgements by using Bayesian techniques. Furthermore, the updated results can lead to better performance prediction (Frangopol, 2011). If the existing parameter θ is a random variable with PDF $f'_\Theta(\theta)$ and the monitoring data provide the likelihood function $L(\theta)$, the updated PDF $f''_\Theta(\theta)$ of the parameter θ can be obtained as (Ang and Tang, 2007)

$$f''_\Theta(\theta) = \frac{L(\theta) f'_\Theta(\theta)}{\int_{-\infty}^{\infty} L(\theta) f'_\Theta(\theta) d\theta} \tag{5.10}$$

The updated mean μ''_θ and standard deviation σ''_θ of the random variable θ are

$$\mu''_\theta = \int_{-\infty}^{\infty} \theta f''_\Theta(\theta) d\theta \tag{5.11a}$$

$$\sigma''_\theta = \left[\int_{-\infty}^{\infty} (\theta - \mu''_\theta)^2 f''_\Theta(\theta) d\theta \right]^{0.5} \tag{5.11b}$$

For example, suppose the initial error between the values from the linear regression model and the actual data is normally distributed as shown in Fig. 5.1, and the mean and standard deviation of the error E are 0.0 and 5.0 (i.e., $N(0; 5)$), respectively. The additional information from monitoring shows that the PDF of the error is normal with $N(0; 4)$. Therefore, the prior PDF $f'_E(e)$ and likelihood PDF $L(e)$ are

$$f'_E(e) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{e}{5}\right)^2 \right] \tag{5.12a}$$

$$L(e) = \frac{1}{4\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{e}{4}\right)^2 \right] \tag{5.12b}$$

Since $1/\int_{-\infty}^{\infty} L(e)f'_E(e)de = 16.05,$ the   updated   PDF $f''_E(e)$ is $0.1277\exp(-0.05125e^2)$ as indicated in Equation [5.10]. According to Equation [5.11], the mean and standard deviation of the updated error are 0 and 3.124, respectively. Figure 5.2 shows the prior, likelihood, and updated PDFs.

The Bayesian updating approach has been treated as an attractive tool to establish more rational structural performance prediction and life-cycle analysis. Orcesi and Frangopol (2011) proposed a probabilistic approach for optimum maintenance planning using monitoring data, where time to failure is predicted and updated through the Bayesian process, and optimum maintenance plans with and without updating are compared. In Okasha and Frangopol (2012), an approach using Bayesian updating based on several algorithms (e.g., Metropolis-Hasting algorithm, slice sampling algorithm) is presented for more accurate system performance prediction and life-cycle analysis.

5.2.5   Performance prediction error and loss function

The uncertainty of the monitored data has a significant effect on the monitoring plan (Kim and Frangopol, 2009). The use of the non-usable monitored data for performance prediction can result in monetary loss. This loss can be incurred when damage prognosis and structural performance estimation
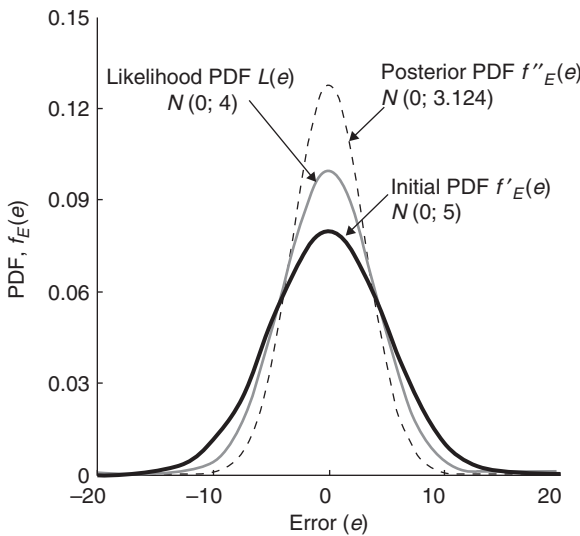
5.2 Prior, likelihood, and updated PDFs of the error between the values from the prediction model and the actual data.

are inaccurate; ultimately, this loss may lead to untimely and inappropriate maintenance actions in addition to unexpected unserviceability of civil infrastructure systems. In order to quantify the monetary loss statistically, the loss function based on prediction error can be used (Taguchi *et al.*, 1988). The prediction error is expressed as the difference between the actual value of parameter θ and its estimated parameter $\hat{\theta}$ (i.e., $|\theta - \hat{\theta}|$) (Ang and Tang, 1984). If the loss function is expressed as

$$f_{\text{Loss}}\left(\theta,\ \hat{\theta}\right) = c_l \cdot \left(\theta - \hat{\theta}\right)^2 \qquad [5.13]$$

where c_l is the monetary loss due to an error in the estimation, the expected loss E_{Loss} can be obtained as

$$E_{\text{Loss}} = \int_{-\infty}^{\infty} f_{\text{Loss}}\left(\theta,\hat{\theta}\right) \cdot f_{\Theta}\left(\theta\right) d\theta = c_l \cdot \left\{ \left(\sigma_\theta\right)^2 + \left(\mu_\theta - \hat{\theta}\right)^2 \right\} \qquad [5.14]$$

where $f_{\Theta}\left(\theta\right)$ is the PDF of θ, and μ_θ and σ_θ are the mean and standard deviation of the parameter θ, respectively. Figure 5.3a and 5.3b show the loss function $f_{\text{Loss}}\left(\theta,\hat{\theta}\right)$ and the PDFs $f_{\Theta}\left(\theta\right)$ of the three cases indicated in Table 5.1. The expected loss E_{Loss} of case 1 has the smallest value, and case 3 provides the largest E_{Loss} among these three cases as shown in Fig. 5.3c.

In general, as the sample size increases, the prediction accuracy can be improved. The loss function considering cost of sampling can be used to find the optimal sample size. Assuming that the loss function is

$$f_{\text{Loss}}\left(\theta,\hat{\theta},n\right) = c_l \cdot \left(\theta - \hat{\theta}\right)^2 + c_s \cdot n \qquad [5.15]$$

where n = number of samples, c_s = cost per sample, and the parameter θ is normally distributed, the optimal sample size n_{opt} to minimize the expected loss is (Ang and Tang, 1984)

Table 5.1 Mean and standard deviation of parameter θ for three cases

	Case 1	Case 2	Case 3	Remarks
Mean μ_θ	$\mu_{\theta,\text{case1}}$	$\mu_{\theta,\text{case2}}$	$\mu_{\theta,\text{case3}}$	$\left\vert\mu_{\theta,\text{case1}} - \hat{\theta}\right\vert < \left\vert\mu_{\theta,\text{case2}} - \hat{\theta}\right\vert = \left\vert\mu_{\theta,\text{case3}} - \hat{\theta}\right\vert$
Standard deviation σ_θ	$\sigma_{\theta,\text{case1}}$	$\sigma_{\theta,\text{case2}}$	$\sigma_{\theta,\text{case3}}$	$\sigma_{\theta,\text{case1}} = \sigma_{\theta,\text{case2}} < \sigma_{\theta,\text{case3}}$

(a)

$$f_{\text{Loss}}(\theta, \hat{\theta}) = c_1 (\theta - \hat{\theta})^2$$

Loss, $f_{\text{Loss}}(\theta, \hat{\theta})$

$\hat{\theta}$

Parameter (θ)

(b)

PDF, $f_{\Theta}(\theta)$

Case 1   Case 2

Case 3

$\mu_{\theta, case1}$ $\hat{\theta}$ $\mu_{\theta, case2} = \mu_{\theta, case3}$

Parameter (θ)

(c)

Expected loss $E_{\text{Loss}} = c_1 \left\{ (\sigma_\theta)^2 + (\mu_\theta - \hat{\theta})^2 \right\}$

Expected loss, E_{Loss}

Case 1   Case 2   Case 3
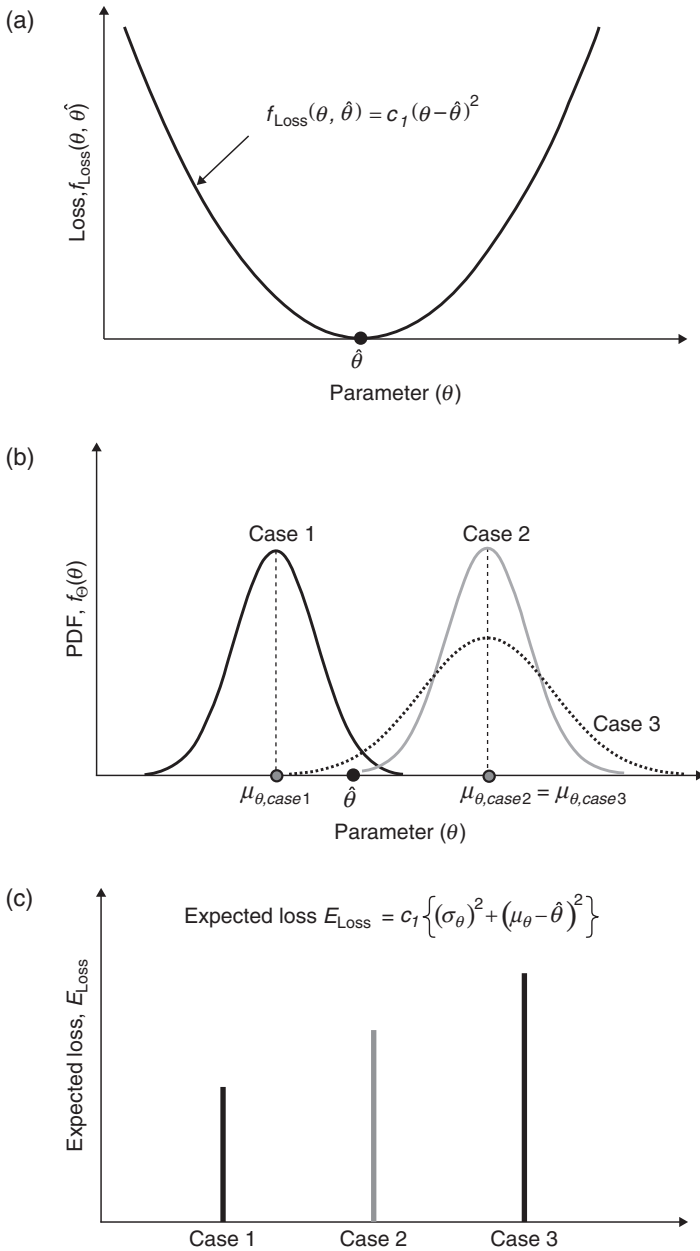
5.3 (a) Loss function; (b) PDFs of parameter θ; and (c) expected losses of three cases defined in Table 5.1.

$$n_{opt} = \sigma_\theta \cdot \sqrt{\frac{c_l}{c_s} - \left(\frac{\sigma_\theta}{\sigma'_\theta}\right)^2} \quad \text{for} \quad \sigma_\theta \cdot \sqrt{\frac{c_l}{c_s}} > \left(\frac{\sigma_\theta}{\sigma'_\theta}\right)^2$$

$$n_{opt} = 0 \quad \text{for} \quad \sigma_\theta \cdot \sqrt{\frac{c_l}{c_s}} \le \left(\frac{\sigma_\theta}{\sigma'_\theta}\right)^2$$

with prior information

[5.16a, b]

$$n_{opt} = \sigma_\theta \cdot \sqrt{\frac{c_l}{c_s}} \quad \text{without prior information} \qquad \text{[5.16c]}$$

where $\sigma'_\theta =$ prior standard deviation of the parameter θ.

When the monitoring system provides daily physical quantities, the optimal monitoring days can be obtained using Equation [5.16]. As an example, consider the error between the value from the linear regression model and the actual data as shown in Fig. 5.1. The loss function is expressed by $f_{\text{Loss}}(e, \hat{e}, t_m) = c_l \cdot (e - \hat{e})^2 + c_s \cdot t_m$, where $c_l = 1000$ and $c_s = 10$. When the standard deviation of the error σ_e is 4.0, the optimal number of monitoring days is $4 \cdot 10 = 40$ days (see Equation [5.16c]). If the prior information indicates $\sigma'_e = 2.0$, the optimal number of monitoring days is 36 days (see Equation [5.16a]).

5.3   Decision analysis based on availability of SHM data

Continuous long-term SHM is needed for reliable assessment and prediction of the structural performance; it can lead to preventing unexpected failure of a structure by applying appropriate maintenance actions on time. However, this is not practical, due to limited financial resources. Therefore, cost-effective SHM is needed. In this section, decision analysis based on the availability of SHM data for cost-effective SHM planning is presented. As mentioned previously (see Section 5.2.3), the expected average availability of SHM data is formulated using the statistics of extremes and the performance prediction based on SHM data.

5.3.1   General concepts of decision analysis

Decision analysis has been widely used for engineering planning and design processes (Dixon, 1966; Benjamin and Cornell, 1970; Saari, 2006). The various factors of a decision analysis can be represented by a decision tree consisting of the sequence of decisions (i.e., alternatives) and the associated probabilities and outcomes. The main purpose of the decision analysis is

to make the best decision by comparing all the possible alternatives. If the decision is expressed in terms of a monetary value, the decision associated with the minimum expected monetary loss (EML) is the solution.

A decision tree with n alternatives is shown in Fig. 5.4a. The rectangular and circular nodes in this figure indicate decision nodes and chance nodes, respectively. The decision tree begins with a decision node where there are n alternatives. Each alternative can lead to several possible outcomes originating from a chance node. The EML of the ith alternative EML(a_i) is (Ang and Tang, 1984):

$$\text{EML}\left(a_i\right) = \sum_{j=1}^{m_i} p_{i,j} \cdot c_{i,j} \qquad [5.17]$$

where m_i is the total number of consequences of alternative a_i, and $p_{i,j}$ and $c_{i,j}$ are the probability and monetary value of the jth consequence associated with alternative a_i, respectively. It should be noted that the outcomes from a chance node are mutually exclusive and collectively exhaustive (i.e., $\sum_{j=1}^{m_i} p_{i,j} = 1.0$). According to the minimum EML, the optimal alternative a_{opt} is determined as that having the minimum EML among n alternatives:

$$\text{EML}\left(a_{\text{opt}}\right) = \min\left\{\text{EML}\left(a_1\right), \text{EML}\left(a_2\right),\ldots,\text{EML}\left(a_i\right),\ldots,\text{EML}\left(a_n\right)\right\} \qquad [5.18]$$

5.3.2   Decision analysis based on availability of SHM data for cost-effective SHM planning

The EML of SHM planning is formulated by using the expected average availability and monitoring cost. The SHM planning associated with minimum EML is selected as the optimum SHM plan. As shown in Fig. 5.4b, each monitoring plan has two events: the monitoring data for performance prediction are usable and not usable during the prediction duration. For the event of usable monitoring data during the prediction duration, the corresponding probability and monetary value are $E_i\left(\overline{A}\right)$ and $c_{i,u}$, respectively. The probability and monetary value associated with the event of non-usable monitoring data during the prediction duration are $1 - E_i\left(\overline{A}\right)$ and $c_{i,\text{nu}}$, respectively. According to Equation [5.17], the EML of monitoring plan i is

$$\text{EML}\left(\text{Plan}\,i\right) = E_i\left(\overline{A}\right) \cdot c_{i,u} + \left[1 - E_i\left(\overline{A}\right)\right] \cdot c_{i,\text{nu}} \qquad [5.19]$$

(a)

Decision node ■
Chance node ●

(b)

Decision node ■
Chance node ●
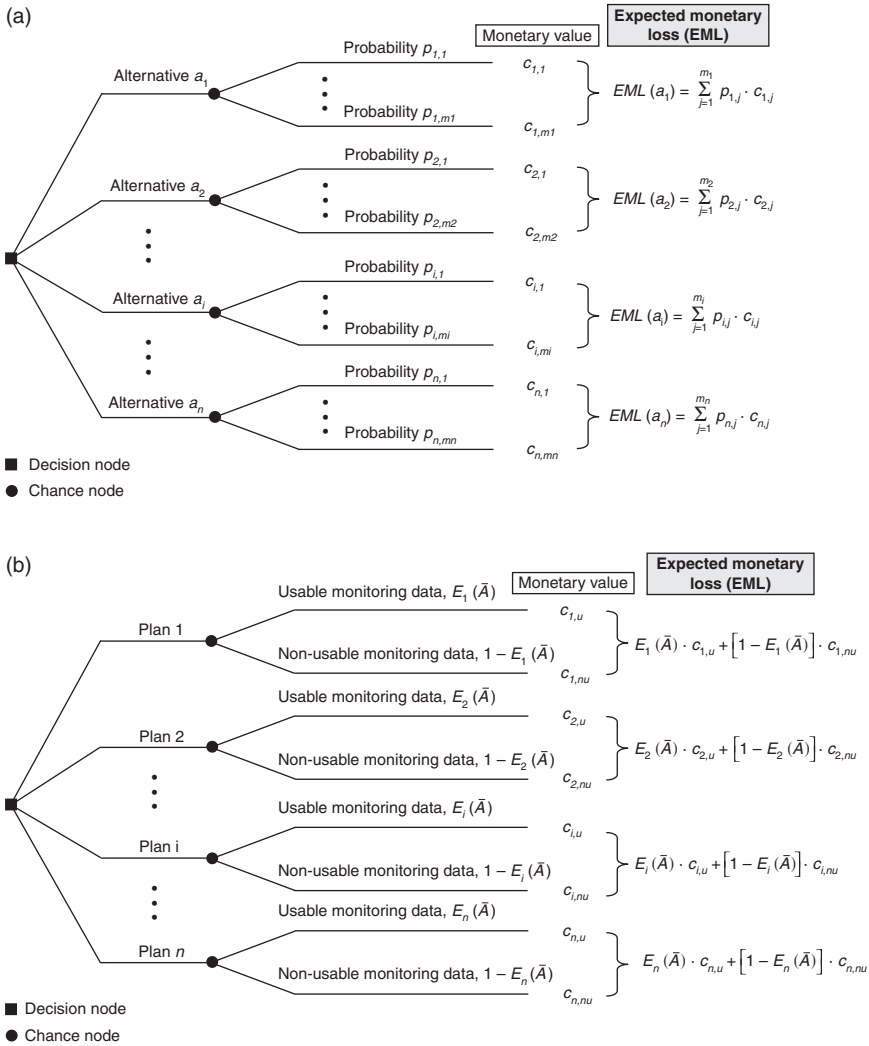
5.4 (a) Generic decision tree; and (b) decision tree for cost-effective SHM planning.

The expected average availability of monitoring data for prediction $E_i\left(\bar{A}\right)$ is defined as shown in Equation [5.9]. According to the minimum EML as indicated in Equation [5.18], a cost-effective SHM plan that provides optimum monitoring and non-monitoring time periods can be established.

As an illustrative example, it is assumed that the monitoring cost is proportional to the monitoring duration, and the monitoring costs per day for the usable and non-usable cases can be estimated using the monitoring cost c_o during t_{mo} days as

$$c_{i,u} = \frac{\left(\left(c_o/t_{mo}\right)\cdot t_{m,i}\right)}{\left(t_{m,i}+t_{f,i}\right)} \qquad \text{for usable case} \qquad [5.20a]$$

$$c_{i,nu} = \frac{\left(\left(c_o/t_{mo}\right)\cdot t_{m,i}\right)}{\left(t_{m,i}+t_{f,i}\right)} + c_{loss} \qquad \text{for non-usable case} \qquad [5.20b]$$

where $t_{m,i}$ = monitoring duration, $t_{f,i}$ = prediction duration of monitoring plan i, and c_{loss} = monetary loss per day due to the use of non-usable monitoring data for performance prediction. Substituting Equations [5.20a] and [5.20b] into Equation [5.19] leads to EML for plan i as

$$\text{EML}\left(\text{Plan}\,i\right) = E_i\left(\overline{A}\right)\cdot\frac{\left(\left(c_o/t_{mo}\right)\cdot t_{m,i}\right)}{\left(t_{m,i}+t_{f,i}\right)} + \left[1 - E_i\left(\overline{A}\right)\right]\cdot\left[\frac{\left(\left(c_o/t_{mo}\right)\cdot t_{m,i}\right)}{\left(t_{m,i}+t_{f,i}\right)} + c_{loss}\right]$$

[5.21]

Assuming c_o = 10 000 and t_{mo} = 100 days, the relation of the ratio of monitoring duration to prediction duration (i.e., t_m/t_f), EML, and the monetary loss c_{loss} is presented in Plate VIa in the color section between pages 374 and 375. Figure 5.5a shows EML versus t_m/t_f for various values of c_{loss} (i.e., 0, 50, 100, 150 and 200). From this figure, it can be seen that as the loss c_{loss} increases, the monitoring plan associated with the larger t_m/t_f is the optimum monitoring plan according to minimum EML criterion. For c_{loss} = 100, the optimum monitoring
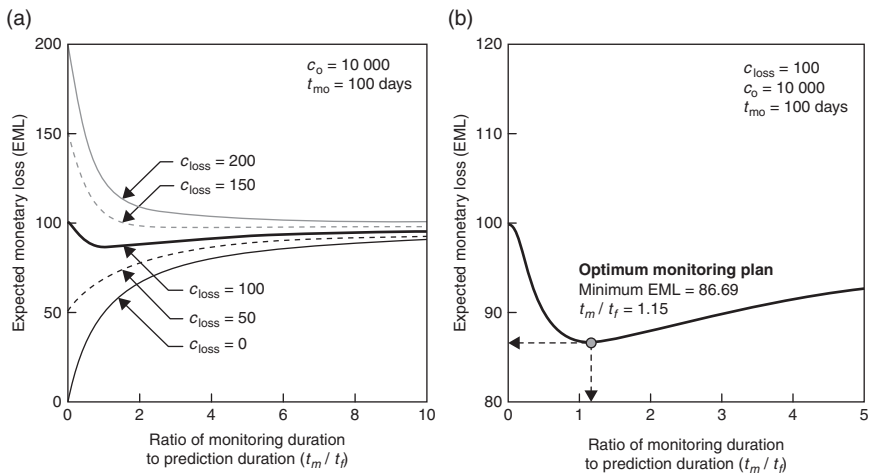
5.5 (a) Relations between t_m/t_f and EML for c_{loss} = 0, 50, 100, 150 and 200 and (b) minimum EML and the associated t_m/t_f for c_{loss} = 100.

plan has the ratio $t_m/t_f = 1.15$, and the associated EML = 86.69 as shown in Fig. 5.5b.

Typically, the decision to ensure the desired level of structural performance during the specific lifetime of a structural system under risk has to be made with incomplete and uncertain information (Ang and De Leon, 2005). The decision analysis has been treated as a rational and efficient tool in risk assessment of civil infrastructures over the last decade. Einstein (1996) introduced an overview of the basic aspects of risk analysis and decision making, including the examples associated with several tunnel projects. The decision-making framework for risk assessment, including discussions on risk acceptance criteria and associated expected loss, has been presented (Ditlevsen, 2003; Faber and Stewart, 2003; Hsu *et al.*, 2012). Furthermore, the risk informed decision analysis for life-cycle analysis under uncertainty can be found in Ang (2011).

5.4   Life-cycle analysis using monitoring data

Accurate service life prediction of a deteriorating civil infrastructure can lead to cost-effective maintenance management for extending its service life. The deterioration process of structures and infrastructures depends on the various environmental and mechanical stressors under uncertainty (Smoak, 2002; Moan, 2005; NCHRP, 2006). The efficient integration of SHM into service life prediction of a deteriorating civil infrastructure can improve the accuracy and reduce the uncertainty associated with the life-cycle analysis, and finally lead to cost-effective and more rational maintenance management (Frangopol, 2011).

5.4.1   Life-cycle performance and cost analysis

Limited financial resources should be allocated in a rational way so that lifetime structural performance can be improved, and the service life of a structure can be extended, ensuring the structural safety over the lifetime (Das, 1999; Stewart, 2006). This requires reliable modeling of loadings, accurate prediction of structural performance, and proper estimation of maintenance management cost over time (Frangopol and Liu, 2006; Pandey *et al.*, 2009). Since time-dependent structural deterioration processes under continuously changing environmental conditions are highly uncertain, reliability-based approaches for life-cycle analysis are necessary (Estes and Frangopol, 1999; Frangopol and Maute, 2003; Kong and Frangopol, 2003, 2005).

The reliability of a structure is expressed using a state function (Ang and Tang, 1984). The state function $g(\mathbf{X})$ is

$$g(\mathbf{X}) = g(X_1, X_2, \ldots X_n) \qquad [5.22]$$

where $\mathbf{X} = (X_1, X_2, \ldots X_n)$ is a vector of random variables. The state of the structure can be determined by $g(\mathbf{X})$ as follows: (a) $[g(\mathbf{X}) > 0]$ = *safe state*; (b) $[g(\mathbf{X}) < 0]$ = *failure state*; and (c) $[g(\mathbf{X}) = 0]$ = *limit state*. The reliability p_S and the probability of failure p_F are

$$p_S = \int_{g(\mathbf{X})>0} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \qquad [5.23a]$$

$$p_F = \int_{g(\mathbf{X})<0} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \qquad [5.23b]$$

where $f_{\mathbf{X}}(\mathbf{x})$ = joint PDF of $\mathbf{X}$. Furthermore, time-dependent reliability $p_S(t)$ at time t is formulated from Equation [5.23a] as

$$p_S(t) = \int_{g[\mathbf{X}(t)]>0} f_{\mathbf{X}(t)}[\mathbf{x}(t)] \, d\mathbf{x}(t) \qquad [5.24]$$

Maintenance actions can improve the structural performance (e.g., reliability) and extend the service life of a structure. Figure 5.6 shows the effect of maintenance on the structural reliability. The service life of a structure t_{life} can be defined as the time when the reliability p_S reaches the predefined threshold reliability $p_{S,\mathrm{thres}}$. As shown in this figure, the first and second maintenance actions lead to improvement of the reliability by $\Delta p_{S,\mathrm{ma1}}$ and $\Delta p_{S,\mathrm{ma2}}$, respectively. Accordingly, the service life extension under the first maintenance is $(t_{\mathrm{life,ma1}} - t_{\mathrm{life},o})$, and under the first and second maintenances is $(t_{\mathrm{life,ma2}} - t_{\mathrm{life},o})$.
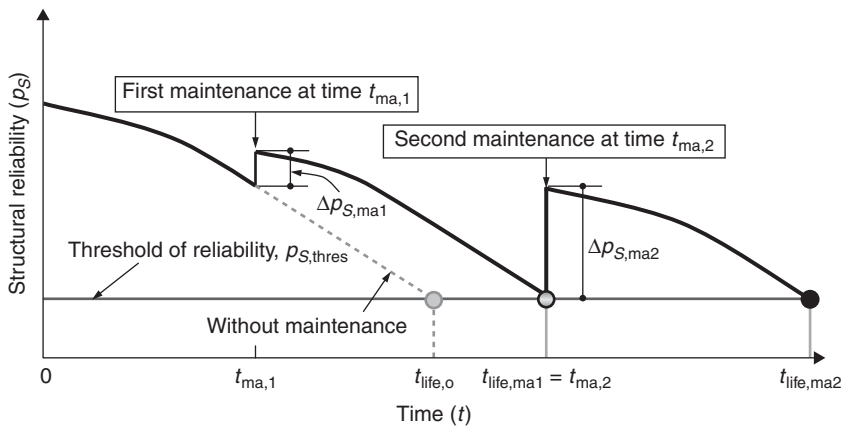
5.6 Time-dependent reliability profile with and without maintenance action.

The types and times of maintenance actions have to be determined considering the maintenance cost, the improvement of structural performance, and the extension of the service life. Through an optimization process, the reliability-based life-cycle analysis can be performed to find the optimum types and times of maintenance actions. The minimum expected total cost is the life-cycle optimization criterion for the deteriorating civil infrastructure (Frangopol *et al.*, 1997; Frangopol and Maute, 2003; Kong and Frangopol, 2005). The general formulation of the expected life-cycle cost c_{ET} is (Frangopol *et al.*, 1997):

$$c_{ET} = c_{INI} + c_{INS} + c_M + c_{FAIL} \qquad [5.25]$$

where c_{INI} = initial cost (i.e., design and construction cost), c_{INS} = inspection cost, and c_M = maintenance cost. The expected cost of failure c_{FAIL} in Equation [5.25] is defined as (Frangopol *et al.*, 1997):

$$c_{FAIL} = p_{F,\,life} \cdot c_{fail} \qquad [5.26]$$

where $p_{F,life}$ = lifetime probability of failure, and c_{fail} = monetary loss due to structural failure. $p_{F,life}$ is defined as the maximum value of p_F during the predefined lifetime.

5.4.2   Effects of SHM on performance and service life prediction

Deteriorating civil infrastructure systems are usually monitored at potential critical locations. The purpose of SHM includes detecting and identifying damages at potential critical locations, assessing and predicting structural performance, and providing information for more reliable maintenance planning (Frangopol *et al.*, 2008). Efficient use of SHM can lead to improvement of accuracy associated with structural performance assessment and prediction, and finally can result in cost-effective maintenance management.

The possible effects of SHM on the prediction of structural performance index and service life under uncertainty are illustrated in Fig. 5.7. In this figure, there are three cases, as follows:

- *Case I* is associated with initial structural performance index prediction without updating based on monitoring data.
- *Case II* indicates that SHM is performed at time t_{mon}, and the mean values of the updated structural performance index $E(P_{II})$ and service life $E(t_{life,II})$ are less than the mean of initially predicted performance

index $E(P_I)$ and service life $E(t_{\text{life},I})$, respectively. In this case, it is clear that inaccurate prediction of structural performance index can delay the maintenance and endanger the survival of the structure (Frangopol, 2011).

- *Case III* shows that the updating based on SHM can lead to prediction of higher mean values of structural performance index $E(P_{III})$ and service life $E(t_{\text{life},III})$ than $E(P_I)$ and $E(t_{\text{life},I})$, respectively, and unnecessary maintenance actions can be avoided (Frangopol, 2011).

The PDFs of the structural performance index and of the service life of these cases are illustrated in Fig. 5.7. The updating using SHM data can reduce the dispersions of both structural performance index and service life prediction.

5.4.3   Efficient use of monitoring data for structural performance prediction

If the SHM is applied, the expected total cost will be (Frangopol and Messervey, 2007)

$$c'_{\text{ET}} = c'_{\text{INI}} + c'_{\text{INS}} + c'_{M} + c'_{\text{FAIL}} + c_{\text{mon}} \qquad [5.27]$$
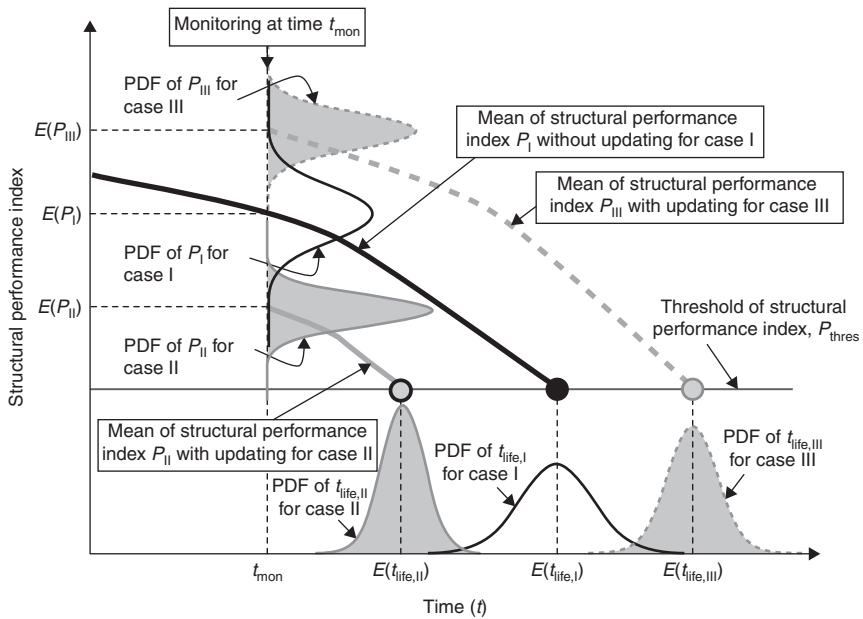
5.7 Prediction of structural performance index with and without updating based on SHM data.

where c_{mon} = monitoring cost. The superscript $'$ indicates costs in Equation [5.25] affected by monitoring. The benefit of SHM B_{mon} can be determined as $c_{ET} - c'_{ET}$ (i.e., difference between the expected costs associated with Equations [5.25] and [5.27]). If the application of SHM does not provide a benefit (i.e., $B_{mon} < 0$), it will be difficult for structure managers to justify adopting SHM. For this reason, efficient use of SHM should be considered to maximize its benefit.

Liu *et al.* (2009) have introduced a practical approach to assess the performance of a structure based on SHM data through formulation of the state function as

$$g = s_o - (1+e) \cdot \tau(t) \cdot s_{mon} \qquad [5.28]$$

where s_o = predefined stress (e.g., the maximum stress obtained from experimental studies on strains), s_{mon} = stress derived from monitored data on strains, $\tau(t)$ = time-dependent function at future time t that is used to predict future component stress, and e = measurement error in the monitored data. This error is assumed to be normally distributed with zero mean value. The time-dependent function $\tau(t)$ at time t is defined as the ratio of the expected largest stress during future time period t to the largest stress obtained during the monitored period (Liu *et al.*, 2009). For example, if the appropriate distribution of the monitored stress s_{mon} is associated with Type I asymptotic form, the CDF of s_{mon} becomes Equation [5.3a]. The largest stress s_{mon,N_t} among the stresses induced by the expected number of heavy vehicles N_t during a specified future time can be predicted, if s_{mon,N_t} is the characteristic largest value (i.e., $1 - F_{Ymax}(s_{mon,N_t}) = 1/N_t$) as follows

$$s_{mon,N_t} = \alpha_{max} - \frac{\ln\left[-\ln\left(1-(1/N_t)\right)\right]}{\beta_{max}} \qquad [5.29]$$

Therefore, the time-dependent function $\tau(t)$ is (Liu *et al.*, 2009)

$$\tau(t) = \max\left\{\frac{s_{mon,N_t}}{\max\left\{s_1, s_2, ..., s_{n_t}\right\}}; 1.0\right\} \qquad [5.30]$$

where s_i = monitored stress induced by ith heavy vehicle, and n_t = total number of heavy vehicles during monitoring duration.

As an illustrative example, monitored data of an existing bridge (Bridge I-39, Northbound Wisconsin River Bridge) in Wisconsin, USA,
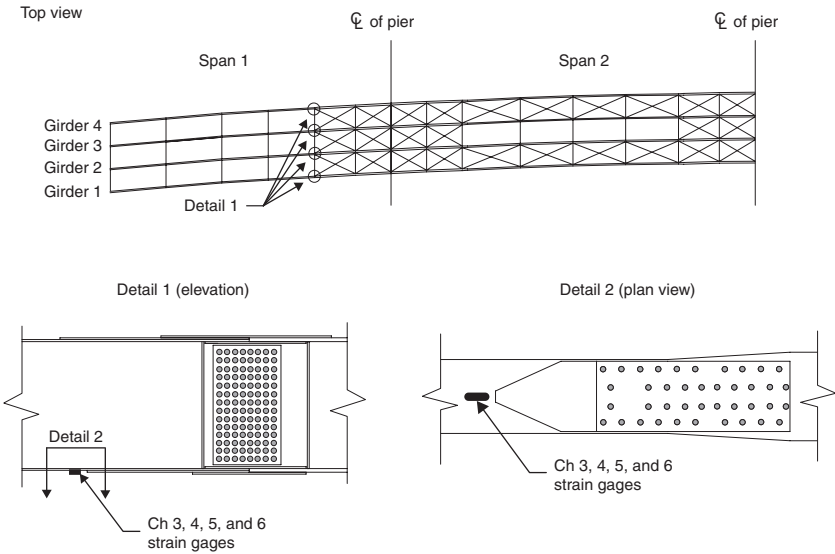
5.8 Sensor locations on the I-39 Northbound Wisconsin River Bridge. (*Source*: adapted from Mahmoud *et al.*, 2005.)

are used. This bridge is a five span continuous steel plate girder bridge with slightly curved girders, as shown in Fig. 5.8 (Mahmoud *et al.*, 2005). In this example, the monitoring data from channels 3 to 6 installed on the bottom face of the bottom flange of each girder are used (see Fig. 5.8). The histograms and the best-fit distributions (i.e., generalized extreme value PDF) of the monitored stress from channel 3 (i.e., CH 3) of girder 4 under right and left loading are shown in Fig. 5.9. Through determination of the best-fit PDF and the associated parameters of monitored stress, the state function for each girder can be formulated from Equation[5.28] as (Liu *et al.*, 2009)

$$g = s_o - (1+e) \cdot \tau(t) \cdot \left(\frac{N_r}{N_t} \cdot s_{\text{mon},r} + \frac{N_l}{N_t} \cdot s_{\text{mon},l} + \frac{N_s}{N_t} \cdot s_{\text{mon},s} \right) \qquad [5.31]$$

where $s_{\text{mon},r}$, $s_{\text{mon},l}$, and $s_{\text{mon},s}$ are monitored stresses of component *i* induced by right, left, and side-by-side lane loading, respectively. Based on the initial monitored data, the total number of heavy trucks crossing the bridge on the right lane, left lane, and side-by-side in the next *t* years are assumed $N_r = 2500t$, $N_l = 1000t$, and $N_s = 30t$, respectively. Figure 5.10 shows the time-dependent exceedance probabilities $P(s_{\text{mon}} > s_o)$ (i.e., probability associated with $g < 0$) for girders 1, 2, 3, and 4. The exceedance probability can

(a)

Histogram
N_r = 636 Trucks
Stress Bin = 0.5 MPa
Mean = 23.71 MPa
St. dev. = 4.17 MPa

Ch 3 of girder 4
under the right lane
loading

Generalized extreme
value (GEV) probability
density function

Probability density function

Monitored live load stress induced by truck traffic, $s_{mon,r}$ (MPa)

(b)

Ch 3 of girder 4
under the left lane
loading

Histogram
N_l = 249 Trucks
Stress Bin = 0.5 MPa
Mean = 13.03 MPa
St. dev. = 6.08 MPa

Generalized extreme value (GEV)
probability density function

Probability density function

Monitored live load stress induced by truck traffic, $s_{mon,l}$ (MPa)

5.9 Histogram and the associated best-fit PDF of monitored data from
CH 3 of girder 4 under (a) the right-lane loading, and (b) the left-lane
loading. (*Source*: adapted from Liu *et al.*, 2009b.)

serve in the structural performance assessment and prediction (Liu *et al.*,
2009). Using this approach, the reliability importance assessment and pre-
diction for monitoring planning at system level was performed by Kim and
Frangopol (2010).
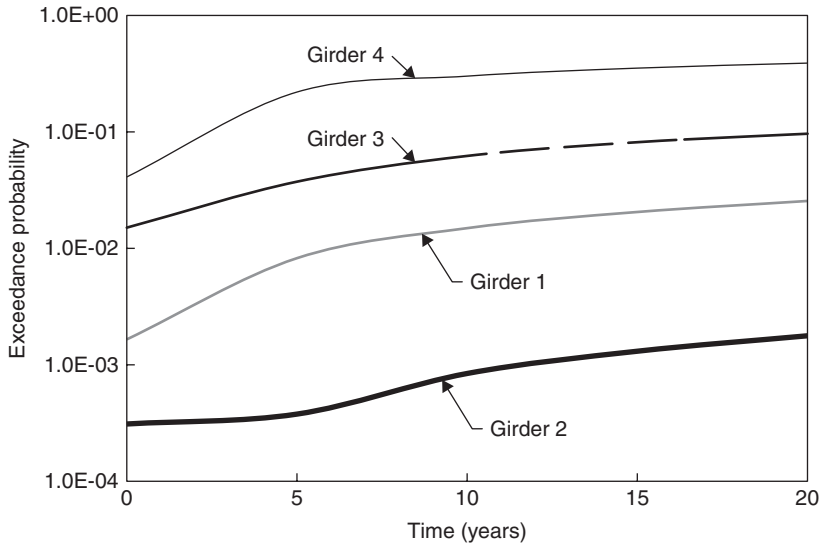
5.10 Time-dependent exceedance probabilities of girders 1, 2, 3 and 4.

5.5   Conclusions

This chapter has presented statistical and probabilistic approaches for efficient prognosis based on SHM information. The concepts of the statistics of extremes and decision analysis are employed to establish cost-effective monitoring planning considering the loss due to use of non-usable monitoring data. In order to quantify this loss, the loss function is introduced. Furthermore, general concepts of life-cycle analysis and possible effects of SHM on life-cycle analysis are presented. Finally, a practical approach to assessing and predicting structural performance using SHM data is presented.

The concepts and approaches presented in this chapter are useful for efficient prognosis, cost-effective, and more reliable maintenance management of civil infrastructure systems. However, significant future effort has to be devoted to developing multi-disciplinary methodologies including development of advanced SHM technology, statistical and probabilistic computational framework for SHM data acquisition and interpretation, and integrated life-cycle approach for optimal monitoring and maintenance (Farrar and Worden, 2007; Frangopol, 2011).

5.6   Acknowledgements

5.7   References

Ang, A. H-S. (2011). 'Life-cycle considerations in risk-informed decisions for design of civil infrastructures'. *Structure and Infrastructure Engineering*, Taylor & Francis, **7**(1–2), 3–9.

Ang, A.H-S and De Leon, D. (2005). 'Modeling and analysis of uncertainties for risk-informed decisions in infrastructures engineering'. *Structure & Infrastructure Engineering*, Taylor and Francis, **1**(1), 19–31.

Ang, A.H-S. and Tang, W.H. (1984). *Probability Concepts in Engineering Planning and Design Volume II*. New York, NY, John Wiley & Sons.

Ang, A.H-S. and Tang, W.H. (2007). *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*. 2nd edn. New York, Wiley.

Benjamin, J.R. and Cornell, C.A. (1970). *Probability, Statistics, and Decision for Civil Engineers*. New York, NY, McGraw-Hill.

Brownjohn, J.M.W. (2007). 'Structural health monitoring of civil infrastructure'. *Philosophical Transactions of the Royal Society A*, Royal Society Publishing, **365**(1851), 589–622.

Chang, P.C., Flatau, A. and Liu, S.C. (2003). 'Review paper: health monitoring of civil infrastructure'. *Structural Health Monitoring*, SAGE, **2**(3), 257–267.

Das, P.C. (1999). 'Prioritization of bridge maintenance needs'. *Case Studies in Optimal Design and Maintenance Planning of Civil Infrastructure Systems*, D. M. Frangopol, ed., Reston, VA, ASCE, 26–44.

Ditlevsen, O. (2003). 'Decision modeling and acceptance criteria'. *Structural Safety*, Elsevier, **25**(2), 165–191.

Dixon, J.R. (1966). *Design Engineering: Inventiveness, Analysis, and Decision Making*. New York, NY, McGraw-Hill.

Doebling, S., Farrar, C., Prime, M. and Schevitz, D.W. (1996). *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Review*. Los Alamos National Labs Report LA-13070-MS UC-900.

Einstein, H.H. (1996). 'Risk and risk analysis in rock engineering'. *Tunnelling and Underground Space Technology*, Elsevier, **11**(2), 141–155.

Estes, A.C. and Frangopol, D.M. (1999). 'Repair optimization of highway bridges using system reliability approach'. *Journal of Structural Engineering*, ASCE, **125**(7), 766–775.

Faber, M.H. and Stewart, M.G. (2003). 'Risk assessment for civil engineering facilities: critical overview and discussion'. *Reliability Engineering & System Safety*, Elsevier, **80**(2), 173–184.

Farhey, D.N. (2005). 'Bridge instrumentation and monitoring for structural diagnostics'. *Structural Health Monitoring*, *SAGE*, **4**(4), 301–318.

Farrar, C.R. and Lieven, N.A.J. (2007). 'Damage prognosis: the future of structural health monitoring'. *Philosophical Transactions of the Royal Society A*, Royal Society Publishing, **365**(1851), 623–632.

Farrar, C.R. and Worden, K. (2007). 'An introduction to structural health monitoring'. *Philosophical Transactions of the Royal Society A*, Royal Society Publishing, **365**(1851), 303–315.

Frangopol, D.M. (2011). 'Life-cycle performance, management, and optimization of structural systems under uncertainty: accomplishments and challenges'. *Structure and Infrastructure Engineering*, Taylor & Francis, **7**(6), 389–413.

Frangopol, D.M., Lin, K.Y. and Estes, A.C. (1997). 'Life-cycle cost design of deteriorating structures'. *Journal of Structural Engineering*, *ASCE*, **123**(10), 1390–1401.

Frangopol, D.M. and Liu, M. (2006). '*Life-Cycle Cost and Performance of Civil Structures*'. McGraw-Hill 2006 Yearbook of Science and Technology, McGraw-Hill, New York, 183–185.

Frangopol, D.M. and Maute, K. (2003). 'Life-cycle reliability-based optimization of civil and aerospace structures'. *Computers and Structures*, Elsevier, **81**(7), 397–410.

Frangopol, D.M. and Messervey, T.B. (2007). 'Risk assessment for bridge decision making'. *Proceedings of the Fourth Civil Engineering Conference in the Asian Regions, CECAR 4, Taipei, Taiwan, 25–28 June 2008* (invited paper); in ASCE Tutorial & Workshop on Quantitative Risk Assessment, Taipei, Taiwan, 25–28 June 2007, 37–42.

Frangopol, D.M. and Messervey, T.B. (2009). 'Maintenance principles for civil structures'. *Chapter 89 in Encyclopedia of Structural Health Monitoring*, C. Boller, F-K. Chang and Y. Fujino, eds., John Wiley & Sons Ltd, Chichester, UK, Vol. 4, 1533–1562.

Frangopol, D.M., Strauss, A. and Kim, S. (2008). 'Use of monitoring extreme data for the performance prediction of structures: general approach'. *Engineering Structures*, Elsevier, **30**(12), 3644–3653.

Glaser, S.D., Li, H., Wang, M.L., Ou, J. and Lynch, J. (2007). 'Sensor technology innovation for the advancement of structural health monitoring: a strategic program of US-China research for the next decade'. *Smart Structures and Systems*, Techno Press, **3**(2), 221–244.

Gumbel, E.J. (1958). *Statistics of Extremes*. New York, NY, Columbia University Press.

Hsu, W-K., Tseng, C-P., Chiang, W-L. and Chen, C-W. (2012). 'Risk and uncertainty analysis in the planning stages of a risk decision-making process'. *Natural Hazards*, Springer, **61**(3), 1355–1365.

Kim, S. and Frangopol, D.M. (2009). 'Optimal decision making of structural health monitoring under uncertainty', *Proceedings of the Tenth International Conference on Structural Safety and Reliability,* ICOSSAR2009, Osaka, Japan, 13–17 September 2009, CRC Press, Taylor & Francis Group, 621–627.

Kim, S. and Frangopol, D.M. (2010). 'Optimal planning of structural performance monitoring based on reliability importance assessment'. *Probabilistic Engineering Mechanics*, Elsevier, **25**(1), 86–98.

Kim, S. and Frangopol, D.M. (2011). 'Cost-effective lifetime structural health monitoring based on availability'. *Journal of Structural Engineering*, *ASCE*, **137**(1), 22–33.

Kong, J.S. and Frangopol, D.M. (2003). 'Life-cycle reliability based maintenance cost optimization of deteriorating structures with emphasis on bridges'. *Journal of Structural Engineering, ASCE*, **129**(6), 818–828.

Kong, J.S. and Frangopol, D.M. (2005). 'Probabilistic optimization of aging structures considering maintenance and failure cost'. *Journal of Structural Engineering, ASCE*, **131**(4), 600–616.

Kwon, K. and Frangopol, D.M. (2010). 'Bridge fatigue reliability assessment using probability density functions of equivalent stress range based on field monitoring data'. *International Journal of Fatigue*, Elsevier, **32**(8), 1221–1232.

Kulkarni, S.S. and Achenbach, J.D. (2008). 'Structural health monitoring and damage prognosis in fatigue'. *Structural Health Monitoring*, *SAGE*, **7**(1), 37–49.

Kunreuther, H. (2002). 'Risk analysis and risk management in an uncertain world'. *Risk Analysis: An International Journal*, Society for Risk Analysis, **22**(4), 655–664.

Lambert, J.H., Matalas, N.C., Ling, C.W., Haimes, Y.Y. and Li, D. (1994). 'Selection of probability distributions in characterizing risk of extreme events'. *Risk Analysis: An International Journal*, Society for Risk Analysis, **14**(5), 731–742.

Li, J. and Zhang, Y. (2006). 'Interactive sensor network data retrieval and management using principal components analysis transform'. *Smart Materials and Structures*, Institute of Physics Publishing, **15**(6), 1747–1757.

Liu, M., Frangopol, D.M. and Kim, S. (2009). 'Bridge system performance assessment from structural health monitoring: a case study'. *Journal of Structural Engineering*, ASCE, **135**(6), 733–742.

Mahmoud, H.N., Connor, R.J. and Bowman, C.A. (2005). 'Results of the fatigue evaluation and field monitoring of the I-39 Northbound Bridge over the Wisconsin River'. *ATLSS Report No. 05–04*, Lehigh University, Bethlehem, PA.

Moan, T. (2005). 'Reliability-based management of inspection, maintenance and repair of offshore structures'. *Structure and Infrastructure Engineering*, Taylor and Francis, **1**(1), 33–62.

NCHRP (2006). 'Manual on service life of corrosion-damaged reinforced concrete bridge superstructure elements'. NCHRP-report 558, *Transportation Research Board*, *National Cooperative Highway Research Program*, Washington D.C.

Okasha, N.M. and Frangopol, D.M. (2012). 'Integration of structural health monitoring in a system performance based life-cycle bridge management framework'. *Structure and Infrastructure Engineering*, Taylor & Francis, **8**(11), 999–1016.

Orcesi, A.D. and Frangopol, D.M. (2010). 'Optimization of bridge management under budget constraints: role of structural health monitoring'. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, **3**(2202), 148–158.

Orcesi, A.D. and Frangopol, D.M. (2011). 'Optimization of bridge maintenance strategies based on structural health monitoring information'. *Structural Safety*, Elsevier, **33**(1), 26–41.

Pandey, M.D., Yuan, X.-X. and van Noortwijk, J.M. (2009). 'The influence of temporal uncertainty of deterioration on life-cycle management of structures'. *Structure and Infrastructure Engineering*, Taylor & Francis, **5**(2), 145–156.

Rosenkrantz. W.A. (1997). *Introduction to Probability and Statistics for Scientists and Engineers.* McGraw-Hill Companies, Inc.

Saari, D.G. (2006). 'Fundamentals and implications of decision-making'. *Chapter 5 in Decision Making in Engineering Design*, K.E. Lewis, W. Chen and L.C. Schmidt, eds., New York, NY, ASME, 35–42.

Smoak, W.G. (2002). *Guide to Concrete Repair*. Books for Business, New York, Hong Kong.

Stewart, M.G. (2006). 'Spatial variability of damage and expected maintenance costs for deteriorating RC structures'. *Structure and Infrastructure Engineering*, Taylor & Francis, **2**(2), 79–90.

Taguchi, G., Elsayed, E.A. and Hsiang, T.C. (1988). *Quality Engineering in Production Systems*. New York, NY, McGraw-Hill.

Zárate, B.A., Caicedo, J.M. Yu, J. and Ziehl, P. (2012). 'Deterministic and probabilistic fatigue prognosis of cracked specimens using acoustic emissions'. *Journal of Constructional Steel Research*, Elsevier, **76**(9), 68–74.

5.8   Appendix: Notation used

$\overline{A}$	average availability of monitoring data for prediction
c_{fail}	monetary loss due to structural failure
c_1	monetary loss due to an error in the estimation
c_{loss}	monetary loss per day due to the use of non-usable monitoring data for prediction
c_{ET}	expected life-cycle cost
c_{INI}	initial cost
c_{INS}	inspection cost
c_{FAIL}	failure cost
c_{M}	maintenance cost
$E(\overline{A})$	expected average availability of monitoring data for prediction
E_{Loss}	expected loss
e_t	error at time t
f_{Loss}	loss function
$f'_{\Theta}(\theta)$	prior PDF of parameter θ
$f''_{\Theta}(\theta)$	updated (posterior) PDF of parameter θ
$f_p(t)$	prediction function in terms of time t
$f_X(x)$	joint PDF of random variables X
$F_X(x)$	CDF of random variable X
$F^{-1}_X(x)$	inverse CDF of random variable X
$g(X)$	state function consisting of random variables X
$L(\theta)$	likelihood function of parameter θ
n_t	total number of heavy vehicles during monitoring duration
N_t	total number of heavy vehicles during a specific future duration
p_{exd}	exceedance probability
p_F	probability of failure
p_S	reliability
P	structural performance index
q_t	monitored physical quantity at time t

s_{mon}	stress derived from monitored data on strains
t_{life}	service lifetime
t_f	prediction duration
t_m	monitoring duration
$Y_{max,n}$	largest value among n existing observations
$Y_{min,n}$	smallest value among n existing observations
α	location parameter for the CDF of Type I asymptotic distribution
β	scale parameter for the CDF of Type I asymptotic distribution
μ_θ	mean of random variable θ
σ_θ	standard deviation of random variable θ

6

System-level design of a roaming multi-modal multi-sensor system for assessing and monitoring civil infrastructures

R. BIRKEN, J. ZHANG and G. SCHIRNER,
Northeastern University, USA

Abstract: This chapter discusses the system-level design of a roaming multi-modal multi-sensor system and puts the associated general problems into a real-life context, based on the case study of Versatile On-board Traffic-Embedded Roaming Sensors (VOTERS) – a sensing system to monitor the structural health of roads and bridge decks. The design challenges of this and similar distributed sensing systems are addressed. First, the background of comparable sensing systems is reviewed, then an overview of the VOTERS multi-modal multi-sensor system with its system-level requirements for sensing, synchronizing, processing, and storing is presented. The focus then shifts to details of the hierarchical distribution of work, the organization and structure of handling large data amounts including a communication framework, and the challenges of enabling sensor fusion. All aspects are considered in the context of achieving a fully automated system.

Key words: middleware, distributed sensor system, sensor fusion, time synchronization.

6.1   Introduction

With the advances in sensor technology and improvements of algorithms for data processing, sensing systems are more widely employed. They are an important means of measuring physical quantities in virtually every domain: industrial automation, environmental monitoring, and infrastructure monitoring, to just name a few examples. Sensing systems have different deployment strategies. A classical approach uses stationary mounted sensor systems, for example in industry automation settings (e.g. sensing process variables). In recent years, mobile applications of groups of sensors have gained research interest in the form of wireless sensor networks (WSNs), in which sensors combine sensing capability with some local computation and

wireless communication. WSNs typically employ many sensors with same sensing capabilities, with peer-to-peer communication for self-organization. Their strength lies in covering large areas through many (mostly stationary) sensors where the distributed sensing is combined into a single unified view of the covered area. Conversely, roaming sensor systems are sensor systems that are moved through and survey the area of interest (e.g. by being mounted onto a survey vehicle). As a result, fewer sensors can be used to cover a larger area. However, roaming sensor systems are typically used for observing slow-changing physical phenomena. This chapter discusses the system-level design of such **roaming** sensor systems.

Sensing challenges in roaming sensor systems stem from the physical process to be monitored. Depending on the process, more than one sensor type may be needed to gain a complete view of the measured process. As an example, to assess road surface status a camera image showing surface cracks may need to be supplemented with a ground penetrating radar (GPR) image to examine the road's subsurface health condition. This drives the need for heterogeneous or **multi-modal** sensors. To form a complete view, data across the sensing domain need to be fused into a single coherent picture. Since roaming sensing systems survey areas of interest, it is important to maximize the coverage in a single pass, which can be achieved by mounting multiple sensors of the same type, for example in an array configuration. This drives the need for **multi-sensor** systems that combine measurements of many identical sensors into one view of a larger spatial area.

In consequence, the design of a roaming multi-modal multi-sensor system has to deal with many challenges. **Large** amounts of raw data need to be stored on the survey system during a survey. The data across homogenous multi-sensor compositions and heterogeneous multi-modal sensors need to be combined through **sensor fusion**. However, as sensors are physically distributed on the survey vehicle, data across sensors have to be **spatially correlated** to ensure the use of measurements of the same physical feature. The system has to allow algorithms for sensor fusion and special correlation to evolve and for raw data from earlier surveys to be reprocessed. For this, raw data have to be stored at a central location for later processing. To achieve this, **centralized storage** of all raw data are desirable. To control which area should be surveyed, a **centralized coordination** of survey vehicles is desirable.

This chapter describes the system-level design of a roaming multi-modal multi-sensor system and puts the general problems into a real-life context. Based on the case study of VOTERS – a sensing system to monitor the structural heath of roads and bridge decks – this chapter describes how the general challenges outlined above can be met. VOTERS utilizes

a fleet of survey vehicles to gather information about a larger geographic area (city, state). VOTERS addresses the complexity challenge by hierarchically distributing the work across fleet management, vehicle management, subsystems and sensors. Each vehicle contains multi-modal sensors mounted to the vehicle. Spatial correlation and sensor fusion are enabled by strict time synchronization across the subsystems in a vehicle. Sensor data are organized into streams and stored locally on the vehicle during a survey. After completing a survey, the vehicle returns to its base and the data are uploaded to a centralized location for post-processing. In consequence, VOTERS offers an efficient way to survey roads and bridge decks over a large geographic area, automating data collection, upload, and processing. It offers geo-spatially and temporally correlated survey data with convenient central access, and provides invaluable information for timely infrastructure investment decisions.

This chapter is organized as follows: Section 6.2 motivates the application example of monitoring the structural health of our transportation infrastructure. Section 6.3 reviews the background of sensing systems. Section 6.4 overviews the VOTERS multi-modal multi-sensor system with its system-level requirements for sensing, synchronizing, processing, and storing. Section 6.5 shows in more detail the hierarchical distribution of work. Section 6.6 discusses our organization and structure of handling large data amounts. Section 6.7 looks at the challenges of enabling sensor fusion. Finally, Section 6.8 concludes the chapter with a general summary and outlook for future work.

6.2   Need for health monitoring of transportation infrastructure

Civil infrastructure construction and maintenance represent a large societal investment. Despite being the lifeline of commerce, civil infrastructure is just at the beginning of benefiting from the latest advances in sensor technology. According to the latest ASCE report card[1] the United States' (US) infrastructure scores only a D, and it is estimated that $2.2 trillion in investments are needed over 5 years to bring the condition of the nation's infrastructure to a good condition.

There are four million miles of roads and nearly 600 000 bridges in the US[2] requiring a broad range of maintenance activities. The ASCE report card[1] gives bridges a grade C commenting: 'More than 26%, or one in four, of our nation's bridges are either structurally deficient or functionally obsolete. While some progress has been made in recent years to reduce the number of deficient and obsolete bridges in rural areas, the number in urban areas is rising. A $17 billion annual investment is needed to substantially improve current bridge conditions. Currently, only $10.5 billion is spent annually on the construction and maintenance of bridges.'

The same report gives roads only a grade of D-, commenting that 'Americans spend 4.2 billion h a year stuck in traffic at a cost to the economy of $78.2 billion, or $710 per motorist. Poor conditions cost motorists $67 billion a year in repairs and operating costs. One-third of America's major roads are in poor or mediocre condition and 45% of major urban highways are congested. Current spending of $70.3 billion per year for highway capital improvements is well below the estimated $186 billion needed annually to substantially improve conditions'. In Reference [1] it is estimated that an investment of $930 billion is needed over 5 years to bring bridges and roads to an acceptable level.

The US faces a monumental problem of infrastructure management in scheduling and implementing maintenance and repair, and in prioritizing expenditure within budgetary constraints. The efficient and effective performance of these operations is crucial to ensuring roadway safety, preventing catastrophic failures, and promoting economic growth. Roadway work zones used for assessment and repair are a major source of traffic congestion, which results in lost productivity and wasted fuel. It is a critical need to make the right roadway and bridge deck repairs in the right place and at the right time. However, current inspection methods and strategies (Section 6.2.2) used to characterize roadway and bridge deck conditions are not well suited to fulfill this need, because they typically only inspect small localized areas, only periodically (order of years, if at all), and mainly through visual (surface) inspection.

6.2.1   Roadway and bridge deck defects

Bridge deck and pavement deterioration frequently takes place below the surface and cannot be evaluated by visual means (Fig. 6.1). Concrete deck deterioration includes delamination arising from chloride induced rebar corrosion, cracking caused by alkali silica reaction (ASR), and cracking caused by overloading or excessive vibration. Pavement deteriorates due to internal moisture damage, debonding, and loss of subsurface support. Reinforced concrete (RC) or prestressed concrete (PC) bridge decks are often overlaid with an asphalt concrete or Portland cement concrete. The presence of the overlay makes it more difficult to detect the subsurface deterioration, and the overlay can also develop damage due to debonding. Pavement layers are subjected to extensive abrasion and deterioration from service loading (e.g. traffic) and environmental attacks (e.g. freeze-thaw, rain, road salts), and thus are subject to deterioration.

Common types of roadway damage are transverse cracks, longitudinal cracks, tracking, corrugation, potholes, delamination, and seepage. Transverse cracks occur more often than longitudinal cracks and can start with a fine crack of less than 0.5 mm in width and of less than 2 cm in depth.
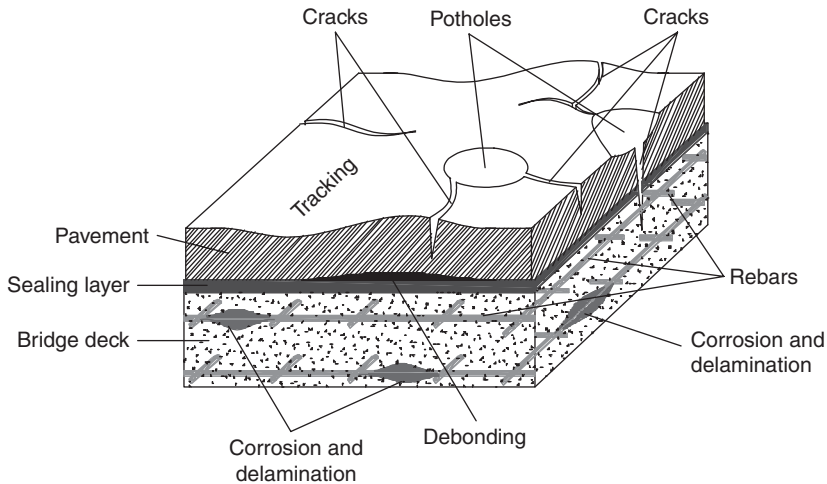
6.1 Some common defects and deteriorations found in concrete bridge decks with asphalt pavement overlay.

Such cracks are hardly visible when it is sunny, but are visible after rain due to the vaporization of the surface water that leaves water in the cracks. Small cracks need to be treated to prevent them from developing into larger cracks. Large cracks often have widths of more than 1 mm, depths of 5 cm, and run meters in length. If large cracks are not sealed, delamination and scaling will follow. If the adhesion between pavement and concrete deck decreases, the overlay may debond from the deck's top surface. The loss of adhesion may be caused by seepage from cracks or potholes. Local debonding may span only several square centimeters, and can be difficult to detect because the pavement surface remains intact. Large area delaminations may develop into large cracks at the pavement surface and eventually cause large potholes and loss of pavement. Feedback effects can complicate and accelerate damage progression. Cracks and potholes are often accompanied by seepage. Water enters into the overlay through cracks. The adhesion between asphalt and concrete deck is extremely vulnerable to water penetration. Water within cracks of a pavement will stay and seep. This is most harmful to asphalt pavement.

6.2.2   Current inspection methods

Traditional bridge deck inspection methods, such as chain drag,[3] half-cell potentials,[4] and chloride contents[4] are slow, require traffic delay causing road closures, and are often not effective. Higher speed technologies such as GPR, infrared thermography, and scanning impact-echo have been

developed and used to some extent by highway agencies to meet their needs for bridge deck condition assessment.[5-12] These technologies may incur the need for traffic closures, or provide insufficient spatial data coverage, which have reduced their acceptance and reliability.

More modern roadway and bridge deck inspection technologies are vehicle-based. They range from profile-based single or dual channel air-coupled GPR[13] and fallen weight deflectometers (FWD) to vehicles with lane-wide digital surface (e.g. video, LIDAR) and subsurface (e.g. GPR array) condition survey systems. Many of the currently commercially available systems can be operated within normal traffic flow, but are used only for inspection of localized areas with no repeat measurements in mind. A need for improved roadway and bridge deck inspection methods and devices using low-cost, faster, and easy-to-deploy sensor technology remains. In addition, a need to manage and jointly consider data from multiple service providers, locations, and inspection dates is highly desired.

In Section 6.4 a mobile multi-modal multi-sensor system is introduced that includes a number of vehicles of opportunity (VOO), which continuously monitor the network-wide health of roadways and bridge decks in an economic way. With the vehicle roaming embedded in the regular traffic, a continuous and non-disruptive measurement becomes feasible. The collected data are geo-spatially and temporally correlated at a central location and provides invaluable information for timely infrastructure investment decisions.

6.3   Sensor systems background

Sensor systems have emerged rapidly in recent decades to accommodate the growing needs of collecting, monitoring, analyzing, and responding to various physical sources and phenomena. It becomes more critical as sensing systems help improve the efficiency and accuracy by automating the process and eliminating human error compared to traditional human systems. WSN, and mobile WSN, is an active research topic related to the concepts and principles of the design we propose and has received much attention.[14-16] Proposed system designs and methodologies to address the inefficiency of heterogeneous sensor systems range from specific applications in personal area networks to generic scalable software architectures. In particular, the field of WSN in geo-sensors shares similar goals and targets with our design. For instance, Reference 17 has designed a data representation abstraction for large-scale geo-sensor data with a centralized monitoring system for air pollution. Even though work and research have been intensively explored, our system differs from them in sensor fusion and sensing topology.

Sensor fusion is a technique to combine multiple sources of sensory data to create and improve the result better than the effort of individual

sensors alone.[18] However, sensor fusion concepts and techniques have not received much attention in the geospatial domain. Geological data collected by roaming sensors to observe physical phenomena can be used to significantly help study, understand, and improve the geo-affected areas. For example, ground penetration radar systems and optical solutions can be equipped to use different wave lengths to evaluate the surface and subsurface conditions of the road.[19,20] Our design specifically targets the road surface and subsurface quality and health inspections for preventive maintenance, which could change completely the traditional method of US road system inspections, due to the innovative sensor fusion framework proposed in our design,[21] and would allow for the first time monitoring of the actual deterioration process. A variety of arrays of sensors, with very different specifications and characteristics, are deployed and their complementary data are combined to help evaluate the same objective with different sensing modalities. For instance, numerous microphones, accelerometers, trigger wheels, ground penetration radar devices, infrared sensors, inertial measurement units (IMUs), and global positioning systems (GPS) are all mounted on the designated vehicle to compose a roaming unit to collect data simultaneously. Our distributed heterogeneous architecture differs in comparison to similar approaches with respect to the number, types, and modalities of sensors. In Reference 22, only accelerometers are used to monitor the pothole on the ground. In Reference 23–25, only a limit number of homogeneous sensors are considered to fulfill the requirements of sensing one particular task. These approaches fail in accurately measuring the task quantitatively, and often are affected severely by the noise due to the lack of sensor variety.

As part of sensor fusion, multiple levels of data fusion take place, such as raw, intermediate and high, depending on the amount of processing power and data integration.[26] In our design, three processing levels are deployed, namely sensor, vehicle, and centralized managers; all are built with sensor fusion in different granularity. The raw data from multiple homogeneous sensors are collected to perform the first level of sensor fusion, raw data processing, and data reduction. Afterwards, necessary information is transmitted together on a vehicle, so that on-board heterogeneous sensor fusion is executed. Lastly, multiple vehicles transmit processed data back to the centralized manager for post-processing. This is the highest level of sensor fusion and data mining that can be performed with the aid of a geographic information system (GIS) backed by a geospatial and temporal database. During the three-fold processing strategy, plug-ins are used to carry out all the tasks algorithmically. No similar level of data fusion schemes has been developed or researched. Reference 22 has proposed the concept of pushing processing to different granularity; however, they in general do not have a concrete implementation, and no sensor fusion is considered to improve the results.

The sensing topology is defined as the structural hierarchy built to cope with the heterogeneous sensor framework for multi-tier processing and decision-making strategy. The communication infrastructure is essential as the backbone of the sensing topology to provide a robust, scalable, flexible and fault-tolerant distributed roaming system. In health care applications[27,28] presented their architectures with a centralized server managing multiple instances of mobile sensors. Reference 22 has designed a queuing software layer for transmitting data with the open Wi-Fi as the network media. Reference 16 has proposed a generic querying interface built on hardware directly, operating system (OS) dependent or application dependent. However, these designs do not all consider the combined effect of distributed computing environments with roaming heterogeneous sensors and various network media, which in turn makes the communication systems for their systems fairly simple and straightforward.

Due to the nature of a distributed roaming sensor system, the selection of multiple communication media is inevitable. Ethernet, Wi-Fi, and cellular network are incorporated into one design as the unified communication channel from vehicle managers to the centralized manager. Ethernet and Wi-Fi have higher bandwidth and lower latency, with a relatively stable network condition, but are not ubiquitously available citywide. In contrast, a cellular network has a fairly low speed and high latency, but good coverage in metropolitan areas. Therefore, Ethernet and Wi-Fi are used for bulk data handling, where large amounts of data need to be uploaded from vehicle to the server. Meanwhile, the cellular channel is used to dynamically monitor and control the vehicle in real time. The term 'real time' here is not defined by a preset numerical time constant, but instead by qualitative expressions such as 'instantaneously.' Hence, it leads the necessity to have a software layer that can communicate seamlessly across multiple platforms. A middleware framework has been adopted in our system to fulfill this requirement. The sensing topology determines that the middleware solution is most suitable and cost-efficient, given data marshaling/de-marshaling, Quality of Service (QoS), real-time features, etc.

Unlike other system frameworks, where mobile units are entirely autonomous and the centralized server is not responsible for managing or configuring these mobile units,[16,22] the centralized manager can be used to track, manage, control, and configure multiple vehicle managers. Moreover, it provides advanced features such as geo-referencing, sensor snapshot, and other operations on temporal and spatial data requiring real-time management with the sophisticated geospatial database. GIS data processing has been widely used in the field of geo-sensors[29,30] to help identify features and information with the aid of specialized GIS analysis features backed by a database. However, they do not all incorporate any real-time control and feedback mechanism, concentrating only on post-processing the related data.

6.4   VOTERS mobile sensor system overview

The VOTERS project (www.neu.edu/voters) provides a framework to complement periodical localized inspections of roadways and bridge decks with continuous network-wide health monitoring.[21] In order to enable continuous monitoring, VOTERS uses a set of sensors mounted on a fleet of vehicles called VOOs. VOOs are traffic-embedded, i.e. they will roam through daily traffic while collecting data. This eliminates hazardous, congestion-prone work zones that are typically set up to gather these critical inspection data sets. At the same time, utilizing VOOs for collection improves safety for the driving public and inspection personnel. Most importantly, using traffic-embedded VOOs opens the opportunity to monitor much more frequently larger stretches of the roadway infrastructure, which provides maintenance decision makers and researchers with a temporal and spatial data set not available with current inspection methods.

VOOs are considered mobile roaming sensor platforms that are driving around a city or other area, collecting inspection information as a by-product. Taxis, buses, postal service vehicles, city or private vehicles, transportation, or other fleet vehicles are all good candidates to be VOOs. Each of them would be equipped with an autonomous multi-modal multi-sensor (acoustic, electromagnetic, and optical) VOTERS sensing system (Section 6.4.1), which only requires power from the host vehicle. Each VOTERS system will be in communication with a control center via a wireless communication link (Section 6.5.1) while collecting data. Once returned to operation base, bulk-collected data will be transferred via a higher speed communication link, such as 802.11n.

This framework allows the collection of sensor data at traffic speeds that contain roadway and bridge deck surface and subsurface condition information. Over time the VOOs will achieve a continuous network-wide health monitoring of roadways and bridge decks. One key element is the accurate registration of all data streams in time and space (Section 6.7.1). The collected and/or on-board processed data will be transferred to the control center for further analysis, visualization, and decision making, i.e. prioritizing areas to be repaired. An important additional benefit will be the creation of time-lapse data sets that allow monitoring and analysis of the deterioration process over time, thereby providing experimental results that can be compared with and used to improve existing life-cycle models.[31–34]

The success of this concept will rely on the VOTERS sensing system being economical, light-weight, compact packaging, and non-interfering with the normal operation of the VOOs. It further depends on smart algorithms that can reduce the amount of data to be transferred from the VOOs to the control center at various levels throughout the system (Section 6.6), because the

bandwidth of current cellular technology is insufficient or cost-prohibitive for the large amount of data collected.

6.4.1   Multi-modal sensing

VOTERS has developed four new prototype sensing systems that collect data containing surface and subsurface (maximum of 1 m deep) condition information of roadways and bridge decks to locate and map defects at traffic speed, and is currently testing their performance. Each sensing system contains several multi-sensor systems in the acoustic, optical, and electromagnetic domains:

- Acoustic technology that uses tire-induced vibration and sound waves to determine surface texture and subsurface defects such as debonded asphalt layers. The waves are recorded with directional microphones[35–38] and a newly developed dynamic tire pressure sensor (DTPS).[39,40]
- Improved air-coupled GPR array technology that will map subsurface defects such as corroded rebar, trapped moisture, voids, and the pavement layers (thicknesses and electromagnetic properties).[19,20]
- Millimeter-wave radar technology (24 GHz) for the near-surface inspection of roadways and bridge decks focusing on pavement condition change detection and surface features covering the roadway, such as ice, water, or oil.[41]
- Video technology used to capture surface defects and automatically analyze any increase in defects over time.[42]

Prototypes of these sensor systems have been developed and deployed on the VOTERS prototype vehicle. Their design, specifications, and functionality are described in the references given above. Here the focus is on how these sensor systems are integrated into the overall mobile VOTERS sensing system, the types of data collected, how the data relates to the roadway and bridge deck defects we want to map, how their large size influenced the system design and architecture, and how they are routed through the VOTERS system.

6.4.2   Data types

In this multi-modal sensor system, special consideration has to be given to the type of data collected by the sensor system/domain. The type influences potential processing and storage requirements. Figure 6.2 depicts example sources for the data types considered in the framework.

Single Data Point: The simplest data type is a single data point per location x and time t. The spatial and temporal interval to the next data point depends on the trigger mode (distance or time) of the system, and can vary. An example of such a data stream is the amplitude of the millimeter-wave radar (Fig. 6.2d). Initial processing can be as simple as using a threshold or change detection, but also windowing, transformation to the frequency domain, and looking for a pattern in a finite subset of such a data stream.

Continuous Samples: Another data type is a continuously sampled sensor (time triggered only) e.g. a microphone that collects acoustic waves (Fig. 6.2c). A common way to process such data is to transform them into the frequency domain to perform an analysis, or to do a principle component analysis. In order to detect the differences between two frequency curves in Fig. 6.2c, one solution is to create a feature vector for each sample set using appropriate sampling strategies (e.g. uniform distributed sampling) to represent the data.

Distance Triggered Finite Length Series: A third data type is the collection of a finite length time series at distance triggered intervals (trace) e.g. as the GPR array sensor system (Fig.6.2b). GPR traces can be analyzed in one dimension through change detection from one trace to the next or by quickly inverting them for the underlying model (e.g., thickness d_i and electromagnetic properties of the roadway layers: conductivity σ and dielectric constant ε). A collection of GPR traces can also be simulated[43–46] and processed in two dimensions, looking for patterns in the *x-t* images with a variety of pattern recognition methods including artificial neural networks.[47–50,43–46]

2D Data: Video or image data (2D arrays) capturing a defined field of view is the fourth data type (Fig. 6.2e). Here, image processing algorithms can be applied to detect cracks.[42]

The diverse set of data types influence the system design in many aspects: synchronization across sensors, spatial relationship, sensor fusion processing/feature extraction, as well as in the form of centralized displaying the results.

6.4.3   Bulk data requirements

While a VOO is roaming through the traffic collecting data, the raw collected data need to be stored on the vehicle. Once returned to the vehicle base, the raw data will be uploaded to a central location for further processing. Preserving raw data streams enables repeatedly post-processing the valuable survey data for fusion/feature extraction algorithm development.
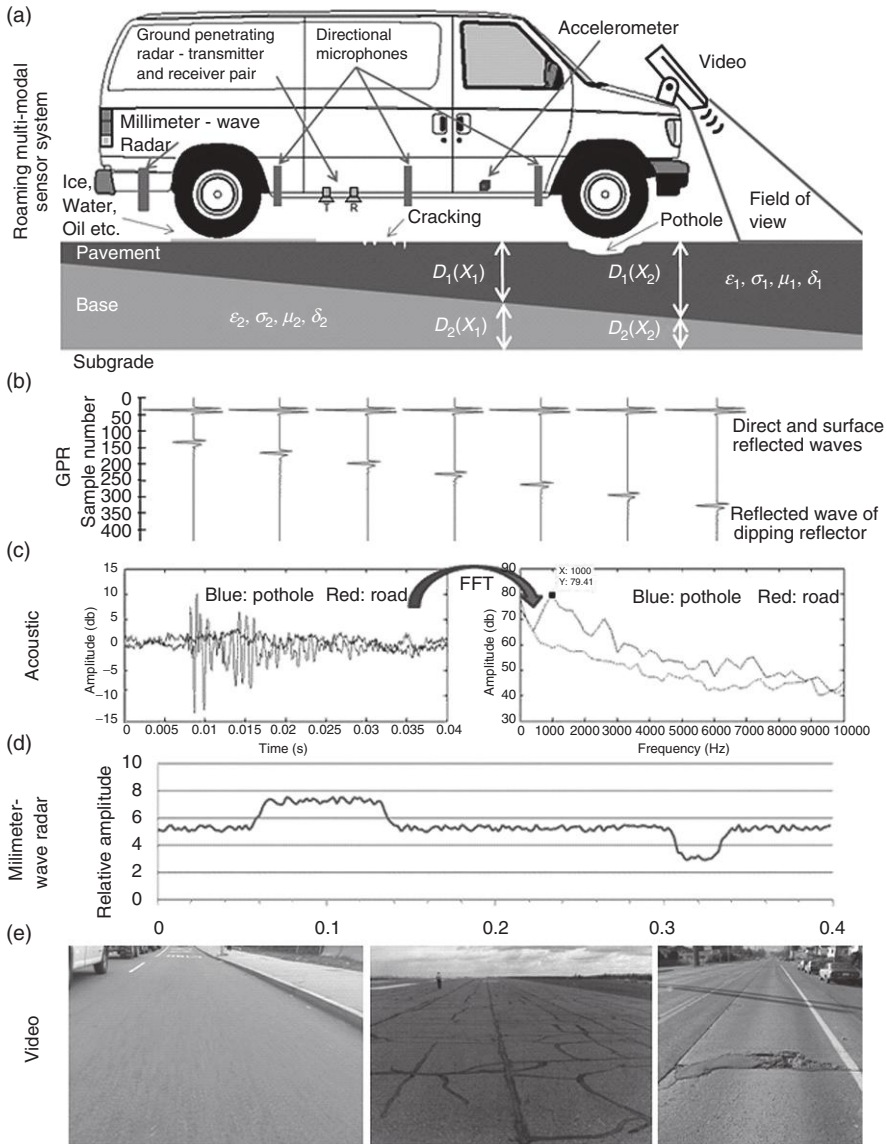
6.2 (a) Schematic of VOTERS roaming sensor system (van) equipped with an access point and multi-modal sensor systems. They collect data streams sensing roadways for surface and subsurface defects, layer thickness, and properties. Four different types of data are collected: (b) GPR traces (c) acoustic waves (d) millimeter-wave radar, and (e) video images.

At the same time, this increases the storage and data handling demands for the overall system. This section overviews data rates for each sensor and the complete system.

A VOO will collect sensor data from five different domains. Table 6.1 shows the data rates for these sensors. The most data intense sensors are the GPR array, with more than 300 GB/h, and the video camera, with more than 450 GB/h. Both sensors will be distance triggered to capture the same area regardless of the speed at which the VOO travels. The trigger distances are set based on coverage of the individual sensor allowing for 15% overlap between measurements. The highest rates are produced when the VOO travels at maximum speed of 100 km/h. Conversely, acoustic microphones, DTPS, and the millimeter-wave radar are time triggered. They collect at a sampling rate of 25 μs providing sufficient resolution even at maximum speed. Their collection speed is not reduced at lower velocities, since these sensors still lead to much smaller data rates. In addition to the sensors, the vehicle's position, inertia, and tilt information is recorded. The amount of data collected for position-related information, however, is negligible compared to the main sensor systems. In total, a single VOO produces 781 gigabytes per hour when driving at 100 km/h (highway speed).

The above calculations show the data volumes for a single VOO. To compute the demands for a complete fleet of five vehicles that concurrently roam an area of interest, we assume that each vehicle travels at 50 km/h on an average. We further assume that each vehicle collects 4 h of new data every day, as vehicles may not be utilized the whole time and may re-traverse sufficiently surveyed areas. With these assumptions, a VOTERS fleet produces about 8 TBytes/day. This demands significant efforts for storage and backup of the centralized survey data.

Table 6.1 Worst-case scenario of data rates per hour for VOTERS sensing system, assuming a maximum velocity of 100 km/h for the distance triggered systems

Sensor	Max sensor	Min trigger interval	Points/ sensor / trigger	Size/ point/ (byte)	Data rate (GB/h)
GPR array	16	0.01 m	1024	2	305.2
Acoustic microphones	4	25 μs	1	4	2.1
Dynamic tire pressure	2	25 μs	1	4	1.1
Millimeter-wave radar	10	25 μs	1	4	5.4
Video images	1	1 m	5018400	1	467.4
Positioning data	1	0.2 s	4	4	0.0003
Total					*781.1*

6.5   Hierarchical multi-tiered architecture

The previous section has introduced the VOTERS system as an overview with its heterogeneous sensors, the types of sensor data, and has described the large data demands. This section discusses how to address the challenges of complexity. It describes the hierarchical organization, the communication infrastructure, and introduces how to deal with large data streams.

6.5.1   Multi-tiered architecture

VOTERS employs a multi-tiered communications architecture to split the system complexity and responsibilities into separate hierarchy levels. Figure 6.3 shows a three-tiered communication architecture suitable for a fleet of vehicle-based sensor systems.

The topmost level in the VOTERS system, the first tier, deals with the whole fleet, at once offering a centralized control. The tasks at this tier are further split across two entities: MAP-GIS and Fleet Manager. Mapping and prognosis (MAP)-GIS is the interface to the collected data, offering off-line processing and data visualization. Conversely, the Fleet Manager performs the daily operative management of coordination of vehicles, status monitoring, and commanding of vehicles. Each vehicle in turn is controlled by a VOO Manager, which coordinates individual subsystems at the third tier in the system architecture. Each subsystem controls one or more sensors. High-data rate sensors (video and GPR) are managed by an exclusive subsystem, while low data rate sensors share a subsystem. The hierarchical structure allows a specialization for each tier in terms of communication structure, hardware, and processing. It further eases scalability in extending additional sensors or vehicles. With integrating multiple systems developed by different project partners, heterogeneous OS requirements come into place. As shown in Fig. 6.3, the VOO Manager runs under QNX OS, while other subsystems may run any combination of QNX, Linux, or Windows. QNX is selected for components with hard real-time requirements, whereas Windows is adapted for better graphic computation and driver support, and Linux is used as the generic low-cost OS for deployment.

The MAP-GIS on top of Fig. 6.3 is the portal for the VOTERS system, enabling end-users to access and review roadway condition information. It is designed to allow users to visualize their datasets and results in an intuitive, web-enabled, geo-referenced interface that supports their ability to make well-informed decisions. The communications between the clients and server are handled by the ArcGIS server software,[51] and clients connect to it via customized web applications, e.g. Silverlight.
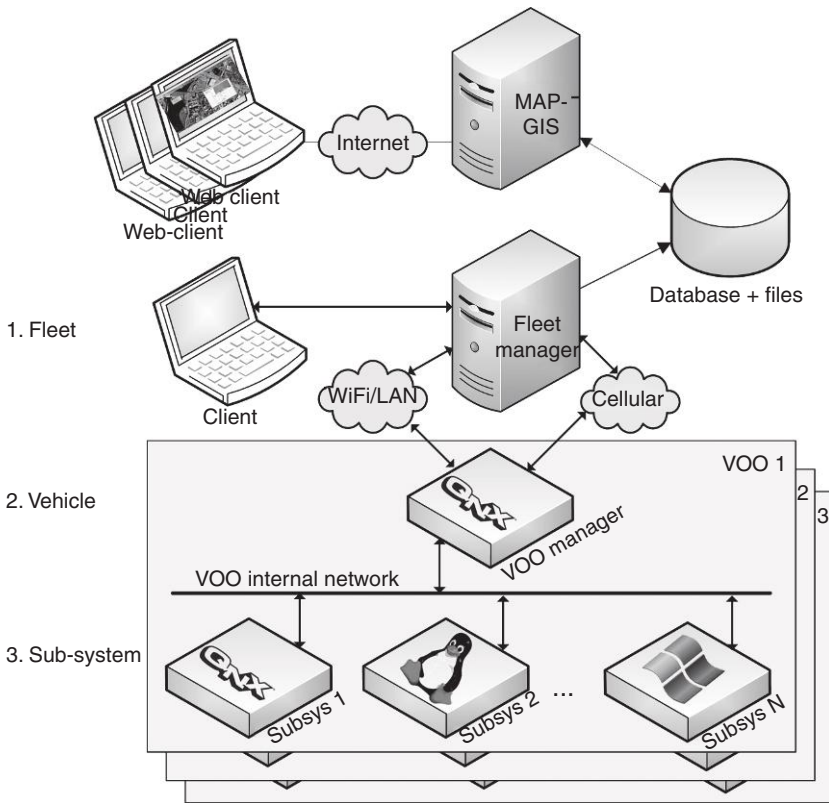
6.3 Multi-tiered communication architecture.

One of the advantages of a multi-tiered architecture is the distribution of responsibilities across levels. For example, the Fleet Manager communicates only with the VOO Manager on the vehicle, which in turn autonomously manages the communication with the subsystems (through the VOO internal network – a local area network (LAN)) and represents them unified to the outside. The VOO Manager is responsible for controlling the operations inside VOO as a self-operating unit. For example, the VOO Manager coordinates VOO-wide state transitions, e.g. as triggered by geo-fencing information. In result, a VOO can operate, after receiving the initial instructions, independently of an active Internet connection. This increases the robustness and reliability when there is no, or poor, outside network connection.

To further understand communication requirements, Fig. 6.4 shows the fundamental message types in the system. Three types of messages are designed to carry the overall communication between Fleet Manager and VOO Manager:
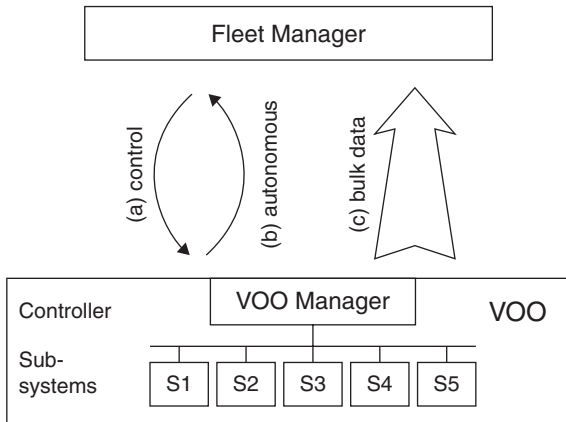
6.4 Communication message types.

1. Control messages are initiated by the Fleet Manager establishing two-way communication with the VOO for purpose of controlling, acknowledgement, and configuration. Examples of control messages include the configuration target surveys, geo-fencing information defining areas that require surveying, the request of snapshot data (currently measured data), and configuration changes to sensor systems inside the VOO.
2. Autonomous messages are sent by the VOO Manager to the Fleet Manager without a specific request. These messages include periodic status updates (e.g. location beacon), system warnings and errors, or warnings based on on-board data analysis (e.g. when black ice is detected).
3. Bulk data deals with uploading the survey data from the VOO to a centralized storage center after the VOO has returned to the central location. It is desirable that the bulk data are automatically uploaded when the vehicle has returned and is connected to a high speed communication link.

The demands on the communication framework differ by message type. Control and autonomous messages are typically small in volume. However, they have a large diversity, since they represent a wide set of features (will require data structures of many types to be transferred). Conversely, bulk data will deal with large data amounts (see Section 6.4.3), but with mostly identical data types.

6.5.2   Communication framework

The design of the communication framework to support a roaming sensor system such as VOTERS faces many challenges. On the one hand, diverse

data types have to be transmitted with a low latency for control and autonomous messages whereas, on the other hand, large volumes of bulk data have to be efficiently transferred where latency is less of a concern. Simultaneously, the system has to be tolerant to sustain autonomous operations when network connectivity is temporarily unavailable. Additional challenges arise from the need to support multiple OSs, as subsystem developers may prefer a particular OS. We generally target Linux; however, we also need to support Windows 7 due to driver availability, and QNX for systems with internal hard real-time requirements.

We recognize the differences in the types of messages to be transferred and thus target two specialized communication schemes. For dealing with control and autonomous messages, VOTERS utilizes distributed systems middleware, and for bulk data transfers it uses a dedicated file transfer protocol (FTP). Both communication schemes are seamlessly integrated into our overall communication framework. By utilizing object-oriented design principles and abstraction, VOTERS hides complexity from sensor developers.

Different options exist for realizing distributed systems middleware. Options range from manually implemented socket communication on top of transmission control protocol/Internet protocol (TCP/IP), and remote procedure calls (RPC) to dedicated object request brokers. For the realization of the VOTERS communication infrastructure, we selected the common object request broker architecture (CORBA)[53] implementation the Ace Orb (TAO).[54] In CORBA, communication interfaces are strictly defined in an Interface Definition Language (IDL), simplifying the management of large varieties of data types (important for control/autonomous messages). It combines messaging and thread management, and has options for real-time messaging and QoS, allowing us to prioritize important messages. It has been widely accepted in the telecommunications industry.

Figure 6.5 shows the software application stack for the VOTERS system. At its bottom (0), the OS (QNX, Linux, Windows) provides basic access to the network interfaces with user datagram protocol (UDP)/TCP and FTP protocols. On top of that, adaptive communication environment (ACE)[52] provides multi-tasking and timing primitives in an OS abstracted fashion. At the heart of the VOTERS system is the CORBA[53] implementation, TAO.[54] It provides object-oriented remote communication, allowing for easy construction of distributed systems. A local call (to a *client*) on one machine will be translated in one or more network messages which, transferred to a different machine, will invoke a server (who implements a *skeleton*). On top of TAO, the VOTERS base classes provide common implementations across all subsystems, such as status management, provisioning, and debug. Finally, the application on top only implements the system specific aspects,

VOO Manager

(4) Application
(3) VOTERS Base_Classes
(2) ORB: TAO Skel	Client
(1) ACE
(0) Operating system/drivers

Fleet Manager

(4) Application
(3) VOTERS Base_Classes
(2) ORB: TAO Skel	Client
(1) ACE
(0) Operating system/drivers

← UDP/TCP →
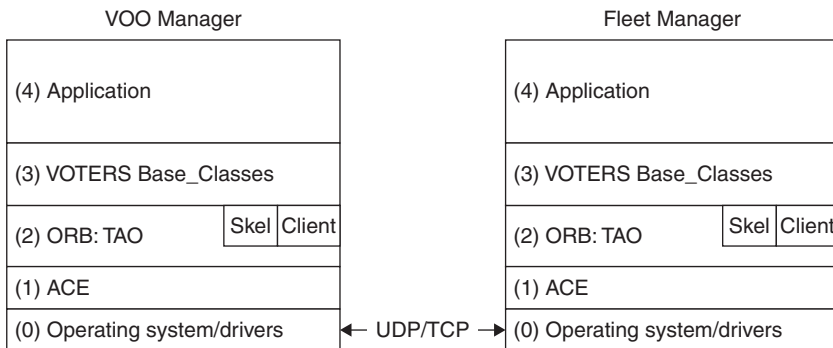
6.5 VOTERS software stack.

such as configuration or management of individual sensors. Utilizing a common software stack across all nodes in the distributed system makes the VOTERS system easier scalable and maintainable.

6.5.3 Communication media for control/autonomous messages and bulk data

For the selection of the communication media, we assume that the VOOs will be parked at a centralized *base*. They start operating from there to gather surveys. With this, two separate operation modes have to be considered: *at-base* when the vehicle is parked at its base station, and *roaming* when the vehicle is traveling either gathering a survey or traveling to/from the survey site.

When in *roaming* mode, it is desirable to control and observe the progress of vehicles, which demands a wireless connection. Cellular networks are suitable candidates, such as wideband code division multiple access (W-CDMA), 4G long-term evolution (LTE) or worldwide interoperability for microwave access (WiMAX). Using an existing cellular network dramatically simplifies deployment and reduces the effort. Hot-spot-based networks such as Wi-Fi are less suitable as their range is shorter and offers limited availability. In our roaming sensor system scenario, the bandwidth produced by all sensors is too high, with over 750 GBytes/h to be transmitted on any cellular network. Furthermore, the application does not demand full coverage real-time information as the observed processes (structural state of the infrastructure) change very slowly. We therefore opt to transmit only control and autonomous messages on a regular basis through the cellular connection, and keep the option of transmitting spot data on demand. This significantly reduces the network bandwidth requirements.

During *at-base* operation, the bandwidth requirements are immense. When a vehicle returns, the collected bulk data have to be uploaded to a centralized server. This requires high-bandwidth connections, but poses low requirements on latency. As the vehicle is stationary at this point, both wired connections and hot-spot-based Wi-Fi communication are suitable. Gigabit Ethernet is very suitable, as it provides sufficient bandwidth, around 100 Mbytes/s, at a very low cost. Further increasing the bandwidth for bulk data transmission is not meaningful, as it matches the bandwidth of affordable high-volume storage systems – especially considering if two to five vehicles upload simultaneously. Using a wired connection has benefits of simple scalability, low cost, and high bandwidth. However, it requires the vehicle operator to plug in the cable when returning to the base. A more seamless solution can utilize Wi-Fi technology (with 802.11n reaching 600 Mbit/s, or the planned 802.11ac being planned for >1 GBit/s). Even though the theoretical maximum network speeds are comparable to Gigabit Ethernet, actual data rates will be much lower, due to range and interference.

In our system, heterogeneous communication architecture is conceived and implemented by combining wired and wireless network media. During *roaming* operation a wireless connection (3G cellular) is used for control messages in a real-time fashion. Similar to other designs,[22,23] we only transmit essential and concise messages over the cellular network. In the next release, a 4G LTE module or a WiMAX component can be quickly integrated, as the underlying communication link is independent of the VOTERS software system. A wired Gigabit Ethernet connection is used during *at-base* operation, primarily for large volume bulk data transfer. In the future, we will look into Wi-Fi 802.11n to further reduce the effort of the driver for operating a VOO.

6.6   Bulk data handling

This section discusses how VOTERS organizes the bulk data being collected during the surveys. It first discusses processing levels, then overviews our hierarchical structuring of bulk data, and closes with a definition of plug-ins for processing the collecting data.

6.6.1   Processing levels

The large amount of data collected by each VOO, combined with the limited bandwidth cellular networks, prohibits uploading raw data to the Fleet Manager while a VOO is collecting survey data. As an initial stage we follow a store, upload, and post-process paradigm. The collected bulk data

Table 6.2 Processing levels

Level	Description	Real-time capability	Scope/fusion
1	Sensor subsystem processing	Possible	Single domain, single sensor
2	On-board processing on VOO controller	Time delayed	Multiple domains, multiple sensor, local geometry
3	MAP processing, off-board	Post-processing	Multiple domains, multiple sensor, local and global geometry, and multiple times (repeats)

are locally stored on the vehicle and uploaded to the Fleet Manager, once the VOO is parked, via a wireless or wired Ethernet connection. Then the MAP-GIS can post-process the data. There are multiple levels of processing possible (Table 6.2), which the MAP can perform. However, Level 1 and 2 processing can ultimately be performed on the VOO, as explained below.

The first level performs processing of raw data from a single sensor stream of a single domain. Such processing is usually very domain specific and does include only the knowledge of time relationship between data points. One can assume that no geometry information is available at this level, not even the relative offsets between sensors of the same domain.

The second level adds knowledge of the geometry, allowing for algorithms with spatial dependencies to operate on multiple channels (same domain) of a single sensor system, e.g. all channels of the GPR array. At this second level it is even possible to fuse data streams from multiple domains and extract knowledge analyzing them jointly.

The third level can be performed only at the MAP, as it requires that data from multiple domains are available, properly geo-registered for joint processing or comparison. In addition, the MAP will have access to data from multiple VOOs and data collected at the same location at different dates. This allows for the detection of changes to the roadway and bridge deck conditions over time using all available data.

The processing solution has to be built up over time; algorithms to analyze such data sets do not exist yet, because such data sets had not been acquired before the VOTERS project had completed its prototype. Once reliable processing algorithms have been developed, additional processing options to the off-board post-processing on the MAP will become available. Level 1 processing, as described above, can then be implemented in hardware or software on the sensor subsystem for immediate processing as the

data stream comes in. This has the potential of reducing the amount of data that needs to be uploaded to the MAP at the end of a day dramatically. In a similar way, Level 2 processing can be moved to the BOSS, as this VOO controller has access to the data from all sensor subsystems and the positioning system.

6.6.2   Bulk data hierarchy

Each subsystem within VOTERS contains one or more sensors. The raw data of the sensor, as well as processed data from one or more sensors, may be recorded as bulk data. To distinguish between different formats and semantic meanings of different bulk data, we define *streams*. A *stream* is a set of data in the same format and same semantic meaning. Note that one sensor can produce multiple streams. One stream can be the raw unprocessed stream, such as video images of the road surface. Another stream may be processed from the raw stream and show an image of detected cracks. Again, although both stem from the same sensor, they are considered separate streams.

In order to identify streams, stream types need to be defined. Each stream instance can then be associated with a stream type. This enables the MAP/GIS to identify how to handle or display the bulk data.

The stream type defines only the type of data. More information is needed to identify how the stream is captured. A stream can consist of one or more files. An example of a stream file is a single image of a video stream. Each image of a video can be stored separately (e.g. as a JPEG image). Multiple image files make up a stream. Proper definition of the streams needs to be complemented by definitions of the storage locations and automatically generated directory names, so that the upload of the bulk data can be triggered from the MAP and completed automatically. This requires several
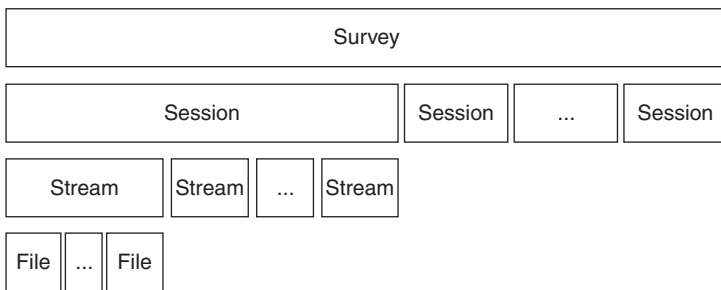
6.6 Bulk data hierarchy.
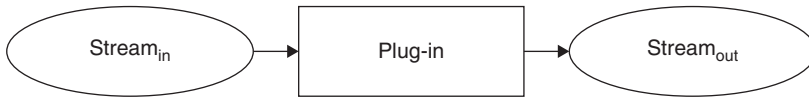
6.7 Plug-in processing streams.

additional definitions (Fig. 6.6) and the fully automated transfer of meta-data alongside the bulk data.

A *survey* defines a data collection run of the vehicle, from the time it leaves the parking location to its return. A survey may consist of one or more sessions. A *session* defines the contiguous recording in the same con-figuration. Therefore, a survey may consist of one or more sessions. A new session is started each time the system configuration changes, or after a brief stop of the recording (e.g. when passing over an area of non-interest or parking temporarily). Each subsystem is responsible for recording its own data. The base directory of data storage is defined throughout the VOTERS system, and each subsystem has to conform to these standard configurations to achieve maximum automation. The stream files will be located in a direc-tory that follows the naming convention: /var/voters/<VooNr>/<Survey>/< SessionNr>/<SubSystem>/<Sensor>/<Stream><StreamQual>.

Following a common guideline of storage location not only allows auto-mation of bulk data upload, it also simplifies debugging in field analysis as all systems will store data in the same manner.

6.6.3   Plug-ins for data processing

For displaying of streams, for processing of the raw data inside the streams, and for processing of processed streams, plug-ins need to be defined. Plug-ins will be designed to process from one stream type (e.g. unprocessed data) to another stream type (e.g. feature extracted data) while being executable at all processing levels as long as the input stream matches. The generic pro-cessing of streams through plug-ins is depicted in Fig. 6.7.

It is desirable to have the plug-ins developed in a way that they are able to run on-board, off-board and through the web-based interface to the MAP. This leads to the concept of a flexible plug-in, which operates on a well-defined data stream and outputs another well-defined processed data stream. This allows developing plug-in algorithms using recorded data at Level 3 executing on the MAP (FleetManager). With the maturing of the algorithm, and given the requirement, the plug-in can be moved to Level 2 or 1 for processing on the VOO or even the sensor subsystems. Implementations can be in software or hardware such as FPGAs or GPUs. To allow sensor fusion, a plug-in can natu-rally operate on multiple data streams. An example is illustrated in Fig. 6.8.
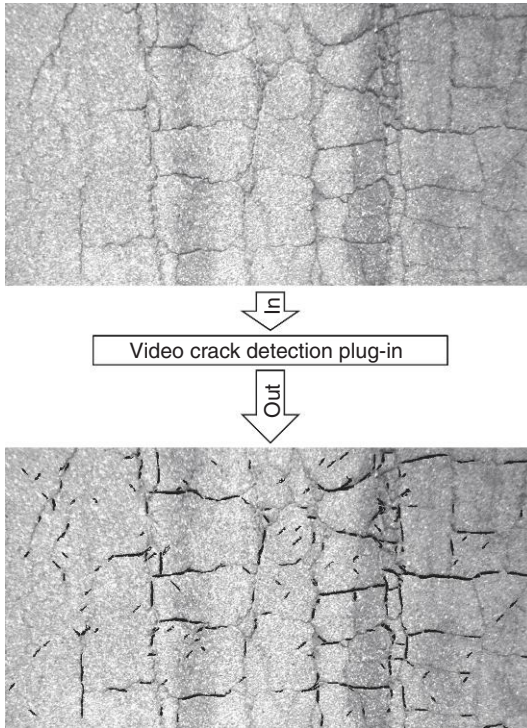
6.8 Example crack detection plug-in operating on black and white video image input stream and creating binary image data stream of detected cracks and a text file containing statistics on the detected crack classes, e.g. 326673 total pixels, 1171 pixels or 0.36% transverse cracks, 76 pixels or 0.02% longitudinal cracks, 1201 pixels or 0.37% transverse cracks, and 2448 pixels or 0.75% diagonal cracks.[42]

6.7   Enabling sensor fusion

The sensor systems are spatially distributed on the VOO and may have all very different triggering requirements or varying sampling intervals. Each sensor subsystem may be operated from a separate single board computer. The challenge is to design a system that allows for each data point to be geo-referenced as accurate as possible to allow for spatial and temporal comparison of different data sets (domains) from one VOO.

It is of critical importance to assign an accurate position to each data point. To achieve this, we rely on two facts: (a) the position of sensors relative to the VOO is known and (b) a tight time synchronization is in place. Then, we time stamp every data point collected at each sensor, as well as the positioning data stream recording the location of the vehicle. Then, the correlation between the sensor streams can be computed time delayed.

The synchronization of the sensor data streams with the positioning data requires the following features:

- Each VOO must have an accurate on-board positioning system. In our case we fuse a decimeter accuracy GPS, a distance measurement instrument (DMI) and an IMU to achieve this requirement.
- Each VOO must have an accurate world time. In our case this is accomplished with a GPS timing board. Other options are available.
- All computers on a VOO need to be time synchronized within a certain tolerance.
- All VOOs need to be time synchronized to the MAP within a certain tolerance.
- The location of each sensor on a VOO must be known with respect to the local positioning system.

6.7.1   Timing synchronization

In order to coordinate distributed subsystems to agree upon a unified notion of system time so that events and tasks can be recorded and executed in an ordered fashion, a synchronization scheme of the timing on each system in a VOO has to be created to propagate reference clock time within a certain range.

Options for time synchronization differ in cost and accuracy. Broadly, one can categorize synchronization approaches into hardware implementations, software implementations, and software implementations with hardware assistance. In Reference [55], hardware timing synchronization with a GPS timing board as a reference is introduced. Global time synchronization is achieved through GPS, and the reference time is distributed via a dedicated connection (single wire). Each subsystem requires a dedicated receiver. Local time synchronization achieves accuracies <1 ns. However, the cost with dedicated hardware is high; it also implies efforts for driver development. Pure software driven methods are widely used and are of research interest. Network Timing Protocol (NTP)[56] has been designed to compensate for clock drift and jitter on the Internet infrastructure. It allows synchronization across many subnets and routers. NTP achieves time synchronization with a few milliseconds accuracy. Precise Timing Protocol (PTP): IEEE 1588[57] has been developed to significantly improve the accuracy. Implementations of IEEE 1588 over Ethernet within a subnet have shown accuracies to the microsecond level.[58–60] PTP has been further developed to improve stability in network switching and routing, and PTP v2 was released in 2008.[61] Further accuracy improvements can be achieved by using hardware assistance, which involve network interface

cards (NIC) with hardware time stamping. They avoid jitter induced by software time stamping due to varying software loads. In result, accuracy in sub-microsecond level can be achieved using the hardware time stamping in NIC drivers[62] and Reference [63]. More and more NICs integrate hardware time stamping. With this, the additional cost of hardware support becomes negligible. This makes PTP with hardware assist a very interesting candidate for synchronization within a vehicle.

To understand the requirements to the timing synchronization, we now look at the demands for sensor fusion in the VOTERS application case. VOOs are designed to collect data at speeds up to $V_{max} = 100$ km/h. At the same time, the desired spatial correlation between two sensors in post-processing is 1 cm. This means that two sensors observing the same physical object should not be more than 1 cm apart. This 1 cm correlation requirement was derived internally from desired sensor resolutions.

With the maximum vehicle speed given, and the special correlation requirement between two sensors, the timing synchronization requirement can be computed (as we will use time stamps to subsequently correlate location):

$$t_{StampDeltaMax} = \frac{s_{CorrDelta}}{V_{max}} \qquad [6.1]$$

$$= \frac{0.01\,\text{m}}{100\,\text{km/h}} \qquad [6.2]$$

$$= 359\ \mu s \qquad [6.3]$$

Considering the desired spatial correlation between streams of $s_{CorrDelta} = 1$ cm and the maximum vehicle speed during recording $V_{max} = 100$, the jitter in time stamping data must not be larger than $t_{StampDeltaMax} = 359$ μs

Multiple errors may be introduced into the $t_{StampDeltaMax}$, e.g. the error in timing synchronization between systems in a VOO, and the timeliness of time stamping the data after they have been collected. In other words $t_{StampDeltaMax}$ indicates a maximum budget of jitter in time stamps. Timing synchronization between subsystems is only one aspect in the overall budget. In order to achieve time stamping with a jitter of less than 359 μs, the maximum difference between two subsystem clocks must be in the lower μs.

With accuracies in lower μs range, PTP becomes our candidate of choice for time synchronization. Some subsystems may have a lower resolution and subsequently pose less stringent final spatial correlation requirements. Those

systems may employ a second timing synchronization scheme (e.g. NTP) if drivers for PTP are not available. Nonetheless, the overall system requirement is constrained by the most critical timing demands.

In VOTERS timing solution, a local synchronization hierarchy with GPS as stratum N and the VOO Manager as stratum $N + 1$ is designed and illustrated in Fig. 6.9. The first step is that the VOO Manager acquires an accurate world time from the GPS receiver timing board, which becomes the common VOO time. There are only very loose requirements on the accuracy of the global synchronization and it is used only to correlate the date of time of recordings between vehicles. Conversely, strict timing requirements (low µs) exist for distributing the vehicle global time to all sensor subsystems. In this case, the common VOO time is essential for correlating sensor data from different subsystems. To accomplish this we use PTP. For subsystems that use an OS without PTP driver, we fall back onto NTP. In the current design, the video collection system has a second-level timing requirement and its OS does not support PTP (Windows 7). The PTP is used for all other subsystems to achieve required timing constraints.

Figure 6.10 show initial results of software-based time synchronization. A GPS timing board is used as the timing reference between the master node and the slave node. The result is obtained between the VOO Manager and a subsystem with PTP (V2) protocol on a 100 Mbps LAN for 5 h with injected traffic in Linux 2.6 without priority scheduling. We observed a maximum jitter of 12 µs with a standard deviation of 2.0375 µs. The observed jitter is small enough to achieve our triggering and time stamping requirements.
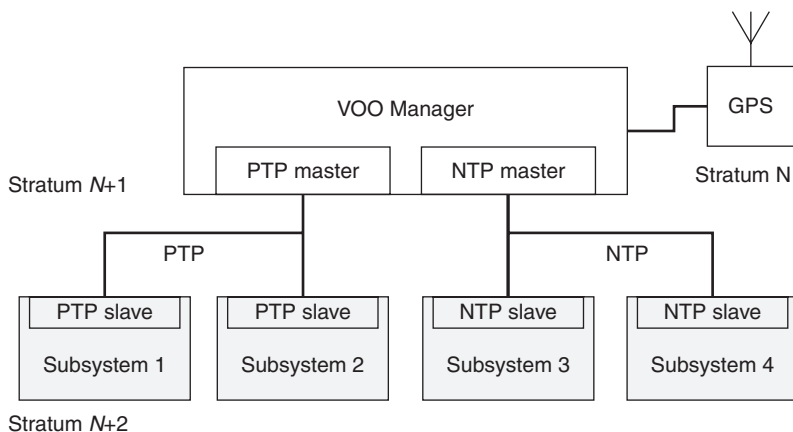
6.9 Software-based timing synchronization.
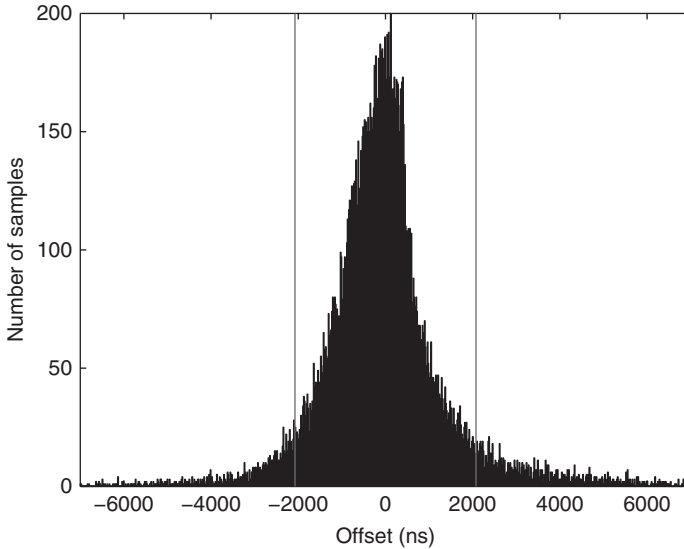
6.10 Results of software-based timing synchronization.

6.7.2   Sensor location calibration

All sensors on a single VOO are spatially distributed (Fig. 6.2). Since we synchronize all sensor data streams including the positioning data stream by accurate time stamping we need to correct for their actual locations in relationship to the positioning system, more specifically to the location of the GPS antenna.

This is accomplished by first defining the GPS antenna location on a VOO as the local coordinate system origin. Then we record the relative (x,y,z) location of each sensor in a geometry file during the physical installation of the sensors on the VOO. Now the geospatial location of each sensor can be calculated by first pulling the global position of the GPS antenna through the time synchronization of both sensor subsystems and then applying the local offset.

6.8   Conclusion

This chapter has introduced the challenges of a roaming multi-modal multi-sensor system that deals with the heterogeneity of sensors, sensor fusion, large amounts of bulk data, and high processing demands for data processing. The chapter introduced the case study of VOTERS – a sensing system to monitor the structural heath of roads and bridge decks. Based on the case study, the chapter introduced a hierarchical system architecture that splits responsibilities across three communication tiers, creating a flexible and

scalable system architecture. The chapter discussed strategies for handling of bulk data, organizing it hierarchically into surveys, sessions, streams, and files. It offered insight into the VOTERS approach to enabling sensor fusion with strict in vehicle time synchronization and known displacement of sensors.

The chapter has introduced the VOTERS project, which aims to assist in assessing the health status of the US road infrastructure. Through utilizing advanced sensor techniques, deployed onto VOOs, which are roaming traffic-embedded, a continuous network-wide road surface monitoring becomes feasible. It has outlined the basic framework and introduced major integration challenges and approaches. Initial demonstrations using a single VOO have been extremely promising showing the tremendous value of the data collected. We have demonstrated the value of our control and communication framework, which enables automatic collection and upload of bulk data, keeping the interaction with the vehicle operator to an absolute minimum.

With the outlined scalable architecture, we are looking forward to a rapid extension of our road surface database. We will build up our repository of data with the collected data runs. We are looking forward to more researchers who will assist in the evaluation of the recorded data and develop new data processing plug-ins.

6.9   Acknowledgements

6.10   References

1. '2009 Report Card for America's Infrastructure', ASCE, 2009.
2. '2008 Status of the nation's highways, bridges, and transit: conditions and performance', FHWA, 2008.
3. D04 Committee (2007), 'Practice for measuring delaminations in concrete bridge decks by sounding', ASTM International, ASTM D4580-03.
4. C09 Committee (2008), 'Test method for water-soluble chloride in mortar and concrete', ASTM International, ASTM C1218/C1218M-99.
5. C. L. Barnes and J.-F. Trottier (2004), 'Effectiveness of ground penetrating radar in predicting deck repair quantities', *Journal of Infrastructure Systems*, vol. **10**, no. 2, pp. 69–76, June 2004.
6. N. Gucunski, F. Romero, S. Kruschwitz, R. Feldmann, A. Abu-Hawash and M. Dunn (2010), 'Multiple complementary nondestructive evaluation technologies

for condition assessment of concrete bridge decks', *Transportation Research Record: Journal of the Transportation Research Board*, vol. **2201**, no. 1, pp. 34–44, December 2010.

7. F. Jalinoos and R. Arndt (2009), 'Structural health monitoring by periodic NDT: NDT for bridge maintenance', *Materials Evaluation*, vol. **67**, no. 11, pp. 1300–1307, November.

8. K. Maser (1991), 'Bridge deck condition surveys using radar: case studies of 28 New England decks', *Transportation Research Record*, no. 1304.

9. K. Maser, J. Doughty and R. Birken (2011), 'Characterization and detection of bridge deck deterioration', in *Proceedings of the Engineering Mechanics Institute*, Boston, MA.

10. N. Gucunski, R. Feldmann, F. Romero, S. Kruschwitz and H. Parvardeh (2010), 'Comprehensive condition assessment of bridge decks by multimodal NDE', in *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2010,* March 8, 2010–March 11, 2010, San Diego, CA, United states, 2010, vol. 7649, p. The Society of Photo–Optical Instrumentation Engineers (SPIE); American Society of Mechanical Engineers.

11. K. Vaghefi, R. C. Oats, D. K. Harris, T. (Tess) M. Ahlborn, C. N. Brooks, K. A. Endsley, C. Roussi, R. Shuchman, J. W. Burns and R. Dobson (2012), 'Evaluation of commercially available remote sensors for highway bridge condition assessment', *Journal of Bridge Engineering*, vol. **17**, no. 6, pp. 886–895, November 2012.

12. S. Yehia, O. Abudayyeh, S. Nabulsi and I. Abdelqader (2007), 'Detection of common defects in concrete bridge decks using nondestructive evaluation techniques', *Journal of Bridge Engineering*, vol. **12**, no. 2, pp. 215–225.

13. K. R. Maser and A. Rawson (1993), 'Network bridge deck surveys using high-speed radar: case studies of 44 decks', Presented at the Bridge Management 2: Inspection, Maintenance Assessment And Repair. *Papers Presented at the Second International Conference on Bridge Management Held 18–21 April 1993*, University of Surrey, Guildford.

14. J. Eliasson, C. Zhong and J. Delsing (2011), 'A heterogeneous sensor network architecture for highly mobile users', in *Wireless Communication and Sensor Networks (WCSN), 2010 Sixth International Conference on*, pp. 1–6.

15. W. Chen, D. Wei, X. Zhu, M. Uchida, S. Ding and M. Cohen (2005), 'A mobile phone-based wearable vital signs monitoring system', in *Computer and Information Technology, 2005. CIT 2005. The Fifth International Conference on*, pp. 950–955.

16. V. Casola, A. Gaglione and A. Mazzeo (2009), 'A Reference Architecture for Sensor Networks Integration and Management', *GeoSensor Networks Proceedings Third International Conference, GSN 2009*, pp. 158–168.

17. Y. J. Jung, Y. K. Lee, D. G. Lee, K. H. Ryu and S. Nittel (2008), 'Air pollution monitoring system based on geosensor network', in *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008.IEEE International*, vol. **3**, pp.1370–1373.

18. W. Elmenreich (2002), *Sensor Fusion in Time-Triggered Systems*, Dissertation, University of Vienna, pp. 157.

19. C. Oden and R. Birken (2012), 'A GPR array system for traffic embedded monitoring of bridges and roadways', *Symposium on the Application of Geophysics to Engineering and Environmental Problems*, vol. **25**, no. 1, pp. 123–123, January 2012.

20. C. Oden and R. Birken (2011), 'Array GPR system for continuous monitoring of bridges and roadways', in *Proceedings of the Engineering Mechanics Institute*, Boston, MA.

21. R. Birken, M. L. Wang and S. Wadia-Fascetti (2012), 'Framework for continuous network-wide health monitoring of roadways and bridge decks', in *Proceedings of Transportation Systems Workshop*, 5–8 March, Austin, TX, CD-ROM.

22. J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden and H. Balakrishnan (2008), 'The pothole patrol: using a mobile sensor network for road surface monitoring', in *ACM MobiSys*.

23. B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan and S. Madden (2006), 'CarTel: a distributed mobile sensor computing system', in *Proceedings of the 4th international conference on Embedded networked sensor systems*, pp. 125–138.

24. E. Kanjo, S. Benford, M. Paxton, A. Chamberlain, D. S. Fraser, D. Woodgate, D. Crellin and A. Woolard (2007), 'MobGeoSen: facilitating personal geosensor data collection and visualization using mobile phones', *Personal and Ubiquitous Computing*, vol. **12**, no. 8, pp. 599–607, August. 2007.

25. B. Jiang, A. P. Sample, R. M. Wistort and A. V. Mamishev (2005), 'Autonomous robotic monitoring of underground cable systems', in *Advanced Robotics, 2005. ICAR'05. Proceedings. 12th International Conference on*, pp. 673–679.

26. D. L. Hall and J. Llinas (1997), 'An introduction to multisensor data fusion', *Proceedings of the IEEE*, vol. **85**, no. 1, pp. 6–23, January. 1997.

27. A. Mughal, M. Kanjee, and H. Liu (2010), 'Mobile healthcare infrastructure with QoS and security', *Mobile Wireless Middleware, Operating Systems, and Applications*, pp. 462–473.

28. R. Chakravorty (2006), 'A programmable service architecture for mobile medical care', in *Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006*, p. 5 pp.–536.

29. B. Resch, M. Mittlboeck, F. Girardin, R. Britter and C. Ratti (2009), 'Real-time geo-awareness–sensor data integration for environmental monitoring in the city', in *Advanced Geographic Information Systems & Web Services, 2009. GEOWS'09. International Conference on*, 2009, pp. 92–97.

30. K. Aberer, S. Sathe, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings and L. Thiele (2010), 'Opensense: open community driven sensing of environment', in *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming*, pp. 39–42.

31. L. Galehouse, J. S. Moulthrop and R. G. Hicks (2003), 'Principles of pavement preservation: definitions, benefits, issues, and barriers', *TR News*, no. 228, September 2003.

32. L. B. Stephens and United States (1985), Office of Highway Planning, *Road Surface Management for Local Governments : Resource Notebook : Final Report*. Washington, D.C.: Federal Highway Administration.

33. J. Walls and M. Smith (1998), 'Life-cycle cost analysis in pavement design', FHWA report FHWA-SA-98-079.

34. D. M. Frangopol and M. Liu (2007), 'Maintenance and management of civil infrastructure based on condition, safety, optimization, and life-cycle cost∗', *Structure and Infrastructure Engineering*, vol. **3**, no. 1, pp. 29–41.

35. V. V. Saykin (2011), 'Pavement macrotexture monitoring through sound generated by the tire-pavement interaction', Northeastern University.

36.   Y. Zhang, J. Gregory McDaniel and M. L. Wang (2013), 'Estimation of pavement micro-texture with acoustic measurement through principal component analysis', *Journal of Transportation Engineering,* 10.1061/(ASCE)TE.1943-5436.0000617, 04013004..

37.   Y. Zhang, X. Ma, J. G. McDaniel and M. L. Wang (2012), 'Statistical analysis of acoustic measurements for assessing pavement surface condition', *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. **8347**, p. 36.

38.   Y. Lu, Y. Zhang, Y. Cao, J. G. McDaniel and M. L. Wang (2013), 'A mobile acoustic subsurface sensing (MASS) system for rapid roadway assessment', *Sensors*, vol. **13**, no. 5, pp. 5881–5896.

39.   Q. Wang, J. G. McDaniel, N. X. Sun and M. L. Wang (2013), 'Road profile estimation of city roads using DTPS', *Proc. SPIE 8692, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, vol. **8692**, pp. 86923C-1.

40.   Q. Wang, Y. Zhang, N. X. Sun, J. G. McDaniel and M. L. Wang (2012), 'High power density energy harvester with high permeability magnetic material embedded in a rotating wheel', *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 83470V–83470V–6.

41.   D. Vines-Cavanau, D. Busuioc, R. Birken and M. Wang (2012), 'Millimeter-wave nondestructive evaluation of pavement conditions', *Presented at the SPIE Symposium on Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring*, p. 83472B–83472B–8.

42.   S. Ghanta, R. Birken and J. Dy (2012), 'Automatic road surface defect detection from grayscale images', *Proceedings of the SPIE – The International Society for Optical Engineering,*, vol. **8347**, 12 pp..

43.   K. Belli, C. M. Rappaport, H. Zhan and S. Wadia-Fascetti (2009), 'Effectiveness of 2-D and 2.5-D FDTD ground-penetrating radar modeling for bridge-deck deterioration evaluated by 3-D FDTD', *IEEE Transactions on Geoscience and Remote Sensing*, vol. **47**, no. 11, pp. 3656–3663, November 2009.

44.   B. K, W.-F. S, and R. C, 'Model based evaluation of bridge decks using ground penetrating radar', *Computer-Aided Civil and Infrastructure Engineering*, vol. **23**, no. 1, pp. 3–16.

45.   G. R. Olhoeft and S.S. Smith III (2000), 'Automatic processing and modeling of GPR data for pavement thickness and properties', *Proceedings of SPIE – The International Society for Optical Engineering*, vol. **4084**, pp. 188–193.

46.   M. H. Powers and G. R. Olhoeft (1995), 'GPRMODV2; one-dimensional full waveform forward modeling of dispersive ground penetrating radar data, version 2.0', U.S. Geological Survey, OFR – 95–58.

47.   N. .Attoh-Okine and N. O. Busby (1995), 'Use of artificial neural networks in ground penetrating radar applications in pavement evaluation and assessment', *Proceedings International Symposium on Non-Destructive Testing in Civil Engineering (NDT-CE)*, pp. 93–100.

47.   R. S. Freeland (2007), 'Subsurface characterization using textural features extracted from GPR data', in *Transactions of the ASABE*, vol. **50**, no 1, pp. 287–293.

49.   M. Heiler, S. McNeil, and J. Garrett Jr. (1995), 'Ground-penetrating radar for highway and bridge deck condition assessment and inventory', in *Nondestructive Evaluation of Aging Bridges and Highways, 6 June 1995–7 June 1995*, Oakland, CA, USA, 1995, vol. 2456, pp. 195–206.

50. S. Moysey, R. J. Knight and H. M. Jol (2006), 'Texture-based classification of ground-penetrating radar images', *Geophysics*, vol. **71**, p. 111.

51. 'ArcGIS (2012) – Mapping and Spatial Analysis for Understanding Our World.' Online]. Available: http://www.esri.com/software/arcgis. (Accessed: 29 August 2012).

52. Object Management Group (1995), *The Common Object Request Broker : Architecture and Specification.* (Framingham, Mass.): Object Management Group.

53. 'TAO: The ACE ORB.' (2012) [Online]. Available: http://www.cs.wustl. edu/~schmidt/TAO.html. (Accessed: 30 August2012).

54. D. C. Schmidt (1993), 'The ADAPTIVE Communication Environment: An Object-Oriented Network Programming Toolkit for Developing Communication Software', pp. 214–225.

55. D. A. Grejner-Brzezinska, R. Da, and C. Toth (1998), 'GPS error modeling and OTF ambiguity resolution for high-accuracy GPS/INS integrated system', *Journal of Geodesy*, vol. **72**, no. 11, pp. 626–638.

56. D. L. Mills (1991), 'Internet time synchronization: the network time protocol', *IEEE Transactions on Communications*, vol. **39**, no. 10, pp. 1482–1493.

57. 'IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems (2002)', *IEEE Std 1588–2002*, pp. i –144.

58. J. Eidson, J. Mackay, G. M. Garner and V. Skendzic (2007), 'Provision of precise timing via IEEE 1588 application interfaces', in *IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, 2007. ISPCS 2007*, pp. 1–6.

59. T. Neagoe, V. Cristea and L. Banica (2006), 'NTP versus PTP in computer networks clock synchronization', in *2006 IEEE International Symposium on Industrial Electronics*, vol. **1**, pp. 317–362.

60. A. Machizawa, T. Iwawma and H. Toriyama (2008), 'Software-only implementations of slave clocks with sub-microsecond accuracy', in *Precision Clock Synchronization for Measurement, Control and Communication, 2008. ISPCS 2008. IEEE International Symposium on*, pp. 17–22.

61. 'IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems' (2008), *IEEE Std 1588–2008 (Revision of IEEE Std 1588–2002)*, pp. c1–269.

62. P. Loschmidt, R. Exel, A. Nagy and G. Gaderer (2008), 'Limits of synchronization accuracy using hardware support in IEEE 1588', in *IEEE International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, 2008. ISPCS 2008*, pp. 12–16.

63. P. Ohly, D. N. Lombard and K. B. Stanton (2008), 'Hardware Assisted Precision Time Protocol. Design and case study', *Proceedings of LCI International Conference on High-Performance Clustered Computing*. Urbana, IL, USA: Linux Cluster Institute, vol. **5**, pp. 121–131.

This page intentionally left blank

Part II

Case studies in assessing and monitoring
specific structures

This page intentionally left blank

7
Sensing solutions for assessing and monitoring bridges

J. M. W. BROWNJOHN, K-Y. KOO and N. DE BATTISTA,
University of Exeter, UK

Abstract: Bridges are popular candidates for condition assessment and long-term monitoring because they are critical infrastructure components, while being accessible, visible and usually publicly owned and operated. They represent the widest application of sensing technology, from traditional, mature and reliable to novel, experimental and (perhaps) less reliable. Being able to trial the novel alongside the mature provides valuable opportunities to advance the field of sensor technology for all types of structure so that technologies viewed as experimental 10 or 20 years ago are now reliable and commercially viable. The sensors described here are used to sense directly or indirectly both environmental and loading parameters and to measure directly the response through various measures of deformation at various timescales.

Key words: bridge, wireless, performance monitoring instrumentation.

7.1   Introduction

Condition assessment and structural performance (or health) monitoring are two types of investigation used in the lifecycle management and operation of bridges.

Condition assessment is relatively routine and predominantly involves visual inspection with desk studies based on guidance documents (The Highways Agency, 1995), but can also on rare occasions include various forms of non-destructive evaluation (NDE) technology such as ground penetrating radar (GPR), impact echo (IE) and imaging (Gucunski *et al*., 2010). These are intensive short-term campaigns that evaluate the state of a bridge at an instant in time. A different form of condition assessment involves controlled load testing with heavy vehicles, e.g. for direct evaluation of load capacity (Bakht and Jaeger, 1990) or dynamic testing to evaluate modal flexibility (Brownjohn *et al*., 2003; Catbas *et al*., 2006) leading to model calibration and indirectly to capacity estimation. Such exercises are usually heavily manned and subject to severe time constraints.

207

On the other hand, monitoring implies continuous automated measurements of time series representing loading and response over periods of time that can vary from days to a lifetime, the primary aim being to establish statistical properties and performance bounds of loading and response, and characterise their relationships. Out-of-bounds performance and variations in, and deviations from, these relationships can be used to signal and diagnose performance problems.

Some of the same sensing technologies used in monitoring are also used in assessment exercises, for example measurements of strain and deflection during load testing, and measurements of vibrations during a modal survey. The most elaborate monitoring systems (e.g. for Stonecutters Bridge, (Wong and Ni, 2009)) can also perform continuous condition assessment due to density, comprehensive coverage of the instrumentation and direct links to 'physics-based' analytical models.

A good example of bridge assessment where the widest possible range of NDE, visual inspection and advanced sensing solutions has been applied together is the International Bridge in Wayne, New Jersey, tested in 2010 and 2011 by multiple international research agencies (http://cait.rutgers.edu/ibs-2011-agenda) as part of a study related to the US Federal Highway Administration Long Term Bridge Performance Programme. In this study, while time series were collected, all measurement activities were manned and variation of performance over time was not studied. In particular the NDE inspection processes are intensively manual and there is no opportunity for automation.

Apart from NDE and visual assessment, sensing technologies used in bridge assessment are a subset of those used in long-term monitoring exercises. Hence, this chapter focuses on sensing technologies used in monitoring where both serial and analogue data are collected autonomously at a range of fixed and variable sample rates.

7.2   Performance metrics or measurands and their uses in assessment

The range of parameters measured covers both direct measures of response (e.g. deflection) and indirect measures of loading (e.g. wind speed). Only in exceptional cases is it possible to measure directly external loading on a structure that drives response; even for earthquakes it is problematic to capture the true 'free-field' ground motions.

Deformations in response to loading can be observed directly in the form of absolute and relative position and derivatives, which are commonly measured as displacement, velocity, acceleration and inclination. Acceleration and angle are measured with respect to a global inertial frame of reference, whereas displacement and velocity are relative to a fixed local reference.

Global positioning system (GPS) is the only possible exception, but in its usual 'differential' mode (Casciati and Fuggini, 2009; Meng and Huang, 2009) a local reference (base station) with known location is required to remove errors common to receivers in the area of the structure.

There are myriad forms of accelerometer but all depend on motion or tendency for motion of a proof mass (tending to remain at rest) relative to a fixture that is forced to accelerate with the structure. The choice of instrument depends on budget along with frequency range and resolution, given that bridge vibration modes of interest can span frequencies from 0.05 to 50 Hz and range from 1 µg to 0.5 g. For civil structures 'seismic' grade piezo-electric accelerometers and force balance servo-accelerometers are popular (Brownjohn, 2007), although low-cost and low power micro-electro-mechanical systems (MEMS) devices are increasingly viable for dense and wireless arrays as performance specifications such as noise and resolution improve (Jo *et al.*, 2011).

For displacement there is a wide range of different technologies. As well as GPS, these include mechanical devices, e.g. linear variable differential transformer (LVDT) and pull-wire devices requiring direct contact (Brownjohn *et al.*, 1994; Nassif *et al.*, 2005), a plethora of devices depending on electromagnetic radiation such as natural, light emitting diode (LED) and laser light and radar.

Strain is a common derived response parameter, sensed by a wide range of technologies including the strain-dependent resistor (which is also dependent on temperature) and the robust (but slow sampled) vibrating wire gauge (Barr *et al.*, 2004). Strain gauges using fibre optic technology and operating at a range of length scales and spatial densities have the advantages of resilience to water and electromagnetic interference (Glisic and Inaudi, 2008). Direct and mechanical measurement of absolute stress is relatively rare, but devices sensing internal axial forces can be used, such as the new Austrian tunnelling method (NATM) pressure cells (Brownjohn and Moyo, 2001). Non-contact devices for sensing cable tensions (Wang *et al.*, 2006) have recently been used for cable-supported bridges.

Major forms of loading derive from weather conditions, resulting in pressure loads and internal temperature changes leading to deformations. Sensors for weather and environment are well developed and borrowed from meteorology, the most usual form being wind speed and direction sensors (Xu and Xia, 2012), both mechanical (cup and vane or propeller) and without moving parts (Pitot tube and sonic devices). While direct pressure measurements are common in wind tunnel testing, they are very rare at full-scale (Frandsen, 2001; Isyumov *et al.*, 1984).

Temperature is a major environmental factor on performance both globally and locally, with sensing technology including thermistors, thermocouples, vibrating wire devices and fibre optics.

Service loading derives from vehicles (cars, trucks, trains), fluids (aqueducts and pipelines) and pedestrians (foot bridges). For highway bridges, weigh-in-motion (WIM) is occasionally used to capture snapshots of vehicle weights (Cebon, 1991) by measuring individual axle forces when the vehicle passes over the WIM location, while optical systems can be used for automated or manual vehicle counting and recognition (Fraser *et al.*, 2011), a technology that has applications on footbridges.

While WIM, toll and optical information contribute to a full picture of vehicle loads, direct observation of deformation in long span bridges can be used to infer actual loads in transit.

7.3   Instrumentation in notable bridge monitoring projects

Table 7.1 presents information on notable major bridge monitoring projects, dating back to Tacoma Narrows. There many other sources of information, such as proceedings of Structural Health Monitoring of Intelligent Infrastructure (SHMII) conferences, Encyclopedia of Structural Health Monitoring (SHM), books, e.g. Xu and Xia (2012) and specific papers such as (Ko and Ni, 2005) and books, e.g. (Andersen and Fustinoni, 2006).

With the implication that monitoring means automated recording of time series over long periods of time, monitoring effectively began with introduction of instrumentation for recording (rather than just observing) motion, such as the seismograph, whose invention is usually credited to John Milne in 1880, although there were several less successful earlier designs, e.g. (Mallett, 1846).

The first well-documented monitoring of bridges was a series of one-day seismograph recording exercises on San Francisco·Oakland Bay and Golden Gate Bridges from 1934 to 1937 by Dean Carder of the US Coast and Geodetic Survey (Carder, 1937), and the second was the motion-picture recording of the original Tacoma Narrows Bridge (University of Washington, 1954) before its collapse in November 1940.

Certainly many bridges have been equipped with permanent seismometer arrays in seismic regions, particularly in California (Table 7.1 provides a few examples) but our focus is on more comprehensive instrumentation arrays for recording normal operational loading and performance as well as the extremes. Except for the effects of differential support, in the case of long span bridges earthquakes are probably less of a threat to bridge superstructures than strong winds, for which more exotic forms of instrumentation have been deployed. In particular, long span bridges have usually been instrumented to study problematic wind-induced vibrations, such as Pasco·Kennewick (Bampton *et al.*, 1983) and Deer Isle·Sedgwick Bridges (Bampton *et al.*, 1986) in USA and Great Belt Bridge (Frandsen, 2001) in Denmark, for which

Table 7.1 Notable major bridge monitoring projects

Bridge	Span	Built	Instrumented	Strain	Displacement	GPS	Wind	Acceleration	Other motion	Temperature	Inclination	Other
Tacoma Narrows (University of Washington, 1954)	853	1940	1940: 4 months		Motion picture							
Pasco-Kennewick (Bampton et al., 1983)	229	1978	1978: 2 years				6 × triax prop	15 × servo				
Deer Isle-Sedgewick (Bampton et al., 1986)	330	1939	1981:3 years				6 × triax prop + 2 × skyvane	12 × servo			1	
Vincent Thomas Bridge (Yun et al., 2008)	457	1964	Mid 1980s					26				
Golden Gate Bridge (Pakzad et al., 2009)	1280	1937	2006 only					56–64				
New Cape Girardeau Bridge (Celebi, 2006)	2003	350	2005					84 in structure and ground				
New Carquinez Bridge (Kuruta et al., 2012)	2003	728	2010		3 × ext		3 × speed, direction	70 CSMIP, 66 wireless		3 × air temperature, humidity		
Confederation Bridge (Cheung et al., 1997)	11 × 250	1997	Construction	Mechanical + VWG	55 × survey			76		50 × thermocouple		Ice force, corrosion
Commodore Barry (Catbas and Aktan, 2009)	501	1974		54 × ERSG + 148 × VWG	17 × crackmeter		2 × sonic + radiation, rain, RH	16 × capacitive		201 × thermistor	36 × VW tiltmeter	2 × WIM
Tuas Second Link (Brownjohn and Moyo, 2001)	92	1997	Construction	12 × VWG, 12 × VWG stress				1 × triaxial		44 × thermocouples		

(Continued)

Table 7.1 Continued

Bridge	Span	Built	Instrumented	Strain	Displacement	GPS	Wind	Acceleration	Other motion	Temperature	Inclination	Other
Great Belt (Frandsen, 2001)	1624	1998	1998 (6 weeks)				1 × cup + vane	2				Wind pressures
Foyle (Sloan et al., 1992)	233	1984	1990/191	8 × ER	2 × laser + 2 × ext		Anemometer			6 × thermocouples		
Humber (Brownjohn et al., 1994)	1410	1981	1990/1991		6 × ext and optical		13 × speed, 15 × dirn, 1 × lift	12 × servoc, 4 × piezo		2 × thermistor	5	
Second Severn Crossing (Macdonald et al., 1997)	528	1996	Construction		3 × optical		1 × sonic	8 × servo-acc		5 × thermistor		
Nahmae (Koh et al., 2009)	404	1973	1998	10 × ERS, 44 × VWG	4 × ext		2 × triax prop	2 × triax, 12 × uniax			12 × tiltmeter	
Jindo (Koh et al., 2009)	344	1984	1998	38			2	5 × cap, 15 × piezo, 2 × seis			4 × tiltmeter	
Sohae (Koh et al., 2009)	470	1993		12 × static, 82 × dynamic	10 × ext, 4 × laser		2 × anemometer	10 × accelerometer, 2 × seismometer	24×acc on cable	14	6 × tiltmeter	
Gwangan (Koh et al., 2009)	500	1994			3 × ext, 1 × laser		3 × prop, 1 × sonic	20 × accelerometer, 2 × seismometer		14	4 × tiltmeter	
Youngjong (Koh et al., 2009)	300	2000	2001	122 × static, 175 × dynamic	4 × ext, 3 × laser		4 × anemometer	3 × triax, 14 × biax, 12 × uniax		33	10 × tiltmeter	
Second Jindo Bridge WSN (Jo et al., 2011)	344	2006	2010				1 × triaxial sonic	triax: 10 × high res, 100 × low res		3		Light sensor
Jiangyin (Ko and Ni, 2005b)	1385	1998	1998		Total station			72				14 × EM cable stress, 12 load pins,

Jiangyin (Zhou et al., 2006)	1385	1998	2005			8	2 × sonic	36		116 × FBG
Binzhou Yellow River (Li et al., 2006)	300	2003	2003	Multiple FBG fibre optic		4	1 × sonic, 1 × propellor	46		
Sutong Yangtze	1088	2008	Construction	211	96 × ext	5	4 × 'aerovane'	28		169 × structure + 12 × air 8
Runyang Yangtze South Bridge	1490	2005	Construction	120	2	9	2 × 'aerovane'	100		44 × structure + 6 × air
Maanshan Yangtze	2 × x1080	2010	Construction	268	6 + 9 × 'deflectometer'	9	2 × 'aerovane'	80		52 × structure + 2 × air 2
Tsing Ma (Wong, 2007)	1377	1997	Construction	110	10 × level, 2 × ext	14	6 × sonic and prop	19		115
Ting Kau (Wong, 2007)	475	1998	Construction	88	2 × ext	7	7 × sonic and prop	45		83
Kap Shui Mun (Wong, 2007)	430	1997	Construction	30	5 × level, 2 × ext	6	2 × sonic and prop	3		224
Akashi Kaikyo (Sumitro, 2001)	1991	1998		72 × FBG	3 × ext, 6 × TMD	8	1 × triaxial	6 × triaxial seismometer	3 × girder velocity	3
Tatara Bridge (Sumitro, 2001)	890	1999					15 × uniax 1 × triax seismometer			

Abbreviations: Vibrating wire gauge, VWG; electrical resistance strain gauge, ERSG; fibre Bragg grating, FBG; tuned mass damper, TMD; California Strong Motion Instrumentation Program accelerometer, CSMIP.

Further examples are given in (Ko and Ni, 2005a) and (Andersen and Fustinoni, 2006).

For some bridges there is no information about sensors instrumentation in public domain.

remedial measures were introduced as a result of the studies. Concerns about wind-induced response have remained a strong driver for monitoring that has included the full range of wind sensors mentioned above.

For wind-sensitive structures that respond dynamically, and this includes long span bridges having mode frequencies below 1 Hz, the fluctuating component of wind (i.e. turbulence) is as relevant as the mean value; hence the frequency response of wind and pressure sensors is as important as it is for motion sensors. In this respect, sonic anemometers (which have no moving parts) are regarded as the best performing devices but are sometimes used in tandem with mechanically robust cup/vane and combined propeller/vane devices for long-term monitoring (Wong, 2004).

Concerns about in-wind performance drove the monitoring of the Humber Bridge in 1990 and 1991. In a collaboration between University of Bristol and Politecnico di Milano (Brownjohn *et al.*, 1994), the bridge was used to validate software to simulate dynamic in-wind performance of a proposed deck section design for the Stretto di Messina Bridge.

In recent years monitoring exercises have also been used to study cable vibration problems such as at Oresund Bridge (Larsen and Andersen, 2007), requiring solutions for recording cable vibrations that have included direct fixing of wired and wireless (Feltrin *et al.*, 2010; Nagayama *et al.*, 2010) accelerometers, strain gauges, proximeters (Humber) and non-contacting optical systems (Caetano *et al.*, 2007; Nassiff *et al.*, 2005).

Earlier systems invariably comprised accelerometers and anemometers, due to the emphasis on wind-induced response. Modern systems employ a wide range of sensors operating with both analogue and digital signals and at a range of sample rates. Also, technologies that emerged in the 1990s, such as GPS and fibre optics, are now 'mature' and systems for measuring configuration (effectively via displacement) are increasingly important.

So far wireless sensors sampling at dynamic rates (to record vibrations) have rarely been used in permanent monitoring installations, but real world tests of the technology include Golden Gate (Kim *et al.*, 2007) and Jindo Bridge (Jo *et al.*, 2011). The Jindo Bridge exercise is, to date, the most remarkable exercise on the deployment of a large-scale network of wireless sensors capable of recording at high data rates with synchronisation errors measured in micro-seconds.

7.4   Case study on condition assessment and performance monitoring: Tamar Bridge

The Tamar Bridge in the UK provides a good demonstration of many of today's relevant technologies for SHM, due to the fusion of four sub-systems operating at different timescales and with different objectives. Together these provide a comprehensive overview of the bridge performance that

supplies not only useful information for bridge operation but also a test for the sensor, acquisition and data management systems.

Not all types of sensors are represented on the bridge; for example, there is no strain monitoring using static, dynamic or fibre optic sensors. The system focuses on recovering quasi-static variation of bridge configuration and dynamic responses, along with reliable information on thermal, wind and traffic loading.

7.4.1 Tamar Bridge: original design and subsequent strengthening and widening

The Tamar Bridge forms a vital transport link over the River Tamar carrying the A38 trunk road from Saltash in Cornwall to the city of Plymouth in Devon. The original bridge (Fig. 7.1), opened in 1961, was designed by Mott Hay and Anderson as a conventional suspension bridge with symmetrical geometry, having a main span of 335 m and side spans of 114 m and, with anchorage and approach spans, the overall length is 642 m. The towers were constructed from reinforced concrete, and have a height of 73 m with the deck suspended at half this height. The towers sit on caisson foundations founded on rock.

Main suspension cables are 350 mm in diameter. Each consists of 31 locked-coil wire ropes and carries vertical locked-coil hangers at 9.1 m centres.

7.1 Tamar Bridge before and after strengthening and widening.

The main cables are splayed at anchorages and anchored some 17 m into rock. The stiffening truss is 5.5 m deep and composed of welded hollow boxes. The original three-lane deck, spanning between cross trusses, was of composite construction with a 150 mm deep reinforced concrete slab on five universal beams surfaced with hand-laid mastic asphalt 40 mm thick.

When opened in 1961, Tamar was the longest suspension bridge in the UK and the first to be built since World War II. It was initially carrying approximately 4000 vehicles a day with a maximum gross weight of 24 tons, but was upgraded in 2000 and 2001 to carry lorries up to 40 tonnes, incorporating modifications to major components, i.e. deck, cables and boundary conditions.

Deck

The appointed consultant for the strengthening and widening upgrade (Acer, now Hyder) proposed replacement of the main deck with a lightweight orthotropic steel deck, and construction of temporary relief lanes cantilevered off the bridge truss, to act as a supplementary diversion route while the main deck was being replaced. These additional lanes became permanent, and several other modifications were made.

The truss was strengthened by the installation of supplementary inverted U-shaped parallel elements fitted below the bottom chord, and by welding additional steel plates at key locations.

Approximately 2800 tonnes of structural steel was added in the deck replacement, cantilevers and additional cables, but with removal of the old composite main deck with heavy concrete slab, the final weight of the suspended structure rose by just 25 to 7925 tonnes. The deck replacement process was completed by December 2001, and the bridge now carries about 50 000 vehicles per day.

Additional stay cables

Eighteen new (nominal) 100 mm diameter locked-coil cables were added primarily to help carry the additional dead load of the new cantilever lanes and associated temporary works. They also reduced the extent of required truss strengthening and restored some 400 mm of the original hogged profile of the main deck, which had been lost due to main cable creep over the previous 40 years.

During and after the upgrade, in specific weather conditions involving rain, some additional stay cables exhibited oscillations of up to 100 mm amplitude. Figure 7.2 shows a schematic of the 18 additional stay cables. P1 north and S2 south exhibited large amplitude vibrations during construction as tension varied, and after construction the longer cables (3 and 4) on occasion exhibited vibrations of 40 mm amplitude. 'Water butt' dampers subsequently deployed provided a simple but effective means of controlling vertical plane vibrations of the hangers.

(a)

S3 S1   S2 S4   P4 P2   P1 P3
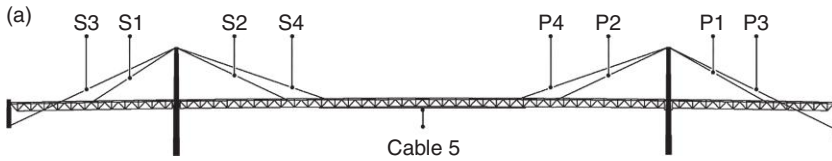
Cable 5

(b)
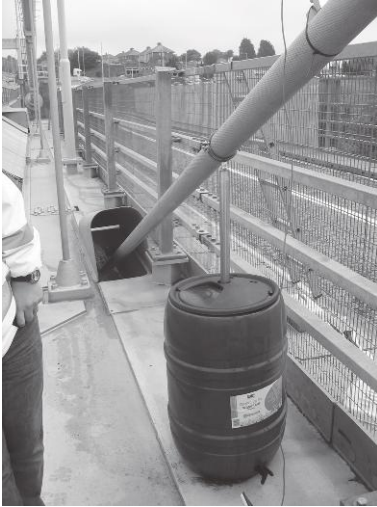
7.2 Schematic of additional stay cable locations and cable damper.

Span continuity, bearings and expansion joints

In the original configuration, bearings at the Saltash tower comprising ver-tical swing links and lateral thrust bearings allowed for longitudinal move-ment and rotation about a vertical axis to allow deck sway, together with an expansion joint in the roadway. At the Plymouth tower the arrangement was the same but with a link connecting the trusses either side of the tower.

During the upgrade, the link at the Plymouth tower was severed, and the load path to provide longitudinal restraint on the main span is via the can-tilever deck sections which are connected to the main truss via shear boxes. These cantilevers are continuous across the Plymouth tower, and incorpo-rate movement joints at the Saltash tower.

7.4.2   Evolution of the monitoring system and reasons for sensor choice

Monitoring systems were originally installed to track bridge configura-tion during the upgrade, and these original sensors still represent the majority of data channels in the present day system. Since late 2006

dynamic data channels have been added, followed by a total station in 2009. In 2010 a relatively simple wireless sensor network was added to track configuration. All these sub-systems have been modified during the 11 years of operation, and all have had their own performance and reliability problems.

Subsystem 1: Fugro SMS

A structural monitoring system (SMS) was installed by Fugro Structural Monitoring to monitor cable loads, structure and environment temperatures, and wind speed and profile. The purpose of the SMS was and is to provide engineering information on the performance and condition of the bridge during and after the widening and strengthening. As such, the main use of the SMS was for tracking cable loads and deck level during the strengthening works. Data were also available from a WIM system, which is no longer serviceable.

The level sensing system comprises a fluid pressure system with sensing stations at 1/8 span centres and is based on a similar system installed on bridges in the Lantau Fixed Crossing. The height measurements were specified to be accurate to +/5 mm and sampled at 1 Hz.

The sensors used in the SMS include the following (shown in Fig. 7.3):

- Anemometers to measure wind speed and direction.
- Fluid pressure-based level sensing system to measure deck vertical displacement.
- Temperature sensors for main cable, deck steelwork and air.
- Extensometers and resistance strain gauges to measure loads in additional cables.
- Electronic distance measurement between tower tops.

Deck and cable temperature sensors are platinum resistance thermometers (PRT) on stainless steel shim glued in place. Air temperature sensors consist of a temperature probe with radiation shield.

Additional cable tensions are measured by resistive strain gauges attached to main tensioning bolts at deck anchor points by epoxy or micro-welding. Gauges are arranged in pairs 180° apart around the bolt, each pair comprising an axial element and an element to measure hoop strain. The four gauges are connected to a full Wheatstone bridge, with the hoop gauges providing the temperature compensation.

Wind speed is measured mechanically by cup anemometers at the top of Saltash Tower (where direction is also measured), the deck level of Saltash tower and Saltash approach.

Tower separation is measured by an electronic distance measuring device (EDM) attached to the upper portal of the Plymouth tower.

(a)
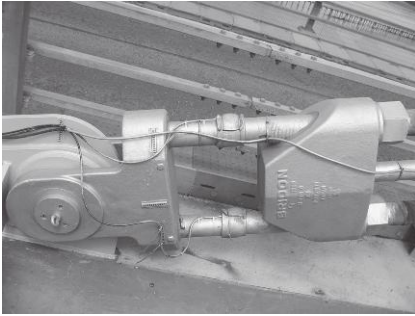
(b)

(c)

(d)
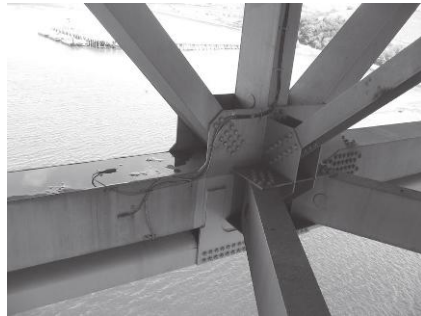
7.3 Fugro SMS sensors. (a) Cable tension; (b) wind speed and air temperature; (c) suspension cable temperature; (d) truss girder temperature.

At present, the main concerns for bridge operation are safety, not just in the bridge but in the adjoining Saltash tunnel. Closed circuit television (CCTV) cameras and image tracking software are used to avoid and manage dangerous traffic situations, and wind data from the SMS are used to determine when the bridge should be closed to high-sided vehicles.

The SMS was overhauled in summer 2007 when extra data channels were added and faulty channels, including the levelling system and some stay cable tension sensors, were fixed. However, many of these sensors (such as level sensors, EDMs) are now unserviceable.

Subsystem 2: Vibration Engineering Section (VES)
dynamic monitoring system (DMS)

Vibration Engineering Section began monitoring the dynamic response of the bridge deck and cables in 2006. This monitoring system includes eight Honeywell QA650 accelerometers arranged to measure biaxial cable vibrations, and three QA750 accelerometers installed close to the midspan of

(a)

(b)

(c)

(d)
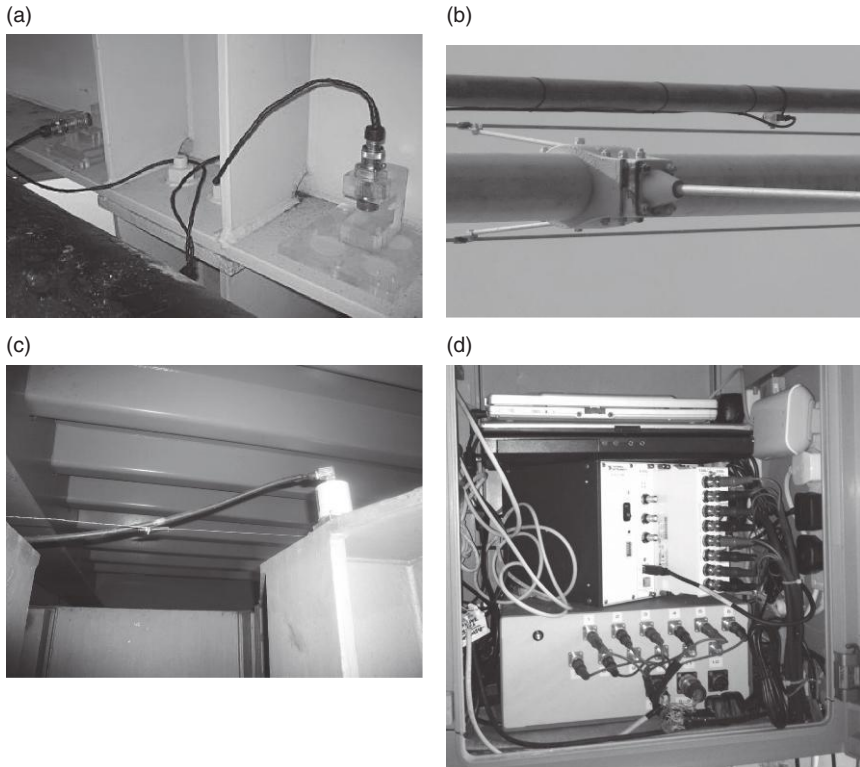
7.4 DMS components. (a) Deck force balance accelerometers; (b) stay cable servo-accelerometers; (c) extensometer at Tamar expansion joint; (d) data acquisition system: dual laptops, SCXI chassis, power supply and modem.

the bridge deck arranged to measure horizontal motion as well as vertical motion on each side of the truss (Fig. 7.4). The QA750/650s are force balance DC 'servo-accelerometers' so that they can measure quasi-static acceleration. They also resolve accelerations in micro-g range.

Three ASM-WS12 'potentiometer' or 'pull-wire' extensometers were installed (lower left view in Fig. 7.4), two to track relative motion between the deck truss and tower pylon on each side, and one for motion between main and side span via the thrust bearing extensions.

All 12 channels were hardwired to a small instrumentation cubicle above the anchorage chamber at the Plymouth side. Analogue voltage signals have since 2006 been collected by a National Instruments (NI) SCXI format data acquisition system (DAQ), which was originally connected to a router in the bridge control room via fibre optic cable and recycled repeater nodes. The original fibre optic link has been replaced with a power-line modem.

(a)   (b)
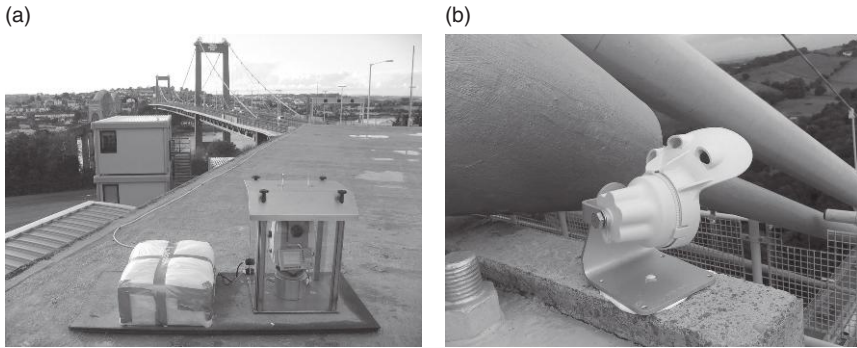
7.5 Total positioning system components (a) RTS on control room roof;
(b) prism reflector on Saltash tower pylon.

Subsystem 3: robotic total station (RTS)

Examination of accelerometer data and quasi-static deck extension, deck
level and cable tension data in 2008 indicated that deck configuration played
an important role in modifying dynamic response and was itself driven in a
complex way by various forms of loading. With award of a major research
grant from UK's Engineering and Physical Sciences Research Council to
fund SHM on suspension bridges, the DMS was extended to incorporate
an RTS (Koo *et al*., 2010). RTS was chosen instead of GPS following dis-
cussion with and advice from researchers at Technische Universität Graz.
For a relatively short span the expected quasi-static movements in all three
directions could be challenging to measure by GPS in real-time kinematic
(RTK) mode, which has an accuracy around 15 mm. To measure with sim-
ilar spatial resolution to the level system (which stopped working about a
year after its 2007 repair) would require expensive multiple antennae and
present problems of fixing them securely, in terms of access and theft as
well as in terms of structural rigidity, and multi-path. RTS has better res-
olution in space but poorer resolution in time, limited to about 1 Hz for a
single target, but since fast sampling was not required, the much slower and
irregular sampling of RTS could be tolerated. Hence, a single Leica TCA
1201M was deployed to a fixed location on the control room roof, with 15
reflector targets mounted at critical points on the bridge (Fig. 7.5). Quasi-
static response with periods up to 20 s could be recovered from integration
of accelerometer signals.

   The TCA was originally tested on the Plymouth tower upper portal, which
offered the best location for target visibility, but the high frequency dynamic
sway of the tower prevented reliable locking on reference reflectors and
recovery of azimuth angles.

(a)

(b)

(c)
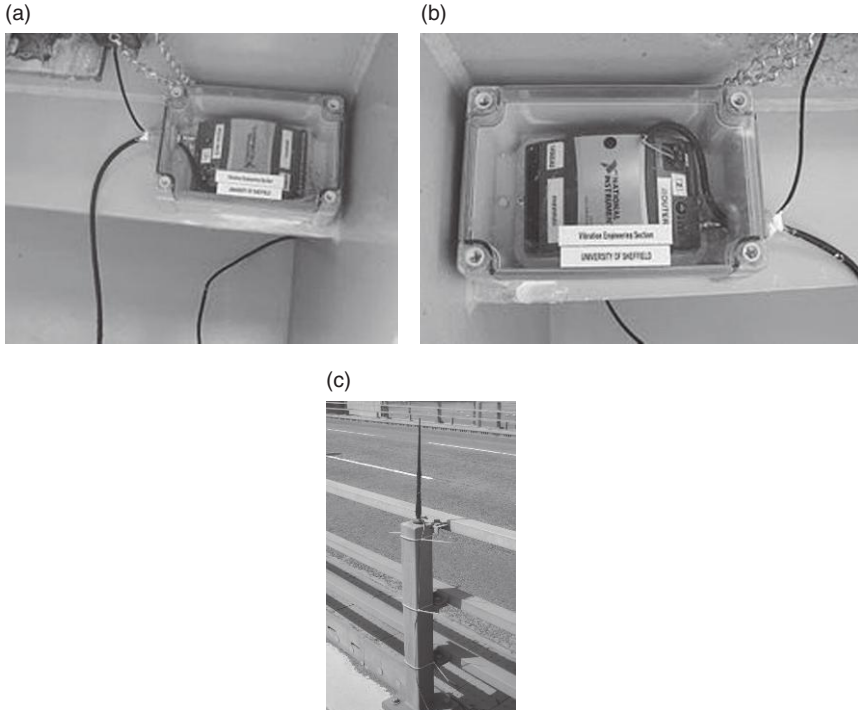
7.6 Wireless links for extensometer signals. (a) WSN node at Saltash tower; (b) router node at Plymouth tower; (c) router antenna.

Subsystem 4: wireless sensor node (WSN)

Due to problems with the wired connections to the extensometers resulting in an earth loop that could not be fixed, an NI WSN operating on the 2.4 GHz frequency range using the IEEE 802.15.4 protocol was installed to transmit the data from the extensometers to the control chamber. The extensometers were connected to an NI WSN-3202 4-channel, 16-bit analogue input node with the DAQ laptop connected to a to a NI WSN-9791 Ethernet gateway node. The distance between the sensing location (the Saltash tower expansion joint) and the DAQ laptop was approximately 450 m, greater than the transmission range of the NI WSN nodes; so an intermediate NI WSN-3202 node was installed and configured to act as a router in a two-hop network. The router was located on the main span near the Plymouth tower, where there was a readily-available DC power supply. This still left a 330 m hop over the main span, longer than the specified operational range of the NI WSN nodes, requiring 10 dBi high-gain, omni-directional antenna on each of the three nodes. System components are shown in Fig. 7.6.
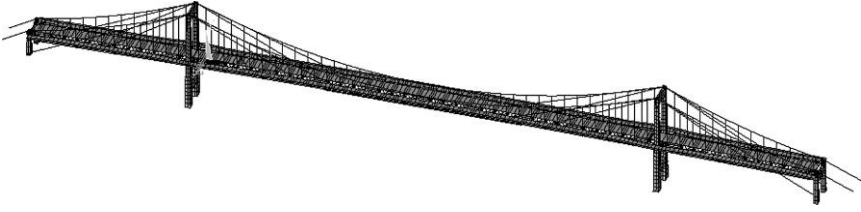
7.7 Finite element model of Tamar Bridge.

Present system

The present system, fusing the four sub-systems, is shown in Plate VII in the colour section between pages 374 and 375. Not all these channels are operational, or work intermittently, as will be described later. Images from CCTV cameras on each of the towers are also collected with the environment and performance data to help interpret anomalous performance.

7.4.3   Other studies supporting the monitoring

To support the monitoring studies, finite element models of the bridge have been created. The most effective model of the bridge, shown in Fig. 7.7 has been developed in ANSYS and used for simulations of effects of thermal and traffic loading on the bridge (Westgate and Brownjohn, 2010). To calibrate any modelling of the bridge, an ambient vibration survey of the bridge was carried out in April 2006 (Brownjohn *et al.*, 2007). This exercise collected acceleration signals from both sides of the deck in lateral and vertical directions as well as lateral and longitudinal motions in the towers, followed by operational modal analysis to recover modal properties specific to the day and time of the measurements. While the modal frequencies and damping ratios are now known to exhibit considerable variation according to environment and loading conditions, the mode shapes should be relatively stable and allow for tracking of specific mode type in the subsequent monitoring.

7.4.4   Data management

Data management is described elsewhere (Koo *et al.*, 2011); data on local PCs are replicated in the VES server and loaded in the Sheffield SHM database (SSDB). The SSDB system includes processing algorithms coded in MATLAB calculating derivatives of raw time series and their combinations, present in columns of database tables that can be interrogated using MATLAB or a web interface and viewed in near real time.

7.5   Monitoring results illustrating sensor characteristics

This section describes performance of sensors and the implications for data interpretation.

7.5.1   SMS data

Not all data channels have been operational throughout the monitoring, but reliable tension data have been available for several of the additional stays. Level data have been fully functional for a relatively short period from mid 2007, during which it provided insights into the deformation performance of the bridge.

Figure 7.8 shows different scales of level data. First the relationship with structure temperature is clear, as is the daily variation in standard deviation of level hour by hour, showing the relatively quiet weekend periods. The effects of rush-hour traffic enhancement and temperature are evident in the 24 h plot, while the zoom on an hour of data hints at heavy vehicle crossings (the large negative spikes). The 4 s sampling period for the sensor is too coarse to be confident of such information. The plots do not show that absolute levels appeared to drift and then shift; the value, but unreliability, of these data led to installation of the RTS system.

The relationships with structure temperature, stay cable tension and deck level are illustrated in Plate VIII in the colour section between pages 374 and 375. Other data show that the deck sag/hog is lopsided in elevation and accompanies a significant longitudinal expansion, while the linearity of correlations with temperature depends on which sensor location is used. The opposite slopes are due to the opposite directions of the stays in relation to the longitudinal expansion.

7.5.2   DMS data

The 11 accelerometers were deployed from the end of 2006 and originally interpreted by the *in situ* data processing and modal analysis software. Plate IX in the colour section between pages 374 and 375 shows deck and cable signals displayed using that interface. The lateral deck signal is too weak in the time series but appears in the spectrum for a higher order lateral mode (red). All eight cable sensors were working at the time, although signal shifts are evident in some cable accelerometer channels. The build-up and decay of signals is also clear; these data predate installation of water butt dampers on all the additional stays.

(a)

(b)
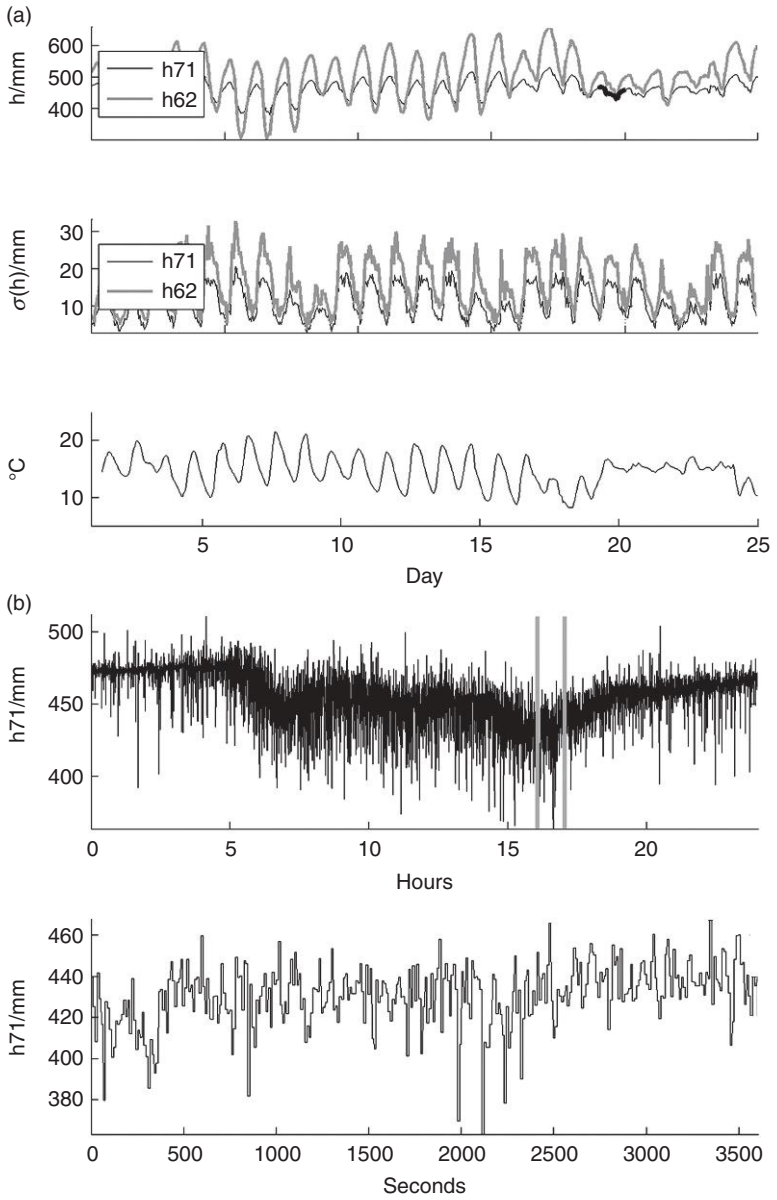
7.8 Level sensing data from SMS at increasing levels of resolution.
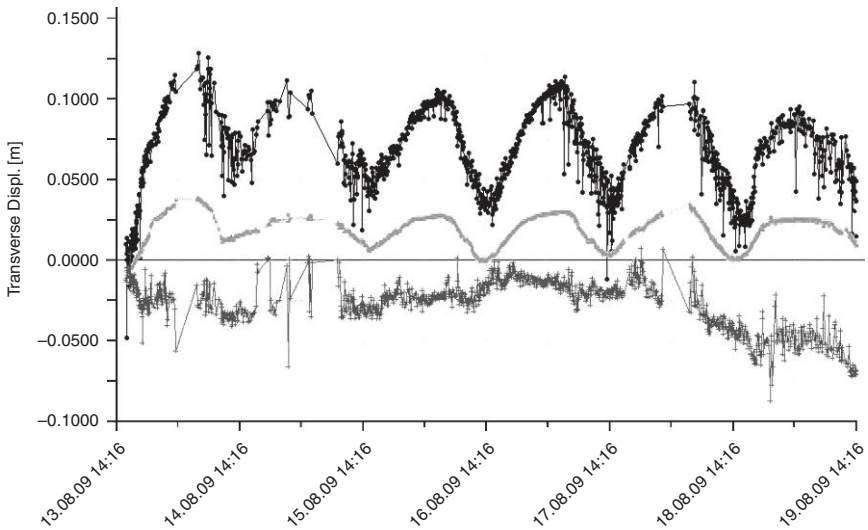
7.9 RTS solutions for midspan.

7.5.3   RTS data

A sample of 6 days of RTS data is shown in Fig. 7.9 for the midspan reflector. Unlike the level data, these signals comprise three components, when they are available. There is significant scatter, in vertical/lateral directions, possibly due to their being a 'snapshot' of quasi-static plus dynamic response. The axial signals, which are barely influenced by dynamics, are much clearer and more stable. The three components combine EDM and theodolite components of the RTS, but for axial data the weight of EDM data is close to 100%. Signal loss is usually due to weather effects and the inability to discriminate closely spaced targets. Nevertheless, the data are more valuable and reliable than the level data, and more cost effective and accurate than GPS data would be. Signal precision for the EDM is approximately 2 mm + 2 ppm (i.e. about 2 mm for furthest reflector) and for the theodolite it is 1 arc-s, or 3.3 mm for the furthest reflector.

7.5.4   WSN data

Plate X in the colour section between pages 374 and 375 shows the monitoring interface and extensometer data recorded over the first 6 months of operation. The longitudinal displacement of the bridge deck had a clear diurnal pattern as well as a seasonal trend, with the deck closer to the tower during day time and the warmer summer months. Also noticeable are sections of missing data from both extensometers. These were due to a recurring

Table 7.2 Tamar sensors and availability

Sensor type	Number of units	Availability
Cable load cell (pair)	20	12 half-sensors unserviceable from varying times in programme
Level sensor	10	Progressively unserviceable from 2008
Deck temperature	2 2 3	Two unserviceable
Cable temperature	1	
Air temperature	1 1	
Wind speed	4	
Wind direction	2	
EDM	1	Intermittent, not used
Deck accels	3	I failed and replaced
Cable accels	8	Four failed 2008, two more failed from 2012. Not needed
Extensometers	2	
RTS reflectors	16	

problem with the power supply to the monitoring systems from the Bridge's electrical circuitry. There were also a number of instances where the data from the south extensometer were clearly erroneous, most likely due to the extensometer cable getting stuck and not being pulled back by the spring mechanism. So, while reflecting the actual movement may not be perfect, due to the nature of the device, resolution is as good as the resolution of the DAQ.

7.5.5   Issues with sensors

The data presentation has highlighted some issues with sensors: as the bridge itself requires maintenance, so does a SHM system. At the time of setting up the DMS, the level sensors and several of the cable 'load cells' (i.e. combinations of strain gauges) were not functioning. These functioned well for a while after the 2007 repair, but problems recurred and some cable load data were clearly anomalous (as identified by ad-hoc cable vibration measurements). Wind and temperature sensors have remained both reliable and valuable. Table 7.2 summarises sensor types and availability.

For the DMS, water ingress into the wiring junction boxes has been the significant problem, and some deck accelerometers have been replaced, while at the time of writing two cable accelerometer pairs have ceased to function, probably for the same reason. The earth loop problem with extensometers has been avoided via the WSN.

For the RTS, the principal issue has been the failure to locate distant or close reflectors, resulting in missed data points, a particular problem in bad weather. Despite this, the RTS has been a powerful tool for tracking bridge deformation.

For the WSN there have been few problems, except occasional data loss due to signal strength.

Reliability of computers has been problematic. While the computers (so called 'Toughbooks') themselves have survived for up to 5 years, software issues have caused occasional hanging, so that we now use internet-linked remote reset hardwired to the computer power switch.

7.5.6   Structural health monitoring (SHM) applications

While the Fugro SMS was installed for tracking structural behaviour during the strengthening and widening, the present system was developed to support development of data mining technologies for SHM, which considers sudden or evolving changes in structural condition. Such changes can be revealed through changes in time varying response patterns. For quasi-static response stresses, strains, deformations correlate with wind, traffic or thermal loads so that changes in the correlations or outliers can indicate structural changes, and such changes can be enhanced by sophisticated analysis of the time series (Omenzetter *et al*., 2004). Modal parameters(e.g. natural frequencies) themselves can be viewed as response parameters which vary with environmental and loading conditions. Clear and unambiguous identification of a modal parameter change can be very useful for diagnosing the underlying change in structural condition.

So far with Tamar, the obvious performance anomalies have been directly linked to traffic overloads, leaving the more subtle changes to be spotted among the noise of normal operational environment and load changes. Hence, recent research on Tamar performance data has focused on filtering out the 'normal' variations by creating meta-models fitting the observed variations of load/environment with response using the technique of co-integration (Cross *et al*., 2012).

7.6   Conclusion and future trends

Experience from this bridge (and other structures our group has monitored) has led to the following observations.

Accelerometers are extremely versatile instruments capable of providing time series and response statistics in isolation and modal parameters collectively (although less reliable estimation is possible from single devices). Unless it is necessary, and possible, to observe variations in mode shape,

a few well-placed accelerometers can be used with reliable modal survey information (using larger temporary arrays) to provide global response via modal expansion.

Accelerometers can also provide quasi-static displacement and rotation information when high quality low noise DC instruments are used. Hence, for a given budget it would be better to use fewer but better and more costly accelerometers.

There have been no significant problems with hard wiring accelerometers but locations have been limited due to cost, primarily of the accelerometers but to some extent of the cabling. In cases where low frequency sampling is adequate or where embedded processing of single point measurements suffices, there are real benefits for wireless devices. A particular example is measurement of stay cable response.

The wind sensors used have been simple and rugged devices. While sonic anemometers have no moving parts, our experience in using them on tall structures (transmission tower, tall building) is that they can per-form badly in heavy rain and suffer from electrical interference, e.g. during thunderstorms.

Measurement of displacement or configuration is obviously very impor-tant for a bridge, particularly a long span bridge, and there is no single per-fect technology for the purpose (Brownjohn and Meng, 2008). The popular solution is to use GPS antennae and receivers, as deployed in the majority of Chinese bridge monitoring exercises (Ko and Ni, 2005). As described in the decision process for Tamar, GPS sensors are relatively expensive sin-gle point devices that have the advantage of relatively high speed synchro-nous sampling. They are subject to unusual and non-Gaussian noise, e.g. due to cycle slip and multi-path, resulting in very large errors, and even with-out those, accuracy is around 15 mm without post processing. For Tamar, RTS was chosen, which has its own problems, but there were no significant problems with loss of information due to the very slow sample rate: a mea-surement with the motor-driven unit finding the measuring locations of 15 reflectors takes over 10 min. This leaves a gap between the longest possible period of displacement data integrated from accelerometer signals (about 20 s) occupied by timescales for crossing of law-abiding heavy vehicles and rapidly changing weather conditions.

For relative motion, e.g. across expansion joints, there is a range of tech-nology including mechanical devices and non-contacting devices, usually some form of laser. The former suffer from abuse (and birds), the latter from line of sight interruption.

Fibre optic sensing technology, mainly used for strain measurement, has advanced considerable in the last decade, particularly the spectrum analy-ser/logging devices. Fibre optic sensors (e.g. fibre Bragg grating and Brillouin scattering devices) are a natural choice for extended steel structures and

are sensibly adopted for Chinese long span bridges. However, strain gauges have hardly been used in this study, which focuses on global issues.

Above all, the greatest need is for development of systems for managing, interpreting and applying the performance data. This is a major challenge, which is still hardly addressed by any research group in a systematic and effective way (Koo *et al.*, 2011).

The trend in long span bridge monitoring is for ever-larger and more complex systems incorporating a wide range of sensors types providing both analogue and (directly) digital data at a range of fixed and variable sample rates.

7.7   References

Andersen, J. E. and Fustinoni, M. (2006). Structural health monitoring systems. COWI-Futurtec, Lyngby, Denmark.

Bakht, B. and Jaeger, L. G. (1990). Bridge testing – a surprise every time. *Journal of Structural Engineering* **116**, No. 5, 1370–1383.

Bampton, M. C. C., Ramsdell, J., V, Strope, L. A., Athey, G. F. and Abbey, O. B. (1983). Pasco-Kennewick cable stayed bridge wind and motion data. US Department of Transportation Federal Highway Administration Report FHWA/RD-82/067.

Bampton, M. C. C., Ramsdell, J. V., Graves, R. E. and Strope, L. A. (1986). Deer Isle-Sedgwick suspension bridge. Wind and motion analysis. Report FHWA/RD-86/183.

Brownjohn, J. M. W. (2007). Noise characteristics of sensors for extreme low level vibration. In: *IMACXXV*, Orlando, FL.

Brownjohn, J. M. W., Bocciolone, M., Curami, A., Falco, M. and Zasso, A. (1994). Humber bridge full-scale measurements campaigns 1990–1991. *Journal of Wind Engineering and Industrial Aerodynamics* **52**, 185–218.

Brownjohn, J. M. W. and Meng, X. (2008). Methods for measuring structural deflection and applications to bridge deck performance monitoring. In: *The Fifth International Conference on Bridge Maintenance, Safety and Management*, IABMAS2010, Seoul, Korea.

Brownjohn, J. M. W. and Moyo, P. (2001). Monitoring of Malaysia-Singapore Second Link during construction. In: *Second International Conference on Experimental Mechanics*, 528–533.

Brownjohn, J. M. W., Moyo, P., Omenzetter, P. and Lu, Y. (2003). Assessment of highway bridge upgrading by dynamic testing and finite-element model updating. *Journal of Bridge Engineering* **8**, No. 3, 162–172.

Brownjohn, J. M. W., Pavic, A., Carden, E. P. and Middleton, C. J. (2007). Modal testing of Tamar suspension bridge. In: *IMACXXV*, Orlando, FL.

Caetano, E., Silva, S. and Bateira, J. (2007). Application of a vision system to the monitoring of cable structures. In: *Seventh International Symposium on Cable Dynamics*, 225–236.

Carder, D. S. (1937). Observed vibrations of bridges. *Bulletin Seismological Society of America* 27, 267.

Casciati, F. and Fuggini, C. (2009). Engineering vibration monitoring by GPS: long duration records. *Earthquake Engineering and Engineering Vibration* **8**, 459–467.

Catbas, F. N., Brown, D. L. and Aktan, A. E. (2006). Use of modal flexibility for damage detection and condition assessment: Case studies and demonstrations on large structures. *ASCE Journal of Structural Engineering* **132**, No. 11, 1699–1712.

Catbas, F. N. and Aktan, A. E. (2009). Development of a Monitoring System for a Long-Span Cantilever Truss Bridge. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, pp. 1–11.

Cebon, D. (1991). Multiple-sensor weigh-in-motion: theory and experiments. *Transportation Research Record* No. 1311, 70–78.

Celebi, M. (2006). Real-Time Seismic Monitoring of the New Cape Girardeau Bridge and Preliminary Analyses of Recorded Data: An Overview. *Earthquake Spectra* **22**, 609–630.

Cheung, M. S., Tadros, G. S., Brown, T., Dilger, W. H., Ghali, A. and Lau, D. T. (1997). Field monitoring and research on performance of the Confederation Bridge. *Canadian Journal of Civil Engineering* **24**, No. 6, 951–962.

Cross, E. J., Koo, K. -Y., Brownjohn, J. M. W. and Worden, K. (2012). Long term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing* http://dx.doi.org/10.1016/j.ymssp.2012.08.026.

Feltrin, G., Bischoff, R., Meyer, J. and Saukh, O. (2010). Structural monitoring with wireless sensor networks: lessons learned from field deployments. In: *5th International Conference on Bridge Maintenance, Safety and Management (IABMAS)*, Philadelphia, PA, USA. 1019–1026.

Frandsen, J. B. (2001). Simultaneous pressure and accelerations measured full-scale on the Great Belt East suspension bridge. *Journal of Wind Engineering and Industrial Aerodynamics* **89**, No. 1, 95–129.

Fraser, M., Elgamal, A., He, X. F. and Conte, J. P. (2011). Sensor network for structural health monitoring of a highway bridge. *Journal of Computing in Civil Engineering* **24**, No. 1, 11–24.

Glisic B. and Inaudi D. (2007). *Fibre Optic Methods for Structural Health Monitoring*. Wiley. Chichester, UK, ISBN: 978-0-470-06142-8.

Gucunski, N., Romero, F., Kruschwitz, S., Feldman, R., Abu-Hawash, A. and Dunn, M. (2010). Multiple complementary nondestructive evaluation technologies for condition assessment of concrete bridge decks. *Transportation Research Record* No. 2201, 34–44.

Isyumov, N., Davenport, A. G. and Monbaliu, J. (1984). CN Tower: Model and full scale response to wind. *IABSE 12th Congress*, International Association for Bridge and Structural Engineering. The secretariat is c/o ETH Zurich, Switzerland, 3–7 September.

Jo, H., Sim, S.-H., Mechitov, K. A., Kim, R., Li, J., Moinzadeh, P., Spener, B. F., Park, J. W., Cho, S., Jung, H.-J., Yun, C.-B., Rice, J. A. and Nagayama, T. (2011). Hybrid wireless smart sensor network for full-scale structural health monitoring of a cable-stayed bridge. In: *Sensors and Smart Structures Technology for Civil, Mechanical, and Aerospace Systems*, San Diego, CA. 1–15.

Kim, S., Pakzad, S., Culler, D., Demmel, J., Fenves, G., Glaser, S. and Turon, M. (2007). Health monitoring of civil infrastructures using wireless sensor networks. 254–263.

Ko, J. M. and Ni, Y. Q. (2005a). Technology developments in structural health monitoring of large-scale bridges. *Engineering Structures* **27**, 1715–1725.

Ko, J.M., Ni, Y.Q., (2005b). Upgrade of a structural health monitoring system for the suspension Jiangyin Bridge. In: *Korea Infrastructure Safety and Technology Corporation*, Gyeongju, Korea.

Koh, H.-M., Lee, H.-S., Kim, S. and Choo, J. F. (2009). Monitoring of Bridges in Korea. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, Chichester, UK, pp. 1–23.

Koo, K.-Y., Brownjohn, J. M. W., List, D., Cole, R. and Wood, T. (2010). Innovative structural health monitoring for suspension bridges by total positioning system. In: *Fifth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2010)*, Philadelphia, PA. 544–551.

Koo, K.-Y., de Battista, N. and Brownjohn, J. M. W. (2011). SHM data management system using MySQL databse with MATLAB and Web interfaces. In: *5th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-5)*, Cancun, Mexico.

Lark, R. J., Howells, R. W. and Barr, B. I. G (2004). Behaviour of post-tensioned concrete box girders. *Proceedings of the Institution of Civil Engineers, Bridge Engineering* 157 (BE4), 71–81.

Larsen, A. and Andersen, J. E. (2007). Identification of the source, impact and damping of stay-cable vibrations. In: Cunha, A. and Caetano, E. (Eds.), *Experimental Vibration Analysis of Civil Engineering Structures EVACES'07*, Porto.

Li, A.Q., Miao, C.Q. and Zhao, L. (2003) The health monitoring system for the Runyang Yangtse River Bridge. In: *Structural health Monitoring and Intelligent Infrastructure*, SHMII1.

Li, H., Ou, J., Zhao, X., Zhou, W., Li, H. and Zhou, Z. (2006). Structural health monitoring system for the Shandong Binzhou Yellow River highway bridge. *Computer-Aided Civil and Infrastructure Engineering* 21, No. 4, 306–317.

Macdonald, J. H. G., Dagless, E. L., Thomas, B. T. and Taylor, C. A. (1997). Dynamic measurements of the Second Severn Crossing. *ICE Proceedings Bridge Transport* 123, No. 4, 241–248.

Mallett, R. (1846). On the objects, construction, and use of certain new instruments for self-registration of the passage of earthquake shocks. *The Transactions of the Royal Irish Academy* 21, 107–113.

Meng, X. and Huang, W. (2009). Global navigation satellite systems (GNSSs) for monitoring long suspension bridges. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, Chichester, UK, pp. 1–19.

Nagayama, T., Jung, H.-J., Spencer, B. F., Jang, S., Mechitov, K. A., Cho, S., Ushita, M., Yun, C.-B., Agaha, G. A. and Fujino, Y. (2010). International collaboration to develop a structural health monitoring system utilizing wireless smart sensor network and its deployment on a cable-stayed bridge. In: *5th World Conference on Structural Control and Monitoring*, Tokyo, Japan. 1–10.

Nassiff, H. H., Gindy, M. and Davis, J. (2005). Comparison of laser doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *NDT & E International* 38, No. 3, 213–218.

Omenzetter, P., Brownjohn, J. M. W. and Moyo, P. (2004) Identification of unusual events in multi-channel bridge monitoring data'. *Mechanical Systems and Signal Processing* 18, No. 2, 409–430.

Sloan, T. D., Kirkpatrick, J., Boyd, J. W. and Thomson, A. (1992). Monitoring the inservice behaviour of the Foyle Bridge. *The Structural Engineer* 70, No. 7, 130–134.

Sumitro, S. (2001). Current and future trends in long span bridge health monitoring system in Japan. In: *NSF Workshop on Health Monitoring of Long Span Bridges*, UC Irvine, CA. 1–12.

The Highways Agency (1995). Design Manual for Roads and Bridges Vol. 3 Highway Structures: Inspection and Maintenance Section 4 Assessment Part 14. HMSO, BA44/95. London.

University of Washington (1954). Aerodynamic stability of suspension bridges with special reference to the Tacoma Narrows bridge. Bulletin No. 116 University of Washington Engineering Experiment Station.

Wang, G. D., Wang, M. L., Zhao, Y., Chen, Y. and Sun, B. N. (2006). Application of magnetoelastic stress sensors in large steel cables. *Smart Structures and Systems* **2**, 155–169.

Westgate, R. and Brownjohn, J. M. W. (2010). Development of a Tamar Bridge finite element model. In: *IMACXXVIII*, Jacksonville, FL.

Wong, K. Y. (2004). Instrumentation and health monitoring of cable-supported bridges. *Structural Control and Health Monitoring* **11**, No. 2, 91–124.

Wong, K. Y. (2007). Design of a structural health monitoring system for long span bridges. *Structure and Infrastructure Engineering* **3**, No. 2, 169–185.

Wong, K.-Y. and Ni, Y. Q. (2009). Modular architecture of SHM systems for cable-supported bridges. *Encyclopedia of Structural Health Monitoring*. John Wiley & Sons, Chichester, UK, pp. 1–17.

Xu, Y-L. and Xia, Y. (2012). *Structural Health Monitoring of Long Span Bridges*. Spon Press. ISBN: 978-0-415-59793-7.

Zhou, H.F., Ni, Y.Q. and Ko, J.M. (2006). Analysis of structural health monitoring data from the suspension Jangyin Bridge. *Proceedings 3rd European Workshop on Structural Health Monitoring*, Granada Spain.

Sensing solutions for assessing and monitoring seismically-excited buildings

A. MITA, Keio University, Japan

Abstract: Several sensing solution systems for assessing and monitoring seismically-excited buildings are reviewed. Although the importance of monitoring buildings has been widely recognized following the devastating earthquake of magnitude 9 in east Japan in 2011, the installation of a structural health monitoring system in a building is still expensive, limiting the use of such systems to tall, or otherwise important, buildings. Using sensors only for earthquakes may not be practicable for normal buildings. As a potential solution for this difficulty, a system using sensor agent robots is introduced.

Key words: structural health monitoring (SHM), smart sensor, sensor agent robot, tall building, earthquake.

8.1   Introduction

For the first time a seismometer recorded the large acceleration time history of the near field earthquake in 1940 at the El Centro Terminal Substation Building's concrete floor in California. Before this event, most building structures in seismic zones had been designed assuming static horizontal forces applied to building floors, because strong motion data were not available. Since then, many strong motion records have been obtained. The California Strong Motion Instrumentation Program (CSMIP, see the URL listed in reference) is a representative modern program that provides a funding scheme for installing and maintaining sensors in buildings. The data obtained at these buildings are accessible by anyone. There are many other programs that monitor buildings – the USGS has operated the National Strong-Motion Project (USGS, see the URL listed in reference) for many years.

In Japan, the Building Research Institute operates the Strong Motion Observation (BRI, see the URL listed in reference) and provides valuable data, including those taken during the 2011 east Japan earthquake of magnitude 9.0 that occurred on March 11, 2011. However, most buildings are

234

privately owned; hence, recorded data are generally not released to the public. The BRI provides an amazing dataset from a 55-story building in Osaka that was about 760 km away from the epicenter. The maximum displacement amplitude of the 52nd floor was about 1.4 m. This large shaking had not been expected before the event, as it was far away from the epicenter. The shaking was revealed by sensors installed in the building.

Extensive strong motion data have also been recorded in seismically isolated buildings. The Japan Society of Seismic Isolation (JSSI) immediately formed the Response-Controlled Buildings Investigation Committee to conduct detailed investigation of the performance of seismically isolated buildings and buildings with dampers to prepare action plans if necessary. Three working committees under this committee were formed. They are: the Base-Isolated Building Design Committee, the Response Evaluation of Base-Isolated Buildings Committee, and the Supplemental-Damped Building Investigation Committee. The second committee was established to evaluate the performance of seismically isolated buildings during the earthquake, using recorded data. Twenty seismically isolated buildings from the Sendai to Tokyo area were selected and examined. The number of stories ranged from 2 to 21. The buildings investigated are listed in Table 8.1, along with information on their isolators, dampers, and stories as reported by JSSI (2012). The peak value of the acceleration recorded at the foundation and the maximum deformation of the isolator are also provided. A summary of the report is seen in the paper prepared by (Kasai *et al.*, 2013) that was accepted for publication by Earthquake Spectra.

There was another notable episode during this earthquake. A structural health monitoring (SHM) system installed in a 40-story building in Tokyo, the World Trade Center Building, was reported to be of great help in assessing the safety of the building immediately following the earthquake. As a result, the state of the building was immediately announced to the tenants and appropriate and immediate decisions were taken, such as resumption of elevators. In most other tall buildings, visual inspections by experts were needed to ascertain the safety of the buildings and elevator systems before resumption of normal services. These visual inspections took hours and days to complete. The SHM system saved hours of time for the residents and tenants.

8.2   New roles for sensing and monitoring systems in buildings

For many years, accelerometers have been used for recording strong motion data. However, due to recent advances in sensors and networks, the concept of SHM has become popular in structural engineering. The SHM system records structural responses during large earthquakes, as well as other harsh

Table 8.1 List of seismic isolation buildings investigated by JSSI

Number	Location	Isolator and damper	Stories[a]	PFA[b] (cm/s^2)	D$_{max}$[c] (cm)
1	Sendai, Miyagi	High damping rubber	3/0	301	12
2	Sendai, Miyagi	Natural rubber, steel damper	9/2	289	17
3	Sendai, Miyagi	High damping rubber	6/2	381	18
4	Sendai, Miyagi	High damping rubber	5/0	345	11
5	Mito, Ibaraki	Natural rubber, lead damper, steel damper	21/0	402	14
6	Tsukuba, Ibaraki	Natural rubber, lead rubber, steel damper	7/0	327	6
7	Funabashi, Chiba	Natural rubber, viscoelastic damper	3/0	150	8
8	Funabashi, Chiba	High damping rubber, lead rubber	4/0	170	5
9	Chiba	High damping rubber	8/0	219	5
10	Chiyoda, Tokyo	Natural rubber, viscous wall	10/0	218	3
11	Chiyoda, Tokyo	Natural rubber, lead rubber, oil damper	11/2	104	5
12	Minato, Tokyo	Natural rubber, slider, semi-active and passive oil damper	13/3	65	14
13	Koto, Tokyo	Natural rubber, oil damper, rotational friction damper	12/0	100	7
14	Koto, Tokyo	Lead rubber	6/0	137	8
15	Chofu, Tokyo	Natural rubber, oil damper, steel damper	2/0	143	4
16	Yokohama, Kanagawa	High damping rubber, oil damper	7/0	147	7
17	Yokohama, Kanagawa	Natural rubber, semi-active and passive oil damper	7/2	71	12
18	Yokohama, Kanagawa	Synthetic rubber, slider	4/0	97	4
19	Yokohama, Kanagawa	Natural rubber, oil damper, steel damper	20/0	69	9
20	Kanagawa	Natural rubber, slider, oil damper	14/1	49	10

[a]Number of stories. [b]Peak foundation acceleration. [c]Maximum deformation of isolator.

external loads such as strong winds. The SHM system is different from other strong motion observation systems. While strong motion observation systems are mainly intended to record accurate and reliable data for analysis later, the SHM system data are intended to be used during the event as well

as after the event. A typical SHM system may provide the following new functions:

- Real time evacuation information is provided to residents to guide them to safe places during a harsh earthquake.
- Immediate health assessment is conducted during and after the event to provide the damage level to the building. The assessment results help resume normal functions of the building such as elevators, escalators, and security systems as soon as possible.
- Degradation of the building is assessed and the results are recorded so that economical and timely renovation of the structural system is possible for prolonging its service life.
- Time histories of structural performance as well as associated renovation information are recorded in an electrical carte. This information will be used when the building is sold to another owner. Thus, the value of the building with this carte will be higher than those without.

8.3   Structural health monitoring (SHM) systems for buildings

Historically, SHM systems for buildings trace their origins to strong motion observation systems using high-precision accelerometers. Because the number of sensor types used in this field is not large and most observation systems are custom-designed, the cost of a system is rather high. Furthermore, as there is no common data management system and storage format internationally, professional, individual care to maintain the systems is required, which adds to their cost.

In the initial phase of developing SHM systems for buildings, database systems and internet access have been added to conventional observation systems. Initially, there were no diagnostics tools installed in the system. With time, systems gradually evolved to include many diagnostics tools and other sensor types. A typical concept of the early SHM system is presented in Fig. 8.1.

A pioneering SHM system was installed in the seven-story base-isolated building built in 2000 on the Hiyoshi Campus, Keio University. The building is continuously monitored. The system consists of accelerometers and displacement sensors to observe the response of rubber isolators and the superstructure. The system is still being operated, and has gradually evolved to include many diagnostics tools. It was successful in recording the building's response to the 2011 east Japan earthquake. An overview of the system is depicted in Fig. 8.2. The trajectory of the isolator during the earthquake obtained from this system is plotted in Fig. 8.3. Details of the evolved version of the system are explained in the report by Okada (2011).
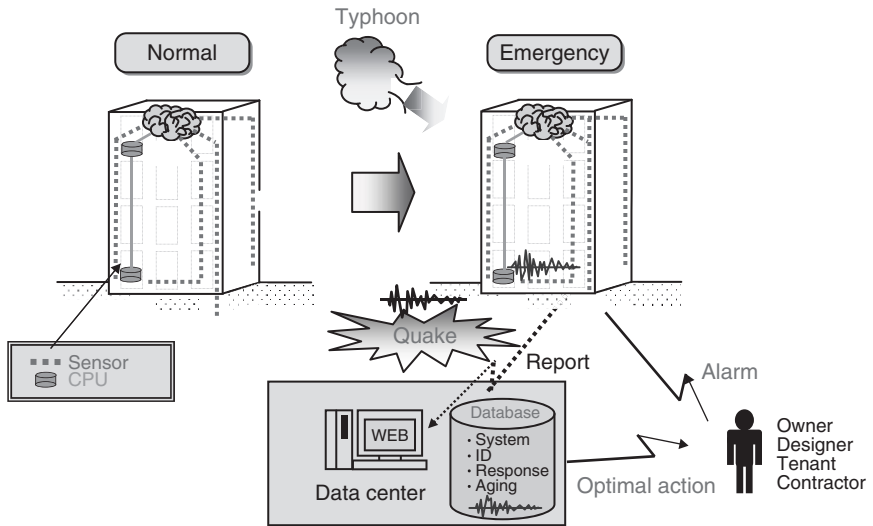
8.1 Typical concept of an early Internet-based SHM system.
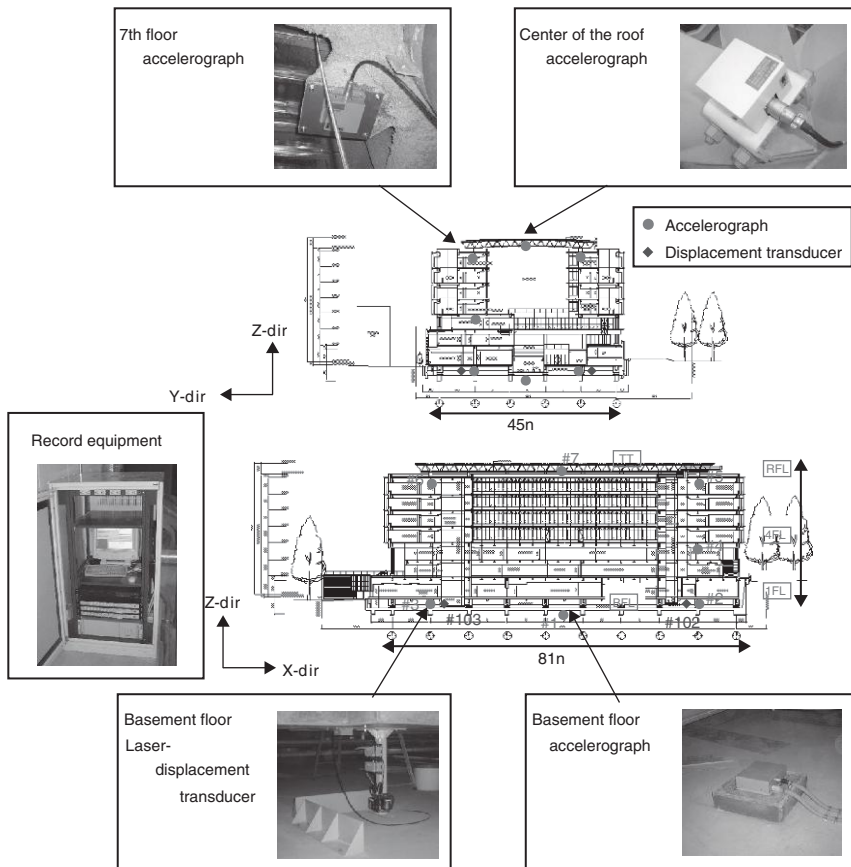
8.2 Seven-story base-isolated building (Hiyoshi Campus, Keio University) with SHM system installed.
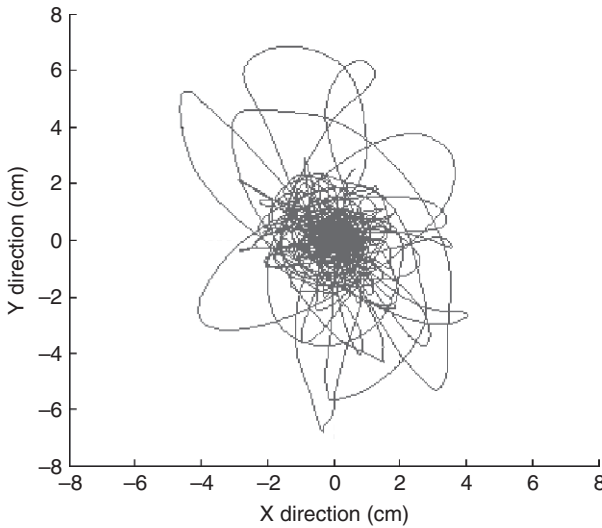
8.3 Trajectory of the isolation layer for the 2011 east Japan earthquake.

The 632 m tall Shanghai Tower is currently under construction in China. An SHM system has been designed and is being constructed by a joint venture from the Tongji University, Hong Kong Polytechnic University, and Tongji Architectural Design (Group) Co. Ltd. (the details of this project are available at Xu, *et al.* (2012).) The layout of the sensing system is presented in Fig. 8.4. The SHM system is intended to monitor the performance of the structure under wind, seismic, and temperature loadings during construction as well as during service. The system has over 400 transducers of 12 different types, including anemometers, wind pressure sensors, accelerometers, seismographs, strain gauges, thermometers, global positioning system (GPS), total stations, theodolites, inclinometers, digital video cameras, and crack sensors. This is one of the largest SHM projects for a tall building in the world.

In the next phase of development of SHM systems for buildings, the importance of common data formats and common analysis tools has been recognized by the community as critical. In 2008, a consortium of 14 companies, led by Keio University (PI: Akira Mita), has been established to develop a common data management platform for SHM. The system has a common data model for the SHM data and uses smart sensors. The smart sensors, shown in Fig. 8.5, are equipped with microprocessor boards with the Linux operating system installed so that necessary communication between sensors and the data server is immediately started when the sensors are connected to the network. The prototype system has recently evolved into a commercial system, with feasibility tests now being conducted.
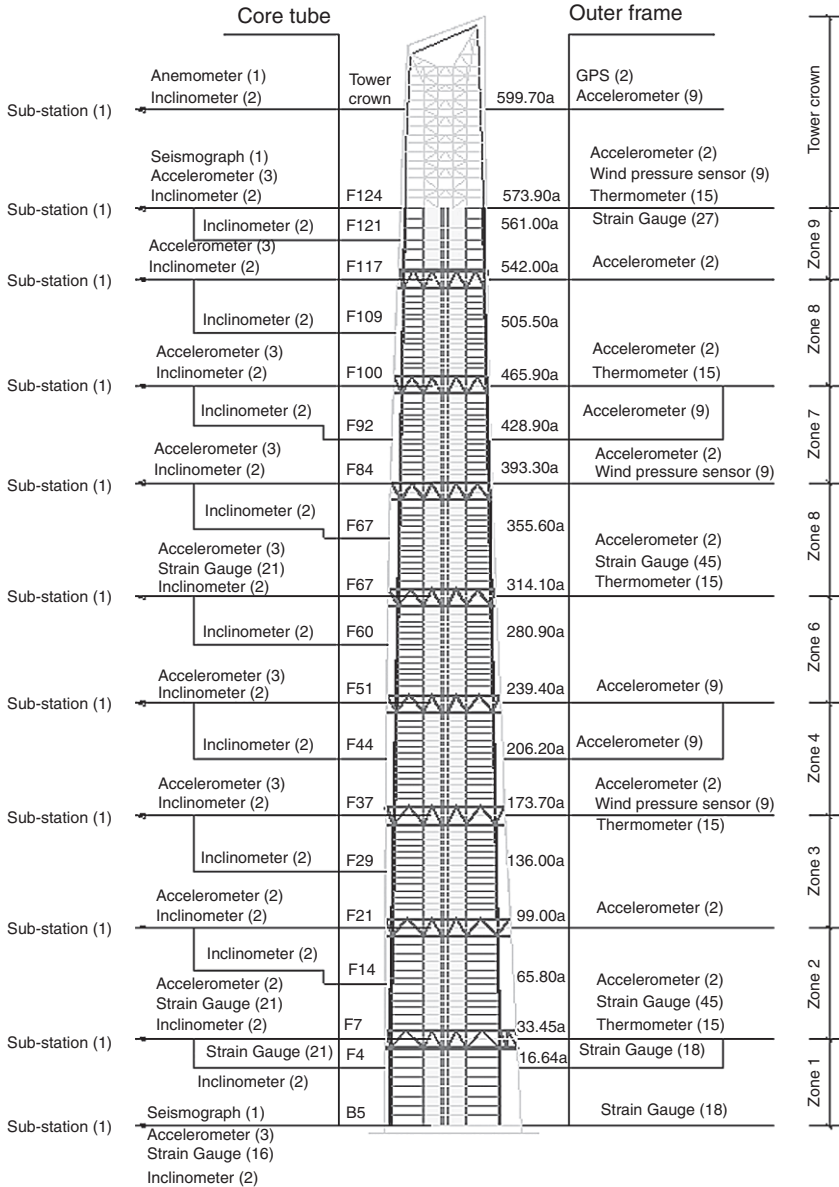
8.4 Instrumentation layout in Shanghai Tower (Xu *et al.*, 2012).

SHM systems using smart sensors are now evolving to include the use of sensor agent robots, which are small 'pet' robots for collecting information on environments in a building. An outline of an SHM system that employs sensor agent robots is depicted in Fig. 8.6. The details of this system are

SUZAKU-V

Power supply

A/D converter

Accelerometer

K-SHM-006
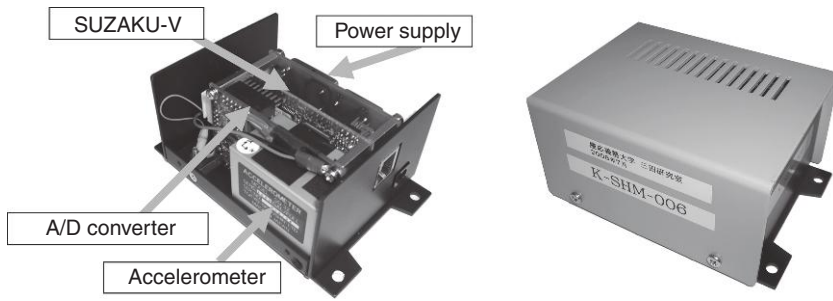
8.5 Prototype smart sensor unit.

Buildings

Analysis

Surrounding buildings' information

Details of building's safety

Data and Diagnosis Server
Database
Web server
System identification

Earthquake data

Sensor gateway
Connection with sensors
NTP server
Data conversion

Earthquake data

Buildings
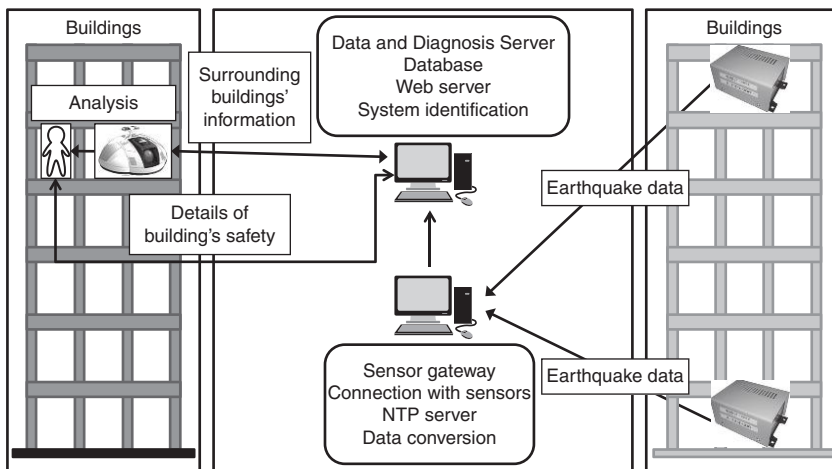
8.6 Use of sensor agent robots and SHM systems with smart sensors for active evacuation.

available at Mita and Ise (2012). A prototype sensor agent robot used in this system is shown in Fig. 8.7. The sensor agent robots are able to measure the response of the building. Using the data obtained by the robots, active evacuation guidance is provided considering the level of damage and other surrounding conditions. The safety level of other buildings will be notified through communication with the SHM system.

An example of a commercial SHM system developed for emergency control (VissQ: see the URL listed in reference) is shown in Fig. 8.8. This system is installed in the World Trade Center Building in Tokyo, where the response during the 2011 earthquake was monitored. VissQ provides data to meet emergency control center to quickly understand the structure's safety. As a result, the tenants did not need to evacuate. The recovery procedures were immediately initiated thanks to the information from the system. After this

8.7 Sensor agent robot 'e-bio' with a notebook PC and a laser range finder.
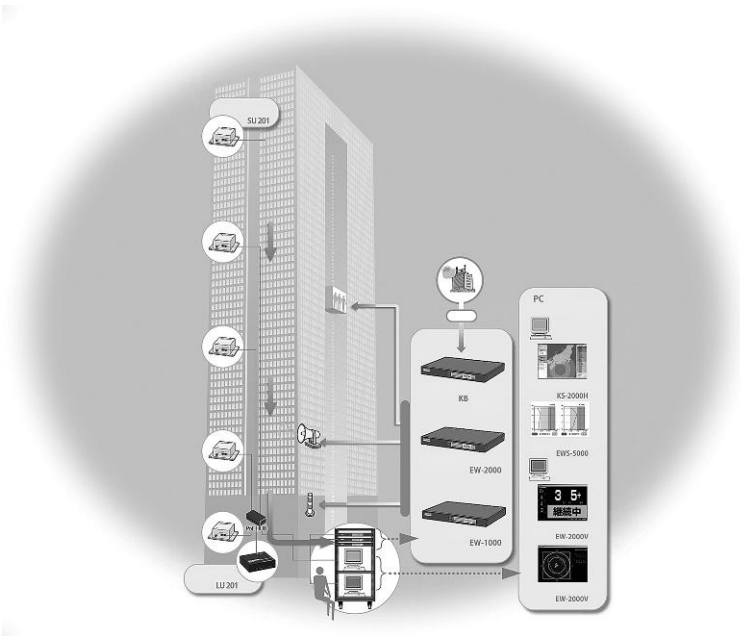
8.8 Commercial SHM system for emergency control.
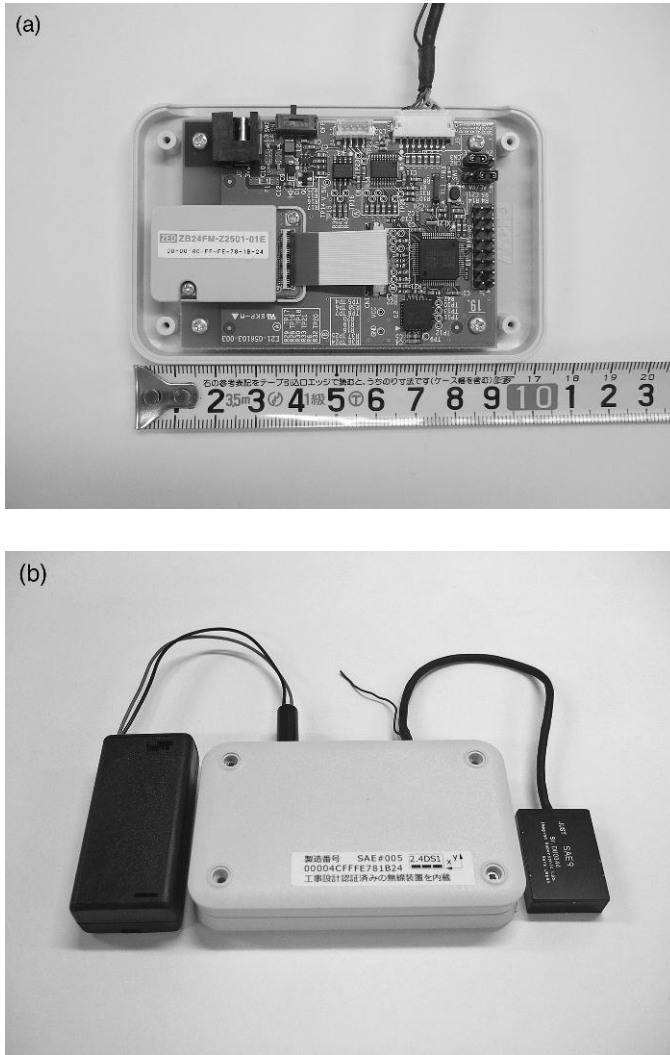
8.9 Smart VA sensor (a) and smart AE sensor (b).

event, the usefulness of SHM system has been recognized. Thanks to this heightened awareness, many tall buildings are now being equipped with this type of SHM system.

8.4   Smart sensor devices to detect local damage

We define a smart sensor as a sensor device with processing and communication capability. Many smart sensor devices have been proposed and

Table 8.2 Outline of smart VA and AE sensors

Targeted behavior	Structure's dynamic properties	Concrete crack
Sensor and monitoring method	Global monitoring by Smart VA Sensor	Local monitoring by smart AE Sensor
Estimation	Three acceleration indices (the number of zero-crossing points, the maximum acceleration value and the sum of the absolute acceleration values)	Numbers of threshold excess for the four different levels in AE signals
Location of sensors	The top, medium level and/or the base of the building	The foot of the shear walls, the columns and/or the edge of the girders where the concrete crack may occur

developed. The sensor device depicted in Fig. 8.5 is a typical example. En *et al.* (2010) developed two types of smart sensor devices for monitoring damage in the reinforced concrete buildings. They are the vibration sensor (VA sensor) and the acoustic emission sensor (AE sensor). An outline of the smart sensors is listed in Table 8.2. The photos are shown in Fig. 8.9. Those sensor devices minimize the data transmitted to the network by analyzing those data to identify local damage using the processor on the device. Thus quick diagnosis is possible.

8.5   Conclusion

Sensing systems for monitoring buildings have entered a new era. In the past, most monitoring systems were aimed at recording responses to strong ground motion, to verify the structural design and identify deficiencies of the structural system. In this new era, the purpose of monitoring has been widened to using the monitored information to evaluate the structural integrity for emergency control in real time as well as for maintenance. In this chapter, several activities associated with SHM were introduced. Following the 2011 earthquake, the importance of monitoring buildings has been widely recognized by engineers and the public.

However, the installation of an SHM system is still expensive, limiting their use to tall, or otherwise important, buildings. Installing sensors only for earthquakes may not be practicable for normal buildings. The SHM system should be integrated into systems for other purposes, such as environmental control systems. Such integration of sensor and monitoring systems will spread the use of sensing and monitoring in buildings.

8.6   References

BRI: http://smo.kenken.go.jp/ (accessed on 11 September 2012).

CSMIP: http://www.conservation.ca.gov/cgs/smip/Pages/about.aspx (accessed on 11 September 2012).

En K, M Nakamura, T Yanase, S Ikegaya and K Yoneyama (2010), 'Structural health monitoring system applied to RC buildings with smart sensors and wireless network', *Proc. of 5th World Conference on Structural Control and Monitoring*, 12–14 July 2010, Tokyo, 5WCSCM-019.

JSSI, Report of Response-Controlled Buildings Investigation Committee, *JSSI*, 26 January 2012.

Kasai, K., A. Mita, H. Kitamura, K. Matsuda, T. A. Morgan, and A. W. Taylor (2013), 'Performance of Seismic Protection Technologies During the 2011 East Japan Earthquake', *Earthquake Spectra*, Vol. **29**, No. S1, pp. S265–S293.

Mita, A. and D. Ise (2012), 'Active Evacuation Guidance and Structural Health Monitoring System for Buildings using Sensor Agent Robots', *Proc. 4th Asia-Pacific Workshop on Structural Health Monitoring (4APWSHM)*, Melbourne/Australia, 5–7 December 2012.

Mita, A., H. Sato and H. Kameda (2011), 'Platform for structural health monitoring of buildings utilizing smart sensors and advanced diagnosis tools', *Journal of Structural Control and Health Monitoring*, Vol. **17**, No. 7, pp. 795–807.

Okada K (2011), 'Internet-based structural health monitoring system', *Structure*, No. 119, pp. 60–61 (in Japanese).

USGS: http://nsmp.wr.usgs.gov/ (accessed on 11 September 2012).

VissQ (Visual Sensor System for Quakes) catalog, Hakusan Corporation, http://www.hakusan.co.jp/. (accessed on 11 September 2012).

Xu, Y.L., S.S. Law, Y.Q. Ni, Y. Xia, and S. Zhu, X.L. Ding, J.N. Cao and H.Y. Tam (2012), 'Performance-Based Health Monitoring of Large Civil Structures', *Proc. ANCRiSST2012*, 27–28 July 2012, Bangalore, India.

9

Sensing solutions for assessing and monitoring super-tall towers

Y. Q. NI, The Hong Kong Polytechnic University, Hong Kong

Abstract: The Canton Tower (CT) is a super-tall structure with a total height of 610 m. To ensure the safety and serviceability of this landmark structure during construction and operation, a sophisticated long-term structural health monitoring (SHM) system consisting of more than 700 sensors of 16 types has been designed and implemented for real-time monitoring of the structure at both in-construction and in-service stages. It provides a unique engineering paradigm for monitoring and assessing mega-structures using sensory technology, and a test-bed for an SHM benchmark study of super-tall structures with use of real-world monitoring data. This chapter describes the design, implementation, operation, and management of this life-cycle SHM system, and the integration of the SHM system with the vibration control system and the renewable energy technology deployed on the CT. The monitoring data from the CT during more than ten typhoons and earthquakes (including the recent devastating Great East Japan Earthquake) are presented, and an SHM benchmark problem with the instrumented CT being a host structure is outlined.

Key words: super-tall structure, structural health monitoring (SHM), sensory system, seismic and typhoon-induced response, benchmark study.

9.1 Introduction

Recent years have witnessed the growth of super-tall structures, built or being built in a number of densely urbanized cites worldwide. Examples are the Burj Khalifa, 828 m high, in Dubai, United Arab Emirates; the Pagcor Tower, 665 m high, in Manila, Philippines; and the Digital Media City Landmark Building, 640 m high, in Seoul, Korea; the Russia Tower, 612 m high, in Moscow, Russia. In China, at least three skyscrapers over 600 m in height (the Shanghai Tower, 632 m high; the Pingan International Finance Center, 646 m high; and the Wuhan Greenland Center, 606 m high) are currently being built. With the increasing complexity of modern skyscraper structures, civil engineers have faced the challenging task of ensuring not only the life-cycle safety of these structures but also their robustness in

246

resisting natural and/or man-made hazards such as earthquakes, typhoons, fires, and so forth at both in-construction and in-service stages. Structural health monitoring (SHM) provides a viable technique to address these challenges.

The applications of SHM technology to building structures are not as widespread as its applications to bridge structures. The significance of implementing SHM systems for large-scale bridges, to secure structural and operational safety and to issue early warnings on damage or deterioration prior to costly repair or even catastrophic collapse, has been well recognized by bridge administrative authorities. In Hong Kong, SHM technology for bridges has evolved for over 15 years since the implementation of the 'Wind and Structural Health Monitoring System' on the suspension Tsing Ma Bridge in 1997. All five long-span cable-supported bridges in Hong Kong (the Tsing Ma Bridge, the Kap Shui Mun Bridge, the Ting Kau Bridge, the Western Corridor Bridge, and the Stonecutters Bridge) have been instrumented with sophisticated long-term SHM systems (Wong, 2004; Ko and Ni, 2005; Wong, 2007; Wong and Ni, 2009a), and several investigations on using the monitoring data for bridge health and condition assessment have been carried out (Ko *et al.*, 2009; Wong and Ni, 2009b; Ni *et al.*, 2010b; Wong and Ni, 2011; Zhou *et al.*, 2011a; Zhou *et al.*, 2011b; Ni *et al.*, 2012b; Xia *et al.*, 2012; Ye *et al.*, 2012). The successful implementation and operation of SHM systems for the bridges and the experience gained from practice and research over the past 15 years have also promoted extended applications of this technology from long-span bridges to high-rise structures. The instrumentation system for the Canton Tower (CT), 610 m high with over 700 permanently installed sensors, is such an engineering paradigm of the application of SHM technology to high-rise structures (Ni *et al.*, 2009).

Long-term SHM of high-rise structures, especially for super-tall towers and out-of-codes buildings, is very helpful in understanding the real performance of these structures under abnormal loading conditions and ensuring their safety in the whole life-cycle. It provides the most authentic information for assessing structural integrity, serviceability, and reliability. Continuous awareness of the evolution of the structural condition via long-term monitoring allows for making informed decisions regarding the maintenance and management of the instrumented structures. Not only does it enable the identification of structural deterioration at the earliest possible stage, but also tells whether evacuation of the occupants of the building is necessary immediately after a major hazardous event. SHM data are also significant for verifying the parameters and assumptions adopted in seismic- and wind-resistant design of super-tall and out-of-code structures. For example, it is expected that the wind characteristics above and below the maximum gradient wind level (it is at the height of 450 m according to the Chinese code). However, the field measurement data of wind characteristics

and wind-induced structural responses at heights above 450 m were scarcely available. With SHM systems instrumented on super-tall structures, the wind characteristics (wind speed, wind direction, wind pressure, and their derivatives such as wind spectrum, turbulence intensity, turbulence integral length scale, gust factor) above the maximum gradient wind level can be monitored long-term and compared with those below the maximum gradient wind level. Correlating the measured wind characteristics with the corresponding structural responses enables the verification of wind-resistant design. The increasingly available data entail improved wind-resistant design of super-tall structures in future.

9.2   Structural health monitoring (SHM) system for the Canton Tower

The CT, formerly named Guangzhou New TV Tower, located in Guangzhou, China, has an assured place among the super-tall structures worldwide by virtue of its total height of 610 m. As shown in Fig. 9.1, it consists of a 454 m high main tower and a 156 m high antenna mast. The main tower is a tube-in-tube structure, consisting of a steel lattice outer structure and a reinforced concrete inner structure. The outer structure has a hyperboloid form, which is generated by the rotation of two ellipses, one at the ground level and the other at an imaginary horizontal plan 454 m above the ground. The tightening caused by the rotation between the two ellipses forms the characterizing 'waist-line' of the tower. The cross-section of the outer structure is 50 m × 80 m at the ground, 20.65 m × 27.5 m (minimum) at the waist level (280 m high), and 41 m × 55 m at the top (454 m high). The outer structure is made up of 24 inclined concrete-filled tube columns, which are transversely interconnected by steel ring beams and bracings. The inner structure is an ellipsoidal shape with a constant cross-section of 14 m × 17 m throughout its height. The centroids of the outer structure and the inner structure are different in the plane. The inner structure and the outer structure are connected at 37 floors. The antenna mast is a steel structure founded on top of the main tower. The lower part of the antenna mast is a steel lattice structure with an octagon cross-section. The diagonal length of the octagon is 14 m at the bottom and decreases as the height rises. The upper part of the antenna mast is a steel box structure. The form of the cross-section varies with height, being square, hexagon and square again. The side length of the square cross-section on the top of the antenna mast is only 0.75 m. The CT serves various functions – television and radio transmission, sightseeing, catering, and entertainment embracing an orbital Ferris wheel, a ceremony hall, observatory decks, 4D cinemas, revolving restaurants, skywalk, etc. With the completion of structural construction in May 2009, the CT was open for the 2010 Asia Games.

9.1 The Canton Tower (CT).

To ensure the safety and serviceability of this landmark structure during construction and operation, a sophisticated long-term SHM system was designed and implemented by The Hong Kong Polytechnic University for real-time monitoring of the CT at both in-construction and in-service stages (Ni *et al.*, 2009). Figure 9.2 illustrates the integrated in-construction and

in-service SHM system, which consists of more than 700 sensors including a weather station (air temperature, humidity, barometric pressure, rainfall), a total station, a global positioning system (GPS) with two rover receivers, a seismograph, accelerometers, anemometers, wind pressure sensors, fiber Bragg grating (FBG) strain and temperature sensors, vibrating wire strain gages and temperature sensors, electrical resistance temperature sensors (thermometers), corrosion sensors, digital cameras (a vision-based displacement measurement system), tiltmeters, and theodolites. Table 9.1 summarizes the sensors deployed on the CT for in-construction and in-service monitoring. Among them, a total of 200 FBG sensors have been deployed on the main tower and the antenna mast, as shown in Fig. 9.3, to provide real-time and long-term monitoring of dynamic strain and temperature. A hybrid tethered and wireless data acquisition network in conjunction with 13 data acquisition units (DAUs) during in-construction

Table 9.1 Sensors deployed on CT for in-construction and in-service monitoring

No.	Type of sensors	Monitoring item	Number of sensors	Sampling rate (Hz)
1	Weather station	Temperature, humidity, rain, air pressure	1	1
2	Anemometer	Wind speed, wind direction	2	1
3	Wind pressure sensor	Wind pressure	4	50
4	Accelerometer	Acceleration	22	100
5	Seismograph	Earthquake ground motion	1	100
6	Vibrating wire gage	Strain and temperature	404	1
7	Thermometer	Temperature of structure	136	1/60
8	Optical fiber strain sensor	Dynamic strain	144	50
9	Optical fiber temperature sensor	Temperature of structure	56	50
10	GPS	Displacement	2	10
11	Digital camera system	Displacement	3	60
12	Laser plummet	Inclination	1	50
13	Optical fiber tiltmeter	Inclination	2	50
14	Corrosion sensor	Corrosion of reinforcement	2	–
15	Total station	Leveling	1	–
16	Theodolite	Elevation	2	–

Web-based online monitoring system

Control center

In tower

User group

Internet

ADSL modem

RJ45 Firewall RJ45

Structural health evaluation system

R145

Manager on site

Data management system

RJ45

Inspection and maintenance system

Data processing and control system

RJ45

Wireless transmission system

(In-constructon monitoring)

(In-service monitoring)

Wired transmission system

Data logger GPS

Data logger Anemometer

Data logger Strain gage

Data logger Temperature sensor

Internet

Internet

Emergency management system

Data analysis and post-processing system OCTS

Department in charge

Email

Message

Alarm buzzer

Emergency response procedure

Unit conversion

Structural analysis

Decision making

Signal processing

Digital filtering

Feature extraction

Spectrum analysis

....

OCTS

Data logger Network camera

Data logger Seismograph

9.2 SHM system for CT.

9.3 FBG sensors for dynamic strain and temperature monitoring: (a) Main tower; (b) Antenna mast.

(Continued)
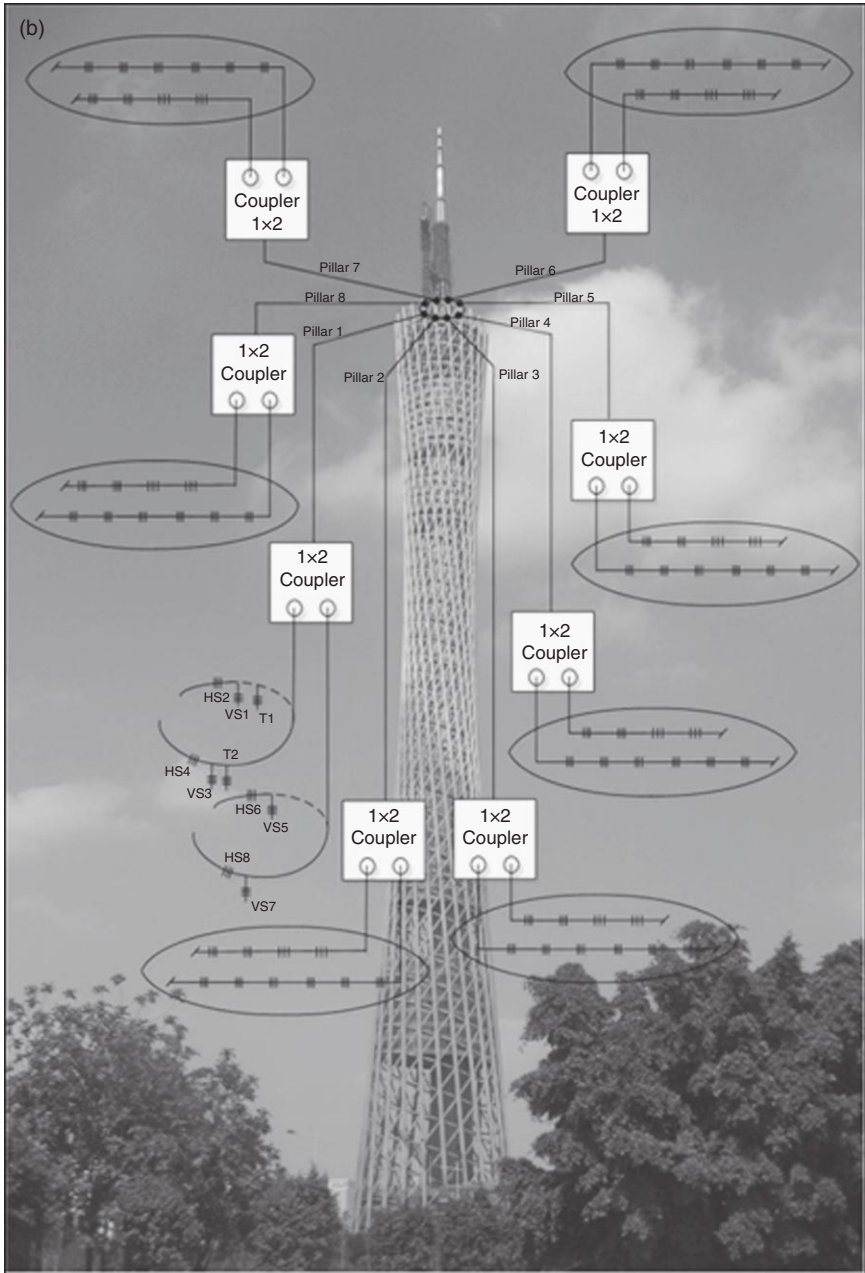
9.3 Continued

monitoring and five DAUs during in-service monitoring has been adopted in the SHM system.

As detailed later, the on-line SHM system has been devised to possess the following features: (i) modular architecture for easy maintenance and upgrade; (ii) life-cycle SHM with the integration of in-construction monitoring and in-service monitoring; (iii) dual function of on-line health monitoring and real-time feedback control with the integration of SHM and vibration control; (iv) integration with renewable energy technology (solar photovoltaic and wind turbine systems) to monitor the power generation efficiency and operational condition; (v) innovative sensors and customized design fit for special circumstances; (vi) hybrid tethered and wireless data transmission networks customized for harsh operational conditions; (vii) user-friendly graphical user interface for easy operation; (viii) innovative structural health evaluation methodologies catering for structural maintenance and management purposes; (ix) all-round protection customized for severe surrounding environments; (x) remote expert access through web-based data collection; and (xi) popularization of scientific knowledge through a virtual reality system integrated with sightseeing.

9.3   Integrated SHM and vibration control

The SHM system for the CT has been designed to have a special function of monitoring and verifying the effectiveness of vibration control devices installed on the structure. It is a unique and interesting use of SHM. A hybrid control system, consisting of two tuned mass dampers (TMDs) coupled with two active mass dampers (AMDs), as illustrated in Fig. 9.4a, was installed at a floor 438 m from the tower base for mitigating wind-induced vibration of the main tower, while two TMDs, as shown in Fig. 9.4b, were suspended at the heights of 571 and 575 m respectively for vibration suppression of the antenna mast. Each of the water-tank TMDs at 438 m level weighs 600 tons, which is about 0.3% of the total weight of the structure (the tower weighs about 200 000 tons and its fundamental frequency is about 0.1 Hz); the allowable stroke is ±1.2 m (restrained at ±0.8 m). Each of the AMDs weighs 100 tons and the stroke is ±2.0 m. The AMDs are seated on top of the water-tank TMDs. The two TMDs inside the antenna mast are suspended at heights of 571 and 575 m, respectively. Each TMD, weighing 2 tons (the antenna mast weighs approximately 1800 tons), is a lead-filled steel ball with a diameter of 0.65 m. Its allowable stroke is ±40 cm.

To command the AMDs (made from linear motion actuators), it is necessary to establish a structural response feedback system that will provide

(a)

(b)
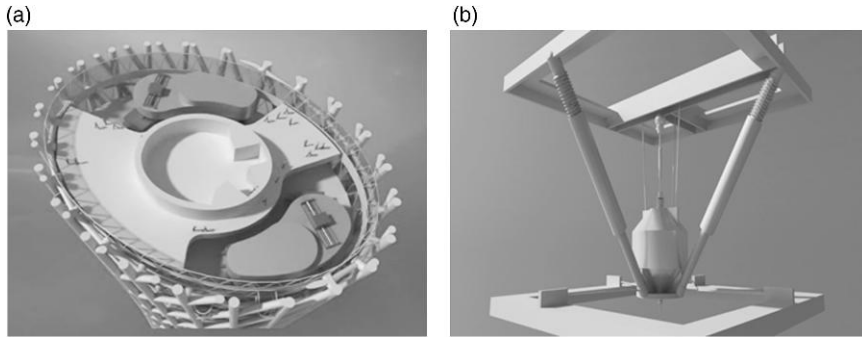
9.4 Vibration control devices for CT: (a) TMDs + AMDs for main tower; (b) Suspended TMDs for antenna mast.

comprehensive information for real-time vibration control. As illustrated in Fig. 9.5, the SHM system for the CT has been devised to integrate with the vibration control system so that reliable and real-time monitoring data can be obtained for feedback vibration control, thus enhancing the effectiveness of the control system. With such integration, the signals from the anemometers and the seismograph of the SHM system will be provided on-line to the vibration control system for making decisions on activating or locking the control system, which is designed for wind-induced vibration control only. In addition, the signals from the ad hoc transducers specific for feedback vibration control will also be transmitted to the SHM center and compared with the measured structural response signals by the SHM system to detect possible faults of the control-specific transducers. Table 9.2 lists the total 20 transducers deployed specifically for real-time feedback vibration control. Twelve accelerometers are installed on the main tower, water tanks, and antenna mast for measuring acceleration responses: four accelerometers are deployed respectively at the 1/4, 1/2, and 3/4 heights of the main tower and at the water-tank level (438 m in height), with one at each level; four accelerometers are placed at the centroids of the water tanks with two on each water tank; and four accelerometers are mounted on the antenna mast at the levels of 529 and 578 m high with two at each level. Four velocity meters are used to monitor the velocity of the main tower, with one at each of the 1/4, 1/2, 3/4 heights of the main tower and one at the water-tank level (438 m in height). Four displacement transducers are employed for measuring displacements of the water tanks, with two for each water tank. As shown in Fig. 9.5, the SHM system is also integrated with the renewable energy systems deployed on the structure (wind turbines at the height of 168 m and solar photovoltaic panels at the heights between 438 and 443 m) for signal acquisition and control.

Solar photovoltaic panels

Wind turbine
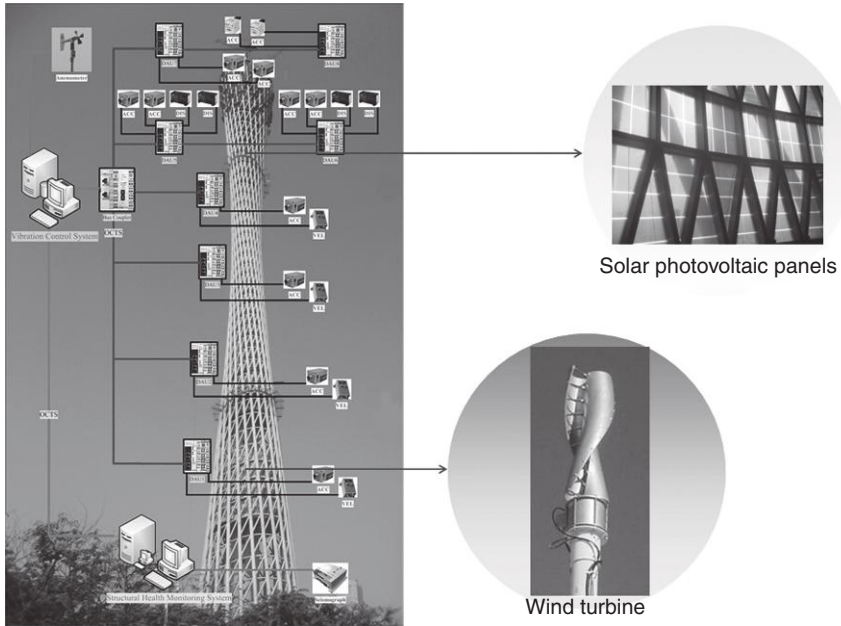
9.5 Integration of SHM with vibration control and renewable energy technology.

Table 9.2 Sensors deployed for feedback vibration control of CT

No.	Type of sensors	Monitoring item	Number of sensors	Sampling rate (Hz)	Location
1	Accelerometer	Acceleration	12	100	¼, ½, ¾ height of main tower, 438 m (water-tank level), 529 m, 578 m
2	Velocity meter	Velocity	4	100	¼, ½, ¾ height of main tower, 438 m (water-tank level)
3	Displacement transducer	Displacement	4	20	438 m (water-tank level)

9.4   Verification of long-range wireless sensing technology

The super-tall CT also provides a unique test-bed for investigating the optimal design of sensor networks and efficient techniques for long-range wireless monitoring (Ni *et al.*, 2011a; Ye and Ni, 2012; Zhang *et al.*, 2012). The

hybrid tethered and wireless communication network of the SHM system for the CT enables the verification of wireless monitoring data. Wireless systems have been implemented in the CT for both static and dynamic response monitoring. A wireless system is operated for synchronous acquisition of strain and temperature data and real-time data transmission from the DAUs to the site office as shown in Fig. 9.6. The vibration of the structure is monitored mainly by using a wired cabling network, while a wireless system is also adopted *in situ* for complementary vibration monitoring (Ni *et al.*, 2011a). The wireless sensing prototype, based on the Stanford Module (Wang *et al.*, 2007), adopts an integrated hardware and software design to implement a simple star topology wireless sensor network. Its wireless sensing unit consists of three functional modules: sensing interface, computational core, and wireless communication channel. In the prototype implementation, the base station of the wireless sensing system is located at the ground level of the CT to receive the acceleration response data from the wireless sensing units placed at different heights up to the top of the main structure. Figure 9.7 shows a comparison of the ambient vibration responses (acceleration time-histories and power spectral densities) of the CT at the height of 225 m monitored by the wireless system and the tethered system, respectively.

9.5   Sensor fusion for SHM

Because the CT has been instrumented with sensors in 16 types, this project offers a platform to explore sensor and data (information) fusion for SHM. It is difficult to calibrate/verify the dynamic displacement measurement data acquired by a GPS. Such calibration or verification is usually carried out in an indirect way by doubly integrating the acceleration measurement data. However, this approach cannot provide an accurate reference of dynamic displacement due to unknown initial displacement and numerical error (Smyth and Wu, 2007). A vision inspection system (VIS), as shown in Fig. 9.8, has been developed and used together with the GPS system for dynamic displacement measurement of the CT (Ni *et al.*, 2010a). It is composed of an industrial digital camera, an extended zoom lens, a laptop (or desk) computer, a Giga LAN wire, and tailor-made software. In this system, the Prosilica GigE camera is connected with the Navitar 24X Zoom Extender lens, and the software adopts the pattern matching technique. With the help of the digital image processing software, the lens can zoom into any target distinct from its surroundings (with different contrast), and in real time track the coordinates of the target that are continuously being streamed to the computer. When the targeting location is accessible, a super-bright (light-emitting diode) LED lamp can be fixed at the targeting location to enable measurement in night or bad weather as well.
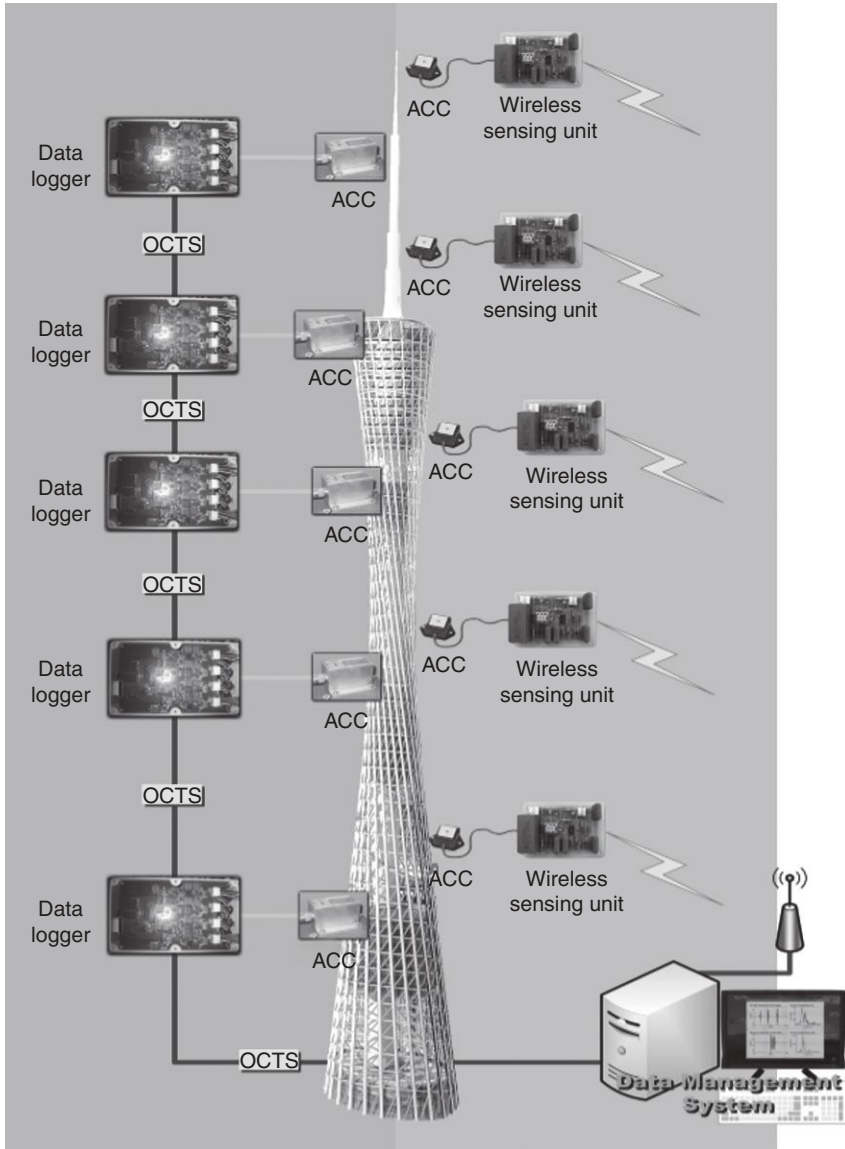
9.6 Wireless data acquisition.

Image processing in the software system involves two phases: an off-line learning phase in which the template is processed, and a matching phase that can be executed in real time. The tailor-made software includes target calibration, projection of the captured image, and calculation of the actual displacement using the scaling factor and the number
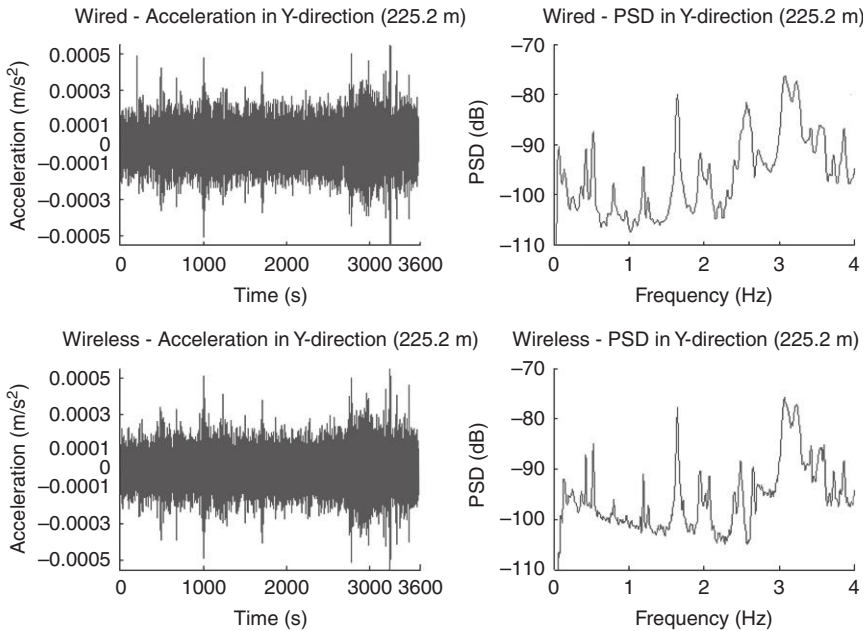
9.7 Comparison of ambient vibration responses monitored by wireless and tethered systems.

of pixels moved, and display and storage of the calculated displacement. The quantity (amount) of information to be processed in real time depends on the number of pixels per frame and the number of frames per second. The region of interest (ROI) for target recognition does not need to cover the whole image frame. It is only necessary to trace the target, reducing the amount of information to be processed in real time. The image processing is performed only within the confined region, and thus the time for image processing is minimized.

The devised VIS is capable of remote long-distance (500–1000 m) dynamic displacement measurement with a sampling frequency between 5 and 60 Hz. Plate XI in the color section between pages 374 and 375 shows the horizontal displacement responses of the CT at the top of the main tower measured by GPS after eliminating spikes and abnormal shifts (generated due to specific positioning of the satellites and electro-magnetic interference in the site) and by VIS. It is found that the VIS-measured displacement response time-history coincides well with the GPS-measured displacement response time-history after corrections, while the former does not exhibit spikes and abnormal shifts. Therefore, the data measured by VIS can directly be used to verify and correct the GPS-measured displacement responses.
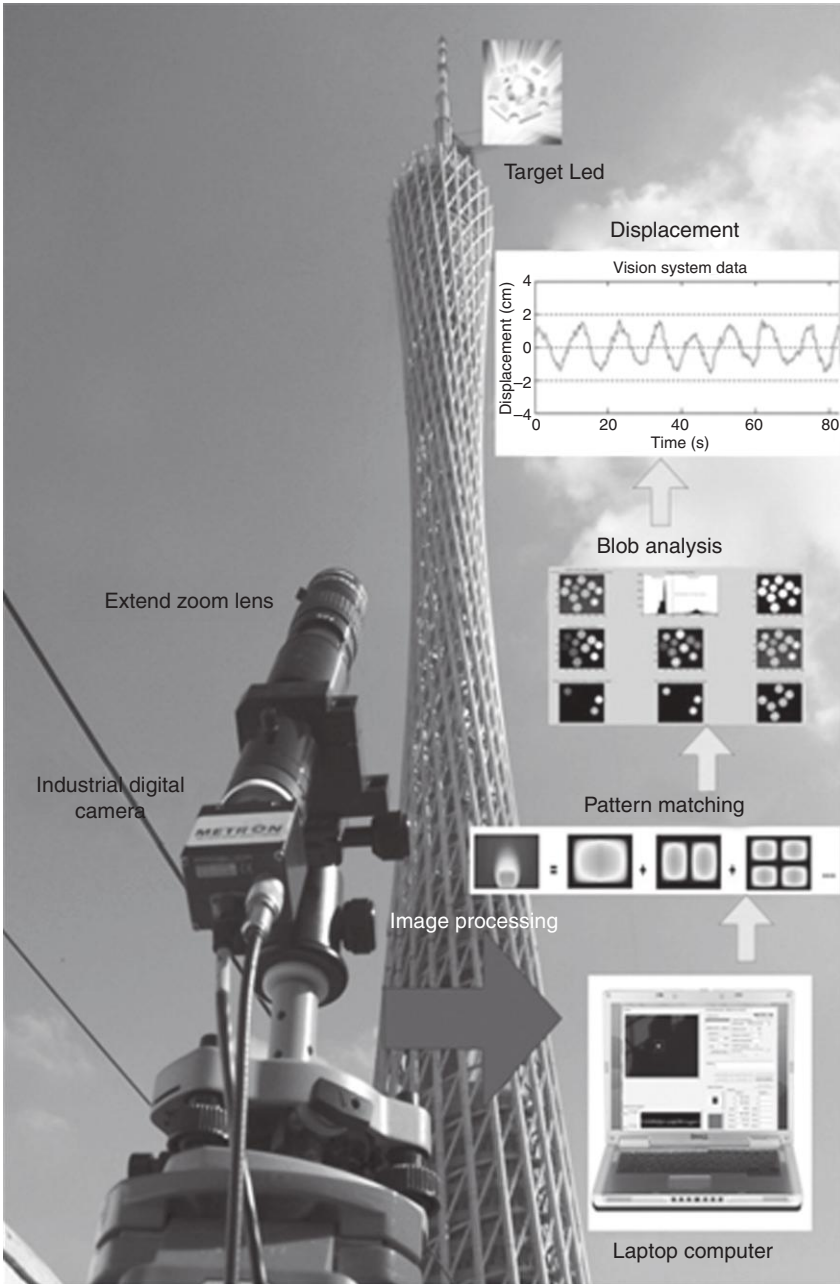
9.8 A Vision inspection system (VIS).

The SHM system for the CT is a long-term monitoring system. Some of the sensors and sensing cables are located on the surface of structural components or outside. To ensure long-term stable system operation, all-round protection is required for all sensors, sensing cables, and data acquisition and transmission components (Fig. 9.9). To avoid the problems of paint-off and weak weld strength, all of the pre-embedded pieces were welded on the steel members prior to installing the sensors and sensing cables on to the outer structure. Detachable stainless steel cases and wire ducts were utilized to protect the sensors and sensing cables, while aluminum alloy protection cases were adopted for the optical fiber sensors. Specific sealant was used to seal the apertures. These measures made the sensor system not only presentable in outlook, but also waterproof and dustproof. Pre-protection measures for the embedded sensors are also necessary prior to sensor installation to prevent damage to the sensors (e.g., vibration wire strain gages, which were embedded into the inner structure). Sensing cables to the DAUs were safeguarded by zinc plating steel tubes. The DAUs were protected by special industrial mainframe-boxes to ensure lightning protection, electromagnetic interference (EMI) prevention, moisture-proofing, temperature control, etc. Aluminum alloy wire ducts were adopted to protect the sensing cables between the control center and the sub-stations. To date, the SHM system has performed well while the CT was struck by lightning several times in the past 5 years.

The CT, designed with the function of sightseeing and cultural entertainment, is deemed to be an ideal and unique place for tourist sightseeing and at the same time to offer visitors opportunities to learn science. The tower allows tourists to understand the SHM's and vibration control systems' operations, and the scientific principles behind them. In addition, visitors can view its real-time data. Display terminals are distributed at various popular zones, such as at the high-tech exhibition and ceremony halls. The terminals are connected by local networks and controlled by the SHM system at the control center. Both professional and educational versions of the interactive software demonstrating the SHM system have been developed and provided at the display terminals.

9.6   Monitoring data during typhoons and earthquakes

Because the data acquisition system is operated automatically to continuously acquire monitoring data, the SHM system has detected the structural responses of the CT during a number of typhoon and earthquake events (including the devastating Tohoku Earthquake in Japan in 2011) in the past 5 years.

All round protection for long-term operation

Structural health monitoring system

1. Installation of pre-embedded pieces in workshop;
2. Painted pre-embedded pieces;
3. Protection of sensor installed in steel column;
4. *In-situ* installation of fiber optic sensor;
5. Aluminum alloy protection case for fiber optic sensor;
6. Protection of fiber optic sensor;
7. Protection of vibration wire strain gage;
8. Installation of surface-type vibrating wire strain gage;
9. Protection of embedment-type vibrating wire stain gage;
10. *In-situ* installation of embedment-type vibrating wire strain gage;
11. Protection of accelerometer;
12. Zinc plating steel tube for protection of sensing cable in inner structure;
13. Stainless steel wire duct for protection of sensing cable in circumferential direction;
14. Stainless steel wire duct for protection of sensing cable in vertical direction;
15. Aluminum alloy wire duct for protection of sensing cable in longitudinal direction;
16. Protection of data logger.

9.9 All-round protection for long-term operation of SHM system.

9.6.1   Responses during typhoons

As the city of Guangzhou is located at the edge of the most active typhoon generating area in the world, the SHM system has monitored the wind properties and structural responses of the CT during 12 typhoons since 2008. They are the Neoguri Typhoon (19 April 2008), the Kammuri Typhoon (6 August 2008), the Nuri Typhoon (22 August 2008), the Hagupit Typhoon (24 September 2008), the Molave Typhoon (16 July 2009), the Koppu Typhoon (24 September 2009), the Haima Typhoon (23 June 2011), the Nockten Typhoon (29 July 2011), the Nanmado Typhoon (31 August 2011), the Nalgae Typhoon (4 October 2011), the Vicente Typhoon (24 July 2012), and the Kaitak Typhoon (17 August 2012). Table 9.3 shows the information measured by the anemometer installed on top of the main tower during the 12 typhoons. Figure 9.10 illustrates the wind rose diagrams obtained during six typhoons in 2008 and 2009. Figure 9.11 shows the normalized spectra of the longitudinal wind speed component obtained using the measurement data from the anemometer located on the top of the main tower during the Nockten and Haima Typhoons, and the corresponding fitted curves by the von Karman Spectrum. Figure 9.12 shows the dynamic displacement responses (upper: east–west direction; lower: south–north direction) measured by the GPS at the top of the main tower during the Nuri Typhoon and the structural acceleration responses (upper: east–west direction; lower: south–north direction) measured by the accelerometers positioned at a height of 386 m from the tower base during the Hagupit Typhoon.

Table 9.3 Measured information of twelve typhoons buffeting CT

Typhoon	Date	Duration (min)	Wind direction	Maximum 10-min mean wind speed (m/s)
Neoguri	19 April 2008	1440	Southeast	18.89
Kammuri	6 August 2008	1080	Southeast	28.60
Nuri	22 August 2008	1440	Northeast	17.93
Hagupit	24 September 2008	1560	Southeast	29.45
Molave	16 July 2009	1560	Southeast	24.67
Koppu	24 September 2009	1680	Southeast	20.17
Haima	23 June 2011	960	Southeast	14.33
Nockten	29 July 2011	1440	Southeast	11.65
Nanmado	31 August 2011	1440	Northwest	10.16
Nalgae	4 October 2011	1935	Northeast	11.23
Vicente	24 July 2012	2820	Southeast	31.11
Kaitak	17 August 2012	1200	Northeast	17.65
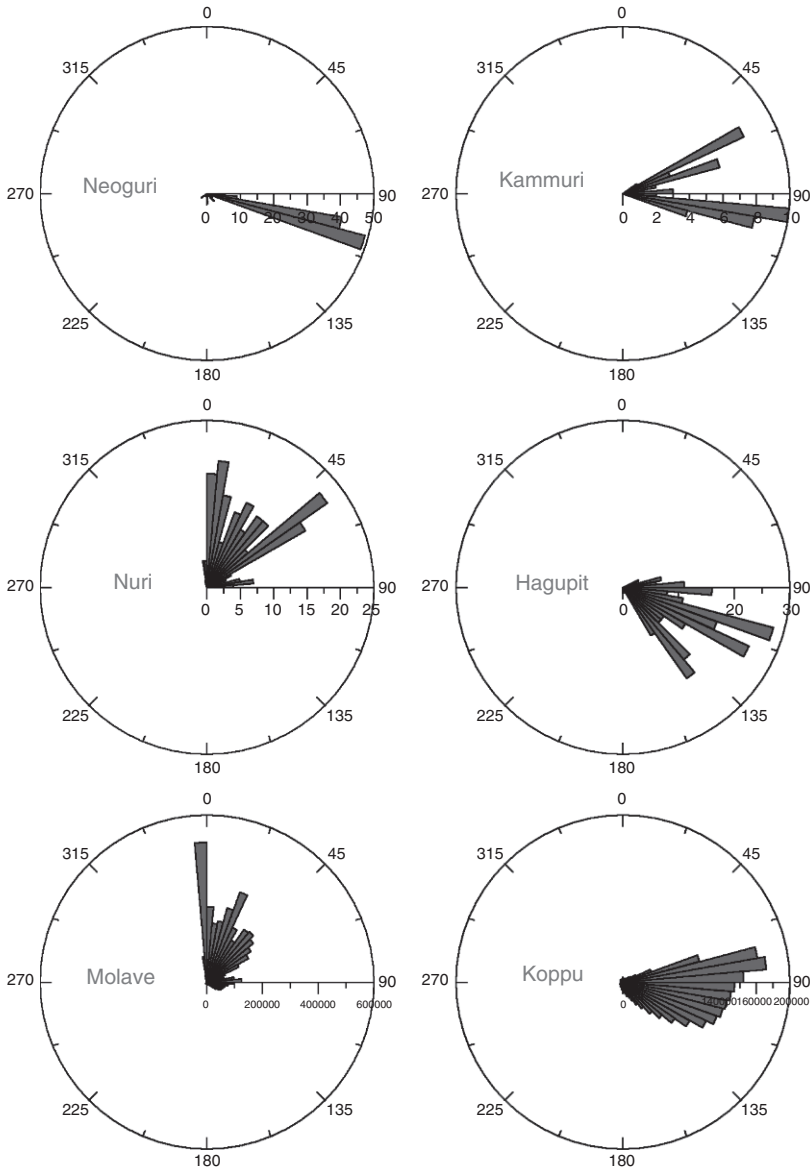
9.10 Wind rose diagrams obtained during six typhoons.

9.6.2   Responses during earthquakes

From 2008 to 2012, the SHM system has successfully monitored the seismic responses of the CT during more than ten earthquake events. Table 9.4

(a)

(b)
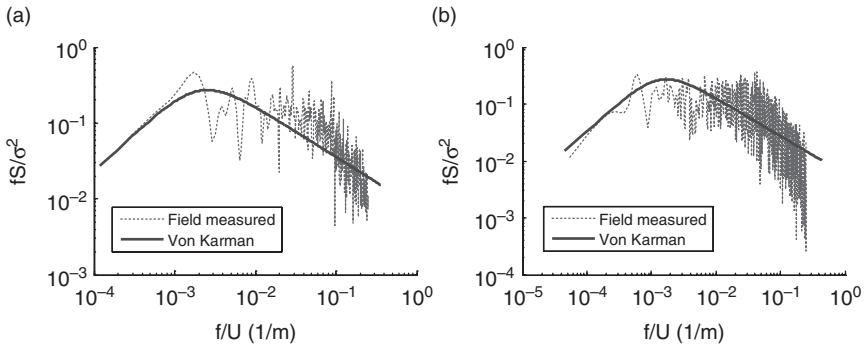
9.11 Normalized spectrum of longitudinal wind speed component: (a) Nockten Typhoon; (b) Haima Typhoon.

lists the information of 11 earthquakes, during which structural dynamic responses have been detected. Figure 9.13 shows the measured dynamic strain responses in the vertical direction during the Wenchuan Earthquake (12 May 2008). The seismic wave traveled from the epicenter at Wenchuan to the site of the CT in about 7 min, and the seismic wave propagation velocity was estimated to be approximately 3155 m/s. Figures 9.14 and 9.15 show the measured acceleration responses of the CT during the devastating Tohoku Earthquake (11 March 2011) and during the Shan State Earthquake (24 March 2011), respectively. Since the epicenter of the earthquake (Tohoku, Japan) is far from the site of the CT (the distance between the epicenter and the site is more than 3000 km), the structural response during the Tohoku Earthquake is quite small. The seismic response of the CT during the Shan State Earthquake (the distance between the epicenter and the site is about 1800 km) is much larger than the response during the Tohoku Earthquake.

By comparing the structural dynamic responses of the CT during the 'short-distance' earthquakes (Shenzhen and Heyuan Earthquakes), 'middle-distance' earthquakes (Pingtung, Hualien and Kaohsiung Earthquakes) and 'long-distance' earthquakes (Shan State, Tohoku and Sumatra Earthquakes) in both time and frequency domains, it is found that (Ni *et al.*, 2012a): (i) the seismic responses of the CT under the 'long-distance' earthquakes are dominated by low-frequency (0.08–0.2 Hz) components, while the seismic responses under the 'short-distance' earthquakes are dominated by the higher-frequency (1.0–2.3 Hz) components and the frequency components are more plentiful. The seismic responses under the 'middle-distance' earthquakes contain the frequency components between 0.08 and 1.0 Hz; (ii) during the 'long-distance' earthquakes, the maximum seismic acceleration responses in general increase gradually from the bottom to the top (the maximum acceleration response

(a)
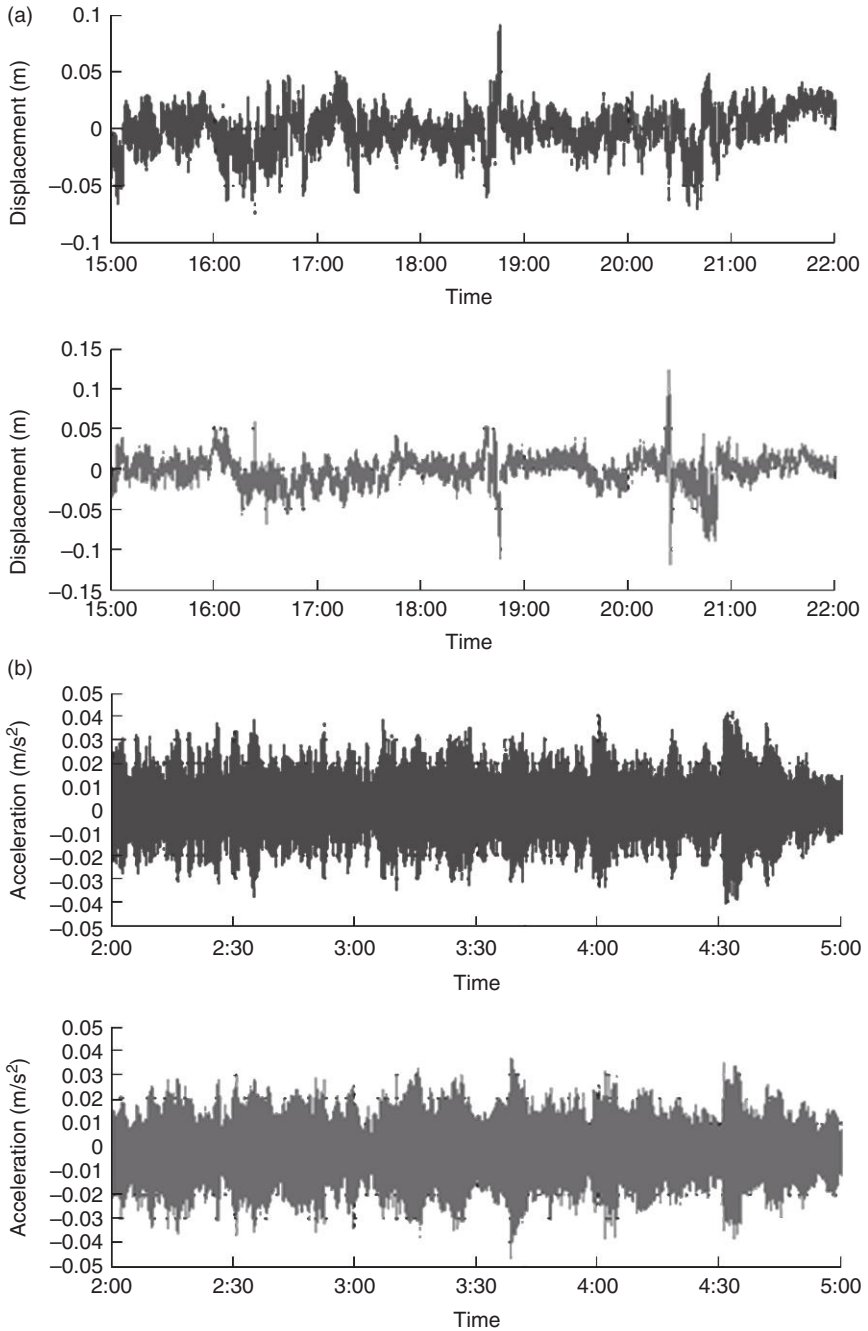
9.12 Measured dynamic displacement and acceleration responses during typhoons: (a) Displacement during Nuri Typhoon; (b) acceleration during Hagupit Typhoon.

Table 9.4 Eleven earthquakes monitored by SHM system on CT

Earthquake	Date	Magnitude (Richter scale)	Depth (km)	Distance to CT (km)
Wenchuan, China	12 May 2008	8.0	19	1330
Hualien, Taiwan	19 December 2009	6.8	44	870
Kaohsiung, Taiwan	4 March 2010	6.4	6	880
Shenzhen, China	19 November 2010	2.8	23	90
Tohoku, Japan	11 March 2011	9.0	24	3220
Shan State, Burma	24 March 2011	7.2	20	1800
Heyuan, China	16 February 2012	4.8	13	160
Heyuan, China	17 February 2012	3.5	9	160
Pingtung, Taiwan	26 February 2012	6.0	20	765
Sumatra, Indonesia	11 April 2012	8.6	20	3140
Sumatra, Indonesia	11 April 2012	8.2	20	3300
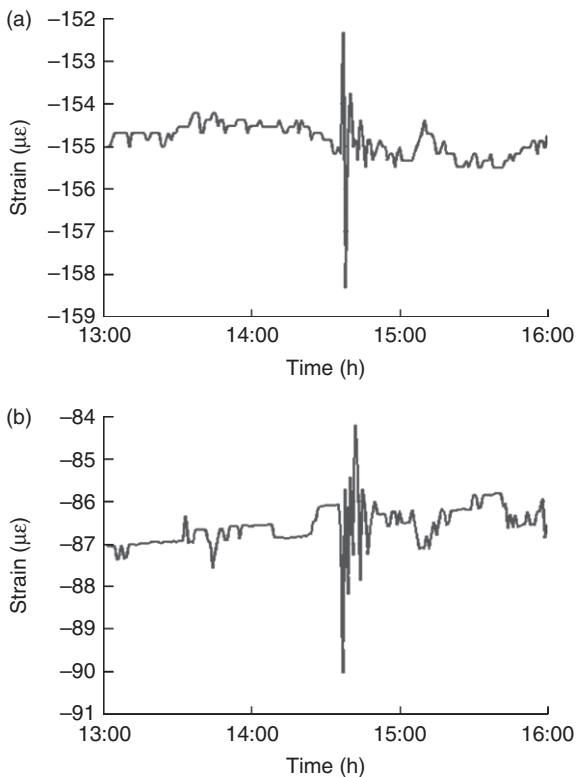
9.13 Measured strain responses during Wenchuan Earthquake (12 May 2008): (a) vertical strain at 173.2 m high; (b) vertical strain at 303.2 m high.

(a)

(b)
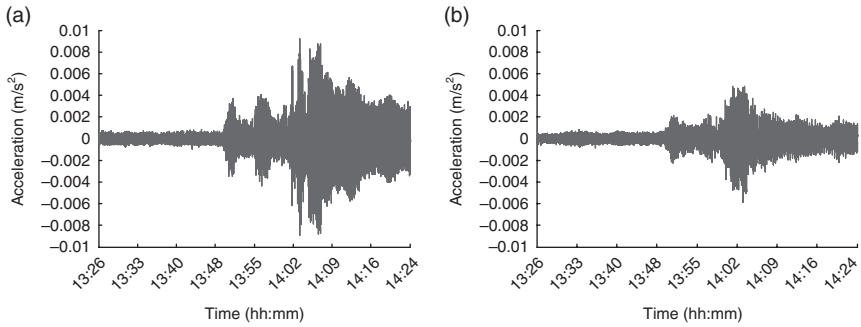
9.14 Measured acceleration responses during Tohoku Earthquake (11 March 2011): (a) in east-west direction at 443.6 m high; (b) in south-north direction at 443.6 m high.

(a)

(b)
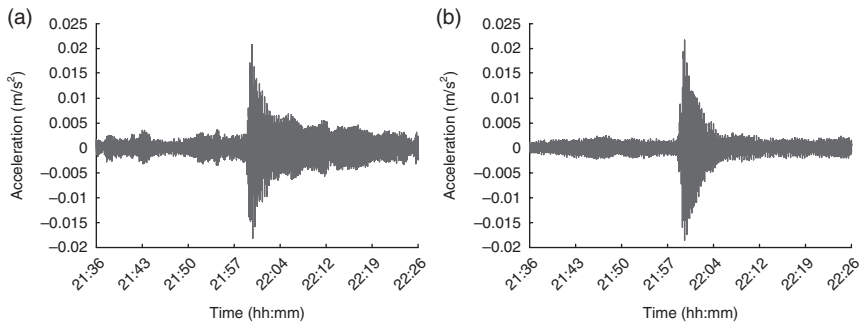
9.15 Measured acceleration responses during Shan State Earthquake (24 March 2011): (a) in east-west direction at 443.6 m high; (b) in south-north direction at 443.6 m high.

occurs at the top level). During the 'short-distance' earthquakes, however, the distribution of the maximum acceleration responses does not conform to such a law; instead it exhibits complex and irregular response characteristics (the maximum acceleration response does not occur at the top level). In the latter case, the high frequency effect becomes significant.

9.7   Strategy for structural health and condition assessment

The structural health and condition assessment strategy for the instrumented CT has been defined in three levels: (i) on-line structural condition evaluation; (ii) target-oriented verification and evaluation; and (iii) off-line structural health and safety assessment. The on-line structural condition evaluation is mainly intended to compare the static and dynamic measurement data with pre-determined thresholds and patterns to provide

a prompt evaluation on the structural condition. The target-oriented verification and evaluation refers to monitoring-based verification/evaluation of temperature-induced deformation and thermal stress, fatigue life of welded steel connections, wind loading and wind effects below and above the gradient wind level, and human comfort during typhoons, among others. The off-line structural health and safety assessment involve a variety of model-based and data-driven damage diagnostic and prognostic algorithms including dynamic-based methods, static-based methods, and hybrid methods. The structural assessment algorithms that have been validated using the field monitoring data from the instrumented bridges in Hong Kong (Ni, 2013) are being implemented with adaptations to achieve the above targets.

9.8   SHM benchmark study

Varieties of structural health and damage detection methods have been proposed by different investigators in the past two decades. However, the feasibility of these methods for real-world applications, especially for applications to large-scale structures, has rarely been examined (Ko and Ni, 2005; Brownjohn and Owen, 2006; Catbas, 2009). A gap still exists between research and the practice in this field, which impedes broader applications of SHM techniques in civil engineering community (Brownjohn, 2007; Farrar and Lieven, 2007; Farrar and Worden, 2010). It is important to establish an SHM benchmark problem with regard to a full-scale structure using field measurement data, aiming to provide an international platform for direct comparison of various algorithms and methods. The participants thus have opportunities to test their SHM techniques using real-world data from a full-scale structure and recognize the obstructions in real-life implementations.

Under the auspices of the Asian-Pacific Network of Centers for Research in Smart Structures Technology (ANCRiSST), an SHM benchmark problem for high-rise structures has been developed by taking the instrumented CT as a host structure. This SHM benchmark problem aims to provide an open platform to researchers and practitioners in the field of SHM for examining the applicability and reliability of their methods to a real high-rise structure with the use of field monitoring data. To facilitate the benchmark study, a reduced-order finite element model (FEM) of the CT has been developed and uploaded in conjunction with typical field measurement data acquired from the structure to the benchmark website (http://www.cse.polyu.edu.hk/benchmark/) for the investigation of modal identification, model updating, loading identification, comfort assessment, SHM-oriented optimal sensor placement, and damage detection. The field monitoring data include 24-h measurement data from 20 accelerometers, one anemometer and one

temperature sensor. The reduced-order FEM shown in Fig. 9.16 is deduced from a validated full-scale 3D FEM consisting of 122 476 elements, 84 370 nodes and 505 164 degrees of freedom (DOFs). With a total of 185 DOFs, the reduced-order FEM comprises 37 beam elements and 37 nodes, having 5 DOFs each. As evidenced in Table 9.5, a good agreement between the reduced-order and full-scale FEMs has been achieved in terms of both modal frequencies and mode shapes for the first 15 modes (comparison between the mode shapes obtained from the reduced-order and full-scale FEMs is given in Ni *et al.*, 2012c). The field monitoring data of the CT during a number of typhoon and earthquake events have also been shared with collaborators worldwide for advanced SHM research. To summarize the outcomes of this collective research, a special session on the SHM benchmark study was organized during the 8th International Workshop on Structural Health Monitoring held on 13–15 September 2011, Stanford, California, USA (Ni *et al.*, 2011b), and a special issue on the SHM benchmark study has been published in the international journal *Smart Structures and Systems* (Ni, 2012).

9.9   Conclusion

The development and implementation of an SHM system capable of fully achieving the objective indices about structural health status (integrity,
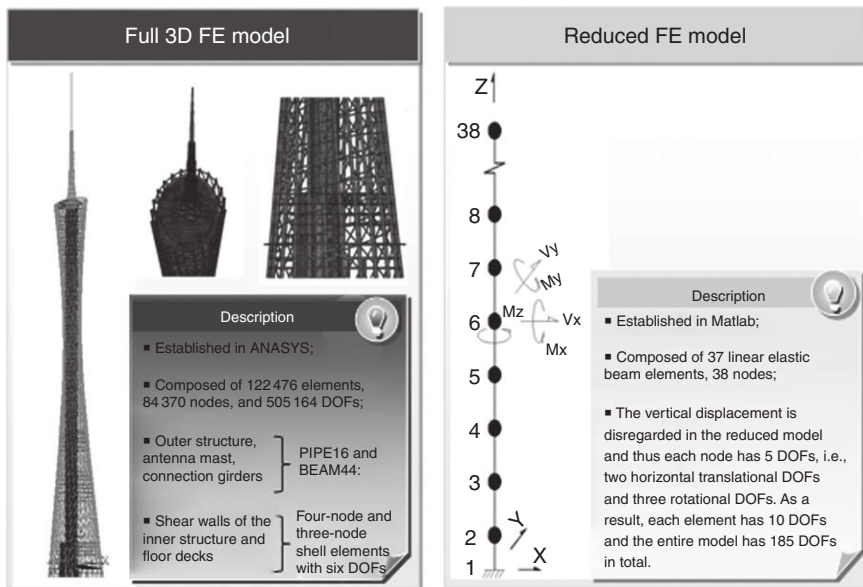
9.16 Development of reduced-order FEM.

Table 9.5 Modal properties of full-scale and reduced-order FEMs

Mode number	Natural frequency (Hz)			MAC (%)
	Full-scale model	Reduced-order model	Difference (%)	
1	0.110	0.110	0.42	99.98
2	0.159	0.159	0.19	99.97
3	0.347	0.347	0.10	99.53
4	0.368	0.368	0.13	99.52
5	0.400	0.399	0.16	99.55
6	0.461	0.460	0.13	99.86
7	0.485	0.485	0.02	99.39
8	0.738	0.738	0.02	99.29
9	0.902	0.902	0.05	99.36
10	0.997	0.997	0.02	99.43
11	1.038	1.038	0.03	98.99
12	1.122	1.122	0.02	99.41
13	1.244	1.244	0.03	98.31
14	1.503	1.503	0.00	96.76
15	1.726	1.726	0.01	97.50

durability, and reliability) pose technological challenges at different levels, ranging from producing innovative sensors to developing advanced diagnostic and prognostic methodologies. It requires the fusion of technologies concerning sensing, communication, data acquisition, signal processing, data mining, information management, computing, decision making, etc. In this chapter, such a comprehensive long-term SHM system for high-rise structures has been showcased by taking the instrumentation system for the CT (610 m high) as an engineering paradigm. After describing the design, implementation, operation, and management of the life-cycle SHM system for the CT, the integration of the SHM system with the vibration control system and the renewable energy technology deployed on the structure has been outlined.

With the on-line SHM system deployed on the CT, the structural dynamic responses during over ten near-field and far-field earthquakes and over ten typhoon attacks have been monitored during 2008 and 2012. The plentiful field monitoring information about the dynamic responses of this supertall structure is not only very helpful for understanding and investigating in-depth the dynamic performance of this structurally complex skyscraper, but also justifies the usefulness and effectiveness of an on-line SHM system for monitoring the loading and response parameters of high-rise buildings under extreme events. Based on the unique field measurement data and well-established FEMs, an SHM benchmark study has been initiated by taking the instrumented CT as a host structure. With the shared measurement data and FEMs, participants can pursue the investigation of modal

identification, model updating, loading identification, comfort assessment, SHM-oriented optimal sensor placement, and damage detection by applying their own methods/algorithms.

9.10 Acknowledgments

9.11 References

Brownjohn, M.W. (2007), 'Structural health monitoring of civil infrastructure', *Philosophical Transactions of the Royal Society A*, **365**(1851), 589–622.

Brownjohn, J.M.W. and Owen, J.S. (2006), 'Suggestions for future research, development and application of bridge health monitoring systems', in: Cruz, P.J.S., Frangopol, D.M. and Neves, L.C. (eds.), *Bridge Maintenance, Safety, Management, and Life-Cycle Performance and Cost*, Taylor and Francis Group, London, UK (CD-ROM).

Catbas, F.N. (2009), 'Structural health monitoring: applications and data analysis', in: Karbhari, V.M. and Ansari, F. (eds.), *Structural Health Monitoring of Civil Infrastructure Systems*, Woodhead Publishing, Cambridge, UK, 1–39.

Farrar, C.R. and Lieven, N.A.J. (2007), 'Damage prognosis: the future of structural health monitoring', *Philosophical Transactions of the Royal Society A*, **365**(1851), 623–632.

Farrar, C.R. and Worden, K. (2010), 'An introduction to structural health monitoring', in: Deraemaeker, A. and Worden, K. (eds.), *New Trends in Vibration Based Structural Health Monitoring*, International Centre for Mechanical Sciences, Udine, Italy, 1–17.

Ko, J.M. and Ni, Y.Q. (2005), 'Technology developments in structural health monitoring of large-scale bridges', *Engineering Structures*, **27**(12), 1715–1725.

Ko, J.M., Ni, Y.Q., Zhou, H.F., Wang, J.Y. and Zhou, X.T. (2009), 'Investigation concerning structural health monitoring of an instrumented cable-stayed bridge', *Structure and Infrastructure Engineering*, **5**(6), 497–513.

Ni, Y.Q. (2012), 'Special issue on SHM benchmark for high-rise structures: preface', *Smart Structures and Systems*, **10**(4–5), i-ii.

Ni, Y.Q. (2013), 'Structural health monitoring of large-scale structures: from diagnostics to prognostics', Keynote Speech, *Proceedings of the 2013 SPIE Smart Structures/NDE Conference on Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 10–14 March 2013, San Diego, California, USA (CD-ROM).

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., Wai, T.T. and Wong, K.Y. (2010a), 'Long-distance remote measurement of static and dynamic displacement for large-scale structures using a vision inspection system', in: *Proceedings of the 3rd Asia-Pacific Workshop on Structural Health Monitoring*, Tokyo, Japan (CD-ROM).

Ni, Y.Q., Ye, X.W. and Ko, J.M. (2010b), 'Monitoring-based fatigue reliability assessment of steel bridges: analytical model and application', *Journal of Structural Engineering*, *ASCE*, **136**(12), 1563–1573.

Ni, Y.Q., Li, B., Lam, K.H., Zhu, D.P., Wang, Y., Lynch, J.P. and Law, K.H. (2011a), 'In-construction vibration monitoring of a supertall structure using a long-range wireless sensing system', *Smart Structures and Systems*, **7**(2), 83–102.

Ni, Y.Q., Xia, Y., Lin, W. and Ko, J.M. (2011b), 'SHM benchmark for high-rise structures: description of host structure and measurement data', in: Chang, F.K. (ed.), *Proceedings of the 8th International Workshop on Structural Health Monitoring*, DEStech Publications, Lancaster, Pennsylvania, USA, 2406–2413.

Ni, Y.Q., Li, W.R. and Ko, J.M. (2012a), 'Monitoring of dynamic responses of a super-tall structure during earthquakes', in: *Proceedings of the International Conference on Earthquake Research Challenges in the 21st Century*, 18–21 May 2012, Harbin, China (CD-ROM).

Ni, Y.Q., Xia, H.W., Wong, K.Y. and Ko, J.M. (2012b), 'In-service condition assessment of bridge deck using long-term monitoring data of strain response', *Journal of Bridge Engineering*, *ASCE*, **17**(6), 876–885.

Ni, Y.Q., Xia, Y., Lin, W., Chen, W.H. and Ko, J.M. (2012c), 'SHM benchmark for high-rise structures: a reduced-order finite element model and field measurement data', *Smart Structures and Systems*, **10**(4–5), 411–426.

Smyth, A. and Wu, M. (2007), 'Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring', *Mechanical Systems and Signal Processing*, **21**(2), 706–723.

Wang, Y., Lynch, J.P. and Law, K.H. (2007), 'A wireless structural health monitoring system with multithreaded sensing devices: design and validation', *Structure and Infrastructure Engineering*, **3**(2), 103–120.

Wong, K.Y. (2004), 'Instrumentation and health monitoring of cable-supported bridges', *Structural Control and Health Monitoring*, **11**(2), 91–124.

Wong, K.Y. (2007), 'Design of a structural health monitoring system for long-span bridges', *Structure and Infrastructure Engineering*, **3**(2), 169–185.

Wong, K.Y. and Ni, Y.Q. (2009a), 'Modular architecture of structural health monitoring system for cable-supported bridges', in: Boller, C., Chang, F.K. and Fujino, Y. (eds.), *Encyclopedia of Structural Health Monitoring*, John Wiley & Sons, Chichester, UK, 2089–2105.

Wong, K.Y. and Ni, Y.Q. (2009b), 'Structural health monitoring of cable-supported bridges in Hong Kong', in: Karbhari, V.M. and Ansari, F. (eds.), *Structural Health Monitoring of Civil Infrastructure Systems*, Woodhead Publishing, Cambridge, UK, 371–411.

Wong, K.Y. and Ni, Y.Q. (2011), 'Structural health monitoring of a suspension bridge', in: Bakht, B., Mufti, A.A. and Wegner, L.D. (eds.), *Monitoring Technologies for Bridge Management*, Multi-Science Publishing, Essex, UK, 365–390.

Xia, H.W., Ni, Y.Q., Wong, K.Y. and Ko, J.M. (2012), 'Reliability-based condition assessment of in-service bridges using mixture distribution models', *Computers and Structures*, **106–107**, 204–213.

Ye, S.Q. and Ni, Y.Q. (2012), 'Information entropy based algorithm of sensor place-ment optimization for structural damage detection', *Smart Structures and Systems*, **10**(4–5), 443–458.

Ye, X.W., Ni, Y.Q., Wong, K.Y. and Ko, J.M. (2012), 'Statistical analysis of stress spec-tra for fatigue life assessment of steel bridges with structural health monitoring data', *Engineering Structures*, **45**, 166–176.

Zhang, T., Wang, D., Cao, J., Ni, Y.Q., Chen, L.J. and Chen, D. (2012), 'Elevator-assisted sensor data collection for structural health monitoring', *IEEE Transactions on Mobile Computing*, **11**(10), 1555–1568.

Zhou, H.F., Ni, Y.Q. and Ko, J.M. (2011a), 'Eliminating temperature effect in vibra-tion-based structural damage detection', *Journal of Engineering Mechanics, ASCE*, **137**(12), 785–796.

Zhou, H.F., Ni, Y.Q. and Ko, J.M. (2011b), 'Structural damage alarming using auto-associative neural network technique: exploration of environment-tolerant capacity and setup of alarming threshold', *Mechanical Systems and Signal Processing*, **25**(5), 1508–1526.

10

Sensing solutions for assessing and monitoring dams

C-H. LOH, National Taiwan University, Taiwan

Abstract: This study presents the monitoring and assessment technology of a large arch dam, the Fei-Tsui arch dam in Taiwan, from both seismic response data and ambient vibration measurement collected from a wireless sensing system. Subspace identification (considering inputs) and stochastic subspace identification (output-only) are both used to extract the dynamic characteristics of the dam. Variation of the system natural frequencies of the dam with respect to reservoir water level is investigated. The system transfer function between the earthquake input and the response of the dam body is also investigated using the ARX model. By combining ambient vibration survey, seismic response data, and static monitoring, the safety assessment of the dam structure can be made.

Key words: subspace identification (SI), stochastic subspace identification (SSI), ARX model, wireless sensing system, arch dam.

10.1 Introduction

Monitoring technology plays an important role in securing the integrity of a structural system and maintaining its longevity. It consists of three aspects: (1) instrumentation with sensors; (2) methodologies for obtaining meaningful information concerning the structural health monitoring (SHM); and (3) early warning from the measured data. SHM refers to the use of *in situ* structural response from nondestructive sensing and analyzes the system characteristics for detecting structural changes that may indicate damage or degradation. However, detecting structural damage and identifying damaged elements in a large complex structure is a challenging task, since the *in situ* measured data of a large civil engineering structure such as a building, a bridge, and a dam are assumed to be imprecise (because of noise corruption) and often incomplete (due to economical reasons). Due to the complexity of the structures and the importance of such uncertainties as environmental effects and model uncertainties, it is necessary to develop some effective and efficient approaches not only to determine damage occurrence and location

275

for SHM in practice but also to set an early warning threshold before disaster can occur.

SHM techniques can generally be categorized as techniques that rely on either parametric or nonparametric system identification approaches. The parametric system identification approaches[1-2] are based on the use of a parametric structural model, which provides the potential to both detect and localize structural damage. This method initially involves constructing a parametric structural model, and then using that model to elucidate the structural behavior and examine the changes in the structural response. For example, Peeters *et al.*[3] used a black-box model to describe the variation of eigenfrequencies as a function of temperature. Damage can be detected if the eigenfrequency of the new data exceeds certain confidence intervals of the model. SHM techniques that rely on nonparametric system identification approaches, in which a priori information about the nature of the model is not needed, have a significant advantage when dealing with real-world situations in which the selection of a suitable class of parametric models to be used for identification purposes is difficult. For a complex civil infrastructure, SHM using nonparametric system identification techniques is increasingly attractive. For example, the application of artificial neural networks (ANNs) to system identification[4-7] and health monitoring[8-10] has received considerable attention over the last two decades.

In recent years, system identification techniques have been applied, both to seismic response data of actual structures and to continuous monitoring data of structures subjected to environmental loadings, for evaluating structural safety. Useful information has been obtained, particularly through post-earthquake analyses for damage detection. However, these techniques have scarcely been used on the actual seismic response of a dam-reservoir system for the following major reason: there have been very few observations undertaken, especially intensive strong-motion array instrumentation on a dam, to collect responses during earthquakes. Also, the high uncertainties in the observations, especially the effect of height on reservoir water, may influence the results. When a dam-reservoir system is subjected to an earthquake, hydrodynamic pressures are set up due to the vibration of the dam and reservoir. These dynamic water pressures and the deformation of the dam interact with each other, and therefore any satisfactory analytical method of analysis should treat the system as a coupled dynamic interaction problem. It is believed that the post-earthquake evaluation of dam response based on the instrument records is of importance to dam engineers, even though this dam-reservoir interaction occurs.

For safety evaluation or damage detection of a structure, both static and dynamic response data can be used. To collect the dynamic response data of a large civil infrastructure, such as a dam, some practical limitations may be

encountered. For example, it is difficult to excite a large civil structure in such a manner that will elicit changes in response of sufficient magnitude that the damage can be identified unless the structure is subjected to large earthquake excitation. Actually, monitoring the seismic response of a structure is an event-based technique, whereas most dynamic techniques are based on measuring ambient vibration of structures, which typically involve collecting a very low level response (around $10^{-2} \sim 10^{-6}$ g). The ambient vibration survey of a structure can also provide useful information for structural safety evaluation. Also, to overcome the limitation of dynamic-based monitoring techniques, long-term static deformation monitoring can be used.

For SHM of a dam, three different response measurements can be presented: earthquake response measurement, ambient vibration measurement, and continuous monitoring of static deformation of the dam. To extract the trend of the static deformation, nonlinear principal component analysis (NPCA) (the Auto-Associate Neural Network) and NARX-Neural Networks are presented. By using these methods, the residual deformation between the estimated and the recorded data is generated, and through statistical analysis the threshold level of the static deformation of the dam can be determined based on the normality assumption of the residual deformation.[11–14] In this chapter, different monitoring systems, such as the static deformation measurement system, the earthquake monitoring system, and the ambient vibration measurement system of the Fei-Tsui arch dam (Taiwan) will be introduced. The wireless sensing system is used to collect ambient vibration data from the dam for analysis. The multivariate data processing technique (using stochastic subspace identification (SSI)) is used to extract the dynamic characteristics of the dam. Comparison of the dynamic characteristics of the dam from both earthquake response data and the ambient datasets is made, and the safety assessment of the dam from this study is discussed.

10.2   Past monitoring effects of dams

Reporting on the actual earthquake motion of an arch dam is very limited. However, the importance of differential motion at a dam site is well recognized. As early as 1964, the seismic response of the Tonoyama Arch Dam in Japan was reported in Reference 15. Fourier analysis revealed that the amplitude of motion at the left abutment was $2 \sim 3$ times greater than that at the left abutment for frequencies over 4.0 Hz. Amplification studies of the Pacoima Dam, from aftershocks of the San Fernando earthquake, were also studied.[16] The aftershock of the 1976 Friuli earthquake in Italy was also measured at the Ambiesta Arch Dam.[17] The average ratios of horizontal maximum velocity at dam crest level to that at the base of the dam ranged from 1.88 to 3.11. The Whittier earthquake of California on 1 October 1987

triggered all 16 accelerometers installed on the Pacoima dam.[18] Preliminary reports indicated that the peak accelerations at the dam base were in the order of 1.0 gal, while those at about 80% of the dam's height at the dam–foundation interaction interface were in the order of 2.0 gal. In Switzerland, earthquake monitoring of large arch dams has been conducted during the past 10 years.[19] Only a limited amount of dam response data were collected. All these indicated a very limited instrumentation on the dam structures. Recently, continuous monitoring of large arch dams has become more and more important, for example Cabril Dam in Portugal[20] and Roode Elsberg Dam in South Africa.[21] The most common sensor for monitoring uses a high sensitivity accelerometer with 1000 Hz sampling rate and low noise measurement, by which the time-varying system natural frequencies are identified. The results from continuous earthquake response monitoring of the dam can provide useful information for its safety assessment.

From a review of current dam monitoring instruments, items of monitoring for dam safety have been selected, according to the safety and condition of each dam. The parameters to be measured, and the appropriate instruments, are as follows:

- Leakage or seepage loss: drainage hole,
- Deformation (dam body as well as embankment): plumb lines,
- Uplift pressure and pore pressure: piezometers,
- Earthquake motion: seismographs.

Different types of gages are also used, including strain-gage type, vibration wire type, etc. Variations in measurement with respect to time are investigated for appropriate safety control.

10.3   Measurement systems of Fei-Tsui arch dam

Built in April 1987, the Fei-Tsui concrete arch dam is 122.5 m high and 510 m long, and is constructed with a layout of three-centered double curvature with variable thickness. The dam was constructed such that the dam body was divided vertically into 29 strips, each about 17.5 m wide. A water seal is installed between each pair of adjacent strips. The capacity of this reservoir is about 400 million m.[3] Since this dam is located in a very active seismic zone area of Taiwan, both dynamic and static monitoring systems are deployed. Two different types of instrumentation were deployed. First, a well-instrumented strong-motion array was established on the dam in 1991, in order to monitor its seismic behavior. For monitoring the dynamic properties of the dam during earthquake, 11 tri-axial accelerometers (AC-63 tri-axial accelerometer) were deployed in the Fei-Tsui dam (as shown in Fig. 10.1). Five of these instruments are installed along the abutment (SD1 ~

(a)

(b)

BLOCK No.
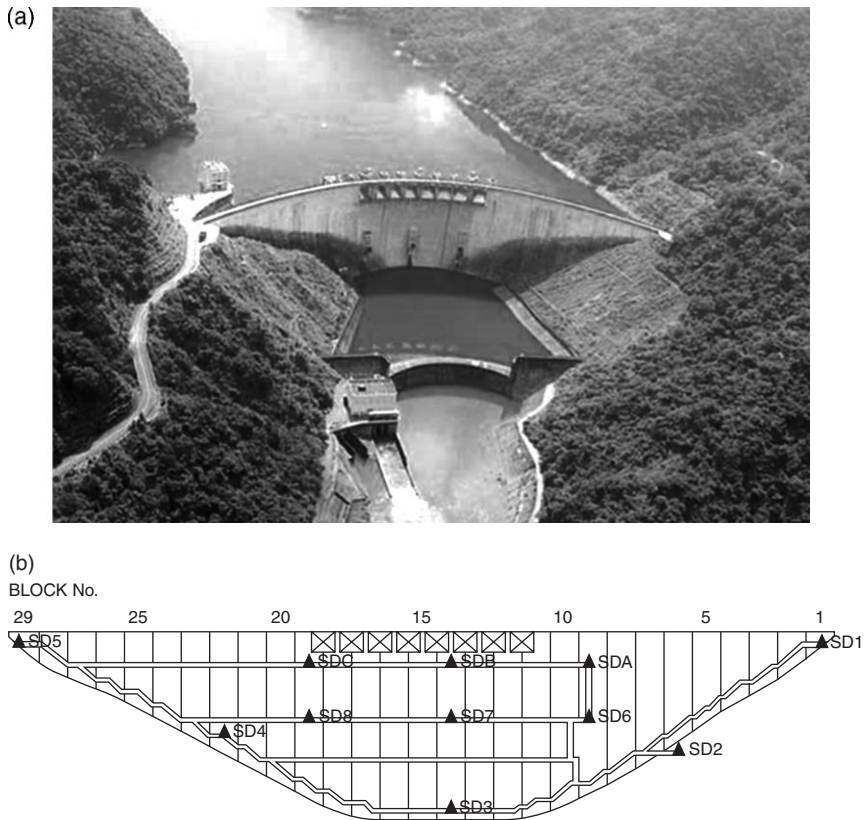
10.1 (a) Photo of Fei-Tsui arch dam (Taiwan) and (b) the distribution of accelerometer.

SD5), three of them (SD6 ~ SD8) are installed on the 115 m gallery and the rest (SDA, SDB and SDC) are deployed on the 150 m gallery.[22] All the records are in digital signal, with a sampling rate of 200 Hz. The location and orientation of each accelerometer is also shown in Fig. 10.2. About 400 m away (downstream side) there are six accelerometers along the canyon surface to detect the free field seismic ground motion. Figure 10.3 plots the earthquake events that triggered the earthquake mentoring system with respect to the reservoir level for that event.

Apart from dynamic measurement, there are several static measurement systems in this Feu-Tsui Dam. One important static measurement system is for monitoring dam deformation. Consider the base point along the dam–foundation interface as the reference (or fixed point): using a plumb line to measure the relative deformation of two different height levels of the dam can provide the deformation measurement of the dam. Figures 10.4a and 10.4b show three vertical profiles (along vertical lines of NPL1, NPL2 and
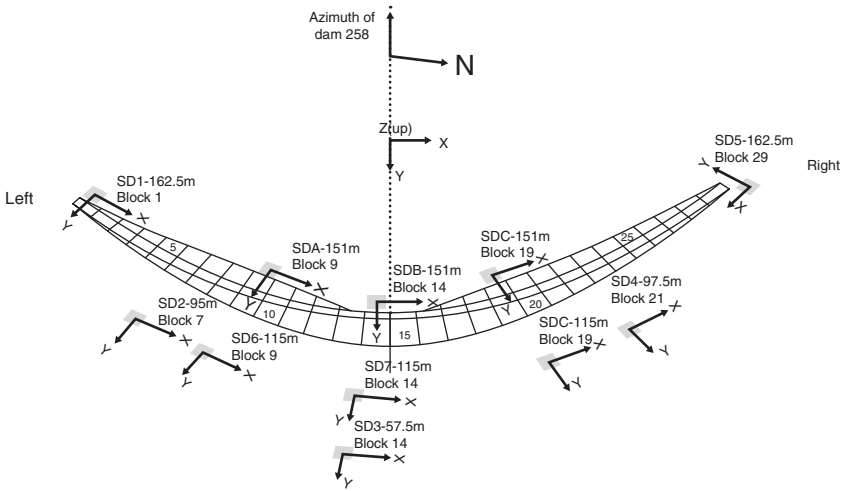
10.2 Location and orientation of each accelerometer.
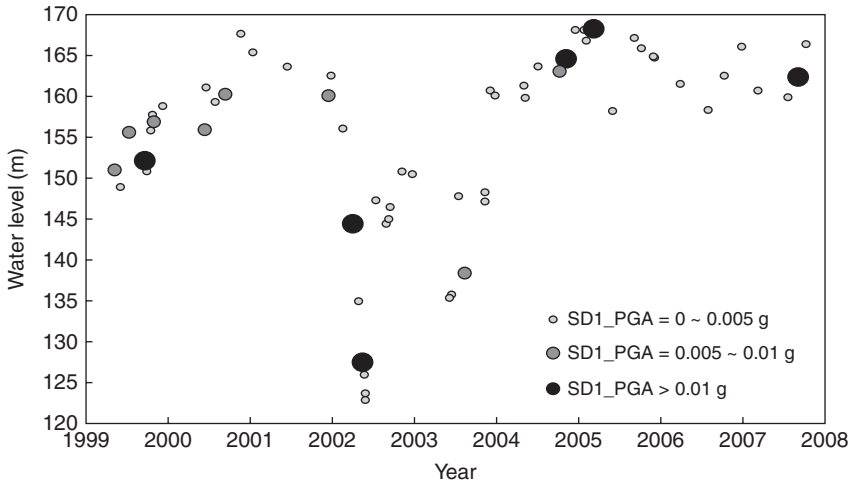
10.3 Plot of the earthquake event with respect to water level that triggered the earthquake monitoring system of the dam.

NPL3) where plumb lines were installed along the dam height. Along each profile, for example in profile NPL2Y, there are five measurement points (at levels 172.5, 150, 115, 90 and 57.5 m) to measure the radial deformation (Y-direction) of the dam. The daily dam deformation data (1 sample/day) and the temperature data were obtained.

All these static measurements were collected starting from January 1, 1987, and there was a total of 13 locations, which covered the entire dam face, to
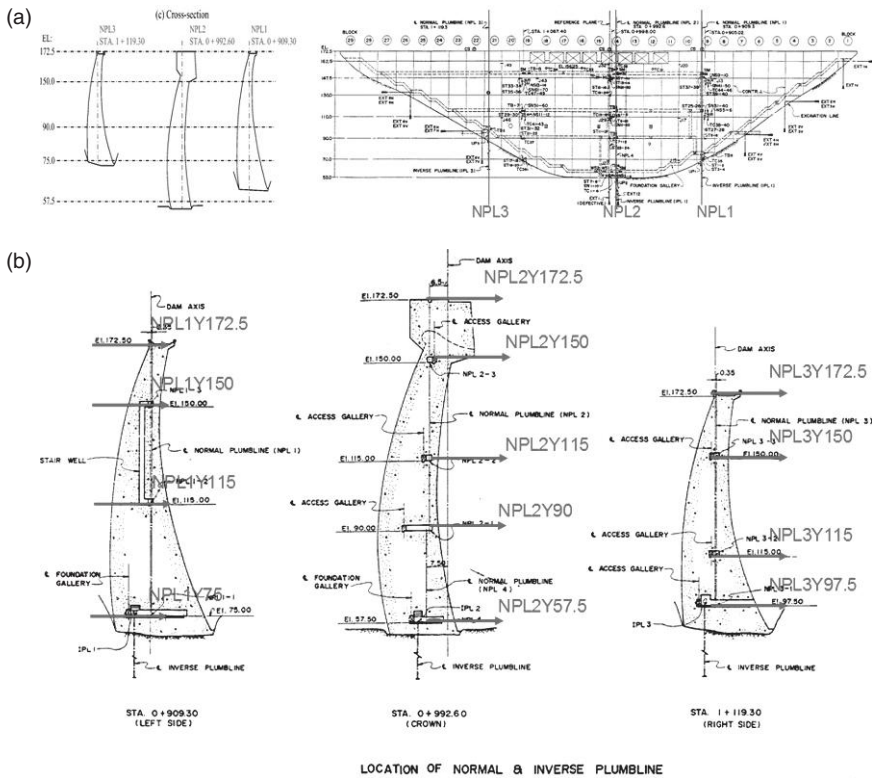
10.4 (a) Location of three vertical plumb lines to measure the static deformation. (b) Three vertical plumb lines (NPL1, NPL2, and NPL3) along the dam height and the measurement location.

measure deformation. Figure 10.5 plots the measured deformation of the dam along profiles NPL1, NPL2, and NPL3. The abscissa of Fig. 10.5 indicates the number of point not time (from 1 January 1987 to 30 April 2009, a total of 7719 data points were collected). The daily maximum and minimum temperature at the dam site were also measured, as shown in Fig. 10.6, and the recorded daily water level of the reservoir is shown in Fig. 10.7. The relationship among the static deformation, the water level, and the temperature distribution of the dam body is complex and unknown. Cybenko[23] and Funahashi[24] rigorously demonstrated that, by using neural networks, even with only one hidden layer, it can uniformly approximate any continuous function. This theoretical basis for modeling the relationship among the static deformation, the water level, and the temperature distribution of the dam body by static networks, is therefore sound. Although the static deformation has almost no change during a very short time, it changes with time over long-term continuous observation. Therefore, long-term static deformation
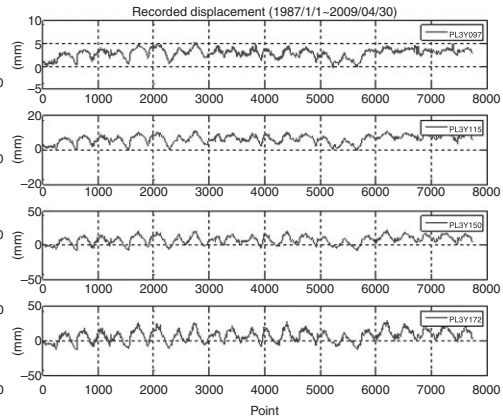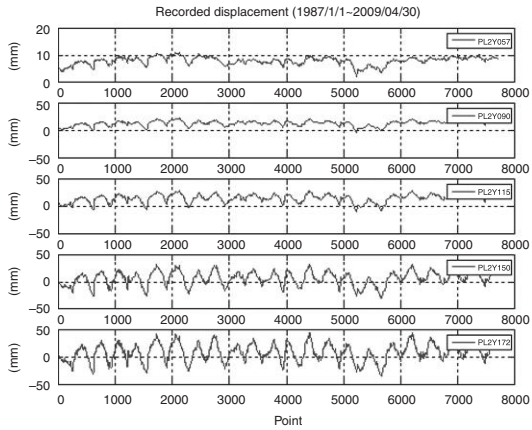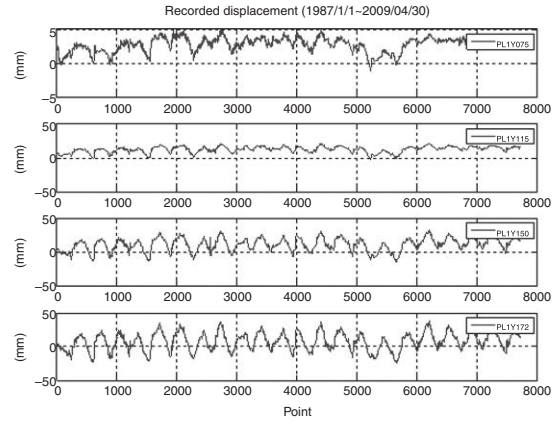
10.5 Recorded static radial deformation of Fei-Tsui arch dam along plumb line NPL1, NPL2, and NPL3 lines at different levels.

Recorded the lowest temperature (1987/01/01~2009/04/30)
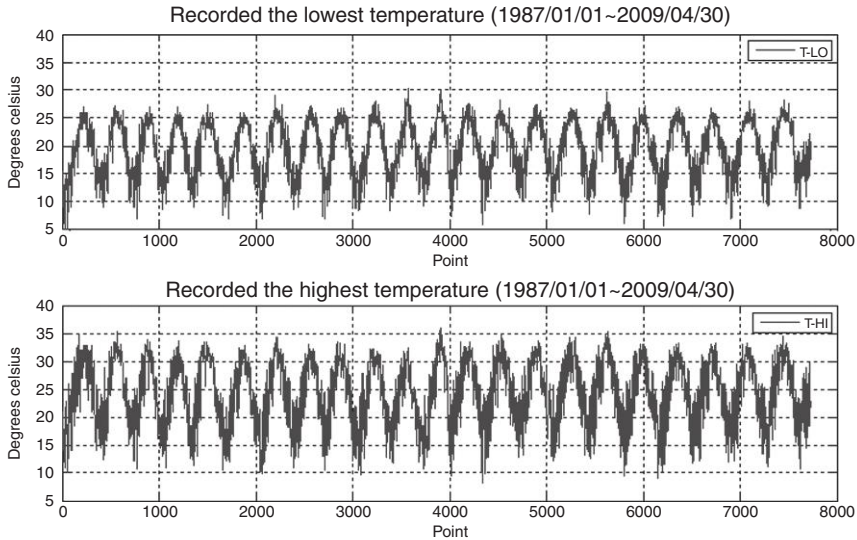
10.6 Recorded lowest and highest temperature at Fei-Tsui arch dam site. (*Source*: From 1 Jan 1987 to 30 April 2009.)
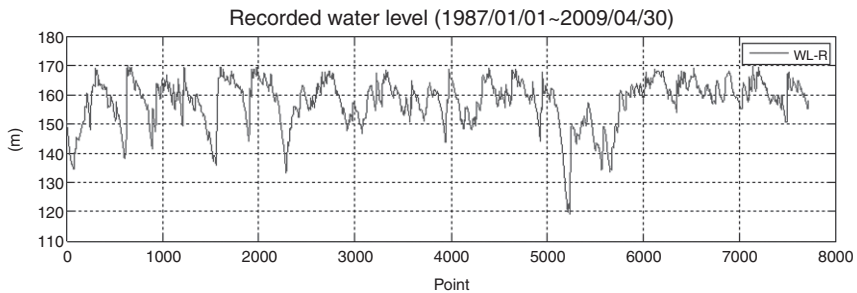
10.7 The recorded daily water level for Fei-Tsui dam.

can be approximated dynamically using dynamic networks. Moreover, static deformation data are plentiful but information poor. Linear and nonlinear principal component analyses are particularly well suited to deal with this kind of problem.

In the study by Loh *et al.*[25] two different approaches were applied to extract features of the long-term SHM data of the static deformation of the Fei-Tsui arch dam (Taiwan). These methods include singular spectrum analysis with AR model (SSA-AR) and the NPCA using auto-associative neural network (AANN) method. The SSA is a novel nonparametric technique based on the principles of multi-variance statistics. An AR model is optimized for each of the principal components obtained from SSA, and

Table 10.1 Comparison of the specification of accelerometer, velocity sensor, and other static monitoring systems

GeoSIGAC-63 tri-axial accelerometer		VSE-15D velocity sensor	
Full-scale	±2 g (0.5, 1.0, 3.0, 4.0 g optional)	Full-scale	±0.1 ms
Type	Force balanced accelerometer	Type	Force balanced accelerometer
Sensitivity	10 V/g	Sensitivity	10 or l0 μ Kine
Dynamic range	>120 dB	Dynamic range	Approximately 140 dB
Bandwidth	DC > 100 Hz (50 or 200 Hz)	Bandwidth	0.2~100 Hz
Damping	70% of critical	Damping	100% of critical
		Max. output voltage	+/−10 V

the multi-step predicted values are recombined to make the time series. The NPCA-AANN method is also used to extract the underlying static deformation features of the dam. By using these two different methods, the residual deformation between the estimated and the recorded data was generated, through statistical analysis, and the threshold level of the dam static deformation can be determined.

Besides strong-motion instrumentation and the long-term static deformation measurement, ambient vibration measurement can also be performed along the dam crest. Velocity sensors (VSE-15D) are used for ambient vibration measurement. The VSE-15D sensor is a servo velocity meter produced by Tokyo Sokushin Co., Ltd. Comparison of the sensitivity of VSE-15D (from ambient vibration measurement) and tri-axial seismograph (AC-63) is shown in Table 10.1. The static monitoring sensors of the dam are also listed in the table. The VSE-15D sensor for ambient vibration survey is very sensitive to detect the low level vibration motion, and the linear range (0.2 ~ 70 Hz) is suitable for SHM applications. The power requirement of the VSE-15D is 15 mA @ +/−15 V at about 450 mW. A total of 5 VSE-15D sensors had been installed on the crest of the dam for continuous monitoring of the dam vibration. Discussion of the ambient vibration survey using the wireless sensing system will be presented in the following section.

10.4   Wireless sensing system for ambient vibration measurement

To reduce the cabling issue as well as the cost-related problem, integration of the wireless technology and the embedded computing platform in the structural monitoring system is discussed. There is no permanent

monitoring system for dam ambient vibration. The ambient vibration survey can be done at any time. The concept of using a smart sensing system with SHM technologies has been elaborated in previous literature.[26] This system integrated the WiMMS wireless sensing unit[27] and SHM technologies to perform the function of automatic continuous monitoring. To meet the requirement of continuous monitoring on a structure, the following requirements need to be considered. First, the sensing system is required to measure the ultra-low level analog signal (in voltage) of structural responses under ambient excitation level. Therefore, the hardware design of analog signal sampling needs to be fully considered and the noise level in the sensing system must be reduced. Second, to meet the requirement of long-term data collection, the powering of the sensing platform must be stable and manageable. Additionally, wireless data communication needs to meet the complex environmental conditions. A smart sensing system is developed for this ambient vibration survey.

10.4.1   Wireless sensing system

This smart sensing system includes three major components, namely: sensors, wireless sensing units, and host node (data logger). Ambient velocity meters (i.e. VSE-15D) sense the structural vibration responses and convert the vibration signals into analog signals (in voltage). These voltage signals are fed into the wireless sensing unit (NTU-WSU) and sampled as discrete digitalized data series. The host node can acquire these sampled data through wireless communication from all distributed wireless sensing units. After obtaining the required data, the host node will execute the SHM process and generate a report. The user can access the host node through File Transfer Protocol (FTP) or Internet Explorer (IE) to obtain the structural responses and on-line analysis results. The host node is used to coordinate the wireless sensing nodes and communicate with the user end. NI cRIO-9022 is selected to play this role. NI cRIO-9022 is a high performance real-time controller. This platform includes a Free-scale MPC8347 real-time processor for deterministic and reliable real-time applications and supports interfaces with RS232, Ethernet and USB. The RS232 interface is adapted with a wireless receiver (wireless module) to communicate with wireless sensing units.

10.4.2   Hardware design

In the design of WiMMS there are three components,[28] namely: micro-controller unit (MCU), RAM, and analog/digital converter (ADC). The MCU is an 8 bit ATmega128 microcontroller @ 8 MHz, the external RAM is 128 kB

and the ADC is a 16 bits analog-to-digital converter @ 0–5 V signal range (TI ADS8341EB). Since the ambient vibration signal is an ultra-low level vibration signal, which might be less than 1 mV, to obtain a good sampling a high performance signal conditioner is considered. According to the specifications of VSE-15D, the output signal range is +/−10 V and the resolution is 10 μ Kine (10^{-7} m/s). This resolution can be realized to a minimum output voltage by multiplying its sensitivity (1000 V/m/s), and it is about 0.1 mV. Compared to the capacity of ADC (ADS8341EB), the resolution is 16 bits and its input range is 0–5 V. The signal conditioner is designed based on these requirements.

The function of the signal conditioner is defined as follows: first, amplifying the sensor signal into the desired range with adjustable gain amplifier; second, scaling and shifting the amplified signal into the range of ADC (0–5 V). PGA204 is used to amplify the sensor signal, and INA159 is used for scaling and shifting the signal into the range of ADC. PGA204 is a programmable gain instrumentation amplifier, offering excellent accuracy. The digitally selected gains of PGA204 are 1, 10, 100, and 1000 V/V. The NTU-WSU unit supports two different frequency radio modules; one is 24XStream on 2.4 GHz, and the other is 9XTend on 900 MHz, both of which are produced by Digi International Inc. The users can select a suitable radio module for their applications. The specifications of 24XStream and 9XTend are listed in Table 10.2. The 9XTend module is good for long distance communication but is not allowable worldwide. The 24XStream module is good for short distance applicable for high density sensor layout and is allowable worldwide. The power design of NTU-WSU unit includes three purposes: first is the powering requirement of the sensor (VSE-15D); second is the powering of the embedded system, which includes MCU, RAM, ADC, etc.; third is the powering of the wireless module. The power source of the sensing unit is a rechargeable Li-Battery (3200 mA @ 7.4 V). There are two regulators and one switching power module in the power design of the unit.

To prevent radio noise influencing the embedded system, and to improve the radio communication range, individual regulators for embedded system and radio module are necessary. Two 5V regulators (LP2986 National

Table 10.2 The specification of 24XStream and 9XTend

	24XStream	9XTend
RF frequency	2.4000–2.4835 GHz	902–928 MHz
Transmit power output	50 mW	Variable 1 mW – 1000 mW
Throughput data rate	19 200 bps	115 200 bps
RF data rate	20 000 bps	125 000 bps
Max. indoor range		
Max. outdoor range	(High-gain antenna)	(High-gain antenna)
Receiver sensitivity	−102 dBm	−l00 dBm

Semiconductor Co.) are used to handle the powering of the embedded system and radio module. Figure 10.8 shows the detailed power arrangement of the NTU-WSU unit. The VSE-15D sensor and the PGA204 are powered by a switching power module (EC3SA-05D15), and there is a logic switch to manage the module to save power. A 5 V Regulator_B is used to supply power to the embedded system, which includes MCU, RAM, LATCH, ADC, and INA159. The wireless radio module is powered by an individual 5 V Regulator_A to improve the performance of radio and to prevent radio noise affecting the embedded system. The power of the NTU-WSU unit also supports optional sensors such as linear variable differential transformer (LVDT), strain type accelerometer (Setra 141A), and microelectromechanical systems (MEMS) Accelerometer (Crossbow).

The detail design of hardware is presented in Fig. 10.9. The signals of the VSE-15D sensors are first amplified by the programmable gain amplifier (PGA204), and the gain value of the PGA204 is controlled by the GPIO of ATmega128 (PD.6–7). The amplified signal is then scaled and shifted into the range ADC (0–5 V) by a fixed gain difference amplifier (INA159). The ADC (ADS8341EB) digitalizes these voltage signals into discrete digital data series and transmits them to the MCU (ATmega128) through the serial peripheral interface (SPI) (@ 2 MHz) interface. The MCU will buffer these data series on the external RAM through GPIO with external RAM interface function of ATmega128 and a D-type LATCH. After the sampling process is finished, the MCU will turn off the power to the sensor through GPIO (PB.7) and start to communicate with the host node by radio module. The interface between the MCU and radio module is UART; the data rate of UART depends on the radio module, 19.2 k bps for 24XStream and
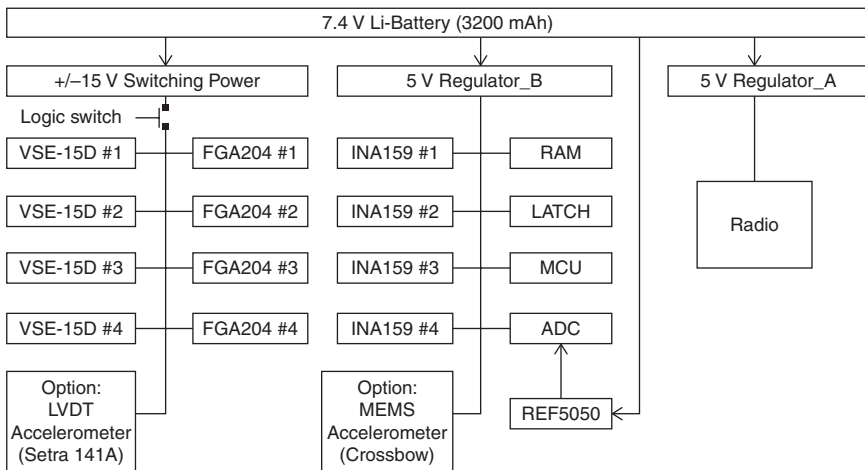
10.8 Power design in NTU-WSU.[29]
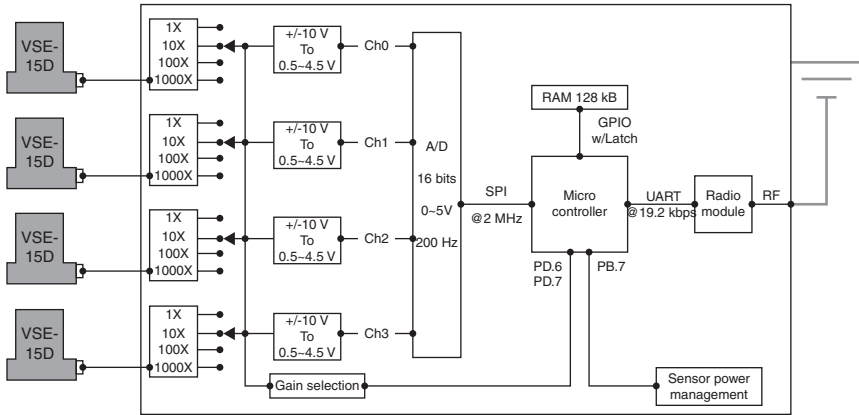
10.9 Hardware design of wireless sensing unit.[29]

Table 10.3 Hardware resources of smart sensing system

	Sensing node	Host node	User end
Name	NTU-WSU-V	cRIO 9022	
System type	Embedded system	Embedded system w/RTOS	PC or laptop etc.
Core	SbitMCU 16 MHz ATmegal28	32 bit processor 533 MHz powerPC	Optional
RAM	128 kB	256MB DDR2	Optional
Flash memory	128 KB	2 GB	Optional
Extend memory	N	USB disk	Optional
Communication1	9XTend or 24XStream w\Router fun.	9XTend or 24XStream	N
Communication2	N	Ethernet (TCP/IP)	Ethernet
ADC sampling	16 bits(+/–10 V) at 200 Hz	N	N

115k bps for 9XTend module. Table 10.3 lists the hardware resources of the sensing node, host nodes, and user nodes; engineers can arrange computations into different spaces with this table.

10.4.3   Software design

The software design of the smart sensing system includes the embedded program of NTU-WSU unit, the embedded LabVIEW RT program of NI cRIO-9022, and the user end computer program. The Sensing Service Routine is the application software of the NTU-WSU unit embedded program. It includes two components, the Data Collecting Service (Client),

and the Computing Service. The Data Collecting Service (Client) works with the Data Collecting Service (Host) of the Host Service Routine (run at Host node), which is used to convert the sensor signal into digital data series, buffering these data in RAM and feeding back these data series via wireless communication. The Host Service Routine is also presented in Fig. 10.7, which is the embedded RT LabVIEW program of the NI cRIO-9022. The NI cRIO-9022 drivers are developed by NI, and users and engineers can use these drivers to implement their application software on cRIO-9022. A series RT Modules includes the functions and drivers to support user programming. The user end is a general purpose computer, such as a PC or a laptop.

Combining the software and hardware design of wireless sensing system, a wireless sensing network for field experiment has been developed for the Fei-Tsui dam. The instrument set-up protocol is shown in Fig. 10.10. The sensing node collects the vibration data, and from the wireless sensing unit the digital data are broadcast to the host node. The host node performs simple analysis of the data (such as Fourier transform). Then, via the wireless router, data are collected by the user node for more detailed analysis. If the field experiment can provide enough power, then the host node can conduct more sophisticated analysis on data fusion.

10.5   Analysis of ambient vibration data

To extract modal information from the output-only data set, output-only system identification techniques can be applied. In this study, the Stochastic
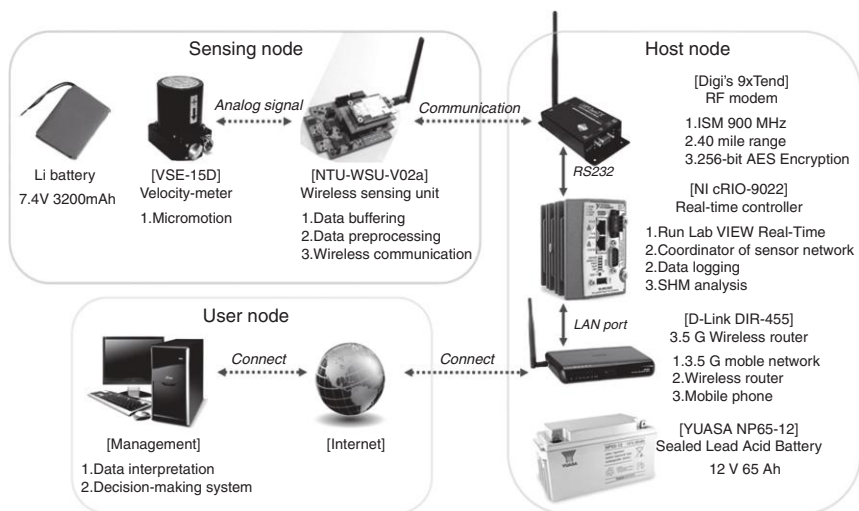
10.10 Instrument set-up protocol for ambient vibration survey in the field.

Subspace Identification (SSI) method, as originally presented by van Overschee and de Moor[30,31] is adopted to identify a stochastic state space model of the dam structural system. The method of analysis of ambient vibration data will be discussed in this section.

Consider a discrete-time stochastic state space model:

$$\mathbf{X}_{k+1}^s = \mathbf{A}_d \mathbf{X}_k^s + \mathbf{w}_k^s \qquad [10.1]$$

$$\mathbf{y}_k^s = \mathbf{C}_c \mathbf{X}_k^s + \mathbf{v}_k^s \qquad [10.2]$$

where $\mathbf{w}_k^s \in \mathbb{R}^{2n \times 1}, \mathbf{v}_k^s \in \mathbb{R}^{i \times 1}$ are white noise. The superscript 's' means 'stochastic' and it implies that the system is excited by the stochastic component. This model can be applied to identify the system by using output-only measurement (e.g. ambient vibration of structures). The purpose of the stochastic identification problem is to determine the system parameters $\mathbf{A}_d$ and $\mathbf{C}_c$ from the measurements of the output $\mathbf{y}_k^s$. It should be noted that the system parameters $\mathbf{B}_d$ and $\mathbf{D}_c$ do not appear in Equations [10.1] and [10.2] because in the stochastic system the input is not actually defined. Since the input of the system is unknown, the system parameters $\mathbf{B}_d$ and $\mathbf{D}_c$ cannot be identified when using the output-only identification algorithms.

In SSI, the Hankel matrix plays an important role. The Hankel matrix comprises a special data arrangement that can be used to extract the data connections between different sensing locations and different sampling times. The output Hankel matrix can be constructed from the output data:

$$\begin{bmatrix} Y_p^s \\ Y_f^s \end{bmatrix} \equiv \begin{bmatrix} Y_1^s & Y_2^s & \cdots & Y_j^s \\ Y_2^s & Y_s^s & \cdots & Y_{j+1}^s \\ \vdots & \vdots & \cdots & \vdots \\ Y_i^s & Y_{i+j-1}^s & \cdots & Y_{i+j-1}^s \\ Y_{i+1}^s & Y_{i+2}^s & \cdots & Y_{i+j}^s \\ Y_{i+2}^s & Y_{i+s}^s & \cdots & Y_{i+j+1}^s \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2i}^s & Y_{2i+j+1}^s & \cdots & Y_{2i+j-1}^s \end{bmatrix} \in \mathbb{R}^{2li \times j} \qquad [10.3]$$

where i is the number of block rows, which is a user-defined index and must be larger than the order $2n$ of the system. Since there are only l DOFs measured, the output Hankel matrix must contain $2li$ rows. j is the number of block columns of the output Hankel matrix. If the sampling length is equal to r then the number j should be equal to $j = r - 2i + 1$ so that all data are used for analysis. According to the expression of Equation [10.3], the output Hankel matrix is divided into the past part, $\mathbf{Y}_p^s \in \mathbb{R}^{li \times j}$, and the future part,

$Y_f^s \in \mathbb{R}^{li \times j}$. Similarly, the process noise w_k^s and measurement noise v_k^s can also be arranged in the Hankel matrices: $W_p^s W_f^s \in \mathbb{R}^{2ni \times j}$ and $V_p^s, V_f^s \in \mathbb{R}^{li \times}$ Moreover, the stochastic state X_k^s can also be divided into past and future parts:

$$X_k^s \equiv \begin{bmatrix} X_1^s & X_2^s & \cdots & X_j^s \end{bmatrix} \in \mathbb{R}^{2n \times j}, X_f^s \equiv \begin{bmatrix} X_{i+1}^s & X_{i+2}^s & \cdots & X_{i+j}^s \end{bmatrix} \in \mathbb{R}^{2n \times j} \quad [10.4]$$

After arranging the output y_k^s and state X_k^s in the block-Hankel matrix, Equations [10.1] and [10.2] can be rewritten as the equations of Hankel matrices:

$$Y_p^s = \Gamma_i X_p^s + H_i^s W_p^s + V_p^s \qquad\qquad [10.5]$$

$$Y_f^s = \Gamma_i X_f^s + H_i^s W_f^s + V_f^s \qquad\qquad [10.6]$$

$$X_f^s = A_d^i X_p^s + \Delta_i^s W_p^s \qquad\qquad [10.7]$$

with

$$\Gamma_i \equiv \begin{bmatrix} C_c \\ C_c A_d \\ C_c A_d^2 \\ \vdots \\ C_c C_d^{i-1} \end{bmatrix} \in \mathbb{R}^{li \times 2n}, H_i^s \equiv \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ C_c & 0 & 0 & \cdots & 0 \\ C_c A_d & C_c & 0 & \cdots & 0 \\ M & \vdots & \vdots & \ddots & \vdots \\ C_c A_d^{i-2} & C_c A_d^{i-3} & C_c A_d^{i-4} & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{li \times 2ni}$$

$$\Delta_i^s \equiv \begin{bmatrix} A_d^{i-1} & A_d^{i-2} & \cdots & A_d & I_{2n} \end{bmatrix} \in \mathbb{R}^{2n \times 2ni}$$

where $\Gamma_i \in \mathbb{R}^{li \times 2n}$ is the extended observability matrix, $\Gamma_i \in \mathbb{R}^{li \times 2ni}$ is the stochastic lower block triangular Toeplitz matrix, and $\Delta_i^s \in \mathbb{R}^{2n \times 2ni}$ is the stochastic reversed extended controllability matrix.

The subspace identification (SI) algorithm is developed based on geometric concepts: a row vector of a block-Hankel matrix is considered as a vector in the j dimensional ambient space (where j is the number of block columns). For example, the rows of each matrix $\tilde{A} \in \mathbb{R}^{p \times j}, \tilde{B} \in \mathbb{R}^{q \times j}$ and $\tilde{C} \in \mathbb{R}^{r \times j}$ are defined as a basis for a linear vector space in this ambient space. Two geometric operations are defined that can be performed with these row spaces:

$$\frac{\tilde{A}}{\tilde{B}} \equiv \tilde{A}\tilde{B}^T \left(\tilde{B}\tilde{B}^T \right)^{\dagger} \tilde{B} \qquad\qquad [10.8]$$

$$\frac{\tilde{\mathbf{A}}}{\tilde{\mathbf{B}}}\tilde{\mathbf{C}} \equiv \left(\frac{\tilde{\mathbf{A}}}{\tilde{\mathbf{B}}^{\wedge}}\right)\left(\frac{\tilde{\mathbf{C}}}{\tilde{\mathbf{B}}^{\wedge}}\right)^{\dagger}\tilde{\mathbf{C}}$$

[10.9]

where $/$ denotes the projection operator, $(\ )^{\mathbf{T}}$ denotes the transpose operator and $(\ )^{\dagger}$ denotes the pseudo-inverse operator. Equation [10.8] defines the orthogonal projection of the row space of the matrix $\tilde{\mathbf{A}}$ on the row space of the matrix $\tilde{\mathbf{B}}$, and Equation [10.9] defines the oblique projection of the row space of $\tilde{\mathbf{A}}$ along the row space of $\tilde{\mathbf{B}}$ on the row space of $\tilde{\mathbf{C}}$.

It is known that the noise effect can be reduced through the computation of the covariance between the future part and past part of the output Hankel matrix. In SSI, a similar result can be achieved by using the projection. The orthogonal projection of the row space of the matrix $\mathbf{Y}_{\mathbf{f}}^{\mathbf{s}}$ on the row space of the matrix $\mathbf{Y}_{\mathbf{p}}^{\mathbf{s}}$ can be calculated by the following formula:

$$\frac{\mathbf{Y}_{\mathbf{f}}^{\mathbf{s}}}{\mathbf{Y}_{\mathbf{p}}^{\mathbf{s}}} \equiv \mathbf{Y}_{\mathbf{f}}^{\mathbf{s}}\mathbf{Y}_{\mathbf{p}}^{\mathbf{sT}}\left(\mathbf{Y}_{\mathbf{p}}^{\mathbf{s}}\mathbf{Y}_{\mathbf{p}}^{\mathbf{sT}}\right)^{\dagger}\mathbf{Y}_{\mathbf{p}}^{\mathbf{s}} = \mathbf{O}_{\mathbf{i}}^{\mathbf{s}},$$

[10.10]

$\mathbf{O}_{\mathbf{i}}^{\mathbf{s}} \in \mathbb{R}^{li \times j}$ is the orthogonal projection matrix. The main theorem of SSI implies that the extended observability matrix $\Gamma_{\mathbf{i}}$ can be found from the result of orthogonal projection:

$$\mathbf{O}_{\mathbf{i}}^{\mathbf{s}} = \Gamma_{\mathbf{i}}\widehat{\mathbf{X}}_{\mathbf{i}}$$

[10.11]

where $\widehat{\mathbf{X}}_{\mathbf{i}} \in \mathbb{R}^{2n \times j}$ is the estimated state sequence in the stochastic system, which is equal to the estimation from the forward non-steady state Kalman filter.[31] Instead of Equation [10.10], the orthogonal projection can be easily expressed in terms of the following LQ decomposition:

$$\begin{matrix} & \mathbf{li} & \mathbf{li} & \mathbf{j} \\ \begin{pmatrix} \mathbf{Y}_{\mathbf{p}}^{\mathbf{s}} \\ \mathbf{Y}_{\mathbf{f}}^{\mathbf{s}} \end{pmatrix} = \begin{matrix} \mathbf{li} \\ \mathbf{li} \end{matrix}\begin{bmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{L} & \mathbf{L}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{Q}_{11}^{\mathbf{T}} \\ \mathbf{Q}_{21}^{\mathbf{T}} \end{bmatrix} \Rightarrow \frac{\mathbf{Y}_{\mathbf{f}}^{\mathbf{s}}}{\mathbf{Y}_{\mathbf{p}}^{\mathbf{s}}} = \mathbf{L}_{21}\mathbf{Q}_{11}^{\mathbf{T}} \end{matrix}$$

[10.12]

$$\mathbf{L}_{21}\mathbf{Q}_{11}^{\mathbf{T}} = \Gamma_{\mathbf{i}}\widehat{\mathbf{X}}_{\mathbf{i}}^{\mathbf{s}}$$

[10.13]

where $\mathbf{L}_{\mathbf{ij}}$ are partitions of the lower triangular matrix from LQ decomposition of the output Hankel matrix, and $\mathbf{Q}_{\mathbf{ij}}$ are the partitions of the orthogonal matrix. The proof of Equation [10.12] can be found in Appendix A.2. Equation [10.13] implies the column space of the extended observability

matrix Γ_i can be obtained from the column space of L_{21}. It is the essence of why these algorithms are called 'subspace' identification algorithms. They retrieve the system matrix as the subspace of projection matrix:

$$\text{column space}(L_{21}) = \text{column space}(\Gamma_i) \qquad [10.14]$$

Equation [10.24] implies that only L_{21} is needed when computing the observability matrix Γ_i and Q_{ij} can be abandoned after the LQ decomposition. The LQ decomposition significantly reduces the computational complexity and memory requirements when computing the projection operator. Once L_{21} are obtained, the system parameters can be determined by the following procedures:

1. Calculate the singular value decomposition (SVD) of L_{21} and determine the 'temporary' system order N by observing the quantity of the singular values in S:

$$L_{21} = USV^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \approx U_1 S_1 V_1^T \qquad [10.15]$$

The extended observability matrix Γ_i can be determined as the first N columns of U:

$$\Gamma_i = U_1 \, (\text{if } N = 2n) \qquad [10.16]$$

In linear algebra, the singular value decomposition is an important factorization of a rectangular real or complex matrix, with several applications in signal processing and statistics. The matrix $U \in \mathbb{R}^{li \times li}$ contains a set of orthonormal 'output' basis vector directions for L_{21}, the matrix $V \in \mathbb{R}^{li \times li}$ contains a set of orthonormal 'input' basis vector directions, and the matrix $S \in \mathbb{R}^{li \times li}$ contains only the singular values in its diagonal and it can be separated into two parts, $S_1 \in \mathbb{R}^{N \times N}$ and $S_2 \in \mathbb{R}^{(li-N) \times (li-N)}$. The small singular values S_2 can be neglected and the 'temporary' system order N of the system is determined. Thus a reduced version of the singular value decomposition is described by the matrices $U_1 \in \mathbb{R}li \times N, S_1 \in \mathbb{R}^{N \times N}$ and $V_1^T \in \mathbb{R}^{N \times li}$. There are two purposes for using the truncated singular value decomposition: (1) to reduce the system order, which can save computation time without affecting the accuracy of analysis results; and (2) to reduce the noise effect, because the contribution of small singular values S_2 is considered as noise effect. Order determination is an important step in SI algorithms. For now the system order N should be determined without affecting the accuracy of L_{21} or the analysis

results will be inaccurate and some structural modes will be lost. That is why here we describe N as the 'temporary' system order.

2. Calculate the continuous-time state matrix $\mathbf{A_c}$ and the output matrix $\mathbf{C_c}$ from $\mathbf{\Gamma_i}$:

$$\mathbf{A_d} = \underline{\mathbf{\Gamma}} \; \overline{\mathbf{\Gamma}}_i = \begin{bmatrix} \mathbf{C_c} \\ \mathbf{C_c \dot{A}_d} \\ \vdots \\ \mathbf{C_c A_d^{i\text{-}1}} \end{bmatrix} \begin{bmatrix} \mathbf{C_c A_d} \\ \mathbf{C_c A_d^2} \\ \vdots \\ \mathbf{C_c A_d^{i\text{-}1}} \end{bmatrix} \qquad [10.17]$$

$$\mathbf{A_c} = \log \mathbf{m} (\mathbf{A_d}) / \Delta \mathbf{t} \qquad [10.18]$$

where $\mathbf{\Gamma}_i \in \mathbb{R}l(i-1) \times n$ denotes $\mathbf{\Gamma_i}$ without the last l rows, $\mathbf{\Gamma}_i \in \mathbb{R}^{l(i\text{-}1)\times n}$ denotes $\mathbf{\Gamma_i}$ without the first l rows, and log m( ) is the logarithm operator of matrix. The output matrix C_c can be determined from the first l rows of $\mathbf{\Gamma_i}$.

3. Calculate the matrix of eigenvalue $\mathbf{\Lambda}$ and eigenvector $\mathbf{\Psi}$ from $\mathbf{A_c}$:

$$\text{eig}(\mathbf{A_c}) = \begin{bmatrix} \mathbf{\Psi} & \mathbf{\Lambda} \end{bmatrix} \qquad [10.19]$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \; \lambda_1^*, \; \lambda_2, \; \lambda_2^*, \; \cdots \; , \lambda_n^*) \in \mathbb{C}^{2n \times 2n}$ is the matrix of eigenvalues, $\mathbf{\Psi} \in \mathbb{C}^{2n \times 2n}$ is the matrix of eigenvectors, $\lambda_k \in \mathbb{C}$ is the eigenvalue of k-th mode, and $*$ is the complex conjugate operator. It should be noted that the eigenvalues and eigenvectors occur in complex conjugated pairs and the eigenvalue is associated with the natural frequency and damping ratio:

$$\lambda_k, \lambda_k^* = -\omega_k \xi_k \pm i\omega_k \sqrt{1 - \xi_k^2} \qquad [10.20]$$

4. Determine the natural frequency ω_k and damping ratios ξ_k from λ_k :

$$\omega_k = \sqrt{\mathbf{a}_k^2 + \mathbf{b}_k^2} \, (\text{rad}/\text{s}), \xi_k = \frac{-a_k}{\omega_k} \qquad [10.21]$$

with $\mathbf{a_k} = \text{Re}(\lambda_k) \in \mathbb{R}, \mathbf{b_k} = \text{Im}(\lambda_k) \in \mathbb{R}$

5. Determine the mode shape $\mathbf{\Phi_k}$ from $\mathbf{C_c}$ and $\mathbf{\Psi}$:

$$\Phi = \mathbf{C_c} \Psi \qquad [10.22]$$

where $\Phi = \begin{bmatrix} \phi_1 & \phi_1^* & \phi_2 & \phi_2^* & \cdots & \phi_n^* \end{bmatrix} \in \mathbb{C}^{l \times 2n}$

is the matrix of mode shapes. The elements in the vector Φ_k are always complex numbers in practice. It can be imagined that the absolute value of the complex number is interpreted as the amplitude and the argument is interpreted as the phase of a sine wave of given frequency ω_k. Steps 2–5 show how the characteristics of the continuous-time system can be determined from the parameters of the discrete-time system.

10.6 Results of the ambient vibration survey of the dam

To conduct the ambient vibration survey of the dam, VSE-15D velocity sensor was used. Because of the large dimensions of dam body and the limited number of sensors, as well as the difficulty of wireless communication in the tunnel of the dam body, the ambient vibration experiment needed therefore to be divided into several steps. The experiment was carried out in four different steps. The overall measurement location of the dam and the corresponding sensor location, are summarized in Fig. 10.11. Details of the sensor layout for each step are shown in Plate XII in the color section between pages 374 and 375 (Step 1: with four sensors on the right-hand side of the dam crest; Step 2: with four sensors on the left-hand side of the dam crest and five sensors in the first corridor of the dam body; Step 3: with five sensors in the first corridor and four in the second corridor; Step 4: with four in the second corridor and four in the third corridor). Sensor data synchronization is required among all test set-up. Vibration data were collected through a wireless communication module with a sampling rate of 200 Hz. The SSI technique was used. Figure 10.12 shows the stability diagram of the analysis of the measurement. Table 10.4 shows the identified natural frequencies and damping ratios of the first eight modes. Plate XIII in the color section between pages 000 and 000 shows the identified three-mode shape of the dam.
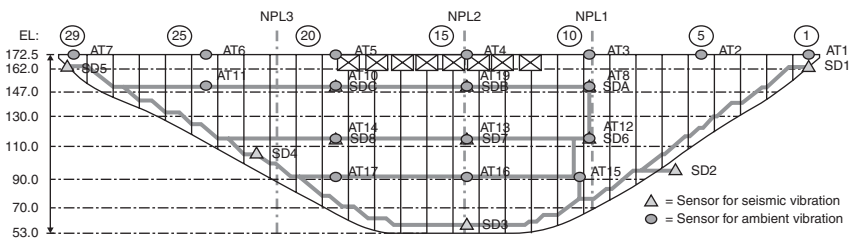
10.11 Distribution of VSE-15D sensor for ambient vibration survey of the dam.
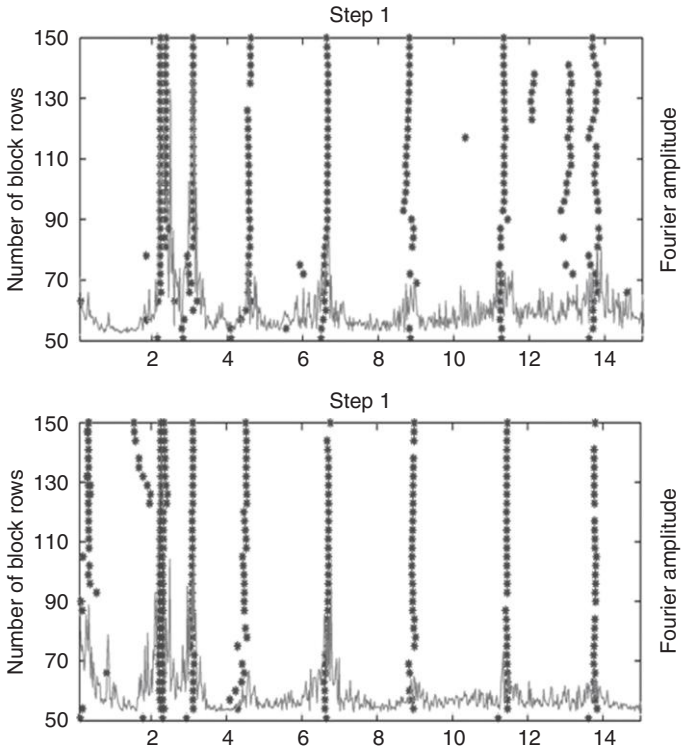
10.12 Generated stability diagram from the ambient vibration data (Step 1 and Step 4)

10.7   Analysis of earthquake response data of Fei-Tsui arch dam

For system identification using input and output measurements, the SI can be used. In recent years, various versions of subspace method have been used to identify the state space model of linear systems. The basic concept of the SI algorithm is the exploitation of the state as a finite-dimensional interface between the past and the future. By using geometric concepts some system characteristics can be revealed by geometric manipulation of the row spaces of certain matrices. First, the input and output data are arranged into two distinct block-Hankel matrices. The projection theorem is then employed to avoid the influence of noise and extract the observability matrix from block equations. Generally, mathematical tools, including pseudo-inverse and singular value decomposition (SVD), are used to reduce the model and then compute system parameters. The well-known algorithms include canonical variable analysis (CVA),[32] N4SID,[33] MOESP, and IV-4SID.[34] A unified treatment of most of these algorithms has been given by van Overschee and de Moor.[31]

Table 10.4 Identified system natural frequencies and damping ratios of the dam from each test set-up

		Natural frequency (Hz)								Damping ratio (%)							
Step 1	Ave.	2.22	2.38	3.09	4.59	6.65	8.81	11.29	13.78	0.93	2.05	1.15	1.52	1.30	1.04	1.37	1.78
	Std.	0.002	0.003	0.002	0.024	0.008	0.027	0.048	0.047	0.07	0.15	0.13	0.31	0.14	0.09	0.69	0.51
Step 2	Ave.	2.23	2.38	3.08	4.61	6.73	9.00	11.35	13.77	2.04	3.09	1048	0.92	1.51	1.24	0.85	1.42
	Std.	0.021	0.009	0.009	0.068	0.050	0.013	0.001	0.053	0.91	0.36	0.34	0.44	0.32	0.08	0.12	0.40
Step 3	Ave.	2.23	2.34	3.08	4.59	6.63	8.98	11.31	13.65	1.23	1.53	1.45	1.75	0.97	0.94	1.14	1.23
	Std.	0.006	0.014	0.003	0.026	0.019	0.012	0.009	0.018	0.74	0.33	0.17	0.12	0.21	0.12	0.15	0.09
Step 4	Ave.	2.23	2.34	3.09	4.48	6.71	8.95	11.45	13.79	1.10	2.62	1.74	2.26	1.25	1.20	1.08	1.07
	Std.	0.004	0.025	0.010	0.039	0.003	0.016	0.005	0.009	0.16	0.67	0.18	0.24	0.05	0.09	0.11	0.06

Consider a discrete-time system:

$$\mathbf{X}_{k+1} = \mathbf{A_d}\mathbf{X}_k + \mathbf{B_d}\mathbf{u}_k + \mathbf{w}_k \qquad [10.23]$$

$$\mathbf{y}_k = \mathbf{C_c}\mathbf{X}_k + \mathbf{D_c}\mathbf{u}_k + \mathbf{v}_k \qquad [10.24]$$

To identify the parameter matrices $\mathbf{A_d}, \mathbf{B_d}, \mathbf{C_c}$ and $\mathbf{D_c}$ of the systems, not only the output data $\mathbf{y}_k$ but also the input data $\mathbf{u}_k$ need to be used for system identification.

Similarly to the stochastic system, for deterministic system the input data $\mathbf{u}_k$ can also be arranged in the Hankel matrix:

$$
\left[\frac{\mathbf{U_p}}{\mathbf{U_f}}\right] \equiv
\left[\begin{array}{cccc}
\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_j \\
\mathbf{u}_2 & \mathbf{u}_3 & \cdots & \mathbf{u}_{j+1} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{u}_i & \mathbf{u}_{i+1} & \cdots & \mathbf{u}_{i+j-1} \\
\hline
\mathbf{u}_{i+1} & \mathbf{u}_{i+2} & \cdots & \mathbf{u}_{i+j} \\
\mathbf{u}_{i+2} & \mathbf{u}_{i+3} & \cdots & \mathbf{u}_{i+j+1} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{u}_{2i} & \mathbf{u}_{2i+1} & \cdots & \mathbf{u}_{2i+j-1}
\end{array}\right]
=
\left[\begin{array}{cccc}
\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_j \\
\mathbf{u}_2 & \mathbf{u}_3 & \cdots & \mathbf{u}_{j+1} \\
\hline
\mathbf{u}_{i+1} & \mathbf{u}_{i+2} & \cdots & \mathbf{u}_{i+j} \\
\mathbf{u}_{i+2} & \mathbf{u}_{i+3} & \cdots & \mathbf{u}_{i+j+1} \\
\mathbf{u}_{i+3} & \mathbf{u}_{i+4} & \cdots & \mathbf{u}_{i+j+2} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{u}_{2i} & \mathbf{u}_{2i+1} & \cdots & \mathbf{u}_{2i+j-1}
\end{array}\right]
\equiv
\left[\frac{\mathbf{u_p^+}}{\mathbf{u_f^-}}\right] \in \mathbb{R}^{2mi \times j}
$$

$$10.25]$$

where $\mathbf{U_p} \in \mathbb{R}^{mi \times j}$ is the past input Hankel matrix and $\mathbf{U_f} \in \mathbb{R}^{mi \times j}$ is the future input Hankel matrix. The matrices $\mathbf{U_p^+} \in \mathbb{R}^{m(i+1) \times j}$ and $\mathbf{U_f^-} \in \mathbb{R}^{m(i-1) \times j}$ are defined by shifting the border between $\mathbf{U_p}$ and $\mathbf{U_f}$ one block row down. Moreover, two special Hankel matrices consisting of both input and output data are defined as[35]:

$$\Xi_\mathbf{p} = \left[\frac{\mathbf{U_p}}{\mathbf{Y_p}}\right] \in \mathbb{R}^{(m+1)i \times j}, \Xi_\mathbf{p}^+ = \left[\frac{\mathbf{U_p^+}}{\mathbf{Y_p^+}}\right] \in \mathbb{R}^{(m+1+2)i \times j} \qquad [10.26]$$

The matrix form of the deterministic model can be expressed as:

$$\mathbf{Y_p^d} = \Gamma_\mathbf{i}\mathbf{X_p^d} + \mathbf{H_i^d}\mathbf{U_p} \qquad [10.27]$$

$$\mathbf{Y_f^d} = \Gamma_\mathbf{i}\mathbf{X_f^d} + \mathbf{H_i^d}\mathbf{U_f} \qquad [10.28]$$

where

$$
\mathbf{H_i^d} \equiv
\left[\begin{array}{ccccc}
\mathbf{D} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{C_c}\mathbf{B_d} & \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{C_c}\mathbf{A_d}\mathbf{B_d} & \mathbf{C_c}\mathbf{B_d} & \mathbf{D} & \cdots & \mathbf{0} \\
\mathbf{M} & & & \ddots & \\
\mathbf{C_c}\mathbf{A_d^{i-2}}\mathbf{B_d} & \mathbf{C_d}\mathbf{A_d^{i-3}}\mathbf{B_d} & \mathbf{C_c}\mathbf{A_d^{i-4}}\mathbf{B_d} & \cdots & \mathbf{D}
\end{array}\right] \in \mathbb{R}^{li \times mi} \qquad [10.29]
$$

and $\mathbf{H}_i^d$ is the low block triangular Toeplitz matrix.

When only the modal properties (natural frequency, damping ratio and mode shape) of the structure are needed, there exists another numerical implementation for SI. The 'multivariable output-error state space' algorithm (MOESP) is employed to extract the column space of the extended observability matrix $\mathbf{\Gamma}_i$ from the LQ decomposition of the Hankel matrix[34]:

$$\begin{array}{cc} & \begin{array}{cccc} mi & (m+1)i\ li & & j \end{array} \\ \begin{bmatrix} \mathbf{U}_f \\ \Xi_p \\ \mathbf{Y}_f \end{bmatrix} = \begin{array}{c} mi \\ (m+1)i \\ li \end{array} \begin{bmatrix} \mathbf{L}_{11} & 0 & 0 \\ \mathbf{L}_{21} & \mathbf{L}_{22} & 0 \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{L}_{33} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{11}^T \\ \mathbf{Q}_{21}^T \\ \mathbf{Q}_{31}^T \end{bmatrix} \end{array}$$

[10.30]

$$\left(\frac{\mathbf{Y}_f /_{\mathbf{U}_f} \Xi_p}{\mathbf{U}_f} \right) = \frac{\mathbf{\Gamma}_i \tilde{\mathbf{X}}_i}{\mathbf{U}_f} = \mathbf{L}_{32} \mathbf{Q}_{21}^T$$

[10.31]

$$\text{column space}\left(\mathbf{L}_{32}\right) = \text{column space}\left(\mathbf{\Gamma}_i\right)$$

[10.32]

Finally, only the $\mathbf{L}_{32}$ factor is needed for identification of the system parameters $\mathbf{A}_d$ and $\mathbf{C}_c$. Once the $\mathbf{L}_{32}$ are obtained from the LQ decomposition of the Hankel matrix in Equation [10.30], the system parameters can be determined from the SVD:

$$\mathbf{L}_{32} = \mathbf{USV}^T = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \quad \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^T$$

$$\mathbf{\Gamma}_i = \mathbf{U}_1 \left(\text{if } N = 2n\right)$$

[10.33]

$\mathbf{A}_d$ and $\mathbf{C}_c$ can be determined $\mathbf{\Gamma}_i$ modal properties of the system can finally be identified.

As mentioned before, order determination is an important component of SI algorithms. As the true system order is often unknown, a practice in the modal analysis is to calculate the modal parameters for increasing order N [35]. If N is higher than the true system order, the noise mode will be involved, but the mathematical poles thus obtained are different for different models if the noise is purely white. Hence, poles of the physical system can be detected by comparing the modal parameters for different model orders. To identify a suitable system order and to separate the true modes from the noise modes, for each specified model order N, the SVD needs to be used. To distinguish the true modes from the noise modes, first determine the system order N based on the SVD, see Equation [10.33] in the SI:[36]

$$\mathbf{USV^T} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}$$ [10.34]

with $\mathrm{diag}(\mathbf{S}_1) = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_N \end{bmatrix}^T \in \mathbb{R}^{N \times 1}$,

   $\mathrm{diag}(\mathbf{S}_2) = \begin{bmatrix} \mathbf{s}_{N+1} & \mathbf{s}_{N+2} & \cdots & \mathbf{s}_{li} \end{bmatrix}^T \in \mathbb{R}^{(li-N) \times 1}$,

The system order N can be determined by assigning a fractional number as a criterion of the SVD $\mathbf{C_r^{SVD}}$, such that:

$$s_N \geq \mathbf{C_r^{SVD}} \cdot s_1 > s_{N+1}$$ [10.35]

The SVD decomposes the data matrix into several components and each singular value represents the importance of the component. This method removes the small components and reconstructs the data matrix with a reduced version of the SVD. It should be noticed that there is no direct relationship between the singular value and the structural mode. As a matter of fact, the number of the singular values will control the number of available modes. So it can be said that the criterion in Equation [10.35] reduces the noise effect by removing the small components from original data but not directly removing the noise modes.

10.8   Results using subspace identification (SI) to seismic response data

For each seismic response datum the SI technique was used to identify the system dynamic characteristics. A total of 84 event data were used. Two different types of input data were considered in this study. One considers the multiple input motions along the dam abutment (from station SD1 to station SD5), the other one considers a single input from one of the stations in the abutment (i.e. either station SD1 or SD5). To select the suitable number of block rows for identification the stability diagram needs to be constructed in advance too. Two seismic event data were used to construct the stability diagram: one is the 2002-3-31 event ($M = 6.8$) and the other is 1999-9-21 event ($M = 7.3$). Figure 10.13 shows the stability diagram developed from these two earthquake events. As compared to the Fourier amplitude spectrum of the dam crest response, multiple inputs can identify higher frequency modes, but the fundamental mode is not so easily determined. On the contrary, if a single input is considered, then the fundamental mode can be clearly identified but higher modes are not. Based on the identified system natural frequencies
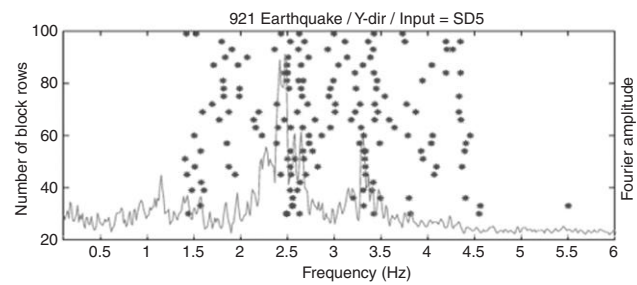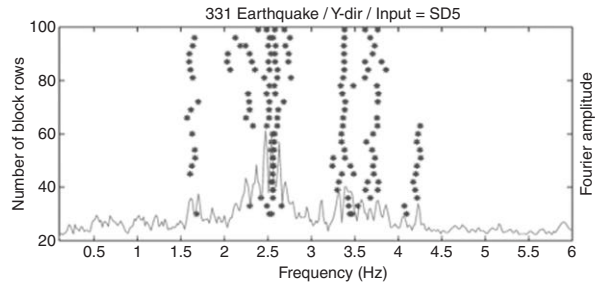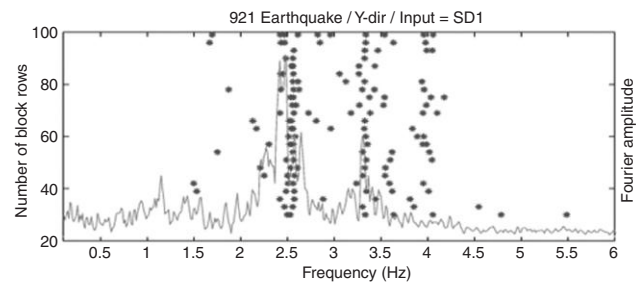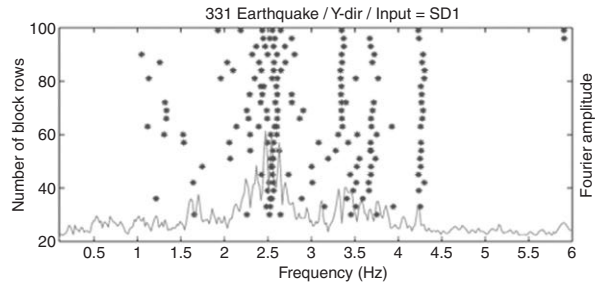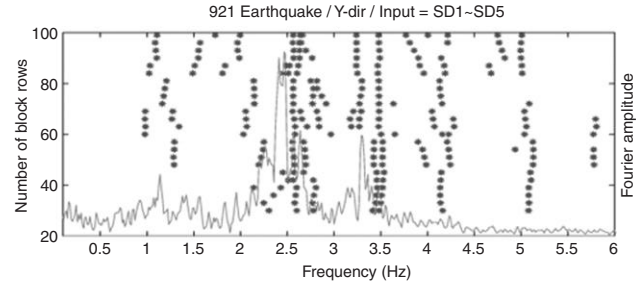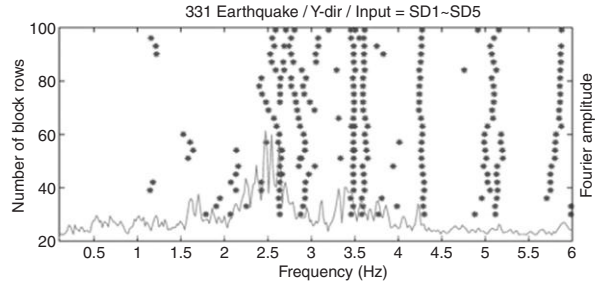
10.13 Stability diagram (using different sets of input motions) from two seismic events: 912 earthquake and 331 earthquake.

(a)
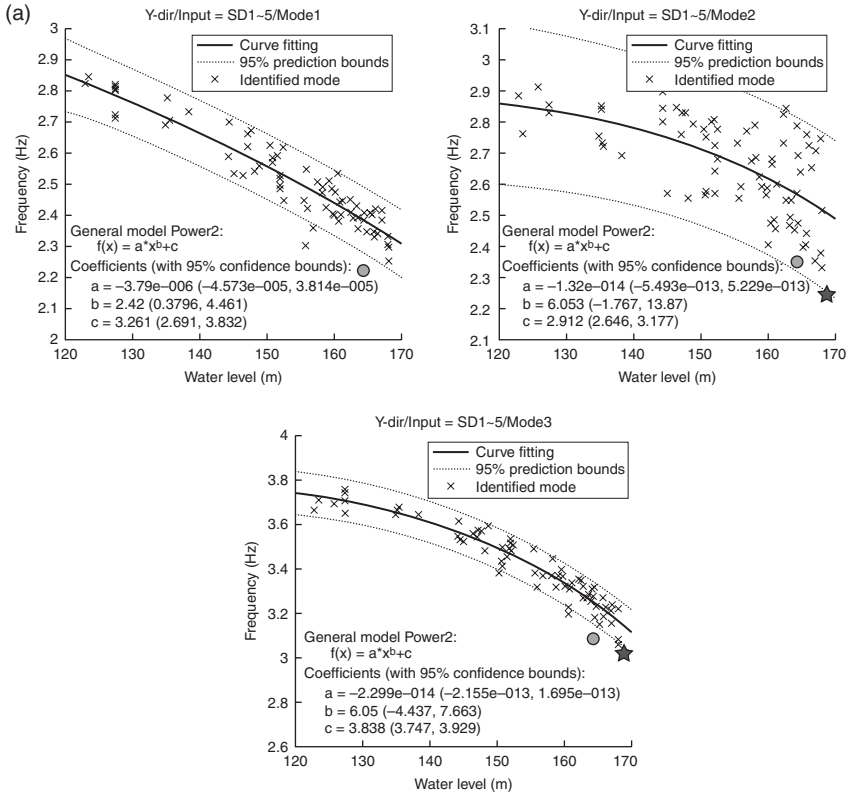
10.14 (a) The relationship between the identified system natural frequencies using multiple inputs (the first three modes) with respect to water height in the reservoir. (b) The relationship between the identified system natural frequencies (the first three modes) using single inputs with respect to water height in the reservoir. (c) The relationship between the identified system natural frequencies (the first three modes) using single inputs with respect to water height in the reservoir.

(*Continued*)

(the first three modes) from all 84 events, the relationship between the identified frequency, $f(x)$, with respect to reservoir water height was generated:[37]

$$f(x) = a\,x^b + c \hspace{4cm} [10.36]$$

where x is the reservoir water depth. Figure 10.14a shows the system natural frequency changes with respect to water depth. The system natural frequencies were identified using multiple inputs. Figures 10.14b and 10.14c show the results of using a single input. Mean and standard deviation of the regression curve are also shown in the figure. The results from ambient vibration and forced vibration are also plotted in the figure for comparison.[38]

(b)

Y-dir/Input = SD1/Mode1

General model Power2:
$f(x) = a*x^b+c$

$a = -1.209e{-}011\ (-1.61e{-}010,\ 1.368e{-}010)$
$b = 4.812\ (2.457,\ 7.168)$
$c = 2.82\ (2.646,\ 2.995)$

Y-dir/Input = SD1/Mode2

General model Power2:
$f(x) = a*x^b+c$
Coefficients (with 95% confidence bounds):
$a = -2.34e{-}014\ (-3.497e{-}013,\ 3.029e{-}013)$
$b = 6.041\ (3.356,\ 8.726)$
$c = 2.917\ (2.764,\ 3.069)$

Y-dir/Input = SD1/Mode3

General model Power2:
$f(x) = a*x^b+c$
Coefficients (with 95% confidence bounds):
$a = -2.719e{-}014\ (-3.04e{-}013,\ 2.497e{-}013)$
$b = 6.046\ (-4.084,\ 8.007)$
$c = 3.746\ (3.616,\ 3.876)$

10.14 Continued

10.9   Results using ARX model to seismic response data

Different from the SI (using state space model), to identify the dynamic behavior of the dam during earthquake excitation, the transfer function between input and output can also be investigated. The multiple input/multiple output discrete-time ARX(MIMO) model was applied. The ARX-MIMO model can be expanded and expressed as:

$$\mathbf{y}(t) = \frac{\mathbf{B}(q^{-1})}{\mathbf{A}(q^{-1})}\mathbf{u}(t) + \frac{1}{\mathbf{A}(q^{-1})}\mathbf{e}(t) \qquad [10.37]$$

where AR refers to the auto-regressive part, $\mathbf{A}(q^{-1})\mathbf{y(t)}$, and X to the extra input, $\mathbf{B}(q^{-1})\mathbf{u(q)}$. Here $\mathbf{y(t)}$ and $\mathbf{u(t)}$ are (nyx1) and (nux1) dimensional column vectors which refer to the multiple output/multiple input case. $\mathbf{A}(q^{-1})$ is an (ny x ny) matrix polynomial, and $\mathbf{B}(q^{-1})$ is an (ny x nu) matrix,[38] i.e.

(c)

Y-dir/Input = SD5/Mode1

— Curve fitting
⋯⋯ 95% prediction bounds
× Identified mode

General model Power2:
f(x) = a*x^b+c
Coefficients (with 95% confidence bounds):
a = −5.464e−011 (−7.033e−010, 5.94e−010)
b = 4.536 (2.269, 6.802)
c = 2.872 (2.673, 3.071)

Frequency (Hz) — Water level (m)

Y-dir/Input = SD5/Mode2

— Curve fitting
⋯⋯ 95% prediction bounds
× Identified mode

General model Power2:
f(x) = a*x^b+c
Coefficients (with 95% confidence bounds):
a = −1.991e−014 (−2.747e−013, 2.348e−013)
b = 6.069 (3.605, 8.533)
c = 2.89 (2.755, 3.026)

Frequency (Hz) — Water level (m)

Y-dir/Input = SD5/Mode3

— Curve fitting
⋯⋯ 95% prediction bounds
× Identified mode

General model Power2:
f(x) = a*x^b+c
Coefficients (with 95% confidence bounds):
a = −2.581e−014 (−2.463e−013, 1.947e−013)
b = 6.048 (−4.402, 7.694)
c = 3.736 (3.631, 3.841)

Frequency (Hz) — Water level (m)

10.14 Continued

$$\mathbf{A}(q) = \begin{bmatrix} a_{11}(q^{-1}) & \cdots & a_{1*ny}(q^{-1}) \\ \vdots & & \\ a_{ny*1}(q^{-1}) & \cdots & a_{ny*ny}(q^{-1}) \end{bmatrix} \qquad [10.38]$$

where the entries a_{ij} are polynomials in the delay operator q^{-1}:

$$a_{kj}(q^{-1}) = \delta_{kj} + a^1_{kj}q^{-1} + \ldots + a^{nakj}_{kj}q^{-ngkj} \qquad [10.39]$$

And $\mathbf{B}(q^{-1})$ in Eq. [10.37] is a $(1 \times n_u)$ polynomial matrix with

$$\mathbf{B}(q^{-1}) = \left[b_{11}(q^{-1}), \, b_{12}(q^{-1}), \ldots, b_{1n_u}(q^{-1}) \right] \qquad [10.40]$$

where the entries $b_{ij}(q^{-1})$ are polynomials in the delay operator q^{-1}:
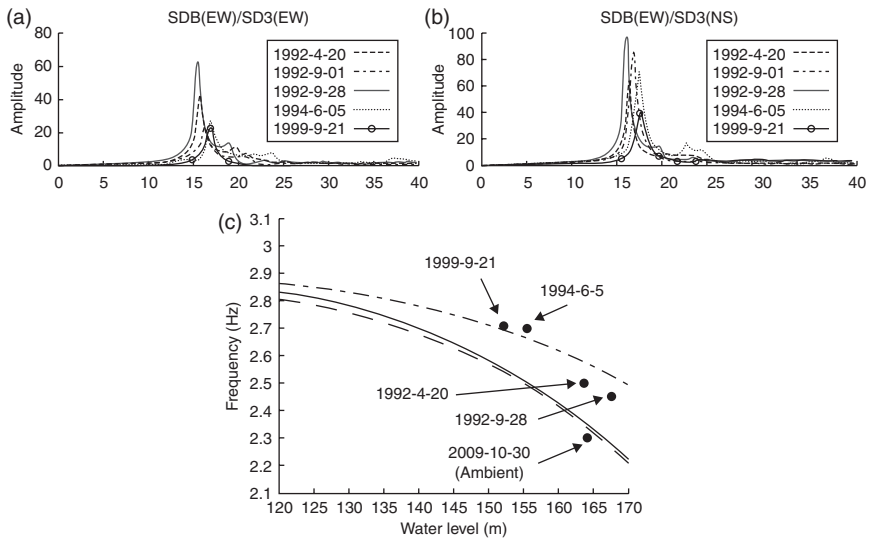
10.15 Identified system transfer function using ARX-MIMO model (a) TF between SBD (EW-direction) and SD3(EW-direction) (b) TF between SBD (EW-direction) and SD3 (NS-direction) (c) Plot of the identified dominant frequency from TF to check with the regression line of frequency-water level relationship from subspace identification.

$$b_{ij}(q^{-1}) = b_{ij}^1 q^{-1} + b_{ij}^2 q^{-2} + \ldots + b_{ij}^{n_{uj}} q^{-n_{uj}} \qquad [10.41]$$

and j denotes the j-th component of inputs and n_{uj} denotes its corresponding order.

Through the recursive least squares method the modal parameters of ARX model can be identified, and the system natural frequency and damping ratio of each structural mode can be estimated through the relation between discrete form of equation of motion (modal equation) and ARX model.

Five earthquake response data were used to estimate the frequency response function (FRF) between Station SDB and Station SD3 (all in east–west direction or up–down stream direction) from the selected five earthquake events, as shown in Fig. 10.15a. Figure 10.15b also shows the estimated FRF between Station SDB (east–west direction) and Station SD3 (north–south direction). The change of dominant frequency in FRF indicated the difference of water level in the reservoir. Comparison on the identified system natural frequency using either SI method (mean curves) or ARX method is shown in Fig. 10.15c.

10.10 Conclusion

In this study an intensive SHM system on a dam structure is introduced. This SHM system covers both the static and dynamic measurement of the dam. Three different monitoring methods, including earthquake response monitoring, ambient vibration survey, and long-term static deformation monitoring of an arch dam (Fei-Tsui arch dam, Taiwan), are studied in this report. For dynamic response measurements (earthquake response and ambient vibration), the multivariate data processing technique is used with the response measurements so as to extract the dynamic features of the system. For ambient vibration measurement, wireless sensing technology is used to collect ambient vibration data of the dam. This wireless sensing system can provide the requirements for autonomous SHM, which include (1) reliable wireless communication data, (2) accurate signal conditioner hardware for ambient vibration sensors (VSE-15D), and (3) capability for continuous long-term monitoring. The SSI method is then applied to identify the system frequencies, damping ratios, and mode shapes. For the identification of dam properties from seismic response data of the dam, considering the non-uniform excitation of the seismic input, two different approaches are used, the SI and the MIMO discrete-time ARX model with least squares estimation. Applying the SI technique to data collected from the dam responses for 84 earthquake events, the change of the identified system natural frequencies with respect to water level in the reservoir is investigated. The results from both the seismic response and the ambient vibration survey can be used for safety assessment of the dam.

10.11 References

1. Doebling, S.W., C.R. Farrar, M.B. Prime and D.W. Shevitz (1996), Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review, *Tech Rep LA-13070-MS, Los Alamos National Laboratory*, Los Alamos, NM USA.
2. Staszewski, W. (1996), Gearbox vibration diagnostics – an overview, *Proceedings of the 8th OMADEM-96)*, Sheffield, England, 16–18 July 1996.
3. Peeters, B. and De Roeck, G. (2001), One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake Engineering and Structural Dynamics*, **30**: 149–171.
4. Chen, S., Billings, S.A. and Grant, P.M. (1990), Non-linear system identification using neural networks. *International Journal of Control,* **55**(1):1191–1214.
5. Masri, S.F., Chassiakos, A.G. and Caughey, T.K. (1992), Structure-unknown nonlinear dynamic systems: identification through neural networks. *Smart Materials and Structures,* **1**:45–56.
6. Masri, S.F., Chassiakos, A.G. and Caughey, T.K. (1993), Identification of nonlinear dynamic systems using neural networks. *Journal of Applied Mechanics,* **60**:123–133.

7. Kumar, M.V., Omkao, S.N., Ganguli, R., Sampath, P. and Suresh. S. (2006), Identification of helicopter dynamics using recurrent neural networks and flight data. *Journal of American Helicopter Society,* **51**(2):164–174.

8. Szewezyk, P. and Hajela, P. (1994), Damage detection in structures based on feature-sensitivity neural networks. *Journal of computing in Civil Engineering*, *ASCE*, **8**(2):163–179.

9. Pandy, P.C. and Barai, S.V. (1995), Multilayer perception in damage detection of bridge structures. *Computers & Structures,* **54**(4): 597–608.

10. Masri, S.F., Smyth, A.W., Chassiakos, A.G., Caughey, T.K. and Hunter, N.F. (2000), Application of neural networks for detection of changes in nonlinear systems. *Journal of Engineering Mechanics*, *ASCE,* **126**(7):666–676.

11. Sohn, H., Dzwonczyk, M., Straser, E.G., Kiremidjian, A.S., Law, K.H. and Meng, T. (1999), An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge. *Earthquake Engineering and Structural Dynamics*, **28**: 879–897.

12. Sohn, H., Worden, K. and Farrar, C.F. (2001), Novelty detection under changing environmental conditions. *SPIE's Eighth Annual International Symposium on Smart Structures and Materials*, Newport Beach, CA. (LA-UR-01-1894).

13. Giraldo, D.F., Dyke, S.J. and Caicedo, J.M. (2006), Damage detection accommodating varying environmental conditions. *Structural Health Monitoring*, **5**(2), 155–172.

14. Hsu, T.Y. and Loh, C.H. (2010), Damage detection accommodating nonlinear environmental effects by nonlinear principal component analysis. *Structural Control & Health Monitoring*, 2009 **17**: 338–354.

15. Okamoto, S., Yoshida, M. Kato, K. and Hakuno, M. (1964), Dynamic behavior of an arch dam during earthquake, Report of Institute of Industrial Science, The University of Tokyo, Vol. **14**, No.2, 54–119.

16. Mickey, W., Perez, V. and Cloud, W.K. (1974), Amplification study of the Pacoima dam fro aftershock of the San Fernando earthquake, *Proceedings of 5th WCEE*, Rome, Vol. **1**, 755–762.

17. Castoldi, A. (1978), Contribution of the surveillance to the evaluation of the seismic efficiency of dams: example of the Ambiesta dam, Seminar on Constructions in Seismic Zones, Bergamo, Vol. 4, 107–118.

18. CSMIP Strong Motion Records from the Whittier, California earthquake of 1 October 1987, Report OSMS 87-05, California, Department of Conservation, CDMG, Office of Strong Motion Studies, 1987.

19. Proulx1, J. and Darbre, G.R. (2008), Earthquake response of large arch dams observational evidence and numerical modeling, *Proceedings of the 14WCEE*. October Beijing.

20. Oliveriva, S., Espoda, M. and Camara, R. (2012), Long-term dynamic monitoring of arch dam: The case of Cabril dam, *Proceedings of 15WCEE*, Lisbon.

21. Hoyo, P. and Oosthuiz, P. (2009), Dynamic testing of a concrete arch dam, *Proceedings of 3rd International Operational Modal Analysis Conference, IOMAC'09*, 95–100.

22. Loh, C.H. and Wu, T.S. (1996), Identification of Fei-Tsui arch dam from both ambient and seismic response data. *Soil Dynamics and Earthquake Engineering*, **15**: 465–483.

23. Cybenko, G. (1989), Approximation by superpositions of a sigmoidal function. *Math. Control Signals System*, **2**: 303–314.

24. Jolliffe, I.T. (2002), *Principal Component Analysis* (second edition). New York: Springer.

25. Loh, C. H., C.H. Chen and T.Y. Hsu (2011), Application of advanced statistical methods for extracting long-term trends in static monitoring data from an arch dam. *Journal of Structural Health Monitoring*, **10**(6): 587–601.

26. Lu, K.C., Weng, J.H. and Loh, C.H. (2009), Turning the building into a smart structure: Integrating health monitoring, *Proceedings of the SPIE 16th International Symposium on Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Vol. **7292**, San Diego, USA, March 2009.

27. Wang, Y., Lynch, J.P. and Law, K.H. (2007), A wireless structural health monitoring system with multithreaded sensing devices: design and validation. *Structure and Infrastructure Engineering*, **3**(2): 103–120.

28. Lynch, J.P., Law, K.H., Kiremidjian, A.S., Carryer, E., Kenny, T.W. *et al.* (2002), Validation of a wireless modular monitoring system for structures. *Proceedings SPIE Smart Structures and Materials: Smart Systems for Bridges, Structures, and Highways*, San Diego, CA **4696**(2), 17–21.

29. Lu, K.C., Weng, J.H. and Loh, C.H. (2010), Development of improved wireless sensing system for SHM, *Proceedings of ASCE Earth & Space 2010 Conference*, Honolulu, 14 March–17 March, 2010.

30. Van Overschee, P. and De Moor, B. (1993), Subspace algorithm for the stochastic identification problem. *Automatica*, **29**(3): 649–660.

31. Van Overschee, P. and De Moor, B. (1996), *Subspace Identification for Linear System: Theory-Implementation-Applications*. Dordrecht (Netherlands): Kluwer Academic Publishers.

32. Larimore W.E. (1994), The optimality of canonical variate identification by example, *Proceedings of the SYSID'94*, 4–6 July 1994, Copenhagen, Denmark, Vol. **2**, pp. 151–156.

33. Van Overschee, P. and De Moor, B. (1994), N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica,* **30** (1): 75–93.

34. Verhaegen, M. (1994), Identification of the deterministic part of MIMO state space models given in innovations from input–output data. *Automatica,* **30** (1): 61–74.

35. Willems, J. (1987), From time series to linear systems. *Automatica*, Part 1: **22**(5): 561–580; Part II:22(6), 675–694; Part III: 23(1), 87–115.

36. Peeters, B. and De Roeck, G. (1999), Reference-based stochastic subspace identification for output- only modal analysis. *Mechanical Systems and Signal Processing,* **13**(6): 855–878.

37. Weng, J.H. (2010), Applications of Subspace Identification in System Identification and Structural Damage Detection, Ph.D. Thesis (Advisor: C. H. Loh), Civil Engineering Department, National Taiwan University, June 2010.

38. Loh, C.H. and Wu T.S. (2000), System ID of Fei-Tsui arch dam from forced vibration and seismic response data. *Journal of Earthquake Engineering*, **4**(4): 511–537.

11

Sensing solutions for assessing and monitoring tunnels

N. A. HOULT, Queen's University, Canada and K. SOGA,
University of Cambridge, UK

Abstract: This chapter introduces the most common sensors used for the
monitoring of tunnels. It is divided into sections that cover typical tunnel
monitoring applications: (i) construction in soft ground;
(ii) construction in rock; and (iii) in-service and long-term monitoring.
A case study is used for each application to illustrate the use of both
common and newly developed sensing technologies in tunnel monitoring.
The chapter concludes with a summary of the sensors discussed, future
trends in sensing for tunnels and useful references.

Key words: monitoring, tunnelling, soft ground, rock, sensors,
displacements, strains, pressures, construction, deterioration.

11.1   Introduction

There are two overarching reasons to monitor tunnels: (i) to minimize
the associated risks especially to the public; and (ii) to minimize the cost
of construction/operation/maintenance. Monitoring can also offer other
advantages, such as providing insights into ground response, feedback for
construction control, verification of design assumptions, assurance of tunnel
lining performance, and evidence of any adverse effects on the surrounding
area (ITA, 2011). Monitoring has been used extensively in tunnel applica-
tions for decades, ever since the development of the observational method
of tunnelling where feedback from the monitoring system is used to inform
and potentially modify design and construction (Peck, 1969; Nicholson *et al.*,
1999). However, when choosing an appropriate monitoring system, a myriad
of factors need to be considered, including whether the tunnel is being con-
structed or is already in service, the ground conditions, and the condition of
surrounding infrastructure, among others. A wide variety of sensor technol-
ogies is also available from traditional surveying techniques to newly devel-
oped wireless sensor networks (WSNs) and distributed fibre optic sensor
systems. The choice of monitoring approach is dependent on the situation

309

and the information required by engineers to minimize both the risk and the cost.

This chapter is written for someone who has a tunnel that potentially requires monitoring. It is subdivided into two main sections: construction monitoring and in-service monitoring. In both sections, some of the major issues most often encountered are highlighted, and appropriate monitoring technologies are discussed. The chapter concludes with a brief discussion of future trends and other sources of information. Case studies are used throughout to illustrate potential monitoring solutions. It should be noted that a wide variety of sensing technologies is available and geotechnical monitoring has itself been the subject of books (e.g. Dunnicliff, 1993). As such, the goal of this chapter is not to be exhaustive in its coverage of the technologies available, but instead to highlight the more commonly used technologies and some newly developed technologies through case studies. It is also worth noting that temperature differentials and extremes often play a significant role in both the behaviour of the infrastructure being monitored (e.g. British Tunnelling Society and Institution of Civil Engineers, 2004) and the monitoring equipment (e.g. Cheung *et al.*, 2010). As such, the monitoring of temperature, which is not dealt with explicitly in this chapter, should be considered whenever changes in temperature or extreme temperatures are expected.

11.2 Construction monitoring in soft ground tunnelling

There are several different approaches to constructing tunnels in soft ground, as outlined in Table 11.1 but, as noted by Dunnicliff (1993), the primary reason for monitoring tunnels is to ensure tunnel face stability and to prevent possible ground failure. However, more recently, monitoring of deformation has become increasingly important due to the need to prevent any excessive ground movements that may damage nearby structures (both below and above surface). There are a number of issues related to stability and deformation that potentially need to be monitored, the most significant of which are: (i) excessive displacements leading to surface settlements, damage to surrounding infrastructure and excessive deformation of the tunnel itself; (ii) unexpected loads acting in and on the supports; and (iii) changes in pore water pressures leading to seepage and hence changes in loading.

11.2.1 Excessive displacements

The need to measure surface displacements, damage to surrounding infrastructure and tunnel displacements, as with all monitoring applications, is

Table 11.1 Construction methods for soft ground tunnels

Technique	Brief description
Cut and cover	In this technique a trench is dug for most or all of the length of the tunnel. It is typically used to create underground stations. The side walls of the trench are supported and form the walls of the tunnel and a structural roof is placed over the tunnel. Chief among the monitoring issues with this type of tunnelling are excessive deformations of the side walls causing damage to the surrounding infrastructure.
Conventional tunnelling	Excavation of a tunnel without the aid of a shield (see below). Mechanical excavators are used to remove material from the tunnel face. The material is then removed from the tunnel (i.e. mucking) and the tunnel is supported using primary support elements such as steel arch supports, sprayed or cast-in place concrete. It is also known as NATM, SCL, or SEM. The method is flexible enough to create complex underground structures such as cross passages. However, larger ground movement is expected and therefore monitoring is essential when a tunnel is constructed using this method.
Shield method	This approach reduces the need for access from the ground surface and the tunnel can be dug under existing infrastructure and/or obstacles. A shield at the tunnel excavation face provides support to the face while material is excavated and removed out through the already constructed tunnel. Tunnel lining is put in place as the shield is advanced forward in order to support the already excavated tunnel. Historically, shields protected workers who were digging at the excavation face, but now they are commonly part of a TBM, which excavates and removes material while simultaneously placing the tunnel lining. Once again issues of ground settlement and damage to surrounding infrastructure are important to monitor when using this technique; however, tunnel stability is also a significant issue to ensure that the tunnel does not collapse.
Open face	The open-face approach is a type of shield method construction where the tunnel face does not require short-term support during excavation. The excavated sections of the tunnelling are lined as with other types of shield method construction.
Earth pressure balance	Another variation on the shield approach where the tunnel face is not capable of supporting itself even over short periods. In this approach, the pressure the TBM exerts on the tunnel face is balanced against the earth pressure by controlling the rate at which the TBM moves forward. The exerted face pressure is controlled by the pressure of the excavated soil inside the tunnel chamber behind the cutter face. The ground settlement can be minimized by effective control of the face pressure.
Slurry shield	A further variation on the shield approach where not only is the tunnel face not self-supporting but the soil has high water levels/pressures. It is often used in very soft grounds. The soil is mixed with a bentonite slurry at the face of the shield machine so that face pressure can be maintained and the soil/slurry mixture can be removed through pipes that run the length of the shield machine.

a function of the consequences of excessive deformations. In most urban applications where the density of infrastructure that can be damaged by excessive displacements (e.g. roads, buildings, retaining walls) is high, monitoring is generally a primary requirement. The extent and magnitude of expected deformations is a function of the tunnelling method. For example, Mair (1996) and Mair and Taylor (1997) suggest that volume losses from open-face tunnelling can range between 1% and 3%; however, these losses can be reduced significantly (to less than 1%) if earth pressure balance tunnelling machines are used (Mair, 2008). Additionally, displacements can be controlled through the use of compensation grouting or other ground improvement techniques, in which case the use of monitoring is essential to determine the amount of grout or hardening agents required. Settlements due to cut and cover tunnelling are a function of the stiffness of the excavation support as well as the construction processes (Hung *et al.*, 2009). Damage to surrounding infrastructure can include cracking of masonry structures and pipelines. When monitoring surrounding infrastructure, it is critical to establish the condition of the infrastructure before construction begins, both to accurately determine the effects of the construction and in case of legal action. This raises the question, discussed in greater detail elsewhere (Dunnicliff, 1993), of who should be responsible for installing and maintaining the monitoring system. Finally, excessive tunnel displacements can result in the tunnel not being fit for purpose, or being an indication of potential tunnel collapse.

Monitoring techniques

There are a number of instruments for measuring displacements including: (i) surveying equipment; (ii) probe, rod, fixed borehole and multipoint extensometers; (iii) crack gauges; (iv) liquid level and electro-level gauges; (v) fibre optics; (vi) inclinometers; (vii) convergence gauges; and (viii) photogrammetry (see Table 11.3 for a list of available sensors). Figure 11.1 illustrates a number of the potential issues associated with tunnel construction and the sensor technologies that can be used to monitor them. It is worth noting that while monitoring equipment placed on the surface can be used to obtain pre-construction measurements, equipment used within the tunnel (except for sliding micrometers – see Section 11.5.5) often does not capture initial displacements which occur just ahead of and behind the face of the tunnel excavation. As such, monitoring equipment within the tunnel can usually provide measurements only after it is installed and is generally used to monitor long-term behaviour (Kavvadas, 2005).

One of the most common methods for measuring displacements in the tunnel and of points on the ground is surveying using traditional instruments or more frequently total stations. Total stations can either be manually

Rotation of structures
 - Surveying
 - Inclinometers
 - Photogrammetry

Cracking of structures
 - Crack meters
 - Photogrammetry

Pore water pressures
 - Piezometers

Displacement of surface
 - Surveying
 - Extensometers
 - Liquid levels

Displacement of subsurface
 - Extensometers
 - Inclinometers

Distortion of tunnel geometry
 - Surveying
 - Extensometers
 - Fibre optics

Support/lining stresses
 - Strain gauges
 - Pressure cells

11.1 Construction issues and associated monitoring technologies.

operated by a surveyor taking measurements using targets that are either temporary or fixed, or set up to operate automatically. One such example of an automatic system is an automated total station used in the London Underground, where the instrument takes frequent measurements of a pre-determined pattern of targets (Wright, 2010). Surveying techniques, with correctly placed targets, can be used to estimate ground settlements, movement of surrounding infrastructure and deformations within the tunnel. These systems can also be used to provide real-time data, which can be crucial when using compensation grouting as discussed by Mair (2008). As with any monitoring technology, it is important to know both the accuracy and precision offered by the equipment so that these tolerances are accounted for in the design. An added complication with measurements taken within the tunnel is that the fixed point of reference can be very far away (sometimes outside the tunnel) and the intermediate measurement locations (usually wall mounted brackets) are themselves subject to long-term displacement.

A second frequently used category of sensors for measuring ground settlements in the subsurface are extensometers. Broadly speaking, extensometers measure changes in the length of an object and so a wide variety of extensometers is available. Probe extensometers consist of a magnetic probe, a series of magnets placed at various depths down a borehole and an instrument for taking readings. The probe is lowered into the borehole and as it encounters the magnets a reading is taken. The change in distance between each magnet, and thus the amount or relative ground settlement, can be determined. Rod extensometers, as seen in Fig. 11.2, allow the distance between two points (the anchor point and the reference head) to be measured using either a depth gauge or a displacement transducer. They are typically used to measure the movement of the tunnel wall relative to a fixed
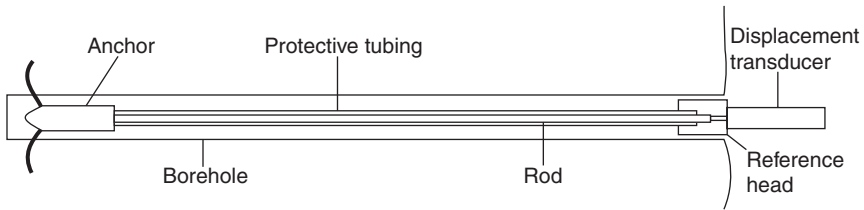
11.2 Rod extensometer.

point some distance away to determine the zone of influence of the tunnel, but they can also be used to measure surface displacements. Fixed borehole extensometers are similar to rod extensometers. Multipoint extensometers can refer to probe, rod and borehole extensometers that have multiple measurement locations rather than just measuring the distance between two fixed points. Extensometers can be used to be measure ground settlements, ground movement relative to the tunnel and tunnel convergence (the change in tunnel diameter).

For the most part, surveying techniques provide the most robust way to monitor the displacement of surrounding infrastructure. However, the locations of targets are limited by line of sight to the instrument/total station. Traditional crack gauges offer an inexpensive but effective way to monitor the effect of tunnel construction on surrounding infrastructure. These gauges consist of two pieces where each piece is affixed on either side of a pre-existing crack and the relative movement between these two pieces can be tracked using a grid system (as illustrated in Fig. 11.3). There are two main disadvantages of traditional crack gauges: (i) they must be read manually and so are not suitable for real-time feedback systems, and (ii) the precision of the measurements is dependent on the person taking the readings, and so for areas where very small displacements are critical they may not be appropriate. A second option for measuring the development of cracks is the use of demountable mechanical (DEMEC) strain gauges. In this case, fixed points are attached to the structure and the change in displacement between these points is measured using a portable device with a dial gauge or digital readout. The DEMEC system is capable of measuring crack widths down to 0.001 mm, although the accuracy and precision are limited to 0.01 mm (Burland *et al.*, 2001a). Alternative crack sensors that use displacement transducers and vibrating wire strain gauges (as illustrated in Fig. 11.3) are available, although with a significant increase in system cost. In Section 11.7.2 a wireless crack sensor is introduced, which can provide real-time results and can be installed faster than conventional wired sensors.

Electronic liquid level (or electro-level) sensors are made up of a series of pots containing liquid with each connected to a central pot with a constant height of liquid (Dunniclff, 1993). The liquid level in each pot is measured

11.3 Crack gauge. (*Source*: Picture courtesy of itmsoil.)

by a displacement transducer attached to a float and as the structure settles the level of liquid in each pot shifts, allowing displacements to be measured. They are very popular for monitoring differential settlements of structures. For example, these gauges were used on a masonry railway bridge in Bologna to determine the amount of compensation grouting required to ensure that the masonry bridge did not crack (Mair, 2008).

Distributed fibre optic strain sensing systems such as Brillouin optical time domain reflectometry (BOTDR) and Brillouin optical time domain analysis (BOTDA) are fibre optic measurement techniques that allow the strain in a fibre optic cable to be measured along the full length of the cable (see also Section 11.7.2). If the fibre optic cable is fixed between two points and the length between the points is known, the strain can be integrated to get a displacement. If the fixed points are placed around the circumference of a tunnel, displacements around the circumference can be measured. These systems have been used to measure displacements in existing tunnels during the construction of new tunnels in both Singapore (Mohamad *et al.*, 2012) and London (Mohamad *et al.*, 2010). The BOTDR technique showed good agreement with the expected behaviour (Mohamad *et al.*, 2010, 2012). One of the appealing aspects of this system is the ease of installation and the number of data points that can be provided. However,

as discussed in Section 11.7.2, system robustness is still an issue that needs to be addressed.

Inclinometers (otherwise known as tiltmeters or clinometers) measure changes in inclination. They can be used in a number of tunnel applications, including measuring rotations due to settlements at the surface as well as in surrounding infrastructure. Traditional inclinometers use a level bubble to indicate tilt; however, more modern inclinometers use accelerometers, which allow the data to be recorded and transmitted. The recent development of microelectromechanical sensors (MEMS) inclinometers has opened up the potential for relatively inexpensive wireless inclinometers, as introduced in the case study in Section 11.7.2. Inclinometers can also be used to measure lateral ground movement in the area of tunnel construction. Boreholes are drilled and a casing is inserted in the hole against which inclination measurements can be taken (Machan and Bennett, 2008). Two types of systems exist: traversing and in-place/stationary. The traversing system consists of a single inclinometer on the end of a cable, which is lowered down the borehole and readings are taken at specific heights along the casing. The in-place stationary system uses a series of inclinometers that are left in the casing at predetermined heights and measurements are taken over time.

Convergence gauges are used to measure changes in the tunnel's profile during construction. Traditionally, these measurements have been taken using tape extensometers and measuring between fixed points on the tunnel wall. While this approach is still common, the measurement of tunnel wall movement is now also done using surveying techniques or potentially by using Lidar as discussed in Section 11.4.1.

Photogrammetry uses photographs to determine initial geometry as well as changes in geometry. A variation of this technique, known as particle image velocimetry (PIV), has been used successfully with digital images to measure displacements in geotechnical applications (White *et al.*, 2003), including field monitoring of a retaining wall on the London Underground during the construction of the Channel Tunnel Rail Link (Take *et al.*, 2005).

11.2.2   Unexpected loads acting in and on the supports

If the loads in the tunnel lining due to the surrounding earth pressures exceed the design values, this can result in deformation of the tunnel lining until a new equilibrium situation is reached, excessive damage to the tunnel lining occurs or, in the worst case, the tunnel collapses. In cut and cover construction, if the loads in the struts that support the excavation walls exceed the design values, there is the potential for excessive deflections of the excavation or collapse of the excavation. As a result, there is sometimes a need to monitor such conditions, especially when new construction techniques are being used or when there is high variability in the surrounding conditions.

11.4 Vibrating wire strain gauge. (*Source*: Picture courtesy itmsoil.)

Monitoring techniques

Two of the most commonly used techniques for measuring the load in the supports for a tunnel are (i) strain gauges, and (ii) pressure cells.

Traditionally the two types of strain gauges used in geotechnical and structural monitoring are vibrating wire and resistive strain gauges. However, in recent years fibre optics has also been used to measure strains. Vibrating wire strain gauges, as seen in Fig. 11.4, are currently the most widely used strain measurement technology and are available from most geotechnical monitoring firms. Batten and Powrie (2000) used vibrating wire strain gauges to measure the prop loads in an excavation in London to determine the effects of temperature and construction stage on the loading so that they could create models to predict the response. One of the issues they noted was the effect of end conditions on the strain measurements and thus the estimate of load in the prop. The use of fibre optics to measure strains is dealt with in Section 11.5.4. An important consideration when using strain gauges to determine the load in a support is how to estimate the stiffness of the structure in order to estimate the stress. For a steel structure, such as the struts monitored by Batten and Powrie (2000), this can be straightforward, but for a concrete or reinforced concrete structure, where creep and

shrinkage effects need to be accounted for, it can be quite difficult to use strains to estimate stresses. Acerbis *et al*. (2011) present an approach for estimating concrete stress based on strain measurements, which takes into account these various time dependent effects. Alternatively, they should be used to monitor change in strain (or stress) rather than evaluating the actual strain (or stress).

Pressure cells can be installed in tunnel linings to measure the radial stress (the pressure exerted on the tunnel by the surrounding ground) or tangential stress (the hoop stress within the lining). In general, they are more suitable for evaluating stress changes rather than the actual stress. These cells can give a realistic indication of the stress changes in segmental linings as well as a sprayed concrete lining (SCL); however, as noted by Clayton *et al*. (2002), careful attention must be paid to the installation of these cells as well as to the effects of temperature, crimping and concrete shrinkage. Hashimoto *et al*. (2009) used pad type earth pressure cells to determine the short- and long-term pressures acting on a tunnel lining due to the effects of grouting and varying soil type. Their results showed that the bending moment in some linings may be overestimated using conventional design approaches. Thus, while a chief concern may be the overloading of the tunnel lining, this type of monitoring may also allow for a refined design to be used. However, it should be stressed that careful evaluation of the performance of pressure cells is needed before making any use of the values obtained.

11.5 Spade pressure cell. (*Source*: Picture courtesy itmsoil.)

Low profile push-in pressure cells, or spade cells (Fig. 11.5), can also be used to measure horizontal soil pressure in normally consolidated and over-consolidated clays. They are composed of two thin pieces of metal welded together around the edges that form a long thin cell into which oil is inserted. The cell, which has a shape similar to a spade, can then be pushed into the clay. The pressure of the soil is then measured as a change in oil pressure within the cell using a pneumatic or vibrating wire pressure transducer (Richards *et al.*, 2007). Because the cell is installed by displacing the surrounding soil, stress concentrations are created locally around the cell, which results in overestimates of the soil stress (Ryley and Carder, 1995). Correction factors for the readings from spade cells based on the level of over-consolidation and the type of clay have been proposed (e.g. Ryley and Carder, 1995; Richards *et al.*, 2007).

11.2.3   Pore water pressures

The pore water pressure in the surrounding soil influences both the potential for seepage into the tunnel as well as the loading on the tunnel lining due to changes in effective stress (Shin *et al.*, 2005; Wongsaroj *et al.*, 2007). At the same time, attempts to control leakage can lead to an increase in pore water pressure acting on the tunnel, which may lead to structural failure (Parks *et al.*, 1986). Additionally, the effect of pore water pressure on non-circular tunnel linings can be significant (Shin *et al.*, 2005). The amount of soil consolidation around the tunnel is governed by the amount of seepage into the tunnel (i.e. whether the tunnel causes the soil to act as drained or partially drained) (Wongsaroj, 2006; Laver, 2011; Laver and Soga, 2011). As such, there may be a need for short- and long-term monitoring of pore water pressure to determine the effects on the tunnel behaviour as well as the long-term settlement of the surrounding ground as discussed in Section 11.6.1.

Monitoring techniques

There are a variety of techniques used to monitor water pressure including: (i) open standpipe piezometers; (ii) vibrating wire piezometers; (iii) pneumatic piezometers; and (iv) fibre optic piezometers.

The open standpipe piezometer typically consists of a borehole with a narrow tube (10–20 mm) inserted in it (although there are push-in versions that do not require a borehole). At the base of the tube is a filter to let water in. Once the tube and filter have been installed, the hole is backfilled with sand around the filter (to allow water to flow freely) and with grout over the remaining depth of the borehole. A dipmeter (an electric gauge with a circuit that closes when it comes in contact with water) is then used to determine the height of the water in the pipe and thus the groundwater pressure at the

11.6 Vibrating wire piezometer. (*Source*: Picture courtesy itmsoil.)

location of the filter. A modification to this design has been proposed that uses a fully grouted tube (Mikkelsen and Green, 2003) and tested with good success (Contreras *et al.*, 2007). The open standpipe piezometer is especially useful for obtaining information about pore water pressures before tunnel construction.

Two limitations of the open standpipe piezometer are that the readings must be taken from above, and that time is required for the water levels to equilibrate, meaning that once tunnelling has started these readings could be inaccurate (British Tunnelling Society and Institution of Civil Engineers, 2004). The remaining three piezometers have all been designed to overcome these limitations (although with a corresponding increase in sensor cost). The vibrating wire piezometer, as seen in Fig. 11.6, uses a vibrating wire strain gauge attached to a diaphragm (which is exposed to the pore water on the other side) to measure changes in water pressure. As the water pressure changes so too does the curvature of the diaphragm and the tension in the attached wire. A relationship between wire tension and water pressure can thus be determined. These sensors can be monitored remotely; however, this technique cannot measure quickly changing water pressures and may require an initial settling period (British Tunnelling Society and Institution of Civil Engineers, 2004). The pneumatic piezometer uses a diaphragm that has air on one side and is exposed to the groundwater on the other. Once the air pressure and water pressure are in equilibrium, a measurement of the air pressure is taken. One of the drawbacks of the system is that it cannot measure negative water pressures. The fibre optic piezometer also uses a diaphragm to measure pressure. In this instance, the distance between the diaphragm and a fixed point is measured using fibre optics. As the water pressure changes, the distance between the diaphragm and the fixed point changes, resulting in a corresponding change in fibre optic reading.

11.3   Case study: Jubilee Line extension, London, UK

In 1993 work began to extend the Jubilee Line of the London Underground east from Green Park station to Stratford (site of the 2012 summer Olympics). As part of this construction, an extensive monitoring campaign was undertaken, as detailed by Burland *et al.* (2001a, b) to observe the response of infrastructure to tunnelling. Burland *et al.* (2001b) present a series of 27 case studies, including both the background to the problem and the results from the monitoring. The monitoring technologies used as part of the Jubilee Line extension included surveying techniques (both precise levelling and total stations), extensometers, crack meters, photogrammetry, and electrolevels. Electrolevels consist of three electrodes submerged in a fluid. As the level is tilted, the amount of fluid covering each electrode changes, as do the corresponding voltage measurements, which allows the tilt of the level to be determined.

One of the instrumented structures in this study was Elizabeth House, consisting of three reinforced concrete tower blocks as seen in Fig. 11.7. Two of the three blocks (two ten-storey structures to the south) share a common raft foundation while a seven-storey structure to the north has a separate raft foundation, as illustrated in Fig. 11.7, and were monitored using a variety of techniques from August 1994 until March 2000 (Burland *et al.*, 2001b). The two running tunnels (5.6 m in diameter) of the Jubilee Line
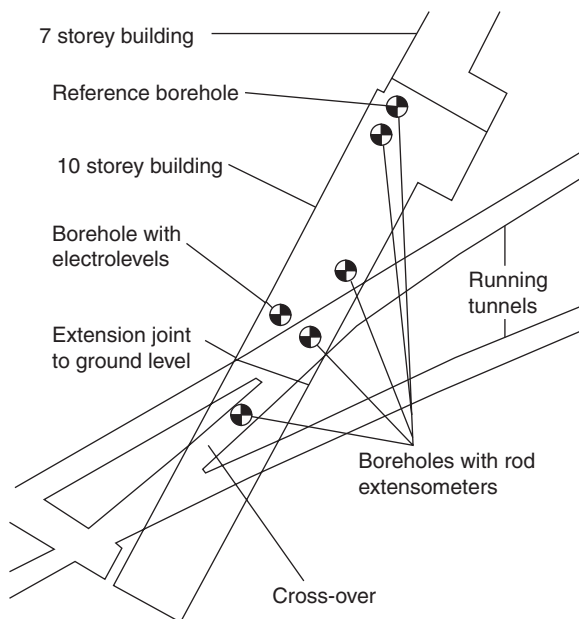
11.7 Elizabeth House layout.

extension as well as a crossover tunnel were constructed underneath the ten-storey block using the New Austrian Tunnelling Method (NATM) with sprayed concrete liners. No special protective measures were applied to protect Elizabeth House, and so the monitoring system was used to assess the impact of the tunnelling on the structure above. The monitoring technologies used to track the movement of Elizabeth House above ground included surveying techniques, crack gauges and photogrammetry. Additionally, six boreholes were drilled through the foundation slab of the ten-storey block to measure vertical and horizontal ground movement. One of the boreholes contained electrolevels, while the other five contained strings of rod extensometers (Fig. 11.7). Interestingly, the rod extensometer data from the borehole farthest from the excavation at the deepest point was used to correct the data from the precise levelling exercise.

The precise levelling and taping were undertaken in the underground parking garage at a level of 7.2 m below the surface (Burland *et al.*, 2001b). Because the entire foundation slab moved during the course of construction, the surveying measurements were all corrected using data from a rod extensometer placed away from the tunnels as mentioned above. The surveying reference point on the foundation slab itself had settled by 11 mm by the end of monitoring in 2000, albeit largely due to underground construction that was not part of the tunnelling project, suggesting that measurements based on this point without correction would have been erroneous. The choice of a baseline for monitoring data is an important consideration and it is not necessarily straightforward as illustrated by this example. A further complication during the monitoring exercise occurred when squatters took over the parking area of the building (Elizabeth House was vacant during the construction) and their dogs made it impossible for surveying data to be acquired for a period until the squatters were evicted. Although this is a rather unusual occurrence, access to the site and the safety of both the personnel and the monitoring equipment is an important consideration when developing a monitoring strategy. Overall, the survey data indicated that the foundation slabs were quite flexible in the longitudinal direction (perpendicular to the tunnels) and stiff in the transverse direction (parallel to the tunnels) but that this flexibility did not cause any damage to the building. The measured settlements were also compared to the predicted settlements calculated using the model proposed by Potts and Addenbrooke (1997) and very good agreement was found, suggesting that the model provides realistic predictions.

To measure potential damage to the facade of Elizabeth House, two techniques were employed: surveying and photogrammetry (Burland *et al.*, 2001b). Both of these techniques required targets to be placed on the facade of the building in order to be able to track the movement of these

points with time. Photogrammetry required that photos were taken from a number of different locations in front of the building and then those photos were used to develop a 3-D model of the building facade. Photos were taken before construction commenced, after the westbound tunnel was constructed and then finally after the eastbound tunnel was constructed. The survey data were used as a reference for the photogrammetry data since the survey data could be tied to a fixed reference point away from the construction. Being able to calibrate the photogrammetry results using the survey data was found to have a significant impact on the accuracy that could be achieved using photogrammetry (Burland *et al.*, 2001b). Using the data from the photogrammetry system, it was determined that the horizontal strains that developed in the building facade after construction were all compressive and not significant enough to cause damage (~250 microstrain).

11.4   Construction monitoring in rock tunnelling

The design of tunnels in rock can be quite different from the design of tunnels in soils. For example, as noted by Kavvadas (2005), while soft ground tunnel designs often aim to minimize surface displacements, rock tunnel designs in mountainous regions often attempt to maximize these displacements so as to minimize the stresses on support structures within the tunnel. However, while design concepts may vary, there are many similarities in terms of the type of monitoring technologies used. As such, this section will discuss the available monitoring technologies and applications but will also refer the reader back to the previous section where sensing technologies overlap. The key areas for monitoring in rock tunnels are: (i) displacements of the surface, the tunnel and surrounding infrastructure; and (ii) support loads and stresses. Typical methods for construction in rock are described in Table 11.2.

11.4.1   Displacements

As stated above, using the conventional tunnelling method (CTM) (also known as the NATM, SCL, and the sequential excavation method (SEM)), large displacements are allowed to occur after tunnel excavation so that the size and strength of the required support can be optimized. In this case, measurements of displacement both at the surface and within the tunnel are key to determining whether the tunnel design is successful and whether modifications to the design are required. However, in urban areas where the density of infrastructure significantly limits the amount of allowable displacement, tunnelling techniques must instead focus on controlling this

Table 11.2 Construction methods for rock tunnels

Technique	Brief description
Drilling and blasting	The tunnel is advanced by drilling holes into the tunnel face into which a carefully controlled amount of explosives are inserted. Once the explosives are detonated, the rubble is removed and the new section of the tunnel is stabilized using one or a combination of support options (e.g. shotcrete, rock bolts, arches etc.). The tunnel can either be blasted to the full height or for bigger tunnels the face may be excavated in stages, a technique known as heading and benching.
Conventional tunnelling (or NATM)	Mechanical excavators (or drilling and blasting – see above) are used to remove material from the tunnel face. The material is then removed from the tunnel (i.e. mucking) and the tunnel is supported using primary support elements such as steel arch supports, sprayed or cast-in place concrete and rock bolts. This approach is concerned with taking advantage of the stress in the rock to minimize the requirement for additional structural support. However, because of the large variability of rock types and stresses within the rock mass, this method is heavily reliant on monitoring to provide the optimum solution.
TBM	This technique is very similar to the use of a TBM in soft ground where material is removed using cutting teeth at the face of the tunnel and transported back through the machine. The type of TBM to be used depends on the requirement for rock support once material has been excavated where some TBMs do not place any supports while others place a concrete liner.

displacement. This is accomplished by controlling the displacement at the tunnel face by installing a stiff temporary lining close to the tunnel face and ensuring that the final lining is placed quickly (Kavvadas, 2005). In either case displacement monitoring is the key to the success of these projects, as the extent of displacement often dictates the tunnel support strategy (Hoek, 2001).

Monitoring techniques

Many displacement monitoring techniques used in rock tunnelling are the same as those used in soft ground tunnelling including: (i) surveying equipment (ii) extensometers (iii) crack gauges (iv) liquid level gauges (v)

fibre optics (vi) inclinometers, and (vii) convergence gauges, all of which are discussed in Section 11.2.1. Two monitoring techniques that are especially useful for rock tunnelling are sliding micrometers and Lidar. Sliding micrometers are discussed in the case study in Section 11.5 while Lidar is discussed below.

A relatively new approach to displacement monitoring is the three dimensional laser scanning technique, or Lidar, which can be used to obtain an accurate picture of the tunnel surface. Fekete *et al.* (2010) used a phase-based Lidar technique that was capable of acquiring approximately 1 million data points a second over distances of up to 80 m. They detail the use of a Lidar system in several tunnels in Norway, including a tunnel that was actively being constructed using the drill and blast construction technique. They suggest that the Lidar system can be set up and operated by one person using a tripod and that readings should be taken at an optimum distance of 0.5–1 tunnel diameter from the face of the excavation. Each measurement stage (including setup and scanning) took less than 7 min. A distance measurement accuracy of within 4 mm was achieved at distances of up to 25 m. The two most significant drawbacks of the one position set up were occlusion (i.e. what cannot be seen at the laser location cannot be measured) and scan bias (i.e. discontinuities parallel to the line of sight). Fekete *et al.* note that the system can be used to estimate shotcrete liner thickness (using pre- and post-installation scans), construction efficiency (amount of material removed for a given process), rock bolt locations, and areas of potential leakage (by measuring changes in the intensity of reflected light).

11.4.2   Excessive loads acting in and on the supports

The loads developed in the supports of rock tunnels are a function of the stresses in the rock prior to excavation, the stiffness of the supports relative to the rock mass, the deflection of the rock before the supports are installed, and the condition of the rock mass including the presence and orientation of any faults. However, to ensure that the support system has been designed correctly, and especially if the observational approach is being used, there is a need for monitoring of the loads and stresses in the support structures.

Monitoring techniques

Loads in rock tunnels can be measured using a variety of techniques including (i) vibrating wire strain gauges, (ii) fibre optic sensors, and (iii) instrumented rock bolts. Vibrating wire strain gauges and fibre optic sensors were discussed in Section 11.2.2. It is once again important to note that if strain

is to be used to estimate stress or load, the modulus and long-term behaviour of the rock mass and/or concrete must be characterized accurately. There are numerous models for the rock mass modulus in the literature as reviewed by Hoek and Diederichs (2006), who also present an empirical model of their own.

The instrumented rock bolt consists of a section of threaded rod with a vibrating wire strain gauge installed in a small hole drilled down the centre of the bar. This section of threaded rod can then be connected to the rest of the bar using couplers, and the rock bolt can be installed as usual. Readings can then be taken of the stress level in the rock bolt.

11.5   Case study: monitoring of the construction of a new tunnel in rock in Switzerland

The following case study will examine some of the monitoring technologies used during the recent construction of a rock tunnel through the Swiss Alps.

11.5.1   Background

The Gotthard Base Tunnel, when in operation in 2015, will be the world's longest tunnel. It consists of twin rail tunnels running through the Swiss Alps. Each tunnel is approximately 57 km and the total network consists of 153 km of tunnels including access shafts and cross-tunnels (Ehrbar, 2008). Boring operations in the eastern tunnel and the western tunnel were completed in October 2010 and in March 2011, respectively. The overburden above the tunnel ranges from 1000 m to more than 2000 m and the rock type varies along the length of the tunnel. The project is divided into five sections: (i) Erstfeld; (ii) Amsteg; (iii) Sedrun; (iv) Faido; and (v) Bodio, as illustrated in Fig. 11.8. There are two 'multi-function stations' in the tunnel located at Sedrun and Faido where trains can cross over to the other tunnel and where a cross-cavern and access tunnel allows for evacuation. Both conventional tunnelling (drill and blast) and tunnel boring machines
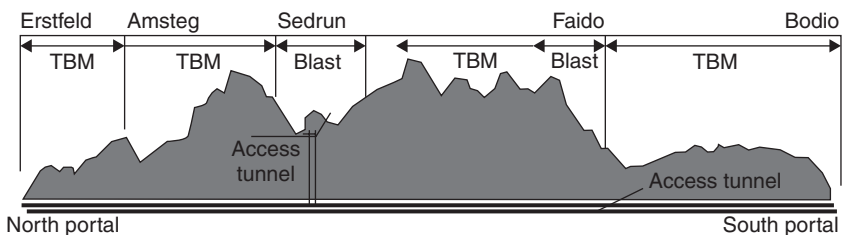
11.8 Tunnel sections of the Gotthard Base Tunnel.

(TBM) have been used depending on the condition of the rock. Ehrbar (2008) noted that conventional tunnelling is more appropriate for use in areas with highly variable rock and, as such, it was used in a 1 km section of tunnel in the Sedrun section where squeezing rock was present with maximum deformations of 750 mm. Conventional tunnelling was also used in the vicinity of the concrete dam at Nalps. Most of the remaining tunnelling was accomplished with TBMs.

11.5.2   Monitoring systems

Because of the extent of the Gotthard Base Tunnel project, a wide variety of monitoring systems were required including: (i) surveying techniques; (ii) fibre optic displacement sensors; (iii) reverse head extensometers; and (iv) seismic monitoring systems.

11.5.3   Surveying techniques

In the Faido section of the tunnel total stations using the Leica Tunnel Measurement System (TMS) were used for both laying out and monitoring of the tunnels (Leica Geosystems AG, 2004). The Faido station section of the tunnel was constructed using the drill and blast technique with heading and benching. A total station with over 300 targets was used to monitor tunnel displacements twice a week, with maximum displacements of 500 mm being measured (Leica Geosystems AG, 2004).

Despite the depth of the tunnel in many areas, surface settlements were expected to be caused by water previously trapped in the rock mass draining into the new tunnels and causing subsidence of the rock (Studer, 2011). These surface settlements had the potential to damage three dams (Curnera, Nalps and Sta. Maria) in the vicinity of the tunnel. As noted by Studer, the potential for these settlements was recognized before construction and so measures were taken to minimize them through systematic and rapid sealing of the tunnel. A three-level monitoring programme was undertaken in the area around the dams, which included the use of total stations, tachymeters, multiple extensometers, and GPS measuring stations to monitor terrain in the area of the dams for movement, as outlined by Studer (2011). The monitoring system was installed to capture the behaviour of the ground surface both before and during the construction of the tunnels in the vicinity of the dams, a monitoring period that was expected to last a minimum of 12 years. The main challenge of installing such a system was to ensure that it was robust enough to remain operational over the 12-year monitoring period. The measurements from each set of sensors were sent to a central computer station for processing and checking by engineers. To minimize the effects of temperature, measurements were taken at night and averaged over time. One of the

challenges with the system was that there were no 'fixed' points, as the entire terrain in the area of the tunnelling moves. Instead, differential movements between sensors were used and a baseline reading was taken in the first year when the tunnels were still several kilometres away. Interestingly, the baseline monitoring showed that the valleys were not static and instead moved seasonally. Once the tunnels approached the area permanent deformations of up to 60 mm were observed; however, in the vicinity of the dams the movement was much smaller, resulting in very low associated risk.

11.5.4   Fibre optic sensors

The south portal of the tunnel located at Bordio was, due to physical constraints, constructed through a layer of loose rock. In order to monitor movements in the exterior concrete buttresses and interior tunnel linings, a fibre optic displacement sensor system was installed (Inaudi and Glisic, 2008). SOFO is a French acronym for 'Surveillance d'Ouvrages par Fibres Optiques', which translates to 'monitoring of structures with fibre optics'. SOFO systems use two fibre optic cables; one is attached to the structure and tensioned between two points, while the other is left free for temperature compensation. Typical gauge lengths for these systems are between 200 mm and 10 m, with resolutions of 2 μm (Inaudi and Glisic, 2008). The system was configured to automatically measure displacements, which could then be correlated to the loading in these structures. The system was installed to monitor conditions during construction; however, the fibre optic cables will be left in place to enable future monitoring, should the need arise.

11.5.5   Reverse head extensometers

In the Sedrun section of the tunnel there was a need to measure the extrusion (axial displacement ahead of the tunnel face) of the tunnel due to the large amount of rock squeezing (Steiner, 2007). As discussed by Hoek (2001), measurements of the extrusion are often crucial in terms of making decisions about deformations and support requirements (both for the face and walls). Although the contract initially called for the use of a sliding micrometer, the extreme pressure conditions in this area of the tunnel meant that this equipment broke down quickly. Instead, a reverse head extensometer, which consists of a data logger and 6–500 mm long displacement transducers, was used (Steiner, 2007). The unit is installed by drilling a borehole into the rock face, inserting the data logger and displacement gauges and grouting the borehole. The data logger is inserted at the end of the borehole farthest away from the tunnel face so that it is not destroyed by the tunnelling operation until all of the displacement transducers have been destroyed by

the advancing tunnelling process. The reverse head extensometer has two advantages over traditional sliding micrometers in terms of data acquisition times and data provided. With a traditional sliding micrometer, readings have to be taken manually at the tunnel face, which disrupts the driving of the tunnel for approximately one and a half hours while readings are taken (Steiner, 2007). On the other hand, Steiner indicates that readings from the reverse head extensometers were taken in approximately 10 min. The sliding micrometer can only provide data at discrete points in time (when construction is stopped), whereas the reverse head extensometer can provide continuous monitoring at data rates as high as 60 Hz. However, the data can be accessed only through either a wireless or wired data connection that requires access to the tunnel face (or more specifically, the data cable of the reverse head extensometer).

11.5.6   Seismic monitoring systems

The multi-function station at Sedrun was initially assumed to be located in very good rock, but during construction it was discovered that a major fault ran through the area (Hagedorn and Stadelmann, 2007). An 8 m high void was created above the tunnel when loose fine grained quartz broke free from the fault during tunnelling. Steel support structures that were put in place to support the surrounding rock deflected by up to 1 m radially, which caused the frames to yield. The stress redistribution due to tunnelling also gave rise to micro-tremors and rock bursts, which caused an increase in the number of events monitored by the pre-existing Swiss Digital Seismic Network (Hagedorn and Stadelmann, 2007). Because of potential effects that these rock burst events could have on the tunnel construction, eight surface seismic monitoring stations and two tunnel stations were added to better localize and monitor the magnitudes of the seismic events. This system was monitored in real-time so that the correct authorities could be notified if action was required. Several events that caused damage to the tunnel (such as invert heave) were directly connected to significant seismic events. The seismic data were used to develop 2-D and 3-D dynamic models to determine the effect the seismic events had on the concrete linings, which had been designed before these events (Hagedorn and Stadelmann, 2007).

11.6   In-service and long-term monitoring

Monitoring of tunnel construction often focuses on the construction phase, but there is a need for in-service and long-term monitoring for a variety of reasons. This section will examine three reasons for ongoing monitoring of tunnels and surrounding areas: (i) changing loads and conditions; (ii)

deterioration; and (iii) construction of surrounding infrastructure. Although environmental monitoring is an important aspect of in-service tunnel operation and maintenance, it will not be discussed here as it goes beyond the scope of this chapter.

11.6.1　Changing loads and conditions

Over time there is the potential for ground and loading conditions to change, which in turn impacts settlements and loads within tunnel linings. As discussed below, changes in pore water pressure in soils and unstable regions in rock masses can lead to changes in deformations and loading.

Monitoring techniques

When it comes to monitoring settlements and stresses due to changing ground and loading conditions, the technologies employed are the same as they would be for initial construction: (i) surveying equipment; (ii) piezometers; (iii) pressure cells; (iv) fibre optics; and (v) extensometers. These monitoring techniques have all been presented before; however, it is worth introducing two applications of long-term monitoring in soft ground and rock below.

Wongsaroj *et al.* (2007) give an example of a project where monitoring was performed to determine the long-term effects of tunnelling in soft ground. The monitoring system (initially installed as part of the London Underground Jubilee Line extension project discussed in Section 11.3) consisted of shallow surface monitoring points, inclinometers, rod extensometers, pneumatic piezometers and combined pneumatic piezometers and earth pressure cells (Standing and Burland, 2006). The monitoring data indicated that over time the area directly above the tunnel swelled, whereas large regions on either side of the tunnel consolidated while soil towards the surface settled as a rigid body. These settlements occur because the tunnel acts as a drain for the surrounding soil, which then changes the pore water pressure. By extensively monitoring the area, the researchers were able to develop a model to predict the extent of the settlements as well as the drainage mechanism. It is also clear from this research that long-term settlements are not necessarily negligible and may need to be monitored in areas where settlement sensitive infrastructure is present. As noted previously, a change in pore water pressure also has the potential to change the loading on the tunnel lining (Shin *et al.*, 2005) as well as long-term ground deformation (Wongsaroj, 2006; Wongsaroj *et al.*, 2007; Laver, 2009; Laver and Soga, 2011).

Long-term effects are not restricted to tunnels in soils. Kontogiammi *et al.* (2008) discuss the case of the Messochora tunnel failure, where 4 weeks after excavation a delayed failure of the tunnel in rock occurred. Analysis of the monitoring data (geodetic techniques and rock extensometers) pointed

to the fact that the support system was at a critical equilibrium, which was exceeded by minor perturbations caused by work in other areas of the tunnels. The monitoring also allowed the researchers to detect a redistribution of stress to surrounding parts of the tunnel after the failure. Unfortunately, this failure was determined only later as the monitoring data were not being actively analysed. Konotogiammi *et al.* recommend that ongoing monitoring be employed in tunnels where weak rocks may continue to displace if triggered by slight changes in stresses.

11.6.2   Deterioration

Deterioration monitoring is a significant issue for most infrastructure assets that are coming to the end of their design life, and nowhere is this truer than for tunnels. Assets such as bridges are expensive to replace, and so keeping them in service for as long as possible is prudent. However, although typical design lives for tunnels are specified between 60 and 150 years (British Tunnelling Society and Institution of Civil Engineers. 2004), many tunnels are almost impossible to replace, due to not only the cost but also the disruption involved. Thus, methods of detecting deterioration are required so that action can be taken to keep tunnels in service. Different types of deterioration can develop within a tunnel, and for each type of deterioration there are a number of causes. Deterioration mechanisms vary depending on the material used to support the tunnel (e.g. steel, cast iron, concrete, masonry or timber). The Federal Highway Association (2005) has produced a guide for the inspection of tunnels that provides an extensive list of deterioration mechanisms. The guide lists deterioration mechanisms based on supporting material (e.g. deterioration mechanisms for concrete include scaling, cracking, spalling and staining) and provides guidance for measuring the extent and severity of the deterioration. A number of challenges exist when it comes to monitoring deterioration, not the least of which is determining its cause. A second challenge is determining whether the deterioration, once spotted, is the result of a single event (such as a construction defect) or an ongoing problem. One of the key elements of most infrastructure management strategies is the use of visual inspection. In this case, structures are examined visually by inspectors periodically to detect signs of deterioration such as cracking, spalling and discoloration. The use of sensors can help to provide additional information as discussed below; however, sensors should be used to supplement visual inspection rather than to replace it.

Monitoring techniques

As noted above, deterioration mechanisms are quite varied and can be difficult to detect if there are no visual signs of distress. However, a number

of approaches to deterioration monitoring have been used in the past and are currently under development including: (i) surveying techniques; (ii) the use of conventional sensors; (iii) fibre optics; (iv) inspection vehicles; (v) WSNs; (vi) acoustic emission; and (vii) video and image analysis. The first three technologies have been discussed in other sections and so this section will focus on the remaining four techniques.

Inspection vehicles have existed for decades; however, recent advances in non-destructive testing techniques have meant that these vehicles are now potentially capable of taking more accurate measurements faster. Idoux (2005) gives details of a vehicle first used to inspect the Paris Metro that is equipped with a laser scanner, an infrared scanner, a telemeter assembly and an odometer. The laser scanner enables the location of 0.2 mm wide cracks from 6 m away while the infrared scanner measures differences in surface temperature down to 0.02°C allowing subsurface defects to be detected.

WSNs are composed of nodes and gateways. The nodes are typically battery powered and have one or more sensors attached to them as well as a radio transmitter. The gateways are similar to data loggers and communicate wirelessly with the nodes to acquire the sensor data. An example of the use of a WSN to monitor a tunnel is given in the case study in Section 11.7.

Acoustic emission monitoring is the use of sensors to measure the audio/vibration signal given off as a material breaks down. The technique can be used to identify certain tunnel deterioration mechanisms, such as cracking of concrete and fracturing of rock. One of the major challenges with this technique is to develop a relationship between acoustic emission events and material deterioration, which typically requires calibration. Acoustic emission (AE) sensors are normally hardwired and powered using a continuous power source because the sensors often have to be on all the time. However, one of the recent developments in this area was the creation of WSN-based acoustic emission systems (Grosse *et al.*, 2008).

Digital image analysis has seen increasing use in deterioration detection over the last decade. The London Underground, for example, has trialled systems using digital cameras mounted to helium-filled balloons to inspect shafts (Wright, 2010). Soga *et al.* (2010) present a method of creating image 'mosaics' of tunnel images that allow curved 3-D surfaces to be accurately represented in 2-D (without warping effects). This technique will allow tunnel managers to store digital images from tunnel inspections in a database that is searchable based on location rather than image number. Techniques have also been developed by researchers to detect deterioration in images, and a review of these techniques can be found in Jahanshahi *et al.* (2009).

11.6.3   Construction of surrounding infrastructure

While emphasis earlier in the chapter was placed on the need to monitor surrounding infrastructure at the same time as constructing a tunnel, another important consideration is the monitoring of tunnels during construction of other infrastructure, especially in densely populated areas. This can be an issue in cities such as New York and London, where a dense web of underground infrastructure exists and sections of it are more than a century old.

   This is also important for newer infrastructure, as illustrated by the case of the Taipei Rapid Transit System (Chang *et al.*, 2001). In this case, a transit tunnel had been completed for less than a year when a deep foundation was started alongside the twin 5.6 m diameter tunnels. The monitoring scheme involved the use of inclinometers (installed both within the tunnels as well as on the diaphragm wall of the adjacent construction), survey points and crack gauges. The researchers used the monitoring data to determine that the excavation caused significant damage to the tunnel. Interestingly, they found that inclinometers were not the right choice for measuring lining displacement. This is a classic mistake in monitoring, whereby a sensor is chosen because it is available or commonly used but does not provide the data required. Fortunately, the researchers had redundant data from other sensors that allowed them to correctly interpret the data from the inclinometers.

11.7   Case study: monitoring of an existing tunnel for deterioration in London, UK

The following case study introduces a number of newer technologies that can be used for existing tunnels to monitor the impact of deterioration on tunnel performance and behaviour.

11.7.1   Background

The London Underground network consists of approximately 181 km of tunnels, some of which date back to the nineteenth century. One of the major ongoing concerns of London Underground Ltd. (LUL) is to ensure that the tunnel geometry is not compromised, which could in turn affect the running of trains. Tunnel geometry could be affected by deterioration of the tunnel lining or by a change in the loading on the tunnel lining due to new construction. In either case monitoring is often required to ensure that potential changes in geometry do not adversely affect the safe operation of the Underground network.

(a)

Inner diameter : 3810 mm
Outer diameter : 4146 mm

Concrete segments

Wedge →
segment

(b)

Localized
spalling

Fibre optic
sensor cables
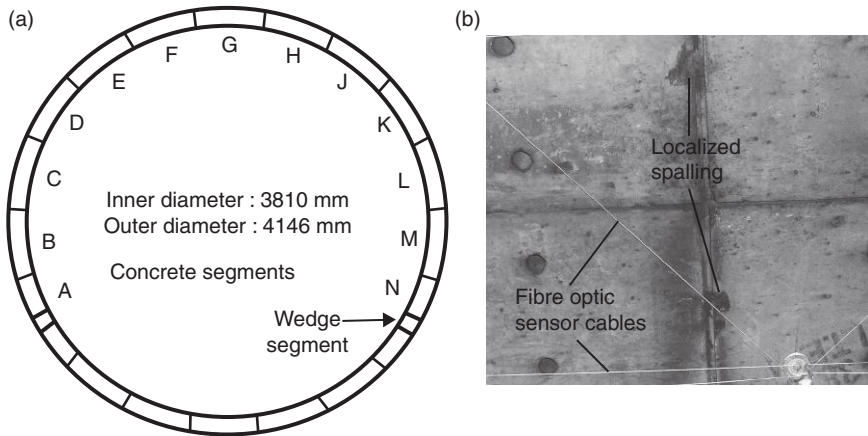
11.9 (a) Tunnel geometry and (b) localized deterioration.

The following case study examines a monitoring project undertaken on the Jubilee Line in a section of expanded concrete tunnel between Baker Street and Bond Street stations. The tunnel is lined with precast concrete panels in rings of 20 panels with wedge segments, as seen in Fig. 11.9a. This section of tunnel has localized sections of concrete spalling and cracking in the panels, as illustrated in Fig. 11.9b, relating back to issues with its initial construction in the 1970s (Wright, 2010). When the tunnels were initially constructed using Greathead shield tunnelling methods, a lack of strength in the Lambeth Group material in this section led to a loss of material in the shoulders of the tunnel, which was later grouted (Lyons, 1979). It is this material loss and subsequent grouting that is believed to have caused the ovalling of the tunnels (Lyons, 1979) that has since resulted in the observed spalling and cracking.

To determine whether the observed deterioration is worsening with time, Tubelines, who manages this section of the Underground for LUL as part of a public-private partnership, decided to install a monitoring system. Three different monitoring systems were used within the same section of tunnel, including a vibrating wire strain gauge system (Wright, 2010; Bennett *et al.* 2011), a fibre optic strain sensor system (Cheung *et al.*, 2010), and a WSN (Bennett *et al.*, 2010). The three systems have advantages and disadvantages in terms of the data they provide as well as their cost, reliability, and installation time. Each system will be presented briefly and then compared.

11.7.2   Monitoring systems

The vibrating wire strain gauge system was the only commercially installed one of the three systems. It consisted of a series of vibrating wire strain
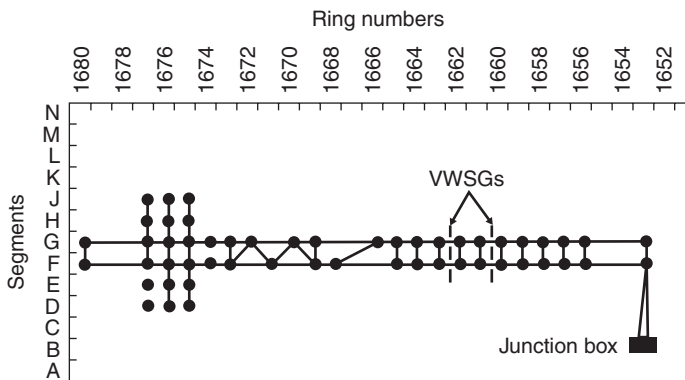
11.10 Optical fibre sensor layout. (*Source*: From Cheung *et al.*, 2010.)

gauges installed across the joints between the panels (Bennett *et al.*, 2010). Each strain gauge was wired into a data logger installed at the top of a vent shaft. The data from the vibrating wire strain gauges allowed for a comparison between these conventional sensors and the fibre optic strain sensing system as well as the WSN.

The fibre optic strain sensing system used the BOTDR technology (Cheung *et al.*, 2010). BOTDR allows distributed strain measurements with an accuracy of approximately 50 με to be taken along the full length of a fibre optic cable with a gauge length of approximately 1 m. The fibre optic system was installed over the course of 12 days. The analyser for this system (a N8510 BOTDR analyser by Advantest Corporation) was placed at the top of a nearby vent shaft (the same vent shaft that housed the data logger for the vibrating wire strain gauges). As a result of the distance between the analyser and the area of the tunnel being monitored, two types of fibre optic cable were used. In the section between the analyser and the region of interest, general-purpose loose tube cable was used as this type of cable is more robust. In the area where measurements were taken, which is illustrated in Fig. 11.10, 900 μm fibre containing one single-mode optical fibre to allow for accurate strain sensing was used. As indicated in Fig. 11.10, there were two locations in the optical fibre layout where the optical fibres and the vibrating wire strain gauges overlapped, which allowed for a direct comparison of the two systems. This system was monitored between February and June, 2008, over which time nine datasets were taken.

The WSN consisted of 26 nodes (which used the Crossbow MicaZ platform) installed on five separate precast concrete tunnel rings along the length of the tunnel as illustrated in Fig. 11.11. There were three different types of node: (i) inclinometer nodes; (ii) linear potentiometric displacement transducer (LPDT) nodes; and (iii) relay nodes installed at the locations illustrated
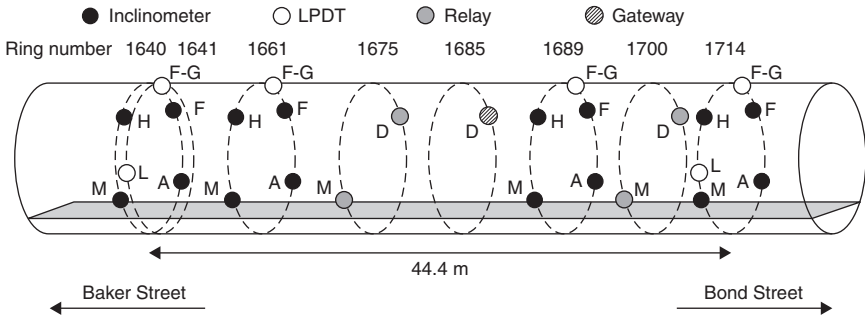
11.11 WSN layout. (*Source*: From Bennett *et al.*, 2010.)

in Fig. 11.11. The inclinometer nodes had both MEMS inclinometer sensors and MEMS-based relative humidity and temperature sensors. The LPDT nodes were placed over the joints between the precast concrete panels to measure joint movement (several LPDTs were installed at the same location as the vibrating wire strain gauges to allow for a direct comparison of the data from the two sensors). The gateway, part of which was placed on ring number 1685 as per Fig. 11.11, was a two-part system with the WSN radio on the gateway in the tunnel connected to a computer at the top of a vent shaft via Ethernet cable. Having the computer for the gateway at the top of the vent shaft meant that it could be accessed during the day (whereas the rest of the network could be accessed only during engineering hours) and that it could be attached to a wireless modem resulting in real-time access to the data.

11.7.3   Data comparison

The data from the vibrating wire strain gauges are compared to the data from the BOTDR system in Fig. 11.12. The displacements measured by two vibrating wire strain gauges attached at the crown of precast concrete panel rings number 1660 and 1662 are plotted against the displacements calculated from the strains measured by the BOTDR system at the same location. Four different results are given for the BOTDR system designated as $w = 0.30$ through 0.450. Unfortunately, in order to be able to measure the change in joint opening, the fibre optic cable had to span between two fixed points at a distance of 0.6 m apart, which is less than the measurement length of the BOTDR system of 1m. This can be corrected mathematically (Klar *et al.*, 2006); however, part of this correction involves calibration of the parameter, w, which varies with the level of applied strain. Thus, the measured displacement for a range of values of w is plotted.

   It can be seen that the results from both systems are in generally good agreement, in that both systems capture the fact that the distance between
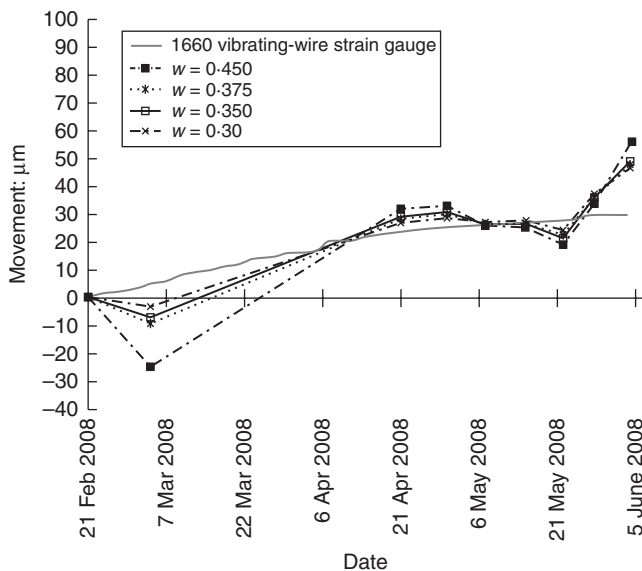
11.12 Comparison of vibrating wire strain gauge and BOTDR system on ring 1660. (*Source*: From Cheung *et al.*, 2010.)
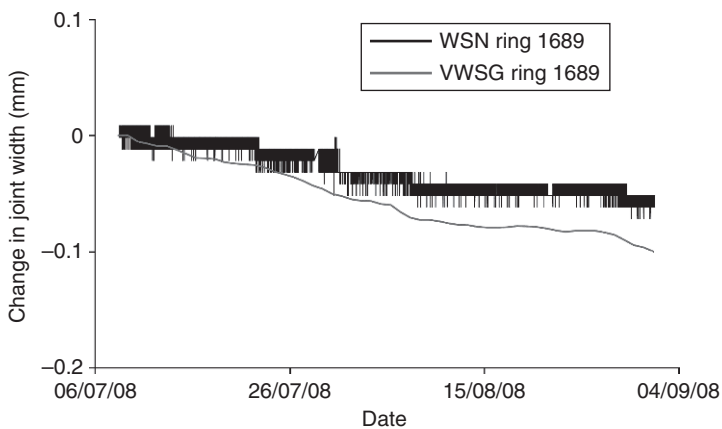
11.13 Comparison of vibrating wire strain gauge and WSN system. (*Source*: From Bennett *et al.* 2010.)

the panels increases over time (20–50 microns in 4 months). However, one can also see the importance of having an accurate correction factor if gauge lengths of less than 1 m are to be used with the BOTDR system, as there is a factor of two differences between the displacement calculated using $w = 0.30$ versus $w = 0.450$. It should also be noted that temperature compensation

is also important for the BOTDR system as the measured strain changes by 23.8 με/°C (Mohamad, 2008), which is equivalent to 14.3 μm/°C for this system. The results from one of the vibrating wire strain gauges are compared with the results from the LPDT WSN node at the same location in Fig. 11.13.

One can see that both systems capture the general behaviour of the joint correctly (50–90 microns of movement in 2 months) although there is a difference in the magnitude of the measurement. This disparity was attributed to the difference in sensor locations relative to the movement of the precast panels. It is believed that the precast panels rotate about both their cross-sectional and longitudinal axes. Thus, unless multiple sensors were placed on the same panel or the sensors were co-located, there will be a difference in the measured displacement. This situation points to a need to develop monitoring systems that are both inexpensive and can be installed quickly so that complex changes in tunnel geometry can be captured accurately.

11.7.4   Overall comparison

Overall, the vibrating wire strain gauges proved to be the most robust monitoring system as it suffered the fewest failures and data outages. This is perhaps to be expected as it was a commercial installation using a proven technology as opposed to the other two systems which were research prototypes. The major limitations of the vibrating wire strain gauge system was the length of time required for installation as well as the limited amount of data provided by the discrete point sensors. The BOTDR technology can provide much more data as readings can be taken along the full length of the fibre optic cable. However, the system suffered from robustness issues as the cable was broken and had to be repaired on two occasions (Chueng *et al*, 2010). Additionally, if measurements are required over gauge lengths shorter than a metre or with resolutions lower than 50 με, the system may not be able to meet those requirements. However, fibre optic strain analysers with better accuracy and smaller gauge lengths have become available in recent years. The WSN offered faster installation times (the network was installed in 10 days while it was estimated by Bennett *et al*. (2010) that an equivalent wired system would have taken 14 days to install. However, the network also suffered from robustness issues as several of the inclinometer nodes failed and had to be replaced. Further improvements on innovative monitoring technologies (such as robust hardware, installation techniques) are needed to build confidence in using such systems for commercial use. When choosing a monitoring system it is important to remember the advice of Dunnicliff (1993) who suggested that the most appropriate monitoring system should be the most robust product available, while at the same

time balancing this against the need for enough data to validate complex models.

11.8   Sensing technology summary

Table 11.3 is a summary of the sensing technologies that have been discussed in this chapter with information about applications, approximate accuracy and approximate cost. It should be noted that both the accuracy and cost columns are highly dependent on the choice of sensor and the application. The accuracy column is intended to give a sense of the best accuracy that can be achieved with that sensor. The cost column is intended to give the reader a sense of the relative cost of sensors and is not intended as a definitive guide as sensor costs can vary significantly depending on the quality of the sensor. It should also be noted that the prices are based on standard products excluding accessories and associated components and do not reflect the actual price that may be quoted when ordering. Additionally, the costs do not include installation, which can have a significant impact especially for borehole sensors.

11.9   Future trends

As with most infrastructure monitoring applications, the trend with tunnel monitoring is towards systems that are capable of delivering large quantities of data in real or near real-time. Systems such as Cyclops for continuous displacement monitoring or distributed fibre optic sensing systems or Lidar provide the engineer, owner or facility manager with access to thousands of data points. This provides the opportunity to perform complex analysis and verify it against a rich data set. The result of this could be more refined designs or better optimized tunnel management than was previously possible. However, there is also a very real risk of reaching a point of data saturation, where there is too much information to manage or even comprehend, at which point there is a risk of the information provided by these monitoring systems being ignored. As such, research and development must not only be focused on developing sensors that can deliver more data, but also on developing methods of presenting this data to decision makers in a useful form. One potential way of doing this is through integration with a geographic information system (GIS). This approach is especially useful for tunnels, because of the potential size of the monitored area. However, it is also a challenge to implement with tunnels since locating the sensor data in space can be a difficulty. For example, a WSN may have many inclinometer nodes installed by a technician. If the technician did not take careful notes during installation, the location of individual nodes may not be clear, making the data from that node useless.

Thus physically locating the source of the data from new sensors will be critical to allow interpretation of results, but recording those location will also be critical.

A second challenge with the development of new monitoring technologies for tunnels will be ensuring that they are robust enough to survive in the field. Dunnicliff (1993) noted the importance of using monitoring technology that is both robust and the simplest solution to a problem. Although new technologies often do not meet the requirement for simplicity, they must meet the robustness requirement. Thus, work needs to be done to ensure that technologies such as distributed fibre optic sensors and WSNs are robust enough for field deployment.

11.10   Sources of further information and advice

There are several places that the interested reader can turn for further information about geotechnical monitoring or specific products. The classic text is 'Geotechnical instrumentation for monitoring field performance' by John Dunnicliff (Dunnicliff, 1993). This text not only covers monitoring technologies but also considerations related to monitoring, such as how to design a monitoring system and who should be responsible for the system. The text has not been updated since 1993; however, Dunnicliff has created a newsletter on monitoring called 'Geotechnical Instrumentation News' or GIN where developments in the industry since 1993 are discussed. GIN can be accessed online through http://www.bitech.ca/instrumentation_news.php.

There are also guidelines published by professional organizations such as the British Tunnelling Society and Institution of Civil Engineers (2004), whose 'Tunnel lining design guide' contains a section on monitoring. The Federal Highway Administration (FHWA) also includes a section on monitoring in their 'Technical Manual for Design and Construction of Road Tunnels – Civil Elements' (Hung *et al*., 2009). Finally, the International Tunnelling and Underground Space Association has developed a guidance document entitled 'Monitoring and Control in Tunnel Construction' (ITA, 2011) that covers many of the aspects of designing and interpreting the results of a tunnel monitoring system.

Additional sources of information include monitoring company and equipment manufacturer websites. Most of the technologies mentioned in this chapter have freely available data sheets and example cases that can be accessed through a web search.

11.11   Acknowledgements

The authors thank Paras Patel of ITMSoil for providing Figs 11.3 to 11.6 as well as pricing information for some of the sensors in Table 11.3, and Rory

Table 11.3 Available sensors for tunnel monitoring

Sensor	Measurement application	Estimated accuracy	Approximate cost[a] (GBP/USD)
Borehole extensometer	Movement of one point in the soil/rock/tunnel relative to another point	~0.02 mm	£1000/$1600
Convergence gauges	Change in dimension (e.g. ovalling of a tunnel)	~0.02 mm	£1800/$2600
Crack gauges	Relative movement across a crack	Manual: ~0.5 mm Electronic: ~0.01 mm	£10/$16 £250/$400
DEMEC gauges	Movement across a crack	~0.01 mm	£400/$650
Electro-level	Differential settlements/inclination	~0.1 mm/m	£200/$320
Fibre optic piezometer	Water pressure	~1 kPa	£600/$1000
Fibre optics	Distributed strain and displacement	BOTDR: ~50 microstrain	£50k/$80k (analyser) Cable cost varies
Fixed extensometer	Displacements relative to a fixed point including tunnel walls relative to the surrounding soil, convergence and surface points	~0.02 mm	£1900/$3000
Image analysis	Non-contact measurement of crack widths, relative displacements and strain fields	~ 0.01 pixels (physical disp. requires a scale factor)	Varies depending on system
Inclinometers	Change in inclination/rotation of an object	~ 10 arc s	£4000/$6500
Instrumented rock bolts	Strain (and stress) in anchors	~1 microstrain	£450/$700
Liquid level	Differential settlements/inclination	~0.3 mm/m	£1100/$1750
Multipoint extensometer	Probe, rod and borehole extensometers that have multiple measurement locations	Varies	£2000/$3200
Open standpipe piezometer	Water pressure	~10 mm	£50/$80
Pneumatic piezometer	Water pressure	~1 kPa	£100/$160

(*Continued*)

Table 11.3 Continued

Sensor	Measurement application	Estimated accuracy	Approximate cost[a] (GBP/USD)
Pressure cell	Soil pressure – horizontal pressure for spade type pressure cell	~0.1 kPa (depending on total range)	£400/$650
Probe extensometer	Relative displacement of points in the soil	~2 mm	£350/$550
Reverse head extensometer	Extrusion of the tunnel face/relative displacement	~0.5 mm (depending on total range)	Not available from all sensor suppliers
Rod extensometer	Displacements relative to a fixed point	~0.02 mm	£1750/$2800
Sliding micrometer	Extrusion of the tunnel face/relative displacement	~0.002 mm/m	£500/$800
Total Station and targets	Used to measure the displacement and rotation of tunnel linings and surface infrastructure.	Distance Reflectors: ~1 mm Reflectorless: ~2 mm Angles ~1 arc s	Total station: £2.5 and up $4k and up
Vibrating wire strain gauge	Strain measurement in structural members (vibrating wire strain gauges are used in a variety of applications)	~1 microstrain	£80/$130
Vibrating wire piezometer	Water pressure	~0.1 kPa	£350/$550
WSNs	Various including: strain, displacement, inclination, temperature	Varies	£500/$800 (per node)

[a]The prices are based on standard products excluding accessories and associated components and do not reflect the actual price that may be quoted when ordering. Additionally the costs do not include installation.

O'Rourke and Chris Fielding of DATUM for providing pricing information for Table 11.3.

11.12 References

Acerbis, R., Asche, H., Barbieri, G. and Collotta, T. (2011). 'Recommendations for converting strain measured in concrete to stress'. *Geotechnical Instrumentation News*, **65**: 29–33.

Batten, M. and Powrie, W. (2000). 'Measurement and analysis of temporary prop loads at Canary Wharf underground station, east London'. *Proceedings of ICE Geotechnical Engineering*, **143**(3), 151–163.

Bennett, P.J., Kobayashi, Y., Soga, K. and Wright, P. (2010). 'Wireless sensor network for monitoring transport tunnels'. *Proceedings of the Institution of Civil Engineers Geotechnical Engineering*, **163**(GE3), 147–156.

British Tunnelling Society and Institution of Civil Engineers. (2004). *Tunnel Lining Design Guide*. Thomas Telford, London: 195 p.

Burland, J.B., Standing, J.R. and Jardine, F.M. (2001a). *Building Response to Tunnelling – Volume 1: Projects and Methods*. Thomas Telford, London: 344 p.

Burland, J.B., Standing, J.R. and Jardine, F.M. (2001b). *Building Response to Tunnelling – Volume 2: Case Studies*. Thomas Telford, London: 881 p.

Chang, C-T., Chieh-Wen Sun, C-W., Duann, S.W. and Hwang, R.N. (2001). 'Response of a Taipei rapid transit system (TRTS) tunnel to adjacent excavation'. *Tunnelling and Underground Space Technology*, **16**(3): 151–158. DOI: 10.1016/ S0886-7798(01)00049-9.

Cheung, L.L.K., Soga, K. Bennett, P.J., Kobayashi, Y., Amatya, B. and Wright, P. (2010). 'Optical fibre strain measurement for tunnel lining monitoring'. *Proceedings of the Institution of Civil Engineers Geotechnical Engineering*, **163**(GE3), 119–130.

Contreras, I.A., Grosser, A.T. and VerStrate, R.H. (2007). 'The use of the fully-grouted method for piezometer installation'. *Proceedings of the Seventh International Symposium on Field Measurements in Geomechanics. FMGM, 2007*. Boston, MA. ASCE Geotechnical Special Publication 175, 20 p.

Dunnicliff, J. (1993). *Geotechnical Instrumentation for Monitoring Field Performance*. John Wiley & Sons, New York.

Ehrbar, H. (2008). 'Gotthard base tunnel, Switzerland experiences with different tunnelling methods'. *Proceedings of the 2nd Brazilian Congress of Tunnels and Underground Structures*, São Paulo, Seminário Internacional 'South American Tunneling', 14 p.

Federal Highway Administration. (2005). *Highway and Rail Transit Tunnel Inspection Manual*. Federal Highway Administration, Washington.

Fekete, S., Diederichs, M. and Lato, M. (2010). 'Geotechnical and operational applications for 3-dimensional laser scanning in drill and blast tunnels'. *Tunnelling and Underground Space Technology* **25**(5): 614–628. DOI: 10.1016/j. tust.2010.04.008.

Grosse, C., McLaskey, G., Bachmaier, S., Glaser, S.D. and Krüger, M. (2008). 'A hybrid wireless sensor network for acoustic emission testing in SHM'. *Proceedings of the SPIE – The International Society for Optical Engineering*, **6932**: 693238-1-9.

Hagedorn, H and Stadelmann, R. (2007). 'Gotthard base tunnel rock burst phenomena in a fault zone, measuring and modelling results'. *Proceedings of the World Tunnel Congress 2008 – Underground Facilities for Better Environment and Safety*, India: 419–430.

Hashimoto, T., Ye, G.L., Nagaya, J., Konda, T. and Ma, X.F. (2009). 'Study on earth pressure acting upon shield tunnel lining in clayey and sandy grounds based on field monitoring'. *The 6th International Symposium on Geotechnical Aspects of Underground Construction in Soft Ground*, Shanghai, Taylor and Francis Group, London: 307–312.

Hoek, E. (2001). 'Big tunnels in bad rock'. *Journal of Geotechnical and Geoenvironmental Engineering*, **127**(9): 726–740.

Hoek, E. and Diederichs, M.S. (2006). 'Empirical estimation of rock mass modulus'. *International Journal of Rock Mechanics & Mining Sciences*, **43**(2): 203–215. DOI: 10.1016/j.ijrmms.2005.06.005.

Hung, C.J., Monsees, J., Munfah, N. and Wisniewski, J. (2009). *Technical Manual for Design and Construction of Road Tunnels – Civil Elements*. Federal Highway Administration, Washington, 694 p.

Idoux, M. (2005). 'Multisensor system for tunnel inspection'. *Proceedings of the SPIE – The International Society for Optical Engineering*, **5640**(1): 303–311

Inaudi, D. and Glisic, B. (2008). 'Overview of fibre optic sensing applications to structural health monitoring'. *Proceedings of the 13th FIG symposium on Deformation Measurement and Analysis and the 4th IAG Symposium on Geodesy for Geotechnical and Structural Engineering*, Lisbon, 10 p.

ITA. (2011). *Monitoring and Control in Tunnel Construction*. International Tunnelling and Underground Space Association, Lausanne.

Jahanshahia, M.R., Kelly, J.S., Masria, S.F. and Sukhatme, G.S. (2009). 'A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures'. *Structure and Infrastructure Engineering*, **5**(6): 455–486. DOI: 10.1080/15732470801945930.

Kavvadas, M.J. (2005). 'Monitoring ground deformation in tunnelling: Current practice in transportation tunnels'. *Engineering Geology*, **79**(1–2), 93–113. DOI: 10.1016/j.enggeo.2004.10.011.

Klar, A., Bennett, P.J., Soga, K., Mair, R.J., Tester, P., Fernie, R., St. John, H. and Torp-Peterson, G. (2006). 'Distributed strain measurement for pile foundations'. *Proceedings of the Institution of Civil Engineers, Geotechnical Engineering*, **159**(3): 135–144.

Kontogianni, V., Papantonopoulos, C. and Stiros, S. (2008). 'Delayed failure at the Messochora tunnel, Greece'. *Tunnelling and Underground Space Technology*, **23**(3):232–240. DOI: 10.1016/j.tust.2007.04.005.

Laver, R.G. (2011). *Long-Term Behaviour of Twin Tunnels in London Clay*. Ph.D. Thesis, University of Cambridge, Cambridge.

Laver, R.G. and Soga, K. (2011). 'Numerical simulation of long-term twin-tunnel behaviour at St James's Park', *Proceedings of the 7th International Symposium on Geotechnical Aspects of Underground Construction in Soft Ground (IS-ROMA 2011)*.

Leica Geosystems AG. (2004). 'Gotthard Base Tunnel: Tunnel technology for the future'. Reporter: The Magazine of Leica Geosystems, 50.

Lyons, A.C. (1979). 'The jubilee line 2. Construction from Baker street to Bond street exclusive and from Admiralty Arch to Aldwych'. *Proceedings of the Institution of Civil Engineers*, **66**(3): 375–394.

Machan, G. and Bennett, V.G. (2008). *Use of Inclinometers for Geotechnical Instrumentation on Transportation Projects – State of the Practice*. Transportation Research Board, Washington, 79 p.

Mair, R.J. (1996). 'General report on settlement effects of bored tunnels'. In *Geotechnical Aspects of Underground Construction in Soft Ground* (eds R. J. Mair and R. N. Taylor), Balkema, Rotterdam, pp. 43–53.

Mair, R.J. and Taylor R.N. (1997). 'Bored tunnelling in the urban environment: State-of-the-art report and theme lecture'. *Proceedings of the 14th International Conference on Soil Mechanics and Foundation Engineering*, Hamburg 4, 2353–2385.

Mair, R.J. (2008). 'Tunnelling and geotechnics: new horizons'. *Géotechnique*, **58**(9), 695–736.

Mikkelsen, P.E. and Green, E.G. (2003). 'Piezometers in fully grouted boreholes'. *Symposium on Field Measurements in Geomechanics, FMGM 2003*, Oslo, 10 p.

Mohamad, H. (2008). *Distributed Optical Fibre Strain Sensing of Geotechnical Structures*. PhD thesis, University of Cambridge.

Mohamad, H., Bennett, P.J., Soga, K. Mair, R.J. and Bowers, K. (2010). 'Behaviour of an old masonry tunnel due to tunnelling induced ground settlement'. *Géotechnique*, **60**(12), 927–938.

Mohamad, H., Soga, K., Bennett, P.J., Mair, R.J. and Lim, C.S. (2012). 'Monitoring twin tunnel interactions using distributed optical fiber strain measurements' *Journal of Geotechnical and Geoenvironmental Engineering*. **138**(8), 957–967.

Nicholson, D., Tse, C., Penny, C., O'Hana, S. and Dimmock, R. (1999). *The Observational Method in Ground Engineering: Principles and Applications*. CIRIA, London, Report 185.

Parks, P., Francis, J., Gorlov, A., Gorlova, E. and Guertin, J.D., Jr. (1986). *Water Intrusion Problems in Transit Tunnels*. Final Report (UMTA-MA-06-0156–86-1), to US Department of Transportation, UMTA Technical Assistance Program.

Peck, R.B. (1969). 'Advantages and limitations of the observational method in applied soil mechanics'. *Géotechnique*, **19**(2), 171–187.

Potts, D.M. and Addenbrooke, T.I. (1997). 'A structure's influence on tunnelling induced ground movements'. *Proceedings of the Institution of Civil Engineers, Geotechnical Engineering*, **125**(2): 109–125.

Richards, D.J., Clark, J., Powrie, W. and Heymann, G. (2007). 'Performance of push-in pressure cells in overconsolidated clay'. *Proceedings of the Institution of Civil Engineers Geotechnical Engineering*, **160**(1), 31–41.

Ryley, M.D. and Carder, D.R. (1995). 'The performance of push-in spade cells installed in stiff clay'. *Géotechnique*, **45**(3), 533–539.

Shin, J.H., Potts, D.M. and Zdravkovic, L. (2005). 'The effect of pore-water pressure on NATM tunnel linings in decomposed granite soil'. *Canadian Geotechnical Journal*, **42**(6):1585–1599.

Standing, J.R. and Burland, J.B. (2006). 'Unexpected tunnelling volume losses in the Westminster area, London'. *Géotechnique*, **56**(1), 11–26.

Steiner, P.R. (2007). 'Displacement measurements ahead of a tunnel face using the RH Extensometer'. ASCE Geotechnical Special Publication, No. 175: *Proceedings of the Seventh International Symposium on Field Measurements in Geomechanics*, 8 p.

Studer, M. (2011). 'Automatic Monitoring of large dams in the Swiss Alps during construction of the Gotthard base tunnel (57 km)'. *Proceedings of the International*

Symposium on Dams and Reservoirs under Changing Challenges – 79th Annual Meeting of the International Committee on Large Dams, Lucerne, June 2011, 6 pp.

Take W.A., White D.J., Bowers K.H. and Moss N. (2005). 'Remote real-time monitoring of tunnelling induced settlement using image analysis'. *Proceedings of the 5th International Symposium on Geotechnical Aspects of Underground Construction in Soft Ground*, Amsterdam, the Netherlands, 771–777.

White, D. J., Take, W. A. and Bolton, M. D. (2003). 'Soil deformation measurement using particle image velocimetry (PIV) and photogrammetry'. *Géotechnique*, **50**(7): 619–631.

Wongsaroj, J., Soga, K. and Mair, R.J. (2007). 'Modelling of long-term ground response to tunnelling under St James's Park, London'. *Géotechnique*, **57**(1): 75–90.

Wright, P. (2010). 'Assessment of London underground tube tunnels investigation, monitoring and analysis'. *Smart Structures and Systems*, **6**(3), 239–262.

Mapping subsurface utilities with mobile electromagnetic geophysical sensor arrays

R. BIRKEN, Northeastern University, USA and
M. ORISTAGLIO, Yale University, USA

Abstract: This chapter reviews the remote sensing of buried utilities and explores its uses in today's commercial devices and in research systems that may become commercially available in the next ten years. Our survey concentrates on techniques that use electromagnetic (EM) waves to probe a few meters underground. These technologies have advanced dramatically during the last decade by combining broadband sensors with modern positioning systems and sophisticated signal processing. Sensors moving just above the ground, while transmitting and recording induction signals at kilohertz frequencies or ground-penetrating radar (GPR) at megahertz to gigahertz frequencies, have so far been easiest to deploy and interpret.

Key words: utility mapping, ground-penetrating radar, induction.

12.1   Introduction

Two events of the late twentieth century highlighted the value of 'mapping before digging' in the planning, construction, and maintenance of modern civil infrastructure. In 1998, the U.S. Congress enacted The Transportation Equity Act of the Twenty-first Century, which called for state-by-state implementation of a new notification system aimed at preventing construction accidents. The 'One-Call' system required utility companies to provide a free service, accessed by regional telephone numbers, to mark locations of buried utilities before digging. One year later, a Purdue University research project sponsored by the Federal Highway Administration (FHWA) reported its finding that mapping of utility lines before highway construction had generated substantial savings – about 2% of total construction costs – in 71 pilot projects totaling over $1 billion in Virginia, North Carolina, Ohio, and Texas (Lew, 1999). The Purdue study provided both validation and a commercial boost to the emerging discipline called subsurface utility engineering (SUE). In its most comprehensive form, SUE aims to create accurate subsurface

347

(a)

Air traffic control, 1030, 1090 MHz
GPS, 1215–1240 MHz
Wireless phone, 900 MHz

FM radio
VHF TV

Loran   AM radio

Speed gun
10.525, 24.25 GHz

RFID
120–150 kHz   13.56 MHz   433 902–928 MHz

3.1–10 GHz

Bridge deck and roadbed

Broadband EM induction

Utility mapping

Deep infrastructure/environment
Ultra wideband GPRs

100 Hz	1 kHz	10 kHz	100 kHz	1 MHz	10 MHz	100 MHz	1 GHz	10 GHz	100 GHz	
10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}	10^{11}	Frequency, Hz

1000 km 100 km 10 km 1 km 100 m 10 m 1 m 10 cm 1 cm 1 mm

Wavelength in vacuum

(b)

dB relative to 1 pT/Hz$^{1/2}$

Horizontal magnetic field (dB)

Measured in Stanford, CA

Summer

Winter

1 Hz	10 Hz	100 Hz	1 kHz	10 kHz	100 kHz

Frequency

dB relative to thermal noise 288°K

Radio wave noise power (dB)

Atmospheric
night

Business

Atmospheric
day

Residential

Quiet rural

Rural

100 kHz	1 MHz	10 MHz	100 MHz	1 GHz

Frequency

12.1 (a) EM spectrum for remote sensing of infrastructure. (b) Natural and human EM noise. (left) Natural horizontal magnetic field noise at inductive frequencies (Meloy, 2003). (right) Radio-wave noise power at radar frequencies in different environments (solid) compared to atmospheric noise (Bianchi and Meloni, 2007).

maps by combining data from existing maps or construction plans with information from three sources: surface surveys, noninvasive remote sensing methods, and minimally invasive exposure of utilities by hand digging or vacuum excavation at selected locations (Stevens and Anspach, 1993; Jeong *et al.*, 2004). In 2002, the American Society of Civil Engineers (ASCE) incorporated SUE practices into a National Consensus Standard titled ASCE C-I 38-02, *Standard Guidelines for the Collection and Depiction of Existing Subsurface Utility Data* (American Society of Civil Engineers, 2002).

This chapter reviews the state of the art of systems for noninvasive mapping of civil infrastructure. Electromagnetic (EM) sensors are currently the remote sensing technology of choice for probing the shallow underground; Fig. 12.1 shows the useful spectrum, which covers a broad band from quasi-direct current to frequencies near a gigahertz. Sensors include EM induction (EMI) coils, ground-penetrating radar (GPR) antennas, and radio-frequency identification (RFID) tags. Most commercial off-the-shelf systems consist of a single sensor pair – a transmitter and a receiver – geared for handheld surveys in which an individual carries the unit over the survey area, often making readings over a marked grid. To speed up this process, sensors can be mounted on pushcarts or small vehicles and combined with Global Positioning System (GPS) devices to record data and positions automatically. There are, however, limits to the accuracy and efficiency of using individual sensors to map large areas. Some limitations are inherent in the sensors themselves, which are antennas consisting of wire elements designed to broadcast (transmitter) and detect (receiver) EM waves in a designated frequency range. Mobile sensors must be compact, lightweight, and capable of recording accurately while on the move. Other limitations are set by interference from natural and human EM noise in various environments (Fig. 12.1b).

In the last decade, the next generation of technology for mapping infrastructure has emerged with the development of compact, broadband sensors combined with modern positioning systems in mobile sensor arrays. These new array systems can move quickly over large areas and could, in principle, produce 3D digital maps of a regional utility network within a relatively short period. Figure 12.2 shows a schematic of such a mobile system. The main components are (1) an array of sensors capable of recording data while in motion, (2) a platform and vehicle for transporting the sensors, (3) a positioning system to track the vehicle, and (4) software to merge and process the sensor and positioning data to create subsurface images or maps. Widespread commercial deployment of such systems during the next decade may herald a step change in our ability to intelligently manage civil infrastructure.

Our review will cover practical aspects of EM remote sensing of buried utilities and conduits. We start with the basic physics of EM probing of the underground, including the electrical properties of soils and the technical requirements for mapping materials of the size and shape found in civil infrastructure. We cover aspects of the commercial market for utility mapping in the One-Call System and SUE. We then describe several next-generation sensor arrays and show examples of recent large-scale surveys. We conclude with a discussion of future directions in mobile sensor technologies and barriers to their widespread commercial use.
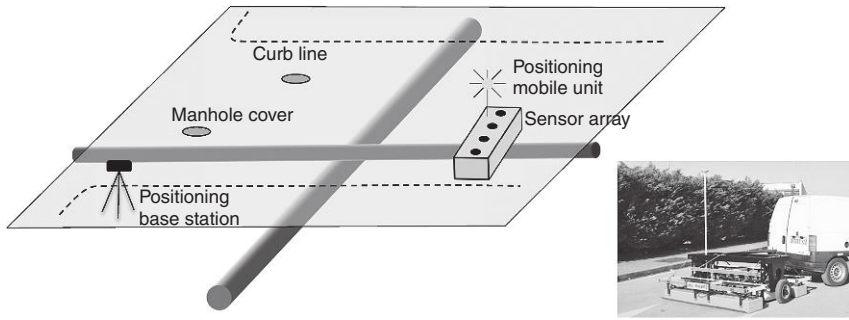
12.2 Elements of a mobile sensor system for mapping civil infrastructure. Photo at right shows a mobile multi-channel GPR array manufactured by Ingegneiria Dei Sistemi (IDS, www.idscorporation.com).

12.2 Physical concepts of passive and active EM remote sensing

Figure 12.3 shows a schematic vertical section through a city street overlying buried utility lines. Telecommunications (phone and cable) are usually closest to the road surface, at depths ranging from about 4 to 12 inches; gas and electric are next in depth, in the range from 12 to 36 inches; water, steam, sewer, and storm drain lines, usually in that order, lie at the bottom, generally at depths between 36 and 72 inches. Electric and telecommunications utilities are often contained in rectangular duct banks carrying several individual lines. The compositions of the lines and conduits vary widely, with increasing use today of non-metallic materials.

There are two basic ways of mapping buried utility lines with noninvasive EM sensors (Fig. 12.4). In the first, called 'passive' (or 'quasi-passive'), a signal from the line itself is detected at the surface. For example, electrical current flowing along a buried power line generates a magnetic field that can be detected at the surface with a magnetometer. The field oscillates at the frequency of the current, which is 60 Hz for most lines in the US. This variation enables it to be distinguished from Earth's natural magnetic field, which is nearly constant, and from stray fields at other frequencies generated by electronic equipment, though not of course from other 60 Hz signals. The field is strongest directly above the line and decreases with distance at a predictable rate – features that can be used to map the line's horizontal location and its approximate depth. This picture becomes more complicated when there are several power lines in the ground, running in different directions and carrying different amounts of current; this is often the case in crowded utility corridors. In addition, power lines often run parallel to other buried utilities that can channel secondary 60 Hz currents created by the primary signal through EMI.

0.0
Electric line
Telecommunications line
Electric
duct bank
Gas
Telecommunications
duct bank
0.5
2
Water
1.0
Steam
4
Sewer
1.5
Storm drain
6 ft
2.0
Depth, m
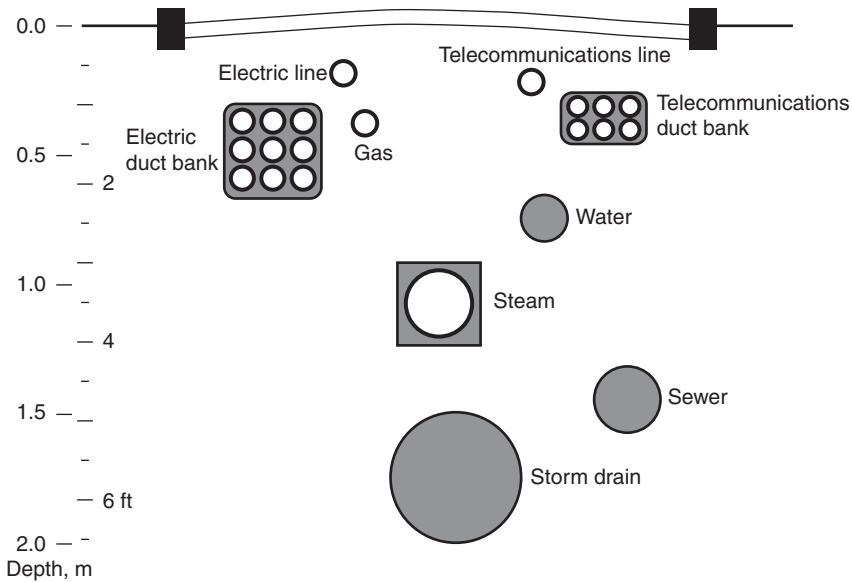
12.3 Schematic cross-section through a typical layout of utilities beneath a city street. (*Source*: after Macaulay, 1976.)

To overcome some of these limitations, a simple modification of the passive method involves using a current generator to inject electrical current at specific frequencies onto different buried lines or conduits. In this method, each line must be accessible somewhere at the surface or below ground – for example, at a valve cover or in a manhole. We call this method 'quasi-passive' because it requires a secondary source of current, but the physics of detection is essentially the same as the passive method. The quasi-passive method allows more flexibility and precision in locating individual lines by using a set of widely spaced frequencies, usually in the kHz range; it also lends itself to rapid mapping of networks of lines by mobile sensors. Leakage of signal onto neighboring lines, however, remains a problem (see the section on field examples below).

Active methods use EM sources and receivers making no contact with buried lines. The most intuitive is GPR, which operates by bouncing radio waves off buried objects (Fig. 12.4, right). Also called ground-probing radar, GPR uses a transmitting antenna to broadcast short pulses of EM energy into the ground and a receiving antenna to detect the echoes returning to the surface. At microwave frequencies from about 40 MHz to several GHz, EM signals propagate in soil as true waves, similar to radio waves in air that travel at essentially the speed of light in vacuum – 299 792 458 m/s or about 0.3 m (1 foot) per nanosecond (ns). The speed of radio waves in all materials, including soil, is less than in vacuum; the reduction factor is called the

material's index of refraction. The index of refraction of air near the ground is actually about 1.00026 (Budden, 1988). The index of most soils lies in the range of about 2 to 4, so the speed of radio waves in soil is two to four times slower than in air. For example, in soil with an index of refraction of 3, radio waves travel at a speed of about 0.1 m/ns.

As radio waves penetrate soil, buried objects whose electrical properties differ from those of the soil generate radar echoes. Metal objects generate the strongest radar echoes, but nearly all other materials, including plastics, cements, and ceramics, can generate detectable echoes. In addition, a local region of soil whose properties differ from the average, such as a wet patch in otherwise dry soil, can also cause echoes. In fact, one of the major difficulties in interpreting GPR is to distinguish radar echoes caused by objects of interest, such as buried utility lines, from the background clutter of echoes generated by other subsurface objects or by variations in the road bed and soil. A technique that helps greatly with GPR interpretation is to scan the antennas over a dense grid of stations covering the area, so that the echoes can be processed into images showing the shapes and locations of buried objects. This process resembles synthetic-aperture radar (SAR) techniques used to create radar images from satellites; it is the basis of the mobile imaging GPR systems described below.

There are two useful rules of thumb for pulse-echo GPR imaging in the geometry where transmitter and receiver antennas are nearly coincident.
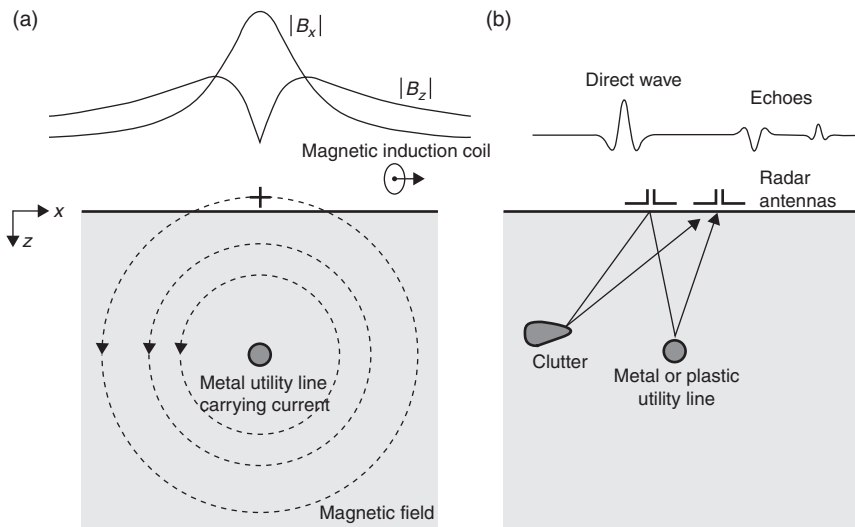
12.4 Remote sensing of buried utility lines. (a) Passive detection of a low-frequency magnetic field using EM induction. (b) Active interrogation of the subsurface with a radio-wave pulse using ground-penetrating radar.

The first is that the spatial sampling, the distance between GPR stations, should be no more than one fourth of the EM wavelength in the soil at the dominant frequency of the radar pulse (Grasmueck *et al.*, 2005). The second is that the maximum resolution of SAR imaging (the minimum feature size resolved in an image) is also about one fourth of the dominant wavelength in recorded echoes. At 1 GHz, the upper end of the useful GPR spectrum, the latter rule gives a resolution of about 2.5 cm (1 inch) in soil with an index of refraction of 3. This degree of resolution is more than sufficient for most civil engineering applications. Unfortunately, such high frequencies usually penetrate much less than a meter in moist soil. A more useful dominant frequency for GPR to detect buried utilities is 250 MHz, which can penetrate 2–3 m in many soils and provide a resolution of about 10 cm in soil with an index of refraction of 3.

Two significant effects limit the range of GPR. The first is an exponential attenuation of the signal amplitude caused by absorption of EM waves in soil materials – clay, sand, salty or fresh water, and organic matter. Mathematically, the intrinsic attenuation rate is proportional to the soil's average electrical conductivity and inversely proportional to its index of refraction. An approximate formula for the attenuation rate in decibels per meter (dB/m) is

$$-1640\frac{\sigma}{n}(\text{dB/m})$$

where σ is the soil conductivity in S/m, n is its index of refraction, and the decibel scale is defined as

$$dB = 20\log_{10}\left(\frac{A_f}{A_i}\right)$$

where A_f and A_i are the final and initial amplitudes, respectively, of the radar pulse. For example, a radio wave in soil with an index of 3 and conductivity of 0.02 S/m (resistivity ρ of 50 ohm-m) decays at the rate of about 10 dB/m. The best modern GPR systems can achieve a dynamic range of about 60 dB, which means that the waves can travel about 6 m in soil before 'falling off the radar screen.' This corresponds to a range of about 3 m for detecting buried objects with GPR – i.e., 3 m down to the object and 3 m back to the surface. For most soils, the conductivity is determined mostly by the presence of salt water, either in the pore space or adhering to clay particles. Radar therefore works best in dry, sandy soils that have electrical resistivity values well above 50 ohm-m. Resistivity values of common types of sandy-clay soils found in the United States range from about 20 to 100 ohm-m. Wet

clay soils can have a resistivity of 10 ohm-m or less. (For comparison, the resistivity of sea water is about 0.2 ohm-m.) Although frequency does not appear explicitly in the formula for the attenuation rate of GPR, the conductivity of moist soil does depend on frequency, and may increase by more than an order of magnitude over the GPR bandwidth (see Section 12.3).

The other effect limiting GPR is diffuse scattering caused by random variations in soil or roadbed properties. This mechanism is particularly important in radar propagation through complicated, layered roadbeds, or through coarse granular soils. Because of its complex physics (Rechtsman and Torquato, 2008), diffuse scattering is difficult to quantify, but it probably increases by at least a few percent the intrinsic exponential attenuation rate of radio waves in the near surface. In practice, the combination of intrinsic attenuation and scattering limits the effective range of GPR to depths of about 2 m or less in all but the most dry, uniform soils.

EM signals at lower frequencies – in the kilohertz (kHz) range or lower – can also be used in active subsurface remote sensing with transmitting and receiving antennas at the surface. These frequencies penetrate deeper into conductive soil, with an intrinsic attenuation rate that varies as the square root of frequency,

$$-0.02\sqrt{\sigma f} \text{ dB/m}$$

Thus, a 10 kHz EM signal traveling in 50 ohm-m soil attenuates at about 0.3 dB/m, so that a 60 dB induction system could in principle detect objects at a depth of 100 m. Unlike with radar, however, the physics of EM remote sensing at kHz frequencies is no longer pulse-echo detection. First, EM wavelengths are extremely long at kHz frequencies (1000 m in air at 300 kHz); second, EM propagation in conductive soil at these frequencies is highly dispersive, rendering the idea of a coherent propagating pulse nearly meaningless. The physics at these frequencies is closer to that of inductive metal detection: electrical currents are induced on buried conductive objects by the time-varying magnetic field (the primary field) of the loop of wire that serves as the transmitting antenna; the induced currents generate a secondary magnetic field; and the total field, the sum of the primary and secondary fields, is recorded by a second loop of wire that serves as the receiving antenna. After recording, primary and secondary fields can often be distinguished by their spatial patterns and phase shifts relative to the transmitter current.

Inductive remote sensing is most effective in locating buried metal objects and much less sensitive than radar to variations in soil conditions. In fact, the ability of induction to detect buried metal objects can actually improve

as soil conductivity increases, because any secondary currents flowing in the soil will tend to flow toward the most conductive objects in the subsurface, which are usually metallic utility lines or conduits.

The physical concepts described in this section can be made more precise by looking at simple solutions of Maxwell's equations in the earth.

12.3   Physics of EM waves in the shallow subsurface

Radar and induction methods for probing objects in earth generate and detect classical EM fields described by solutions of Maxwell's equations in conductive media. This section presents simple models to illustrate the physics of EM subsurface remote sensing and imaging. Many general texts on applied electromagnetism cover similar ground in more detail and at a more relaxed pace. Two excellent introductory texts are Kraus and Fleisch (1999) and Ulaby *et al.* (2010); a more advanced text with emphasis on fields in the earth and other inhomogeneous media is Chew (1990). There is a vast literature on radar and EMI for remote sensing and non-destructive evaluation. A standard general reference for radar is Skolnik (2002). Daniels (2004), Butler (2005), and Jol (2009) offer recent treatments of GPR systems and technology; Wang and Oristaglio (2000a) and Oristaglio *et al.* (2001) cover advanced aspects of GPR imaging. The two-volume monograph edited by Nabighian (1988) is still the most comprehensive reference on geophysical applications of EMI methods.

We consider solutions of Maxwell's equations for fields varying harmonically (sinusoidally) in time as

$$A(\mathbf{r},t) = R\{A(\mathbf{r},\omega)e^{-i\omega t}\},$$

where $A(\mathbf{r}, t)$ is a real function giving the amplitude of a field component at location $\mathbf{r}$ and time t, $A(\mathbf{r}, \omega)$ is the complex harmonic field amplitude at angular frequency $\omega = 2\pi f$, where f is frequency in cycles per second (Hz); $i = \sqrt{-1}$, and $R\{\cdot\}$ denotes the real part of the quantity in braces. Any physical field varying in time can be represented as a sum or integral of harmonic components. We use Cartesian coordinates with position vector $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$, where $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ are unit vectors in the coordinate directions, and $\{x, y, z\}$ are the coordinates of $\mathbf{r}$. The radial distance from the origin is the length of the position vector, $r = \sqrt{x^2 + y^2 + z^2}$, while $\hat{\mathbf{r}} = \mathbf{r}/r$ is a unit vector in the radial direction. We use the coordinate z to represent depth in the Earth.

Maxwell's equations for the complex electric field $\mathbf{E}(\mathbf{r}, \omega)$ and magnetic field $\mathbf{B}(\mathbf{r}, \omega)$ vectors are

$$\nabla \times \mathbf{E} = i\omega\mathbf{B}, \tag{12.1}$$

$$\nabla \times \mathbf{B}/\mu = (\sigma - i\omega\,\epsilon)\mathbf{E} + \mathbf{J}_s, \tag{12.2}$$

where the ultimate sources of the EM fields are impressed currents, represented by the current density $\mathbf{J}_s(\mathbf{r},\omega)$. In radar, these sources are short pulses of current on a transmitting antenna that broadcasts radio waves at megahertz (MHz) frequencies; in induction, the sources are currents varying at kilohertz (kHz) frequencies on a loop of wire that creates a primary magnetic field.

The material properties of the subsurface – soil, rock, and buried objects – are given by the electrical permittivity ϵ, conductivity σ, and permeability μ. In the most general anisotropic medium, these properties can all be tensor functions of position and frequency. We will assume for simplicity that electrical anisotropy is negligible, so that the electrical properties are all scalar quantities, and that μ assumes its free-space value $\mu_0 = 4\pi \times 10^{-7}$ ohm-s/m, which holds for most soils. (Exceptions are certain magnetic soils, usually derived from weathering of volcanic rocks with high concentrations of iron-rich minerals that have a large magnetic permeability; see Dabas *et al.*, 1992.) It is also convenient to write $\epsilon = \epsilon_r\epsilon_0$, where $\epsilon_0 = 1/(c_0^2\mu_0) \approx 8.85 \times 10^{-12}$ s/ohm-m is the permittivity of free space, and ϵ_r is the medium's relative permittivity, sometimes called its dielectric value or dielectric function. Here, $c_0 = 1/\sqrt{\mu_0\epsilon_0}$ is the speed of light in free space. Another useful quantity is $\eta_0 = \sqrt{\mu_0/\epsilon_0} \approx 377$ ohms, called the free-space impedance.

The term $(\sigma - i\omega\epsilon)\mathbf{E}$ in the second of Maxwell's equations represents electrical current that flows in material media (in response to the electric field) as a combination of conduction currents $(\sigma\mathbf{E})$ and displacement currents $(-i\omega\epsilon\mathbf{E})$. Conduction current flows in electrically conductive material, such as metal objects, salt water and clay-rich soil; displacement current exists wherever a time-varying electric field is present, but is the dominant component of the total current mainly in insulating materials, such as plastics, dry sand, and solid rock, and in fresh water.

12.3.1   Propagation and diffusion

Maxwell's equations combine to give the vector wave equation

$$\nabla \times \nabla \times \mathbf{E} - k^2\mathbf{E} = i\omega\mu\mathbf{J}_s \tag{12.3}$$

where $k^2 = \omega^2\mu\epsilon + i\omega\mu\sigma$. In homogeneous, source-free regions, Equation [12.3] has plane-wave solutions of the form

$$\mathbf{E}(\mathbf{r},\omega) = \mathbf{E}_0 e^{i\mathbf{k}\cdot\mathbf{r}} \tag{12.4}$$

where $\mathbf{k}$ is a complex propagation vector and $\mathbf{E}_0$ is a complex amplitude vector, subject to

$$\mathbf{k}\cdot\mathbf{k} = k^2 = \omega^2\mu\epsilon + i\omega\mu\sigma; \quad \mathbf{k}\cdot\mathbf{E}_0 = 0 \tag{12.5}$$

Assume that the plane wave travels along the z-axis (depth), and is polarized along the y-direction; then the electric field is given by

$$\mathbf{E}(\mathbf{r},\omega) = \hat{\mathbf{y}}E_y(z,\omega) = \hat{\mathbf{y}}E_0 e^{ikz}, \tag{12.6}$$

where

$$k = k_R + ik_I = \omega\sqrt{\mu\epsilon}\left(1 + i\sigma/\omega\epsilon\right)^{1/2} \tag{12.7}$$

is the wave's complex propagation constant, E_0 is its (real) amplitude at the surface $z = 0$, and $\hat{\mathbf{y}}$ is a unit vector in the y-direction. In general, the propagation constant (Equation [12.7]) describes a wave that moves at phase velocity k_R/ω and decays at the rate k_I in the positive z-direction (when $k_R, k_I, \omega > 0$ and time-dependence $e^{-i\omega t}$ is restored).

Radar regime

The expression for the propagation constant simplifies considerably in the wavelike (radar) and diffusion (induction) regimes, respectively. The radar regime is defined by the condition $\sigma/\omega\,\epsilon \ll 1$, which gives

$$k \approx \omega\sqrt{\mu\epsilon}\left(1 + i\frac{\sigma}{2\omega\epsilon}\right) = \frac{\omega}{c} + i\alpha \tag{12.8}$$

where

$$c = \frac{1}{\sqrt{\mu\epsilon_0\epsilon_r}} = \frac{c_0}{n}, \text{ and} \tag{12.9}$$

$$\alpha = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon}} = \frac{\eta_0\sigma}{2n} \tag{12.10}$$

where $n = \sqrt{\epsilon_r}$ is the index of refraction of the material. The full form of the field as a function of depth and time is

$$E_y(z,t) = R\{E_o e^{-\alpha z} e^{i\omega z/c} e^{-i\omega t}\} = E_o e^{-\alpha z} \cos \omega(z/c - t), \qquad [12.11]$$

which represents a sinusoidal wave that moves with speed $c = c_0/n$ in the positive z-direction and attenuates exponentially at rate α. If the electrical properties do not themselves depend explicitly on frequency, then both the speed and attenuation rate of radio waves are independent of frequency in this regime. This means that a superposition of harmonic waves, such as a radar pulse, retains its shape as it propagates through the medium. In practice, however, moist soils exhibit dispersion in the microwave range. The index of refraction varies by several percent, and the effective conductivity varies by more than an order of magnitude within the frequency range of GPR. This dispersion does not significantly alter radar pulse shapes in soil, but does increase the attenuation rate of radar pulses as their central frequency rises to GHz frequencies.

Electrical properties of soils at microwave frequencies

When both conductivity and permittivity vary with frequency, it is convenient to express the physical mechanisms generating displacement and conduction currents as a single quantity called the complex permittivity, defined by

$$\mathbf{J} = (\sigma - i\omega\epsilon)\mathbf{E} = -i\omega(\epsilon + i\sigma/\omega)\mathbf{E} = -i\omega\hat{\epsilon}(\omega)\mathbf{E}. \qquad [12.12]$$

One reason for this identification is that simple phenomenological models are available to represent the combined effects of polarization and conduction in real materials, including soils, in a way consistent with the requirement of causality. For example, Debye models represent the causal complex permittivity $\hat{\epsilon}$ as a sum of rational functions,

$$\hat{\epsilon} = \epsilon_0 \left(\epsilon_\infty + \sum_p \frac{a_p}{1 - i\omega\tau_p} \right) + i\frac{\sigma_{DC}}{\omega}, \qquad [12.13]$$

where ϵ_∞ is the limit of the relative permittivity at infinite frequency and σ_{DC} is the DC electrical conductivity. For use in the equations above, a Debye model gives the following results for relative permittivity and (effective) conductivity,

$$\epsilon_r = \mathcal{R}\{\hat{\epsilon}/\epsilon_0\}; \quad \sigma = I\{\omega\hat{\epsilon}\}. \tag{12.14}$$

Figure 12.5 shows measurements of the permittivity and conductivity of wet soils at frequencies from 1 MHz to 1 GHz (Hipp, 1974), along with fits to the measurements with a two-term Debye model (Wang and Oristaglio, 2000b). In this range, the permittivity varies by 10–20%, and the conductivity increases by more than an order of magnitude.

Induction regime

The induction regime is defined by the condition $\sigma/\omega\epsilon \gg 1$; the propagation constant simplifies to

$$k = k_R + ik_I \approx \sqrt{i\omega\mu\sigma} = (1+i)\alpha', \tag{12.15}$$

where

$$\alpha' = \sqrt{\frac{\omega\mu\sigma}{2}} \tag{12.16}$$

The expression for the propagating harmonic field is then

$$E_y(z,t) = E_o{}^{-\alpha'z}\cos\omega(z/c' - t), \tag{12.17}$$

where

$$c'(\omega) = \sqrt{\frac{2\omega}{\sigma\mu}} \tag{12.18}$$

is the phase velocity of the plane wave, which depends strongly on frequency.

Attenuation in conductive media is often expressed as the skin depth δ, the distance in which the plane-wave amplitude decays by $e^{-1} \approx 0.37$ (–8.7 dB) of its initial value. From Equation [12.17], we have

$$\delta = 1/\alpha' = \sqrt{\frac{2}{\omega\mu\sigma}} \approx 503.3\sqrt{\frac{\rho}{f}}. \tag{12.19}$$

A plane wave in the inductive regime travels about 7 skin depths before falling below a 60 dB threshold for detection. But since skin depth varies inversely with frequency in this regime, the detection range can in principle be increased to any desired value by lowering the frequency. In practice, external human or natural noise eventually sets a detection threshold. The

tradeoff for increased depth of detection in the induction regime is that the wavelength,

$$\lambda = \frac{2\pi}{k_R} = 2\pi \cdot \delta = 3162.3 \sqrt{\frac{\rho}{f}}, \qquad [12.20]$$

is a multiple of the skin depth; thus, the resolution of inductive systems degrades quickly as the depth of penetration increases. A crude rule of thumb is that the resolution of a fixed-frequency inductive system is about 5% of its maximum depth of penetration. Systems broadcasting and detecting signals that penetrate to a depth of 100 m in the ground will be able to resolve an unknown feature at that depth only to within about ±5 m. Inductive systems can, however, give more precise location estimates than the rule of thumb by using parametric methods that assume a specific size and shape for the object (for example, a long, linear wire) and invert measurements to determine only the object's horizontal position and depth (Section 12.3.3). Of course, if the assumptions used in a specific parametric method are not true for the actual object being located, the location estimates can be far off the mark.

12.3.2 Sources

The EM field generated by a GPR antenna, or by current flowing along a short length of wire in an inductive transmitter or along a short section of buried metal utility line, is modeled well by the field of an ideal electric dipole. Moreover, the field of an arbitrary distribution of current can always be represented as a superposition of current dipoles. To study this solution, let the dipole be at the origin of coordinates and point in the y-direction. Then the electric field satisfies the vector wave equation

$$\nabla \times \nabla \times \mathbf{E} - k^2(\omega)\mathbf{E} = \hat{y}i\omega\mu\bar{I}(\omega)\delta(\mathbf{r}), \qquad [12.21]$$

where $\bar{I}(\omega)$ is the current moment of the dipole in units of amp-m (a current dipole is the idealization of current I flowing along a short length of wire dl).

The electric field of the dipole is most easily expressed in spherical coordinates (r,θ,ϕ) around the dipole axis (Fig. 12.6, top left). In this coordinate system, r is the distance from the dipole to the observation point, θ is the angle between the dipole axis (y-axis) and the radius vector, and ϕ is the azimuthal angle. For convenience, the azimuthal angle is measured counterclockwise between the z-axis and the projection of
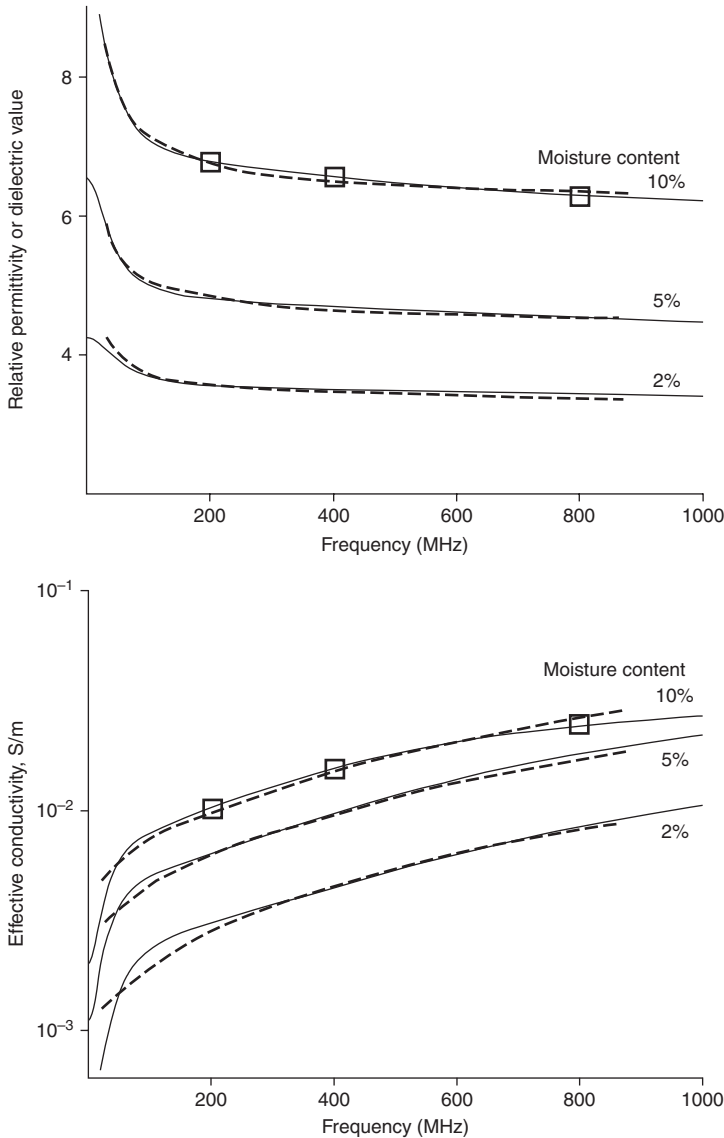
12.5 Electrical properties of moist sandy-loam soils. Dashed curves are measurements (Hipp, 1974); solid curves are fits to the dispersive properties with a two-term Debye model.

the radius vector onto the (x,z)-plane. The geometry is easiest to visualize in the (x,z)-plane perpendicular to the dipole axis; there is azimuthal symmetry in this plane, and the EM field is quasi-2D. The vector fields are given by

$$\mathbf{B}(\mathbf{r}, \omega) = B_\phi \hat{\phi}, \quad \text{and} \quad \mathbf{E}(\mathbf{r}, \omega) = E_r \hat{\mathbf{r}} + E_\theta \hat{\theta} \qquad [12.22]$$

where $(\hat{r}, \hat{\theta}, \hat{\phi})$ are unit vectors in the spherical coordinate directions; the field components are

$$B_\phi = -\mu \bar{I}(\omega)\left(-ik + \frac{1}{r}\right)\frac{e^{ikr}}{4\pi r}\sin\theta, \qquad [12.23]$$

$$E_\theta = -i\omega\mu \bar{I}(\omega)\left(1 + \frac{i}{kr} - \frac{1}{k^2 r^2}\right)\frac{e^{ikr}}{4\pi r}\sin\theta, \qquad [12.24]$$

$$E_r = -i\omega\mu \bar{I}(\omega)\left(\frac{2i}{kr} - \frac{2}{k^2 r^2}\right)\frac{e^{ikr}}{4\pi r}\cos\theta \qquad [12.25]$$

Note that in the x, z-plane, the unit vector $\hat{\theta} = -\hat{\mathbf{y}}$, so that E_θ is just the electric field perpendicular to the plane. The terms proportional to r^{-1} in each expression represent the radiation field or far field of the dipole, which dominates the other terms if $kr \gg 1$. The terms with higher (inverse) powers represent the near field, which has the same form as that of a static electric dipole. The other panels of Fig. 12.6 are explained in Section 12.3.4 below on GPR propagation in the ground.

12.3.3   Model for passive utility locating

The most useful model for understanding the passive, or quasi-passive, mapping of utilities is the field of an infinite line current. This model is a good approximation to the field set up when current is injected onto a long metal pipe in soil; its magnetic field can be obtained by integrating the field of an electric current dipole (Equation [12.23]),

$$B_\phi = -\mu \bar{I}(\omega)\left(-ik + \frac{1}{r}\right)\frac{e^{ikr}}{4\pi r}\cos\theta,$$

along the infinite length of the line current. In the limit of steady current $\omega = 0$, the magnetic field of a current dipole reduces to

$$B_\phi = \frac{\mu I dl}{4\pi r^2}\sin\theta, \quad \text{or} \quad \mathbf{B} = \frac{\mu I dl \times \hat{r}}{4\pi r^2}, \quad (\omega = 0), \qquad [12.26]$$

(a)

(b)
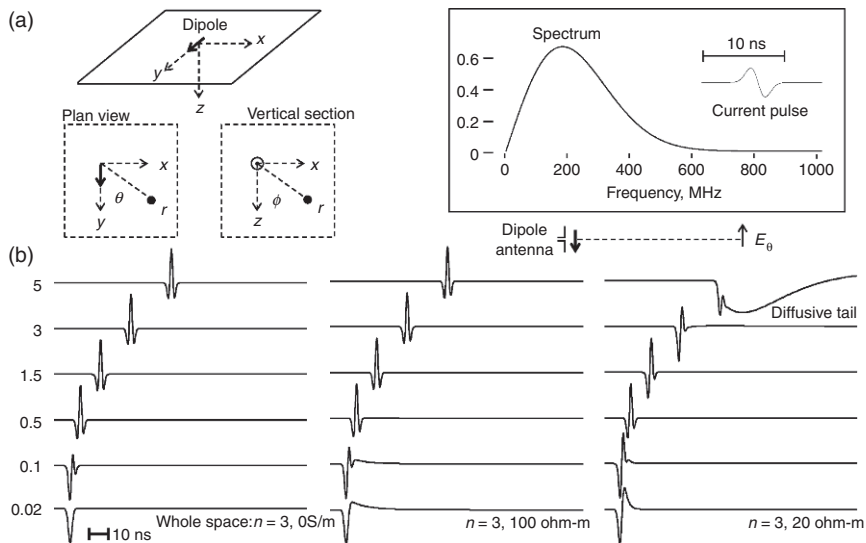
12.6 Radio-wave propagation from a dipole antenna. (a) This panel shows the source waveform and its spectrum. (b) These panels show the field radiated by the dipole in regions with different electrical conductivities.

which is the classic Biot-Savart law (Kraus and Fleisch, 1999) for the magnetic field of a steady current along a short length of wire *dl*. Integrating this equation gives the formula for the magnetic field of a steady line current,

$$B_\phi = \frac{\mu I}{2\pi r},$$ [12.27]

or

$$B_x = \frac{\mu I}{2\pi} \frac{z}{x^2 + z^2}, \quad \text{and} \quad B_z = -\frac{\mu I}{2\pi} \frac{x}{x^2 + z^2}.$$ [12.28]

The origin is, for convenience, the point where the pipe intersects the measurement plane (the z coordinate of ground level is thus negative). These equations hold for any finite conductivity of the soil and show that when measured along a profile along the surface the horizontal field peaks directly above the line current, while the vertical field goes through zero.

The magnetic field at finite frequency is

$$B_\phi = \frac{\mu k I}{4i} H_1(kr),$$ [12.29]

where H_1 is the first-order Hankel function of the first kind. For large and small values of kr, the Hankel function has the following forms,

$$H_1(kr) \approx \frac{2i}{\pi kr}, \quad (kr \to 0); \qquad [12.30]$$

$$H_1(kr) \approx -\sqrt{\frac{2}{\pi}} \frac{e^{ikr - i\pi/4}}{\sqrt{kr}}, \quad (kr \to \infty). \qquad [12.31]$$

The small-argument expansion reproduces the static limit as $\omega \to 0$; the large-argument expansion determines the decay rate of the magnetic field at large distances from a linear current in conductive soil. The detection limit for an individual pipe is reached when its magnetic field falls below a threshold set either by the sensitivity of the surface magnetometer or by background noise. The detection limit is therefore determined by the magnitude of the current and by the product of the frequency of operation and the soil conductivity. Figure 12.7 shows the amplitude of the magnetic field directly above a pipe carrying 1 mA of current as the pipe's depth increases. Currents impressed onto pipes in the quasi-passive locating method are generally in the mA range.

The static limit, Equation [12.27], gives a simple method for estimating the depth of a buried pipe from magnetic field measurements at two different heights above the ground. Modern inductive pipe locators have several internal coils for measuring the magnetic field. A typical arrangement has two pairs of coils: each pair measures the vertical component and one horizontal component of the magnetic field – for example, (B_x, B_z) as shown in Fig. 12.7 – at a fixed height above the ground. Some devices contain a fifth coil that measures the remaining horizontal component (B_y in Fig. 12.7) at a point between the two other coil pairs. If there is only a single current-carrying line in the subsurface, and the locator is arranged so that $B_z = B_y = 0$, then the instrument is directly above the line and the y-sensor points in the line direction. In this configuration, the depth of the line is given by

$$D = \frac{d}{a-1} - h, \qquad [12.32]$$

where $a = \left| B_x^{(2)} / B_x^{(1)} \right|$ is the ratio of the magnetic fields at the two B_x sensors, d is their separation, and h is the height of the lower sensor above the ground. The more accurate expression (Equation [12.29]) can be used to compute the depth, but requires knowledge of the soil conductivity. The

right panel of Fig. 12.7 shows the error in the estimated depth using the static limit and Equation [12.32]. The percentage error increases with the depth of the pipe and with the conductivity-frequency product. (A more precise analysis would take into account effects at the earth-air interface, but it can be shown that these effects are negligible when the conductivity-frequency product is less than 20 000 S/m-s.)

12.3.4   Detection of utility lines with GPR

The electric field given by Equations [12.24] and [12.25] is a good approximation to the field of a GPR antenna lying on the surface of the earth. In the (x,z)-plane (the plane perpendicular to the axis of the dipole at its midpoint), the electric field has only a y-component,

$$E_y(\mathbf{r},\omega) = i\omega\mu\overline{I}(\omega)\left(1 + \frac{i}{kr} - \frac{1}{k^2 r^2}\right)\frac{e^{ikr}}{4\pi r}, \qquad [12.33]$$

where $r = \sqrt{x^2 + z^2}$. This expression does not include effects at the earth-air interface, which modifies the amplitude of the field radiated in different directions into the ground (see Equation [12.38] below). A few wavelengths from the source, the first term dominates, giving the radiation field,

$$E_y(\mathbf{r},\omega) \approx i\omega\mu\overline{I}(\omega)\frac{e^{ikr}}{4\pi r}. \qquad [12.34]$$

We can use Equation [12.34] to illustrate the difference between radar propagation in non-conductive and conductive soil by computing the radar wave radiated from an electric dipole driven by a short pulse of current in the way most modern GPR systems operate. The time-dependent field is an integral superposition of harmonic solutions

$$E_y(\mathbf{r},t) = R\left\{\frac{1}{\pi}\int_0^\infty E_y(\mathbf{r},\omega)e^{-i\omega t}d\omega\right\}. \qquad [12.35]$$

In conductive media, the integral must be evaluated numerically. Figure 12.6 shows the evolution from the near field to the far field for three different values of the whole-space conductivity: 0, 0.01, and 0.05 S/m (a relatively conductive medium); in each case, the medium has an index of refraction of 3 ($\epsilon_r = 9$). The source current is the first derivative of a Gaussian function, whose frequency content is centered at 200 MHz.

(a)

d
h
D

x

z

Horizontal magnetic field (nT)

σ f

0
1000
2000
5000
10 000
20 000

Depth of pipe (m)

(b)

B_z
B_x
B_y
d

σ f, S/m.s

% Error in
estimated depth

−10
−20
−30
−40
−50
−60
−70

Actual depth (m)
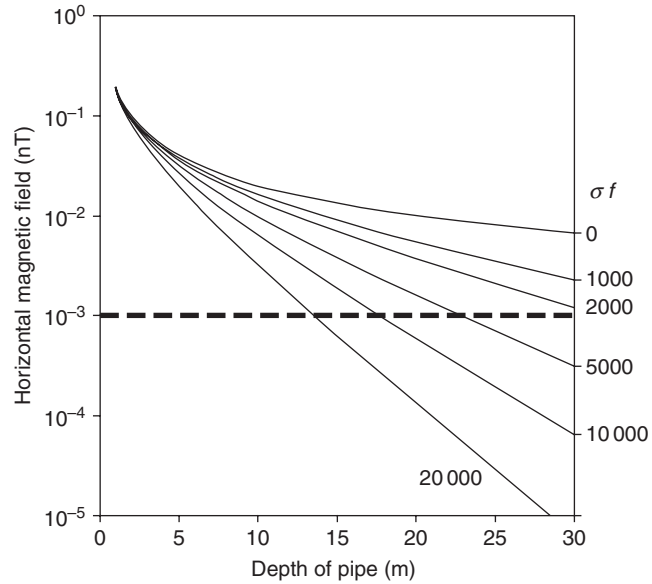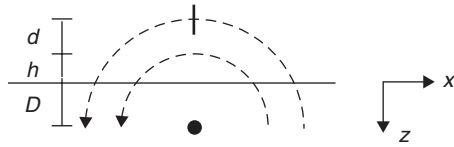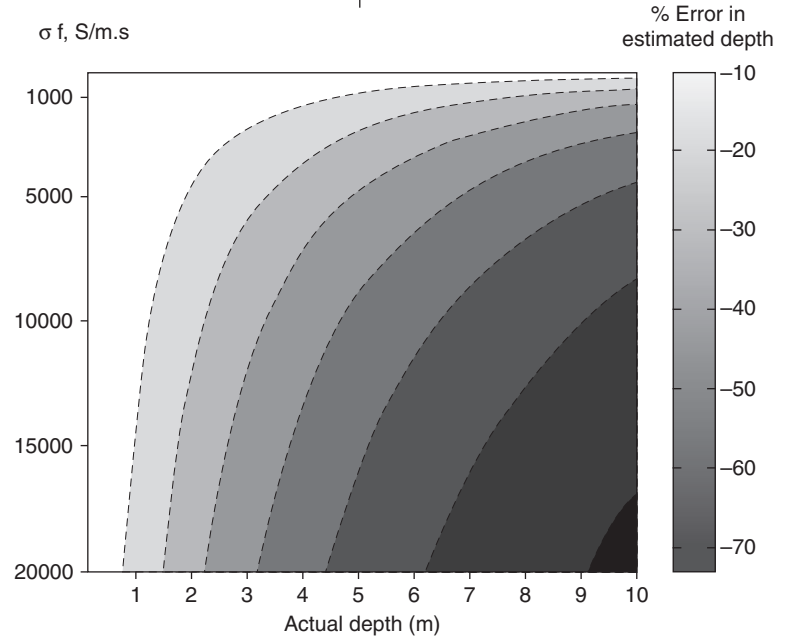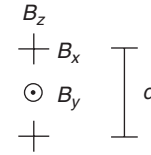
12.7 (a) Magnetic field above a current-carrying pipe at different depths, for typical values of conductivity times frequency. Dashed line is a 1 pT noise limit. (b) Error in the depth to the pipe estimated by Equation [12.32].

Consider first the non-conductive medium (left panel). In the near field, the propagated waveform is a combination of the source current and its integral, whereas in the far field the waveform is proportional to the time-derivative of the current waveform, with the propagation delay, $r/c = nr/c_0$,

$$E_y(\mathbf{r},t) \approx -\frac{\mu}{4\pi r}\bar{I}'(t-r/c), \quad (r\to\infty),$$ [12.36]

where $\bar{I}' = d\bar{I}/dt$. The far-field approximation is already accurate to a few percent at a distance from the source of 1.5 m, which equals three wavelengths at the central frequency of the waveform.

Distortion of the propagating pulse in a conductive medium is clearly visible in the waveforms in the far right panel. In particular, at 5 m from the source, the waveform acquires a long tail of diffusive energy, which arises because the frequency of 200 MHz is at the transition from diffusive to wavelike behavior in this medium.

The middle panel shows waveforms in soil of conductivity 0.02 S/m. Distortion of the waveform is negligible; the main effect in the far field is an exponential attenuation of the signal,

$$E_y(\mathbf{r},t) \approx -\frac{\mu e^{-\alpha r}}{4\pi r}\bar{I}'(t-r/c), \quad (r\to\infty),$$ [12.37]

where α is given by Equation [12.10]. In the middle panel of Fig. 12.6, the attenuation is compensated in the plotting to allow better comparison of waveform shapes. Equation [12.37] is perhaps the most important single equation in GPR and can be used to derive many other results, including formulas for synthetic-aperture GPR imaging (Oristaglio *et al.*, 2001).

The presence of the earth-air interface modifies the whole-space formulas presented above. Fortunately, when GPR antennas are placed close to the ground (within a quarter wavelength), the main effect of the interface is to change the radiation and receiving patterns of antennas. For example, the electric field in the plane perpendicular to the dipole is well approximated by

$$E_\theta = -i\omega\mu\bar{I}(\omega)\left(1+\frac{i}{kr}-\frac{1}{k^2r^2}\right)\frac{e^{ikr}}{4\pi r}\sin\theta\cdot F(\phi),$$ [12.38]

where

$$F(\phi) = \frac{2n\cos\phi}{\left(1-n^2\sin^2\phi\right)^{1/2}+n\cos\phi},$$ [12.39]

and ϕ is the angle between the normal to the earth-air interface and the radius vector (see Fig. 12.6, top left). A derivation and full treatment of the 3D radiation pattern of an interfacial dipole can be found in Engheta *et al.* (1982).

GPR range equation

An approximate range equation for GPR follows from the model in Fig. 12.8, in which the transmitting and receiving antennas are close together, directly above a pipe at depth z in the ground. The following expression based on [12.38] is a good approximation to the amplitude of the returned echo:

$$A_e = C_s \left(\frac{e^{-\alpha z}}{z} \right) R_{cs} \left(A_i \frac{e^{-\alpha z}}{z} \right) = C_s \cdot R_{cs} \cdot A_i \frac{e^{-\eta_0 \sigma z / n}}{z^2}, \qquad [12.40]$$

where A_i is the amplitude of the wave transmitted into the ground, $\eta_0 \approx 377$ ohm is the free-space impedance; σ and n are the conductivity and index of refraction of the ground; R_{cs} is the radar cross-section of the pipe; and C_s is a system gain factor, which may depend on the central frequency of the system. The radar cross-section R_{cs} is a measure of an object's ability to generate an echo and has units of m^2. The first three terms from right to left represent: (1) propagation from the transmitter to the pipe; (2) reflection at the pipe; and (3) propagation back to the antenna. Equation [12.40] is not accurate when objects are within a few wavelengths of finite-sized antennas used in actual GPR systems. A simple modification is to add an effective standoff to the depth, letting $z = z + H$, where H is determined empirically to normalize amplitudes at some fixed distance from the antennas. In tests of commercial broadband GPR antennas with central frequencies in the range 200–800 MHz, we have found that $H = 1$ m gives reasonable results.

The condition for detecting the echo is that the amplitude ratio A_e / A_i should be within the system dynamic range dB_s expressed in decibels; that is,

$$20 \log_{10} \left(A_e / A_i \right) = 20 \log_{10} \left(R_{cs} + C_s \right) - 40 \log z - 3272 \left(\sigma / n \right) z > dB_S, \quad [12.41]$$

where frequency-dependent quantities, such as $\sigma(\omega)$, $n = \sqrt{\epsilon(\omega)}$, and C_s, can be evaluated at the central frequency of the source without much loss of accuracy. According to Equation [12.41], the depth of penetration of radar increases as the soil's electrical conductivity decreases (its resistivity increases) and as its index of refraction increases. In moist soils, the natural increase of conductivity with frequency in the microwave range (Fig. 12.5)
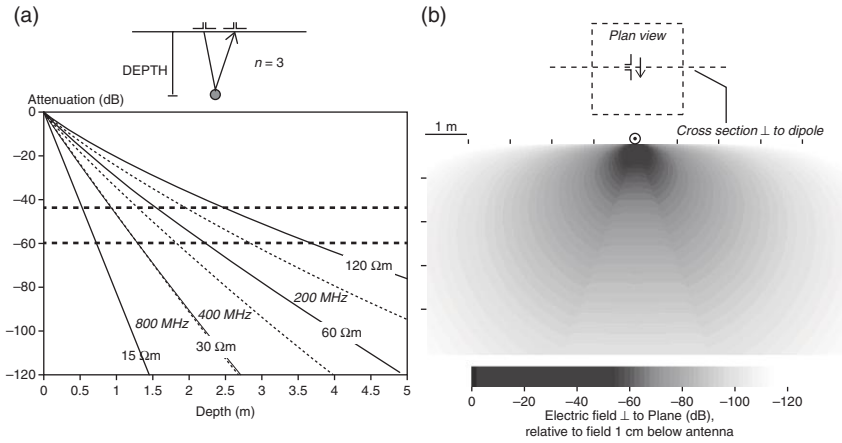
12.8 GPR range in conductive soil. (a) Estimated amplitude of a radar echo from a pipe versus its depth of burial in soils with different conductivities. (b) Amplitude radiation pattern of a dipole antenna into the ground.

reduces the depth of penetration of GPR at higher frequencies. In addition, properties of the radar system itself, such as antenna efficiency and dynamic range, vary with frequency, causing depth of penetration to vary. For example, the efficiency of electronics generally decreases as frequency increases in the microwave range.

The conductivity of typical sandy-clay soils found in the United States ranges from about 0.01 to 0.05 S/m, corresponding to resistivities from 100 to 20 ohm-m. The index of refraction of nearly all soils falls in the range from 1.4 to 6, corresponding to dielectric values from about 2 to 36. These properties may vary considerably from place to place (Doolittle *et al.*, 2010), and need to be understood locally for best use of GPR. For example, we have found through many surveys that soils in the New York metropolitan area generally have conductivity values between 0.02 and 0.04 S/m and refractive indices between 3 and 4.

The radar cross-section R_{cs} depends on the contrast in electrical properties between the soil and object, the size and shape of the object, and its internal structure. It is often the most difficult quantity to estimate in the range equation. The simplest model that can be solved analytically is an interface between two infinite layers with different electrical properties, which can be used to model the echo amplitude generated at a horizontal interface between two different types of soil below the surface. Let a harmonic vertically traveling plane wave, with its electric field polarized parallel to the interface (as in Equation [12.6]), be incident on the interface. Then a reflected plane wave (echo) is generated at the interface, and the ratio

of the amplitudes of the reflected and incident plane waves is given by the reflection coefficient,

$$R = \frac{\sqrt{\hat{\varepsilon}_1} - \sqrt{\hat{\varepsilon}_2}}{\sqrt{\hat{\varepsilon}_1} + \sqrt{\hat{\varepsilon}_2}}$$

where $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ are the values of complex electrical permittivity in the upper and lower layers, respectively. The complex permittivity depends on both the dielectric value and the electrical conductivity of a material, and it depends on frequency. For non-conductive materials, this expression reduces to the standard reflection coefficient for light at normal incidence between two media with indices of refraction n_1 and n_2,

$$R = \frac{n_1 - n_2}{n_1 + n_2}$$

The reflection coefficient at an interface, which is dimensionless, illustrates how the echo amplitude depends locally on the contrast in properties between two regions. For finite-sized objects embedded in soil there is no simple expression for the radar cross-section. The intuitive picture for an object illuminated in the far field by radar whose wavelength is small compared to the object's linear dimensions is that the cross-section is the product of the reflection coefficient at the object's surface times the surface area (m^2) illuminated by the radio wave. This simple picture breaks down when an object is comparable in size to the radar wavelength, as in GPR applications, and echoes are generated by a combination of scattering from the object's surface and from its internal structures (a conduit carrying a utility line), with interference among echoes generated at different locations.

Some general observations can nevertheless be made. Large objects embedded in soil usually have larger radar cross-sections than small objects composed of the same material. Also, metal objects have larger radar cross-sections than other objects of similar size, because the conductivity of metals is many orders of magnitude higher than the conductivity of soils. Plastic, ceramic, and cement objects have radar cross-sections because their dielectric values differ from those of soils (but are usually within a factor 10 of soil values). At radar frequencies, water has a dielectric of about 80, one of the largest values of any material; this makes it a strong radar scatterer (a property exploited in weather radar). The best way to determine R_{cs}, along with other system parameters, is by calibration in a test pit.

Figure 12.8 illustrates the GPR range equation, calibrated by measurements made in a test with a 200 MHz GPR. The system had a dynamic range

of about 44 dB (150:1) in the amplitude of echoes that could be registered above system noise (Birken *et al.*, 2005). The range of the system's analog-to-digital converter was actually 60 dB, but the realized dynamic range was limited by system noise. A controlled test with an 8-in. diameter metal pipe buried 4 ft (1.2 m) deep in dry sand $(\sigma \approx 0, n = 4)$, determined a value for $C_s \cdot R_{cs}$ of about 1 (m²) for the test configuration. Thus, the curves in Fig. 12.8 follow Equation [12.41], with this normalization.

The solid curves labeled 15, 30, and 60, and 120 ohm-m show echo amplitude versus depth for a pipe buried in soil with an index of refraction of 3 and with the indicated resistivity values. The ideal detection limit occurs at the point where these curves intersect the −60 dB line; the practical limit is at the system noise level of −44 dB. These curves show that a pipe with a radar cross-section of 1 m² could be detected at depths of 0.5, 0.8, 1.2, and 2.5 m in soils with resistivities of 15, 30, and 60, and 120 ohm-m, respectively. Reducing system noise to make available the full dynamic range of 60 dB increases these depths of detection to 0.7, 1.3, 2.2, and 3.6 m.

The dashed curves show the effect of frequency on the GPR depth of investigation. These curves, labeled 200, 400, and 800 MHz, correspond to radar propagation in one type of soil (a sandy-loam with 10% moisture content) with pulses centered at the indicated frequencies. Because of electrical dispersion in this frequency range, pulses with different center frequencies sense different effective dielectric and conductivity values for the same soil. The specific values used for this figure are shown by the squares in Fig. 12.5. Because conductivity rises with frequency, the depth of penetration of high-frequency GPR is significantly reduced.

The presence of the earth-air interface has only a modest effect on these formulas constructed for propagation in a whole space. The right panel of Fig. 12.8 shows the amplitude radiation pattern of a dipole antenna at the interface between air and soil that has an index of refraction of 3 and a resistivity of 100 ohm-m. The presence of the interface causes the radiation pattern to vary with angle, visible in this plot as a slight modulation of the stronger radial decay.

12.4   Commercial services, systems, and sensors

There are over 20 million miles of underground pipelines, utility lines, and conduits in the United States (Sterling, 2000). Much of this civil infrastructure has been installed in the last 60 years, with original intended lifetimes of 50 years or more; for large portions of the network, maps or as-built plans are either inaccurate or non-existent. In the last decade, two major commercial markets have been built around the need to locate underground utility lines when planning their maintenance, upgradation, and protection. These two markets are the One-Call System and SUE.

12.4.1   The One-Call system for utility locating

Utility locating in the United States is a service consortium of companies that locate underground utilities before construction or digging. The consortium is organized under the One-Call, or 'Call-Before-You-Dig,' system mandated by The Transportation Equity Act for the Twenty-first Century, enacted by Congress on 9 June 1998 (TEA 21, Title VII, Subtitle C, Comprehensive One-Call Notification).

'One-Call' refers to a single number published in phone books that will access a free service to locate any buried utility. A call to the system, which is required by law before any digging, generates a locate ticket that is sent to a service provider who dispatches a technician to the site to locate buried lines and place appropriate markings on the ground. The service is free to the person making the call; its cost is built into utility rates.

One-Call is the largest commercial outlet for underground mapping services. The market has grown steadily since the passage of The Transportation Equity Act and is now estimated at over $1 billion per year, but it is regional and highly fragmented. Services are provided not only by companies that specialize in utility locating, but also by public or private utility companies themselves as well as by companies engaged in construction or excavation and in geophysical or engineering services. Accurate figures are hard to assemble, but it appears that no single company at present has more than 10% of the national market. Margins are low, and today there is relatively little differentiation in technology. Accuracy standards vary by state, but generally require that locators mark a corridor 45–60 cm wide above a utility line, with the line ideally located beneath the center line of the marked corridor. Specifying the depth to the line is generally not required, but is considered a best practice.

Prevention of costly, sometimes deadly, digging accidents was the motivation for creating the One-Call system. Comprehensive statistics on accidents involving all types of utilities are not available, but the American Gas Association has estimated that damage from *reported* digging accidents involving just major gas pipelines averaged about $40M per year from 1994 through 1999 (Kalisch, 2000). Highway construction in particular is a leading cause of damage to the nation's underground networks (see the NTSB Report 'Protecting Public Safety through Excavation Damage Prevention,' NTSB/SS-97/01). One-Call has helped to lower accident rates; but, when costs associated with construction delays caused by mislocated or unlocated lines are added to actual physical damage caused by accidents, the yearly tally probably still exceeds a billion dollars.

In addition to preventing accidents, there is now growing evidence that comprehensive 3D subsurface mapping during the planning phase of construction can significantly reduce construction costs through better design

and fewer change orders. Within the next decade, it is possible that a combination of new legislation setting higher standards for One-Call systems, pressure from insurance companies to better manage risks, and reductions in the cost of 3D mapping will move the utility locating market in the direction of the more comprehensive practices embedded in the ASCE SUE guidelines (ASC C-I 38-02). Such a move would stimulate more comprehensive mapping of utility networks with combinations of mobile sensors.

Several factors may drive this trend:

- *Crowded Rights of Way* – Underground rights of way are becoming crowded, and competition for limited space is growing. Installation of the 'last mile' of broadband Internet service and maintenance of existing infrastructure in cities will demand high levels of construction in already dense utility corridors.
- *Non-Conductive Utilities and Uncertainty in Locates* – Buried utilities now include non-conductive materials, such as plastic, glass and clay, which are nearly invisible to the metal detectors used in One-Call locating. Even conductive utilities, when densely packed, are vulnerable to misidentification with existing technology.
- *Need for Precision in Directional Drilling* – Directional drilling, also called trenchless technology, is being used more and more to emplace utilities in crowded corridors. Planning and executing a directional drilling plan requires an accurate 3D map of existing lines and other possible obstructions.
- *Lack of Permanent Records* – One-Call locating services provide only temporary surface markings, which are later washed away or obscured. To prevent this waste, some municipalities are adopting digital archiving of utility maps according to SUE standards.

12.4.2 Subsurface utility engineering

SUE is defined by the ASCE as 'a branch of engineering practice that involves managing certain risks associated with utility mapping at appropriate quality levels, utility coordination, utility relocation design and coordination, utility condition assessment, communication of utility data to concerned parties, utility relocation cost estimates, implementation of utility accommodation policies, and utility design' (American Society of Civil Engineers, 2002).

SUE combines civil engineering, surveying, and geophysical methods to obtain information about the location of underground utilities with different levels of reliability and accuracy (Anspach, 1995, 1996; Zembillas, 2003). The project owner determines the right balance between cost and risk. Knowing the capabilities and limitations of SUE methods and the associated risk allows for proper management of a subsurface construction project. Accurate

3D locations of buried utilities are useful not just to enhance safety, but also to save costs by expediting design and construction (Jeong *et al.*, 2004).

The technologies of SUE include geophysical remote sensing, surface surveying and mapping, computer-aided design (CAD) and geographic information systems (GIS). SUE engineers certify utility information in accordance with a standard classification scheme that allows a better allocation of risk between the project owner, project engineer, utility owner and construction manager. The end product is a complete utility map that can be integrated into the engineering design of a construction project.

The FHWA has been promoting the use of SUE since 1987 as a means to save costs on highway construction projects. Some states, such as Virginia, now require SUE before any highway construction. The original Purdue study commissioned by FHWA in 1996 examined 71 highway projects in four states and estimated savings from use of SUE at about $23 million, or 2% on the total value of the construction (about $1 billion). Extrapolating this figure to the $59 billion estimated for highway construction in 2004 suggests a possible savings of over $1 billion per year from systematic use of SUE. A later study by Jeong *et al.* (2004) found savings of about $12 in construction costs for each $1 of SUE expenditure. A more recent study by Jung (2012) examined 22 SUE projects and eight non-SUE projects from various districts of the Pennsylvania Department of Transportation and found similar results: $11.39 was saved for every $1 spent on SUE in selected road projects; a ratio of 1.65% was found as the ratio of SUE cost to total project cost. Construction projects that combine standard SUE practices with 3D geophysical mapping using a new generation of mobile sensors have found even more substantial cost reductions and safety enhancements (Goodrum *et al.*, 2008; Birken, 2009; Zembillas, 2008).

Today, SUE relies mainly on vacuum excavation to expose buried utilities with small potholes or test pits, at sites selected by studying existing maps or by locating utilities with handheld sensors. Location, depth, size and type of utility are also recorded by hand. New maps are then created by extrapolating information between holes or by filling the gaps with geophysical mapping along linear profiles spaced 10–20 feet apart. SUE contracts require that final results adhere to one of four standard quality levels as defined by the ASCE (2002):

- Quality level D (QL-D) consists of information derived from existing records or oral recollection. It is often not adequate to the comprehensiveness and accuracy required to eliminate risks and dangers of conflict with underground infrastructure. This quality level is used in early planning of route selection and in estimating utility relocation costs.
- Quality level C (QL-C) consists of information obtained by surveying and plotting visible above-ground utility features and by using
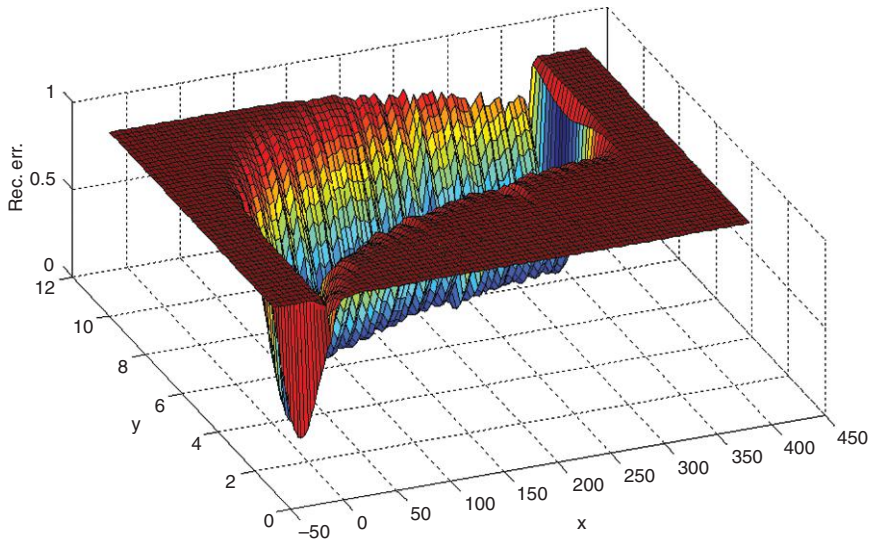
Plate I (Chapter 3) Reconstruction error using KPCA σ = 1.5, *pc* = 180.
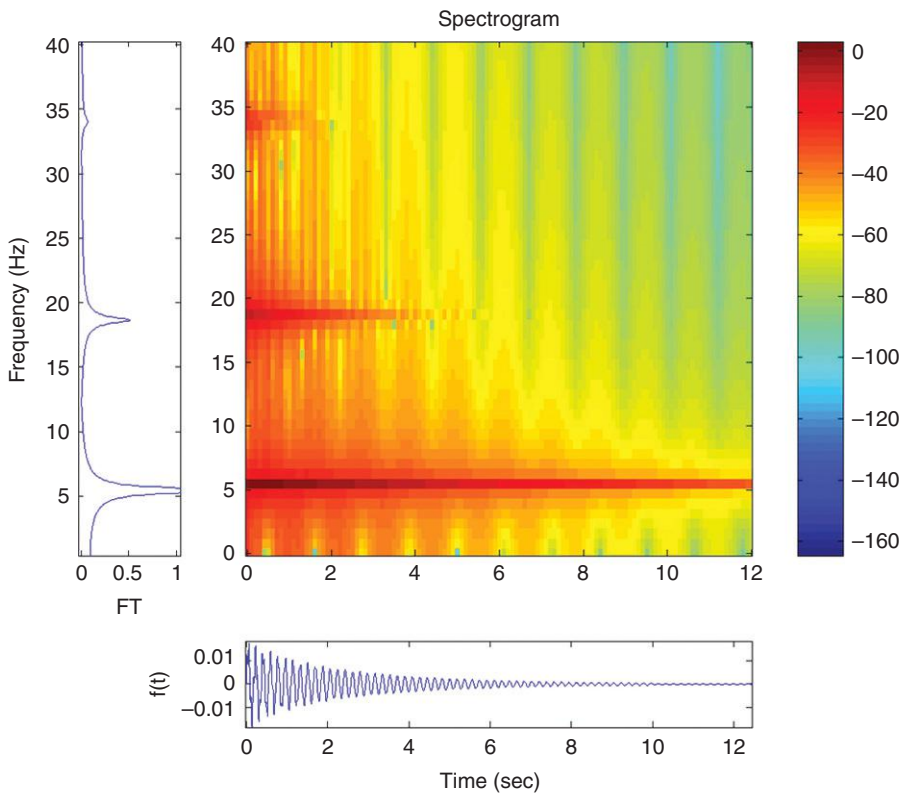
Plate II (Chapter 4) STFT of the measured third floor free-vibration displacement response.
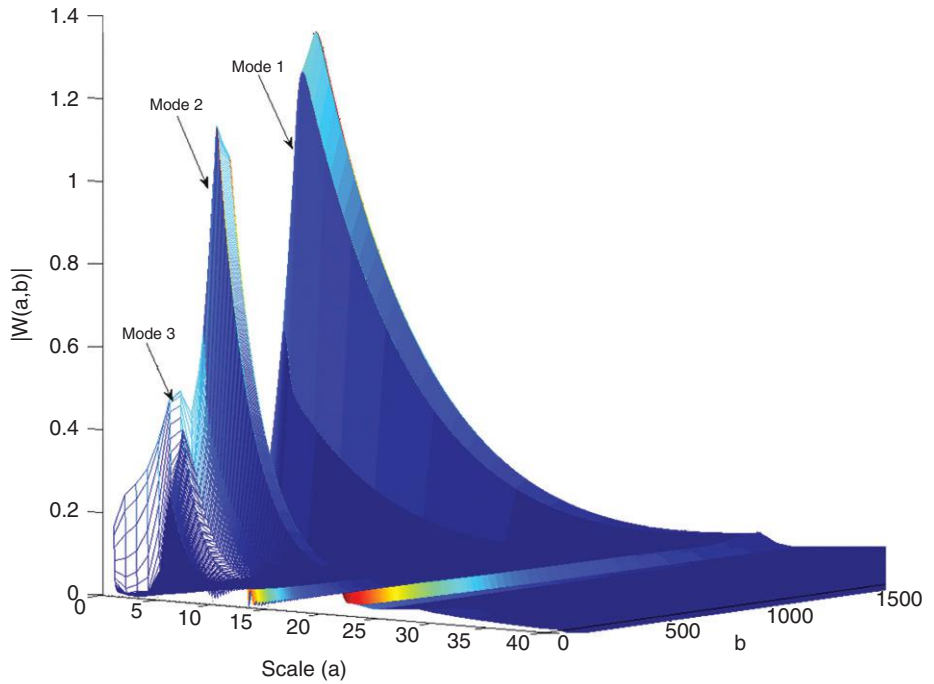
Plate III (Chapter 4) Scalogram (view from below the x-y plane showing existence of three modes and free vibration decrement) of the measured third floor free vibration.
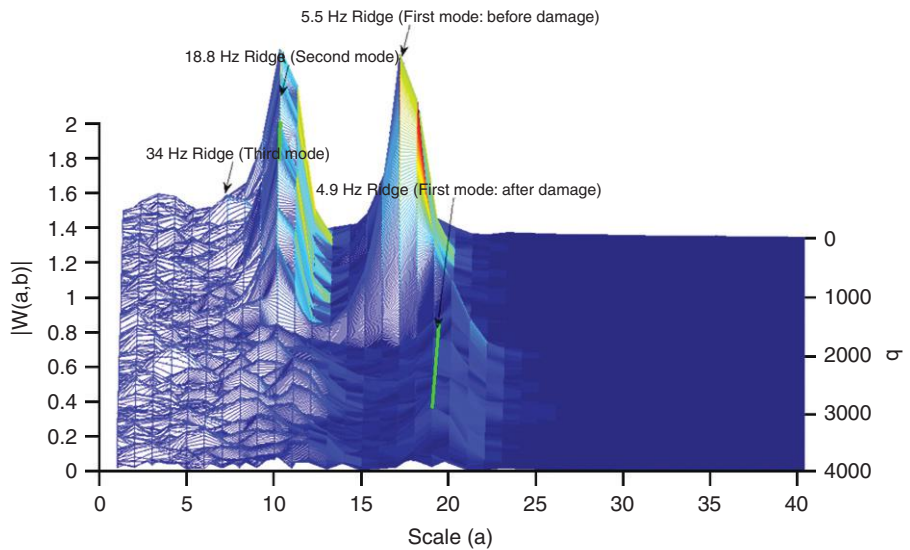
Plate IV (Chapter 4) Scalogram of third floor acceleration response: note the shift in the first mode. Frequency (Ridge) of 5.5 –4.9 Hz after 10 s due to damage.
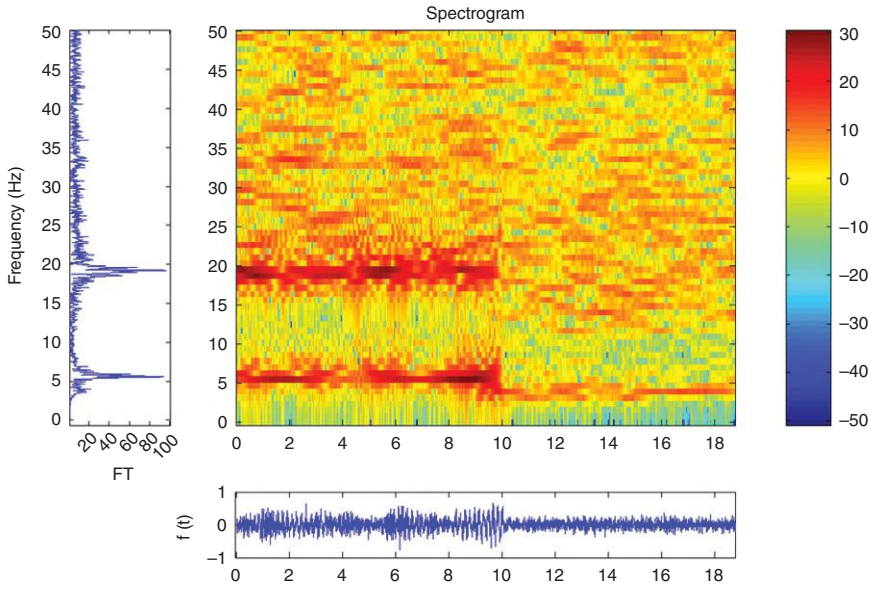
Plate V (Chapter 4) Spectrogram of the third floor acceleration response to white noise excitation.
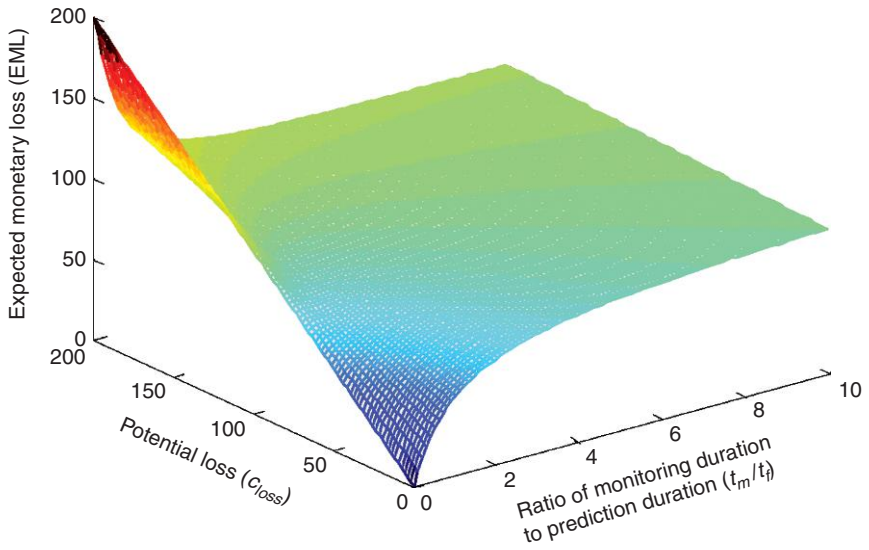
Plate VI (Chapter 5) (a) Relation among t_m/t_f, EML and c_{loss}.
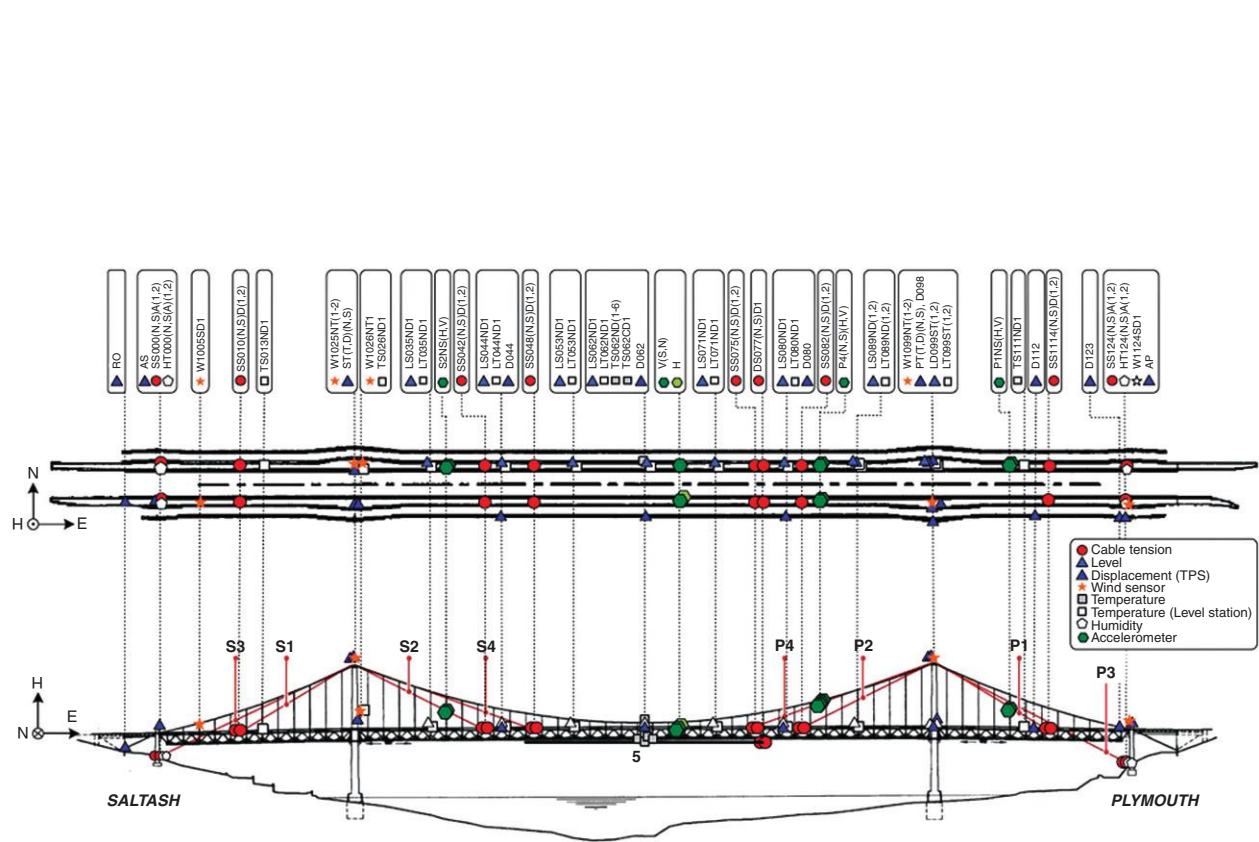
Plate VII (Chapter 7) Tamar SHM system overview.
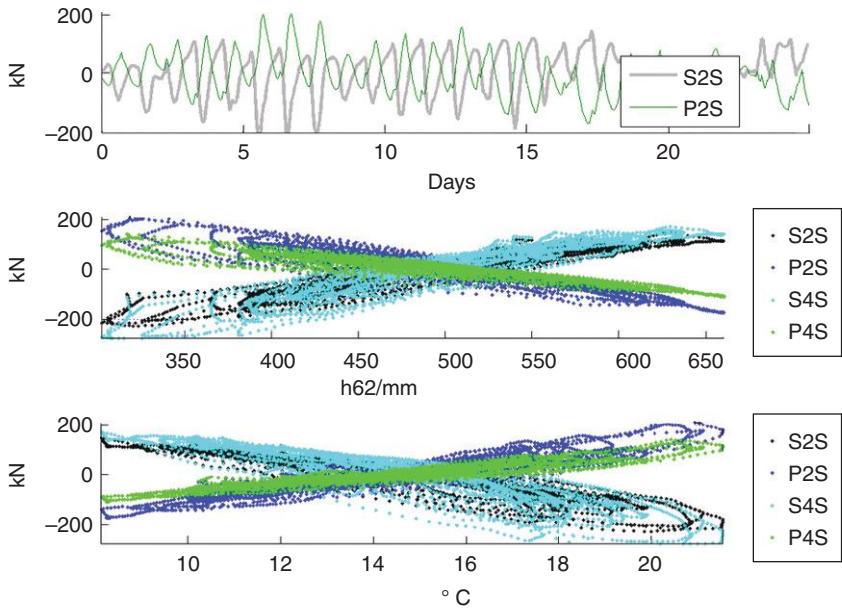
Plate VIII (Chapter 7) correlation of stay cable tension with deck deflection and temperature.
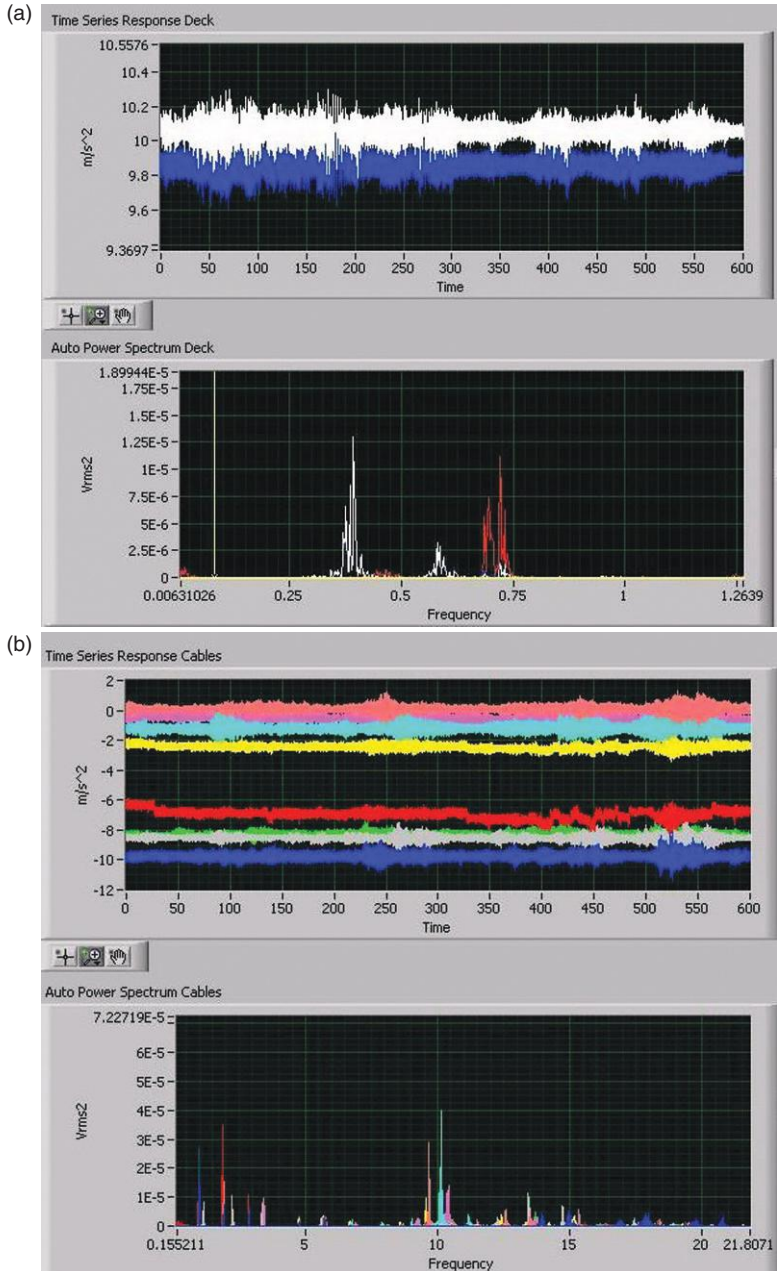
Plate IX (Chapter 7) Deck (a) and cable (b) acceleration signals in time and frequency domains.

(a)

(b)
01 Jul 2010 – 31 Dec 2010
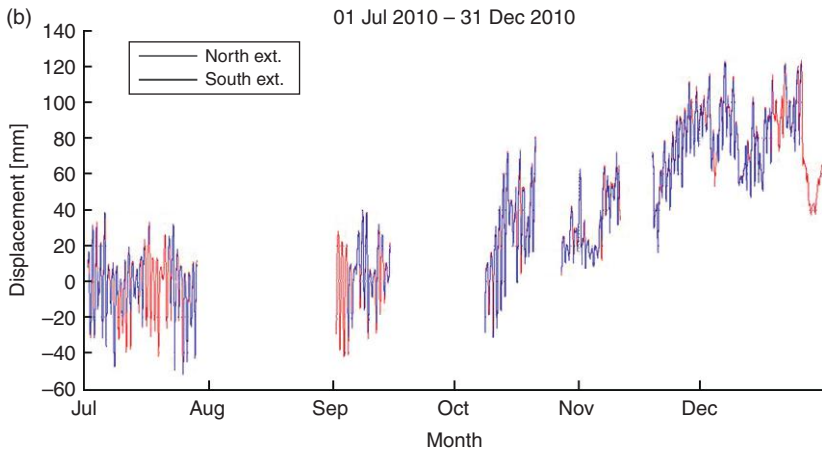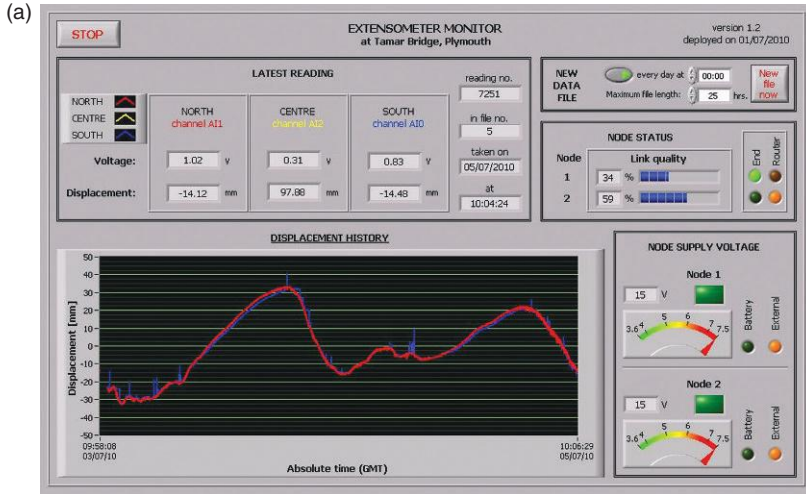
Plate X (Chapter 7) WSN monitoring interface and extensometer data.
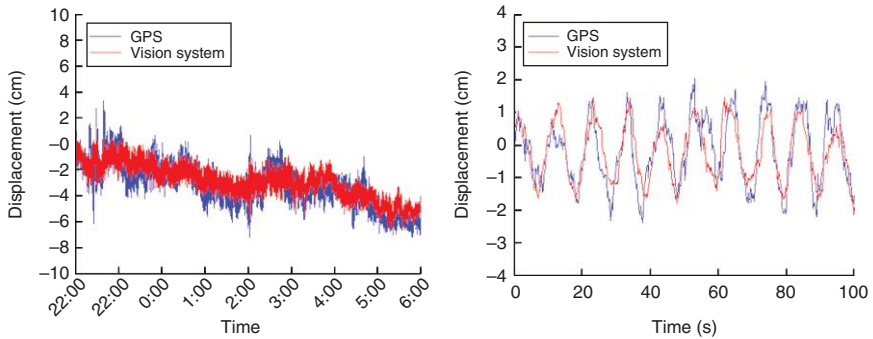
Plate XI (Chapter 9) Comparison of monitoring data obtained by GPS and VIS.
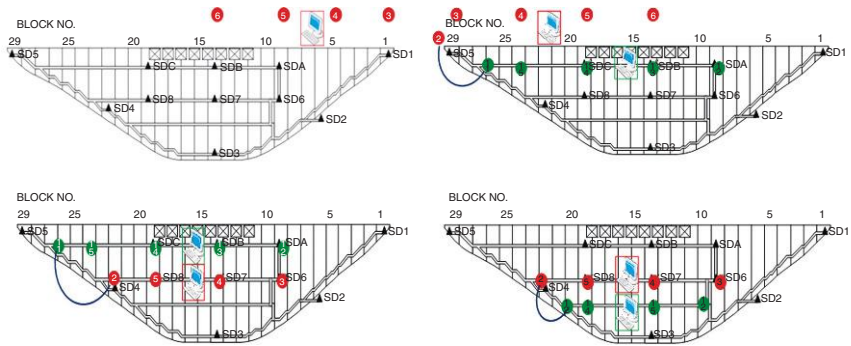
Plate XII (Chapter 10) Four test step on the ambient vibration survey of the dam.
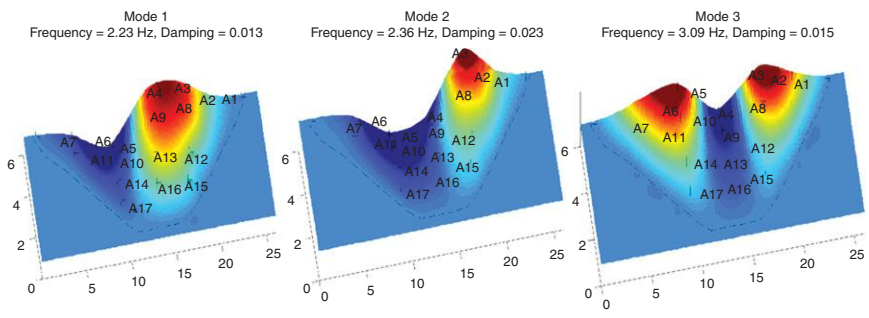
Plate XIII (Chapter 10) Identified three modes of Fei-Tsui arch dam from ambient vibration data.
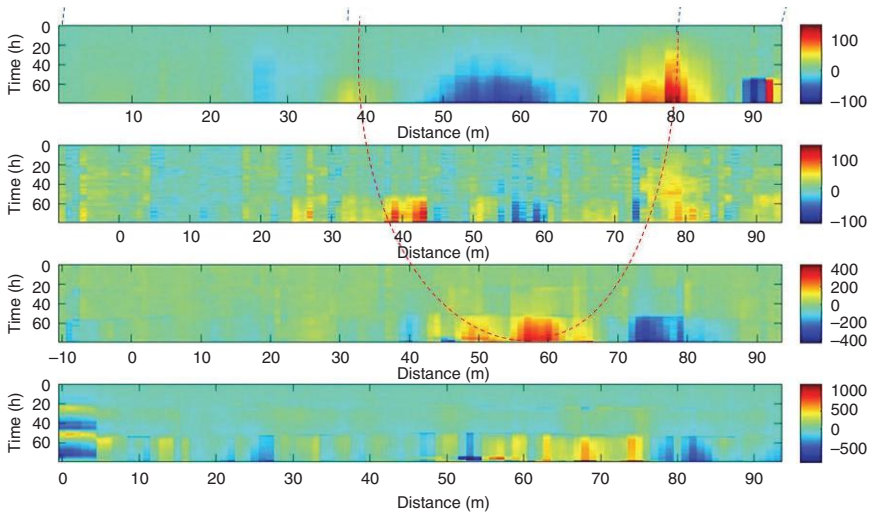
Plate XIV (Chapter 13) Strain Peaks recorded at the boundaries of the failure zone by distributed optical fiber sensors, up to 42 h before failure. (*Source*: Image courtesy of Tencate, geophyConsult and Deltares, Artières *et al.*, 2010.)
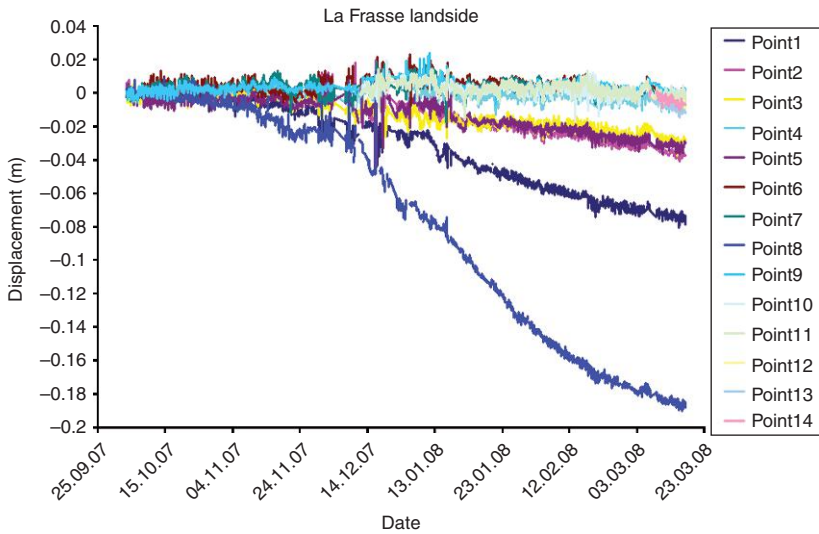
Plate XV (Chapter 13) Landslide reactivation in November 2007. (*Source*: Picture courtesy SMARTEC SA and EPFL).

Plate XVI (Chapter 17) Photo of the MIT-Bluefin hovering autonomous underwater vehicle. (*Source*: Englot *et al.*, 2009.)
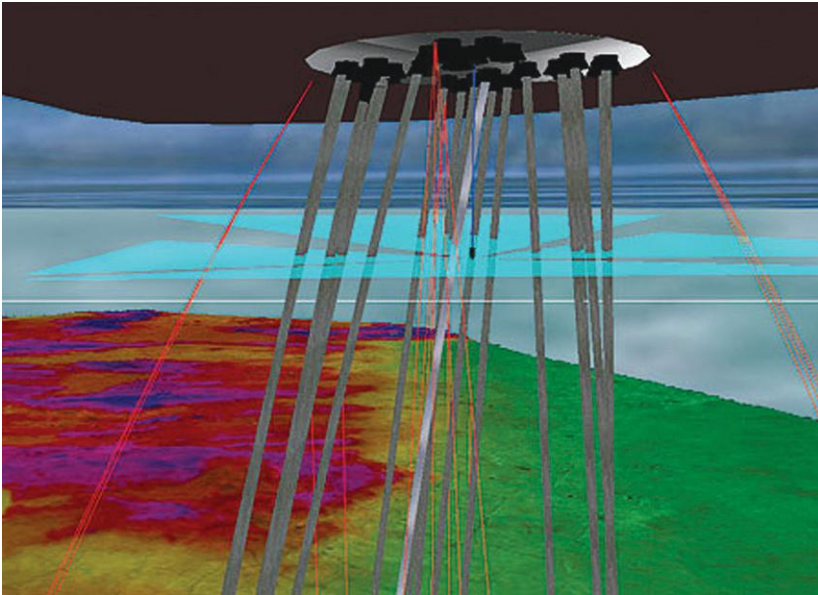
Plate XVII (Chapter 18) Illustration of riser and anchor chain monitoring system.[6]
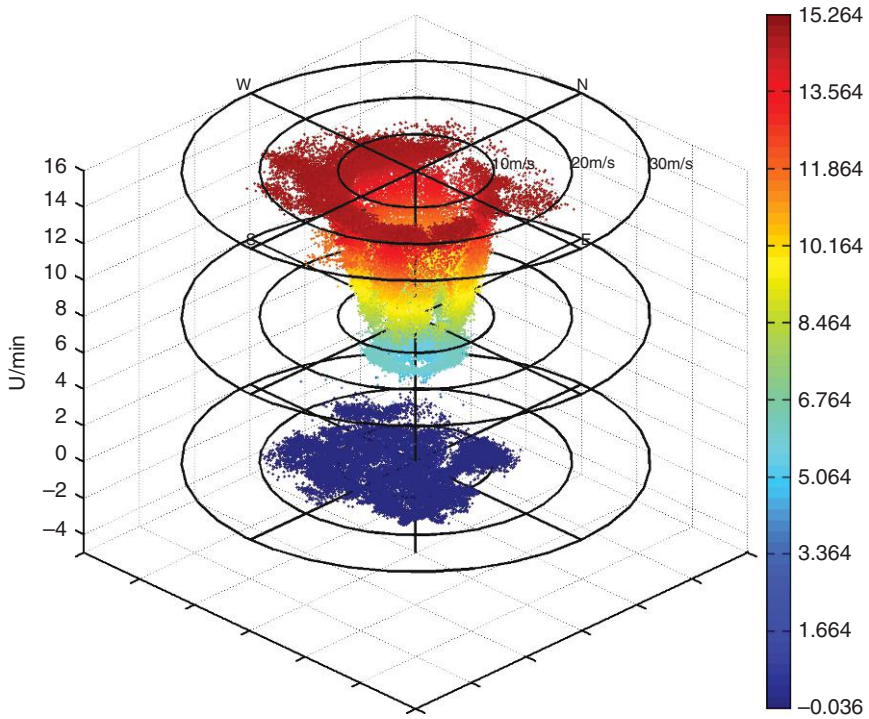
Plate XVIII (Chapter 19) Wind direction (angle) and – wind speed (radius) in combination with rotor speed (z-axis), each point is a 10 min mean.
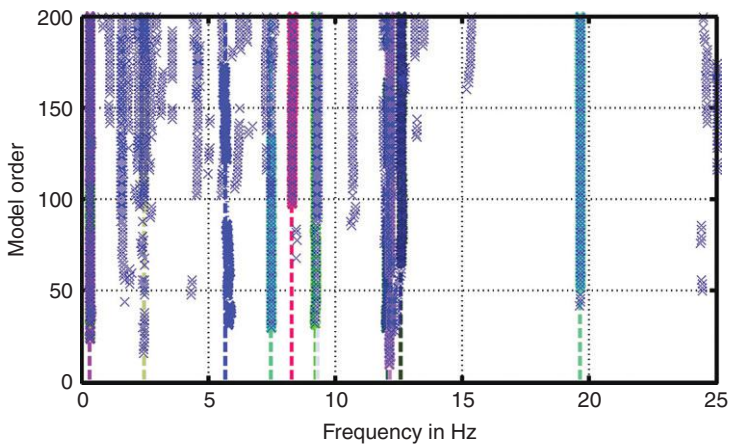
Plate XIX (Chapter 19) Stability diagram for SSI solutions over 200 model orders (one x for each solution) with identified stable frequencies, indicated through colored lines. (Source: Rolfes et al., 2012.)
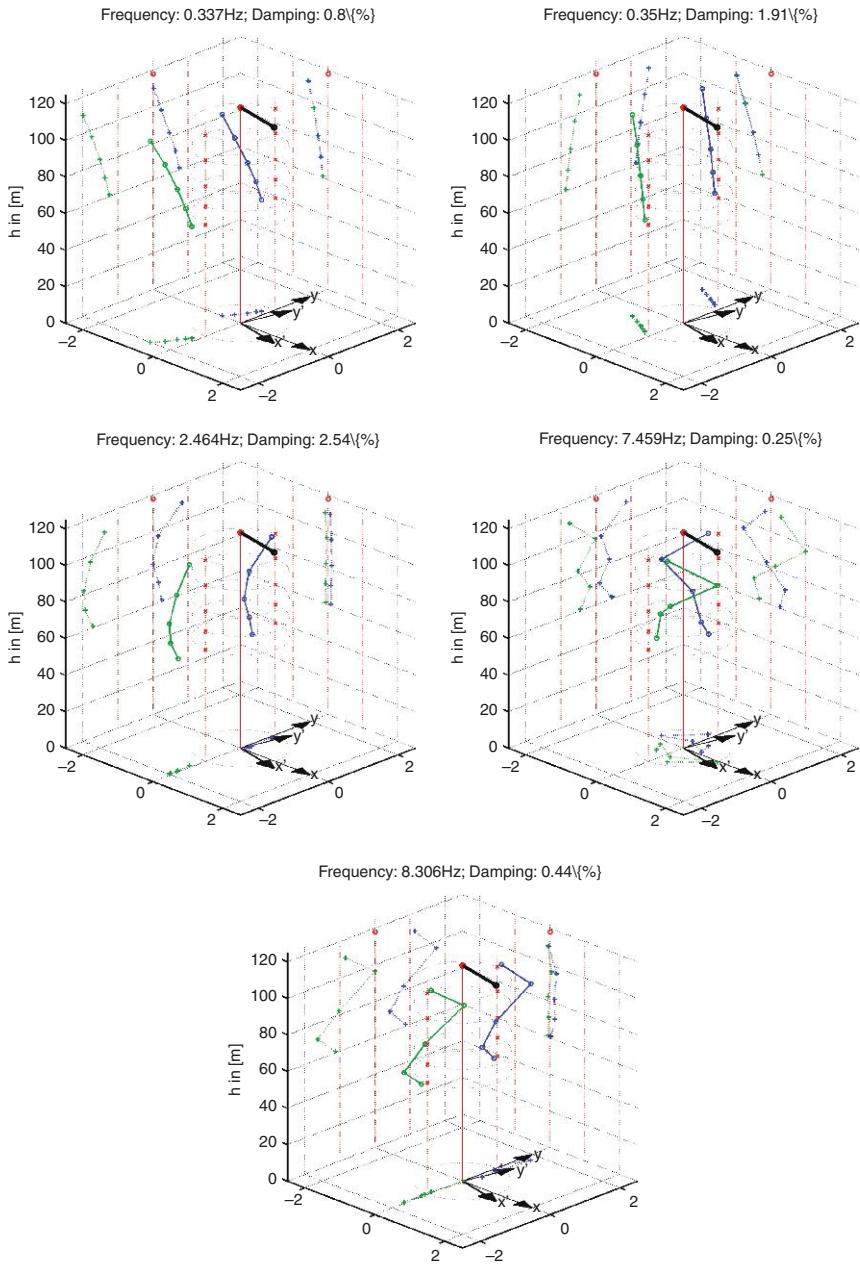
Plate XX (Chapter 19) Three-dimensional visualization of global eigenvectors. One line is composed of all southern sensors (green), one of all northern (blue). The nacelle position is indicated by a rotated coordinate system at the bottom and a black line on top. (*Source*: Rolfes *et al.*, 2012.)

professional judgement in correlating this information to QL-D information. This level has been traditionally used for design.

- Quality level B (QL-B) consists of information obtained through the application of appropriate surface geophysical methods to determine the existence and approximate horizontal position of subsurface utilities. Quality level B data should be reproducible by surface geophysics at any point recorded. This information is surveyed to applicable tolerances defined by the project and incorporated into plan documents.

- Quality level A (QL-A) provides precise horizontal and vertical locations of utilities obtained by actual exposure and subsequent measurement of subsurface utilities at specific points, e.g., at the four corners of an intersection. The specification can also be achieved by verification of previously exposed and surveyed lines. The 3D data of location and other attributes of each exposed line (type, material etc.) are shown on plan documents. Accuracy in vertical location of an exposed utility is typically required to be ±15 mm. Accuracy in horizontal location is set by the mapping accuracy levels defined or expected by the project owner.

The SUE standard thus requires a large jump in accuracy from QL-B, where positions are approximate, to QL-A, where vertical and horizontal tolerances of 15 mm or better are required. At present, these tolerances can be achieved only at spot locations after exposure of the utilities by hand digging or vacuum excavation. Figure 12.9 shows the tradeoff in risk reduction versus cost for the different SUE quality levels. Noninvasive geophysical technologies can provide vertical and horizontal accuracies ranging from about 5 to 15 cm; this is good enough for a 'B+' rating, but does not reach SUE Quality Level A. Geophysical techniques can, however, cover large areas more cost effectively than vacuum excavation. With systematic calibration at vacuum excavation sites, the accuracy of geophysical mapping may eventually approach the QL-A standard (Birken, 2009).

12.4.3   Utility locating equipment

The most widely used utility or pipe locators are handheld EMI units using passive (or quasi-passive) detection and pushcart GPR units with a single transmitter–receiver antenna pair (Fig. 12.10). Induction locators, which can trace most conductive utility lines and estimate their depth, dominate One-Call services. During the last 5 years, GPR systems have penetrated the SUE market, with their ability to locate both conductive and non-conductive lines and provide more accurate depth estimates.

	Required utility location accuracy	
	Vertical	Horizontal
QL-A	±15 mm	< 15 mm
QL-B+	±15 cm	±15 cm
QL-B	some	±60 cm*
QL-C	NA	NA
QL-D	NA	NA

*varies by state
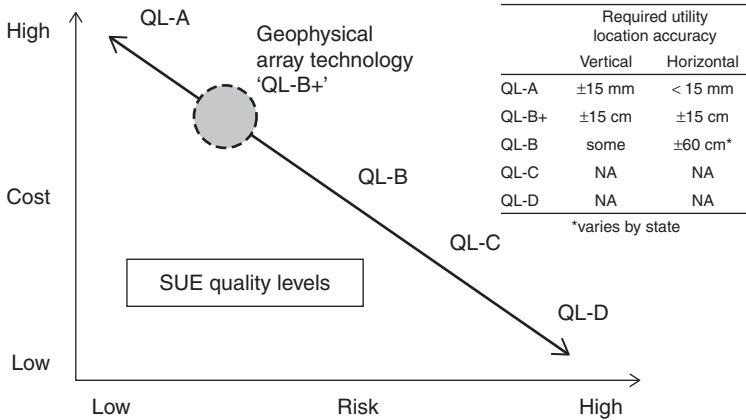
12.9 Risk versus cost for existing SUE quality levels.

EM pipe locators

The basic components of a quasi-passive EM locating system are a current generator and induction coils. Commercial manufacturers offer a large variety of locators: from basic single-frequency systems with one current transmitter and a few tuned receiver coils to sophisticated multi-frequency systems with multiple current transmitters and broadband multi-coil receivers. High-end systems have integrated GPS and antennas for transmitting and receiving signals from RFID markers buried with utility lines (North, 2010).

Single-channel GPR systems

Three manufacturers dominate the worldwide market for commercial single-channel GPR systems: GSSI, Sensors and Software, Inc., and Malå Geoscience. GSSI introduced the first commercial GPR system in 1974; advances over the last 40 years have included improvements in electronics and antenna efficiency, better on-screen displays and control software, and more advanced signal processing. One of the most important improvements was the incorporation of shielding on top of the antennas to ensure that most of the EM energy is radiated into the ground. Introduction of push-cart-mounted GPR systems – first done by the Canadian company Sensors and Software in 2000 – improved logistics dramatically for operation on city streets. Figure 12.10 shows a typical pushcart system.

GPRs are classified as ultra-wideband (UWB) radar systems. According to FCC guidelines (FCC, 2002), UWB refers to any radio technology whose frequency bandwidth exceeds either 500 MHz or a value equal to 20% of the system's center frequency. By comparison, air traffic control radar,
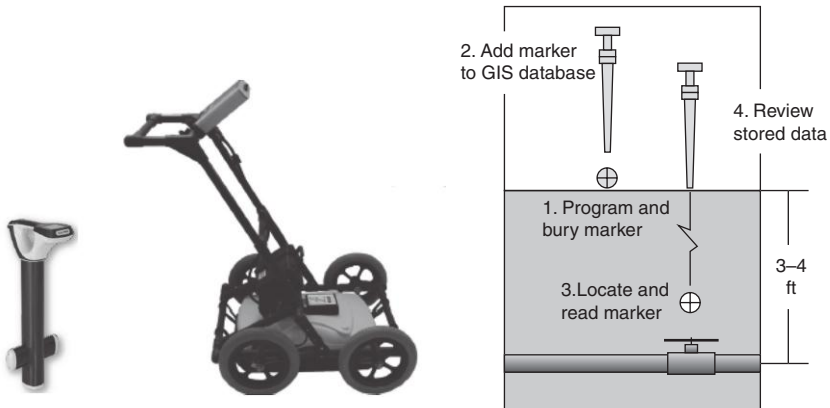
12.10 Locating technologies for One-Call. (Left to right) Inductive magnetic field sensor (vivax-metrotech.com); pushcart-mounted GPR (Sensors and Software, gprlocates.com); and RFID marker and reader (3m.com).

which operates at 1030 MHz, has a bandwidth of only a few percent. The top right panel of Fig. 12.6 shows an idealized transmitted pulse and spectrum of an UWB GPR with a central frequency of 200 MHz. Modern systems are available at center frequencies ranging from about 100 MHz to 3 GHz. A single radar control unit can usually be used over the full range of center frequencies, but antennas of the right size and shape must be used in different frequency ranges. Some systems have integrated cross-polarized antennas, where each antenna package contains dipole antennas in two orthogonal orientations. Full system specifications, along with applications and case histories, can be found on the manufacturer websites.

12.4.4   Services

Many engineering companies provide commercial utility locating services outside the One-Call system. SUE companies themselves often provide such services, but land-surveying and large construction companies are also active in this service market. Many service providers have the full range of handheld locating equipment available, and choose the best tool for the particular situation depending on the budget and complexity of the project. Whereas most One-Call services make only temporary paint marks on the ground, SUE companies and other service providers will often create digital CAD drawings of the located utilities in real-world coordinates, complemented by street maps or aerial photography. These maps can be used for project design or future needs in the same area.

12.5   Mobile sensor arrays

The early 2000s saw the introduction of multi-channel data acquisition systems to control EMI and GPR sensor arrays. The first systems consisted of single-channel systems mounted side by side on pushcarts. Other early systems were a 6-channel GPR array (RIS MF) from IDS and a 3-channel EMI array (EM61-MK2) from Geonics. These were followed by integrated multi-channel systems designed to be moved by a vehicle at low speed in traffic. The first commercial integrated system, called the CART Imaging System, was a 16-channel GPR jointly developed by Malå Geoscience and Witten Technologies. The CART system, which could be operated with either 200- or 400-MHz antenna sets, participated in a series of large-scale mapping projects for Consolidated Edison Company of New York (Con Edison) and other major utility companies from 2000 to 2004. One project, sponsored by a grant from the government of Västerbotten County, Sweden (where Malå has its headquarters), helped in the restoration of the utility network of Lower Manhattan after 9/11 (Birken *et al*., 2002; see the field examples below).

Other GPR arrays introduced in the last 10 years include a 14-channel 400-MHz array called Terravision from GSSI; the MIRA family of systems from Malå; and high-channel, mixed-frequency systems from 3d-Radar (Eide and Hjelmstad, 2002) and IDS Company (Francese and Morelli, 2006). Figure 12.11 shows examples of second-generation GPR arrays from the latter two manufacturers. Because antenna efficiency at radar frequencies depends strongly on size, each system uses interleaved antennas of different sizes to cover a broad frequency range. The STREAM array from IDS uses a pair of 15-channel 200-MHz arrays, offset by half the antenna spacing, to cover a swath 1.74 m wide with profiles 6 cm apart; a separate array covers a 1.2 m swath with eight profiles created from four 200-MHz and four 600-MHz antenna pairs. The Geoscope array from 3d-Radar uses interleaved triangular patch antennas to cover a 2.4 m swath with 200, 400, and 800 MHz, using 7, 8, and 16 antenna pairs, respectively (in the figure, a light triangle is a transmitter; a dark triangle is a receiver). All values refer to the central frequency of these broadband antennas.

At the time of writing, the GSSI Terravision and IDS STREAM systems were the only GPR arrays licensed for operation in the US under the FCC guidelines for UWB radar introduced in 2002 (FCC, 2002). New commercial EMI arrays include the 3-channel GEM-3 and 7-channel GEM-5 systems from Geophex (geophex.com).

12.5.1   Dual-array system

From 2003 to 2007, the Research and Special Programs Administration of the US Department of Transportation sponsored a project to design and

develop a mobile system with both radar and induction sensors for mapping underground utilities. Key goals for this 'dual-array' system were the capabilities to operate in most soil conditions, to cover large areas efficiently, and to map both plastic and metal utility lines, including large conduits at depths of 10 m or more (Birken *et al.*, 2005). The GPR array in the dual-array system was a version of the CART Imaging System. The EMI array consisted of 16 vector magnetometers (induction coils) with a flat frequency response from 1 to 100 kHz. The system could work in quasi-passive mode with clamp-on transmitters that injected current at discrete frequencies onto individual pipes, using galvanic or toroidal clamps, or in active mode using a 3-axis induction coil that rode with the transmitters and operated over the same frequency range as the sensors.

The main new hardware developed for the dual-array system was a broadband vector EM sensor (Fig. 12.12, left) that uses current-feedback induction coils to measure the magnetic field at frequencies from nearly DC to about 100 kHz (Conti, 1992). The response of the coils is actually flat from about 1 to 100 kHz and falls off linearly outside this band. (The original coils had a flat response starting at 100 Hz and were later modified to reduce their sensitivity at low frequencies, so as not to be overwhelmed by ambient 60 Hz noise.) Each of the 16 sensors in the EMI array was a three-axis vector magnetometer. To survey simultaneously with all 48 channels, a special data acquisition system was assembled from commercial components to record the fields as time series that were Fourier analyzed to extract various frequency components. The 16 sensors were arranged in two rows of eight sensors, offset vertically by 0.5 m; each row was 2 m long (Fig. 12.12, center). The trailer housing the sensors was fabricated from non-conductive fiberglass (Fig. 12.12, right).

12.5.2 Positioning

Array systems require accurate positioning to realize maximum efficiency in large-scale mapping of infrastructure. Fortunately, development of mobile sensor arrays has been accompanied by tremendous strides in the accuracy of real-time positioning at street level. For example, two widely available commercial positioning systems can provide the cm-level accuracy needed for construction of seamless 3D GPR images over large complex areas (Birken *et al.*, 2002): robotic total stations (also called laser theodolites) and real-time kinematic GPS (RTK-GPS) systems. Although EMI arrays require less dense spatial coverage to avoid spatial aliasing, horizontal positioning to ±15 cm or better is necessary to reach the highest SUE quality standards.

Laser theodolites measure the time-of-flight and the directional angles of infrared laser pulses traveling between a base station and a reflecting prism mounted on the object to be tracked. These systems can measure linear

(a)

(b)
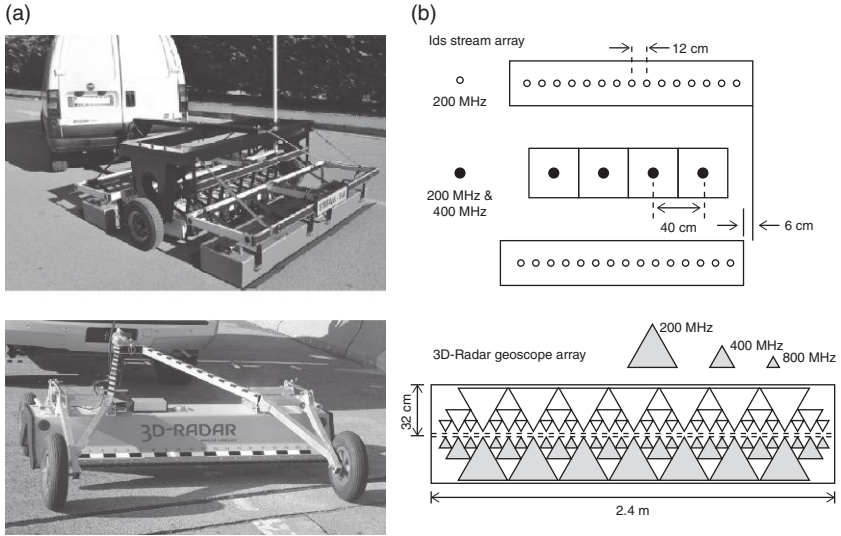
12.11 Second-generation GPR arrays. (a) Photos of STREAM system (www.idscorporation.com) at top and of Geoscope system (www.3d-radar.com) at bottom. (b) Schematics of antenna arrays.
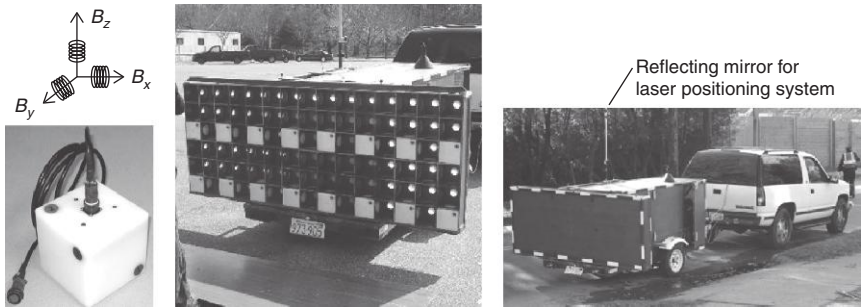
12.12 Magnetic field sensors in the dual-array system. (Left to right) Individual three-axis magnetometers; sensor layout; full system configuration for surveying.

distances with an accuracy of a few millimeters over distances of a kilometer or more and can measure angles to about 1 arc-second. If accurate RTK-GPS is available, robotic total stations with integrated GPS can track a mobile data acquisition platform with sub-cm accuracy in both local and global coordinates.

Integrated navigation systems combining GPS with inertial measurement units and independent distance measurement instruments provide continuous, accurate absolute positioning even in areas unfavorable to

GPS. Examples are the Trimble Applanix POS LV (applanix.com) and the TOPCON IP-S2 (topconpositioning.com). While standard GPS has become a commodity consumer technology, RTK-GPS and laser theodolites cost $100 k or more, roughly the cost of a second-generation GPR array.

12.5.3   Merging sensor data and array positions

There are several different ways of merging sensor and positioning data. The approach described in Burns *et al.* (2004) allows the surveying instrument and sensor array to operate independently. In this method, a total station records the 3D position of a fixed point on the array at frequent intervals along its path, while data collection by the sensor array is triggered independently by a counter wheel that sends a command to the array controller at fixed intervals along the path – for example, every 10 cm. A processing workflow merges the two data streams as follows (Fig. 12.13). An algorithm first computes the 3D paths of the vehicle from positioning data obtained by tracking the prism attached to the sensor platform. A continuous path for the vehicle is estimated by fitting a smooth curve to the collection of 3D locations of the tracked point. The algorithm then uses the fixed geometry of the sensors with respect to the tracked point, along with the distance between trigger points, to distribute data from the sensors over the swath covered by the array. When surveying hilly terrains, 3D locations mapped by the total station must be supplemented by data from tilt sensors to properly locate the sensors relative to the tracked point (MacIntosh *et al.*, 2006).

In an optional final step, sensor locations, and accompanying data, are interpolated onto a regular grid in rectangular chunks tiling the survey area. This last step facilitates subsequent processing, such as SAR focusing or inversion of induction measurements. Moreover, when dealing with large irregular survey areas, this last step yields blocks of data convenient for storage and retrieval. Lehmann and Green (1999) and Grasmueck and Viggiano (2007) describe other approaches to positioning suitable for small, regular areas. The locations of surface features, such as curb lines, manhole covers, and valve covers, are also mapped using the total station, so that sensor data can be located relative to street features.

Figure 12.14 shows one of the first large GPR images made by combining mobile sensors with accurate positioning. Witten Technologies constructed this underground image in December 2000 during a demonstration project for Con Edison; the image covers about 250 000 sq ft beneath intersecting streets in the Bronx, New York. The raw data were collected in dozens of independent passes of the array along the streets during several nights of surveying with a 200 MHz CART system. The final image was produced

(a)   (b)   (c)
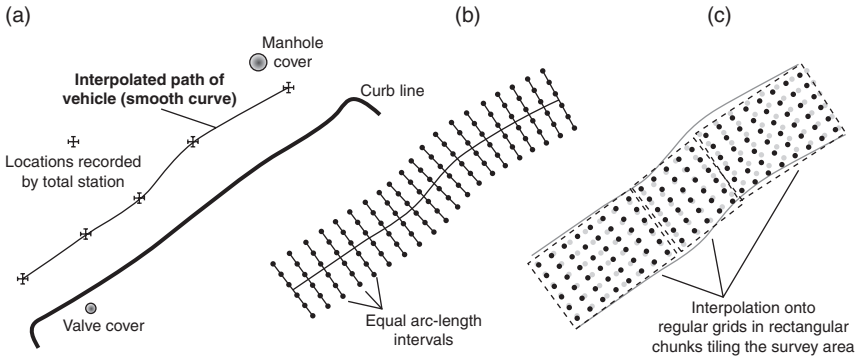
12.13 Method of merging geometry and sensor data to accurately locate measurements in absolute and relative coordinates (Burns *et al.*, 2004).

in two post-processing steps after all the data were collected. First, the data were positioned in street coordinates using the algorithm described above; second, the raw data were synthetically focused into a planar slice at a depth of 12 inches below street level. Each pixel in the focused image receives contributions from sensor locations that are within about 6 m of the imaged point. The complete 3D image consists of 72 horizontal slices created at 1-inch depth intervals in separate focusing steps. Figure 12.19 shows the image slice at 12 inches depth in perspective view; its seamless appearance highlights the accuracy of the positioning method. Buried trolley tracks, utilities, trenches and test pits, clearly visible at this depth and in the rest of the image, illustrate the power of synthetic GPR imaging (Oristaglio *et al.*, 2001).

12.5.4   Other subsurface remote sensing technologies

A variety of quasi-passive methods for general geophysical mapping exploit the broad spectrum of natural EM signals near Earth's surface (Vozoff, 1985; Nabighian, 1988). These natural fields include long-period oscillations of Earth's magnetosphere caused by its interaction with the solar wind; magneto-telluric plane waves that penetrate the earth as they propagate through the atmosphere following lightning strikes (there are about 100 strikes every second worldwide); and very low frequency (VLF) man-made transmissions used to communicate with submarines. These natural fields have not yet been widely used for infrastructure mapping.

   For projects involving deeply buried pipelines, or in areas where surface access is difficult or impossible, remote sensing technologies from the

GPR image 12 inches below street level

Trench carrying 138 kV
electrical feeder line

Buried trolley tracks

Prospect Avenue

East 149th Street

Paved over
test pits

Gas main

Courtesy of ConEdison

50 ft

Southern Boulevard
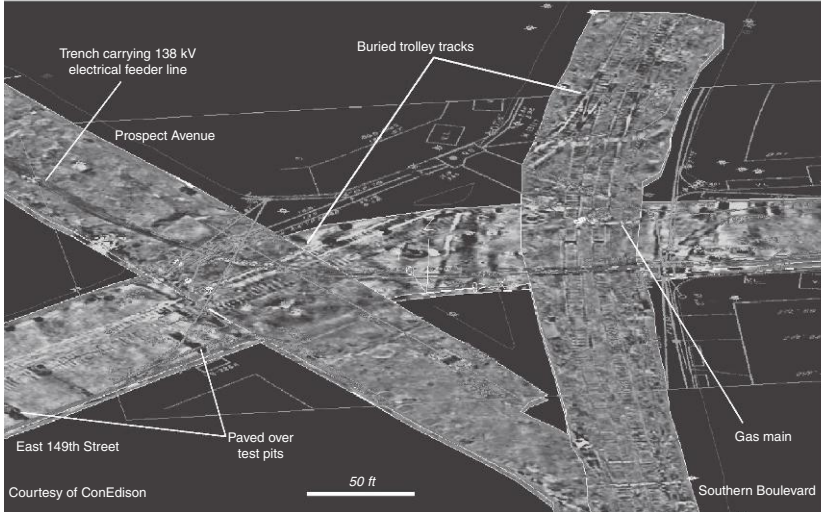
12.14 GPR image of buried utilities surveyed near the intersection of 149th Street, Prospect Avenue and Southern Boulevard in the Bronx, New York.

surface can be supplemented by inertial sensors that crawl or are dragged within the pipeline itself (see geospatialcorporation.com for examples).

12.6   Survey examples

The collapse of the World Trade Center on 11 September 2001 caused extensive damage to the underground utility network of Lower Manhattan (*New York Times*, 5 December 2001). Much of the city south of Canal Street was without electricity and telephone services for nearly a week after the collapse; water and sewer services were also interrupted for several blocks surrounding the site. Service was restored to surrounding areas, including Wall Street, within about one week by running high-voltage feeder cables and wires above-ground to shunt the network around the damaged zone. In all, Con Edison laid 32 miles of cable on downtown streets. All of this cable

eventually had to go underground, along with new water, sewer, gas, and telecommunications lines. From November 2001 to March 2002, the authors participated in a project involving Witten Technologies, Malå Geoscience, Con Edison, and City of New York Department of Design and Construction to help the reconstruction effort by mapping the underground utility network near the Trade Center quadrangle.

A GPR survey with the CART system covered about 400 000 sq ft (37 000 sq m) in 120 h of street work that was spread over 6 weeks and performed mainly at night between 10 p.m. and 4 a.m. The raw data were processed into a continuous high-resolution 3D-radar image descending to a depth of about 6 ft (1.5 m) below street level. Figure 12.15 shows horizontal slices through the image at depths of 24 and 42 inches near the intersection of West and Liberty Street at the southwest corner of the Trade Center quadrangle. Shallower depth slices mainly show features in the roadbed and surface valves and manhole covers.

At 24 inches depth, gas and electric lines are visible in the image; at 42 inches, a network of intersecting water lines is present. The images were used by Con Edison to correct existing infrastructure maps; most of these maps were found to have either wrongly located or omitted numerous underground lines. The corrected maps helped to find clear lanes for placement of new lines and conduits and to determine the best locations of test pits that were dug by hand to directly verify existing lines.

Figure 12.16 shows an example of high-resolution GPR mapping for utility maintenance. In 2003, the City of Miami was faced with the task of maintaining a 24 in diameter water main whose bell joints were found to be leaking along a 2 mile length. Since the joints were irregularly spaced and existing maps were ambiguous, it would have been necessary to dig hundreds of test pits along the length of the pipe just to locate its joints. The images in Fig. 12.16 are horizontal and vertical sections through a 100 ft section of the line; the sections were extracted from a 3D-radar image created over the entire length of the water line from data collected during 3 nights of surveying. Both the joints and the leaks stand out in the radar image. Further analysis allowed setting of priorities to replace immediately those joints that were leaking at the time of the survey.

One of the most important practical advances since the introduction of mobile sensor arrays has been the combination of geophysical mapping techniques with SUE practices, making it possible to create engineering quality CAD drawings from GPR images. Figure 12.17 shows an example from a project performed in downtown Tampa, FL, by Witten Technologies in collaboration with Craig A. Smith & Associates, an engineering firm based in Florida.
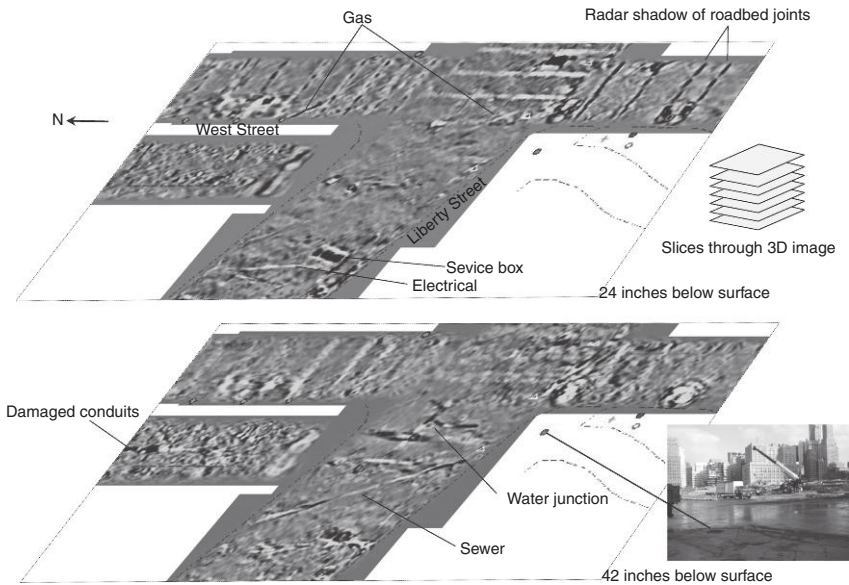
12.15 Slices through a 3D-radar image collected in January 2002 near the intersection of West and Liberty Streets in the World Trade Center quadrangle. (Photo looking NW at lower right.)

12.6.1 Combining GPR with EM induction (EMI)

The first large-scale field test of the dual-array system, combining radar and induction sensors, was carried out in November 2003 through April 2004 for Regional Water Authority of South Central Connecticut. The goal was to map a network of water pipelines emerging from a pumping station near a reservoir in Hamden, CT. The survey covered a triangular grassy area of about 22 000 sq ft (2043 sq m). Radar data were collected with a 200 MHz CART array; broadband EMI data were collected by injecting current on utility lines with transmitters operating at frequencies of 512, 8192, 9820, and 32 770 Hz.

The right panel of Fig. 12.18a shows a slice through the 3D-radar image at a depth of 24 inches. The most prominent feature in this image is a portion of a shallow water line that runs nearly N–S at the southeastern edge of the survey. Figure 12.18b shows maps of different components of the magnetic field recorded over the same area as current sources operating at different frequencies were connected to different access points of the water network. Symbols mark the locations of telephone poles, light posts, and other surface features. EM data were collected data in 56 individual profiles; the spacing
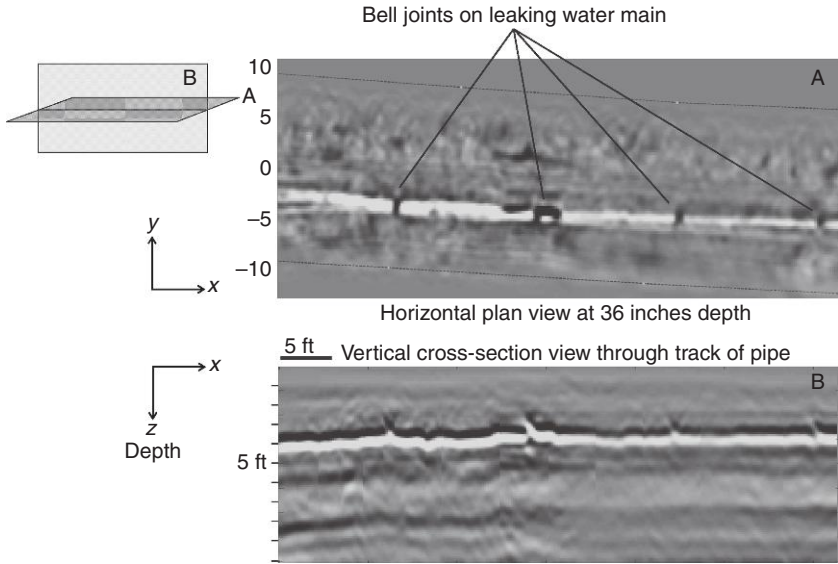
12.16 GPR mapping for utility maintenance. Horizontal and vertical sections (plan and profile views) through a 3D-radar image collected in Miami Beach, FL, show a 36-inch. water main with leaky bell joints.
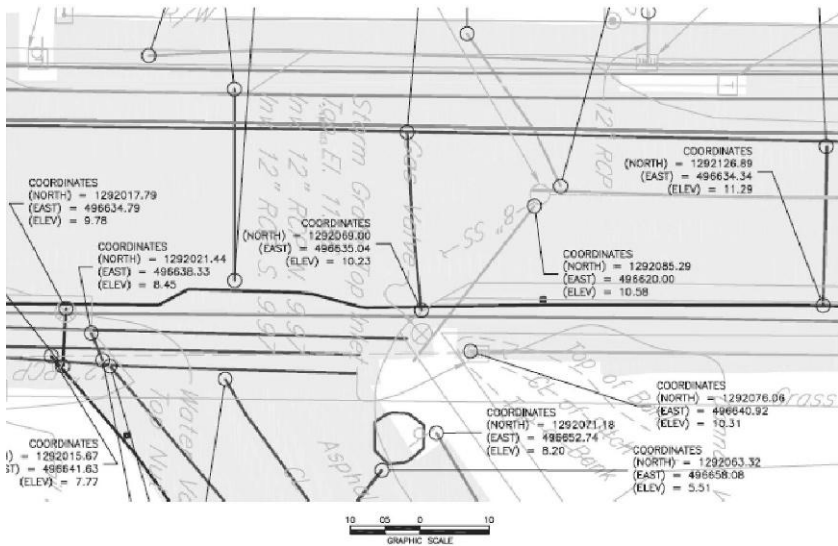
12.17 Engineering CAD drawing of a downtown utility corridor in Tampa, FL, created from a combination of GPR mapping and SUE. (*Source*: Courtesy of MPS3D, LLC.)

between array frames was 12 inches along the profile direction. The 16 vector sensors were arranged in two linear rows, with a spacing of about 30 cm between sensors in each row and a vertical offset of about 50 cm between the rows (see Fig. 12.12, center).

The grayscale image in the left panel of Fig. 12.18b shows the amplitude of the horizontal magnetic field, recorded near ground level over the survey area, at a frequency of 8.2 kHz, the frequency of current injected onto a large, deep water main emerging from a reservoir pumping station at the NE corner of the survey. Strong continuous features in the horizontal field mark high values of the magnetic field amplitude and provide a qualitative guide to the locations of buried lines. For example, the stripe of high values running SW to NE through the image indicates the track of the deep water main.

The center panel shows the vertical magnetic field amplitude at 512 Hz, the frequency of current injected onto a shallow electrical line emerging from the pumping station. The vertical field low running diagonally marks the track of the electrical line (the vertical magnetic field is null directly above a buried line of current). The right panel shows the horizontal field amplitude at 32.8 kHz, the frequency of current injected onto two shallow utility lines running along the eastern edge of the survey area. This high-frequency current, which should ideally have stayed on the two shallow lines, has clearly leaked, probably by inductive coupling, onto the electric line, water main, and other utilities. To obtain quantitative results, inversion software was developed to solve for the locations of a network of current filaments whose magnetic field best matches the surface measurements (Oristaglio *et al.*, 2005).

Figure 12.19 shows another field test of the dual-array system, mapping electrical lines near a complex power junction in the city of Elmsford, NY. The radar and EMI images in this example show the complementary nature of radar and EM probing of the subsurface. The GPR image in the top panel, focused at a depth of 30 inches below the road surface, reveals an electric power line running under the southern side of the street. The radar image beneath the middle of the street, however, is almost completely blanked out by the roadbed's concrete rebar, whose shadows appear as vertical stripes. The small circular feature to the right of center in the image marks a manhole cover accessing an underground service box that is only faintly visible in the image at this depth. A collection of conduits (conduit bank) appears to the left of the service box; here, rebar had been removed from the roadbed to access the conduits for earlier servicing, and the GPR signals could penetrate much deeper in this region. In service boxes for conduit banks, individual stretches of electric line are connected to other lines and rerouted.

The horizontal magnetic field map shown in the bottom panel of Fig. 12.19 was recorded while current at 65 kHz was injected onto a conduit carrying the
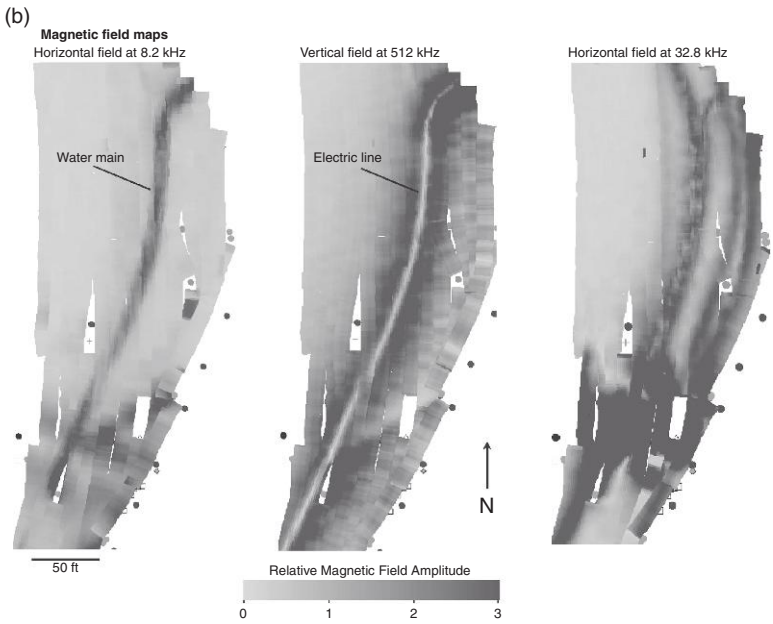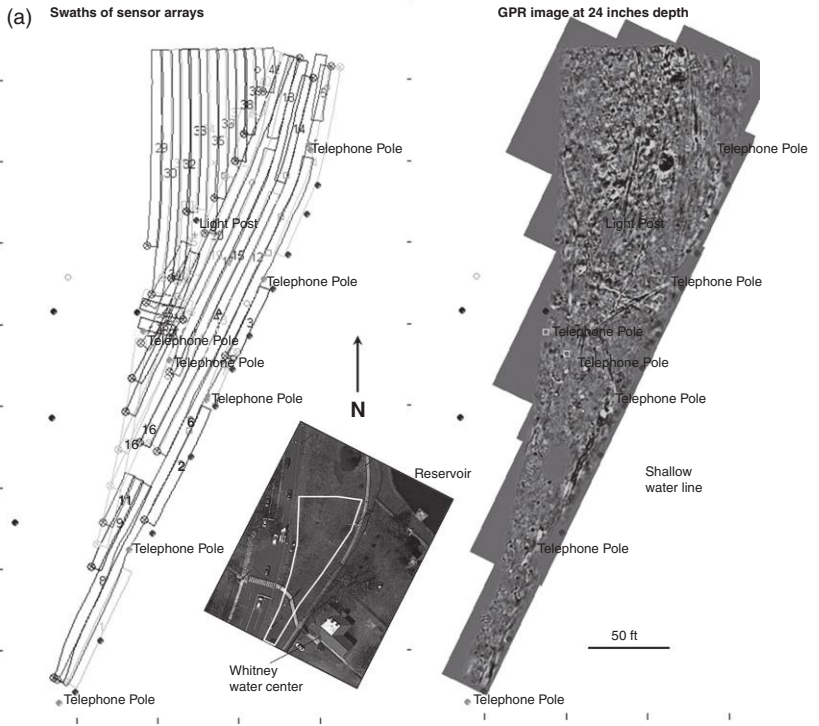
12.18 (a) Mapping a utility network near a water station in Hamden, CT. (Left to right) Swaths of the sensor array; survey area highlighted on an aerial photo of the region; radar image at 24 in. depth. (b) Magnetic field maps.

electric line seen in the radar image. The magnetic field map shows that this particular conduit loops back on itself to enter the service box at its northern end. This loop is not clear in the radar image. The electric-line conduit then bypasses the conduit bank seen in the radar image and proceeds down the street (to the left in the figure), where it intersects with parallel and perpendicular conduits (or branches into separate lines). The bypassing of the conduit bank can be deduced from the EM survey because the current injected onto the electric-line conduit does not 'light up' the bank itself. Despite systematic efforts to track interconnections in crowded corridors, utility companies often find that records become outdated when maintenance repairs are not properly recorded and entered in the network database.

12.6.2   Other case studies

Modern single-channel GPR systems allow the collection of 3D data sets over limited regions with relative ease. Porsani *et al.* (2012), for example, describe a successful project that mapped utilities under a roughly 20 m × 20 m area along a subway tunnel construction site in São Paulo City, Brazil. Case studies using commercial array systems for large-scale mapping of urban infrastructure are now appearing in the open literature. Mooney *et al.* (2010) describe a demonstration project for Con Edison in which a 3D GPR survey was performed along 8 km of city streets in Yonkers, NY, prior to installation of a new high-voltage transmission line. Birken *et al.* (2007) present results of a combined GPR/SUE project covering approximately 500 000 sq m in Istanbul, Turkey. MacIntosh *et al.* (2006, 2007) reported on projects in Florida and Colorado that located deep conductive utilities using the EM array of the dual-array system.

Array GPR systems are also being used for large archaeological surveys. Recent examples include mapping of ancient Roman sites in Italy (Piro *et al.*, 2012), Corsica (Verdonck and Vermeulen, 2011), Austria (Goodman *et al.*, 2011), and the UK (Linford *et al.*, 2010), and mapping of Viking sites in Sweden (Trinks *et al.*, 2008).

12.7   Future of mobile sensor technologies

A challenge for future mobile sensor systems is mapping infrastructure over large areas from vehicles that move at the speed of normal traffic. Existing single or dual-channel GPR systems in the GHz range, mounted behind SUVs, can collect data at 10 cm intervals while moving at highway speeds. These systems are used in roadbed and bridge-desk inspection (Maser, 1991; Maser and Rawson, 1992; Gucunski *et al.*, 2010), but the sparse lateral coverage – one or two profiles across the width of the road – allows only

Radar image at 30 inches depth   Rebar shadow

N   Conduit banks   20 ft   Service box   Electric line

Magnetic field map
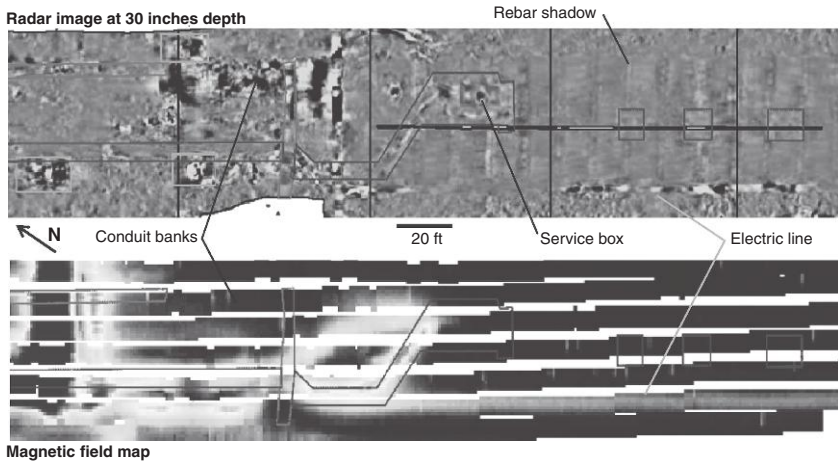
12.19 Complementary nature of radar and EM induction in underground mapping.

coarse quality control of roadbed materials and identification of isolated problems.

Two large barriers to building cost-effective GPR systems for collecting dense data at traffic speeds are (1) the cost of high-speed, 16-bit analog-to-digital converters (ADC) operating at GHz frequencies, and (2) compliance with Federal Communications Commission Rule 02–48 (FCC, 2002), which limits the strength of UWB radar pulses in the GPR band. As a result, many commercial GPR array systems employ off-the-shelf ADCs operating at kHz frequencies in combination with a sampling strategy that uses many transmitter impulses to collect a single GPR trace. Even if fast ADCs become available, pulse repetition frequencies cannot rise much higher than a few hundred kHz and stay in compliance with FCC rules. The value added by array systems in the planning and maintenance of civil infrastructure, and their demonstrated lack of interference with other communication systems operating in this band, may eventually lead to a re-evaluation of these rules.

12.7.1   VOTERS

In 2009, the National Institute for Standards and Technology funded the VOTERS project at Northeastern University. VOTERS, which stands for 'Versatile Onboard Traffic Embedded Roaming Sensors,' provides a framework to shift from periodic, localized inspections of infrastructure to continuous monitoring of roadways and bridge decks across an entire municipal network (Birken *et al*., 2012; www.neu.edu/voters). To achieve this, the project aims at developing sensor arrays that are small and affordable enough

to be mounted on fleets of vehicles that are constantly driving through a city, for example, cabs and trucks. To be compatible with the VOTERS concept, a fast GPR array system must collect data at cm-scale spatial sampling while moving at the speed of normal city traffic. Although the GPR system under development in VOTERS is being customized for inspection of bridge decks and roadway layers (Oden and Birken, 2011), it can provide building blocks for a low-frequency GPR that would locate utilities at near-highway speeds. An EMI system can be added as a complementary sensor array, coupled with an onboard induction transmitter. In principle, such a system could routinely monitor changes in the utility corridors of a major city.

12.7.2   Geophysical array technologies and SUE

In a recent review, Birken (2009) concluded that advances in mobile sensor arrays offer subsurface mapping solutions that complement traditional SUE methods. The article suggested that the combination of SUE Quality Level B with geophysical remote sensing can change the current tradeoff in cost versus risk for construction planning, because modern geophysical sensor arrays can map buried utilities over large areas at a cost (per line foot of mapped utility) that is significantly less than the equivalent cost of vacuum excavation. The lower accuracy of geophysical mapping, compared to direct measurement after vacuum excavation or hand digging, increases construction risk only slightly – or sometimes not at all, if geophysical maps can be calibrated by vacuum excavation. In fact, an easy 'win–win' in SUE projects is to use mobile sensor arrays to precisely target locations for vacuum excavation.

After examining results from several successful pilot projects, the Florida Department of Transportation recently endorsed the combination of remote sensing and vacuum excavation as best practice in managing cost and risk in large construction projects. Realizing the benefits of mobile sensors in practical civil engineering will require further changes in how infrastructure projects of the twenty-first century are planned, funded, and executed.

12.8   References

American Society of Civil Engineers (2002), Standard Guidelines for the Collection and Depiction of Existing Subsurface Utility Data (CI/ASCE 38–02) (Standard No. 038-02), Reston, VA: ASCE, 978-0-7844-0645-8 or 0-7844-0645-6.

Anspach, J. H. (1995), Subsurface utility engineering: Upgrading the quality of utility information, *Proc. Conf. on Adv. in Underground Pipeline Eng. II, 2nd Int. Conf. ASCE*, New York, 813–824.

Anspach, J. H. (1996), Subsurface Utility Engineering: Utility Detection Methods and Applications, *Symposium on the Application of Geophysics to Engineering and Environmental Problems*, SAGEEP, Keystone, CO, 443–449.

Bianchi, C. and Meloni, A. (2007), Natural and man-made terrestrial electromagnetic noise: An outlook, *Annals of Geophysics*, **50**, 435–45.

Birken, R. (2009), Impact of geophysical array technologies on subsurface utility engineering quality levels, *Proc. of ANCRISST,* Boston, MA, *CD-Rom*.

Birken, R., Hansen, T., Deming, R. and Oristaglio, M. (2004), Locating pipelines with a broadband electromagnetic induction array system, *Proc. SAGEEP 2004*, Colorado Springs, Colorado, 1023–1034.

Birken, R., Wang, Ming L. and Wadia-Fascetti, S. (2012), Framework for continuous network-wide health monitoring of roadways and bridge decks, *Proc. Transportation Systems Workshop 2012*, Austin, Texas.

Birken, R., Stearns, R., Zhu, Q. and MacIntosh, S. (2007), Combining 3D GPR and subsurface utility engineering to create accurate utility maps in Istanbul, *Proc. Near Surface Geophysics 2007*, Istanbul, Turkey, B23.

Birken, R., Miller, D., Burns, M., Albats, P., Casadonte, R., Deming, R., DeRubeis, T., Hansen, T. and Oristaglio, M. (2002), Efficient large-scale underground utility mapping in New York City using a multi-channel ground-penetrating imaging radar system, *Proc. SPIE*, 4758, 186–191.

Budden, K.G. (1988), *The Propagation of Radio Waves*, Cambridge University Press, Cambridge.

Burns, M., Derubeis, T., Albats, P., Casadonte, R., Birken, R., Deming, R., Haldorsen, J., Hansen, T., Miller, D.E. and Oristaglio, M.L. (2004), Method for merging position information with measurements and filtering to obtain high-quality images that are positioned accurately with respect to global coordinates: United States patent 6,766,253.

Butler, D.K. (2005), *Near-Surface Geophysics, Investigations in Geophysics Series*, Society of Exploration Geophysicists, Tulsa.

Chew, W. (1990), *Waves and Fields in Inhomogeneous Media*, Van Nostrand Rheinhold.

Conti, U. (1992), Multiple-coil magnetic field sensor with series-connected main coils and parallel-connected feedback coils: United States patent 5,130,655.

Dabas, M., Jolivet, A. and Tabbagh, A. (1992), Magnetic susceptibility and viscosity of soils in a weak time varying field, *Geophysical Journal International*, **108**, 101–109.

Daniels, D. (ed.) (2004), *Ground Penetrating Radar*, 2nd edn., Institution of Electrical Engineers, London.

Doolittle, J., Dobos, R., Peaslee, S., Waltman, S., Benham, E. and Tuttle, W. (2010), Revised ground-penetrating radar soil suitability maps, *Journal of Environmental & Engineering Geophysics*, **15**, 111–118.

Eide, E. and Hjelmstad, J. (2002), 3D utility mapping using electronically scanned antennae array, *Proc. 9th Int. Conf. GPR*, Santa Barbara, CA.

Engheta, N., Papas, C. H. and Elachi, C. (1982), Radiation patterns of interfacial dipole antennas, *Radio Science*, **17**, 1557–1566.

Federal Communications Commission (2002), FCC 02–48, www.fcc.gov/Bureaus/ Engineering_Technology/Orders/2002/fcc02048.pdf.

Francese, R.G. and Morelli, G. (2006), New perspectives in buried utility detection and mapping with a multi-scan GPR system, *Proc. SAGEEP 2006*, Seattle, WA, 439–442.

Goodman, D., Novo, A., Morelli, G., Piro, S. Kutrubes, D. and Lorenzo, H. (2011), Advances in GPR imaging with multi-channel radar systems from engineering to archaeology, *Proc. SAGEEP 2011*, Charleston, SC, 416–422.

Goodrum, P., Smith, A., Slaughter, B. and Kari, F. (2008), Case study and statistical analysis of utility conflicts on construction roadway projects and best practices in their avoidance, *Journal of Urban Planning and Development*, **134**, 2, 63–70.

Grasmueck, M. and Viggiano, D.A. (2007), Integration of ground-penetrating radar and laser position sensors for real-time 3-D data fusion, *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 1, 130–137.

Grasmueck, M., Weger, R. and Horstmeyer, H. (2005), Full resolution 3D GPR imaging, *Geophysics*, **70**, 1, K12–K19.

Gucunski, N., Romero, F., Kruschwitz, S., Feldmann, R., Abu-Hawash, A. and Dunn, M. (2010), Multiple complementary nondestructive evaluation technologies for condition assessment of concrete bridge decks, *TRB: Journal of the Transportation Research Board*, No. 2201, National Research Council, Washington, DC, 34–44.

Hipp, J.E. (1974), Soil electromagnetic parameters as functions of frequency, soil density, and soil moisture, *Proc. IEEE*, **62**, 98–103.

Jeong, H.S., Abraham, D.M. and Lew, J.J. (2004), Evaluation of an emerging market in subsurface utility engineering, *Journal of Construction Engineering and Manufacturing*, **130**, 225–234.

Jol, H.M.,(ed.) (2009), *Ground Penetrating Radar: Theory and Applications*, Elsevier Science, Boston.

Jung, Y.J. (2012), Evaluation of subsurface utility engineering for highway projects: Benfit-cost analysis, *Tunneling and Underground Space Technology*, **27**, 1, 111–122.

Kalisch, B. (2000), *Achieving Common Ground*, AGA Special Report, www.aga.org.

Kraus, J. and Fleisch, D. (1999), *Electromagnetics*, 5th edn., McGraw Hill, New York.

Lehmann, F. and Green, A.G. (1999), Semiautomated georadar acquisition system in three dimensions, *Geophysics*, **64**, 3, 710–731.

Lew, J.J. (1999), Cost savings on highway projects utilizing subsurface utility engineering, Federal Highway Administration, NTIS No. FHWA/IF00/014, Washington, D.C.

Linford, N., Linford, P., Martin, L. and Payne, A. (2010), Stepped frequency ground-penetrating radar survey with multi-element array antenna: Results from field application on archaeological sites, *Archaeological Prospection*, **17**, 187–198.

Macaulay, D. (1976), *Underground*, Houghton Mifflin, New York.

MacIntosh, S., Birken, R. and Zhu, Q. (2007), Subsurface pipe detection with an array of inductive broadband vector receivers, *Proc. Near Surface Geophysics 2007*, Istanbul, Turkey, B07.

MacIntosh, S., Birken, R. and Zhu, Q. (2006a), Correcting 3D position measurement for geophysical surveys using a tilt sensor system, *Proc. SAGEEP 2006*, Seattle, Washington, CD-ROM.

MacIntosh, S, Tolboll, R., Birken, R. and Zhu, Q (2006b), Large-scale mapping of complex utility lines with a broadband electromagnetic induction array, *Proc. SAGEEP 2006*, Seattle, Washington, CD-ROM.

Maser, K.R. and Rawson, A. (1992), Network Bridge Deck Surveys Using High Speed Radar: Case Studies of 44 Decks: Transportation Research Record No. 1347, Transportation Research Board National Research Council, Washington, DC.

Maser, K.R. (1991), Bridge Deck Condition Surveys Using Radar: Case Studies of 28 New England Decks: Transportation Research Record No. 1304, National Research Council, 94–102.

Meloy, J. (2003), What and Where is the Natural Noise Floor?, http://www.vlf.it/naturalnoisefloor/naturalnoisefloor.htm, accessed 18 November 2012.

Mooney, J.P., Jr., Ciampa, J.D., Young, G.N., Kressner, A.R. and Carbonara, J. (2010), GPR mapping to avoid utility conflicts prior to construction of the M-29 transmission line, *Proc. Transmission and Distribution Conference and Exposition 2010, IEEE PES*, 1–8.

Nabighian, M., (ed.) (1988), *Electromagnetic Methods in Applied Geophysics: Theory (vol. 1) and Applications (vol. 2)*, Society of Exploration Geophysicists, Tulsa.

North, D. (2010), Marking the spot: RFID markers and GPS technology support Virginia DOT roadway construction, CE News, December, 38–40.

Oden, C. and Birken, R. (2011), Array GPR system for continuous monitoring of bridges and roadways, *Proc. Engineering Mechanics Institute Conference*, Boston, MA, 2–4 June 2011.

Oristaglio, M., Miller, D. and Haldorsen, J. (2001), *Ground-Probing Radar, in Scattering*, P.C. Sabatier and E. R. Pike, eds., Academic Press, London.

Oristaglio, M., Birken, R., Hansen, T., Deming, R., Macintosh, S. and Zhu, O. (2005), Efficient large-scale utility mapping with radar and induction arrays, *Proc. SAGEEP 2005*, Atlanta, GA.

Piro, S., Morelli, G., Novo, A., Ceraudo, G. and Goodman, D. (2012), A GPR array system for fast archaeological mapping: Stream X at Aquinam Roman site (Castrocielo, Italy), *Proc. SAGEEP 2012*, Tuscson, AZ, 322–329.

Porsani, J.L., Ruy, Y.B., Ramos, F.P. and Yamanouth, G.R.B. (2012), GPR applied to mapping utilities along the route of the Line4 (yellow) subway tunnel construction in São Paulo City, Brazil, *Journal of Applied Geophysics*, **80**, 25–31.

Rechtsman, M. and Torquato, S. (2008), Effective dielectric tensor for electromagnetic wave propagation in random media, *Journal of Applied Physics*, **103**, 084901/1–15.

Sterling, R.L. (2000), Utility Locating Technologies: A Summary of Responses to a Statement of Need Distributed by the Federal Laboratory Consortium for Technology Transfer, Federal Laboratory Consortium Special Report Series No. 9, Washington, DC.

Skolnik, M. (2002), *Introduction to Radar Systems*, 3rd edn., McGraw-Hill, New York.

Stevens, R.E. and Anspach, J.H. (1993), New technology overcomes the problems of underground system interferences on power projects, *American Power Conference* 55, 323–326.

Transportation Equity Act for the 21st Century, TEA 21, Title VII, Subtitle C, Comprehensive One-Call Notification, enacted June 9, 1998 (http://www.fhwa.dot.gov/tea21/).

Trinks, I., Nissen, J., Johansson, B., Emilsson, J., Fribor, J. and Gustafsson, J. (2008), Pilot study of a new multichannel GPR system MIRA for large scale, high-resolution archaeological prospection at the site of the Viking town Birka in Sweden, ISAP News, Issue 16, July 2008

Ulaby, F., Ravaioli, U. and Michielssen, E. (2010), *Fundamentals of Applied Electromagnetics*, 6th edn., Prenctice Hall, New York.

Verdonck, L. and Vermeulen, F. (2011), 3-D survey with a modular ground-penetrating radar system at the Roman town Mariana (Corsica), *6th Int. Workshop on Advanced Ground Penetrating Radar (WAGPR),* Aachen, Germany.

Vozoff, K. (1985), *Magnetotelluric Methods*: Society of Exploration Geophysicists, Tulsa.

Wang, T. and Oristaglio, M. (2000a), GPR imaging using the generalized Radon transform, *Geophysics*, **65**, 1553–1559.

Wang, T. and Oristaglio, M. (2000b), 3-D simulation of GPR surveys over pipes in dispersive soils, *Geophysics*, **65**, 1560–1568.

Zembillas, N. (2003), Subsurface utility engineering – A technology-driven process that results in increased safety, fewer claims and lower costs, *Proc. Conf. Int. Society of Trenchless Technologies (ISIT)*, CD-ROM, Copenhagen, Denmark.

Zembillas, N.M. (2008), Subsurface utility engineering: A technology-driven process that results in increased safety, fewer design changes, and lower costs, *Proc. of the ASCE International Pipelines Conference*, Atlanta, GA, ASCE.

13

Sensing solutions for assessing the stability of levees, sinkholes and landslides

D. INAUDI, SMARTEC SA, Switzerland

Abstract: Monitoring large-scale structures such as dams and levees has always presented significant challenges, in particular when localized damage needs to be detected and mapped. Traditionally, the only monitoring procedure allowing the detection and localization of damage such as leakages, settlements, cracks, or abnormal joint movements was periodic visual inspection. However, inspecting large structures regularly is a tedious and subjective procedure, which reveals issues only visible to the eye. The interval between inspections can be large and it is difficult to obtain rapid feedback on the conditions of a structure during or immediately after an event such as an earthquake or a large storm. Permanent monitoring systems based on distributed optical fiber sensing technology, laser distance meters, and synthetic aperture radar (SAR), have opened new possibilities to address these needs. They allow integrity monitoring with direct detection, localization, characterization, and immediate reporting of new local conditions. These systems are therefore not only able to measure strain and temperature (answering the 'how much' question) but also to localize possible damage areas (answering the 'where' question). This makes them ideal for monitoring structures where the location of a possible instability or failure is unknown. For example, the sensors can detect and localize a seepage zone in a levee, the onset of a sinkhole, or the formation of a crack in a concrete wall. This also allows a more targeted approach to inspection, since it is now possible to concentrate the interventions on specific areas showing abnormal behavior. These technologies and several field application examples will be presented in this chapter.

Key words: levee monitoring, sinkhole monitoring, landslide monitoring, distributed sensing, distributed optical fiber sensors, instability localization, geotechnical monitoring.

13.1 Introduction

Requirements for structural health monitoring in the last few decades have rapidly increased, stimulating new developments in various sensing technologies. New technologies, such as distributed optical fiber sensing

396

and laser distance meters, global positioning system (GPS), and SAR offer new possibilities in structural monitoring. Distributed deformation sensors (sensing cables) are sensitive to strain and temperature variations at each point along their length. A single sensor can now measure one-dimensional strain fields over the entire length of the structure (levee, landslide, sink-hole area, tunnel, etc.), to provide assurance of integrity, and to directly detect, characterize, and report local strain changes and movements. These sensors are therefore not only able to measure strain (answering the 'how much' question) but can also localize damage areas (answering the 'where' question). This makes distributed fiber optic (FO) sensing ideal for mon-itoring structures where the location of possible instability or failure is unknown a priori. For example, the sensor can detect and localize a seepage zone in a levee, the onset of a sinkhole in a subdivision or the formation of a crack in a tunnel liner. Laser distance meters and SARs (ground- or satellite-based) can be used to measure small displacements over large distances, making them ideal for monitoring large-scale events such a landslides or subsidence from a distance. In this chapter, distrib-uted sensing techniques and components based on Brillouin and Raman scattering, as well as laser distance meters, are briefly introduced and their potential for use in monitoring for geotechnical engineered systems is dis-cussed. Finally, several large-scale case studies are presented, including levee monitoring, sinkhole detection, and landslide observation.

13.2   Detection, localization and quantification of instability

The growing demand for safety awareness over the last few years has stimulated the development of several monitoring techniques capable of detecting early stage events, thus preventing structures from major failures and leading to better knowledge of the structure itself (Glisic and Inaudi, 2006, 2007a). In the field of geotechnical applications such as levees, landslide, sinkholes, and tunnels, where both structural dimensions and the need to forecast damage location represent a major challenge, distributed techniques offer the ability to monitor several kilometers of a structure using a single fiber optic sensor (FOS). Using a limited number of very long sensors, it is possible to successfully monitor the structural and functional behavior of geotechnical structures with a high spatial resolution over large areas using a single interrogator at a reasonable cost (Glisic and Inaudi, 2007b). Other techniques that offer a global cov-erage of a large structure include robotic laser distance meters (Manetti, L. *et al.*, 2008) and ground- or satellite-based SAR (Colesanti, C. and Wasowski, J. (2006).

13.2.1   Distributed optical fiber sensors

The evolution of distributed technologies such as Raman and Brillouin optical fiber sensing systems has led to the development of suitable FOS cables capable of detecting strain and temperature with high accuracy and reliability (Belli *et al*. 2009). Currently, different distributed FOSs are adapted and optimized for installation in various situations, such as embedding in concrete, surface mounting inside existing tunnels, or laying in a trench in the ground to monitor a potential landslide. Distributed sensing cable design addresses several goals, such as protecting the fiber from accidental damage while ensuring the transfer of strain and temperature from the structure to the sensing fiber. Furthermore, a cable can be designed to contain several fibers for redundancy, or for providing different sensing functions. Figure 13.1 shows several examples of distributed sensing cables. The picture (a) represents a strain-sensing cable that is constituted of an optical fiber embedded in composite material tape. The tape is designed for direct gluing on relatively smooth surfaces, such as steel or concrete. The tape shape significantly increases the adhesion area, allowing for an easier transfer of strain from the structure to the fiber and at the same time allowing an easier manipulation of the sensor. The second cable (b) combines two strain-sensing optical fibers embedded in a thermoplastic material with two free optical fibers installed in an embedded tube. The two free fibers can be used for temperature sensing. With a single cable it is therefore possible to separate the effects of strain from the effect of temperature changes that might be occurring simultaneously. This cable is designed for direct embedding in concrete and soil, and allows an accurate strain transfer to the sensing fibers. The two strain fibers and the two temperature fibers also guarantee a minimum redundancy, should one fiber fail for some reason.

   The last design (c) also contains two fibers for strain monitoring and two fibers for temperature compensation. In this case, the transfer of strain from the host structure to the fiber is, however, performed by friction and not by embedding. This allows production of this cable at lower cost, but does not guarantee perfect strain measurement fidelity as with the other two designs. This cable is therefore more suitable for qualitative measurements, e.g. to identify movement zones in geotechnical applications.

   The increasing demand for monitoring geotechnical structures, such as levees, landslide, sinkholes, etc., has encouraged further development of specialized cables, capable of high sensitivity and reliability along with mechanical robustness suitable for the harsh operational environment. This research has led to the development of geotextiles that integrate the optical fiber sensors into them. This type of sensor is particularly suitable for installation by direct embedding in ground/soil. The integration with geotextile sheets offers better mechanical protection and enhanced sensor performance in terms of

(a)

Thermoplastic tape   Optical fiber

13.00

0.20

(b)

8.00

1.30   2.70   2.70   1.30

2.00   2.00   4.00

(c)

Outer sheath

Aramide yarn

Loose tube

FO strain   FO temperature
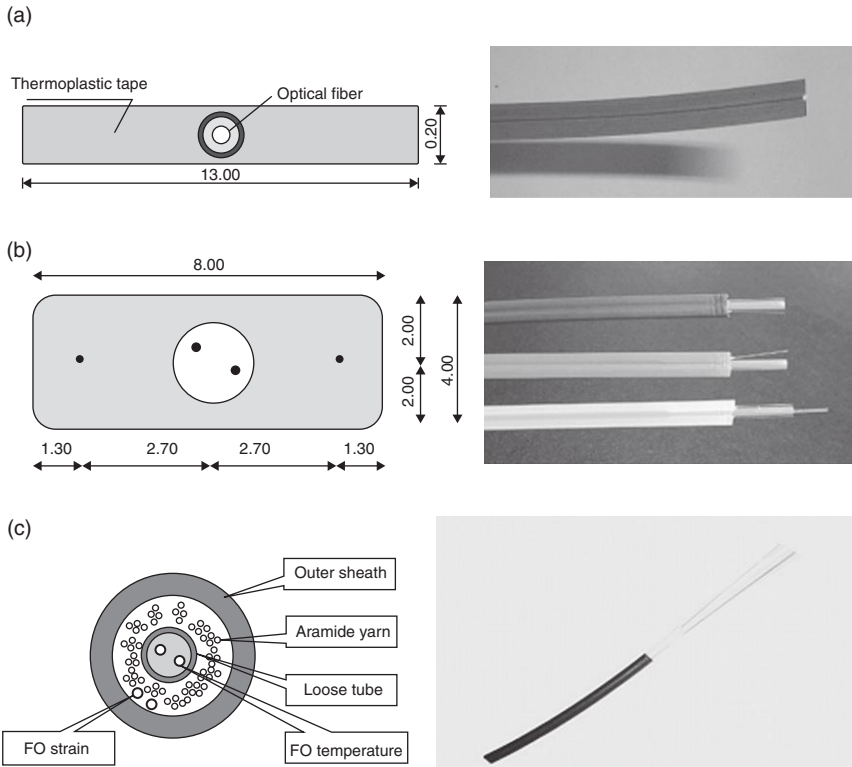
13.1 (a) SMARTape sensor for distributed strain sensing on metallic structures, such as a pipeline. (b) SMARTProfile sensor, suitable both for structural and geotechnical monitoring over long distances. Contains two strain-sensing fibers and two free fibers for temperature compensation. (c) SMARTube distributed sensor for geotechnical applications where direct embedding, both in ground/soil and concrete, is required. (*Source*: Photos courtesy of SMARTEC SA, Switzerland.)

sensitivity and the detection of small events. Cloth-type geotextiles can also be used to guide water from seepage toward the sensing cable to enhance the cable's sensitivity. Figure 13.2 shows an example of an instrumented geotextile.

Instabilities can be identified and localized using FOS by different mechanisms. These mechanisms can be summarized as follows:

- Strain changes in the ground induced by soil movements.
- Relative movements between anchors attached to the ground, a tunnel inner lining, or another retaining structure.
- Detection and localization of cracks produced in concrete or steel by ground movements.

13.2 Geotextile with embedded distributed strain sensor. (*Source*: Photo courtesy of SMARTEC SA, Switzerland.)

- Identification of leakages produced by damage to the retaining structure through temperature anomaly monitoring.
- Changes in tension in retaining structures, such as anchors or geotextiles.

13.2.2   Laser distance meters

The laser distance meter systems allow continuous and autonomous millimeter-scale distance measurements over distances up to approximately 1000 m (Manetti, L. *et al.*, 2008). A typical remote laser system configuration consists of a monitoring laser instrument, targets and personal computer with software. In the case of outdoor installations, the laser instruments are usually mounted in a protection cabinet (see Fig. 13.3). When a large number of measurement points can be observed from a single location, it can be advantageous to use a robotic total station to automatically scan all points with a single instrument. For distances of up to approximately 100 m, it is possible to obtain measurements without the use of specific targets, using the surface reflection of concrete or stone. For larger distances, reflecting sheets or surveying reflectors and targets are needed.

13.3 Laser distance meters, with protective housing, control system, and solar power supply. (*Source*: Picture courtesy SMARTEC SA.)

Laser distance meters and total stations can be applied in several surveying applications, such as landslides, earth settlement, slope instability, tunnel monitoring, tank level monitoring, or civil engineering surveying, where millimeter-scale distance measurements are required.

The presented techniques offer a wide variety of performances and possible applications, and can be advantageously applied in different situations. Table 13.1 summarizes the main aspects of each technology.

Table 13.1 Comparison of main characteristics of sensing technologies

	Distributed FO Raman sensing	Distributed FO Brillouin sensing	Laser distance meters/surveying	Ground-based SAR	Satellite-based SAR
Type of measurement	Temperature measurement points at 1 m distance along cable path	Strain and temperature measurement points at 1 m distance along cable path	Distances between reference point and movable targets	Distance changes between reference point and movable object	Vertical movements of areas or 3D movement of targets
Typical number of measurement points	10 000	20 000	10–100	10 000	Millions
Applications	Seepage Leakage Environmental changes	Ground movements Cracks and joint movements Sinkholes Levee failure	Landslides Movement of structures	Landslides Dam monitoring	Subsidence Settlements
Main application domains	Dams, dykes, levees, sinkholes, glaciers	Dams, dykes, levees, walls, tunnels	Landslides, bridges and buildings	Dams, dykes	Levees, sinkholes
Cost (typical order of magnitude)	20 000–100 000 $	100 000–250 000 $	10 000–100 000 $	100 000–200 00 $	Depends on area and frequency
Typical accuracy	0.1°C	1°C 20 µε	0.5 mm	0.5 mm	1 cm

13.3   Levee monitoring

A levee or dyke is a barrier built in order to retain large water accumulations from rivers, lakes, or seaboards. There are different types of dykes, depending on their manner of construction. In general, the main components include a watertight clay core, a fill material (such as earth or rocks), and, close to the surface, there may be a filtering layer and a reinforced concrete jacket (Glisic and Inaudi, 2007a). Dykes are frequently found on soil with relatively poor mechanical properties. They have a trapezoidal cross-section, being very wide at the base and relatively narrow at the top. The angles of the slopes primarily depend on the construction material and are imposed by stability conditions. In some cases a concrete wall will complement the levee, creating an additional retaining structure on top of the earth structure. Being a barrier for large water accumulations, in order to increase safety and ensure structural reliability, requirements for monitoring of dykes and levees are increasing. The main aims of monitoring are early detection of slope instability, uncontrolled seepage and/or piping (formation of a natural 'pipe' inside the levee), as well as internal erosion due to seepage. Uncontrolled seepage can result from cracking of the concrete jacket or clay core and can be generated by water pressure combined with long-term settlement of the dyke construction. Monitoring for deformations in the jacket, core, and soil allows for the early detection of such failure mechanisms. The stability of slopes depends mainly on the construction material, the water table/level in the dyke itself and the pore pressure in the soil. Pore pressure, therefore, is an important parameter to be monitored.

Due to their large spatial extent and the impossibility of clearly determining a priori the critical sections, distributed sensing systems are particularly suitable for monitoring levee integrity and the localization of damage. This system is mainly intended to measure movement, leakage, crack detection, and average strain, and can replace thousands of discrete deformation and temperature sensors. A schematic showing typical positions for distributed sensor placement in a dyke cross-section is presented in Fig. 13.4. The distributed systems offer the unique capability of monitoring strain and temperature simultaneously: distributed deformation FOS can provide for detection and localization of slow movement in the slopes. At the same time, distributed temperature FOS can provide for seepage or overflow detection. Slow local movement of a slope puts the deformation sensor in tension; seepage or overflow changes the thermal properties of the soil, which are detected by the temperature sensor. In both cases, localization and evaluation of the size of the area involved can be performed. The principles are shown schematically in Fig. 13.4. It is generally possible to cover a length
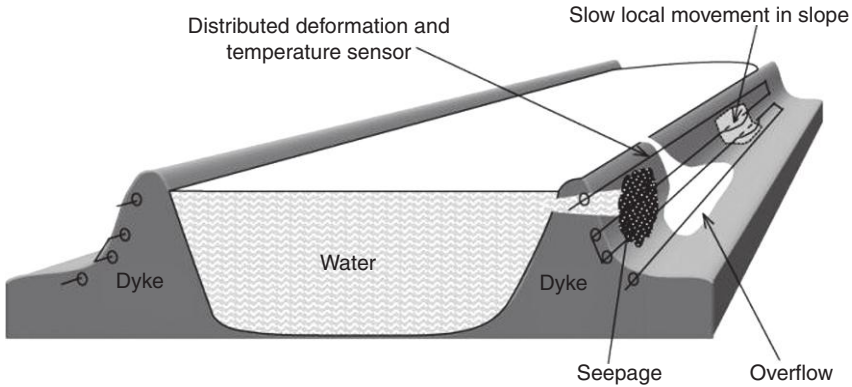
13.4 Schematic representation of distributed deformation and temperature FOS locations in a levee for seepage and overflow detection (distributed temperature FOS) and localization of slow slope movements (distributed strain FOS). (*Source*: Image courtesy of SMARTEC SA, Switzerland.)

of up to approximately 20 km of a given levee with a single measurement instrument.

For such applications, rugged distributed strain sensors prove to be suitable. Moreover, the previously described geotextile solutions enhance the detecting capabilities of the monitoring system both in terms of early stage movement and seepage. In this case, in addition to enhanced monitoring capability, the choice of a geotextile solution can also provide reinforcement capabilities, thus providing a double benefit with enhanced safety.

13.3.1   Application example: iLevees

The ongoing iLevees project 'Intelligent Flood Protection Monitoring Warning and Response Systems,' in the state of Louisiana, has the goal of providing an alerting and monitoring system capable of preventing early stage failure, in terms of both ground instability and seepage. The motivation for the monitoring system is to improve safety awareness, provide sensible information about levees' status and conditions, before, during, and after floods, and to prevent the tragic events such as those that occurred following Hurricane Katrina in 2005. The use of distributed fiber optic sensing will help in overcoming the issue of optimal sensor location, allowing full structure coverage over several kilometers. Continuous long-term monitoring during the complete levee lifetime will allow for the collection of data that can improve our general knowledge of these structures, with unquestionable benefits in future levee designs,

operation, and maintenance. Figures 13.5 and 13.6 show the installation of distributed strain- and temperature-sensing cables in the levee foundations and on top of the I-wall and T-wall sections (the sensors depicted in Fig. 13.1b were used). These sensors allow the detection and localization of events such as levee failure onset, seepage, tunneling, formation of cracks in wall sections, or abnormal joint movements. In addition to fiber optic sensing, the iLevees project includes several other traditional and advanced sensing techniques, including piezometer arrays to monitor the groundwater pressure and satellite-based SAR techniques. This latest technique is particularly interesting for monitoring large areas and detecting settlements. However, since typical satellite scan intervals are only every 15–30 days, this technology is not suitable for real-time alerting and monitoring tasks.

13.3.2   Application example: IJkdijk project

The IJkdijk project in the Netherlands (Artières *et al.*, 2010) is for the development of sensing and monitoring techniques for preventing levee failures. The project has realized a number of laboratory and full-scale tests to
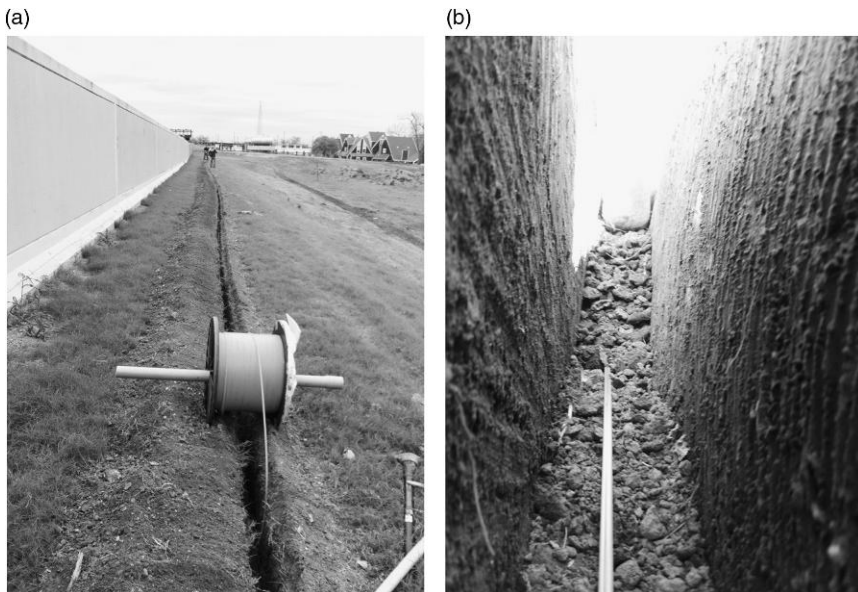
13.5 Installation of distributed FO strain and temperature sensing cables in a test section of levee in New Orleans Louisiana. (*Source*: Images courtesy iLevee Project, SMARTEC, Geocomp, Shannon & Wilson, New Orleans Levee District.)

(a)   (b)

13.6 Installation of distributed FO strain-sensing cables on top of a retaining wall section in New Orleans Louisiana to detect wall movements. (*Source*: Images courtesy iLevee Project, SMARTEC, Geocomp, Shannon & Wilson, New Orleans Levee District.)

validate models and to test different sensing and instrumentation technologies that could allow an early detection and localization of potential levee failure. One of the major field tests was the construction of a full-scale section of levee. The levee was designed with a height of 6 m, a length of 100 m, and a base width of 27 m, with a crest width of 3 m and side slopes of 1:1.5 (V:H) on the dry side and 1:2.5 on the wet side. The levee was built parallel to a local canal levee and on top of 1.3–3 m of clay and peat. The levee core is sand with a thick clay cover.

The levee was instrumented using sensor technologies from ten different participants, varying from acoustic measurements, optical detection (by three different methods), loosely connected microelectromechanical systems (MEMS), thermographic cameras, LIDAR, and three traditional systems for measuring pore pressures and humidity, combined in innovative ways. To enable the calibration of the new techniques and evaluate the test in general, reference monitoring systems consisting of traditional pore pressure meters, humidity meters, cameras, and three vertical Shape Accelerometer Array (SAAs, rigidly connected MEMS inclinometers) were installed as well. Figure 13.7 shows a photo of the test section after the induced failure had occurred. Four distributed fiber optic-sensing cables embedded in a geotextile were buried along the length of the dyke. The lowest was buried

13.7 IJkdijk test levee after collapse. (*Source*: Image courtesy of Deltares.)

under 30 cm of sand. The others were buried under the clay cover of the dyke. A Brillouin scattering reading unit was used to measure strain during the test. In Plate XIV in the color section between pages 374 and 375, it is possible to observe that the distributed sensing system has recorded peaks of strain at the boundaries of the failure zone. Early signs of distress were already visible 42 h before the levee eventually failed. This demonstrates the potential of this technique for early detection and localization of weak spots along a large dyke.

13.4   Sinkhole monitoring

A sinkhole (also known as a sink, shake hole, or swallow hole) is a natural depression or hole in the Earth's surface caused by karst processes (i.e. the chemical dissolution of carbonate rocks or the suffusion process in sandstone). Sinkholes may vary in size from 1–600 m both in diameter and depth, and vary in form from soil-lined bowls to bedrock-edged chasms (Fig. 13.8). Sinkholes may be formed gradually or suddenly, and are found worldwide. It is clear that such phenomena represent a risk to ground stability and a non-negligible safety risk to surface and subsurface infrastructure in the surrounding areas. Such applications require critical area localization, and the use of the discrete sensors is practically difficult, if not impossible, because of the installation complexity and related costs. In these cases, a distributed sensing system is particularly suitable and cost effective (Belli *et al.*, 2009).

13.8 Sinkhole formation. (*Source*: Picture courtesy of Burns and McDonnel.)

FOS technology is well suited for defining a monitored perimeter, where the exact location of where a sinkhole might form, such as in the vicinity of old salt wells and mines, needs to be known precisely. The sensing cable can be simply installed by direct burial in a shallow trench dug along the perimeter. The selection of the appropriate sensing cable represents a key aspect and a big challenge since the cable should be capable of withstanding hostile environmental conditions (e.g. wide temperature variations, direct burial in the earth, as well as being resistant to burrowing rodents). The cable also needs to be sensitive enough to provide early and reliable displacement detection, and capable of optimizing the transfer of forces from the ground to the fiber. Such a trade-off between robustness and sensitivity was obtained through ad hoc cable development and extensive testing to assess both the mechanical reliability and sensitivity of the cable to displacements and external forces.

13.4.1   Application example: salt cavern monitoring, Hutchinson, Kansas

The city of Hutchinson is located in Reno County, Kansas. Hutchinson is on the route of the trans-continental, high-speed mainline of one of the nation's largest railroads. The railway passes near a former salt mine well field, where mining was carried out in the early part of the twentieth century. The salt mining was performed at depths of over 400 ft by drilling wells through the shale bedrock into the thick underground salt beds. Fresh water was pumped into the salt, dissolving it to be brought back to the surface as brine for processing and sale. This solution mining process resulted in the generation of multiple, large underground voids and caverns, which have

been reported to be up to 300 ft tall and over 100 ft in diameter. In places, the shale roof rock over some of these old mine voids has collapsed, forming crater-like sinkholes that can be over 100 ft in diameter and 50 ft deep at the surface. The collapse and sinkhole formation can occur very rapidly, over a period of hours to days. Figure 13.8 is a photograph of a sinkhole that opened up virtually overnight at this site in 2005, from the collapse of a salt cavern that had last been mined in 1929. The potential rapid formation of sinkholes by the collapse of old mine caverns clearly represents an issue for ground stability, and a non-negligible safety risk for surface infrastructure, including the railway.

An area on the site containing old, potentially unstable salt caverns adjacent to sensitive surface infrastructure (e.g. reads, railway lines, or houses) was identified to establish an effective monitoring system in order to provide early stage detection, continuous monitoring, with automatic telemetry. Arrangements were made for alerting via cell phone and email, in the case of ground deformation (strain) that might be the early signs of sinkhole formation. The distributed fiber optic (FO) monitoring system was selected in large part because it provides thousands of monitored points using a single FO sensing cable, all measured at the same time, in a single scan (Shefchik *et al.*, 2011). The sensing cable is directly buried at a depth of approximately 1.4 m in a potential sinkhole area above and around salt caverns with a total path length of over 4 km (see Fig. 13.9). After digging the trench, the silty soil was mechanically compacted, and the sensing cable was placed on the compacted soft ground before the trench was backfilled and compacted. The sensing cable was installed in several segments to provide easier handling during installation, and to adapt to the site by running the cable through several short, horizontally bored segments beneath a large drainage ditch, multiple road crossings, and other obstacles at the surface. All cable segments were later linked to form a single sensing loop by FO fusion splicing. The splices between segments, as well as some extra lengths of non-buried cable, are stored in dedicated, above-ground junction boxes that can be accessed for maintenance as well as for re-routing segments of cable in case a break were to be caused by the formation of a sinkhole. The final layout of the FO cable is shown in Fig. 13.10, showing the software graphical user interface used to display the measurement data.

Following the installation of the cable and completion of all the necessary quality/functionality tests on the sensing cable including a sensor integrity test by means of visual inspection, sensor attenuation test by means of optical time domain reflectometry (OTDR) measurements, and the testing of the quality of the FO splices, the system was ready for commissioning and final handover. The system commissioning mainly consisted of the following:

13.9 Digging of a trench to install the distributed sensing cable. (*Source*: Photo courtesy of SMARTEC.)

- Sensor parameterization to optimize system performance in terms of strain resolution: using the FO system managing software, it is possible to set the length of the sensor, the spatial resolution, the measurement time and a series of instrument parameters that influence the final system performance in terms of strain resolution and accuracy.
- Establish and survey an on-site coordinate system so that lengths along the cable can be related to specific marked locations on the ground: a key aspect in a distributed monitoring project is an established coordinate system that will allow the precise position of an alarm triggered by ground strain to be shown on a computerized map. When ground strain is detected by the cable, the software reports the exact location along

the cable in terms of distance from the end of the cable (essentially at the location of the computer). Luminous, high-visibility signs posted at the site along the cable route indicate lengths from the end of the cable so that responders in the field can quickly proceed to and identify the location indicated by the system. The coordinate system allows the user to define several specific alarm zones along the sensor path with specific thresholds or alert requirements depending on the project needs. All alarm checking and notifications are automatically handled by the system software.

- System functionality check: includes a simulation of ground settlement by artificial imposition of external force. Tests were run in the field along temporarily unburied segments of the cable. By displacing the cable to simulate ground strain or by pulling the cable in tension, system users can gauge and record the response of the system.

The final step for achieving a fully automatic surveillance system is the use of the distributed data management and analysis software that is designed for data storage, processing, representation, and analysis, as well as for the control of single or multiple reading units. The main functions of the software are automatic data acquisition, mapping and graphical visualization of the real-time strain data along the entire cable length, and triggering of warnings of significant ground displacement on the display, as shown in Fig. 13.10. The software stores all information related to a sensor in a single database structure. Multiple users can access the software simultaneously from different PCs (locally or remotely over a modem or local area network).

The algorithm that supports the software is robust to protect against false alarms caused by outlier values or noisy measurements. It provides system-wide corrections for environmental influences and variations. Seasonal variations in temperature can be screened out so that they do not impact the validity and reliability of the measurements. The software is specifically developed to send alerts in case ground deformation exceeds a designated threshold level. In this project, if a threshold is exceeded an alert is triggered and sent by both email and text message to a select list of recipients designated to respond by proceeding to the site for physical assessment to determine whether it may be a sinkhole forming. The alert recipients include key project management, the client's consultant, and local first responders (in this case the Hutchinson Fire Department). In case the warning is not acknowledged, the software automatically sends a reminder to the same list of recipients. The software structure offers a high level of self-diagnostic capability, and provides data and information to the users in an easy and fully understandable format.

13.10 Cable layout and user interface for real-time display of movement locations. (*Source*: Photo courtesy of SMARTEC).

To assess system capabilities in terms of ground deformation detection and alert triggering, some on-site pulling tests were carried out. These tests were intended to evaluate and confirm the integrated system performance, including the functionality of the sensor, reading unit, and data management software. The idea was to apply an external force to segments of the cable in portions of the trench that had not yet been backfilled in order to induce strain and to simulate the symptoms of ground deformation. For one test, a vertical force was applied on a cable section by raising the cable to different heights above the floor of the trench in order to simulate a highly localized ground deformation event – see Fig. 13.11. Different forces were applied to the cable during the test, to simulate different levels of ground displacement, with cable displacements of 15 cm, 30 cm, 61 cm, and 1.2 m. The test was repeated at several different locations to evaluate the capability of the system for reliably determining the exact location of ground strain events. The system proved capable of detecting movements of 20 cm at multiple test locations along the 4 km-long cable. All the tests demonstrated proper functioning of the system, both in terms of ground deformation detection and alert triggering with exact locations. The recorded results and graphs showed how different amounts of deformation of the cable can influence the distribution of strain along the sens-
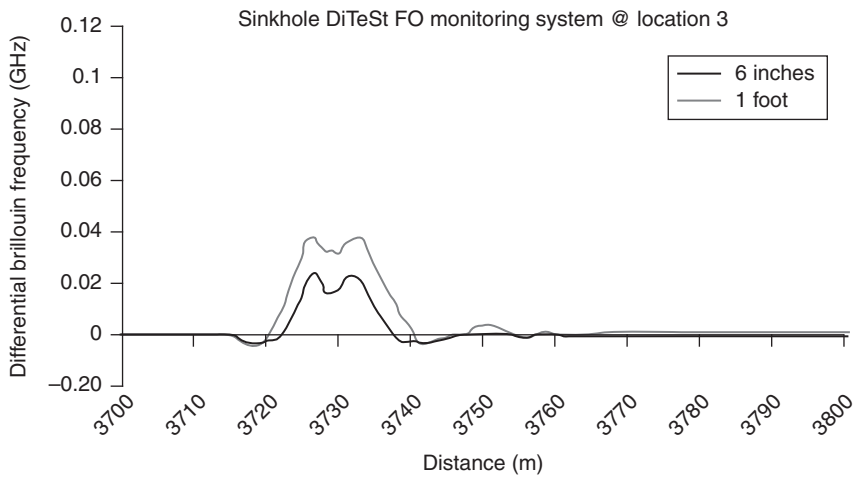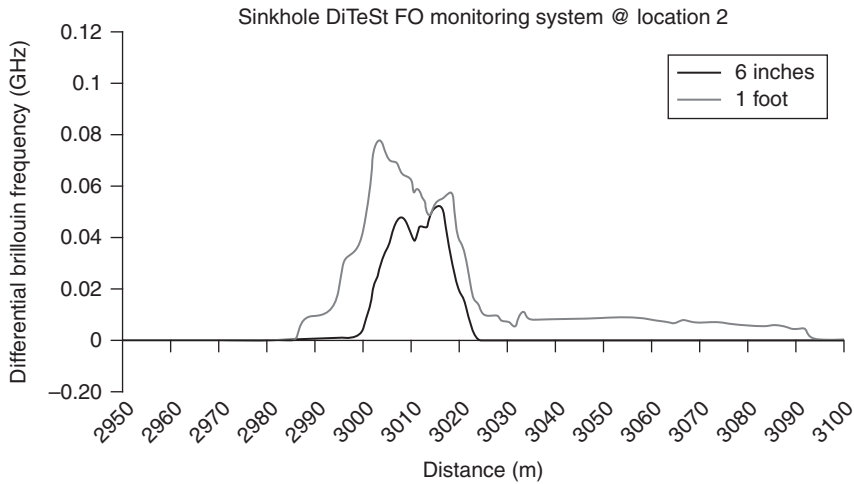
13.11 Results of field testing. Strain recorded for different levels of cable lift. (Shefchik *et al.*, 2011.)

ing cable. The data showed coherent behavior of the system at all of the test locations.

13.5 Landslide monitoring

Landslides are a type of ground instability that can manifest in different forms and is present worldwide in areas with mountainous morphology. These movements of rock or soil are characterized by a slow movement that can suddenly accelerate in the presence of heavy rain, freeze–thaw cycles, or earthquakes. Areas prone to these phenomena are typically

known and regularly inspected; however, it is sometimes necessary to implement a continuous monitoring program when the displacement rate increases or when inhabited areas and communication lines are present below the affected area. Since the moving mass is typically large and inhomogeneous, movements need to be detected at a large number of locations scattered over a large area. Since it is sometimes impossible to find suitable reference points for more conventional sensing technologies, distributed long-range sensing constitutes a valid and effective alternative.

In Fig. 13.12, a typical landslide area above a road is depicted. A distributed optical fiber strain sensor is installed in a serpentine pattern to allow detection of any displacement within the potential landslide area. The cable can either be installed on the surface of the landslide or buried in the ground as previously described for sinkholes and levees.

13.5.1   Application example: Korea landslide

An early application example of this technology was on a landslide area in Korea (Fig. 13.13). The sensing cable was installed above ground and attached to posts anchored within the slide area at regular distances. The anchors where specifically designed to follow the landslide movement, and the cable strain was measured at regular intervals to detect
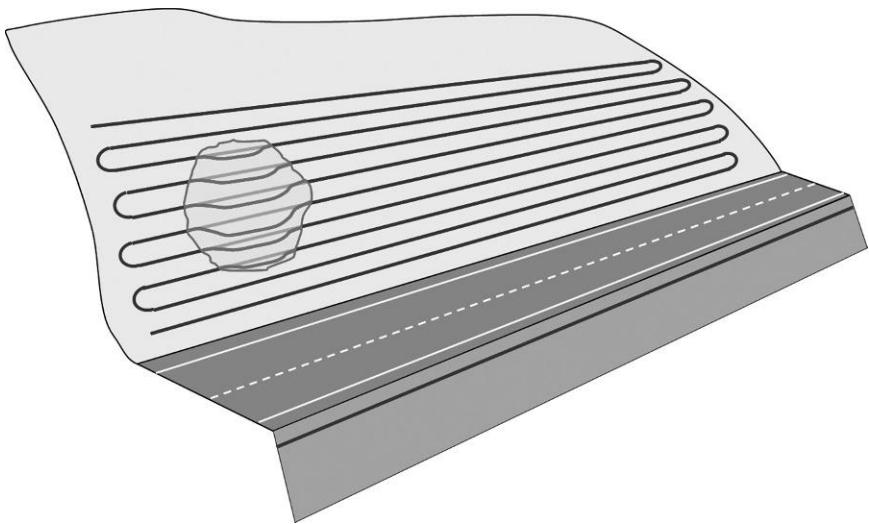
13.12 Installation of a distributed strain-sensing cable in a landslide area for detection and localization of movements. (*Source*: Picture courtesy of SMARTEC.)
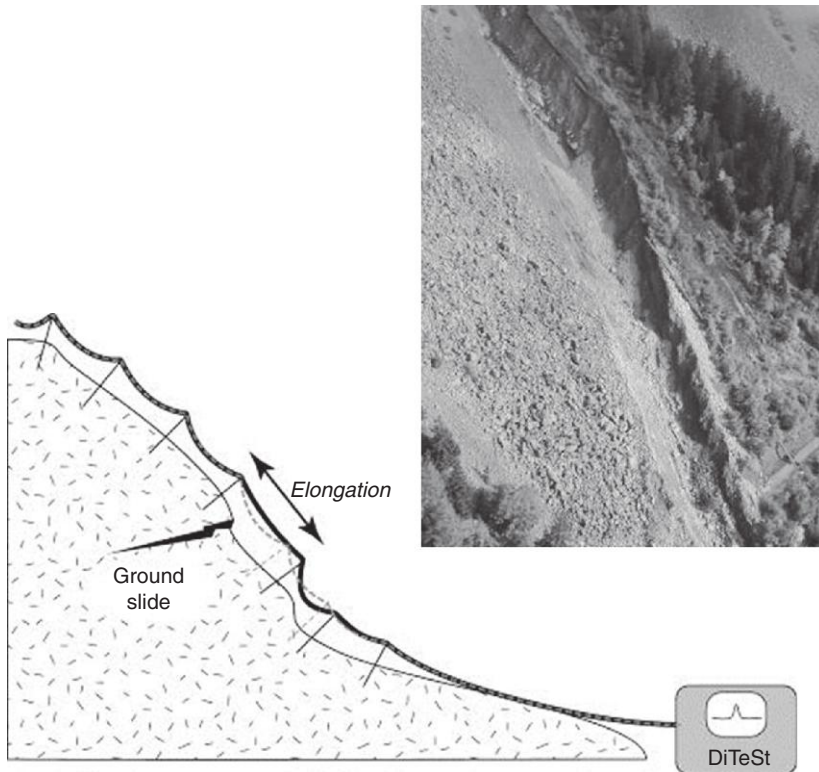
13.13 Landslide area and schematic installation overview. (*Source*: Picture courtesy Luc Thévenaz, EPFL.)

and localize movement areas. A typical measurement result is presented in Fig. 13.14. The comparison of the two measurements, carried out at different times, clearly shows the area affected by the movement. At the location where one anchor was moved relative to one direction, the cable tension increased on one side and decreased on the other side. Despite this capability of detecting strain variations, it is difficult to use such measurement for quantitative analysis of the displacements. Since the strain-sensing cables provide only elongation along their axis, the 3D movement of the landslide can be reconstructed only if a very dense 2D pattern is installed. This is often impractical or too expensive. Therefore, these systems are more useful for localizing the active faults and following the qualitative evolution of the movements to determine if acceleration is taking place. Once a problematic area is identified, it is possible to instrument it with more local sensors, such as laser distance meters or ground-based radar.
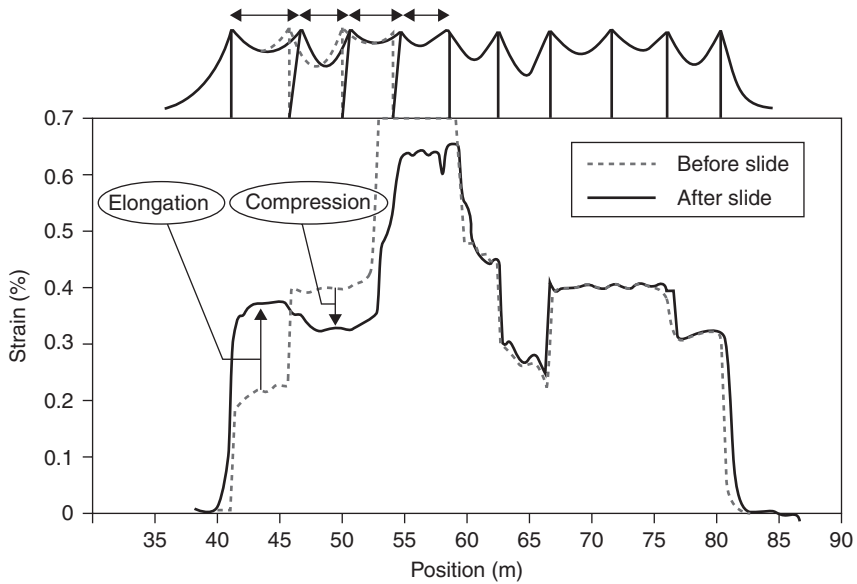
13.14 Comparison of two measurements, showing the location of the active landslide area. (*Source*: Picture courtesy Luc Thévenaz, EPFL.)

13.5.2   Application example: La Frasse landslide

A robotic laser distance meter unit was installed in June 2006 on the 'La Frasse' landslide in the Vaud canton, Switzerland. The landslide involves approximately 40 million cubic meters of material and its width varies between 500 m and 1000 m. This landslide is considered one of the most important and active landslides of the Alps, with average displacements of 30–50 cm/year. A total of 16 points (14 on the landslide plus two reference points outside the sliding area) were equipped with geodetic prisms and measured every 2 h. This landslide is also instrumented with four inclinometers. These are manually measured every 2 months. As shown in Fig. 13.15, the instrument was installed on a concrete pillar on the opposite side of the valley. The measured distances vary between 110 m (Reference 2) and 660 m (Reference 1) (Fig. 13.16). The PC managing the robotic laser distance meter is equipped with a WiFi/GPRS/EDGE card, allowing a real-time and high-speed remote access to the measurements. A direct connection to the 220 VAC power supply was available on-site. Plate XV in the color section between pages 374 and 375 shows the reactivation of the landslide in November 2007. The last important sliding event occurred during

13.15 Robotic laser distance meter installed on a pillar. (*Source*: Picture courtesy SMARTEC SA and EPFL.)

winter 1993–1994, with a total displacement of 3.5 m for this year and maximal sliding speeds of 1 cm/day.

13.6 Future trends

Global climate changes are making it increasingly difficult to reliably predict future extreme event probabilities based on historical data. While we are faced with increased demands on new and existing protective structures, the public expects a continuous improvement in safety and accident prevention. For these reasons, events such as landslides, levee breaches, and sinkholes, which were once considered a pure result of random fate, are
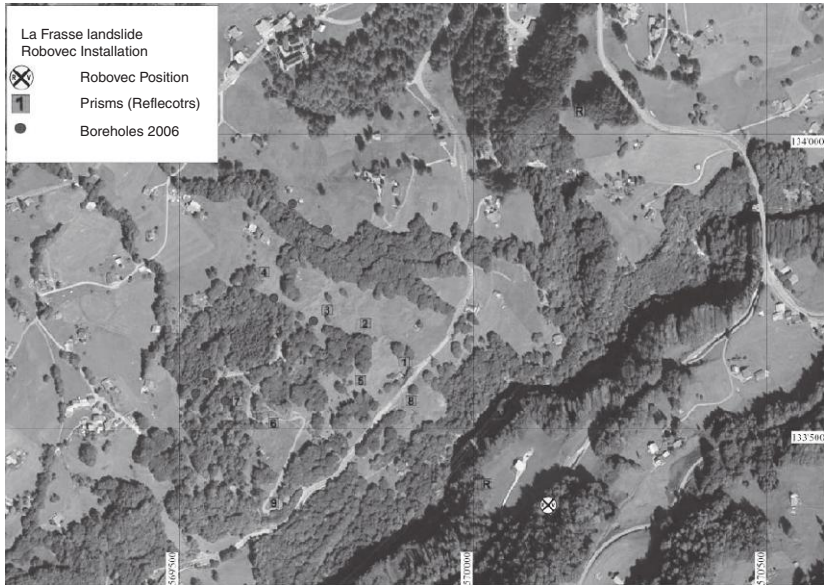
13.16 Overview of the landslide, with targets (1–9) and reference points positions. (*Source*: Picture courtesy SMARTEC SA and EPFL.)

increasingly seen as a failure of the owners to adequately protect the public. In this context it becomes important and valuable to adopt a preventive strategy so that early signs of distress that can point to possible vulnerabilities in case of extreme events can be identified. This new challenge can appear monumental because of the large areas potentially affected by such 'accidents.' A single sinkhole, landslide or levee breach that can happen at any location over tens of kilometers or length can cause an important communication line or transportation system to be interrupted or a large area to be flooded.

The traditional approach consisting of periodic visual inspection of those structures or the instrumentation of only a few isolated sections is not adequate in this new situation. The future trend will certainly move to a more pervasive use of sensing techniques that allow a multi-step approach to monitoring. At first level, the owner is interested in a screening of very large areas to identify possible distress locations. In this phase, the evolution of the damage is typically very slow, and techniques such as satellite-based SAR (Colesanti *et al.*, 2006) can cover very large areas very cost-effectively. FO distributed sensors offer the advantage of continuous measurement and high sensitivity but do require a more extensive installation of hardware on the ground. This can therefore be considered an adequate solution when an area of up to several tens of km has been identified as potentially dangerous.

Finally, traditional sensors, such as laser distance meters, total stations, piezometer arrays, settlement gauges, or tilt meters, can be applied locally in areas where distress has been specifically identified.

The availability of very large amounts of data from satellite-based sensing systems, distributed FOSs, point sensors and visual observation by inspectors and the general public also present a new challenge. It is increasingly important to manage all the data collected, allowing correlations between different types of observation to be found. In this context, geographic information systems (GIS) allow a systematic tagging and retrieval of information based on the physical location where the measurements have been recorded. Significant progress in this domain now allows for the treatment and storage of huge datasets that can be interrogated efficiently to discover small anomalies that would be impossible to discern by a manual analysis of the data. If, for example, the GIS system reveals an increase in strain in a distributed group of sensors, a settlement with a SAR measurement, and a soil crack reported by a neighbor, all at approximately the same location, the management authority would quickly dispatch an inspector to have a closer look there.

The real value and challenge of these applications therefore resides more and more in the ability to fuse and process this flood of data, and to pinpoint the few events that need attention from an expert.

13.7   Conclusions

The preceding case studies demonstrate how technology has advanced sufficiently to achieve blanket coverage for monitoring large-scale structures. The full-scale demonstrations and real world applications show that distributed FO monitoring and laser distance meters are viable and practical technologies for geotechnical applications. The potential for localizing and measuring strain, temperature and displacement over large areas makes these technologies particularly useful for monitoring levees, landslides and sinkholes as well as other geo-structures. These established technologies allow for wide area screening for potential trouble areas in aging structures as well as long-term monitoring for overall structural health. The continual aging of our infrastructure, emerging knowledge of historically unknown weaknesses, and greater potential impacts from extreme natural events, together with increasing population concentrations near potentially hazardous structures, raises the importance of and the need for large-scale structural monitoring and associated potential improvements to public safety. The use of FO and laser-based sensing technologies for these purposes is extremely cost effective in comparison to the near impossible task of providing total coverage using a conventional

sensor-based alternative. These technologies are able to provide real information that can be used to identify weaknesses, and repair and conserve existing geo-structures and infrastructure. Distributed sensing technology is a valuable tool that is particularly cost effective for large-scale monitoring. As our demands on deteriorating infrastructure continue to increase, along with the associated potential liabilities associated with infrastructure failure, opportunities for distributed and laser sensing are sure to follow.

13.8   Sources of further information and advice

Information about geotechnical instrumentation can be found in John Dunnicliff's book *Geotechnical Instrumentation for Monitoring Field Performance*. The book has not been updated recently with the latest sensing technologies, but Dr Dunnicliff regularly includes 'Geotechnical Instrumentation News' in the journal *Geotechnical News* published by BiTech and it is also available online. Every three months, the journal publishes articles and reports on geotechnical instrumentation and related topics.

   Further sources of information include international conferences such as the 'Symposium on Field Measurements in Geomechanics' and the SHMII conference series.

13.9   References

Artières, O., Beck, Y.L., Khan, A.A., Cunat, P., Fry, J.J., Courivaud, J.R., Guidoux, C. and Pinettes, P. (2010), 'Assessment of dams and dikes behavior with a fibre optics based monitoring solution', *International Congress on Dam Maintenance and Rehabilitation*, Zaragoza, Spain, 23–25 November.

Belli, R., Glisic, B., Inaudi, D. and Gebreselassie, B. (2009) 'Smart textiles for SHM of geostructures and buildings', *4th International Conference on Structural Health Monitoring on Intelligent Infrastructure (SHMII-4)*, Zurich, Switzerland.

Colesanti, C. and Wasowski, J. (2006), 'Investigating landslides with space-borne Synthetic Aperture Radar (SAR) interferometry', *Engineering Geology*, **88**, pp. 173–199.

Glisic, B. and Inaudi, D. (2006), 'Distributed fiber optic strain and temperature sensing for structural health monitoring', *IABMAS'06 The Third Int'l Conference on Bridge Maintenance, Safety and Management*, 16–19 July 2006, Porto, Portugal.

Glisic, B. and Inaudi, D. (2007a), '*Fibre Optic Methods for Structural Health Monitoring*', John Wiley & Sons Ltd, Chichester, West Sussex, England.

Glisic, B. and Inaudi, D. (2007b), 'Distributed fiber optic sensors: novel tools for the monitoring of large structures', *Geotechnical News*, **25**(3), pp. 8–12.

Manetti, L., Inaudi, D. and Glisic, B. (2008), '3-Demon monitoring platform: examples of applications in structural and geotechnical monitoring projects' *13th FIG Symposium on Deformation Measurement and Analysis*, Lisbon, 12–15 May.

Shefchik, B. , Tomes, R. and Belli, R. (2011), 'Salt cavern monitoring system for early warning of sinkhole formation', *Geotechnical Instrumentation News*, December, pp. 30–33.

14

Sensing solutions for assessing and monitoring pipeline systems

B. GLISIC, Princeton University, USA

Abstract: Pipeline systems are essential for the well-being and security of the people, the vitality of the economy, and the prosperity of society. Damage to pipelines can lead to potentially disastrous humanitarian, social, economic, and ecological consequences. Therefore, the assessment and monitoring of pipeline conditions are essential. An overview of sensing solutions for that purpose is presented. First, the most common types of pipeline and related damage are introduced. Then, current and emerging sensing solutions are presented, with accent on the applicability of each solution and its suitability to detect defects. The advantages and weaknesses are summarized, and a few illustrative examples briefly presented.

Key words: structural health monitoring of pipeline systems, typical damage and failure modes, current (traditional) sensing technologies, emerging sensing technologies, steel and concrete pipelines.

14.1   Introduction

A pipeline can be defined as a continuous, long, tubular structure used to transport significant amounts of liquid or liquefied materials and gases over long distances.[1] Pipelines allow for the delivery of resources, and provide transportation services such as for gas and liquid fuels, industrial materials, water supply, and sewage, which are essential for well-being and security, the vitality of the economy, and the prosperity of society.

Damage to pipelines can lead to leakage of transported material and reduction (or even ceasing) of the transport of life-sustaining supplies, with potentially disastrous humanitarian, social, economic, and ecological consequences. Therefore, the assessment and monitoring of pipeline conditions

---

[1] Solids can be packed in special containers (capsules) and sent through a pipeline (i.e., pneumatic tube) using compressed air, but this technique is limited to short distances and small quantities of material (e.g., to transport cash within a bank building), and as such is not an object of this study.

422

are essential for early emergency response and for the mitigation of environmental hazards. An overview of sensing solutions serving that purpose is presented in this chapter.

First, the most common types of pipelines and their related damage states are introduced. Then, current and emerging sensing solutions to identify pipeline damage are presented. The physical principles behind the solutions are simply outlined, giving more space to an analysis of the applicability of each solution and its suitability for detecting defects. The advantages and weaknesses of these sensing solutions are summarized, and a few illustrative examples are briefly presented. The chapter closes with future trends, a bibliography, and references.

14.2   Types of pipeline systems

Materials or goods transported by pipelines can be divided into four groups. The first group includes oil and gas: (i) crude oil (e.g., oleoducts); (ii) natural gas; and (iii) refined oil products such as gasoline, aviation gasoline, diesel, liquefied natural gas (LNG), and home-heating oil. The second group includes: (iv) clean water (e.g., aqueducts); and (v) wastewater (e.g., sewage). The third includes other industrial materials, such as: (vi) ammonia; (vii) hydrogen; (viii) carbon dioxide; and (ix) water suspensions of coal and ore (e.g., slurry pipeline). Finally, other materials are transported at a smaller scale using pipelines (e.g., beer, biofuels, etc.). For example, the annual mileage of gas and oil transmission systems in the United States of America in 2009 is given in Table 14.1 (adapted from PHMSA, 2011).

Oil, gas, and water are essential life and society supporting commodities, while sewage pipeline system is an important society supporting service. To identify assessment and monitoring needs and solutions, it is necessary to first understand specific pipeline properties, i.e., dimensions, material components, manufacturing process, construction process, and typical surrounding environments.

14.2.1   Oil and gas pipeline types

Crude oil and gas pipelines are subdivided according to their function as: (a) flowlines; (b) gathering pipelines; (c) transmission (or trunk) pipelines; and (d) distribution pipelines (Kennedy, 1993).

Flowlines are installed in oil and gas fields and connect individual wells to field central storage or field processing facilities. They are relatively short (e.g., a few miles) with outer diameters (OD) that range in size between 2 inch (50.8 mm) and 6 inch (152.4 mm). Flowlines transport a mixture of crude oil, gas, and water from the well to a tank battery, where these materials

Table 14.1 Mileage for gas and oil transmission systems in the United States for year 2009

(i)	(ii-a)	(ii-b)	(ii-c)	(iii)	(vi–vii)	(viii–ix)	(i)+(ii-a)+(iii)+ (vi–vii) + (viii–ix)
Crude oil	Natural gas transmission	Natural gas gathering	Gas distribution (main and service)	Petroleum[a] and refined products	Highly Volatile Liquids (HVL)	CO_2 or other	Total transmission
53 018	304 423	20 561	2 080 047	61 838	57 223	4192	480 694

[a] Condensate, natural gasoline, natural gas liquids, and liquefied petroleum gas (LPG).

Note: Although crude oil is also considered as petroleum, it is separately presented in this table.

Source: Adapted from PHMSA (2011).

are separated. Since they are relatively short, low operational pressures are sufficient for transporting the contents.

Gathering pipelines further transport crude oil or clean gas from a tank battery to a larger long-distance transmission pipeline. In some cases they are directly connected to wells. Gathering pipelines have a larger diameter, typically ranging between 4 inch (101.6 mm) and 16 inch (406.4 mm), and higher pressure is required to move their fluids. Pressurization is ensured by means of pumps (for crude oil) or compressors (for gas).

Flowlines and gathering pipelines are commonly made of steel, and coated externally to protect them against corrosion (Kennedy, 1993). Since corrosive agents are frequently mixed with crude oil and gas in wells, flowlines may be internally coated against corrosion, or made of some corrosion-resistant material such as plastic.

Transmission (trunk) pipelines transport crude oil to refineries or other storage terminals, or they may transport dry clean natural gas to utility companies and other customers. They feature large diameters, typically up to 56 inch (1524 mm) and in some cases even 80 inch (2032 mm). They also extend over very long distances (e.g., several hundred miles). Crude oil transmission pipelines require pumps at points of origin and pumping stations along the pipeline to maintain the pressure required to overcome friction, changes in elevation, and other losses. Gas transmission pipelines also require high pressure, which is ensured by a compressor at the beginning of the pipeline and compressor stations along the pipeline. Thus transmission pipelines are designed to support high operational pressures. They are made of steel, externally coated to protect them against corrosion; such pipelines are also typically buried (Kennedy, 1993). However, in some cases, when burying is impossible for cost or technical issues, the pipelines are built on the surface and may be additionally coated for thermal insulation (e.g., Trans Alaska Oil Pipeline). In some cases, gas and oil can be simultaneously transported. The design of these two-phase pipelines is complex due to several flow regimes that can occur inside the pipelines causing unpredictable pressure drops. Therefore, two-phase pipelines are used only when there is no economical or practical alternative (e.g., offshore pipelines).

Distribution pipelines are part of the network that delivers the gas to end users such as residential houses, industry, and businesses. They have small diameters and are commonly made of plastic.

Refined product pipelines transport oil products such as gasoline, aviation gasoline, diesel, LNG, and home-heating oil from refineries to storage terminals or utility companies. They can be hundreds of miles long, but they generally have smaller diameters, commonly around 8–16 inch (203.2–406.4 mm); however, they can also be as big as 26–28 inch (660.4–711.2 mm). Refined product pipelines operate under higher pressure than transmission

pipelines. In some cases, two different liquid products can be transported, with or without the use of a separator.

An important property of oil and gas pipelines is that they transport pressurized fluids. That is why these pipelines are made of steel and buried when possible. The pipelines are constructed by assembling a large number of individual pipe elements called line pipes (Kennedy, 1993). Depending on the application, various diameters and grades of steel (with different chemical compositions and physical properties) can be used according to specifications frequently proposed by trade associations (e.g., American Petroleum Institute in the USA (API, 2011)). In general, two types of line pipes are manufactured: seamless and welded. The former are made without longitudinal welds, while the latter have one or two longitudinal seams, or even spiral seams along their lengths. Metallic line pipes are externally protected against corrosion with a special coating (e.g., coal tar enamel, fusion-bonded epoxy, etc.), and the coating is additionally protected against mechanical damage using heavy paper wraps, plastic wraps, composite wraps and, in the case of offshore applications, concrete. Line pipes are internally protected against corrosion only if they transport corrosive materials (e.g., flowlines).

Line pipes are joined in the field to form a continuous pipeline. The joints are realized by: (i) welding (the most common); (ii) threaded coupling; and in rare cases by (iii) bell-and-spigot; and (iv) flanged joints. An example of a metallic gas pipeline with welded joints is presented in Fig. 14.1.

14.2.2   Water supply and sewerage pipeline types

There are more than 2.1 million km of water and wastewater pipelines across the USA (CBO, 2002) with water pipelines having the majority, totaling about 1.3 million km. Water supply pipelines are subdivided as per their functions as: (a) transmission lines; (b) distribution mains; (c) distribution networks; and (d) process lines (e.g., in treatment plants).

Untreated water is commonly transmitted from the natural source (e.g., river, lake, glacier, etc.) to the treatment (purification) facility using open surface channels (i.e., aqueducts), covered underground tunnels, and water transmission pipelines. The transmission pipelines are also used to transport clean water from the purification facility to storage facilities such as tanks, reservoirs, and towers. In both cases, the transmission is made over long distances, the pipelines feature large diameters and are usually buried. The water inside the transmission pipelines is mobilized by gravitational force or pumping. Further distribution from storage facilities to system and users is provided using a system of pipes consisting of water mains (i.e., the principal pipeline) and smaller branches that connect the individual users (i.e., distribution network).

14.1 Continuous gas steel pipeline with welded joints (Glisic and Inaudi, 2007). (*Source*: Courtesy of SMARTEC.)

Metallic line pipes were used in the past for water transmission and distribution pipelines, but due to corrosion they have been replaced with a large variety of materials such as concrete pressure pipe (CPP) and plastic. CPP is used in transmission, distribution mains and, to certain extent, process lines. Plastic pipes are used typically for short-distance transmission and distribution networks.

Manufacturing and specifications of CPP are commonly set by some regulatory body or professional association, (e.g., in the USA by the American Water Works Association (AWWA) or ASTM International). In accordance with AWWA, there are five types of CPP (ACPPA, 2011): (i) reinforced CPP (steel-cylinder type), two types of prestressed concrete cylinder pipe (PCCP) (ii) lined and (iii) embedded (iv) reinforced CPP (noncylinder type), and (v) CPP, bar-wrapped (steel-cylinder type). All cylinder type pipes have steel ring joints at their extremities while noncylinder pipes have bell-and-spigot joints.

Ninety out of 100 of the largest water utilities in the United States use PCCP in their water systems (Semanuik and Mergelas, 2006), and PCCP systems have the lowest water main break rate per 100 km than any other pipe material (Prosser, 1996). PCCP consists of a concrete core (i.e., the main structural element), thin steel cylinder (i.e., water barrier), prestressing strands wrapped around the steel cylinder (to guarantee that the core is always in compression), and a mortar coating to protect the prestressing strands from damage and corrosion (Pure Technologies, 2011). The diameter of this type of PCCP (also called 'lined PCCP') ranges between 16 and 60 inch (406.4–1524 mm). Another type of PCCP, called an 'embedded PCCP', consists of a steel cylinder embedded in the concrete core. The prestressing wires are wrapped around the concrete. Typical diameters of embedded PCCP systems range between 30 and 256 inch (762–6502.4 mm).

Sewerage pipelines can be subdivided into the following categories: (a) wastewater (gravity) sewers; (b) wastewater force mains; and (c) process lines (in wastewater treatment plants). Wastewater sewers are large underground networks of pipes that convey blackwater (i.e., wastewater from toilets containing fecal matter and urine), graywater (i.e., wastewater from domestic activities containing residues of washing processes), and stormwater (i.e., wastewater from precipitation events, rain and snow) from individual users and sites to a treatment facility, mainly using gravity but also pumps when necessary. Wastewater force mains convey wastewater from the discharge side of a pump or pneumatic ejector to a discharge point (EPA, 2000). Gravity sewers are made of polyethylene, plastic, or reinforced concrete, while force mains are mainly made of ductile iron, plastic, and CPP. An example of a gravity reinforced concrete pipeline with a bell-and-spigot joint is given in Fig. 14.2.

14.3   Typical damage and failure modes

The structural health condition and performance of a pipeline can be adversely affected over time due to: (i) damage or deterioration induced by environmental degradation, wear, and episodic events such as earthquake, vandalism, or impact; (ii) changes in operational and environmental conditions; or (iii) unintentional design, construction, and maintenance imperfections or errors.

Pipeline defects can be classified as built-in, long-term, and caused by one-time events. Built-in defects are generated during the construction of the pipeline, while long-term defects are the consequence of a deterioration process. Examples of built-in defects are offsets in alignment, bad realization of joints, flattened or ovaled pipes, sags due to settlement, etc. Examples of long-term deterioration processes are corrosion (internal and external), fatigue, excessive hydraulic flows, structural failures, leaks and

14.2 Reinforced concrete gravity pipeline with bell-and-spigot joint.

infiltrations, erosion, soil movements, etc. Finally, examples of one-time events that cause damage to a pipeline are third-party interference, terrorist activity, earthquakes, landslides, heavy construction over the pipe, etc. Factors that in general lead to pipeline degradation are schematically presented in Fig. 14.3.

Pipeline failure occurs on the pipe whose integrity has been adversely affected by the above defects and deterioration processes. Line pipe
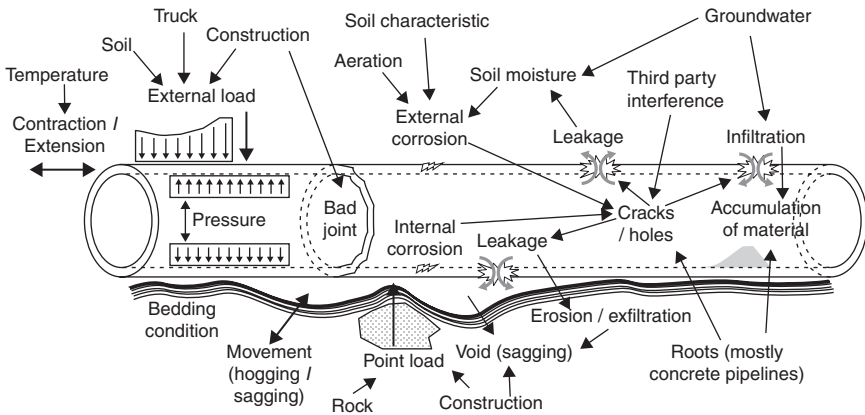
14.3 Illustration of pipeline degradation causes. (*Source*: Adapted from O'Day *et al.*, 1986.)

manufacturing defects, such as bad welds or dimensional imperfections, are frequently discovered during production, as the pipe is tested against standard and proven quality control. However, the pipeline in service will certainly be exposed to a variety of factors that can contribute to failure. The failure of the pipeline should be considered a consequence of malfunctioning of the 'system' (Cosham and Hopkins, 2004): e.g., the corrosion protection system will become faulty due to a combination of several factors such as aging of the coating, an aggressive environment, and rapid corrosion growth. Malfunction of the corrosion protection system will then lead to corrosion failure of the pipeline. Other causes of pipeline malfunction should be analyzed in a similar manner.

Steel and concrete pipelines have different material properties and different construction details; consequently, the damage types and manifestations are also different. That is the reason why they are presented separately.

14.3.1   Steel pipelines

For oil and gas transmission pipelines in Western Europe and North America, the most common causes of damage and failure are mechanical damage and corrosion (Cosham and Hopkins, 2004) – manufacturing problems are usually discovered during production. Steel pipeline defects can have various manifestations and mechanisms, as shown in Table 14.2.

Mechanical damage is induced by installation and assemblage (which cause gouges, coating damage, misalignment of joints, bad welds, etc.), impact (which causes coating damage, gouges, holes, and dents – buckles, wrinkles, etc.), heavy external load (which causes dents, ovalization, bending, etc.),

Table 14.2 Classification of steel pipeline defect types and manifestations

Type	Manifestation
Corrosion	Pitting (localized loss of material)
	Longitudinal, circumferential, or spiral corrosion (along welds, seams, gouges, etc., pits with one dimension significantly longer than the other)
	General (loss of material over an extended area)
Gouges	Longitudinal, circumferential, or spiral 'scars' on pipeline surface
Dents	Locally deformed wall of the pipeline in form of buckles, wrinkles, knobs, etc.; in general they are plain, kinked, smooth on welds, smooth containing gouges, or smooth containing other types of defects
Manufacturing defects	Ovaled shape of the cross-section, rolling imperfection, wrinkle, seam weld defect, etc.
Weld defects	Seam weld defects (during manufacturing or repair) and girth weld defects (e.g., at joints during assemblage of the pipeline)
Deformation	Ovalization of cross-section, bad/sharp angle between line pipes, bending, etc.
Cracking	Discontinuity through the wall of the pipeline with almost no loss of material and no displacement of material
Environmental cracking	Corrosion cracking, which is caused by a combination of conditions that can specifically result in stress corrosion cracking, corrosion fatigue, and hydrogen embrittlement
Hole	Discontinuity through the wall of the pipeline with significant loss or displacement of material

Source: Adapted from Cosham and Hopkins (2004) and El-Sayed (2011).

soil movement (which causes dents, deformation, cracking, etc.), and third-party interference (e.g., vandalism, theft, terrorist acts, etc.).

Corrosion can be internal or external. Internal corrosion is caused by aggressive chemical action of transported media (which causes general corrosion), the wrong selection of pipe material (which causes general corrosion), and bad welding (which causes corrosion along the welds). External corrosion can be the consequence of inappropriate coating (which causes pitting, general corrosion, etc.), inadequate cathodic protection (which causes general corrosion), damaged coating (which causes pitting, corrosion along gouges, etc.), and bad welding. Corrosion, if not stopped, can lead to environmental cracking or puncture. An example of an externally corroded pipeline with flanged joint is given in Fig. 14.4.

14.4 Externally corroded pipeline with flanged joint.

14.3.2   Concrete pipelines

The main causes of long-term deterioration of concrete pipelines are sulfate corrosion resulting from sewer gases, excessive hydraulic flows, thermal strain and deformation, the settlement of soil, leaks and infiltrations, and erosion (Najafi, 2004). The pipeline whose structural integrity has been adversely altered is likely to experience pipe breakage. The manifestation of pipeline breakage can be classified in seven categories as presented in Table 14.3.

The mechanisms of concrete pipeline deterioration are structural (cracks, fractures, etc.), hydraulic (insufficient capacity, flooding, debris, encrustation, and grease), corrosion (sulfate and other chemical corrosion), erosion, and operational problems (roots, blockages, maintenance procedures, etc., Najafi, 2004). The two most significant failure mechanisms are structural and operational.

Structural defect failure mechanisms (e.g., crack, fractures, etc.) occur when the forces around the pipeline exceed ultimate limit states, or the

Table 14.3 Classification of concrete pipeline breakage manifestations

Manifestation	Description
Longitudinal cracks	Occur at springing level, at the crown and invert; a consequence of excessive crushing or ring stresses (e.g., due to external heavy load over the pipe)
Circumferential cracks	Consequence of shear or bending stresses induced by relative vertical and/or horizontal movement of successive line pipe lengths due to differential settlement in soil (e.g., caused by inadequate trench and bedding practices, large voids due to leakage,) and/or permanent ground displacement due to land sliding or earthquake; can also be a consequence of thermal stresses (contraction of pipe) and third-party interference (e.g., accidental breaks); in case of concrete pipes, most likely (but not only) occurring near joints
Tension (diagonal) cracks	Consequence of concentrated load, e.g., due to hard spot beneath the pipe
Broken pipe	Occurs when pieces of a cracked and fractured pipe visibly move from their original position; it is a very serious defect that represents an advanced damage stage of a cracked pipe
Deformed pipe	Occurs when a longitudinally cracked pipe loses the support of surrounding ground; it is a very serious defect that represents an advanced damage stage of a cracked pipe
Socket bursting	Consequence of excessive pressure inside the joint due to the expansion of the jointing material
Holes	Consequence of corrosion or impact

Source: Adapted from O'Day *et al.* (1986), Najafi (2004), Davies *et al.* (2001).

pipeline materials lose the ability to resist existing forces, or both are combined (Najafi, 2004). Slow processes that can cause structural defects, besides corrosion, are infiltration of groundwater through existing defects (e.g., in sewers) and creation of void formation in the soil due to exfiltration (water leaves the pipeline through an existing defect and scours the pipeline). Dynamic forces can also create defects – one-time events such as heavy loading over the pipeline (e.g., heavy truck on a construction site above a pipeline) or earthquake, or smaller cyclic events that occur daily or seasonally (e.g., routine track machinery above ground, frost heave, etc.). An example of the failure of concrete pipeline at the bell-and-spigot joint due to permanent ground deformation is given in Fig. 14.5.

Operational defect failure mechanisms happen when demand increases, or capacity decreases, or both are combined. Demand depends on infiltration
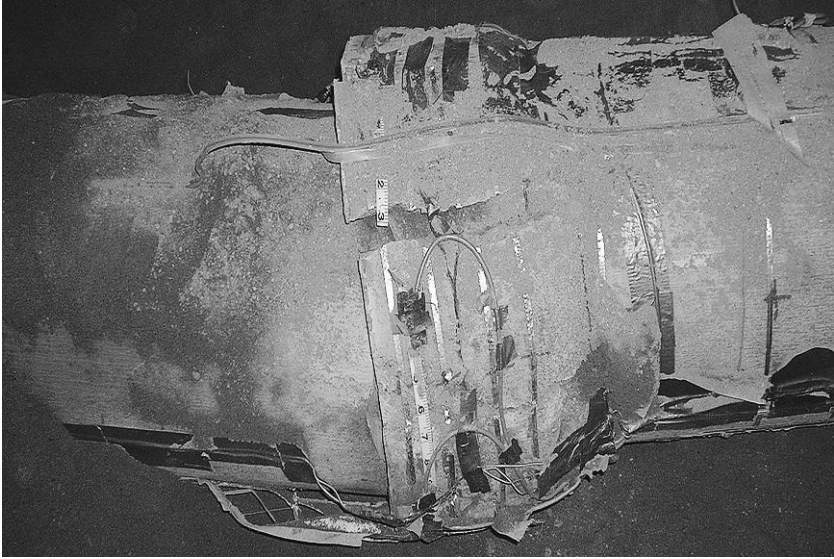
14.5 Crushing of the bell-and-spigot joint of concrete pipeline due to permanent ground deformation.

and inflow, and it increases as the number of structural defects increases. Capacity depends on the effective diameter of the pipeline and its roughness coefficient. Structural defects reduce the effective diameter (e.g., open and misaligned joints and broken sections) and increase the roughness coefficient (e.g., cracks). Root masses, grease and collected debris in sewers additionally decrease the effective diameter.

14.3.3   Example of a pipeline failure

A case of a real-world pipeline failure is presented in this subsection in order to show: (i) the economic losses and the extent and the severity of the resulting damage to society; (ii) the direct causes of the failures; and (iii) the circumstances and contributing factors.

A 30-inch (762 mm) diameter natural gas transmission pipeline (Line 132) broke and burst at the intersection of Earl Avenue and Glenview Drive in the residential area of San Bruno, California, on 9 September 2010. The burst created a crater with an approximate length of 72 feet (22 m) and approximate width of 26 feet (8 m). In the process, a 28 foot long (8.5 m) pipe segment was thrown 100 feet (30 m) away from the crater. Approximately 47.6 million standard cubic feet (1.35 million cubic meters) of natural gas (NTSB, 2011) was released in the air. The released natural gas was ignited and a resulting fire destroyed 38 homes and damaged another

70. Eight people lost their lives, several were injured, and many more had to be evacuated. The cost of repairing the pipeline was approximately $13 500 000, and the loss of natural gas approximately $263 000. The estimation of costs associated with human victims and injuries, and destruction and damage to homes, was disputed in a court at the time of writing this book. It took 95 min for the owner and the operator of the pipeline, the Pacific Gas and Electric Company (PG&E), to stop the flow of the gas and isolate the rupture site. Such a slow response and lack of automatic or remote control safety valves significantly contributed to the severity of the damage (NTSB, 2011). The damage was contained only by the prompt response of the local fire department and emergency responders.

The National Transportation Safety Board (NTSB) determined that the probable cause of the accident was a seam weld flaw in a poorly welded pipe section. The flaw grew over time to a critical size, 'causing the pipeline to rupture during a pressure increase stemming from poorly planned electrical work at the Milpitas Terminal' (NTSB, 2011). As the flaw was visible, it was concluded that the PG&E had inadequate quality assurance and quality control in 1956 during its Line 132 relocation project; this allowed the installation of this defective pipe section. Furthermore, an inadequate pipeline integrity management program failed to detect and repair or remove the defective pipe section (NTSB, 2011).

California Public Utilities Commission's (CPUC) and the U.S. Department of Transportation's exemptions of existing pipelines from the regulatory requirement for pressure testing was identified as a contributing factor to the accident. It is likely that pressure testing would have detected the pipeline defects. An additional contributing factor was the CPUC's failure to detect the inadequacies of PG&E's pipeline integrity management program.

The San Bruno case reveals that the failure happened not only due to a combination of several factors, including initial flaws in the constructed pipe, but also due to the inadequacy of the quality assurance and integrity
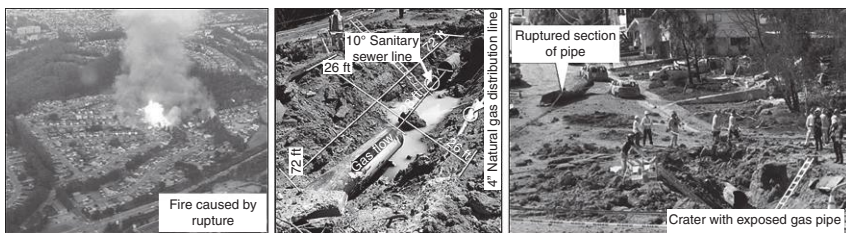
14.6 Photographs of ruptured Line 132 at San Bruno, California. (*Source*: Adapted from NTSB, 2011, courtesy of The National Transportation Safety Board.)

management program of the utility company (PG&E). Exemptions from regulatory requirements were also an issue. Photographs taken after the explosion are shown in Fig. 14.6.

14.4   Current sensing solutions for pipeline systems

In order to assess the structural health condition of pipelines, inspections are regularly performed. Historic experience and recent studies have shown that vigilant inspection offers significant cost savings (Ciocco *et al.*, 2002). Pipelines installed on the surface are visually inspected frequently by a trained professional. Buried pipelines are also inspected visually by traversing the pipeline route on the ground, or patrolling the route with a light airplane (Kennedy, 1993). However, these visual inspections can only identify damage that is visible on the surface and cannot provide an assessment of the internal pipeline condition. Increased safety, economy, and ecological concerns, combined with general technological progress, created settings for the development and application of various sensing solutions that provide for a more accurate assessment of the pipeline health condition. The most frequently applied solutions are presented for steel and concrete pipelines. Other solutions can be found in the literature presented in Sections 14.7 and 14.9.

14.4.1   Steel pipelines

A simple technique used to detect leaks or failures includes hydrodynamic modeling of the pipeline flows using in-line flow meters and pressure sensors. Discrepancies between measurement results and the model indicate potential leaks or failures within the pipeline (Liu, 2003). Another simple technique is based on metering accuracy: leaks and failures within pipe segments can be identified by direct observation of pressure drops and volume loss, based on comparisons of the flow into the segment and the flow out of the segment (Kennedy, 1993). More sophisticated acoustic emission sensors can be used to sense the noise generated by the fluid flowing from a leak in a pressurized pipeline. The acoustic emissions are transmitted by the pipeline and captured by sensors installed on the pipe walls at given intervals. The leakage noise is then compared with the common background noise of the system to detect the leak. The advantages of all these methods are that they are simple and allow for continuous and automated monitoring. A major disadvantage though is that they can indicate damage only after the rupture of the pipe. The additional disadvantages of the first two techniques are the lack of specificity in determining the leakage point (these techniques identify the defective segment between two sensors rather than the defective point within the segment).

For more sophisticated structural health assessment of the pipeline, in-line tools are recommended.

Existing technologies employed for inspection of buried pipelines are mainly based on the use of devices that can be inserted into the interior of the pipe (in-line tools). These devices may contain various types of sensors, with the specific sensors depending on the type of the pipe. The most common device used for inspection of steel pipelines is the so-called pig (Liu, 2003). This is a small package containing one or several types of sensing transducers such as magnetic flux leakage (MFL), remote field eddy current (RFEC), ultrasonic transducers, etc. (Kobayashi *et al.*, 1999). The MFL set-up consists of a circumferential array of strong permanent magnets and a leakage flux detector. The magnets magnetize the pipe wall with a near saturation flux density. Defects in the pipe wall create MFL near the pipe's surface. These leakage fluxes are detected by a leakage flux detector (Hall probes or induction coils moving with the magnets). A scheme of a pig with MFL set-up and MFL principle is shown in Fig. 14.7. RFEC uses low frequency alternating current (AC). Electrical coils on the pig need to generate eddy currents in the pipe walls (Najafi, 2004), while damage (such as corrosion or cracking) alters their flow through the pipe walls and is detected as a variation in the remote magnetic field. This is schematically represented in Fig. 14.7. The through-wall nature of the RFEC technique allows external and internal defects to be detected with similar sensitivity. More information about electromagnetic techniques is given in Chapter 12.

Ultrasonic transducers are used to assess the distribution of the thickness of the pipe walls along the pipe length, and to detect any damage due to corrosion as a change in the wall thickness. Other examples of devices
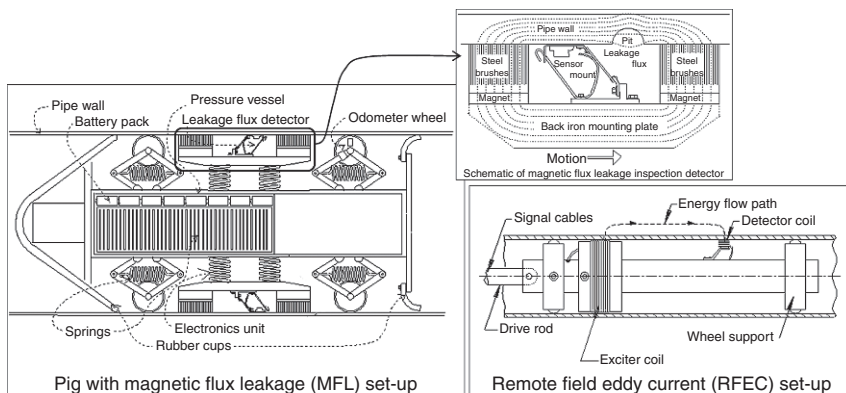
14.7 Schematic representation of pig device with MFL set-up, MFL technique, and RFEC technique. (*Source*: Adapted from (Atherton, 2011), courtesy of Prof. David Atherton, Queen's University, Kingston, Ontario, Canada.)

that can be mounted on the pig are gauging tools and cameras (Kennedy, 1993). Gauging tools measure the inside diameter of the pipe to locate dents, ovalization, changes in diameter, etc. Typically, spring-loaded probes are in direct contact with the pipe internal wall and measure the actual diameters in the cross-sections as the pig moves along the pipeline. The cameras can be set to take pictures in predefined positions (e.g., welds), or at locations where the damage is detected by other tools on the pig. For example, a gauging probe that enters a corrosion pit can trigger the camera to photograph the pit. The cameras can be used in gas pipelines while in service because of the transparency of the gas. The use of cameras in oil pipelines requires the transmission of oil to be stopped and the pipeline to be emptied. The advantage of the use of the in-line tools is an accurate assessment of the pipe health condition. The disadvantage is that they allow for only periodical inspections and cannot be used for automatic and continuous health monitoring.

Acoustic and ultrasonic transducers can also be installed on the surface of the pipe. Piezoelectric elements can be mounted on the walls of the pipe and they can introduce guided elastic waves, called Lamb waves, into the walls of the pipe. The propagation properties of waves, such as attenuation, velocity, or reflections can be correlated with the health condition of the pipe walls. In steel pipes the Lamb waves can be directed (Rose, 1999; Towfighi *et al.*, 2002) and can be made to propagate over distances of up to 1 km. The advantage of ultrasonic methods is that they can interrogate large areas of the pipeline and reliably detect and locate the damage. The main disadvantage is that a large number of piezoelectric transducers require cabling for automated monitoring, which significantly increases the cost of monitoring. More information about acoustic and piezoelectric techniques is given in Chapter 18.

14.4.2   Concrete pipelines

Inspection and monitoring techniques for concrete pipelines depends on the type and purpose of the pipeline. Gravity pipelines have the advantage that they are not filled completely with fluid. Consequently, many inspection techniques involving in-line tools can be applied to gravity pipelines while they are in operation. Different techniques are used for pressurized (e.g., clean water) concrete pipelines.

Gravity pipelines (e.g., sewers) are most commonly inspected using closed-circuit television (CCTV) (Najafi, 2004). Cameras are either simply installed at a manhole access point or are placed on a robot and moved through the pipeline using a robotic system (Sinha and Fieguth, 2006). Cameras installed at a manhole access point have a limited area of observation, while cameras

placed on the robot can be used only if not more than 25–30% of the pipe's depth is filled with water. CCTV provides important information about visible defects in the pipe. However, it may miss very small defects; it covers only the area of the pipe above the water, and physical obstacles frequently found in the pipeline (e.g., debris) can reduce visibility and accessibility to the pipe walls (Najafi, 2004).

Sewer scanner and evaluation technology (SSET) is a significant upgrade of CCTV. Besides ordinary cameras that provide one with a forward view, SSET has side scan camera that allow one to 'open' the cross-section of the pipe. Side scan picture combined with a forward view can allow one to reconstruct 3D images of the pipe interior, and to detect and locate defects automatically. The advantage of the method is greatly improved reliability in detecting the defects. The disadvantage is again related to obstacles in the pipe, which is similar to CCTV.

Besides the use of a camera, ultrasonic systems can also be used when the internal area of a pipeline is accessible (Wirahadikusumah *et al.*, 1998). It is particularly useful in cases where the pipe is significantly filled with liquids, for example 25–75% of the depth. Ultrasonic systems can be combined with CCTV on the same device; CCTV is used above the liquid level, and the ultrasonic system is used below that level. The fundamental principle of the ultrasonic system is similar to the case of still pipelines. One example of an ultrasonic system is the rotating sonic caliper. The ultrasonic pulse is emitted from the source inside the pipe in one radial direction; the pulse amplitude and return time, as a result of reflections from the inner pipe wall, is used to measure pipe thickness and identify pipe cracks (Wirahadikusumah *et al.*, 1998). The ultrasonic source can rotate in all directions, allowing the circumference of the pipe at a location to be completely examined. The advantage of the ultrasonic method is that it can detect cracks that are difficult to spot by cameras. The disadvantage is that for concrete pipes it cannot inspect beyond the inner pipe surface, due to high attenuation of the ultrasonic signal (Duran *et al.*, 2002).

Other sensing solutions that can be used as in-line tools for gravity pipelines (i.e., mounted on robots and pulled through the pipeline) are focused electrode leak location (FELL) technology and laser-based scanning systems. FELL consists of a probe that travels through the pipe and generates an electric field, and an electrode located on the top ground surface. The electrical resistance between the probe and the electrode depends on the condition of the soil: dry soil is non-conductive while wet soil is conductive. Thus leakage and/or infiltration points along the pipelines can be identified and rated based on the intensity and duration of signals originating from the pipeline and captured at the surface electrode. The advantage of the method is reliable detection and rating (or quantification) of damage. The

disadvantage is that it can be applied only during dry weather. The laser-based scanning system can be used to evaluate both the shape of the pipe and the defects that occur on the inner wall surface. This method provides inspectors with very accurate and reliable information, but it is effective only for that part of the pipe above the water level. Another advantage of the method is that the data analysis is fully automated.

The general disadvantage with all of the above solutions that include robotically driven instrumentation is that they are usable only for periodic inspections, and not for continuous, real-time monitoring. Furthermore, robotic driven instrumentation can frequently get stuck in the pipeline.

The sensing solutions for buried pipelines presented so far mainly focus on the assessment of the condition of the pipeline wall. Although some of these solutions can give some information about the pipeline condition behind the wall, this is not their primary objective. For this purpose, several remote-sensing techniques have been developed.

Ground penetrating radar (GPR) emits a coherent beam of radio waves into the observed medium. The waves travel through the medium (e.g., soil) until they hit another medium with different conductivity (e.g., void), whereupon one part of the wave reflects back to the emitter and the other part continues to propagate. The reflected part of the wave contains information on the position of the medium that generated the reflection. GPR can be used to detect delaminations in concrete sewers; however, the main use of GPR is to examine the bedding of the pipeline (Najafi, 2004) (i.e., to detect rocks, voids, and regions of water saturation produced by exfiltration, as potential areas of deformation and damage to the pipeline). The wave propagation distance depends on the medium – it will be immediately attenuated in highly conductive or magnetic materials (e.g., steel), as well as in materials with high dielectric constants. However, in usual soils such as sands, clays, rocks, etc., they have penetration properties that highly depend on the water content of the soil (higher water content translates into lower penetration). Dependence of the performance of the GPR on the soil properties is a disadvantage of the method. Additional disadvantage is the intensive labor necessary to scan larger areas.

Another remote sensing technology for monitoring concrete pipelines deployed above the surface is the infrared thermography system (ITS), which captures thermal images of soil altered by leaks due to the rupture of a pipe (Weil, 1998). However, the implementation of reliable image interpretation is challenging.

CPPs face a different set of issues as compared to the gravity pipes. First, their degradation rate is less dependent on their age, and more a consequence of cumulative effects generated by various external influences

(Najafi, 2004). In addition, the defects that occur on CPPs are not readily visible and the consequences of failure are often very severe.

Remote eddy field current/transforming coupling (REFC/TC) is one of the viable sensing solutions for CPP and in particular for PCCP. The main advantage over conventional REFC is that it senses beyond the inner wall surface (i.e., it reaches the outer wall and the changes beyond the outer wall, external to the pipe). It uses low frequency AC and through-wall transmission to inspect electrically conducting tubes or pipes (such as those embedded in CPP). The REFC/TC is used to detect, localize, and quantify broken prestressing wires in PCCP (Grigg, 2006). It can quantify wire breaks in embedded cylinder pipes, lined cylinder pipes and even noncylinder pipes (Najafi, 2004). The disadvantage of the method is that it is a manned inspection technique, and therefore used only during periodic inspections.

Acoustic emission testing can also be applied to PCCP. The main principle behind the method is that breaking prestressing wires in the pipeline wall, and subsequent slippage of the wire, produces a transient noise that can be detected by acoustic receivers placed along the pipe. The acoustic wave travels through the water and, knowing the speed of the water and the time of flight of the acoustic wave to various receivers installed along the pipeline, it is possible to pinpoint the location of the broken wire. Acoustic emission can also be used to detect leakage, as leaking fluid will create its own acoustic signal. The advantage of the system is that it can be used for on-line monitoring. The disadvantage is that a large number of acoustic receivers are needed to cover large lengths of the pipeline, which increases the cost of monitoring.

14.5   Emerging sensing solutions

Although several sensing technologies are employed for assessing pipeline health condition, none of them is fully suitable for real-time automated operation. Most of these technologies are employed in periodic inspections and does not provide for continuous assessment of the pipeline health condition. In addition, the technologies generally require manual operation and, in some cases, manual data analysis, which adds subjectivity to the process. On the other hand, there is a clear need for monitoring systems that can detect defects and degradation in the pipelines at an early stage, and inform responsible managers about it. The monitoring system should be able to provide the information both automatically and on-demand, and the information should be reliable, objective, and comprehensible. Finally, the technologies used in monitoring should be low cost (to accelerate commercial adoption) and applicable to both new and existing pipelines. With

the rapid development of technologies over the last twenty years, there is a long list of emerging sensing solutions that have promising potential in pipeline monitoring. Three sensing solutions have been identified as particularly promising, as they focus on the early detection of defects and degradation (e.g., corrosion, excessive deformation, strain concentration, etc.) before major damage occurs (e.g., leak or rupture). These three solutions are wireless technologies, distributed fiber optic sensors (FOS), and large area electronics (LAE).

14.5.1   Wireless technologies

Traditional sensors used in structural health monitoring are commonly attached to the reading unit using cables. The reading unit then transfers data to a central data repository, where all the data are collected and analyzed. However, the use of cables between the sensors and the reading unit generates considerable cost, especially in terms of cable installation. For example, in commercial building monitoring, the cost of the cabling can raise the average cost per sensor channel to $5000 (Celebi, 2002). An important attribute of pipelines is their length, which makes the use of traditional wired sensors for continuous monitoring unaffordable or economically unjustifiable.

Wireless technologies are proposed to overcome the cost limitations imposed by traditional wired sensing solutions for large structures (Straser and Kiremidjian, 1998; Lynch, 2002; Spencer *et al.*, 2004). The sensors can communicate with the repository by means of wireless communication, provided an analog-to-digital converter and microcontroller are integrated in the wireless node. This opens the door to a completely new paradigm – to perform distributed and decentralized in-network reading and processing of data instead of traditional centralized reading and processing at a central repository. This approach has been successfully tested on large bridges (Lynch *et al.*, 2006), indicating that similar applications may be possible on pipelines. More details on wireless technologies are given in Chapters 1, 9, and 10, and in this section only the advantages and challenges related to pipeline monitoring are presented.

The great advantages of wireless technologies are their low cost, decentralized in-network data collecting and processing, and large flexibility in interfacing with many existing sensing solutions. In fact, many traditional sensors already used in pipeline health evaluation, such as acoustic emission sensors, and also strain gauges, accelerometers, etc., can be interfaced with wireless nodes. Thus, wireless technologies may enable a wide range of sensors to be massively deployed in structural health monitoring of pipelines. In-network data processing allows for minimized communication between the sensors and the central repository, because the

communication is needed only when the sensors detect defects or degradation processes. This significantly reduces the power consumption of wireless nodes and ensures that only useful data are kept in the central repository.

One important challenge of wireless technologies is power. Power is conventionally provided by cables, or using batteries. In the case of buried pipelines and buried wireless nodes, laying electrical power cable or changing batteries is particularly impractical. However, the wireless nodes feature low power consumption, and that is why on-site power harvesting solutions can be applicable. As the pipeline transports fluids, installation of commercially available power harvesting turbines within the pipe can provide the wireless nodes with power. Other alternatives are providing the power from above the ground surface using wired photovoltaics, or by electromagnetic coupling. Another challenge is high attenuation of wireless signal in the ground. This limitation can be minimized by using appropriate radio frequencies. Ongoing research in the field is intended to overcome challenges related to the use of wireless technologies for pipelines monitoring (e.g., Kim *et al.*, 2010). An example of a wireless node connected to sensors installed on the pipeline and embedded in soil is given in Fig. 14.8.
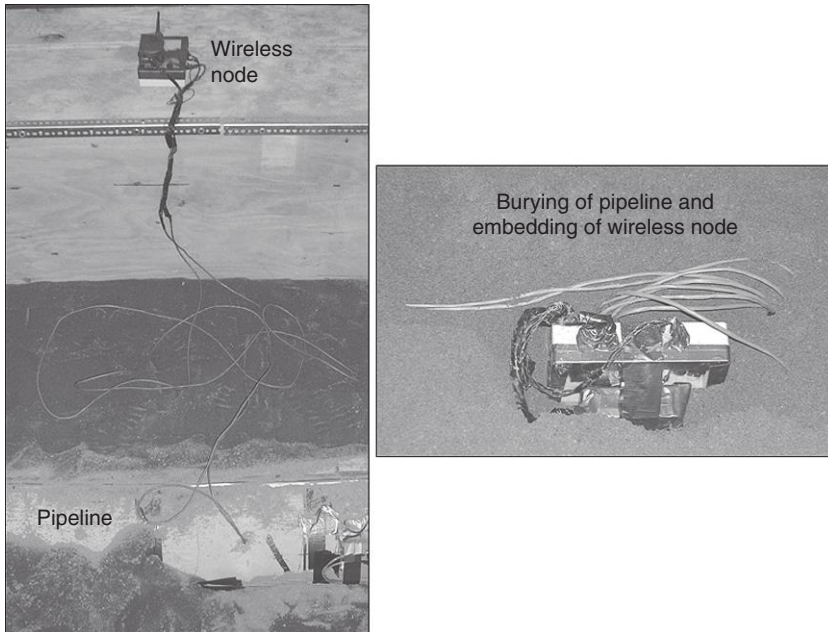
14.8 Example of wireless node connected to sensors installed on pipeline and embedding in soil.

14.5.2   Distributed fiber optic sensing technologies

A distributed sensor (or sensing cable) can be represented by a single cable that is sensitive at every point along its length. Hence, one distributed sensor can replace a large number of discrete sensors. Moreover, it requires only a single connection to transmit the information to the reading unit, instead of a large number of connecting cables required in the case of wired discrete sensors. Finally, distributed sensors are less difficult and more economical to install and operate on large structures. An illustrative comparison between pipelines equipped with distributed and discrete sensors is shown in Fig. 14.9 (it should be noted that this schematic drawing does not refer to a real case – e.g., redundancy is not included).

Distributed fiber optic technologies can be used to sense strain, temperature, and vibration (acoustic waves). There are three main principles for distributed sensing in the domain of FOS: Rayleigh scattering (e.g., Posey et al., 2000), Raman scattering (e.g., Kikuchi et al., 1988), and Brillouin scattering (e.g., Kurashima et al., 1990). These technologies are presented in more detail in various relevant literature (see also Chapters 13 and 18), and this section focuses only on their application to pipelines.

Strain sensors can be installed directly on the pipeline (Inaudi and Glisic, 2010), by bonding, or can be embedded in the soil. Sensors installed on the pipeline can be combined in so-called topologies (Glisic and Inaudi, 2007), which besides the strain level profile along the pipeline, also provide information on general deformation of the pipeline, such as bending and axial deformation (e.g., using 'parallel topology' i.e., topology consisting of three
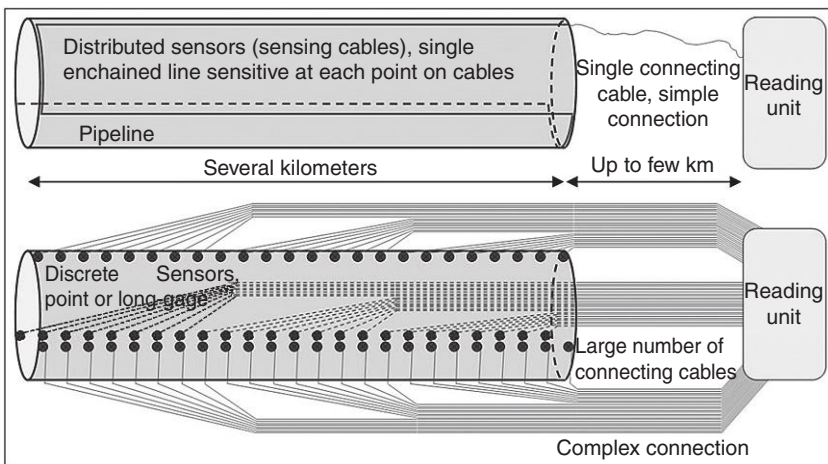
14.9 Distributed vs discrete monitoring, schematic comparison (does not refer to real case, Glisic and Yao, 2012.

or more parallel sensors, as shown in Fig. 14.9). The sensors embedded in the soil cannot provide information on spatial deformation or damage to the pipe, but they can detect and localize the movements in the soil that can potentially imperil the pipe, and point to endangered locations. Although this approach is not direct, it can be very efficient if proven to be reliable, since the embedding in the soil is significantly simpler and faster than bonding sensors onto the pipeline (Glisic and Yao, 2012). Strain and deformation sensing are largely based on stimulated Brillouin scattering effect, as it allows ten-kilometric lengths to be monitored with a single reading unit. Examples of distributed strain sensors installed on concrete pipeline and in the soil are given in Fig. 14.10. A steel pipeline equipped with distributed sensors is shown in Fig. 14.11 (the position of sensors on the pipeline is as shown in Fig. 14.9, only parts of installed 'top' and side sensors are visible in Fig. 14.11).

Sensing the temperature profile along the pipeline can also provide leakage detection capability since pipeline damage is often correlated with leakage that can be indirectly detected as a change of thermal parameters in the surrounding soil (Inaudi *et al.*, 2007; Nikles *et al.*, 2004). Gas leaking from a pipeline generates a cold-spot due to gas pressure relaxation and related gas cooling. Hot liquid leaking from pipelines generates a hot-spot at the location of the leak, because the flowing fluid generally has a higher temperature than the surrounding soil. Finally, if the temperature of the liquid and soil are equal, then a sensing cable can be provided with an additional electrical cable that can heat the entire sensor profile along the pipeline length. Leakage can be detected as an anomaly in heating/cooling of the
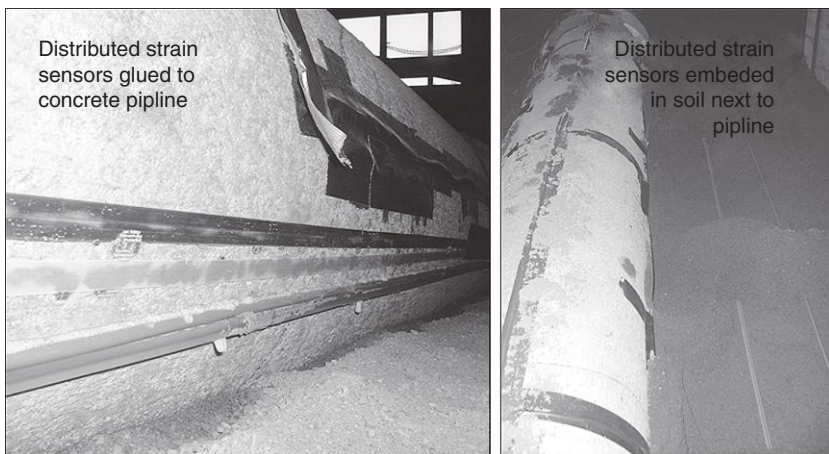
14.10 Various types of distributed strain sensors glued to concrete pipeline and embedded in soil (Glisic and Yao, 2012).
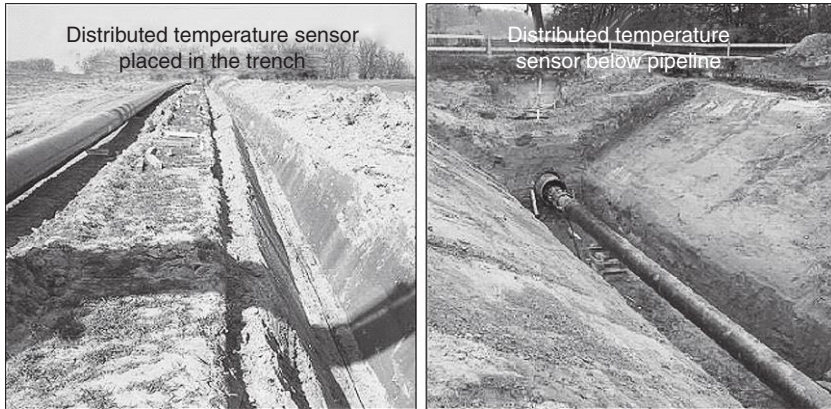
14.11 Distributed temperature sensor for leakage monitoring placed below the pipeline. (*Source*: Adapted from (Nikles *et al.*, 2004), courtesy of Omnisens.)

cable at the location of the leak. Cold and hot spots can be identified using temperature sensors based on Raman or Brillouin scattering. An example of a pipeline equipped with a leakage detection distributed sensor is given in Fig. 14.11.

Raman scattering systems have the advantage that they are cheaper than Brillouin scattering systems and are not cross-sensitive to strain. The advantage of Brillouin scattering systems is that they can cover an order of magnitude longer length of pipeline (hundred-kilometric); however, a challenge is to guarantee the strain-free state of the optical fibers over such long lengths.

Acoustic waves in an optical fiber can be captured and located by analyzing back scattered light. That is why the scattering-based technologies have the potential to be used for acoustic emission sensing and enable detection of third-party interference and leakage. The advantage is that the same sensors used for strain and temperature monitoring can be used for acoustic emission monitoring. For acoustic emission monitoring, it is only necessary to exchange the reading unit.

The greatest advantage of distributed fiber optic sensors is that they are continuous, and consequently all cross-sections of the pipeline are effectively instrumented (see Fig. 14.9). This is very important, taking into consideration the particularly long lengths of pipelines and the uncertainty of the location in which damage can occur. The main challenge for distributed fiber optic sensors is their packaging, which should not generate important losses, and yet it should guarantee safe and durable installation along long lengths of pipeline. Examples of pioneering applications are given in Figs 14.1 and 14.11, and more details about these applications can be found in literature

Measurements of sensors over all length of pipeline
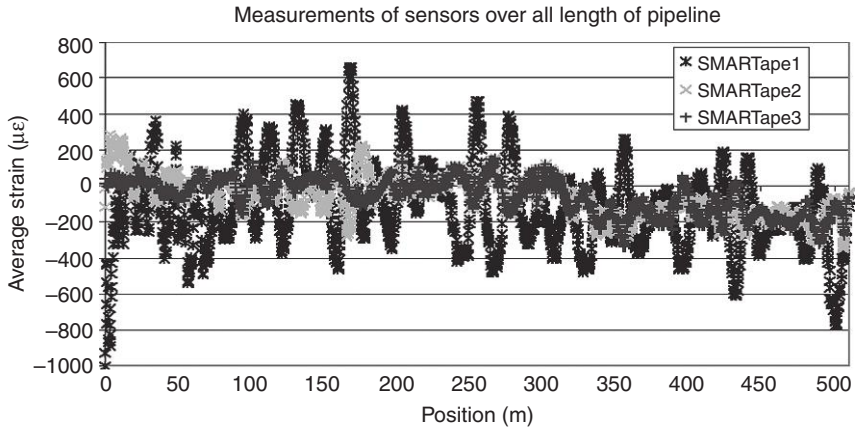
14.12 Strain in pipeline from Fig. 14.1 generated by weight of the soil after the burying of the pipeline, measured by a set of three distributed fiber optic sensors (Glisic and Inaudi, 2007). (*Source*: Courtesy of SMARTEC.)
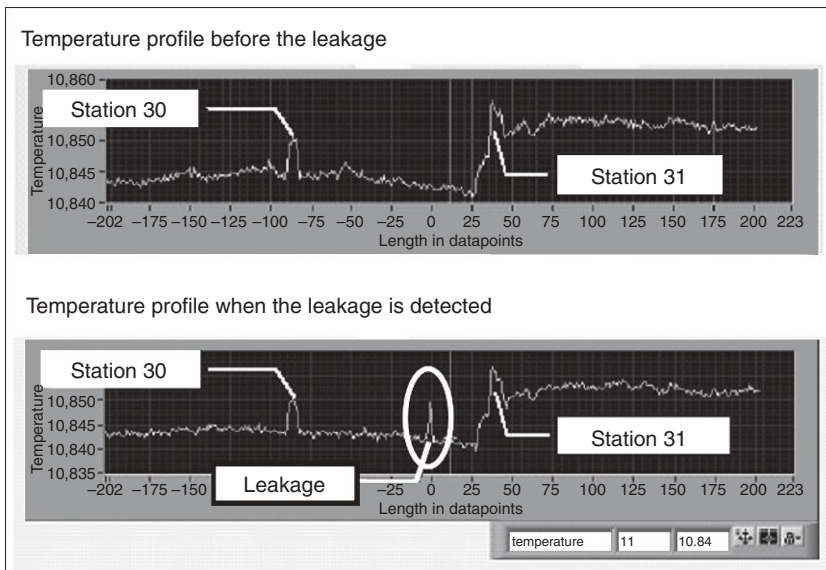
14.13 Leakage detection and localization for pipeline from Fig. 14.11. (*Source*: Adapted from Nikles *et al.*, 2004), courtesy of Omnisens.)

(Glisic and Inaudi, 2007; Nikles *et al.*, 2004). The monitoring results from these two projects are given in Figs 14.12 and 14.13 respectively. Figure 14.12 shows strain distribution along the pipeline from Fig. 14.1, due to burying (Glisic and Inaudi, 2007). The position of the sensors on the pipeline is as

shown in Fig. 14.9. Figure 14.13 shows on-site leakage detection on pipeline from Fig. 14.11 (Nikles *et al.*, 2004).

14.5.3   Large area electronics (LAE)

Today's sensing technology gives pipeline managers access to sparsely spaced sensors. Unfortunately, sparsely spaced sensors do not allow for the reliable early detection of anomalies such as strain concentrations, cracks, localized corrosion, etc. at locations even modestly distant from the sensors. This form of indirect damage detection often requires complex algorithms whose accuracy is challenged by the structural behavior of the pipeline, as also by noise sources from the environment, such as variable loading, temperature, etc. Distributed fiber optic sensors, presented in the previous section, can overcome some of the reliability issues as they have excellent spatial resolution and allow each section of the pipeline to be monitored. Thus, they provide for a means of performing one-dimensional (1D) monitoring. However, defects within the cross-section but far from the sensor will be difficult to detect at an early stage. A technology that enables two-dimensional (2D) monitoring promises to completely overcome the reliability issues related to damage detection.

LAE is an emerging technology that allows a broad range of electronic devices to be integrated on low-cost plastic sheets (Arias *et al.*, 2010; Someya *et al.*, 2008), see Fig. 14.14.

Through the use of micro-fabrication techniques, thin-film transducers (including pressure sensors, temperature sensors, particle sensors, etc.) have been demonstrated (Someya *et al.*, 2004; Graz *et al.*, 2009), and these may be formed into dense arrays spanning relatively large areas (i.e., tens of square meters). This in itself has important implications for structural
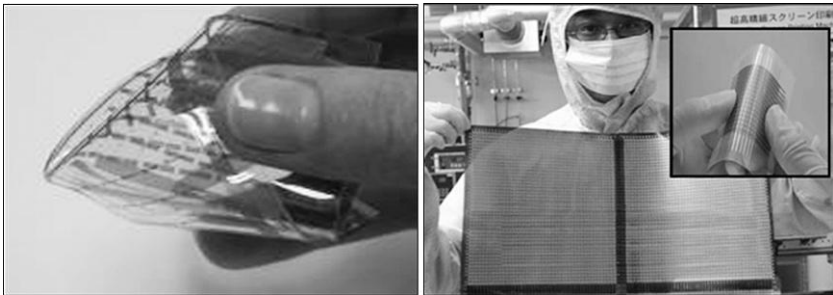
14.14 Examples of large area electronics. (*Source*: Courtesy of Prof. Sigurd Wagner, Princeton University, Princeton, NJ, USA) and large area electronics pressure sensor array, adapted from (Someya *et al.*, 2004), courtesy of Prof. Takao Someya, University of Tokyo, Japan.)

health monitoring. For instance, thin-film resistive strain gauges on polyimide are commercially available, and they are presently one of the most widespread modalities for monitoring civil structures. Such devices are compatible with the substrates and processing used for LAE, and thus thousands of strain gauges can be fabricated onto a single sheet that is both low in cost and highly conformal (Glisic and Verma, 2011). Hence, multi-functional sensing sheets for monitoring strain, temperature, corrosion, pressure, etc. can be created by proper integration techniques. Another important benefit of LAE is that it enables the integration of functional thin-film transistors (TFT). This means that basic circuit functionality is available to facilitate readout from the large number of sensor channels present on a LAE sheet.

The concept of wireless sensor networks is extremely valuable for large-scale monitoring, since it allows monitoring and interrogation over distributed point sensors. Wireless sensor nodes combine sensing, processing, communication, and possibly energy harvesting functionality into a miniature form factor (Chen *et al.*, 2010), that can be embedded in LAE (Glisic and Verma, 2011). This permits the possibility of a large number of nodes for high-spatial-resolution sensing and large coverage area. Recent efforts have focused on minimizing the power consumption of wireless devices so that they are compatible with self-powered operation (through energy harvesting) and physical miniaturization to mitigate the obtrusiveness of the nodes. For instance, piezoelectric and thermoelectric harvesters have been demonstrated along with thin-film batteries and energy-storage super capacitors. Thus, LAE makes self-powered operation possible.

The image to the right in Fig. 14.14 shows an example of an LAE pressure sensing sheet (Someya *et al.*, 2004) that illustrates several important characteristics: (1) a large number of sensors can be integrated with high density; (2) various parameters can be monitored with the same sensing sheet, such as strain, temperature, pressure, corrosion etc.; (3) the cost per sensor can be very low, thanks to inexpensive substrates and fabrication methods (Arias *et al.*, 2010); (4) reading, data analysis, and wireless data transmission devices can be embedded in the sheet, thus distributing the computing over the pipeline; and (5) energy harvesting and storage devices can also be embedded in the sensing sheet. Finally, due to their low thickness and flexibility, the LAE sensors are non-intrusive and, depending on their purpose, they can be installed on the interior or exterior surface of the pipeline, or both. The main challenges of LAE sensors are power supply and their durable installation at characteristic locations along the pipelines.

A summary of current and emerging sensing solutions presented in this chapter is given in Table 14.4.

Table 14.4 Summary of current and emerging sensing solutions

Sensing technique	Measured parameters	Continuity of monitoring	Advantages	Disadvantages
In-line flow metering (for steel pipelines)[a]	Leak detection	Continuous	Simple use	Post-event detection Inaccurate leak localization
In-line pressure metering (for steel pipelines)[a]	Leak detection	Continuous	Simple use	Post-event detection Inaccurate leak localization
Acoustic emission (for steel pipelines and PCCP pipelines)[a]	Leak detection Strands breakage in PCCP	Continuous	Simple use	Large amount of sensors needed to cover entire pipeline Post-event detection for steel pipeline
Magnetic flux leakage – MFL (on a 'pig', for steel pipelines)[a]	Defects in wall, cracks, pits, dents, etc.	Periodic	Accurate assessment of pipeline condition	Only periodic inspections are possible
Remote field eddy current – RFEC (on a 'pig', for steel pipelines)[a]	Defects in wall, cracks, pits, dents, etc.	Periodic	Internal and external defects	Only periodic inspections are possible
Ultrasonic transducers (on a 'pig', for steel pipelines)[a]	Thickness of pipe wall Defects in wall, cracks, pits, dents, ovalization, etc.	Periodic	Accurate assessment of pipeline condition	Only periodic inspections are possible
Gauging tools with camera (on a 'pig', for steel pipelines)[a]	Internal geometry of the pipeline Defects in wall, cracks, pits, dents, ovalization, etc.	Periodic	Accurate assessment of pipeline condition	Cameras require stopping of transmission in oil pipelines Only periodic inspections are possible
Acoustic and ultrasonic transducers (on a wall, for steel pipelines)[a]	Leakage	Continuous	Accurate detection and localization	Large number of sensors Important cabling needed

Closed-circuit television – CCTV (at manhole, for concrete pipelines)[a]	Internal state of pipeline	Continuous	Accurate assessment	Limited area of observation Small damage not visible
Closed-circuit television – CCTV (on a robot, for concrete pipelines)[a]	Internal state of pipeline	Periodic	Accurate assessment	Used only if not more than 25–30% of pipeline is filled Obstacle in pipeline can stop the robot Small damage not visible
Sewer scanner and evaluation technology – SSET (for concrete pipelines)[a]	Internal state of pipeline	Periodic	Improved reliability of damage detection 3D imaging	Used only if not more than 25–30% of pipeline is filled Obstacle in pipeline can stop the robot
Focused electrode leak location – FELL (for concrete pipelines)[a]	Leakage Infiltration	Periodic	Accurate detection, localization and quantification	Post-event detection Obstacle in pipeline can stop the robot Applicable only during the dry weather
Ground penetrating radar – GPR (for concrete pipelines)[a]	Bedding of pipeline, detection of scour, rocks, voids etc.	Periodic	Assessment of bedding and evaluation of associated hazards	Performance depends on soil properties Labor intensive
Infrared thermography system – ITS (for steel and concrete pipelines)[a]	Leakage	Periodic	Accurate localization	Complex data analysis
REFC/Transforming coupling (for CPP pipelines)[a]	Detection, localization and quantification of broken prestressing strands	Periodic	Senses beyond inner surface	Manned, non-automatic operation

(Continued)

Table 14.4 Continued

Sensing technique	Measured parameters	Continuity of monitoring	Advantages	Disadvantages
Wireless technologies (for steel and concrete structures)[b]	Means of reading sensors, computation, and communication	Continuous	Greatly improved coverage of monitoring, data analysis and data management	Power requirement
Distributed fiber optic sensors (for steel and concrete structures)[b]	Strain, temperature, deformation, cracks, leakage	Continuous	High accuracy in measurements, damage detection and localization High spatial coverage	Installation is labor intensive
LAE (for steel and concrete structures)[b]	Multi-parameter monitoring, strain, deformation, pressure, temperature, corrosion etc.	Continuous	Large area coverage Advantages of wireless technologies	Power requirement Installation is labor intensive

[a]Current technologies.
[b]Emerging technologies.

14.6   Future trends

Pipelines require reliable yet affordable sensing solutions, but their large size is a major challenge for both sensor installation and data management. Emerging sensing solutions presented in the previous sections are an adequate response to these challenges; however, the implementation of these technologies is challenging as well. Thus, novel solutions based on innovative or engineered materials could provide complementary solutions. Besides LAE sensing solutions, sensing skins based on carbon nanotubes are being developed (Loh *et al*., 2009) for 2D strain field monitoring. Sensing skins are read using electrical impedance tomographical conductivity mapping techniques, and provide a 2D mapping of damage (Loh *et al*., 2009). Besides strain field assessment, sensing skins can be made multi-functional and detect other defects such as corrosion (Loh *et al*., 2007), and interfaced using wireless nodes (Pyo *et al*., 2011). Another promising approach is the use of multi-functional materials with self-sensing properties of materials. Recent discoveries indicate that the electrical properties of cementitious materials can be used to detect and locate defects such as crack (Chung, 2003; Lynch and Hou, 2005). Successful developments and large-scale implementation could practically transform an entire pipeline into a sensor.

For existing pipelines located in difficult-to-reach locations or buried below the surface, the installation of sensors is challenging and remote sensing technologies is needed for nondestructive evaluation (NDE) of pipeline conditions. The implementation of aerial and satellite-based instrumentation such as the global positioning system (GPS) and satellite-based high resolution systems, as well as light detection and ranging (LIDAR), and to certain extent infrared tomography (e.g., Reed *et al*., 2004), can address these challenges.

For rehabilitation purposes, *in situ* pipe lining technologies have been developed over the last few decades (e.g., AWWA, 2001; Kramer *et al*., 1992). Their aim is to increase the service life of existing pipelines by remote insertion of polymeric linings using inexpensive trenchless construction procedures (e.g., Kramer *et al*., 1992; Davis *et al*., 2007). Rehabilitation using *in situ* pipe lining technologies opens new doors for the application of monitoring systems, as the sensors could be embedded between the lining and the inner wall of the pipeline during the installation of the lining or even integrated in the lining. The benefits are multiple: (i) the monitoring system can help assess the performance of rehabilitated pipelines under hazardous event (e.g., earthquake, etc.) (ii) it will provide reliable real-time information about the pipeline structural health condition after a hazardous event; and (iii) it will detect and localize degradation or defects that may occur after the rehabilitation, pointing to weak areas that need additional attention.
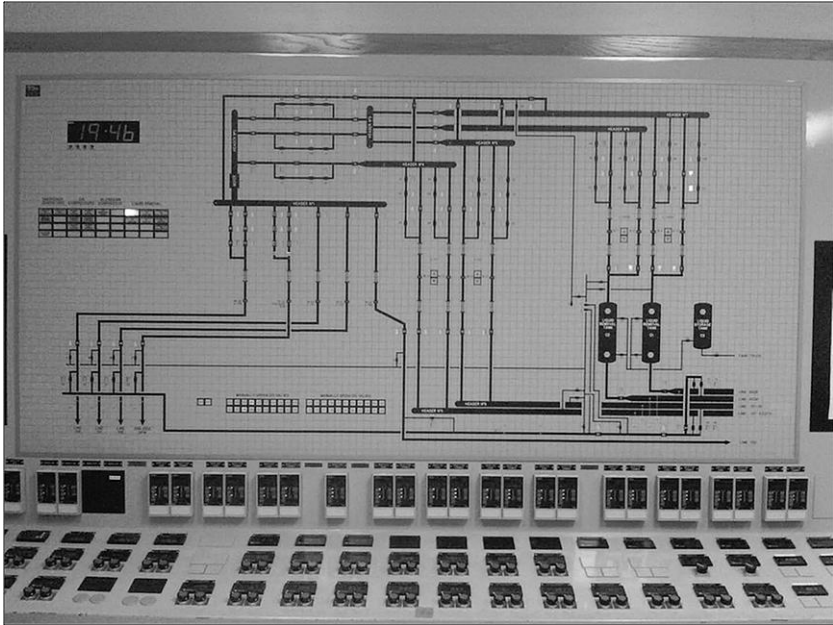
14.15 Examples of a control panel in a SCADA center (NTSB, 2011). (*Source*: Courtesy of The National Transportation Safety Board.)

Rapid developments in the fields of telecommunications, informatics technologies, and computer sciences over the last quarter of the twentieth century have led to the development and implementation of various supervisory control and data acquisition (SCADA) systems for continuous operational monitoring of pipelines (e.g., NTSB, 2011; Reed *et al.*, 2004). However, SCADA mainly encompasses the operation of pipeline processes (flow, pressure, etc.) and, although it can be used to detect some types of damage (e.g., leakage from the flow rate) it is not widely used for structural health monitoring. This, however, indicates one of the future trends – integration of structural health monitoring in existing SCADA systems. Integration can further be leveraged by including geographical information system (GIS) to create a comprehensive system for data visualization, analysis, and damage detection (Reed *et al.*, 2004). Finally, decision-making tools fed by the data from the monitoring system could be developed and integrated as well. An example of a control panel in a SCADA center is given in Fig. 14.15.

To conclude, future trends in sensing solutions for pipeline systems include the development of new sensing technologies, implementation of remote monitoring solutions, leverage of repair and rehabilitation methods to implement monitoring systems, integration of traditional and emerging technologies with existing operational monitoring systems, and development of

decision-making tools. Progress in these areas has the potential to transform the way pipeline health is assessed: implementation of continuous on-line monitoring promises to significantly improve safety and optimize maintenance, and also to significantly reduce the volume of periodical inspections and decrease associated costs.

14.7   Sources of further information and advice

More information concerning oil and gas pipelines can be found in John L. Kennedy's classic *Oil and Gas Pipeline Fundamentals*, published by PennWell Publishing Company, Tulsa, Oklahoma, p. 366 (1993). This book gives an excellent introduction to pipeline manufacturing, design, machinery, construction practices, operation, control, inspection, and regulations. Although the part on inspection is slightly outdated, it presents the main issues found in practice and gives a brief overview of traditional techniques. The book is focused on steel pipelines, as these are mainly used in the oil and gas industry.

The book by Mohammad Najafi *Trenchless Technology: Pipeline and Utility Design, Construction, and Renewal*, McGraw-Hill, New York City, p. 489, (2004) gives more information on concrete pipelines, although it also treats metallic pipelines. It is mostly oriented toward water supply pipelines and sewers, and includes trenchless installation of the pipelines; thus it is a very good complement to Kennedy's book. It presents and explains the causes of pipeline deterioration and the most important traditional inspection techniques.

An overview of monitoring solutions for pipeline systems is given in *Techniques for monitoring structural behaviour of pipeline systems* by Christopher Reed, Alastair J. Robinson, and David Smart, sponsored and published by AWWA Research Foundation and American Water Works Association, Denver CO, p. 247 (2004). Besides monitoring techniques, the publication presents surveying techniques and remote monitoring tools such as GPS, satellite-based high resolution systems, LIDAR, and infrared tomography. Integration of monitoring technologies and decision support methodologies is also presented.

More information about emerging technologies can be found in papers published in specialized journals and conferences. A non-exhaustive list includes *Structural Health Monitoring*, *Journal of Civil Structural Health Monitoring*, *Smart Materials and Structures*, *Structural Control and Health Monitoring*, and *Measurement Science and Technology*; recommended conferences are International Workshop of Structural Health Monitoring (IWSHM), International Conference on Structural Health Monitoring of Intelligent Infrastructure, and SPIE Smart Structure/ NDE. These journals and conferences deal mainly with emerging sensing

technologies and their application to various fields of engineering. Topics more specific to pipelines can be found in other journals (e.g. *Journal of Pressure Vessel Technology*), and proceedings of various conferences on pipelines (e.g. Rio Pipeline Conference and Exposition, International Pipeline Conference (IPC), Pipeline Technology Conference (PTC), ASCE Pipeline Conference, etc.).

Important information regarding pipelines is found in the websites of numerous organizations and agencies, some of which are listed below:

API: American Petroleum Institute, http://www.api.org
ASCE: American Society of Civil Engineers, http://www.asce.org
ASTM: American Society of Testing Materials, http://www.astm.com
AWWA: American Water Works Association, http://www.awwa.org
DOE: US Department of Energy, http://energy.gov
EPA: Environmental Protection Agency, http://www.epa.gov
EPRG: European Pipeline Research Group, http://www.eprg.net
ISHMII: International Society for Structural Health Monitoring of Intelligent Infrastructure, www.ishmii.org
PHMSA: U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration, http://www.phmsa.dot.gov/
PIPE: Professional Institute of Pipeline Engineers, http://www.pipeinst.org
PRC: Pipeline Research Council, http://www.prci.com

14.8   Acknowledgment

14.9   References

ACPPA (2011), http://www.acppa.org/WhatisConcretePressurePipe.pdf, official site of American Concrete Pressure Pipe Association (Accessed on 27 September 2011).

API (2011), http://www.api.org/Standards/, official website of American Petroleum Institute (Accessed on 26 September 2011).

Arias A C, MacKenzie J D, McCulloch I, Rivnay J and Salleo A (2010), 'Materials and applications for large area electronics: Solution-based approaches', *Chemical Reviews*, **110**(1), 3–24.

Atherton D (2011), http://www.physics.queensu.ca/~amg (Accessed on 21 December 2011).

AWWA (2001), 'Rehabilitation of water mains', *Manual M28,* 2nd Edn.*,* Denver, CO, USA, American Water Works Association.

CBO (2002), 'Future investment in drinking water and wastewater infrastructure', United States Congress, Congressional Budget Office, Washington, DC, November 2002.

Celebi M (2002), 'Seismic Instrumentation of Buildings (with Emphasis on Federal Buildings)', *United States Geological Survey (USGS) Report No. 0-7460-68170*, Menlo Park, CA.

Chen G, Fojtik M, Kim D, Fick D, Park J, Seok M, Chen M-T, Foo Z, Sylvester D and Blaauw D (2010), 'A millimeter-scale nearly-perpetual sensor system with stacked battery and solar cells', *Proc. Int. Solid-State Circuits Conference*, San Francisco, California, USA, 288–289.

Chung D D L (2003), *Multifunctional Cement-Based Materials*, New York, NY, USA, Marcel-Dekker.

Ciocco M, Neyhart J, Mandayam S, Jahan K and Clearly D B (2002), 'Ultrasonic imaging of defects in concrete pipelines', *Conference on the Review of Quantitative Nondestructive Evaluation*, American Institute of Physics, **21**, 254–1260.

Cosham A and Hopkins P (2004), 'An overview of the pipeline defect assessment manual (PDAM)', *4th International Pipeline Technology Conference*, 9–13 May 2004, Oostende, Belgium, pp. on conference CD.

Davies J P, Clarke B A, Whiter J T and Cunningham, R J (2001), 'Factors influencing the structural deterioration and collapse of rigid sewer pipes', *Urban Water*, **3**, 73–89.

Davis C A, Hu J, O'Rourke T D and Bonneau A (2007), 'Seismic Performance Evaluation of LADWP Water System Using GIRAFFE. Proceedings', *5th U.S.-Japan Workshop on Seismic Performance of Water Supplies*, Oakland, CA, American Water Works Association Research Foundation and Japan Water Works Association, 13 p.

Duran O, Althoefer K and Seneviratne L D (2002), 'State of the art in sensor technologies for sewer inspection', *IEEE Sensors Journal*, **2**(2), 73–81.

El-Sayed A H (2011), 'Course: Oil and Gas Pipeline Design, Maintenance and Repair, Part 9', Mining, Petroleum & Metallurgical Engineering Department, Cairo University, Egypt, http://www.eng.cu.edu.eg/users/aelsayed/courses.htm (Accessed on 3 October 2011).

EPA (2000), 'Wastewater Technology Fact Sheet: Sewers, Force Main', *EPA 832-F-00–071*, United States Environmental Protection Agency, Office of Water, Washington, DC.

Glisic B and Inaudi D (2007), *Fibre Optic Methods for Structural Health Monitoring*, Chichester, UK, John Wiley & Sons, Inc.

Glisic B and Verma N (2011), 'Very dense arrays of sensors for SHM based on large area electronics', Structural Health Monitoring 2011: Condition-Based Maintenance and Intelligent Structures – Proceedings of the 8th International Workshop on Structural Health Monitoring, **2**, 1409–1416.

Glisic B and Yao Y (2012), 'Fiber optic method for health assessment of pipelines subjected to earthquake-induced ground movement', *Structural Health Monitoring*, **11**(6), 696–711.

Graz I, Krause M, Bauer-Gogonea S, Bauer S, Lacour S P, Ploss B, Zirkl M, Stadlober, B and Wagner S (2009), 'Flexible active-matrix cells with selectively poled

bifunctional polymer ceramic nanocomposite for pressure and temperature sensing skin', *Journal of Applied Physics,* **106**, 034503-1.

Grigg N S (2006), 'Condition assessment of water distribution pipes', *Journal of Infrastructure Systems*, **12**(3), 147–153.

Inaudi D, Glisic B, Figini A and Walder R (2007), 'Pipeline leakage detection and localization using distributed fiber optic sensing', *Proc. of Rio Pipeline Conference 2007*, Rio de Janeiro, Brazil.

Inaudi D and Glisic B (2010). 'Long-range pipeline monitoring by distributed fiber optic sensors,' *ASME Journal of Pressure Vessel Technology*, **132**(1), 011701–01–011701–09.

Kennedy J L (1993), *Oil and Gas Pipeline Fundamentals*, Tulsa, Oklahoma, PennWell Publishing Company.

Kikuchi K, Naito, T and Okoshi T (1988), 'Measurement of Raman scattering in single-mode optical fiber by optical time-domain reflectometry', *IEEE Journal of Quantum Electronics*, **24**(10), 1973–1975.

Kim J, O'Connor S, Nadukuru S, Pour-Ghaz M, Lynch J P, Michalowski R L, Green R A, Bradshaw A and Weiss W J (2010), 'Response of a buried concrete pipeline to ground rupture: A full-scale experiment and simulation', *Proceedings of SPIE Conference: Smart Structures/NDE 2010*, San Diego, CA, 7–11 March 2010.

Kobayashi M, Minato H, Kondo M, Murashita K and Kurashima M (1999), 'NKK ultrasonic pipeline inspection pig', *NKK Technical Reviews*, **80**, 46–50.

Kramer S R, McDonald W J and Thomson J C (1992), *An Introduction to Trenchless Technology*, New York, NY, USA, Van Nostrand Reinhold.

Kurashima T, Horiguchi T and Tateda M (1990), 'Distributed temperature sensing using stimulated Brillouin scattering in optical silica fibers', *Optics Letters*, **15**(18), 1038–1040.

Liu H (2003), *Pipeline Engineering*, Boca Raton, FL, USA, Lewis Publishers.

Loh K J, Kim J, Lynch J P, Kam N W S and Kotov N A (2007), 'Multifunctional layer-by-layer carbon nanotube–polyelectrolyte thin films for strain and corrosion sensing', *Smart Materials and Structures,* **16**, 429–438.

Loh K J, Hou T-C, Lynch J P and Kotov N A (2009), 'Carbon nanotube sensing skins for spatial strain and impact damage identification', *Journal of Nondestructive Evaluation,* **28**, 9–25.

Lynch J P (2002), 'Decentralization of Wireless Monitoring and Control Technologies for Smart Civil Structures', *Ph.D. Thesis (John A. Blume Earthquake Engineering Center, Technical Report #140)*, *Department of Civil and Environmental Engineering, Stanford University, Palo Alto, CA, USA*.

Lynch J P and Hou T-C (2005), 'Conductivity-based strain and damage monitoring of cementitious structural components', *Proceedings of SPIE*, **5765**, 419–429.

Lynch J P, Wang, Y, Loh, K, Yi, J H and Yun C B (2006), 'Performance monitoring of the Geumdang Bridge using a dense network of high-resolution wireless sensors', *Smart Materials and Structures*, **15**(6), 1561–1575.

Najafi M (2004), *Trenchless Technology: Pipeline and Utility Design, Construction, and Renewal*, New York City, USA, McGraw-Hill.

Nikles M, Vogel B H, Briffod F, Grosswig S, Sauser F, Luebbecke S, Bals A and Pfeiffer T (2004), 'Leakage detection using fiber optics distributed temperature

monitoring', *11th SPIE Annual International Symposium on Smart Structures and Materials,* San Diego, CA, USA, 5384, 18–25.

NTSB (2011), 'Pacific Gas and Electric Company Natural Gas Transmission Pipeline Rupture and Fire San Bruno, California, 9 September 2010', *Accident Report NTSB/PAR-11/01, PB2011-916501, Notation 8275C*, Adapted 30 August 2011, National Transportation Safety Board (NTSB), http://www.ntsb.gov/investigations/2010/sanbruno_ca.html (Accessed on 3 October 2011).

O'Day D K., Weiss R, Chiavari S and Blair D. (1986), 'Water Main Evaluation for Rehabilitation/Replacement', Denver, CO, USA, AWWA Research Foundation (90509).

PHMSA (2011), http://www.phmsa.dot.gov/pipeline/library/data-stats, official website of U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration, updated on 31 August 2011 (Accessed on 20 September 2011).

Posey R Jr., Johnson G A and Vohra S T (2000), 'Strain sensing based on coherent Rayleigh scattering in an optical fibre', *Electronics Letters*, **36**(20), 1688–1689.

Prosser D (1996), 'Research, Product Improvement, New AWWA Standards, Compliance Certification & Performance – Contribute to the Increased Usage of Prestressed Concrete Cylinder Pipe', *AWWA ACE Conference Proceedings*.

Pure Technologies (2011), http://www.puretechltd.com/types_of_pipe/pccp.shtml (Accessed on 27 September 2011).

Pyo S, Loh K J, Hou T-C, Jarva E and Lynch J P (2011), 'A wireless impedance analyzer for automated tomographic mapping of a nanoengineered sensing skin', *Smart Structures and Systems*, **8**(1), 139–155.

Reed C, Robinson A J and Smart D (2004), *Techniques for Monitoring Structural Behaviour of Pipeline Systems*, Denver, CO, USA, AWWA Research Foundation and American Water Works Association.

Rose J L (1999), *Ultrasonic Waves in Solid Media*, New York, NY, USA, Cambridge University Press.

Semanuik S and Mergelas B (2006), 'Comparison of Identified Distress in CCP Pipelines Operated by Water Utilities in North America', *ASCE Pipeline Division Specialty Conference (Pipelines 2006): Service to the Owner, 30 July–2 August, Chicago, Illinois, USA*, doi:10.1061/40854(211)21.

Sinha S K and Fieguth P W (2006), 'Segmentation of buried concrete pipe images', *Automation in Construction*, **15**(1), 47–57.

Someya T, Sekitani T, Iba S, Kato Y, Kawaguchi H and Sakurai T (2004), 'A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications', *Proceedings of the National Academy of Science,* **101**(27), 9966–9970.

Someya T, Pal B, Huang J and Katz H E (2008), 'Organic semiconductor devices with enhanced field and environmental responses for novel applications', *MRS Bulletin*, **33**, 690–696.

Spencer B F, Ruiz-Sandoval, M E and Kurata N (2004), 'Smart sensing technology: opportunities and challenges', *Journal of Structural Control and Health Monitoring*, **11**(4), 349–368.

Straser E G and Kiremidjian A S (1998), 'A modular, wireless damage monitoring system for structures', *John A. Blume Earthquake Engineering Center Report No. 128*, Stanford, CA.

Towfighi S, Kundu T and Ehsani M (2002), 'Elastic wave propagation in circumferential direction in anisotropic cylindrical curved plates', *Journal of Applied Mechanics*, **69**(3), 283–291.

Weil G J (1998), 'Infrared thermographic pipeline leak detection systems for pipeline rehabilitation programs', *Proceedings of SPIE*, **3398**, 54–65.

Wirahadikusumah R, Abraham D M, Iseley T and Prasanth R K (1998), 'Assessment technologies for sewer system rehabilitation', *Automation in Construction*, **7**(4), 259–270.

15

Sensing solutions for assessing and monitoring roads

M. L. WANG and R. BIRKEN, Northeastern University, USA

Abstract: The United States faces a monumental infrastructure management problem in the scheduling and implementation of maintenance and repair operations, and in the prioritization of expenditures within budgetary constraints. The efficient and effective performance of these operations is crucial to ensuring roadway safety, preventing catastrophic failures and promoting economic growth. Several viable inspection techniques are discussed focusing on current methods as well as futuristic sensing methods that will speed up the assessment process without closing traffic, and are likely tools for network-wide roadway monitoring systems. Economic benefits cannot be realized until an accurate and timely updated monitoring and assessment system and network is achieved.

Key words: roadway inspection, bridge deck inspection, asset management, nondestructive evaluation (NDE) and testing, sensor technology.

15.1   Introduction

Civil infrastructure construction and maintenance represent a large societal investment. Despite being the lifeline of commerce, civil infrastructure is just at the beginning of benefiting from the latest advances in sensor technologies. According to the latest ASCE report card (ASCE, 2013) the US infrastructure scores only a D+, and it is estimated that a $3.6 trillion investment is needed by 2020.

There are 4 million miles of roads in the United States (FHWA, 2008) requiring a broad range of maintenance activities. The ASCE report card (ASCE, 2013) gives roads only a grade of D, commenting that 'Forty-two percent of America's major urban highways remain congested, and costing the economy an estimated $101 billion in wasted time and fuel annually. While the conditions have improved in the near term, and Federal, state, and local capital investments increased to $91 billion annually, that level of investment is insufficient and is still projected to result in a decline in conditions and performance in the long term. Currently, the Federal Highway

461

Administration (EHWA) estimates that $170 billion in capital investment would be needed on an annual basis to significantly improve conditions and performance.'

The nation faces a monumental problem of infrastructure management in scheduling and implementing maintenance and repair operations, and in prioritizing expenditures within budgetary constraints. The efficient and effective performance of these operations is crucial to ensuring roadway safety, preventing catastrophic failures, and promoting economic growth. Roadway work zones used for assessment and repair are a major source of traffic congestion, which results in lost productivity and wasted fuel. It is a critical need to make the right roadway repairs in the right place and at the right time.

Current inspection strategies characterizing roadway conditions are not well suited to fulfill this need, because they typically only inspect small localized areas, only periodically (order of years if at all), and mainly through visual inspection. Therefore, there is a critical need for technology that can cost-effectively monitor the condition of a network-wide road system and provide accurate, up-to-date information for maintenance activity prioritization. A framework to shift from periodical localized inspections to continuous network-wide health monitoring is required and is evolving.

15.1.1   Roadway and bridge deck defects

Roadway deterioration frequently takes place below the surface and cannot be evaluated by visual means (Fig. 15.1).
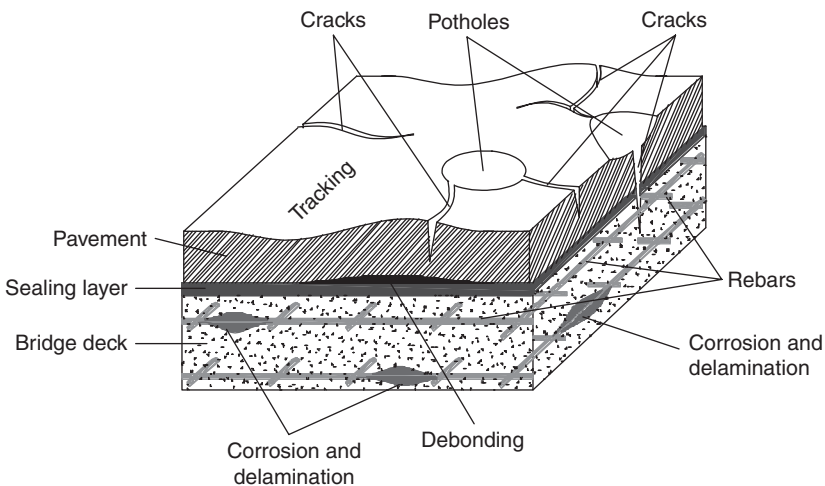
15.1 Some common defects and deteriorations found in concrete bridge decks with asphalt pavement overlay.

   Common types of roadway damage are transverse cracks, longitudinal cracks, tracking, corrugation, potholes, delamination, and seepage. Transverse cracks occur more often than longitudinal cracks and can start with a fine crack of less than 0.5 mm in width and of less than 2 cm in depth. Such cracks are hardly visible when it is sunny, but are visible after rain due to vaporization of the surface water that leaves water in the cracks. Small cracks need to be treated to prevent them from developing into larger cracks. Large cracks often have widths of more than 1 mm, depths of 5 cm, and run for meters in length. If large cracks are not sealed, delamination and scaling will follow. If the adhesion between pavement and concrete deck decreases, the overlay may debond from the deck's top surface. The loss of adhesion may be caused by seepage from cracks or potholes. Local debonding may span only several square centimeters, and can be difficult to detect because the pavement surface remains intact. Large area delaminations may develop into large cracks at the pavement surface and eventually cause large potholes and loss of pavement. Cracks and potholes are often accompanied by seepage. Water enters into the overlay through cracks. The adhesion between asphalt and concrete deck is extremely vulnerable to water penetration. Water within cracks of a pavement will stay and seep. This is most harmful to asphalt pavement.

   The three most common mechanisms that can result in deterioration of concrete and bridge deck include freeze–thaw action, alkali silica reaction, and corrosion of the reinforcing steel. Freeze–thaw action forces frequent expansion and contraction of the bridge deck, resulting in extensive cracking. Alkali silica reaction is a chemical reaction that occurs between the cement and aggregates, inducing a buildup of pressure in the aggregates. This also results in cracking. Corrosion of the reinforcing steel, the most common mechanism for deterioration, occurs in reinforced concrete bridge decks that are exposed to deicing salts during the winter months. The consequence of corrosion is a decrease in the cross-sectional area of the reinforcing steel along with the buildup of rust, also known as hydrated ferric oxide. Rust takes four to five times more volume than the original products of the reinforcing steel. The results are a buildup of internal stresses, cracking, delaminations, and eventual spalling of the concrete. As a result of corrosion, a bridge deck can be replaced at least once, and rehabilitated many times throughout its life cycle. The key to ensuring proper repair or replacement is inspection.

   The objective of this chapter is to review several viable techniques to assess and determine the condition of roadway, including highway pavement and bridge deck, for an early repair. Road defects and distresses both on the surface and under the surface detected by various methods will be reviewed and discussed. Several commercial products and procedures will be listed and discussed wherever possible. The adaptation of a technique lies in its

accuracy, ease of use, and cost effectiveness. Certainly, a favorable choice is a nondestructive testing technique deployed without stopping traffic.

15.1.2   Pothole and other highway pavement distress problems

Potholes and other localized roadway distresses plague urban infrastructures. Pothole patching as shown Fig. 15.2 is the most common and cost-effective method of treatment, but varies widely in its effectiveness. The quality of the treatment is compromised by the need for quick and immediate repair, and the inability to achieve adequate compaction due to low quality materials and small patch areas. This leads to rapid deterioration of repairs, ultimately resulting in wasted time and money.

   One way to improve the life of roadways and repair methods is through the detection of potholes or localized distresses in the stages of early formation. Ideally, this would be achieved through feedback on the condition of an agency's highway pavements using maintenance vehicles equipped with detection devices, rather than relying on user feedback. Early monitoring and detection of both surface and subsurface distresses of highway pavement could be used to identify potholes and localized distresses in the early stages of formation.

15.1.3   ASTM pavement condition assessment

The original AASHO Road Test (National Research Council (U.S.), and American Association of State Highway Officials, 1961) is one of the most comprehensive full-scale accelerated pavement tests conducted in the United States. It was used to come up with empirical equations explaining the physical reasons behind the deterioration of roads. As part of this test the

(a)   (b)
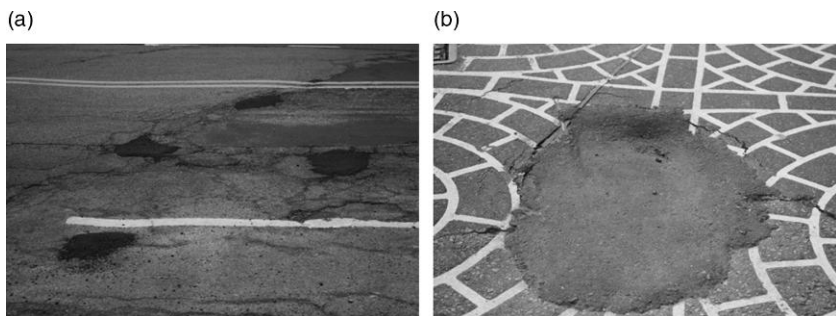
15.2 Pothole distress problem illustrated. (a) Repeated refill of potholes, and (b) the sinking of a repaired pothole.

Pavement Serviceability Index (PSI) has been developed. They calculated visible distresses on-road, such as cracking, patching, slope variance, and rut depth. Profilometer data are obtained to calculate the surface roughness and compared these values to ratings given by experts on a scale of 5 as an assessment of roads (Jain *et al.,* 1971).

$$PSI = 5.03 - 1.9\log\left(1 + SV_t\right) - 1.375RD_t^2 - 0.01\left(C_t + P_t\right)^{0.5} \qquad [15.1]$$

where: *PSI* – serviceability in terms of the present serviceability index, SV_t – slope variance at time T, C_t – crack length in feet per 1000 ft², P_t – patching in ft² per 1000 ft², RD_t – average rut depth in inches

While some disadvantages have been observed with the model, it still remains reliable for assessing pavement condition using actual distresses involved (Pedersen, 2007).

In the late 1970s an alternative approach, named *Paver*, was introduced (Shahin and Kohn, 1981). Later this was introduced as ASTM Standard D 6433 (ASTM, 2011).This standard suggests a pavement condition index (PCI) for assessing pavement. The standard says that this practice covers the determination of roads' and parking lots' pavement condition through visual surveys using the PCI method of quantifying the pavement condition. Our discussion here is focused on roadway pavement.

Before conducting a visual survey, the pavements are divided into pavement branches, which are further divided into pavement sections. While conducting a survey only certain parts of a given section are surveyed. They are called pavement sample units. In general, the area of a pavement sample unit is 2500 ft². The ASTM Standard D6433 recommends that an average of 33%–50% of a pavement section be assessed.

This ASTM standard has identified a total of 19 different types of distresses available at any given instant on a pavement. Each of these distresses is further sectioned based on distress density and severity, and a deduct value is calculated from the charts provided in the ASTM standard. For example, in Fig. 15.3 different curves are shown, one for each severity level of low, medium, and high. Based on the distress density on a particular sample unit, a deduct value is calculated for all the distress of the whole section. Once all the deduct values for all distresses are calculated, the ASTM standard outlines an algorithm for the PCI calculation (Shahin, 1994).

The PCI cannot measure structural distresses, nor does it provide direct measurement of skid resistance or roughness. It can be used only to determine the maintenance and repair needs and their priorities. It provides a feedback on pavement performance for validation or improvement of current pavement design and maintenance procedures (ASTM D5340, 2012).
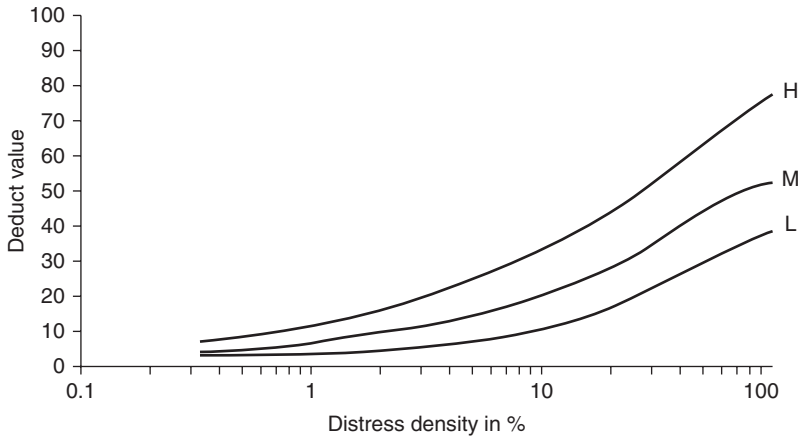
15.3 ASTM Standard 6433 curve for deduct values of longitudinal/
transverse cracks (H – high, M – medium, L – low) (ASTM D6433-11, 2011).

MicroPaver (2007) is a software that can input the PCI survey results and perform pavement condition calculations. In addition, it is equipped with a pavement management system, i.e., modules for pavement deterioration modeling, and cost-effective and budgetary pavement management decisions.

The ASTM standard lays out the right approach for assessing the pavement conditions, but it has multiple human interferences where the user judgement is given more value than strong physics-based logic.

The first opportunity for human interference is at the very beginning of the PCI process, when the decisions have to be made as to which sample units to select for representing a whole pavement section. Having realized that it would be a cumbersome, expensive, and monumental task to collect data manually for all complete sections, the ASTM standard suggests collecting only 33%–50% of it. This leaves a very high probability that the sampled data not reflect the exact condition of the whole pavement section. The second opportunity for interference is in the subjective assessment of the distress severity and density estimations. Any wrong input from the user could lead to erroneous results from survey. Finally, the third and most important part is the calculation of the PCI, and for that we use established deduct values. The deduct curves given in the ASTM standard have been established with expert advice but not based on any strong physics fundamentals.

When the PCI was adopted in ASTM Standard 6433, there was no concept of any automated inspection surveys, and image sensing and non-contact sensing devices were not so advanced (Ponniah *et al.,* 2001).Until recently, most automated surveys were only capable of taking high quality images of the pavement, and still relied on humans to assess the condition of pavement from the pictures interactively. Some companies offer the capability

of automatically identifying all the cracks, patch, and rutting to derive PCI values. Limitations of the PCI method are that it is time consuming, interrupts traffic, poses safety risks for the inspectors, requires trained experts, the results are low resolution (i.e. only a single PCI value per street from intersection to intersection), and the subsurface pavement conditions are not considered at all.

Some measurement systems are trying to be physical-based and measure acoustic or electromagnetic properties to derive the layers of subsurface profiles (Lu *et al.,* 2013; Daniels, 2004). The macrotexture depth (MTD) (Section 15.2.3) and the International Roughness Index (IRI)(2.4) are frequently used to assess the roughness of the surface and comfort at a driving speed.

15.2   Nondestructive evaluation (NDE) techniques for highway pavement assessment

With the advent of technology, the pavement management system has seen a shift from the manual data collection method and analysis to semi-automated data collection and analysis. Several states, and especially agencies, began to adopt the automation of data collection and analysis. It was widely believed that a manual method involved a lot of subjectivity (Smith *et al.,* 1998). Though automated technology was more costly, it was accepted, as the data collection process could be done without affecting traffic too much. The main technology used for automated data collection is optical cameras mounted on the back and the sides of a van, which will take images continuously and save those images so that they can be analyzed separately back at the office (Dondi *et al.,* 2011). There have been attempts to calculate the distresses automatically from the data, but there are many obstacles, such as shadow, sunlight, clarity, etc. of the images (Smith *et al.,* 1998). Considering the wide use of automated data collection, the ASTM standard had updated ASTM D 6433 (ASTM, 2011a) 'Standard Practice for Roads and Parking Lots PCI Surveys' (Shahin *et al.,* 2003). To this date most agencies use the automated data collection involving optical cameras for collecting digital images of a street. Once data are collected, the various distresses are mostly manually evaluated, and after this the agencies use MicroPaver, and other software such as Street Saver (Smith *et al.,* 2007) or CarteGraph (www. cartegraph.com), for calculating the PCI and further required pavement analysis.

15.2.1   Falling weight deflectometer (FWD)

An FWD is used as a testing device for measuring the physical properties of pavement. This is a nondestructive and non-intrusive process, and
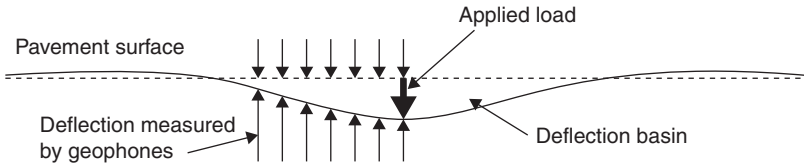
15.4 Schematic of FWD load and deflection measurement. (*Source*: Texas Department of Transportation FWD Technical Advisory.)

is generally preferred over other destructive testing, because tests are not only faster but also do not entail removal of pavement material (Belt *et al.,* 2006). The relevant ASTM standard (ASTM D4694, 2006b) states that the vehicle should be brought to a stop with the loading plate positioned over the desired test location. The FWD process applies dynamic loads to a pavement surface, simulating the magnitude and duration of a single heavy moving wheel load. The peak deflections at each measured location are recorded in micrometers, as shown in Fig. 15.4.

The process measures the vertical deflection response of the surface to an impulse load applied to the pavement surface (ASTM D4694, 2006b). The main components of an FWD system are a force-generating device, a guide system, a loading plate, a deflection sensor, a load cell, and a data processing and storage system (ASTM D4694, 2006b). Major factors affecting pavement deflection when using the FWD technology are the pavement layer thickness, layer material types, material quality, subgrade support, environmental factors, pavement discontinuities, and variability within the pavement structure. It is important to regularly maintain and service FWDs. This will improve equipment performance, longevity, and the quality of acquired data (Belt *et al.,* 2006).

First introduced in Europe, FWD has been in use in the United States since the 1980s. The device is used to measure surface deflection by applying an impulsive force to the pavement surface. As Fig. 15.5 presents, the FWD includes four main systems: (1). an impulsive-force generator that enables application of variable weights to the pavement; (2) a loading plate to spread the impulsive force uniformly through the tested layer surface; (3) sensors for deflection basin determination; and (4) a data acquisition and processing system (Elseifi *et al.,* 2012).

The FWD was once used to simulate the different wheel loading effects on the pavement, so the impact force was designed over a range using corresponding drop heights (Fig. 15.5a). The diameter of the circular loading plate is set at 0.3 m for commercial products (Elseifi *et al.,* 2012). After loading, the deflected surface basin would be measured by transducers. The transducers used in FWDs usually include geophones and seismometers for
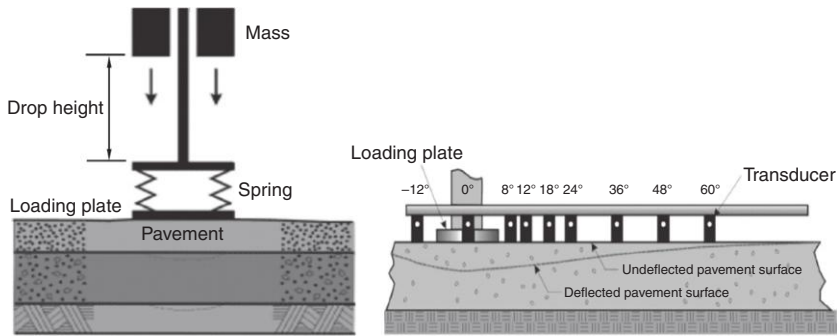
15.5 Diagram of the FWD testing.

different applications. For pavements with thin asphalt layers, the deflection sensors have to be located closer to the load center than in the case of a thicker asphalt pavement (Gurp, 2005).

Based on the FWD principles, a highly mobile modification was proposed in the late 1990s, the rolling wheel deflectometer (RWD). The objective of the RWD is to implement the FWD road test at normal highway speed without traffic interruption. The latest version of the RWD was introduced in 2003, and is shown in Fig. 15.6a (Elseifi *et al.*, 2012). It consists of a 16.2 m long semitrailer applying a standard 80 kN load on the pavement structure over the rear single axle by means of a regular dual-tire assembly. The RWD measures wheel deflections at the pavement surface by means of the spatially coincident method, which compares the profiles of the surface in the undeflected and deflected states. As the RWD travels on top of the pavement, triangulation lasers mounted on a 7.7 m aluminum beam and placed at 2.6-m intervals are used to measure surface deflections. The system has a 100 mm measurement deflection range and has an accuracy of ±0:0254 mm. A spatially coincident method is used to collect the deflected and undeflected states in running mode, as shown in Fig. 15.6b.

Also, a comparison between FWD and RWD has been summarized by Elseifi *et al.* (2012). During RWD testing, laser deflection readings are measured at 15 mm intervals. Irrelevant data, such as measurements collected on top of a bridge, sharp horizontal and vertical curves, and at traffic signals, were removed. Erroneous data may also be obtained if the pavement surface is wet, or in areas with severe cracking at the pavement surface. Valid deflection measurements are then averaged to reduce the effects of truck bouncing and vibrations on the measured deflections.

The speed effects were examined for running tests. It was found that the influence of the testing speed on the measured deflection was minimal. A statistical study on four different speeds had shown no apparent bias involved.

(a)

16.2 m

Laser sensors

Beam defelection system   Cooling and loading system

(b)

Deflection = $[(B_2 - 2 C_2 + D_2) - (A_1 - 2 B_1 + C_1)]$

Where,

$A_1, B_1, C_1,$ = laser readings at A, B and C at time t = 0,

$B_2, C_2, D_2,$ = laser readings at B, C and D after 8ft of travel.

Def = h − h*

8 ft   8 ft   8 ft

D   C   B   A

Point of interest

h = A–2B + C

D   C   B   A

h* = B–2C + D
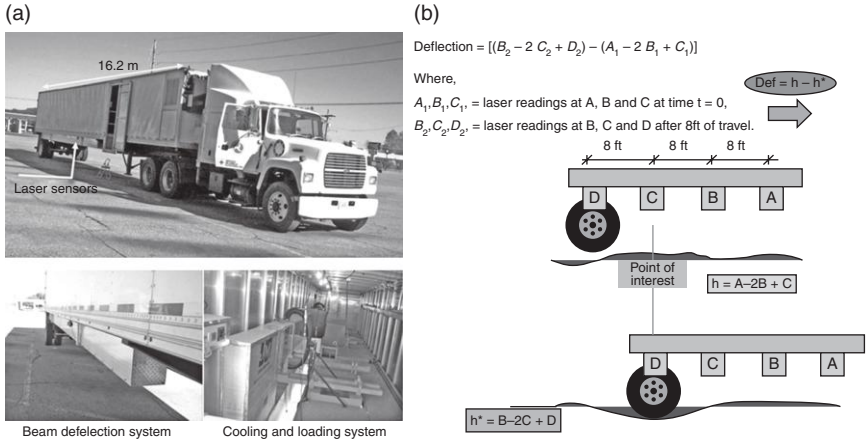
15.6 RWD system.

15.2.2   Ground-penetrating radar (GPR)

GPR is a time-dependent electromagnetic technique that can provide high-resolution 2D or 3D radar images of the subsurface. This geophysical method has been developed over the past 30 years, primarily to investigate the shallow subsurface of the earth, building materials, and infrastructure such as roads and bridges. GPR uses the principle of scattering electromagnetic waves to locate buried objects, and operates in the frequency range between 10 MHz and 3 GHz (Birken and Oristaglio, 2013).

A transmitting antenna radiates an electromagnetic wave that travels through the material at a velocity that is determined primarily by the electrical properties of the material. If the wave hits a buried object or boundary layer between two materials with different electrical properties, then part of the wave's energy is 'reflected' back to the surface, while part of its energy continues to travel downward. The wave that is reflected back to the surface is captured by a receive antenna, and recorded on a digital storage device for later interpretation (Daniels, 2004).

GPR is used routinely to assess the structure and substructure of pavements, bridge decks, airport runways, taxiways, and/or aprons. Properly applied, it can be used to assess the layers, reinforcement, moisture content, presence of voids, and other anomalies in the substructure. Changes in the conductivity and dielectric permittivity associated with these deformations or material property alterations manifest themselves as layered horizons or lateral discontinuities on the GPR records.

Pavement inspections with GPR
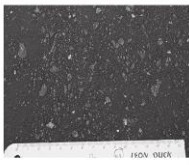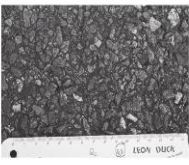
Single channel GPR pavement thickness mapping was one of the first civil engineering applications. Early work was done by Maser and Scullion (1992)

showing that a single horn antenna could accurately map multiple layers of pavement with the added benefits of being operated at highway speed without the need for road closures. Saarenketo and Scullion (2000) illustrated the wide variety of research done in the field of civil engineering pertaining to layer detection. Hugenschmidt (2004) shows an example of GPR being used to detect the thickness of a bridge deck using layer detection and the layer of reinforcement. Loizos and Plati (2007) show layer detection of multiple pavements with GPR systems at different frequencies, and delve into the accuracy of the systems by checking them against the ground truth with good results. The dielectric constant of pavement can be calculated by evaluating the wave velocity within the top pavement layer or the reflection factor of the pavement surface (Saarenketo, 2006). In multi-layer scenarios, reflecting waves may interfere with each other, requiring special data processing approaches to extract the layer information (Lahouar and Al-Qadi, 2008; Zhou *et al.*, 2010).

15.2.3   Macrotexture depth (MTD)

The MTD of a pavement is physically related to the tire/road friction coefficient and the severity of raveling (Fig. 15.7). MTD can be used to gauge frictional properties of the pavement surface (skid resistance) and to detect bleeding or segregation and non-uniformity of asphalt concrete during pavement construction. Low values indicate a smooth surface, and high values a coarser surface.

One traditional method for MTD measurement was introduced by ASTM E965 (ASTM, 2006a) and named the 'sand patch method' or 'volumetric patch method'. This method characterizes surface macrotexture by the quantity 'Mean Texture Depth.' It is straightforward to apply, and the result is three-dimensional. However, this method may give different results for different operators (China and James, 2012), and may be applied to pavement only when traffic is closed. An alternative method, more commonly used now, is the laser-based profilometer method, which is considered to

MTD (mm)	0 ~ 0.4	0.4 ~ 1.2	> 1.2
Physical meaning (level of skid resistance)	Dangerous	Fair	Good
Sample surface			

15.7 Physical meaning of pavement macrotexture. (*Note*: The length of the ruler is 15 cm.)

be more advanced, safe, and economical. Several types of equipment or systems were developed based on laser application. The standard ASTM E2157 – 09 (ASTM, 2009a) proposed this method by using a circular track meter (CT Meter) to measure and evaluate pavement macrotexture profiles for laboratory investigations and actual paved surfaces in the field (ASTM E2157, 2009a; ASTM E1845, 2009b).

Northeastern University has developed methods (Zhang *et al.,* 2012, 2013) to estimate MTD values based on the measurement of acoustic noise between tire and road surface while the vehicle is moving. The MTD is estimated by matching the principal component vector set derived from the pavement with known MTD values from existing roadways or runways. Figure 15.8 represents the pictures of the tested pavements corresponding to the predicted MTD values shown in Table 15.1.

15.2.4   International roughness index (IRI)

The IRI was established in 1986 by the World Bank, and has been widely used to measure road roughness since then. This index also indicates how comfortable drivers and passengers feel in a moving vehicle. It measures long wave length of the profile between 1.3 and 30 m (Sayers *et al.,* 1986) over a continu-
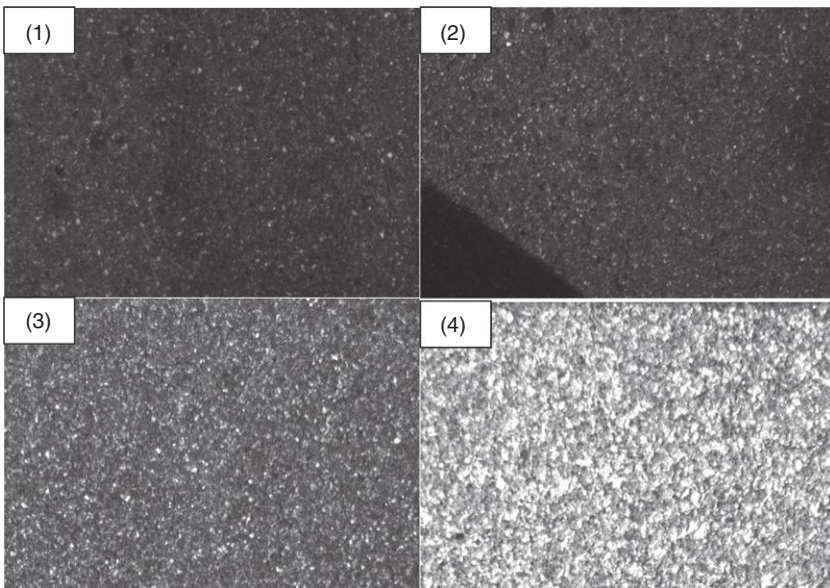
15.8 Pavement macrotexture condition. (*Note*: width × height of each picture is 37 cm × 22 cm.)

Table 15.1 Predicted MTD of pavement from Fig. 15.7

Pavement number	1	2	3	4
MTD by PCA method (mm)	0.5	0.9	1.2	2.5
MTD by energy method (mm)	0.38	0.96	1.08	2.63
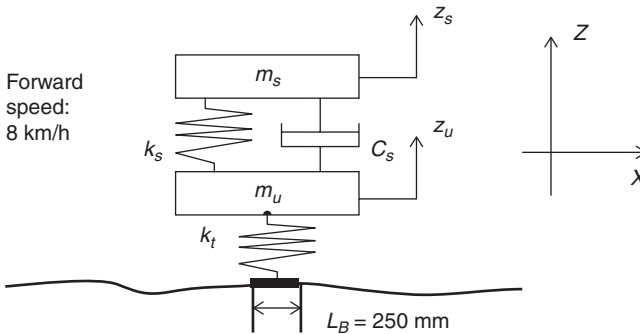
15.9 Quarter car model proposed by Sayers (1995).

ous length of at least 160 m (0.1 mile) (Sayers, 1990). IRI is commonly mea-sured by using a quarter car model (also called a gold car) shown in Fig. 15.9.

K_s is the suspension spring rate; c_s is the suspension damping rate; k_t is the tire spring rate; m_t is the sprung mass (portion of vehicle body mass sup-ported by one wheel); and m_s is the unsprung mass (mass of wheel, tire, and half of axle/ suspension). L_B is the contact length between the tire and the road. The road has a certain geometry, if the profile is known; as the quar-ter car moves over the road, the road geometry viewed as the input to the model moves the spring upward and causes the vibration of this 2 degrees of freedom (DOFs) system. The IRI of a profile is calculated and normalized by the length L as the following:

$$\text{IRI} = \frac{1}{L}\int_0^{L/v}|\dot{z}_s - \dot{z}_u|\,dt \qquad [15.2]$$

where z_s is the height (vertical coordinate) of spring mass; z_u is the height (vertical coordinate) of unsprung mass (Sayers *et al.*, 1986). $\dot{z}_s$ and $\dot{z}_u$ are derivatives of z_s and z_u respectively, representing velocity. When we inte-grate IRI, we integrate relative motion over time, which is defined by the distance traveled.

The question follows: how do people measure a road profile to obtain IRI? Two major methods are used in current profile measurement, which

are (1) manual and (2) mobile measurement. The manual profile measurement requires a person to either push a small cart outfitted with profilometers, or to use crutch-like profilometers to measure the profile while walking, such as the walking profiler G2 (arrb.com.au) from the arrb Group and the Dipstick profiler (dipstick.com) from FaceCo. The surveying speed is up to walking speed. The mobile profile measurement uses a vehicle outfitted with profilometers and accelerometers to measure the profile at driving speed (Lenz, 2011; IMS online; on-road measurement). The road profile is calculated by subtracting the double integral of acceleration from the profilometer displacement. Many technologies are used to measure displacement, such as Pavetesting from UltraTechnologies (ultra-technologies. com), Profile Gauge ViaPPS from ViaTech (viatech.no) and Road Surface Profilometer (RSP) (dynatest.com) from Dynatest. The cost is very high.

15.2.5   Spectral analysis of surface waves (SASW)

Conventional surface wave methods are time consuming and cause traffic interruption because the sensors must be in contact with the pavement surface and be repositioned for each measurement. The SASW method has been used to detect subsurface layer depths and stiffness. Subsurface information is critical for infrastructure foundation survey, diagnosis, and construction quality control for pavement network. Subsurface defects, such as loss of stiffness or layers debonding of pavement, can eventually lead to surface cracks, potholes, subsidence, and other dangerous subsequent failures. Impact echo (IE) is able to identify the debonding and thickness of a shallow top layer (Sansalone and Streett, 1997). Impulse response (IR) can test the overall dynamic stiffness/mobility of the entire pavement structure (Davis, 2003). GPR is best for locating metal materials, such as reinforcement rebars (Bungey, 2004). Chain Drag can find the debonding area by hearing the hollow sound (Costley, 2003). SASW and related methods own the capability of estimating both the thickness and elastic modulus of subsurface layers (Stokoe II and Nazarian, 1983; Krstulovic-Opara *et al.,* 1996).

Since first proposed in 1980s (Stokoe II and Nazarian 1983; Nazarian and Stokoe II, 1985), the SASW has been widely applied in geological tests to estimate the underground soil profile without coring or opening the ground (Nazarian and Stokoe II, 1985; Joh, 1996). It utilizes the phase difference between two sensing channels to measure the dispersion features of a surface wave that propagates horizontally in the soil when it is subject to an impact load. Once the dispersion curve is obtained from the test data, the layer profile and shear velocities can be estimated by the inversion procedure. Extensive research has been done to improve the accuracy and efficiency of

the inversion procedure, due to its decisive role. Conventional trial and error inversion needs forward modeling methods; the stiffness matrix is a typical choice (Kausel and Roesset, 1981). To date, SASW has been extended to investigate pavement (Olson and Miller, 2009; Ismail *et al.,* 2009) and concrete structures (Cho, 2002), originally from geological site testing.

Despite these exciting achievements of SASW, the state of the art of the surface-wave-based methods are considered to be of low efficiency due to the expertise, forward modeling involved iterations, and contact sensing. Recently, attempts were made to enhance the testing speed by using non-contact air-coupled sensors, the microphones, to replace the traditional contact sensors, accelerometers, or geophones (Zhu and Popovics, 2002), in which the radiating surface wave collected by the microphones is used to measure the dispersion curve. Air-coupled SASW has emerged as a prospective potential for fast non-contact testing. The concept of mobile testing by microphones was also presented to perform fast testing on concrete slabs (Ryden *et al.,* 2008).

One major issue that restricts the efficiency of SASW and its air-coupled extension is the iterative inversion process, which is usually time consuming and requires expertise to initialize the profile. Consequently, the SASW method is implemented as a point-to-point, post-processed, stationary test. Researchers are seeking faster, automated inversion algorithms to advance its efficiency (Nazarian *et al.,* 1995; Orozco, 2004; Zomorodian and Hunaidi, 2006). Ryden and Park (2006) proposed a global search algorithm to invert the complete phase-velocity spectrum based on fast simulated annealing (FSA). All these efforts drove the inversion for better certainty with various mechanisms; however, they still rely on the basic iterative procedure of trial and error with theoretical forward modeling. As long as the inversion methods fall into this procedure, fast implementation will be impossible for SASW method and its extensions.

Northeastern University has developed an SASW comparable method that is able to inspect subsurface conditions at normal walking speeds thanks to its use of non-contact sensors and fast air-coupled surface wave analysis algorithms (Zhu and Popovics, 2008; Cao *et al.,* 2011; Lu *et al.,* 2011, 2013). In its current state, the prototype system is mounted on a three-wheel push-cart. A microphone array is mounted underneath the cart. Each microphone is enclosed by sound-proof foam to isolate against noise from the environment. The system uses an electromagnetic hammer to supply an adjustable and traceable impact source.

The typical acoustic signal of surface wave and the signal to noise ratio (SNR) from the system are presented in Fig. 15.10. The extracted acoustic signal with a high SNR is the primary foundation for the subsurface profile estimation.
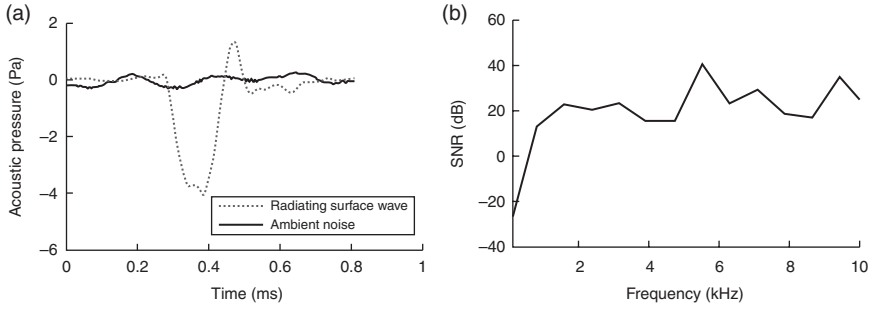
15.10 Typical SNR under field conditions. (a) Acoustic time domain signal compared to ambient noise, and (b) frequency spectrum of acoustic time series.
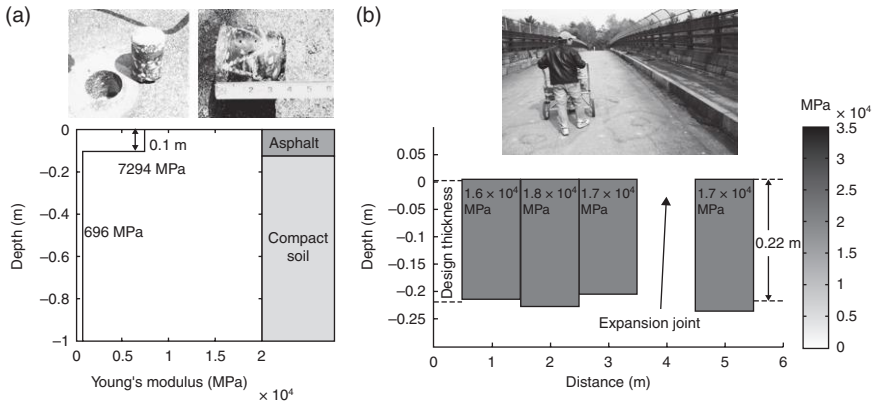
15.11 Field test cases as described in text.

The prototype cart has been tested in the field. One example surveyed asphalt pavement at a site in Rowley MA. The predicted subsurface profile and elastic modulus are shown in Fig. 15.11a. This result agrees with a borehole taken at the site, which is also shown in Fig. 15.11a.

The other field test was performed on a concrete bridge deck at the Pingree Bridge in Rowley MA. Results are shown in Fig. 15.11b in the form of a continuous scan (done at walking speed) with a spatial resolution of 1 m. The predicted depth of 0.22 m coincides with the bridge's design documents. The field test results demonstrate the capability of this new push-cart SASW system for subsurface inspection.

Table 15.2 is a summary of the current commercial sensing technologies for highway pavement condition assessment. There is a trend to select and develop a nondestructive evaluation (NDE) technique to measure and

Table 15.2 Commercial sensing technologies for highway pavement assessment

Location	Commercial sensing technology	Measurement indicator
Pavement surface	Sonic and ultrasonic testing (Hudson and Uddin, 1987)	Longitudinal roughness, rut depth measurement
	Laser-based profilometer (ASTM, 2009a)	Longitudinal transverse cracks, joint faulting, longitudinal roughness, surface texture
	Video imaging technology (Wang, 2000)	Surface cracks, distress data quantification
	Expert system (Micro PAVER) (MicroPAVER 2007; ASTM, 2011a)	PCI prediction, pavement distress condition storage
	Locked wheel method (e.g., dynamic friction tester) (ASTM, 2011b)	International friction index (IFI), skid resistance
	Multi-function pavement evaluation vehicles (e.g., ROSAN System) (Koklanaris, 1998)	MTD, IRI, longitudinal pavement profile, surface texture
Pavement subsurface	Dynamic deflection equipment (FWD) (ASTM, 2006b)	Elastic moduli of individual layers in a multi-layered system
	Seismic method (SASW) (Stokoe II and Nazarian, 1983)	Pavement moduli and layer thickness
	Ground-penetrating radar technology (GPR) (Daniels, 2004)	Pavement layer thickness, dielectric constants, locate steel reinforcement
	Thermal infrared photography (ASTM, 2007)	Void detection, delamination
	Acoustics (IE) (ASTM, 2010)	Cracks, delaminations, voids, honeycombing, debonding

cover vast areas of roads in a network-wide fashion to address future maintenance problems by prioritizing repair and effecting long-term saving.

15.3   Health assessment of bridge decks

The ASCE report card (ASCE, 2013) gives bridges only a grade of C+, commenting that 'Over two hundred million trips are taken daily across deficient bridges in the nation's 102 largest metropolitan regions. In total, one in nine of the nation's bridges are rated as structurally deficient, while the average age of the nation's 607 380 bridges is currently 42 years. The

FHWA estimates that to eliminate the nation's bridge deficient backlog by 2028, we would need to invest $20.5 billion annually, while only $12.8 billion is being spent currently. The challenge for Federal, state, and local governments is to increase bridge investments by $8 billion annually to address the identified $76 billion in needs for deficient bridges across the United States.'

Throughout the life cycle of a bridge, the deck is the part that is the most susceptible to deterioration. This is largely due to deicing salts applied during the winter months causing corrosion of the reinforcing steel. As a result, the deck must be repaired often, and replaced multiple times. To diagnose the condition of the deck, common techniques such as visual inspection and chain drag are the primary methods currently used in the United States.

15.3.1   Qualitative NDE techniques for bridge decks

Visual inspections are completed on a bi-yearly basis, where inspectors provide a numeric condition rating on a scale of 0–9, 9 indicating excellent condition (Graybeal *et al.,* 2002). The inspectors look for deterioration in the wearing surface, cracks, efflorescence, moisture, and rust staining on the deck surface, curbs, and soffit. However, after extensive analyses of the reliability of this method, it was found that problems in the assessment arose for decks that were not either new or had not extensively deteriorated (Barnes and Trottier, 2004). In addition, assessments significantly differed among inspectors for the same bridge deck (Graybeal *et al.,* 2002).

The chain drag method for inspection is a nondestructive method that involves dragging chains across the surface of a bridge deck and listening to audible sounds produced by the vibrating chains. However, analyses of this method showed that the deteriorated areas must be large for chain drag to be successful (Yehia *et al.,* 2008). Additionally, similar to visual inspection, deterioration quantities of the same bridge deck differed greatly among different assessors (Graybeal *et al.,* 2002).

15.3.2   Quantitative NDE techniques for bridge decks

In the next few subchapters we explore half-cell potential (HCP), GPR (Section 15.2.2), IE, and infrared (IR) thermography and their ability to precisely quantify deterioration in bridge decks in a little more detail. Vaghefi *et al.* (2012) provide a brief overview of those and other NDE methods that inspect only the surface of the bridge deck, which are mainly optical methods.

Half-cell potential (HCP)

The HCP method detects areas of active rebar corrosion within a bridge deck. ASTM C876 (ASTM, 1999) provides guidelines to determine these areas. For example, the standard indicates that values below −350 mV have a 90% probability of actively corroding. Similarly, those measurements above −200 mV have a 90% probability of not actively corroding. The set-up for this method is shown in Fig. 15.12. Here, a reference electrode, usually a copper–copper sulfate electrode, sits on the surface of the concrete. A voltmeter is then connected to the reference electrode on one end, and to an exposed rebar on the other end. An increasing amount of corrosion will result in an increased (more negative) voltage reading (Stratfull, 1973).

Bridge deck inspections with GPR

Bridge deck inspections with a high frequency (1–3 GHz) GPR work by rapidly scanning over the surface of the deck, parallel to the direction of vehicular travel. It is the signal reflection at the rebar level that is commonly used to assess bridge deck deterioration due to corrosion of the reinforcing steel.

A single channel GPR equipment set-up using a ground-coupled antenna (alternatively an air-coupled antenna can be used, sacrificing some of the signal detail) is shown in Fig. 15.13. The antenna is attached to a small cart with wheels so that it can be easily pushed along the deck. The antenna is connected to a data acquisition unit that displays the data in real time so that it can be inspected as it is acquired. For bridge decks, it is common to take measurements every one foot across the width of the deck. Additionally, to obtain detailed information for each rebar in the deck, signal traces are commonly collected every quarter inch along the length of the deck. Data
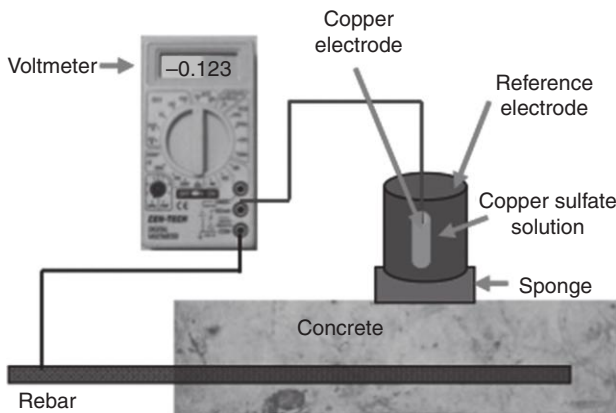
15.12 HCP set-up (Cui *et al.*, 2012).

15.13 GPR system data collection.

processing usually occurs after all data has been collected, and primarily involves extracting the amplitude, time and coordinates of each signal at the rebar level.

Impact echo (IE)

The IE test (Fig. 15.14) is used to determine the location and extent of delaminated areas in reinforced concrete, and/or its thickness (ASTM, 2010). It works by first sending sound waves through the concrete from a mechanical excitation (Wiggenhauser, 2009), or impact. The stress waves, or P-waves, travel through the concrete and reflect back toward the impact source when they come in contact with a concrete/air interface, usually being a delamination or the bottom of the deck. A transducer collects the displacements (with respect to two-way travel time) produced by the reflected waves.

The recorded surface vibrations are transformed from the time domain to the frequency domain to determine the resonant vibrations (Wiggenhauser,

15.14 IE set-up.

2009). The resonant frequency is used to calculate the thickness of the concrete area under investigation:

$$d = \frac{v}{2 \times f}$$

where: d = depth of delamination or thickness of concrete (in), f = frequency of recorded waveform (Hertz), v = prescribed velocity of the stress wave (in/s)

If the computed thickness is equal to that of the deck, the section is free from delaminations. Similar to the HCP measurements, the IE test collects measurements along a grid pattern. These results can also be contour plotted to determine delaminated areas.

Infrared (IR) thermography

IR measures the radiant temperature of the material and has been used for bridge inspection of subsurface issues in concrete. IR surveys reveal voids (air-pockets) and delaminations, because the delaminations interrupt the flow of heat through the slab. Delaminations occur when the concrete above and below the reinforcing steel begins to deteriorate due to

increased stress caused by corrosion of the steel. These delaminations are essentially thin voids in the concrete at the reinforcing steel level, which change the thermal profile of the deck. Delaminated areas and voids due to improper pouring of slabs appear as 'hot spots' in the infrared survey, which are detectable by a sensitive infrared camera (Maser and Roddis, 1990; Maser, 2008).

The IR thermography data collection is carried out according to ASTM D4788-03 (ASTM, 2007) using a high-resolution IR camera and a color video camera operated from an elevated platform attached to a survey vehicle, as shown in Fig. 15.15.

15.15 Infrared survey vehicle set-up. (*Source*: Courtesy of Infrasense Inc.)

The infrared data are collected in a series of passes across each runway, multi-lane road, or bridge deck to cover the full width. The survey produces a series of infrared images, collected at approximately one every 30 cm of vehicle travel. Typical raw images are shown in Fig. 15.16.

After the survey, the data analyst uses customized software to mosaic the images into a single strip image for each lane or section of interest. As part of this analysis process, each image is normalized to a reference temperature so that adjustments are made to local temperature fluctuations caused by wind, passing vehicles, etc. The strip image for each pass is then placed next to those of adjacent passes to produce a thermal image of the entire surface of interest.

The white blotchy areas on the IR images of Fig. 15.16 indicate delaminations or voids. The anomalous areas that appear in each image are outlined with a cursor and subsequently mapped and quantified. Figures 15.17 and 15.18 show both a composite infrared image and the resulting mapped quantities.

Table 3 lists the current commercial sensing technologies for highway bridge deck assessment. Again, emphasis is on a rapid and accurate assessment of corrosion and delamination of the bridge deck.
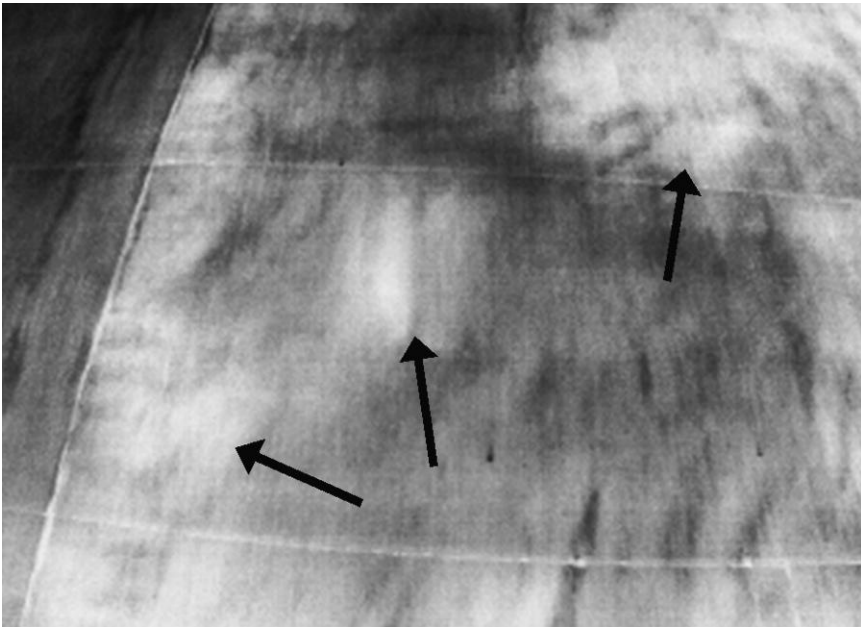
15.16 Raw infrared image. The arrows indicate delaminated areas.
(*Source*: Courtesy of Infrasense Inc.)

Table 15.3 Commercial sensing technologies for highway bridge deck assessment

Location	Commercial sensing technology	Measurement indicator
Deck surface	3D Photogrammetry (Vaghefi *et al.,* 2012)	Missing seal, cracks, and spalls within 2 feet, surface depression, surface roughness, vertical movement of bridge
	Streetview-style photography (Vaghefi *et al.,* 2012)	Missing seal, cracks, and spalls within 2 feet, surface depression, surface roughness
	Optical interferometry (Vaghefi *et al.,* 2012)	Missing seal, cracks, and spalls within 2 feet, surface depression, surface roughness, vibration
	Digital image correlation (DIC) (Vaghefi *et al.,* 2012)	Vertical and transverse movement of bridge, vibration
	Spectral analysis (Vaghefi *et al.,,* 2012)	Missing seal, cracks, and spalls within 2 feet, chemical leaching, surface depression
	Electro-optical (EO) airborne and satellite imagery (Vaghefi *et al.,* 2012)	Missing seal, spalls and surface cracks, surface depression, surface roughness
	LIDAR (Vaghefi *et al.,* 2012)	Missing seal, cracks, and spalls within 2 feet, surface depression, surface roughness, vertical and transverse movement of bridge
	Thermal IR (Maser 2008; ASTM, 2007)	Missing seal, cracks and spalls within 2 feet, surface depression
	Radar (Vaghefi *et al.,* 2012)	Missing seal, cracks, and spalls within 2 feet, surface depression, surface roughness, vibration
	Acoustics (e.g. IE method) (ASTM, 2010)	Surface cracks
Deck subsurface	Thermal IR (Maser 2008; ASTM, 2007)	Moisture in cracks, internal horizontal crack, surface depression, depression with parallel fracture
	Radar (Vaghefi *et al.,* 2012)	Fracture planes/open spaces, corrosion change, chloride content through the depth
	Ground-penetrating radar (GPR) (Daniels, 2004)	Moisture in cracks, fracture planes/ open spaces, surface depression, depression with parallel fracture, corrosion change, chloride content through the depth
	Acoustics (e.g. IE) (ASTM, 2010)	Internal horizontal crack, hollow sound, fracture planes/open spaces, corrosion change
	Half-cell potential (HCP) (ASTM, 1999)	Areas of active rebar corrosion
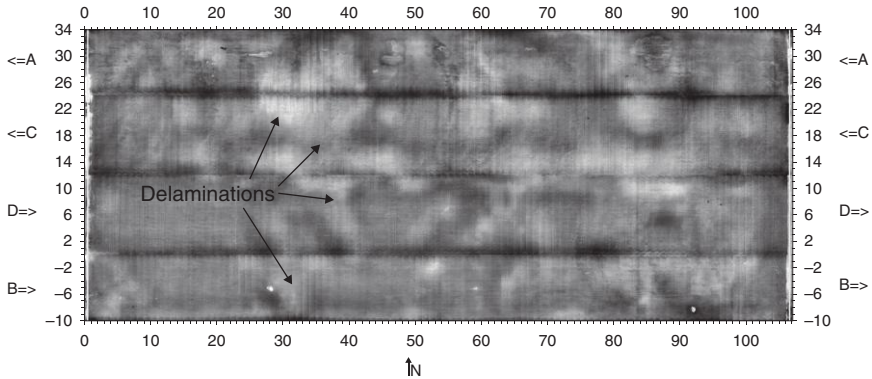
15.17 Example composite infrared image of an entire bridge deck.
(*Source*: Courtesy of Infrasense Inc.)
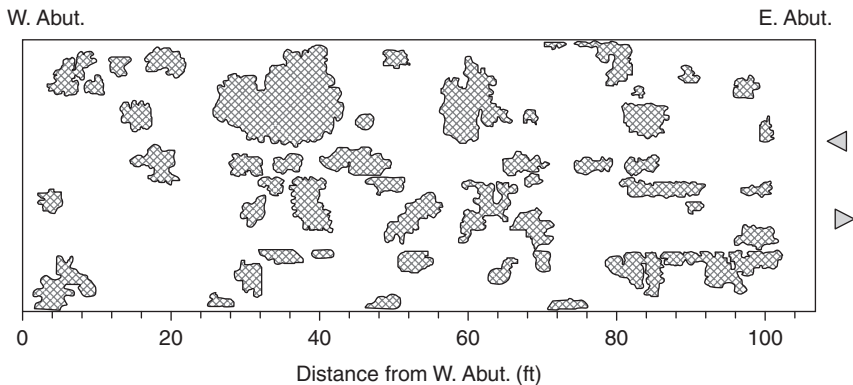
15.18 Mapped result based on example composite infrared image in
Fig. 15.17. (*Source*: Courtesy of Infrasense Inc.)

15.3.3   Condition assessment examples

In recent years NDE methods, especially GPR and HCP, have been coupled
to accurately assess the deterioration levels of numerous bridge decks. For
example, Maser (2009) analyzed 87 bridge decks using a two-level inspection
survey strategy. The first level involved surveying the decks with GPR and
IR, visually inspecting the under-deck, and completing a preliminary analy-
sis of the deterioration levels. The data were collected according to ASTM
standards, and the GPR survey was carried out using twin air-coupled horn
antennas. The GPR data were reviewed to estimate rebar depth and con-
dition, and the IR images were created to determine the overlay deterio-
ration quantities. A Level 2 analysis took place if there was an indication

that the deck had deteriorated more than 10%, and/or that there was uncertainty due to an asphalt overlay. Level 2 analysis was necessary for 38 of the bridge decks, and involved coring and a detailed mapping of delaminated and debonded conditions. An analysis of the cores and mapping showed that, of the defective cores, 83% were correctly identified by either GPR or IR.

Maser *et al.* (2012a) scanned 12 bridge decks in Minnesota with air-coupled GPR, seven of which were also surveyed with infrared thermography. All of the bridge decks were either sounded with chain dragging or had some level of pre-construction repair quantities. The necessary repair quantities were determined using a combination of GPR and IR, and GPR alone were compared to actual repair quantities or sounding in some cases. Very small differences of 3.5% and 4% resulted between the predicted and actual quantities.

Yehia *et al.* (2008) evaluated the deterioration of two nearly identical cast-in-place concrete bridge decks using GPR scans obtained from a 1.5 GHz antenna pair. A chain drag survey was conducted and compared to the GPR results. The deterioration map from the GPR survey indicated that one bridge deck was 35% deteriorated, while the chain drag survey indicated only 21% deterioration. Similarly, the GPR survey of the second bridge deck showed 21% deterioration, while the chain drag survey showed only 13%. While the results are close, it is obvious that the chain drag survey indicated a less severe state of deterioration than the GPR. Cores were also taken at locations where deterioration was found only by GPR, only by chain drag, and common areas. The coring and GPR results agreed 77% of the time, while the coring and chain drag results agreed 23% of the time.

Gucunski *et al.* (2009) evaluated nine in-service bridge decks with various NDE technologies, including GPR and HCPs. The results of one bridge deck were reported, which included a discussion of rebar level attenuation captured using a 1.5 GHz ground-coupled antenna, and its correlation to HCPs in zones of the highest deterioration.

Wamweya *et al.* (2009) studied two bridge decks, one with and one without a bituminous wearing surface. Both were scanned with a 1.5 GHz ground-coupled GPR antenna, with visual inspection and chloride ion analysis of cores to confirm the results. Both bridge decks showed good correlation between rebar level attenuation and chloride levels. For example, chloride ion concentrations of cores taken in areas of high rebar level attenuation were far beyond the threshold of 400 ppm, at which point corrosion of the reinforcing steel is expected.

Gucunski and Nazarian (2010) analyzed a bridge deck with multiple NDE technologies, including ground-coupled GPR and HCP. The GPR attenuation was determined from the rebar level and compared to the HCP through condition assessment maps. The maps indicate that zones near the

central strip and near the joints of the decks were areas of deterioration according to both methods.

Gucunski *et al.* (2012) surveyed a Virginia bridge deck with ground-coupled GPR and HCP, among others. The top rebar reflections were used for the GPR analysis which correlated well with active corrosion. Condition ratings were calculated, on a scale of 0–100, for multiple forms of deterioration including active corrosion, delamination assessment, and concrete degradation. The HCP measurements were used to determine the active corrosion rating, and the GPR amplitudes were used to determine the concrete degradation. A combined rating was also obtained for each bridge by taking an average of the ratings for each mechanism.

Maser *et al.* (2012b) presented a detailed condition study of reinforced concrete slabs extracted from a decommissioned bridge deck. The results of this study show that, using a ground-coupled GPR system to directly image the top rebar, there is a strong correlation between rebar reflection amplitude and both half-cell corrosion potential and concrete delamination. At the optimum threshold for the data set collected in this work, the spatial correlation between HCP and GPR is 90.2%, while the correlation between
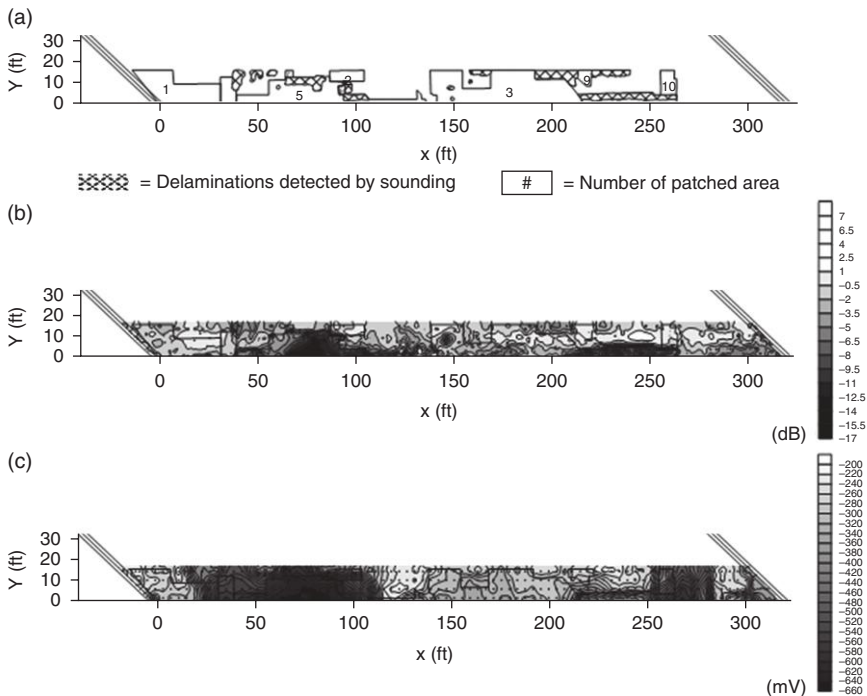
15.19 Bridge A (a) Delamination and patch map. (b) GPR signal attenuation contour plot. (c) HCP contour plot.

GPR and IE (defining delamination) is 79.3%. This difference suggests that the GPR amplitudes are more strongly correlated with the corrosion process than with delamination.

Various bridge decks around the state of Massachusetts were analyzed with GPR and HCP, among others, but due to the most significant correlations determined between the GPR and HCP measurements, only these results are displayed below. The two bridge decks shown below are about 50 years old, contain black reinforcing steel, and are significantly patched, which can be viewed in the appropriately labeled maps below. The GPR data were collected as described earlier (Section 15.3.2.2), and the amplitudes at the rebar level were extracted and spatially mapped in the figures below to assess their variation. The HCP measurements were taken on a four foot grid and are shown in Figs 15.19c and 15.20c.
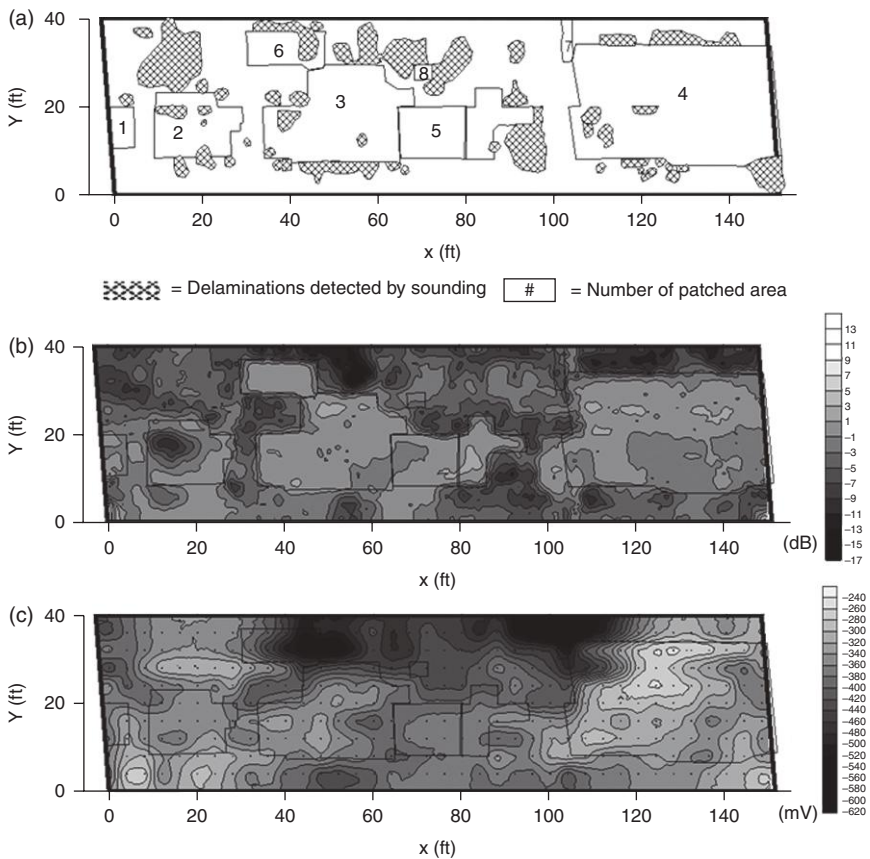
15.20 Bridge B (a) Delamination and patch map. (b) GPR signal attenuation contour plot. (c) HCP contour plot.

Due to the extensive patching of both decks, the locations of each patch were recorded during data collection so that the effects of patching could be analyzed after post-processing. The locations of each patch can be viewed in Figs 15.19a and 15.20a, which are both top-down views of the bridge decks. These locations are also overlaid onto the GPR and HCP contour plots for each bridge deck. This demonstrates the ability of the GPR rebar reflection amplitudes and HCP measurements to detect the patched areas. Additionally, this truly identifies the ability of these methods to detect variations in corrosion levels.

Both maps for each deck clearly identify areas that are significantly corroded, as well as those areas that are healthy. The dark areas signify the corroded areas, and the lighter ones are the healthy areas. By coupling these methods to identify deteriorated areas, departments of transportation can confidently know the deterioration state of the deck. In addition, the plan to repair and replace the deck, and the allocation of the necessary funding to do so, can be completed in a more efficient manner.

15.4   Future trends

Current inspections primarily consist of localized visual evaluations, are highly subjective and time consuming, require traffic blockage, provide limited information, and are not performed on a regular basis. As a result, critical maintenance decisions are often based on limited, subjective, extrapolated, and outdated information. There is a critical need for technology that can cost-effectively monitor the condition of a network-wide road system and provide accurate, up-to-date information for maintenance activity prioritization. Some of the important sensors, such as the accelerometer and global positioning system (GPS), are currently available on every smart phone. These sensors on the vehicles have been used to identify potholes. Accelerometer sensors are fixed to get acceleration along the X, Y, and Z-axes, and they record the GPS coordinates. Whenever a vehicle goes over a pothole, the acceleration change along the vertical axis indicates a significant change (DeZoysa *et al.*, 2007; Eriksson *et al.*, 2008). Microphones fixed on vehicles indicate the presence of potholes by high amplitude sound signals (Mednis *et al.*, 2010). Currently the sensors, such as the accelerometer, which can be separately installed on vehicles to track potholes, are easily available on mobile smart phones. So, using a mobile phone application the user can easily track down the presence of road bumps and potholes (Mohan *et al.*, 2008; Astarita *et al.*, 2012). The main challenge in using smartphone accelerometers for identifying potholes is the need for re-orientation of disoriented devices so that they can extract true vertical accelerations. Another drawback is that smartphone accel-

erometers are above the vehicle suspension, altering the measurements depending on the car.

Though many ideas have been suggested for better road assessment and real-time traffic monitoring, very few have lasted the test of time. Key reasons are people who do not have enough expertise in maintaining and managing roads and transportation conduct, and failure to interpret properly data gathered from road surfacing monitoring and eventually linking it to the pavement management system. Because it is essential to realize that data gathering and sensing are only small parts of road assessment, the real challenge is in the next stage of integration, so that physical sense can be made from the data; tracking of user locations may seriously violate the privacy of the participants, and mobile users are reluctant to join and contribute. It is important to develop trust-building measures among the users. Taking sensor technology, especially in road assessment, from laboratory to the real world requires a large investment although it is a low-cost solution. There are logistical issues like getting cooperation of the entities that operate the vehicles. There are also multiple authorities, municipal, provincial, and central governmental, involved in maintaining the road network. Deployment requires buy-in from all management levels, and appropriate investment.

The Versatile Onboard Traffic Embedded Roaming Sensors (VOTERS) project at Northeastern University provides one such framework to shift from periodical localized inspections to continuous network-wide health monitoring (Birken *et al.,* 2012). Initial research focused on the development of a cost-effective, lightweight package of advanced radar, acoustic, optical, and GPS sensor technology compatible with such a vision. Now that VOTERS sensing technology is installed beneath a prototype vehicle, it can monitor road conditions at both the surface and subsurface levels while the vehicle is navigating through daily traffic going about its normal business, thereby allowing for network-wide frequent assessment of roadways. Once cost-effective sensor systems are available (Cao *et al.*, 2011; Lu *et al.*, 2011; Zhang *et al.*, 2012, 2013) the obstacle to realization of such a vision is the automated processing of the multi-modal data streams (Birken *et al.,* 2013), automated transformation of the results into visualizations, and the fusion of the physical-based data into a quantity comparable to the PCI. Once such network-wide pavement condition maps and data are available, decision makers can use them to better allocate maintenance resources and assess the long-term performance of different roadway materials due to the more frequent surveying of the roads. Such an experimental deterioration dataset at unprecedented small time scales (for slow aging civil infrastructure) can also be used for life-cycle cost analysis and for comparison to its current underlying models.

15.5   References

ASCE (2013),'Report Card for Americas' Infrastructure: American Society of Civil Engineers', http://www.infrastructurereportcard.org/.

Astarita, V., Caruso, M. V., Danieli, G., Festa, D. C., Giofrè, V. P., Iuele, T. and Vaiana, R. (2012), 'A mobile application for road surface quality control: UNIquALroad', *Procedia-Social and Behavioral Sciences*, **54**, pp. 1135–1144.

ASTM (1999), 'Standard Test Method for Corrosion Potentials of Uncoated Reinforcing Steel in Concrete', ASTM C876, ASTM International, West Conshohocken, PA.

ASTM (2006a), 'Standard Test Method for Measuring Pavement Macrotexture Depth Using a Volumetric Technique', ASTM E965-96, ASTM International, West Conshohocken, PA.

ASTM (2006b), 'Standard Test Method for Deflections with a Falling-Weight-Type Impulse Load Device', ASTM D4694, ASTM International, West Conshohocken, PA.

ASTM (2007), 'Standard Test Method for Detecting Delaminations in Bridge Decks Using Infrared Thermography', ASTM D4788-03, ASTM International, West Conshohocken, PA.

ASTM (2009a), 'Standard Test Method for Measuring Pavement Macrotexture Properties Using the Circular Track Meter', ASTM E2157-09, ASTM International, West Conshohocken, PA.

ASTM (2009b), 'Standard Practice for Calculating Pavement Macrotexture Mean Profile Depth', ASTM E1845, ASTM International, West Conshohocken, PA.

ASTM (2010), 'Standard Test Method for Measuring the P-Wave Speed and the Thickness of Concrete Plates Using the Impact-Echo Method', ASTM C1383-04, ASTM International, West Conshohocken, PA.

ASTM (2011a), 'Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys', ASTM D 6433-11, ASTM International, West Conshohocken, PA.

ASTM (2012), 'Standard Test Method for Airport Pavement Condition Index Surveys', ASTM D5340, ASTM International, West Conshohocken, PA.

Barnes, C. and Trottier, J. (2004), 'Effectiveness of ground penetrating radar in predicting deck repair quantities', *ASCE Journal of Infrastructure Systems*, **10**(2), pp. 69–76.

Belt, R., Morrison, T. and Weaver, E. (2006), 'Long-Term Pavement Performance Program Falling Weight Deflectometer Maintenance Manual', No.FHWA-HRT-05-153.

Birken, R., Wang, M. L. and Wadia-Fascetti, S. (2012), 'Framework for continuous network-wide health monitoring of roadways and bridge decks', *Proceedings of Transportation Systems Workshop*, Austin, Texas.

Birken R. and Oristaglio M. (2013), 'Mapping subsurface utilities with mobile electromagnetic geophysical sensor arrays', Wang M. L., Sohn, H. and Lynch J. P. (ed.) *Smart Sensor Technologies for Civil Infrastructures Volume 2: Applications in Structural Health Monitoring*, Woodhead Publishing, UK

Birken R., Zhang J. and Schirner G. (2013), 'System level design of a roaming multimodal multi-sensor system for assessing and monitoring civil infrastructures', Wang M. L., Sohn, H. and Lynch, J. P. (ed.) *Smart Sensor Technologies for*

Civil Infrastructures Volume 2: *Applications in Structural Health Monitoring*, Woodhead Publishing, UK

Bungey, J. H. (2004), 'Sub-surface radar testing of concrete: a review', *Construction and Building Materials*, **18**(1), pp. 1–8.

Cao, Y., Lu, Y., Zhang, Y., McDaniel, J. G. and Wang, M. L. (2011), 'A fast inversion analysis algorithm for the spectral analysis of surface wave (SASW) method', *SPIE Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring*, International Society for Optics and Photonics, pp. 79831E-79831E.

China, S. and James D. E. (2012), 'Comparison of laser-based and sand patch measurements of pavement surface macrotexture', *ASCE Journal of Transportation Engineering*, **138**(2), pp. 176–181.

Cho, Y. S. (2002), 'NDT response of spectral analysis of surface wave method to multi-layer thin high-strength concrete structures', *Ultrasonics*, **40**(1), pp. 227–230.

Costley, R. D. (2003), 'Finding delaminations in concrete bridge decks', *146th ASA Meeting*, Austin, TX, http://www.acoustics.org/press/146th/Costley.htm.

Daniels, D. (2004), *Ground Penetrating Radar*, 2nd Edition, Institution of Electrical Engineers, London.

Davis, A. G. (2003), 'The nondestructive impulse response test in North America: 1985–2001', *NDT & E International*, **36**(4), pp. 185–193.

De Zoysa, K., Keppitiyagama, C., Seneviratne, G. P. and Shihan, W. W. A. T. (2007), 'A public transport system based sensor network for road surface condition monitoring', *Proceedings of the 2007 Workshop on Networked Systems for Developing Regions* (Article No. 9), ACM.

Dipstick Profiler, 'Recognized Worldwide as the Instrument of Record for Accurate and Repeatable Flatness/Levelness Measurements', http://www.dip-stick.com/.

Dondi, G., Barbarella, M., Sangiorgi, C., Lantieri, C. and De Marco, L. (2011), 'A Semi-Automatic Method for Identifying Defects on Road Surfaces', *ASCE Proceedings of 2011 International Conference on Sustainable Design and Construction*, Italy.

Elseifi, M., Abdel-Khalek, A. M. and Dasari, K. (2012),'Implementation of rolling wheel deflectometer (RWD) in PMS and pavement preservation', Report No. FHWA/11.492, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA.

Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S. and Balakrishnan, H. (2008), 'The pothole patrol: using a mobile sensor network for road surface monitoring', *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, ACM, pp. 29–39.

FHWA (2008), 'Status of the nation's highways, bridges, and transit: conditions and performance report to Congress', US DOT FHWA.

Graybeal, B., Phares, B., Rolander, D., Moore, M. and Washer, G. (2002), 'Visual inspection of highway bridges', *Journal of Nondestructive Evaluation*, **21**(3), pp. 67–83.

Gucunski, N., Feldmann, R., Romero, R., Kruschwitz, S., Abu-Hawash, A. and Dunn, M. (2009), 'Multimodal condition assessment of bridge decks by NDE and its validation', *Proceedings of the 2009 Mid-Continent Transportation Research Symposium*: Ames, IA.

Gucunski, N. and Nazarian, S. (2010), 'Material characterization and condition assessment of reinforced concrete bridge decks by complementary NDE technologies', *Proceedings of the Structures Congress*: Orlando, FL.

Gucunski, N., Maher, A., Ghasemi, H. and Ibrahim, F. (2012), 'Segmentation and condition rating of concrete bridge decks using NDE for more objective inspection and rehabilitation planning', *Proceedings of the 6th International Conference on Bridge Maintenance, Safety and Management*: Lake Como, Italy.

Gurp, C. V. (2005), 'Use of falling weight deflectometers in pavement evaluation', Report: European Cooperation in the Field of Scientific and Technical Research, COST 336, European Commission Directorate General Transport, Netherlands.

Hugenschmidt, J. (2004), 'Accuracy and reliability of radar results on bridge decks', *Ground Penetrating Radar, 2004, GPR 2004, Proceedings of the Tenth International Conference*, IEEE, Vol. 1, pp. 371–374.

IMS online, http://www.ims-rst.com/data-collection.shtml.

Ismail, M. A., Samsudin, A. R., Rafek, A. G. and Nayan, K. A. M. (2009), 'In situ determination of layer thickness and elastic moduli of asphalt pavement system by Spectral Analysis of Surface Wave (SASW) method', *Recent Advancement in Soil Behavior, in Situ Test Methods, Pile Foundations, and Tunneling*: Selected Papers from the 2009 Geohunan International Conference August 3–6, 209, Changsha, China, ASCE Publications, Vol. 192, pp. 70–76.

Jain, S. P., McCullough, B. F. and Hudson, W. R. (1971), 'Flexible pavement system – second generation, incorporating fatigue and stochastic concepts', Research Report No. 123-10, Texas Transportation Institute, Texas A&M University.

Joh, S. H. (1996), 'Advances in the data interpretation technique for spectral analysis of surface wave (SASW) measurement', Ph.D. Dissertation, the University of Texas at Austin, pp. 170–182.

Kausel, E. and Roesset, J. M. (1981), 'Stiffness matrices for layered soils', *Bulletin of the Seismological Society of America*, **71**(6), pp. 1743–1761.

Krstulovic-Opara, N., Woods, R. D. and Al-Shayea, N. (1996), 'Nondestructive testing of concrete structures using the Rayleigh wave dispersion method', *ACI Material Journal*, **93**(1), pp. 75–86.

Lahouar, S. and I.M. Al-Qadi. (2008), 'Automatic detection of multiple pavement layers from GPR data', *NDT&E International*, **41**, pp. 69–81.

Lenz, R. W. (2011), 'Pavement Design Guide Manual Notice', http://onlinemanuals. txdot.gov/txdotmanuals/pdm/nondestructive_evaluation_of_pavement_functional_properties.htm.

Loizos, A. and C. Plati. (2007), 'Accuracy of ground penetrating radar Horn-antenna technique for sensing pavement subsurface', *Sensors Journal, IEEE*, **7**(5), 842–850.

Lu, Y., Cao, Y., McDaniel, J. G. and Wang, M. L. (2011), 'Wave number estimation based method on in situ pavement ground truth with near source-receiver sensing', *Proceedings of SPIE*.

Lu, Y., Zhang, Y., Cao, Y., McDaniel, J. G. and Wang, M. L. (2013), 'A mobile acoustic subsurface sensing (MASS) system for rapid roadway assessment', *Sensors*, **13**(5), pp. 5881–5896.

Maser, K. R. and Scullion, T. (1992), 'Automated Pavement Subsurface Profiling Using Radar: Case Studies of Four Experimental Field Sites', Transportation Research Record, no. 1344, Transportation Research Board, National Research Council, Washington, D.C., pp. 148–154.

Maser, K. and Roddis, K. (1990), 'Principles of thermography and radar for bridge deck assessment', *Journal of Transportation Engineering*, **116**(5), pp. 583–601.

Maser, K. (2008), 'Integration of ground penetrating radar and infrared thermography for bridge deck condition evaluation', In *Proceedings of Symposium*. NDE/NDT for Highways and Bridges, SMT, pp. 67–74.

Maser, K. (2009), 'Integration of ground penetrating radar and infrared thermography for bridge deck condition assessment', *Proceedings of the 7th International Symposium on Non-Destructive Testing in Civil Engineering*: Nantes, France.

Maser, K., Carmichael, A., Johnson, B., Keranen, P., Holdhusen, B. and Green, D. (2012a), 'Network level bridge deck condition assessment in Minnesota', *Proceedings of the NDE/NDT for Highways and Bridges*: *Structural Materials Technology (SMT) Conference*: New York, New York.

Maser, K., Martino, N., Doughty, J. and Birken, R. (2012b), 'Understanding and detecting bridge deck deterioration using ground penetrating radar', *Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, **2313**, pp. 116–123.

Mednis, A., Strazdins, G., Liepins, M., Gordjusins, A. and Selavo, L. (2010), 'Roadmic: road surface monitoring using vehicular sensor networks with microphones', *Networked Digital Technologies*, Springer, Berlin Heidelberg, pp. 417–429.

MicroPAVER (2007), 'MicroPAVER: Pavement Management System', KMS & Associates, Inc. http://www.city.pittsburgh.pa.us/district8/assets/07_pavement_mgt_system.pdf.

Mohan, P., Padmanabhan, V. N. and Ramjee, R. (2008), 'Nericell: rich monitoring of road and traffic conditions using mobile smartphones', *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, ACM, Raleigh, North Carolina, pp. 323–336.

National Research Council (U.S.), and American Association of State Highway Officials (1961), The AASHO road test: history and description of project. Washington, D.C.: National Academy of Sciences-National Research Council.

Nazarian, S. and Stokoe II, K. H. (1985), 'In situ determination of elastic moduli of pavement systems by spectral-analysis-of-surface-waves method: practical aspects', Report No. 368-1F, Center for Transportation Research, The University of Texas at Austin, pp. 67–68.

Nazarian, S., Yuan, D. and Baker, M. R. (1995), 'Rapid determination of pavement moduli with spectral-analysis-of-surface-waves method', Report No. TX-94 1243-1, Center for Geotechnical and Highway Materials Research, the University of Texas at El Paso, El Paso, Texas.

Olson, L. D. and Miller, P. (2009), 'Comparison of surface wave tests for pavement system thicknesses/moduli', *GeoHunan International Conference*: *Challenges and Recent Advances in Pavement Technologies and Transportation Geotechnics*, Vol. 189, pp. 174–179.

On-road Measurement – Measurement of Road Surface, http://www.vti.se/en/vti-offers/on-road-measurement/measurement-of-road-surface/.

Orozco, M. C. (2004), 'Inversion method for spectral analysis of surface waves (SASW)', PhD dissertation, Georgia Institute of Technology, pp. 35–114.

Pedersen, N. J. (2007), 'Pavement lessons learned from the AASHTO road test and the performance of the interstate highway system', *Transportation Research Board*, Pavement Management Section.

Ponniah, J., Sharma, B. N. and Kazmierowski, T. J. (2001), 'A critical review of an existing pavement condition rating system', *Fifth International Conference on Managing Pavements*, Seattle, Washington.

Road Surface Profilometer, http://www.dynatest.com/equipment/functional/profiling.

Ryden, N. and Park, C., B. (2006), 'Fast simulated annealing inversion of surface waves on pavement using phase velocity spectra', *Geophysics*, **71**(4), pp. 49–58.

Ryden, N., Lowe, M. J. and Cawley, P. (2008), 'Non-contact surface wave scanning of pavements using a rolling microphone array', *AIP Conference Proceedings*, **975**, pp. 1328–1332.

Saarenketo, T. and T. Scullion (2000), 'Road evaluation with ground penetrating radar', *Journal of Applied Geophysics*, **43**, pp. 119–138.

Saarenketo, T. (2006), 'Electrical properties of road materials and subgrade coils and the use of ground penetrating radar in traffic infrastructure surveys', Dissertation, Oulu University Press.

Sansalone, M. and Streett, W. B. (1997), *Impact-Echo*: *Nondestructive Testing of Concrete and Masonry*, Bullbrier Press, Jersey Shore, PA.

Sayers, M. W., Gillespie, T. D. and Paterson, W. D. (1986), 'Guidelines for conducting and calibrating road roughness measurements', World Bank Technical Paper.

Sayers, M. W. (1990), 'Profiles of roughness', *Transportation Research Record* **1260**, pp. 106–111.

Sayers, M. W. (1995), 'On the calculation of international roughness index from longitudinal road profile', *Transportation Research Record* **1501**, pp. 1–12.

Shahin, M. Y. and Kohn, S. D. (1981), 'Pavement Maintenance Management for Roads and Parking Lots', Report No.CERL-TR-M-294, U.S. Army Construction Engineering Research Laboratory, Champaign, Illinois.

Shahin, M. Y. (1994), *Pavement Management for Airports, Roads and Parking Lots*, Chapman and Hall, New York.

Shahin, M. Y., Cline, G. D. and Burkhalter, J. A. (2003), 'Automated data collection for pavement condition index survey', *Transport Research Board*, Illinois.

Smith, R. E., Freeman, T. J. and Pendleton, O. J. (1998), '*Evaluation of Automated Pavement Distress Data Collection Procedures for Local Agency Pavement Management'*, Texas Transportation Institute, Texas.

Smith, R. E. and Chang-Albitres, C. M. (2007), 'The Impact of Semi-Automated Pavement Distress Collection Methods On Pavement Management Network-Level Analysis Using The MTC STREET SAVER', Texas Transportation Institute, Texas.

Stokoe II, K.H. and Nazarian, S. (1983), 'Effectiveness of ground improvement from spectral analysis of surface waves', *Proceedings of the European Conference on Soil Mechanics and Foundation Engineering*, **1**, pp. 91–94.

Stratfull, R. (1973), 'Half cell potentials and the corrosion of steel in concrete', *Highway Research Record* **433**, pp. 12–21.

UltraTechnologies, http://www.ultratechnologies.com/Pavement_Performance.html.

Vaghefi K., Oats, R. C., Harris, D. K., Ahlborn, T. M., Brooks, C. N., Endsley, K. A., Roussi, C., Shuchman, R., Burns, J. W. and Dobson, R. (2012), 'Evaluation of commercially available remote sensors for highway bridge condition assessment', *Journal of Bridge Engineering*, **17**(6), pp. 886–895.

ViaTech, http://www.viatech.no/ezpublish-4.2.0/index.php/nor/Hjem/ViaPPS.

Walking Profiler G2, http://www.arrb.com.au/Equipment-services/Walking-Profiler-G2.aspx.

Wamweya, A., Torgashov, E. and Anderson, N. (2009), 'Bridge deck assessment using ground penetrating radar (GPR)', *22nd EEGS Symposium on the Application of Geophysics to Engineering and Environmental Problems*: Fort Worth, Texas.

Wiggenhauser, H. (2009), 'Advanced NDT methods for quality assurance of concrete structures', *Non-Destructive Testing in Civil Engineering*, Nantes, France.

Yehia, S., Abudayyeh, O., Abdel-Qader, I. and Zalt, A. (2008), 'Ground-penetrating radar, chain drag, and ground truth: correlation of bridge deck assessment data', *Transportation Research Record*, No. 2044, Transportation Research Board of the National Academies, pp. 39–50.

Zhang, Y., Ma, X., McDaniel, J. G. and Wang, M. L. (2012), 'Statistical analysis of acoustic measurements for assessing pavement surface condition', *SPIE Proceedings*, **8347**, *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, San Diego, California, pp. 83471F-1-83471F-14; doi:10.1117/12.916955.

Zhang, Y., McDaniel, J. G. and Wang, M. L. (2013), 'Estimation of pavement mac-rotexture with acoustic measurement through principal component analysis', submitted to *Journal of Transportation Engineering*, 10.1061/(ASCE)TE.1943–5436.0000617, 04013004.

Zhou, H., Li, S. and Zhu, J. (2010), 'Automatic layer-interface detection of pavement based on matched filter', *Wireless Communications Networking and Mobile Computing (WiCOM), 2010 6th International Conference*, *IEEE*, pp. 1–3.

Zhu, J. and J. Popovics (2002), 'Non-contact detection of surface waves in concrete using an air-coupled sensor', *AIP Conference Proceedings*, **615**, pp. 1261–1268.

Zhu, J. and Popovics, J. S. (2008), 'Non-contact NDT of concrete structures using air coupled sensors', NSEL Report Series, Report No. NSEL-010, University of Illinois at Urbana-Champaign: Urbana, IL.

Zomorodian, S. and Hunaidi, O. (2006), 'Inversion of SASW dispersion curves based on maximum flexibility coefficients in the wave number domain', *Soil Dynamics and Earthquake Engineering*, **26**(8), pp. 735–752.

16

Sensing solutions for assessing and monitoring railroad tracks

P. RIZZO, University of Pittsburgh, USA

Abstract: In railroads, the increase in axle loads, train speed, and traffic rises the risk of damage. Among the several types of damage, transverse defects in the rail head represent the primary cause of accidents. Thus, the effective inspection or the permanent monitoring of railroad tracks is necessary. This chapter reviews the most common sensing methods to detect damage in rails. The nondestructive evaluation techniques suitable for periodic inspection and some approaches for permanent monitoring of the tracks are introduced. The general principles, advantages, and limitations of each of these techniques are described.

Key words: railroad structures, rail head, nondestructive evaluation (NDE), structural health monitoring (SHM), sensors.

16.1 Introduction

In railroad transportation, the increase in axle loads, train speed, and traffic frequency causes ever-increasing contact forces, which may damage rails or vehicles. Rail defects range from internal cracks to surface defects. Each kind of defect has a specific damage fatigue process, and a specific lifetime and safety impact (UIC, 2001; Zerbst *et al.*, 2005; Oukhellou *et al.*, 2008). Among the several types of damage, transverse defects represent the primary cause of accidents. These defects are cracks developing in a direction perpendicular to the rail running direction and include transverse fissures initiated inside the rail head, and detail fractures initiated at the head surface as rolling contact fatigue (RCF) defects. Since these defects cause the weakening of tracks, they may cause derailments.

This chapter seeks to provide a comprehensive overview of the nondestructive technologies and the sensing systems that have been developed to assess rails and to overcome one of the challenges in rail defect detection technology: the discrimination between real defects and rail singularities, such as fishplated joints, welded joints, and switch frog, i.e. the point in the switch where two rails cross. Although the relevant work conducted over the

last two decades will be reported, this chapter focuses on the advancements over the last 5 years. Interested readers can refer to past review papers for comprehensive reports of techniques and practices adopted in the 1990s and the early 2000s (Cannon *et al.*, 2003; Clark, 2004; Barke and Chiu, 2005; Zerbst *et al.*, 2005; Papaelias *et al.*, 2008). The chapter is structured as follows. Section 16.2 illustrates the most common types of defects in rails. Section 16.3 describes the fundamental principles and the application of the nondestructive evaluation (NDE) methods used for periodic inspection of rails. Section 16.4 illustrates the sensing systems proposed for permanent monitoring of tracks. Finally, Section 16.5 reviews the sensing systems used in Japan and Korea for high-speed rail.

16.2   Defects in rails

Safety statistics from the United States Federal Railroad Administration (FRA, 2012) indicate that in the decade 2001–2010 train accidents caused by track, roadbed, and structures caused nearly 9000 derailments. Table 16.1 shows the ten largest causes of these accidents. Transverse/compound fissure and detail fracture-shelling/head check were the two largest causes of damage, accounting for over 1200 accidents, near $300M in cost, which represents 23.6% of direct cost reported by the FRA.

Table 16.1 U.S. Federal railroad statistics relative to the ten largest causes of accidents in the area of track, roadbed, and structures during the decade 2001–2010 (FRA, 2011)

Specific causes: 2001–2010	Count	%	Amount ($)	%
T220- Transverse/compound fissure	789	8.4	153 801 982	12.2
T207- Detail fracture – shelling/ head check	423	4.5	143 960 387	11.4
T109- Track alignment irreg (buckled/sunkink)	300	3.2	103 305 256	8.2
T110- Wide gage (defective/missing crossties)	1815	19.4	101 567 525	8.0
T299- Other rail and joint bar defects	136	1.5	68 578 011	5.4
T221- Vertical split head	390	4.2	47 683 472	3.8
T102- Cross-level track irreg. (not at joints)	292	3.1	46 301 793	3.7
T314- Switch point worn or broken	779	8.3	41 026 680	3.3
T204- Broken weld (field)	67	0.7	38 366 044	3.0
T001- Roadbed settled or soft	261	2.8	38 188 266	3.0
Total	5252	56.1	782 779 416	62.0

The TXXX are train accident cause codes (FRA, 2011).

Similar statistics limited to the year 2010 confirm this trend. In 2010, however, rail buckling caused the largest amount of direct cost, accounting for 15% of damage losses reported by the FRA in the category track, road bed, and structures (Table 16.2). Overall, rail discontinuities are classified according to their orientation with respect to the geometrical planes of the track and the position in the rail section (Sperry, 1989; Lanza di Scalea, 2007; Oukhellou *et al.*, 2008). The head is the part of the rail where defects occur most frequently. Head defects can be distinguished between those having internal origin and those having surface origin. Progressive trans-verse cracking, kidney-shaped fatigue cracks, horizontal cracking with or without transverse cracking of the head, horizontal cracking beneath the gauge corner, and longitudinal-vertical cracking all belong to the first cat-egory. RCF damage, shelling, head checks, head squats, and corrugation originate at the surface. Sometimes, certain anomalies coalesce and consti-tute the most severe damage. For instance, RCF initiates at the surface as horizontal head checks or squats. A few millimeters from the surface, they can turn into a transverse crack and develop a detail fracture (Care *et al.*, 2005, 2006; Thomas *et al.*, 2010). Figure 16.1 shows an example of a few defect types located in the head and an example of buckled rail. Rail-web and rail-foot defects include longitudinal and vertical cracking, cracking occurring at fishplate bolt holes or other holes found in the web, transverse fatigue cracking, and rail-foot corrosion. Some of these anomalies originate during manufacturing, such as the transverse fissure from hydrogen nuclei

Table 16.2 U.S. Federal railroad statistics relative to the ten largest causes of accidents in the area of track, roadbed, and structures during the year 2010

Specific causes: 2010	Count	%	Amount ($)	%
T109-Track alignment irreg. (buckled/ sunkink)	37	5.7	16 998 896	15.0
T220-Transverse/compound fissure	60	9.2	13 809 082	12.2
T207- Detail fracture – shelling/head check	31	4.8	13 145 495	11.6
T001- Roadbed settled or soft	28	4.3	10 194 896	9.0
T102- Cross-level track irreg. (not at joints)	13	2	6 566 251	5.8
T110-Wide gage (defective/missing crossties)	92	14.1	6 420 722	5.7
T204- Broken weld (field)	6	0.9	5 460 860	4.8
T299- Other rail and joint bar defects	6	0.9	4 611 183	4.1
T316-Turnout frog (rigid) worn, or broken	3	0.5	3 524 558	3.1
T210- Head and web sep. (outside jt. bar limit)	28	4.3	3 453 051	3.1
Total	304	46.7	84 184 994	74.4

(a)

(b)

(c)
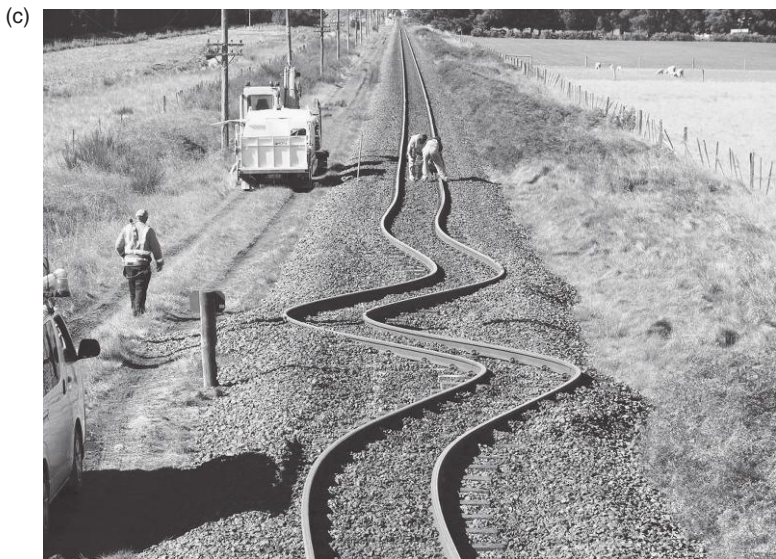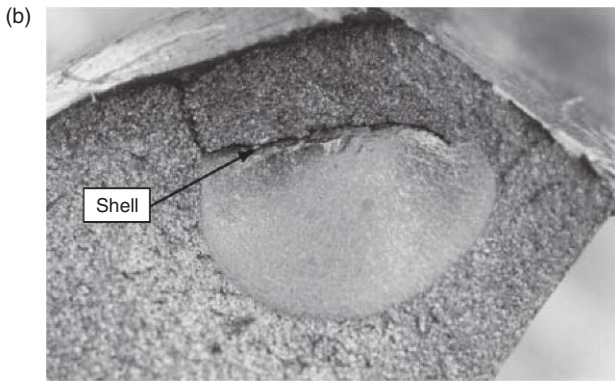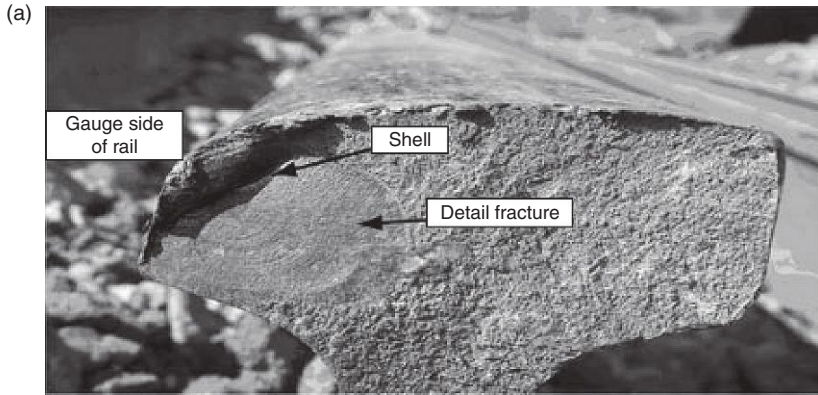
16.1 Example of detail fracture under a shell. Reprinted with permission from (a) TSB (2005a). (b) TSB (2005b). (c) AGU (2010). (d) Verysit (2011).

(Continued)

(d)

16.1 Continued

(Lanza di Scalea, 2007). By looking at the statistics presented in Tables 16.1 and 16.2, it is evident that broken welds are the cause of the most expensive accidents in the rail industry.

There is another problem in rails related to welds. Internal defects such as shrinkage cavities, microporosities, inclusions, and a coarse dendrite micro-structure can affect the structural integrity and fatigue performance of the rails by acting as crack initiation points. As a result, failure of welds can take place rapidly, with no indication of fatigue cracking.

16.3 Nondestructive evaluation of rails

In this section the fundamental principles and the application of the NDE methods used for the periodic inspection of rails are presented.

16.3.1 Eddy currents (EC)

EC-based methods measure a material's response to electromagnetic fields over a specific frequency range, typically a few kHz to several MHz (Shull, 2002). The principle of any EC inspection is illustrated in Fig. 16.2. A
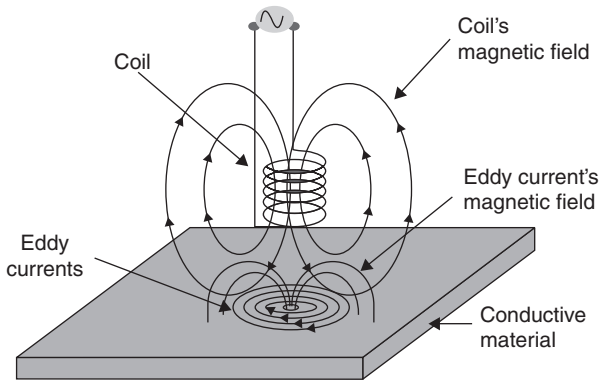
16.2 Schematic principles of the eddy current inspection.

magnetic coil with alternating current induces a time-varying magnetic field that causes an electric current to be generated in the test object. These currents produce small magnetic fields around the material which, by opposing the original field, change the impedance of the magnetic coil. The structural anomalies are flagged by observing the change in the magnetic coil impedance. EC-based inspection is a non-contact method that can be used to inspect the surface or near-surface of conductive materials (Shull, 2002). The method does not require surface preparation, and it is sensitive to lift-off variations, i.e. distances between the probes and the test specimen, which is usually less than few centimeters, and it is characterized by the skin depth, which is the depth of penetration of the alternating current.

In railroad track applications, EC probes are positioned at a constant distance from the surface of the head with particular attention given to any lift-off variations that may occur during inspection (Thomas *et al.*, 2006). For several years, the application of this method was limited to the inspection of surface welds. Advanced EC systems were then developed to identify the presence of RCF, wheel burns, grinding marks, and short-wave corrugation (Thomas *et al.*, 2000; Junger *et al.*, 2004; Pohl *et al.*, 2004, 2006; Thomas *et al.*, 2006; Papaelias *et al.*, 2008). Over the last 10 years, these systems have been integrated in grinding trains, inspection trains, and trolleys to detect RCF damage along a long length of rail (Thomas *et al.*, 2010).

NEWT developed an electromagnetic rail inspection system based on the field gradient imaging (FGI) technology, known as LIZARD®. LIZARD® rail inspection technology can be employed for plane rail, switches and crossings (switchblades) either as a trolley-mounted device or as a hand-held portable system (LIZARD, 2011a, b). In order to reach inspection speeds above 70 km/h, EC probes are combined with ultrasonic testing, which is described in Section 16.3.5. The inspection speed achieved by the

combined ultrasonic/EC system is typically 75 km/h, but higher speeds, up to 100 km/h, have been reported (Thomas *et al.*, 2000; Pohl *et al.*, 2003; Junger *et al.*, 2004; Thomas *et al.*, 2006; Papaelias *et al.*, 2008). Oukhellou *et al.* (1999) designed a system made of non-contact double-coil and double-frequency (10 and 100 kHz) EC sensors for detecting broken rails and large head spalls of the rail, and mounted underneath a subway train. The sensing system, positioned at a 20 mm distance from the head of the rail, was designed to account for vertical and horizontal displacements caused by the bogie dynamics, to withstand 100 km/h maximum speed, and tolerate up to 10 g of acceleration.

To improve damage-detection capabilities, EC-based measurements can be processed by using algorithms such as multiple signal classification (MUSIC) (Mehel-Said *et al.*, 2008). The MUSIC algorithm is based on the eigen-decomposition of the signal covariance matrix, which produces signal and noise subspaces. The approach, which has been tested on a subway line, is particularly effective to isolate shelling. Before the application of the MUSIC algorithm, the same group of investigators at the Centre de Recherche en Automatique de Nancy in France used other algorithms, such as time-heuristic approaches, wavelet analysis, inverse filtering, and independent component analysis to classify defects or singularities (Bentoumi *et al.*, 2003, 2004a, b; Lauer *et al.*, 2004). With the help of prior information extracted from an infrastructure database, Oukhellou *et al.* (2008) designed a Bayesian network to infer the probabilities of real or false (rail singularities) defect classes on the basis of previous decisions.

16.3.2   Alternating current field measurement (ACFM)

The ACFM technique, which can be considered a subset of the EC technology, is a non-contact electromagnetic technique patented by TSC Inspection Systems (TSC, 2011) for detecting and sizing surface-breaking cracks in a range of different materials and through coatings of varying thickness. Single- or multi-element ACFM probe arrays introduce an electric current locally into the structure and measure the associated electromagnetic field. The presence of a crack disturbs the alternate current flowing in the component. In contrast to EC sensors, which are required to be placed at a close and constant distance from the inspected surface, a maximum operating lift-off of 5 mm is possible without significant loss of signal when using ACFM probes. The fact that the signal strength diminishes with the square of lift-off and not with its cube, which is the case for EC sensors, enables the ACFM technology to cope with much greater lift-off (Howitt, 2002; Papaelias *et al.*, 2008).

In 2000, TSC began the development of a portable pedestrian-operated ACFM walking stick system to both detect and size RCF cracking on rails

(Howitt, 2002; Lugg and Topp, 2006; Papaelias *et al*., 2007). One of the most recent advancements was presented by Papaelias *et al*. (2009), which simulated a high-speed ACFM rail inspection using a rotatory test piece containing spark-eroded notches. It was found that the recorded signals were unaffected by the inspection speed provided that the lift-off distance remained constant.

16.3.3   Infrared thermography

The analysis of active or passive infrared (IR) images has recently gained new attention, due to the technological improvement of the detectors' thermal sensitivity and spatial resolution. By exploiting the variation of the thermal properties between the defect and the host structure, experienced thermographers can infer the presence of anomalies from the observation of a single frame or a video sequence regarding the temperature distribution of the structure's surface. IR-based NDE provides a non-contact method able to accomplish full-field defect imaging of virtually any material (Vandone *et al*., 2012).

Greene *et al*. (2007) proposed a method based on the change in IR emission of the rail surface during the passage of a train wheel to detect cracks in rails. The approach utilizes the differential thermography method to observe the IR radiation emitted from a surface during a change in stress condition. When a wheel passes over a section of a track, an elastic bending stress is induced in the rail, which is relaxed once the wheel has moved away. The rail undergoes a minute transient temperature change when loaded: the head is in compression and experiences a slight temperature rise, while the foot is in tension and shows a slight temperature drop. The method was tested on a 1 m long section of rail formed from two welded 0.5 m long pieces. The surface was sprayed with a thin layer of high-carbon matte black paint. The rail section was loaded in three-point bending and preliminary data were presented.

A hybrid method based on EC and IR technology was proposed for rail-surface characterization. The method, called pulsed eddy current (PEC) thermography, consists of delivering short bursts (typically 20 msec–1 sec) of high frequency (50–500 kHz) electromagnetic excitation to the test object. The induced ECs that flow in the material are diverted whenever they encounter a discontinuity. This produces areas with increased and decreased ECs density. Areas with increased density experience higher levels of ohmic heating, which are detected by an IR camera (Wilson *et al*., 2010; 2011). This heating system is more effective than using halogen lamps, whenever the targeted defect is a vertical surface-breaking. Wilson *et al*. (2010) reported on the use of PEC thermography for the detection and characterization of RCF. Although the method is truly non-contact, it requires high power to

produce a significant temperature increase and, like any EC-based system, is sensitive to lift-off distance.

IR technology applied to rail inspection presents several challenges, including variability of the surface cleanliness, maintenance of the correct field of view during motion, ruggedization of the camera, low emissivity of polished steel, contaminants (grease and dirt) on the rail surface, and optical obscuration caused by the various fittings and fixtures used to hold the rail in place.

16.3.4   Magnetic flux leakage (MFL)

MFL consists of magnetizing the test object by means of permanent magnets or DC electromagnets, and scanning its surface with some form of a flux-sensitive sensor (Bray and Stanley, 1997). The magnetic flux lines, coupled into the specimen using metal 'brushes' or air coupling, flow through the material and complete a magnetic path between the pole pieces. When the flux is contained within the test piece, its detection in the air space surrounding the object is very difficult. However, if a structural anomaly disrupts the surface of the magnetized structure, the permeability is changed and the flux leaks from the discontinuity (Mix, 2005). Defects are therefore detected by measuring changes in the rails' magnetic permeability.

Rail inspection using MFL consists of positioning search coils at a constant distance from the rail. Whenever a transverse surface or near-surface defect is present, the ferromagnetic domains in the steel do not support the magnetic field flux and some flux leaks. MFL sensors are suitable for locating near-surface or surface transverse defects, such as RCF cracking, but are ineffective at capturing deep internal cracks and rail-foot corrosion. For fissures MFL is not suitable, because they run parallel to the magnetic flux lines and hence do not cause sufficient flux leakage. For damage located in the web or in the foot, the sensing coils are too distant. MFL is also adversely affected by increasing the inspection speed; as speed increases, the magnetic flux density in the rail head decreases. Clark (2004) reported that above 35 km/h the signal becomes too weak to be able to detect defects.

16.3.5   Ultrasonic testing (UT)

Bulk waves

UT is based on the propagation of ultrasonic stress waves. It is conventionally performed from the top of the rail head in a pulse–echo configuration either manually, using dedicated portable ultrasonic equipment placed on push-trolleys, or by using special test vehicles. Ultrasonic waves are transmitted by transducers located inside a water-filled wheel known as roller search unit or

a sled carrier (Clark 2004; Thomas *et al.*, 2006). The orientation of the transducer with respect to the head varies such that the wave is refracted inside the head at 0°, 37° or 45°, and 70° angles. The reflected or scattered energy of the transmitted beam is then detected using a collection of transducers. The amplitude of any reflections, together with their time of arrival, is utilized to locate and identify the defects and to assess the overall structural integrity of the rail. To reduce the acoustic mismatch due to the presence of air entrapped between the probe wheel and the rail, water can be sprayed on the rail surface using ad hoc sprinklers (Clark, 2004).

Sometimes, to locate the presence of longitudinal defects such as vertical split heads and shear defects, the ultrasonic transducers are positioned to look across the rail head. Although the inspection speed achieved by inspection trains with this method can vary from 40 up to 80 km/h (Clark, 2004; Thomas *et al.*, 2006), the actual inspection speeds can be as low as 15 km/h, owing to the need to remove uncertainties related to false positives. In fact, one of the detrimental issues related to the inspection of rail by means of wheels is the number of false readings, which tends to be very high (Clark, 2004). Test trains can achieve inspection speeds as high as 100 km/h (Aharoni and Glikman 2002; Aharoni *et al.*, 2002; Thomas *et al.*, 2006). However, as pointed out by Papaelias *et al.* (2008), there is limited information available about the quality of the performance of these vehicles at such speeds.

To reduce the number of false positives and increase the inspection speed, more realistic detection thresholds can be set, or the time window of the processed signals can be optimized. However, if the threshold is set too high, the system may generate false negatives. If the time window is set too close to the origin that corresponds to the rail surface, the detector will be prone to excessive noise. On the other hand, if the time window is too far from the origin then sub-surface cracks will go undetected (Papaelias *et al.*, 2008).

In general, ultrasonic test trains perform relatively well in detecting deep surface-breaking and internal defects, particularly in the rail head and web. However, RCF defects that are less than 4 mm deep may not be detected. Moreover, surface damage such as shelling and head checks can prevent the ultrasonic beams from reaching the internal defects, resulting in false negative readings. This occurrence is schematized in Fig. 16.3. Ultrasonic test trains can also miss some defects in the rail foot, especially corrosion, as this part of the rail can only be scanned partially. They also perform relatively poorly when inspecting alumino-thermic welds (Clark and Singh, 2003; Papaelias *et al.*, 2008).

Phased array

Phased arrays use multiple ultrasonic elements and electronic time delays that create beams by constructive and destructive interference. Phased arrays
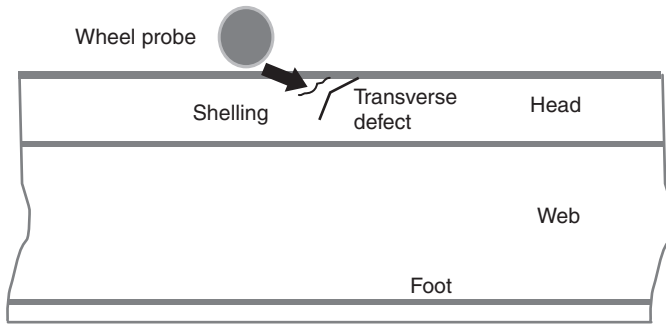
16.3 Scheme of a transverse defect masked by the presence of a shelling, when the rail head is inspected by means of a wheel probe.

provide ultrasonic beams that can be steered, scanned, swept, and focused electronically (Olympus NDT, 2011). They offer more flexibility than single-element ultrasonic transducers because they can achieve (a) beam steering and (b) beam focusing without moving the transducers. Traditional single-element ultrasonic transducers excite a beam propagating along a fixed direction and, if necessary, can focus the beam on a fixed point via acoustic lenses. However, neither beam direction nor focal point can be changed in a single transducer without moving the transducer or replacing the acoustic lens. Phased arrays comprise multiple transducers, which are electronically delayed to build a constructive interference wavefront. By changing the time delays between the elements of the array, the resulting wavefront can be pointed to different directions (beam steering) and also focused at different depths (dynamic focusing) without requiring mechanical motion.

Owing to the large amount of data generated during phased array inspections, data processing is not as straightforward as with conventional ultrasonic transducers, and therefore the inspection speed is lower. For this reason, current research focuses mainly on the development of the signal processing algorithm as well as optimizing the sensing hardware to increase the speed of phased array-based inspection systems. A new in-parallel analysis concept (known as the fast automated angle scan technique (FAAST)) has been developed by Socomate to address the processing problem (Caperet, 2006). The 128-channel system, developed for rail inspection, is capable of processing in real time the data obtained from a multi-element probe. The system can achieve inspection speeds of up to 100 km/h. The main inspection angles are $-70°$, $-35°$, $0°$, $+35°$, and $+70°$. Several research groups in Europe and North America are currently involved in the development of ultrasonic phased arrays for rail inspection applications (Utrata, 2002; Utrata and Clark, 2003; Garcia and Zhang, 2006; Papaelias *et al*., 2008).

The SNCF (French National Railways) uses linear phased arrays to detect and characterize small gas inclusions in the bulk of the rail repairs (Bredif *et al.*, 2008). The probe is moved along the rail axis while an electronic scan is performed in the cross-section of the rail. The shape of the probe is cylindrically curved to naturally focus the ultrasonic beam in the cross-section of the rail; focusing in the incident plane is done by applying electronic delay laws. As some head rails can present a profile different from the nominal one, the probe is partially flexible through a central articulation.

Guided waves

Whenever an ultrasound propagates into a bounded media, a guided ultrasonic wave (GUW) is generated. The wave is termed 'guided' because it travels along the medium guided by the medium's geometric boundaries. GUWs propagate along, rather than across, the waveguide. The advantage of GUW inspection is its ability to probe long lengths of the waveguide, locating cracks and notches from a few monitoring points, while providing full coverage of the waveguide's cross-section (Rizzo, 2010).

GUWs are gaining growing interest for rail inspections as an improvement over wheel-type ultrasonic methods. In general, GUWs are ideal in those monitoring applications that can benefit from built-in transduction, involve moderately large inspection ranges, and require high sensitivity to small flaws. In rail applications, because these waves propagate along, rather than across, the rail they are ideal for detecting the critical transverse defects. They are also potentially not sensitive to surface shelling because they can run underneath this type of discontinuity. Sometimes when the wavelength of GUWs is of the same order of magnitude as the rail curvature radius, they are referred to as surface-guided waves (Hesse and Cawley, 2006). The selection of the proper wavelength is important to achieve a penetration depth sufficient to probe the entire rail head cross-section, to allow for the screening of several meters of rail from a single inspection point, and to reduce sensitivity to non-critical features.

Gharaibeh *et al.* (2011) conducted an experimental and numerical study to select the guided wave modes that enhance the detection of damage. The experiments were carried out to generate a specific flexural mode (the F3 mode) on the head, a specific torsional mode (the T2 mode) in the web, and the flexural F3 mode along the foot. The results showed that the T2 wave mode is sensitive to the discontinuities induced in the web, while the F3 and F2 are unaffected by web damage. Transverse defects were machined in the head and foot. A slot was introduced into the rail head at 4.46 m from the excitation point and was detected when the depth was increased to 2 mm. A transverse slot was induced in the foot at 4 m from the excitation point, with

a depth varying from 1 to 5 mm in steps of 1 mm. No identifiable reflections were recorded until the defect depth reached 5 mm.

In the last 10 years, several groups have proposed different non-contact techniques to generate and detect GUWs. Researchers at the Penn State University (USA) and University of Warwick (UK) developed systems based on the use of electromagnetic acoustic transducers (EMATs) (Hayashi *et al.*, 2003; Dixon *et al.*, 2004; Rose *et al.*, 2004, 2006; Edwards *et al.*, 2006a, b, c, 2007; Fan *et al.*, 2007). An EMAT is made of a permanent magnet or an electromagnet and an electric coil. The coil is driven with a large alternating current at ultrasonic frequency, which generates an AC magnetic field. When the test material is close to the EMAT, ultrasonic waves are generated in the test material through the interaction of the two magnetic fields. The orientation of the magnetic field, the geometry of the coil, and the physical and electrical properties of the material under investigation have a strong influence on the ultrasonic wave generated within the sample. A rail test vehicle with an EMAT probe carriage designed for real-time testing of rails was designed by Tektrend International Inc. This vehicle includes hardware and software and resulted in the RailPro system. The RailPro system uses bulk and guided stress waves (TC, 2011). Another commercial system, based on EMAT technology and called Prism, has been developed by Wavesinsolids LLC and its technological partner TISEC Inc. Prism (Fig. 16.4) has a maximum inspection speed of 15 km/h and it is reported to be capable of detecting large transverse rail-head defects (i.e. equivalent to 20% of the cross-sectional area of the rail) (Papaelias, *et al.*, 2008; WINS, 2011). Although they are widely used in NDE, EMAT sensing might be challenging due to the lift-off distance and to the excessive noise generated when the inspection car is subjected to sudden acceleration.

Another non-contact system for rail assessment is based on the use of laser pulses and air-coupled transducers for the generation and detection of GUWs, respectively. This approach was developed mainly by two groups: Johns Hopkins University in collaboration with TTCI and Tecnogamma SPA (Kenderian *et al.*, 2002, 2003; Kenderian, 2002), and the University of California, San Diego (Lanza di Scalea *et al.*, 2006a, b; 2007a, b; Rizzo *et al.*, 2007; Rizzo *et al.*, 2010). Figure 16.5 shows one of the early prototypes developed at San Diego and tested in the field. The technique exploits the conversion of short duration laser pulses into mechanical waves. A thermoelastic stress is created when the laser energy density is not sufficiently high to induce material ablation or plasma formation on the surface of the object. In this regime, shear mechanical stresses are generated by the thermal expansion caused by the sharp increase of the surface's temperature. Ablation is instead created whenever the laser's power density is high, or when the surface of the irradiated surface is covered with a film of water

16.4 Photo of the EMAT-based Hy-rail system developed at Wavesinsolids (TISEC, 2011).

16.5 Early development of the UCSD laser/air-coupled system.

or gel. In this case the rapid vaporization (ablation) of the film at the surface, or the melting of a small portion of the medium's surface, induces high reaction pressures that can be considered similar to normal stress loading (Ni *et al.*, 2011). Irrespective of the thermoelastic or ablative generation, the guided waves are detected by means of air-coupled transducers oriented toward the incoming waves.

The drawback of any non-contact testing when compared to conventional contact testing is a reduced signal-to-noise ratio. The use of signal processing based on the discrete wavelet transform (DWT) helps in overcoming this problem, as demonstrated by a few groups (Toliyat *et al.*, 2003; McNamara and Lanza di Scalea, 2004; Lanza di Scalea *et al.*, 2005b; 2006a; Rizzo and Lanza di Scalea, 2007; Rizzo *et al.*, 2010).

16.3.6   Visual inspection

The simplest form of visual inspection is performed by experienced personnel walking along the rail track. The success of this practice is dependent on the technician's skills and experience. With the advent of high-speed cameras and image-processing algorithms, the research and development of machine vision for railroad track assessment has thrived. Such inspection technology uses video cameras, optical sensors, and custom-designed algorithms to analyze rail surfaces (Deutschl *et al.*, 2004; Singh *et al.*, 2006; Molina *et al.*, 2011). Machine-vision systems were developed, or are under current development, for both wayside and mobile inspection, including joint bars, surface defects in the rail, rail profile, ballast profile, track gauge, intermodal loading efficiency, railcar structural components, and railcar safety appliances (Molina *et al.*, 2011).

Machine vision based on the use of video cameras can be classified according to functionality into four major groups: (a) track inspection; (b) train inspection; (c) systems for maintenance and operation; and (d) passenger related systems (Papaelias *et al.*, 2008). The principle of any automated visual system is simple: as the train moves over the rail, a high-speed camera captures images, which are analyzed automatically using image-processing software. The analysis aims at identifying objects, and detecting and classifying defects. To detect small defects, the resolution of the captured video image needs to be high, and the blurring effects due to the movement of the camera have to be kept to a minimum. However, as the resolution of the image increases, so does the amount of data acquired, and hence more computational time is needed to complete the analysis. The result is that the speed of inspection can vary from 1 to 320 km/h depending on the type of inspection and the quality of resolution required. Visual-based inspection can measure the rail head profile and percentage of wear, rail gap, moving sleepers, absence of ballast, base plate condition in the absence of ballast,

pincers position, missing bolts, RCF, and rail corrugation (Papaelias *et al.*, 2008). One of the challenges in machine vision installed onto a testing vehicle is the need for periodic cleaning of optics due to the hard railway environment (dirt ejection, dust, etc.) and the need for proper illumination in every light condition (Oukhellou *et al.*, 2008).

Deutschl *et al.* (2004) developed a specific color image acquisition method to automatically detect 3D defects such as surfaces flakes, cracks, and grooves. The Visual Inspection System for Railway (VISyR) is a patented (Marino *et al.*, 2007; De Ruvo *et al.*, 2008) system composed of a line-scan camera and dedicated software. It uses an appropriate illumination set-up to reduce the effects of variable natural lighting conditions and make the system robust against changes in the natural illumination. To synchronize data acquisition, the line-scan camera is triggered by a wheel encoder to set the resolution along the motion direction at 3 mm (independently from the train velocity) and 1 mm along the perpendicular direction (head width). VISyR has the capability of detecting and tracking the rail head by means of a FPGA-based rail detection and tracking block (RD&TB) protocol, which automatically detects and tracks the rail-head center into a video sequence. RD&TB is based on the principal component analysis followed by a multilayer perceptron network classification for computing the coordinates of the center of the rail. The protocol has been designed to perform its task in 5.71 microseconds with an accuracy of 98.5%, allowing an on-the-fly analysis of a video sequence acquired up at 190 km/h (De Ruvo *et al.*, 2008). The VISyR system was tested on a benchmark video sequence of more than 3 000 000 lines, covering a rail network of about 9 km (De Ruvo *et al.*, 2008).

ENSCO Inc. owns a few track-inspection systems that mount high-resolution cameras and continuous laser to measure transverse rail profile, rail wear, and rail cant in real time at a one-foot sample rate up to the maximum speed of the car (ENSCO, 2011a). The joint bar (fishplate) inspection system (JBIS) is a patented vision-inspection technology used by ENSCO to perform automated crack detection on joint bars (fishplates), measure rail gap, and deliver joint bar inventory reports (ENSCO, 2011b). In Europe, non-contact optical-based systems are used by MERMEC group for rail profile and corrugation and for the inspection of track head and fish-plates (MERMEC, 2011)

Ren *et al.* (2011) developed an inspection car mounted with a laser-camera that projects light onto the track and measures gauge and longitudinal irregularity. The car contains a mechanism for absorbing vibration and impact, and the inspection system uses machine vision and the inertia system for measuring the track geometric parameters. The light is projected onto the track by means of a laser; the images of the rail profile are captured by a CCD camera, and the displacement between the vertex

of the rail and the camera are measured in real time through fast image processing.

   In the United States, researchers at the University of Illinois at Urbana Champaign (Sawadisavi *et al.*, 2008, 2009, Molina *et al.*, 2011a, b) proposed the use of a 60 Hz video camera for rail-surface characterization. An example of such processing is illustrated in Fig. 16.6. To characterize the periodic components of the rail, such as frog bolts or joint bar bolts, the middle portion of the video is converted into a panoramic mosaic and transformed in a block-wise manner into the Gabor frequency domain. The periodicity is then estimated in the Gabor frequency domain. Each block's height is identical to the height of the rail-web area (Fig. 16.6b), and each block's response is computed using an overlapping width with its right neighboring (Fig. 16.6c). This block-wise Gabor response is then processed as a one-dimensional signal (Fig. 16.6d). Spectral analysis is subsequently performed to find periodic components (Fig. 16.6e). The MUSIC algorithm is utilized to extract frequencies from a signal containing multiple superimposed signals of different frequencies.

   Another machine-vision technology is under development at IBM (Li *et al.*, 2011). The architecture has four major modules: data acquisition, track condition monitoring, defect severity analysis, and temporal condition analysis. It also has a long-term predictive assessment module. Specifically, the data acquisition module is in charge of capturing videos using multiple cameras mounted on a moving track-inspection vehicle. The image analysis aims at detecting rail components including tie plate, spike, spike hole, anchor, joint bar, and joint bar bolt.

16.3.7   Radiography

Radiographic techniques (specifically X-rays) are one of the few NDE methods that can examine the interior of an object and the only NDE method that works on all materials. An X-ray is a form of electromagnetic radiation with a wavelength in the range of 0.01 to 1 nm, corresponding to frequencies in the range of 3×10^{17} to 3×10^{19} Hz. X-rays have high electromagnetic energy. These rays tend to pass through objects that block visible light. This enables 'seeing' the interior of the material/structure under investigation. The amount of X-rays that pass through a material is dependent on the elemental composition, density, and thickness of the material, and the energy and amount of X-rays used for inspection (Shull, 2002). This method can detect cracks, flaws, and thickness reduction. X-ray-based NDT has the advantages of being accurate, inherently pictorial, adaptable to examine shapes and sizes, and sensitive to the discontinuity that causes a reasonable reduction of cross-section thickness. However, this method car-

(a)

Original image switch point bolts

(b)

Panoramic mosaic from the mid-rail area

(c)

Gabor frequency of the panoramic mosaic

(d)

One-dimensional signal from Gabor frequency of panoramic mosaic

(e)

Spectral analysis on one-dimensional signal
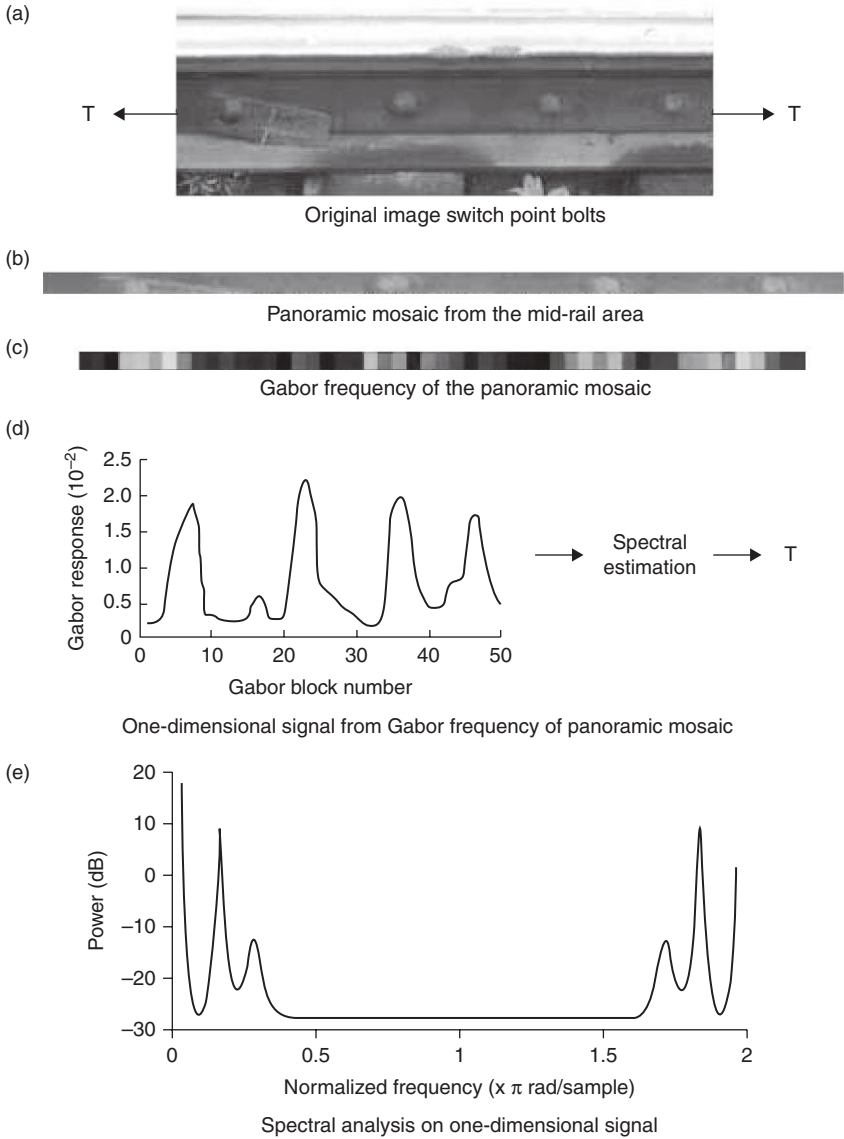
16.6 Turnout component recognition algorithm developed at the University of Illinois, Urbana Champaign (Molina *et al.*, 2011a).

ries the burden of safety hazard concerns, and it can be time consuming and expensive.

Radiographic inspection of rails can be carried out using either gamma or X-ray sources. Gamma rays were common in the past when digital

technology was not available, or was less developed. With the advent of portable digital X-ray detectors, the use of X-ray sources became more commonplace (Papaelias *et al.*, 2008). Radiographic methods are among the most efficient and effective NDE methods to localize and size an internal defect, although owing to the orientation of the detectors with respect to the rails, radiography is not very efficient in detecting transverse rail defects. However, health and safety drawbacks as well as portability of the equipment make the application of this method problematic. As such, X-rays or gamma rays are used only as a means of verification in places where defects have already been detected using other NDE techniques, or in rail areas, such as alumino-thermic welds, and switches and crossings, where inspection with other NDE methods is unreliable (Offereins and Mutton, 2001).

16.4   Structural health monitoring (SHM)

Impedance-based SHM, often termed the electromechanical impedance (EMI) method, is an active sensing method that exploits the relationship between the electrical impedance of a piezotransducers (PZT) and the mechanical impedance of the host structure to which the PZT is bonded or embedded. It has been shown that the value of the admittance, which is the inverse of the electrical impedance, is a function of the stiffness, mass, and damping of the host structure (Sun *et al.*, 1995; Park *et al.*, 2003), the length, width, thickness, orientation, and mass (Wheterhold *et al.*, 2003; Madhav and Soh, 2008) of the PZT, as well as the adhesive utilized to bond the PZT to the structure (Madhav and Soh, 2007). Therefore, any changes in the admittance are indicative of the presence of structural damage, provided the physical characteristics of the adhesive and the PZT remain constant.

   Researchers at Virginia Polytechnic Institute (USA) and Korea Advanced Institute of Science and Technology (South Korea) (Park *et al.*, 2007, 2008b) demonstrated the feasibility of a piezoelectric sensor-based health monitoring technique combined with a two-step support vector machine classifier for railroad track damage identification. The advantage of the proposed methodology is that the system is composed of two PZT patches that are used to exploit the principles of both impedance (frequency range 40–50 kHz) and guided wave propagation (input signal of 3-cycle 50 kHz sine function in a magnitude of 10 V) methods. The approach was proven for the detection of two kinds of damage in a railroad track, namely a hole-damage 0.5 cm in diameter at the web section and transverse cut damage 7.5 cm in length and 0.5 cm in depth at the head section. Two damage-sensitive features were separately extracted for each method: (a) feature 1: root mean square deviations (RMSD) of impedance signatures, and (b) feature II: sum of the square of the wavelet coefficients for maximum energy mode

of guided waves. From these two damage-sensitive features, a two-dimensional damage feature space was made and fed into a two-step support vector machine classifier. The experimental tests conducted in a laboratory on a 140 cm rail section showed a damage estimation rate of 96.67% (Park *et al.*, 2007).

To overcome the limitation of the cost associated with EMI equipment, and increase the damage-detection capability of the EMI method, Park *et al.* (2008) presented a paper on the use of outlier analysis applied to a macro-fiber composite (MFC) impedance-based wireless SHM system. The impedance-based SHM method has some limitations because the measured impedance data may have considerable deviations caused by environmental or operational condition changes, including temperature, humidity, external loadings, and MFC patch-bonding conditions. The outlier analysis is based on the Mahalanobis squared distance by taking root mean square deviation values of impedance signatures as a damage-sensitive feature vector. The approach was experimentally verified by detecting three types of simulated damage in the laboratory. Damage in the head and in the flange was simulated by a clamp of 692 g fixed at several locations. Damage in the web was simulated by magnets of 52 g at various locations.

Prior to being fully deployed in the field, this system should be proven robust to withstand the vibration generated by passing trains and extreme environmental changes. Moreover, the logistics associated with the requirements to connect the sensing system to appropriate hardware should be considered.

16.5   Systems for high-speed-rail inspection

Several countries possess a high-speed-rail network that connects the country's most important cities. Examples are the TGV in France, the Shinkasen in Japan, and the Korean Train Express (KTE) in South Korea, just to mention a few. Besides the inspection necessary to secure the safety and reliability of rail tracks, in high-speed rails one important parameter to be monitored is track irregularity. In fact, irregularities can affect the running behavior of the train, the riding comfort of passengers, and the safety of the running train (Kim *et al.*, 2009).

In Japan, since 1964, track condition monitoring has been accomplished by measuring car body acceleration of commercial trains of the Tokaido Shinkansen high-speed-rail network line using devices called RAIDARSS 2+ and RAIDARSS 3. These devices include three inertial measurement algorithms developed to measure the longitudinal level of tracks using a track condition monitoring device mounted on commercial Shinkansen trains. Either the RAIDARSS 2 + or the RAIDARSS 3 allows checking the track condition several times a day. If the measured accelerations exceed

the predetermined target maintenance values, the measured values and locations of them are automatically reported to the train control center and track maintenance depots (Naganuma *et al.*, 2008). The RAIDARSS systems measure vertical track irregularities using double integration of the axle-box acceleration. The principle is that inertial measurement can be attained by double integration of the acceleration. For example, the vertical position of a wheel can be found by using double integration of the axle-box acceleration. The result provides the longitudinal level due to the wheel being continuously in contact with the rail.

The KTE opened in 2004 and it is able to reach a maximum speed of 300 km/h. The Korea Railroad Research Institute (Kim and Park, 2008; Kim *et al* ., 2009) developed a rail measurement for the KTE to track rail irregularities. Irregularities being measured are longitudinal level irregularity, cross-level irregularity, gauge irregularity, alignment irregularity, and twist irregularity. The inspection system is composed of lasers, cameras, and inertial tools. The spatial position of a measuring sensor is determined by double integration of acceleration transducers. The lasers, cameras, and inertial transducers are installed into the frame of the bogie and they move together. The signal process of the system is divided into time domain and space domain.

16.6   Conclusions

Rail transportation is facing an unavoidable increase in traffic load and train speed. These factors, combined with the aging of existing infrastructures, demand more maintenance more often. Fortunately, there is an escalating progress in the areas of NDE and SHM techniques that aim at addressing the needs of the rail industry.

In this chapter a comprehensive overview of the most common systems for the NDE and SHM of railroad tracks is given. In particular, the chapter focused on the advancements reported in the last 5 years. Methods based on electromagnetic waves, stress waves, and visual inspection were discussed. Then, the sensing systems and the challenges associated with the rail health monitoring by means of the electromechanical impedance method were covered. Finally, a brief overview of the sensing systems used in Japan and Korea to monitor high-speed (>300 km/h) rails was given.

Although significant progress has been made in each of the aforementioned techniques, the author of this chapter believes that none of them can localize surface and internal damage in all rail components (head, web, and foot) with no false positives/negatives and at fast (>100 km/h) inspection speed yet. Current developments may point to one of the following directions: to advance existing solutions especially in lieu of the exponential improvement of the computer processors; to develop novel

methodologies/algorithms; and to integrate systems that combine, for instance, visual inspection ECs and GUWs.

Table 16.3 lists the NDE methodologies discussed here for the detection of damage in railroad tracks. Four parameters are considered: the size of the equipment, the simplicity of the concepts being used, the feasibility of using the technology in terms of permanent monitoring rather than periodic inspection, and cost. With the exception of the SHM implementation column, the other three columns refer to the methods only and do not include any consideration associated with the installation of the equipment into a rail car or train. The column relative to the simplicity refers to the ease of data interpretation and the level of complexity of the fundamental principles behind the NDE technique. Finally, the cost of guided wave testing is strongly dependent upon the technology used for the generation and detection of waves onto the rail. The usage of contact transducers is perhaps the most economical solution. On the other side, the use of pulsed laser and

Table 16.3 NDE methodologies for the detection of damage in railroad tracks

NDE methods	Size of Equipment	Simplicity	SHM implementation	Cost
ECs	L/M	M	L	M
Electromechanical impedance	L	M	H	L/M
Infrared thermography	M	M	L	M/H
MFL	M/H	M	L	M/H
Magnetic particles	L	M	M	L
Radiography	H	M	L	H
UT: bulk waves	M	M	L/M	M
UT: guided waves	L/M	M	M	L/M/H
UT: phased array	M	M	M	L/M
Vibration method	M	H	M/H	L/M
Visual inspection (by walking)	L	H	–	L
Visual inspection (with cameras)	M	H	L	M

H: high. M: medium. L: low.

Simplicity evaluates the level of complexity of the fundamental principles behind the NDT technique and the ease in interpreting the data.

SHM implementation indicates the feasibility of the technology as a permanent system for continuous structural monitoring. A low condition usually indicates that either the technology cannot be used for permanent monitoring or that it covers only a very limited segment of the rail.

The cost information is merely informative and may vary with different countries and with different service companies.

laser vibrometer for the generation and detection, respectively, of the propagating stress waves is the most expensive solution.

16.7   References

AGU (2010). http://blogs.agu.org/landslideblog/files/2010/11/10_10-Canterbury-22.jpg (Last date 4 accessed July 2011).

Aharoni R, Glikman E and Krug G (2002), 'A novel high-speed rail inspection system', *The e-Journal of Nondestructive Testing,* **7**(10), 9, http://www.ndt.net/article/ecndt02/156/156.htm.

Aharoni R and Glikman E (2002), 'A novel high-speed rail inspection system', *Proc. 8th ECNDT*, Barcelona, Spain.

Barke D and Chiu W K (2005), 'Structural health monitoring in the railway industry: a review', *International Journal of Structural Health Monitoring*, **4**(1), 81–94.

Bentoumi M, Aknin P and Bloch G (2003), 'On-line rail defect diagnosis with differential eddy current probes and specific detection processing', *European Physical Journal Applied Physics*, **23**(3), 227–233.

Bentoumi M, Bloch G, Aknin P and Millerioux G (2004a), 'Blind source separation for detection and classification of rail surface defects', *SAEM – Studies in Applied Electromagnetics and Mechanics*, **24**, 112–119.

Bentoumi M, Millerioux G, Bloch G, Oukhellou L and Aknin P (2004b), 'Classification de Defauts de Rail par SVM', *Proc. SCS'04*, Monastir, Tunisia, 242–245.

Bray D E and Stanley R K (1997), *Nondestructive Evaluation. A Tool in Design, Manufacturing, and Service*, Boca Raton, CRC press.

Bredif P, Plu J, Pouligny P and Poidevin C (2008), 'Phased-array method for the UT-inspection of French rail. repairs', *Rev. QNDE*, 27, ed. by D. O. Thompson and D. E. Chimenti, 762–769.

Cannon D F, Edel K-O, Grassie S L and Sawley K (2003), 'Rail defects: an overview', *Fatigue Engineering Materials and Structures*, **26**, 865–887.

Care R, Clark S, Dembosky M and Doherty A (2005), 'Why rails crack, Gauge corner cracking on the British network: Investigation', *The Arup Journal*, **3**, 37–41.

Care R, Clark S, Dembosky M and Doherty A (2006), 'Why rails crack, gauge corner cracking on the British network: analysis', *The Arup Journal*, **1**, 16–19.

Clark R (2004), 'Rail flaw detection: overview and needs for future developments', *NDT&E International*, **37**, 111–118.

Clark R and Singh S (2003), 'The inspection of thermite welds in railroad rail – a perennial problem', *Insight*, **45**, 387–393.

De Ruvo P, De Ruvo G, Distante A, Nitti M, Stella E and Marino F (2008), 'A visual inspection system for rail detection & tracking in real time railway maintenance', *Open Cybernetics and Systemics Journal*, **2**, 57–67.

Deutschl E, Gasser C, Niel A and Werschonig J (2004), 'Defect detection on rail surfaces by a vision based system', *Intelligent Vehicles Symposium IEEE*, **14–17**, 507–511.

Dixon S, Edwards R S and Jian X (2004), 'Inspection of the railtrack head surfaces using electromagnetic acoustic transducers', *Insight*, **46**, 326–330.

Edwards R S, Jian X and Dixon S (2006a), 'Rail defect detection using ultrasonic surface waves', *Rev. QNDE*, Eds D. O. Thompson and D. E. Chimenti, 25, 1601–1608.

Edwards R S, Dixon S and Jian X (2006b), 'Characterisation of defect in the railhead using ultrasonic surface waves', *NDT&E International*, **39**, 468–475.

Edwards R S, Dixon S and Jian X (2006c), 'Depth gauging of defects using low frequency wideband Rayleigh waves', *Ultrasonics*, **44**, 93–98.

ENSCO (2011a) http://www.ensco.com/products-services/rail-technologies/track-inspection-systems/rail-surface-wear-condition.htm.

ENSCO (2011b) http://www.ensco.com/products-services/rail-technologies/track-inspection-systems/track-imaging-systems.htm.

Fan Y, Dixon S, Edwards R S and Jian X (2007), 'Ultrasonic surface wave propagation and interaction with surface defects on rail track head', *NDT&E International*, **40**, 471–477.

Federal Railroad Administration (2011), *FRA Guide for Preparing Accident/Incident Reports,* Appendix C.

Federal Railroad Administration (2012), http://safetydata.fra.dot.gov/OfficeofSafety/.

Garcia G and Zhang J (2006), 'Application of ultrasonic phased arrays for rail flaw inspection', *TTCI Report for the US Department of Transportation*.

Gharaibeh Y, Sanderson R, Mudge P, Ennaceur C and Balachandran W (2011), 'Investigation of the behaviour of selected ultrasonic guided wave modes to inspect rails for long-range testing and monitoring', *Proc. of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **225**, 311–324.

Greene R J, Yates J R and Patterson E A (2007), 'Crack detection in rail using infrared methods', *Optical Engineering*, **46**(5), 051013.

Hayashi T, Song W-J and Rose J L (2003), 'Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example', *Ultrasonics*, **41**, 175–183.

Hesse D and Cawley P (2006), 'Surface wave modes in rails', *The Journal of the Acoustical Society of America*, **120**, 733–740.

Howitt M (2002), 'Bombardier brings ACFM into the rail industry', *Insight*, **44**(6), 372–382.

Junger M, Thomas H M, Krull R and Rühe S (2004), 'The potential of eddy current technology regarding railroad inspection and its implementation', *Proc. 16th World Conf. on NDT*, Montréal, Canada.

Kenderian S, Djordjevic B and Green R E (2002), 'Laser based and air coupled ultrasound as non-contact and remote techniques for testing of railroad tracks', *Materials Evaluation*, **60**, 65–70.

Kenderian S (2002), *Advanced Ultrasonic Techniques to Determine the Structural Integrity of Rail Steel*, PhD Thesis, The John Hopkins University, Baltimore.

Kenderian S, Cerniglia D, Djordjevic B, Garcia G, Morgan R, Sun J and Snell M (2003), 'Rail track field testing using laser/air hybrid ultrasonic technique', *Materials Evaluation,* **61**, 1129–1133.

Kim S S and Park C S (2008), 'High speed rail measurement system of HSR-350x', *Journal of Korean Society for Railway*, Korea, **11**(2), 115–119.

Kim S-S, Park C, Kim Y-G and Park C (2009), 'Parameter characteristics of rail inspection measurement system of HSR-350x', *Journal of Mechanical Science and Technology* **23**, 1019–1022.

Lanza di Scalea F, Bartoli I, Rizzo P and Fateh M (2005a), 'High-speed defect detection in rails by non-contact guided ultrasonic testing', *Transportation Research Record, Journal of the TRB*, **1916**, 66–77.

Lanza di Scalea F, Rizzo P, Coccia S, Bartoli I, Fateh M, Viola E and Pascale G (2005b), 'Non-contact ultrasonic inspection of rails and signal processing for automatic defect detection and classification', *Insight*, **47**(6), 346–353.

Lanza di Scalea F, Bartoli I, Rizzo P and Fateh M (2006a), 'High-speed defect detection in rails by non-contact guided ultrasonic testing', *Journal of Transportation Research Board*, *Transportation Research Record*, **1961**, 66–77.

Lanza di Scalea F, Rizzo P, Coccia, S., Bartoli I and Fateh M (2006b), 'Laser/air-coupled hybrid detection in rail tracks: status of FRA prototype development at UC San Diego', *Journal of Transportation and Research Board*, *Transportation Research Record*, **1943**, 57–64.

Lanza di Scalea F (2007), 'Ultrasonic testing in the railroad industry', *Chapter 15, Part 2 in Nondestructive Testing Handbook, Ultrasonic Testing*, Vol. 7, P.O. Moore Editor, American Society for Nondestructive Testing, 535–540.

Lauer F, Bentoumi M, Bloch G, Millerioux G and Aknin P (2004), 'Ho-Kashyap with early stopping vs soft margin SVM for linear classifiers – an application', *Lecture Notes in Computer Science*, **3173**, 524–530.

Li Y, Otto C, Haas N, Fujiki Y and Pankanti S (2011), 'Component-based track inspection using machine-vision technology', *ICMR'11 Proc. 1st ACM International Conference on Multimedia Retrieval*, Trento, Italy.

LIZARD (2011a), http://www.lizard.co.uk/railinspectionsystem.html (last date 25 accessed June 2011).

LIZARD (2011b), http://www.lizard.co.uk/howitworks.html (last date 25 accessed June 2011).

Lugg M and Topp D (2006), 'Recent developments and applications of the ACFM inspection method and ACSM stress measurement method', *Proc. ECNDT*, Berlin.

Madhav A V G and Soh C K (2008), 'Multiplexing and uniplexing of PZT transducers for structural health monitoring', *Journal of Intelligent Material Systems and Structures,* **19**(4), 457–467.

Madhav A V G and Soh C K (2007), 'An electromechanical impedance model of piezoceramic transducer-structure in presence of thick adhesive bonding', *Smart Materials and Structures,* **16**(3), 673–686.

Marino F, Distante A, Mazzeo P L and Stella E (2007), 'A real time visual inspection system for railway maintenance: automatic hexagonal headed bolts detection', *IEEE Transactions on Systems, Man and Cybernetics-Part C*, **37**(3), 418–428.

McNamara J and Lanza di Scalea F (2004), 'Improvements in non-contact ultrasonic testing of rails by the discrete wavelet transform', *Materials Evaluation*, **62**, 365–372.

Mehel-Saidi Z, Bloch G and Aknin P (2008), 'A subspace method for detection and classification of rail defects', *16th European Signal Processing Conference, EUSIPCO-2008*, Lausanne.

MERMEC (2011), http://www.mermecgroup.com/diagnostics/track-inspection/64/1/track-head-inspection.php.

Mix P (2005), *Introduction to Nondestructive Testing, A Training Guide*, 2nd Edn., Hoboken, New Jersey, John Wiley and Sons.

Molina L F, Resendiz E, Edwards J R, Hart J M, Barkan C P L and Ahuja N (2011a), 'Condition monitoring of railway turnouts and other track components using machine vision', *TRB* 11–1442.

Molina L F, Edwards J R and Barkan C P L (2011b) 'Emerging condition monitoring technologies for railway track components and special trackwork', 2011 ASME/ASCE/IEEE Joint Rail Conference, JRC2011, Pueblo, Colorado, USA.

Naganuma Y, Kobayashi M, Nakagawa M and Okumura T (2008), 'Condition monitoring of Shinkansen tracks using commercial trains', *4th IET Intl. Conf. on Railway Condition Monitoring*, Issue, 1–6.

Ni X, Rizzo P and Daraio C (2011), 'Laser-based excitation of nonlinear solitary waves in a chain of particles', *Physical Review E*, **84**, 026601, 5 pages.

Olympus NDT (2011), *Advances in Phased Array Ultrasonic Technology Applications*, http://www.olympus-ims.com/data/File/advances_book/Applications_Cha1.en.pdf.

Offereins G A and Mutton P J (2001), 'Recent experiences with the performance of aluminothermic rail welds under high axle loads', *Proc. 13th Intl. Rail Track Conference*, Canberra, Australia, 2001.

Oukhellou L, Aknin P and Perrin J-P (1999), 'Dedicated sensor and classifier of rail head defects', *Control Engineering Practice*, **7**(1), 57–61.

Oukhellou L, Côme E, Bouillaut L and Aknin P (2008), 'Combined use of sensor data and structural knowledge processed by Bayesian network: application to a railway diagnosis aid scheme', *Transportation Research Part C*, **16**, 755–767.

Papaelias M Ph, Lugg M, Smith M, Roberts C and Davis C L (2007), 'Detection and quantification of rolling contact fatigue cracks in rails using ACFM technology', *Proc. BINDT 2007 Conference*, Glasgow, UK.

Papaelias M Ph, Roberts C and Davis C L (2008), 'A review on non-destructive evaluation of rails: state-of-the-art and future development', *Proc. IMechE Part F: J. Rail and Rapid Transit*, 222, 367–384.

Papaelias M Ph, Lugg M, Roberts C and Davis C L (2009), 'High-speed inspection of rails using ACFM techniques', *NDT&E International*, **42**, 328–335.

Park G, Sohn H, Farrar C R and Inman D J (2003), 'Overview of piezoelectric impedance-based health monitoring and path forward', *The Shock Vibration Digest*, **35**(6), 451–463.

Park S, Lee J-J, Yun C-B and Inman D J (2007), 'A built-in active sensing system-based structural health monitoring technique using statistical pattern recognition', *Journal Of Mechanical Science and Technology*, **21**, 896–902.

Park S, Inman D J and Yun C-B (2008a), 'An outlier analysis of MFC-based impedance sensing data for wireless structural health monitoring of railroad tracks', *Engineering Structures,* **30**, 2792–2799.

Park S, Inman D J, Lee J-J and Yun C-B (2008b), 'Piezoelectric sensor-based health monitoring of railroad tracks using a two-step support vector machine classifier', *ASCE Journal of Infrastructure Systems*, **14**(1), 80–88.

Pohl R, Erhard A, Montag H-J, Thomas H-M and Wüstenberg H (2004), 'NDT techniques for railroad wheel and gauge corner inspection', *NDT&E International*, **37**, 89–94.

Pohl R, Krull R and Meierhofer R (2006), 'A new eddy current instrument in a grinding train', *Proc. ECNDT 2006*, Berlin, Germany.

Ren S, Gu S, Xu G, Gao Z and Feng Q (2011), 'A new track inspection car based on a laser camera system', *Chinese Optics Letters*, **9**(3), 031202–1.

Rizzo P, Bartoli I, Cammarata M and Coccia S (2007), 'Digital signal processing for rail monitoring by ultrasonic guided waves', *Insight*, **49**(6), 327–332.

Rizzo P (2010), 'Water and wastewater pipe health monitoring: a review', *Advances in Civil Engineering*, Article ID 818597, 13 pages.

Rizzo P and Lanza di Scalea F (2007), 'Wavelet-based unsupervised and supervised learning algorithms for ultrasonic structural monitoring of waveguides', in *Progress in Smart Materials and Structures Research*, Hauppauge, NY: Nova Science Publishers, Inc., Chapter 8, 227–290.

Rizzo P, Cammarata M, Bartoli I, Lanza di Scalea F, Salamone S, Coccia S and Phillips R (2010), 'Ultrasonic guided waves-based monitoring of rail head: laboratory and field tests', *Advances in Civil Engineering*, Article ID 291293, 13 pages.

Rose J L, Avioli M J, Mudge P and Sanderson R (2004), 'Guided wave inspection potential of defects in rail', *NDT&E International*, **37**, 153–161.

Rose J L, Lee C M, Hay T R, Cho Y and Park I K (2006), 'Rail inspection with guided waves', *Proc. 12th Asia-Pacific Conference on NDT*, Auckland, New Zealand.

Sawadisavi S, Edwards J R, Hart J M, Resendiz E, Barkan, C P L and Ahuja N (2008), 'Machine-vision inspection of railroad track', *2008 AREMA Conference Proceedings*, *American Railway and Maitenance of Way Association (AREMA)*, Landover, Maryland.

Sawadisavi S, Edwards J R, Resendiz E, Hart J M, Barkan, C P L and Ahuja N (2009), 'Machine-vision inspection of railroad track', *Proceedings of the TRB 88th Annual Meeting*, Washington, DC, January 2009.

Shull P J (2002), *Nondestructive Evaluation: Theory, Techniques, and Applications*, New York, Marcel Dekker.

Singh M, Singh S, Jaiswal J and Hempshall J (2006), 'Autonomous rail track inspection using vision based system', *IEEE Intl. Conf. on Computational Intelligence for Homeland Security and Personal Safety*, 16–17, 56–59.

Sperry Rail Service (1989), *Rail Defect Manual*, Banbury, Connecticut.

Sun F P, Chaudry Z, Rogers C A, Majmundar M and Liang C (1995), 'Automated real-time structure health monitoring via signature pattern recognition', *Proceedings of SPIE*, **2443**, 236–247.

Thomas H-M, Junger M, Hintze H, Krull R and Rühe S (2000), 'Pioneering inspection of railroad rails with eddy currents', *Proc. 15th World Conf. on NDT*, Rome, Italy.

Thomas H-M, Heckel T and Hanspach G (2006), 'Advantage of a combined ultrasonic and eddy current examination for railway inspection trains', *Proc. ECNDT*, Berlin, Germany.

Thomas H-M, Dey A and Heyder R (2010), 'Eddy current test method for early detection of rolling contact fatigue (RCF) in rails', *Insight*, **52**(7), 361–365.

Toliyat H A, Abbaszadeh K, Rahimian M M and Olson L E (2003), 'Rail defect diagnosis using wavelet packet decomposition', *IEEE Transactions on Industrial Applications*, **39**(5), 1454–1461.

TC (2011), 'Development of a mobile inspection system for rail integrity assessment (TP 13611E), Available at http://www.tc.gc.ca/eng/innovation/tdc-summary-13600–13611e-162.htm (Accessed 2 July 2011).

TISEC (2011), http://www.tisec.com/SI/rail_flaw.php.

TSB (2005a), http://www.tsb.gc.ca/eng/rapports-reports/rail/2005/r05e0059/r05e-0059_photo_1.jpg; (Accessed 2 July 2011).

TSB (2005b), http://www.tsb.gc.ca/eng/rapports-reports/rail/2005/r05e0059/r05e-0059_photo_2.jpg; (Accessed 2 July 2011).

TSC (2011), http://www.tscinspectionsystems.co.uk/ (Accessed 2 July 2011).

Utrata D (2002), 'Exploring enhanced rail flaw detection using ultrasonic phased array inspection', *Review QNDE*, Eds D. O. Thompson and D. E. Chimenti, 21, 1813–1818.

Utrata D and Clark R (2003), 'Groundwork for rail flaw detection using ultrasonic phased array inspection', *Review QNDE*, Eds D. O. Thompson and D. E. Chimenti, 22, 799–805.

Vandone A, Rizzo P and Vanali M (2012), 'Two-stage algorithm for the analysis of infrared images', *Research in Nondestructive Evaluation*, **23**(2), 69–88.

Verysit (2011), http://veryst.com/Train_Derailment_Broken_Rail.html (Last date accessed 4 July 2011).

Wilson J, Yun T-G, Zainal A I, Suixian Y and Darryl A (2010), 'Modeling and evaluation of eddy current stimulated thermography', *Nondestructive Testing and Evaluation*, **25**(3), 205–218.

Wilson J, Tian G, Mukriz I and Almond D (2011), 'PEC thermography for imaging multiple cracks from rolling contact fatigue', *NDT&E International*, **44**, 505–512.

WINS (2011), http://www.wins-ndt.com/ultrasonic_rail_flaw_detection.php (accessed 2 July 2011).

Zerbst U, Madler K and Hintze H (2005), 'Fracture mechanics in railway applications: an overview', *Engineering Fracture Mechanics*, **72**, 163–194.

Sensing solutions for assessing and monitoring underwater systems

P. RIZZO, University of Pittsburgh, USA

Abstract: Infrastructures containing or surrounded by water are an important aspect of modern civilization. Most of them operate under harsh environmental conditions. The periodic inspection or permanent monitoring of wet structures is therefore necessary. This chapter reviews the most common nondestructive evaluation (NDE) and structural health monitoring (SHM) methods for underwater structures. First, the structures and the challenges associated with their inspection are introduced. Then, the NDE techniques of visual inspection, magnetic particles, ultrasonic testing, sonars, radiography, and electromagnetism are illustrated. Finally, the SHM approach to acoustic emission, fiber optics, vibration, and flooded member detection are discussed. The general principles, advantages, and limitations of each are described.

Key words: underwater structures, nondestructive evaluation (NDE), structural health monitoring (SHM).

17.1   Introduction

Subsea infrastructure systems and, broadly speaking, any structure containing or surrounded by water represents an important aspect of modern civilization. These structures can be arbitrarily clustered into five major groups: offshore structures that serve to extract oil or natural gas from the seabed; communication cables; pipelines that transport oil and natural gas to processing facilities; naval vessels (ships and submarines); and waterfront facilities such as piers, retaining walls, and docks. With the advent of renewable energy technology a new type of offshore structure is represented by the offshore wind farms.

The offshore oil and gas industry has a major investment in subsea installations. These structures are vulnerable to internal and external corrosion, manufacturing flaws, ship anchors, and seismic movements of the sea bed. The effect of one or more of these factors, together with the harsh environment in which these installations operate, can shorten the lifetime of component parts and lead to failures that can result in leaks of oil and hydraulic

525

fluids into the sea (McStay *et al.*, 2005). Moreover, the lifetime of most of the early structures was expected to be 25–30 years. However, improved methods of oil recovery, including horizontal drilling, have led to much longer lifetimes, with the consequent need for structures to be kept in operation longer than expected (Baker and Descamps, 1999). As such, the periodic inspection or permanent monitoring of offshore structures is necessary to avert risks for personnel and the environment, and to minimize economic losses. For instance, in the last decade it has been reported that in the Gulf of Mexico alone the number of structural inspections per year was nearly 5000 (Laurendine, 2002; Mijarez, 2006). In fact, the inspection of offshore structures has been a requirement for a considerable time and there are regular procedures in place stipulating that offshore operators need to obtain certification for structural integrity monitoring.

The inspection of underwater structures is challenging, due to tidal and weather conditions, marine growth, visibility, and accessibility. Marine organisms, for instance, form a layer of organic material on the structure itself. In wood structures such as piles used in waterfront installations, these organisms can also be the direct cause of damage (Lopez-Anido *et al.*, 2004). The organisms that cause damage can be divided into fungi and marine borers. Their basic physical and biological characteristics, as well as the type of damage that they cause in marine wood piles, are detailed in Lopez-Anido *et al.* (2004).

Irrespective of the engineering system being assessed, the inspection of underwater structures is carried out applying nondestructive evaluation (NDE) techniques that closely resemble topside NDE tools i.e. tools that are direct applications or modifications of techniques employed in dry conditions (Goldberg, 1996). The modifications usually consist of waterproofing and adaptation of the instruments to the subsea environment (Saebjornsen and Forli, 1983; Mijarez, 2006). According to Hayward (1993), subsea inspection can be classified into three levels:

  Level 1: inspection is performed visually on most or all parts of the structure without removing the marine growth in and around the site.
  Level 2: inspection is conducted visually close to selected areas that are cleaned *a priori*.
  Level 3: detailed assessment of predetermined sites is performed using one or more advanced NDE methods.

Most Level 3 inspection is performed by divers, sometimes in conjunction with a remotely operated vehicle (ROV). When the examination is performed by divers it might be subject to operator error and speed constraints. Accordingly, practical subsea inspection systems must incorporate technically reliable equipment, requiring minimum calibration and maintenance (Hayward, 1993).

The objective of this chapter is to provide a comprehensive presentation of the nondestructive technologies and sensing systems that have been developed for the assessment of underwater structures. Some of these technologies were proved and are used successfully in routine inspections; some of them are at their infancy in terms of their applications for underwater structures. In particular, we will focus on metallic and concrete structures. For inspection/monitoring of timber or masonry structures, the interested reader is referred to Kelly (1999) for a preliminary insight. The present chapter is structured as follows. Section 17.2 briefly describes the characteristics of offshore structures and the challenges associated with their inspection. Section 17.3 describes the fundamental principles and the application of the NDE methods used for the periodic inspection of immersed structures and waterfronts. Section 17.4 illustrates the sensing systems proposed for permanent structural health monitoring (SHM) of underwater structures. Section 17.5 ends the chapter with some concluding remarks and a comprehensive table that associates the NDE methods described in the chapter to their corresponding fields of application.

17.2   Underwater structures: types and challenges

This section briefly describes the characteristics of offshore structures and the challenges associated with their inspection.

17.2.1   Offshore platforms

In 2002 there were around 7400 offshore platforms around the world, 54% of which were located in the Gulf of Mexico (Laurendine, 2002). These installations were constructed for diverse applications and exist in many types, for instance compliant tower, tension leg, sea star, and fixed structures. Fixed structures consist of a superstructure and a substructure. The superstructure is a formation fixed on the jacket substructure and is composed of a series of modules housing drilling and production equipment, a helicopter pad, and living accommodation. The substructure is either a steel tubular jacket or a pre-stressed concrete structure. Tubular jacket substructures are composed of tubular steel members connected at their ends to form tubular joints and crossbeam members that are normally air-filled steel tubes used to strengthen and support their main legs. The joints are formed by welding the outer end of one tube onto the undistributed external side of the other (El-Reedy, 2002; Mijarez, 2006).

Marine riser systems are one of the most important components of an offshore structure, as they link the seabed pipeline with the topside processing equipment. The design of the risers dictates or limits how they can be inspected. Risers can be made of different materials, such as carbon steel,

stainless steel, titanium, composite, and have different geometries, i.e. with/without coating, with/without insulation, with paint or sprayed aluminum coating. As reported by Lozev *et al*. (2005), corrosion is the major contributor to accidents. This is due to the persistent wetting and drying in the splash zone combined with defects in the coatings. Offshore pipeline failure statistics collected since the early seventies show that 55% of all Gulf of Mexico incidents reported to the U.S. Department of the Interior until 2002 had been caused by corrosion damage (Lozev *et al*., 2005). Of this 55%, external corrosion contributed to 70% of damage cases.

Jacket-type steel structures are the most common type of offshore structures. They consist of tubular members welded together to form a steel space frame extending from the seabed to just above the sea surface, in order to provide a template for pile foundations (Aghakouchak and Stiemer, 2001). Tubular joints are prone to fatigue damage, due to the geometry of the connections and the high stress concentration produced by the welds, and due also to the loading conditions exerted by the waves and the wind. Fatigue damage has been one of the most common modes of in-service failure of jacket structures. Field observations and laboratory tests have shown that in tubular connections, a fatigue crack starts at the weld toe and gradually propagates around the intersection and through the tubular wall (Aghakouchak and Stiemer, 2001).

17.2.2   Piles and pier structures

Piles and waterfront facilities, such as piers, wharves, and harbors are usually subjected to four zones of influence, as sketched in Fig. 17.1. The submerged zone is always surrounded by water while the atmospheric zone stays dry. The tidal zone is the area that is above water at low tide and under water at high tide while the splash zone is affected by the wave motion. Waterfront facilities can be made of different materials including concrete, steel, wood, and masonry. Concrete is used in concrete piles to support piers and wharves, decks, and retaining walls. Concrete is also used in pavements, bridge foundations, boat loadings and ramps, breakwaters, undersea cable and pipeline stabilization, and offshore structures (Kelly, 1999). Factors such as corrosion, freeze–thaw cycles, chemical deterioration, and overload can cause premature deterioration of concrete in these harsh environments. Signs of deterioration are cracking, loss of cross-section, spalling, and softening (Kelly, 1999). The damage is generally most severe in the splash and tidal zones, but does occur in all zones. Disintegration and spalling can expose reinforcing steel. Concrete deterioration in the marine environment usually starts as a result of poor construction techniques and inadequate inspection and quality control during construction (Kelly, 1999).
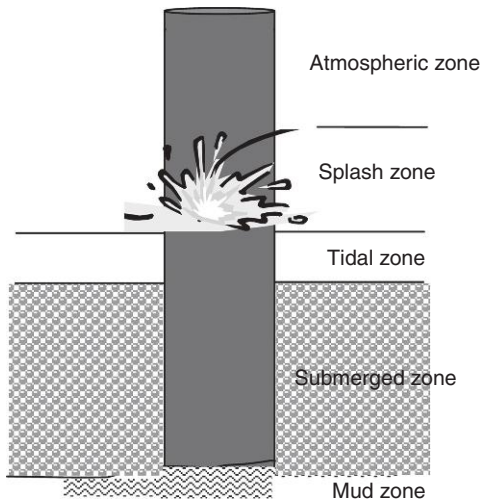
17.1 Various zones of influence on piling. (*Source*: Adapted from Kelly, 1999.)

17.2.3   Submerged pipelines

Underwater pipes are designed to bring natural gas to process terminals or oil to refineries. They are usually made of steel surrounded by an anti-corrosive coating, an insulating layer, a polyethylene pipe, and finally covered by a layer of shotcrete (Wang and Wang, 2010). When the natural gas needs to be transported at cryogenic temperatures ($-160°C$), the subsea pipeline is made of a tripe triple-wall pipe constituted of: (1) a 36% NiFe inner pipe, commonly known as Invar, which allows a significant reduction of the thermal contraction of the inner pipe resulting from the temperature differential between the $-160°C$ operating temperature and the ambient temperature; and (2) two other pipes (intermediate and outer pipes) that protect and insulate the inner pipe (ITP 2011).

   One of the biggest challenges in underwater pipe assessment arises when trying to locate holes that may be only a few millimeters in diameter somewhere along possibly hundreds of kilometers of pipeline, especially if the pipeline is either partially or totally buried in sediment. Traditional techniques for leak inspection consist of filling the pipeline with a solution of water and a chemically, or optically detectable compound, such as a fluorescent element. The pipeline is then followed by an ROV equipped with either a dedicated sensor or a video camera (Barbagelata and Barbagelata, 2002). When leak occurs, the oil or the natural gas transits from the inside pressure to the lower external pressure. The sudden expansion of the fluid mass creates a broadband acoustic source that can be detected by hydrophones

(Zingaretti and Zanoli, 1998; Tian, 2008). Unfortunately, in an offshore environment, the background noise may mask such acoustic waves.

Permanent leak detection systems monitor changes in the pipe's pressure. These systems require pressure sensors to be distributed along the pipeline, adequately powered, and properly maintained. This approach is able to detect substantial leaks only within 100 m (Theakston, 2004; McStay *et al.*, 2005). To identify leak locations in pipelines and risers, most oil company operators run annual inspections of their underwater installations as part of a routine maintenance procedure. The two main techniques employed are based on acoustic and fluorescence measurements. In the acoustic technique, one or more hydrophones are coupled to appropriate signal processing to discriminate between genuine acoustic leak and ambient noise and to minimize false positives or false negatives. If the sensing systems are embedded on an ROV, the signal processing must be also able to discard the noise associated with the thrusters and other ROV devices (McStay *et al.*, 2005). A commercial system based on pipeline acoustic leak detectors was developed in Italy (Barbagelata and Barbagelata, 2002, 2004). It is composed of a hydrophone array, a preamplifier, and a cable driver. The unit may be used in a variety of ways, such as on a vessel-towed 'fish', an ROV installation, or in handled mode by a diver.

As stated in the introduction, there are three levels of underwater inspection (Kelly, 1999).

- *Level 1* – General visual inspection. This serves to detect obvious major damage or deterioration due to overstress, severe corrosion, or extensive biological growth and attack. This type of inspection does not involve cleaning of any structural element, and it provides initial input for an inspection strategy. When performed by divers, the Level 1 effort is a 'swim-by' overview and may include tactile observation (depending on water clarity).
- *Level 2* – Detailed visual inspection. It consists of a complete and detailed visual inspection of select components or subcomponents. It is usually targeted at critical areas of structures or deteriorated areas that may be hidden by surface fouling. The structure is cleaned before or during the inspection. As cleaning is time consuming, Level 2 is generally restricted to areas that are critical or that may be typical of the entire structure.
- *Level 3* – Detailed NDE. This is the most detailed inspection and often requires the use of one of the NDE techniques discussed in the next section. The training, cleaning, and testing requirements will vary depending on the type of damage/defect to be investigated and the type of inspection equipment to be used. In some cases, Level 3 inspections are partially destructive, such as sample coring in wood or concrete,

material sampling, and *in situ* surface hardness might be required. Level III inspections demand considerably more experience and training than the other two levels (Kelly, 1999).

17.3   Nondestructive evaluation (NDE) techniques

In this section the assessment of underwater structures is discussed from the point of view of periodic inspections conducted by divers or by means of ROVs.

17.3.1   Visual inspection

Visual inspection is the most common underwater inspection method (Goldberg, 1996; Na and Kundu, 2002). It often consists of cameras operated by divers or mounted in submersibles or ROVs. The acquired images can be coupled to appropriate image processing algorithms to extract damage-sensitive features and classify defects. One of the challenges associated with visual inspection is the clarity of sea water and the presence of organic material on the structure to be inspected. The inspection of structures without marine cleaning is suitable for detection of gross damage such as large deformations, severed connections, missing members, and debris or gaps in surface coatings or fragmented coatings. When the structure is cleaned then mechanical damage, obvious cracks, corrosion dents, and deformations can be detected (Busby, 1978; Mijarez, 2006). Any visual inspection method carried out without image processing is subjected to the experience and skills of the operators and the inspectors' perception of the defects.

One type of visual inspection is conducted by means of fluorescent dyes. McStay *et al*. (2005) developed a 2000 m depth-rated subsea sensor system, called Fluorotrak, which can detect fluorescein, chemicals, and hydraulic fluids leaking from underwater structures. The system, which is schematically shown in Fig. 17.2, comprises excitation and detection systems as separate, self-contained units. Each unit contains a watertight corrosion-resistant housing with an optical window at one end. An ultrabright light-emitting diode (LED) array serves as the optical source. This array emits light with wavelength between 430 and 500 nm and provides an output power of up to 150 mW. The light is projected in a divergent manner using a lens system through the optical window and into the water column. The resulting fluorescence is collected and digital signal processing is utilized to extract the intensity. The ability to stand off from subsea structures, while rapidly detecting chemicals, makes the system suited to subsea leak inspections.

The use of fluorescent dyes is one of the most common methods used by oil and gas companies to detect leaks. The dye is added to the pipeline or
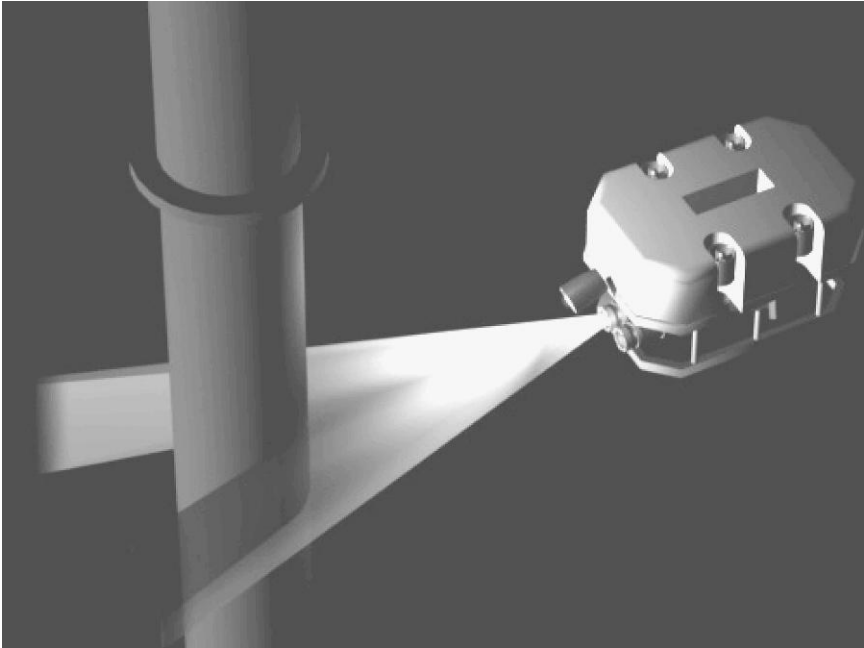
17.2 The Fluorotrak leak detection system. (*Source*: McStay *et al.*, 2005 Fig. 3.)

the riser, either offline or simply by injecting the dye into the existing fluid stream during normal working operation. If the pipeline transports oil, the intrinsic fluorescence of the oil can be used to aid in the detection of leaks (McStay *et al.*, 2005). Other methods for detecting the leaking fluorescent fluid have been developed. These include simple visual observation either by a diver or a camera. Black light systems have been employed to enhance such visual observation. In general these techniques are useful for very large leaks only in which a significant plume is observable. To enhance the detection of the dye and to remove the reliance on the operator's skills, subsea fluorometers can be used. A fluorometer is able to measure the intensity and the wavelength distribution of the fluorescent (So and Dong, 2001). The greater detection sensitivity of such fluorometers (typically ppm) also offers the potential to reduce the volume of dye needed to effect leak detection. In general, existing fluorometers employed for leak detection comprise an excitation source and photodetector with appropriate spectral filters, all housed in a watertight housing (Abu-Zeid *et al.*, 1987; Bartz *et al.*, 1988; McStay *et al.*, 1995; McStay *et al.*, 2007). Laser sources have also been used in marine fluorometers; however, expense, power consumption, and health and safety concerns have restricted their use (Camagni *et al.*, 1988; Yulong *et al.*, 1990).

The biggest drawback of dye-based leak detection is the localized nature of the detection. In practice this feature requires the instruments to be placed in the plume of the fluorescent fluid emanating from a leak. Often this means that the instruments, and hence the supporting ROV, have to be brought close to the subsea structure. The point sensing nature of these instruments can also cause problems when trying to locate leaks rapidly in large volumes of water. To overcome this problem, and to obtain a more representative measurement of fluorescent material in a body of water, optical fiber array systems have been developed (Russell *et al.*, 2001; Huang *et al.*, 2007). Although providing a multitude of sensing locations, these systems still need to have supporting structures, to properly install them from the sensing head to the optoelectronics unit.

Visual inspection is also commonly used to assess waterfront facilities by skilled divers. The inspection frequency depends upon whether the inspection is surface or underwater, and upon the expected rate of deterioration and damage. A typical example requiring more frequent inspection is any area experiencing damage by ships' berthing that results in advanced deterioration to both fender and structural piling (Kelly, 1999). Based on previous studies and reports, Kelly (1999) recommended the following:

- All superstructure and piling/sheet piling above the waterline should be inspected annually.
- Concrete/steel structural members at the splash/tidal zones and downward should be inspected at least every 6 years. As deterioration is discovered, the level of inspection and frequency needs to be increased accordingly.
- Timber members should be inspected at least every 3 years and more frequently if deterioration is discovered or if marine animal infestation is critical (Kelly, 1999).

17.3.2   Magnetic particle

Besides visual inspection, magnetic particle inspection (MPI) has been the most widely used NDE method underwater (Goldberg, 1996; Na and Kundu, 2002). MPI is employed to detect surface and near-surface flaws in ferromagnetic materials, and is used to look for cracking at welded joints and in areas identified as being susceptible to environmental cracking, fatigue cracking, or creep cracking. The process of MPI is relatively simple. A magnetic field is applied to the specimen, either locally or overall, using a permanent magnet, electromagnet, flexible cables, or hand-held products. Fine ferromagnetic particles are then applied onto the specimen's surface. If the material is sound, most of the magnetic flux is concentrated below the material's surface and the particles follow the induced magnetic field

(Shull, 2002; TWI, 2008). However, if a flaw is present, such that it interacts with the magnetic field, the flux is distorted locally and 'leaks' from the surface of the specimen in the region of the flaw.

MPI is carried out by divers and is generally used as a prime NDE technique to define the true length of discontinuities detected visually (Kelly, 1999). It can be applied only to bare metal specimens and is therefore a time-consuming process, as surface cleaning is normally required before its application (Visser, 2002; Mijarez, 2006). Underwater MPI is used primarily as a quality assurance tool to support underwater welding on ship structures. It can also be used to inspect hulls or other magnetic components. Some of the drawbacks associated with MPI are:

1. low sensitivity to detect cracks that run parallel to the magnetic field. In this circumstance there is little disturbance to the magnetic field and it is unlikely that the crack is detected. To overcome this limitation the inspection surface is magnetized in two perpendicular directions. Alternatively, techniques using swinging or rotating magnetic fields can be used to ensure that all orientations of a crack are detectable.
2. residual magnetic fields left after the inspection is terminated may interfere with welding repairs. These can be removed by slowly wiping the surface with an energized alternating current (AC) yoke.
3. deeply embedded flaws cannot be detected.

17.3.3   Ultrasonic testing (UT)

UT is suitable for the detection of subsurface defects, thickness measurements, weld examinations, and the detection of inclusion and internal corrosion (Bayliss *et al.*, 1988; Na and Kundu, 2002). Steel structures such as H-piles, pipe piles, and sheet piles used in waterfront facilities are inspected using bulk waves to measure the metal thickness (Kelly, 1999). Ultrasonic-based systems can also be used to obtain a general condition rating and indication of overall strength of concrete based on sound velocity measurements (Kelly, 1999). The UT method is based on the propagation of ultrasonic stress waves generated by one or more probes that send broadband or narrowband mechanical waves through the structure. Conventional UT such as local thickness gauging uses bulk waves to test a limited region within the ultrasonic probe. This approach, although widely used in both dry and wet structures, is time consuming when applied to large structures. Longitudinal waves are used for ultrasonic thickness measurements in marine risers. The thickness is determined by the time of flight for the ultrasonic signal to reach the back surface of the pipe and to return to the transducer. Through-coating measurements allow coated risers to be inspected without removal of the coating after applying echo-to-echo technique and

A-scan imaging if the coating is well bonded to the metal surface and the thickness of the coating is less than 6 mm (Lozev *et al*., 2005). One form of UT is represented by phased arrays consisting of multiple ultrasonic elements and electronic time delays that create beams by constructive and destructive interference. The beams can be steered, scanned, swept, and focused electronically. To generate a different beam angle, different transducers' wedges or lenses must be used. Electronic scanning permits very rapid coverage of the components to be inspected, typically an order of magnitude faster than a single transducer mechanical system. Owing to the large amount of data generated during inspection with ultrasonic phased arrays, data processing is not as straightforward as with conventional ultrasonic transducers. Marine risers can be scanned faster with a higher probability for defect detection and sizing using phased-array transducers (Lozev, 2001; Lozev *et al*., 2005).

Whenever an ultrasound propagates into a bounded media, a guided ultrasonic wave (GUW) is generated. The wave is termed 'guided' because it travels along the medium guided by the medium's geometric boundaries. GUWs propagate along, rather than across, the waveguide. The advantage of GUW inspection is its ability to probe long lengths of the waveguide, locating cracks and notches from only a few monitoring points, while providing full coverage of the waveguide's cross-section. An application of GUWs applied to marine risers is provided by Guided Ultrasonics Ltd, who perform inspections of the splash zone and in general subsea pipework. The system is made of marinized ring and electronics (GUL, 2011).

17.3.4   Sonars

Sonar measures the time it takes for a burst of sound to travel from a source to a target and back again. By knowing the velocity of sound through the appropriate medium, the distance from the source to the target can be determined (Rizzo, 2010). Sonar can generate an image of the test object.

Researchers at MIT developed the MIT-Bluefin hovering autonomous underwater vehicle (HAUV) shown in Plate XVI in the color section between pages 374 and 375 (Hover *et al*., 2007; Englot *et al*., 2009; MIT, 2011). The system employs a DIDSON (Dual Frequency Identification Sonar) imaging sonar (Belcher *et al*., 1999) and BlueView bathymetry sonar to construct a detailed 3D representation of the test object. Imaging sonar can be considered an underwater camera that uses ultrasonic waves at few MHz frequencies to create images of a nearby object. Instead, bathymetry sonar uses low frequency acoustic waves to map the seafloor. In the MIT system, the sonars are coupled to a Kalman filter simultaneous localization and mapping (SLAM) processing tool. The vehicle was tested to identify mine-shaped training targets planted on the hull of a retired naval aircraft

carrier, the USS *Saratoga*. The test covered a 24 m^2 section of the hull on which nine mine-shaped training targets were mounted.

17.3.5   Radiography

Radiographic techniques (specifically X-rays) are one of the few NDE methods that can examine the interior of an object, and the only NDE method that works on all materials (Shull 2002). An X-ray is a form of electromagnetic radiation with a wavelength in the range of 0.01 to 1 nanometers, corresponding to frequencies in the range of 3×10^{17}–3×10^{19} Hz. X-rays have high electromagnetic energy. These rays pass through objects that block visible light. This allows 'seeing' the interior of the material/structure under investigation. The amount of X-rays that pass through a material is dependent on the elemental composition, density, and thickness of the material, and the energy and amount of X-rays (Shull, 2002). This method can detect cracks, flaws, and thickness reduction. X-ray-based NDE has the advantages of being accurate, inherently pictorial, and adaptable to examine shapes and sizes, and sensitive to the discontinuity that causes a reasonable reduction of cross-section thickness. However, this method carries the burden of some safety concerns. It can be time consuming and expensive. It also requires extensive experience and trained personnel to safely carry out the inspection and properly interpret the images (Zhu and Rizzo, 2010). As radiographic methods were proposed to inspect pipes (see for instance Balaskó *et al.*, 2005), the use of X-ray technology has been extended to monitor any tubular components and welds under water.

The advent of digital radiography represented a leap in this NDE methodology. Digital radiography systems offer the possibility of obtaining images with less strict exposure requirements than those of conventional film systems. Exposure imprecision normally leads to radiographs that are dark or show little contrast. Some of the advantages of digital radiographic systems include rapid image display, reduction of X-ray doses, image processing, automated acquisition, partially or completely automated evaluation, and image storage (Moreira *et al.*, 2010; Pincu and Kleinberger, 2010).

Moreira *et al.* (2010) demonstrated that digital radiography provides superior performance with respect to conventional film technique in terms of image quality indicators (IQIs) and in the detection of small defects. IQI is a parameter that assesses the quality of a radiographic image in terms of sharpness, contrast, and noise. The comparison was conducted by targeting defects in the welds of offshore pipelines. For this study the digital detector array, Varian 2520 V 127 μm, was used. The pipes were manufactured and welded according to the requirements established by API 5L and ISO 3183 specifications (API, 2007; ISO, 2007). The pipes were welded introducing artificial flaws such as longitudinal and transverse cracks, lack

of penetration, and lack of a side wall. The quality of the images was assessed by the following parameters: normalized signal-to-noise ratio at base material, basic spatial resolution, and contrast sensitivity by IQI. The defect visibilities obtained with a digital detector were compared with those obtained with digitalized films. It should be stated that, although the method was validated on pipes that were to be deployed in a marine environment, the experiments presented in this study were performed in dry conditions.

17.3.6   Electric/magnetic methods

Eddy current (EC)-based methods measure a material's response to electromagnetic fields over a specific frequency range, typically a few kHz to several MHz (Shull 2002). The principle is illustrated in Fig. 17.3. A magnetic coil with an AC induces a time-varying magnetic field, which causes an electric current to be generated in the test object. These currents produce small magnetic fields around the material that generally oppose the original field and therefore change the impedance of the magnetic coil. By measuring the change in the impedance of the magnetic coil as it traverses the sample, anomalies in the inspected structure can be detected.

EC-based inspection is a non-contact method that does not require surface preparation. On the other hand, EC suffers from a number of drawbacks, such as that the interrogated material must be a conductor, it is sensitive to liftoff variations (i.e. the distance between the structure and the EC probe), and it is limited to surface or near-surface detection (Shull, 2002). The last limitation is related to the size of the skin depth, which is the
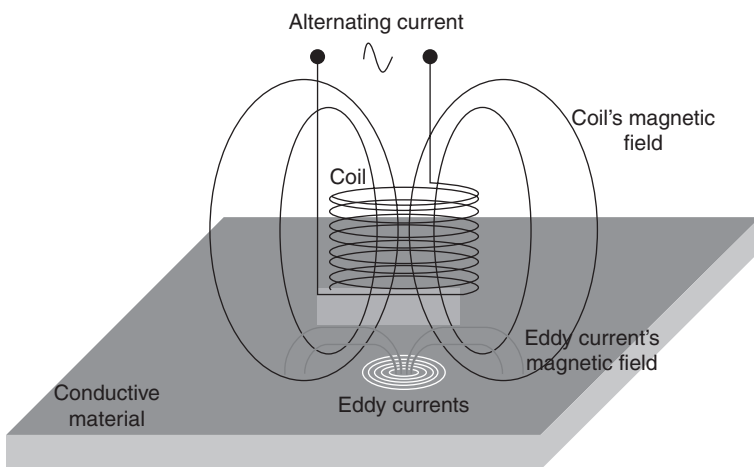
17.3 Schematic of the EC inspection.

depth of penetration of the AC. The skin depth sets the maximum depth to determine defects at a given frequency. To overcome this problem, the remote field eddy current (RFEC) method was introduced by Schmidt *et al*. (1989). RFEC utilizes the existence of a secondary field that completely penetrates the wall thickness, thereby allowing through-thickness inspection (Shull, 2002). The current can travel along a test object, such as a pipe, and be detected by far field coils, which are separated by approximately two times the pipe diameter from the excitation coils.

A pulsed EC technique was proposed for riser inspection by Wassink *et al*. (2000). A stepped or pulsed input signal is used for the detection of corrosion areas under the riser's insulation. This approach allows the detection of wall-thinning areas in the riser without removing the outside protection coatings or any accumulated biomass. The advantages of this technique are its larger penetration depth and the possibility of obtaining quantitative measurements of wall thickness. The technique, which was initially used for the outside inspection of plant pipelines (see for instance de Raad, 1995), can be applied for nominal riser wall thicknesses between 6 and 65 mm and insulation thickness less than 150 mm using probes located within 50–100 mm range (Lozev *et al*., 2005).

The alternating current field measurement (ACFM) technique is another non-contact electromagnetic technique utilized to detect and size surface-breaking cracks in a range of different materials and through coatings of varying thickness. The technique measures the magnetic field perturbations associated with electric field perturbations induced by the presence of a flaw (Raine, 2002). The ACFM perturbations in a uniform magnetic field can be detected with coils parallel or perpendicular to the field, or perpendicular to the surface. Modified systems employ one induction coil for generating surface currents and two small coils to measure the magnetic field in the area surrounding a crack (Mijarez, 2006). The system was originally developed in the early 1990s for subsea and topside inspections of offshore structures without the need to remove the item's protective coating. Nowadays the technique is routinely used by oil and gas industries for structural weld inspection. Other applications include in-service crack detection in naval vessels. Although developed and patented by TSC Inspection Systems (TSC, 2011) initially for routine inspection of structural welds, the technology has been improved further to cover broader applications across a range of industries (Lugg and Topp, 2006; Papaelias *et al*., 2007, 2008).

A few examples of ACFM applied to underwater structures were reported by Marques and Martins (2000) and Asokan *et al*. (2003). The former described the use of ACFM at PETROBRAS, the Brazilian State Oil Company, for the routine structural inspection of offshore platforms. On

the other hand, Asokan *et al*. (2003) developed an underwater manipulator for the inspection of weld seams.

17.4   Structural health monitoring (SHM) of underwater structures

Currently, there is growing interest in developing a scientific process that involves the observation of a structure or mechanical system over time using periodically spaced measurements, the extraction of damage-sensitive features from these measurements, and the statistical analysis of these features to determine the current state of system health. This process is referred to as SHM (Farrar and Worden 2007). Morimoto (2010) illustrated the benefits of using new technologies that could replace periodic inspection by means, for instance, of wireless sensor networks. Morimoto (2010) investigated the environmental and social impacts of applying wireless sensor technology in the United Kingdom water industry. The study was conducted performing a probabilistic cost–benefit analysis, which takes into account future uncertainty. In this section we describe those nondestructive methods that can be implemented in the scientific process of SHM, and therefore ideally implementable online with embedded hardware/software in an automated manner as a system operates.

17.4.1   Acoustic emission

Acoustic emission (AE) is a passive method that monitors the transient stress waves generated by the rapid release of energy from localized sources, e.g. fracture, within a material. The elastic energy propagates as a stress wave (i.e. an acoustic emission (AE) event) in the structure and is detected by one or more sensors attached to or embedded in the structure being monitored (Zhu and Rizzo, 2010). Such an event can be linked to the onset of new damage or to the progression of existing anomalies. AE differs from most other NDE techniques in two key respects. First, the signal has its origin in the material itself, not in an external source. Second, AE effectively detects movement while most other methods detect existing geometrical discontinuities. Different AE sources may produce different AE waveforms. The AE source mechanism results in different received signals if the source is oriented differently with respect to the geometry of the medium or the propagation path to the detector.

AE is suitable for global monitoring, real-time evaluation, and remote sensing, to discriminate among different sources of events (i.e. sources of damage). As with any method related to the propagation of stress waves, AE may suffer from signal attenuation and may be subjected to extraneous noise. It is a contact method, in that it requires contact between the sensing

technology and the structure under investigation. The previous loading history of the structure cannot always be determined, and the presence of existing damage cannot be appreciated. To determine the location of the AE event, multiple sensors are required. Unlike other NDE methods, AE-based methods cannot define the defect size or how serious it might be (Mijarez, 2006). Moreover, offshore environments are generally very noisy and the AE signals are usually weak, especially if the AE sensors are located far away from the AE event. Thus, signal processing and discrimination should be aggressively pursued to successfully monitor underwater structures by means of AE.

The underwater monitoring of existing structures by means of AE requires that transducers and cabling must be retrofitted onto the structures. This procedure needs divers and hence can be time consuming and expensive to install.

17.4.2   Fiber optics

The use of distributed fiber optic sensors based on the Sagnac interferometer (Russell *et al.*, 2001; Huang *et al.*, 2007; Kondrat *et al.*, 2007; Wu *et al.*, 2008) to locate a disturbance applied to its sensing fiber can be exploited to identify and monitor leaks from pipes (Kirkendall and Dandridge, 2004; Manuel *et al.*, 2008). The fiber is fixed on the surface of a pipe to sense the strain caused by the acoustic signals generated by the leaks. The strain produces a change of the refractive index of the fiber core, alters the total optical path length, and induces an optical phase signal in the fiber.

Wang and Wang (2010) presented a hybrid Mach-Zehnder and Sagnac interferometer configuration for detecting broadband vibrations caused by leaks from a high pressure pipe. The sensing system is based on spectrum analysis and signal filtering based on the discrete wavelet transform, and it was tested in a laboratory setting using an underwater waveguide. The light source was a 1050 nm amplified spontaneous emission (ASE) broadband light source. The experimental set-up included a 4500 m long single model sensing optic fiber. The photodetector amplifier had a low pass filter with cut-off frequency of 100 kHz. The performance of the system was demonstrated under different water flow velocities. Huang *et al.* (2007) proposed a scheme with two light paths, where the same optical length is traveled following different sequence paths. Because the propagation lights of the two paths pass through the leaking point at different times, the resulting phase of the two signals differ. After interference, the sensing phase signal is demodulated, and then the leaking point can be acquired from the null frequency of the output spectrum.

Fiber optic sensing based on fiber Bragg grating (FBG) was proposed to measure strains for ship hull monitoring systems. In addition to immunity to electromagnetic fields and resistance to harsh corrosive environments, FBG sensors have the ability to measure multiple physical parameters (Chan *et al.*, 2006). This ability, combined with serial multiplexing of FBG sensors, allows for multiple parameters to be monitored (Murawski *et al.*, 2010). Interested readers are referred to Wang *et al.* (2001) for an application of fiber optics in vessels. Murawski *et al.* (2010) suggested the use of fiber Bragg technology to measure strain in offshore structures. The principle of this approach was proved on the construction legs of an offshore platform simulated using finite element methods.

17.4.3 Vibration methods

Underwater SHM methods based on vibration analysis rely on the natural vibration modes of the offshore structure monitored. These modes, which are excited by natural wind and waves, are dependent on the characteristics of the structure. If the mass of the structure does not change, the decrease in its stiffness, caused by possible damage inflicted to its load-carrying members, will result in shifting of its characteristic vibration frequencies. When vibration methods are deployed, it is assumed that the breakage of the structural members, or large changes of the sediment foundation (or sediment characteristics) result in detectable changes of the natural vibration frequencies. In order to be reliable, the instrumentation (usually accelerometers and data acquisition systems) must be very sensitive. The raw data, when coupled to advanced signal processing methods, are rendered immune to ambient variations (such as tidal waves). Vibration analysis has been applied to confirm structural damage below and above the water line, placing sensors at deck level and as deep as 30 m. For instance, EQE International installed sway monitoring sensors on 13 different platforms in the North Sea (E&P, 2001; Burdekin *et al.*, 2002; Mijarez, 2006).

In tubular joints, when fatigue cracks penetrate through the thickness of the tubular wall, a major reduction in joint stiffness is observed. This stage is usually considered the end of the fatigue life of the joint and a variation in the global natural frequency of vibration might be seen in structural response data.

Fritzen *et al.* (2008; 2010) presented a system for long-term monitoring of offshore wind energy plants. The system contains algorithms for load identification, damage detection under changing environmental and operational conditions, damage localization, sensor fault identification, and operational modal analysis. The system analyzes the data provided by an array of accelerometers positioned on the tower or the foundation of any

wind energy plant. The study presented by Fritzen *et al.* (2010) describes the outcomes of 8 months' monitoring of the tower of a wind energy plant prototype M5000-2.

Vibration analysis applied to offshore structures suffers from the same drawbacks as for above-ground structures: it cannot determine that a crack has occurred until the crack becomes large enough to affect the natural vibration frequency of the structure. It may also suffer from such environmental factors as temperature on the superstructure.

17.4.4   Flooded member detection (FMD) method

One technique that is particularly suitable for SHM implementation aims at detecting flood in underwater structural components that are normally water tight such as dry tubular members. The process is called FMD. The presence of water in the dry tubular member provides an indication of through-wall failure. Although flood may not compromise the overall integrity of the structure, the FMD method is a suitable approach to detect structural damage in tubular elements.

Historically, the detection of flooded members was developed as diver-based inspection or for fully automated ROV use. To this end, four methods were proposed: (1) acoustic impact; (2) thermal profiling; (3) radiographic; and (4) ultrasonic methods. The acoustic impact testing was the earliest method. A mechanical excitation is induced on the member under investigation and the resultant vibration is compared with a known healthy pattern. The approach is fast, with minimal cleaning requirement, and easily implemented in an ROV. However, the external fluid loading and the growth of marine organisms on the member surface can mask the required information and therefore affect reliability. In thermal profiling, a heat source is applied around the member and the resultant temperature is monitored (Hayward *et al.* 1993). The low inspection rates and the difficulty in making the system fully automatic represent the main disadvantages of this approach. Radiographic-based and ultrasonic-based FMD are the main FMD methods. Radiographic-based FMD uses gamma ray and a sensitive detector fixed on opposite forks for a changeable yoke system and mounted across the diameter of the member under inspection. The method exploits the gamma ray absorption properties of materials to detect the presence of water. Radiographic FMD systems are generally attached and interfaced to ROV to mitigate any potential radiation hazards. The commercial product Tracero Diagnosis TM™ (http://www.tracerco.com/) offers radiographic FMD equipment for offshore applications and it is adaptable to ROVs. The main advantages of this approach are that through-thickness defects are found and that the method is readily deployable, even from smaller ROVs. However, the method requires access to both sides of the member, with

extensive surface preparation and pretest calibration. Particular problems areas are: tight cracks, coatings, sealing by marine growth, and a general difficulty of access, principally on old platforms (Mijarez, 2006).

In ultrasonic-based FMD, an ultrasonic pulse is transmitted through the member wall. The presence of a fluid provides an efficient acoustic conduction path, which otherwise cannot exist, and the detection of a return echo from the opposite side of the cylinder is used to confirm through-wall damage (Hayward, 1993). An ultrasonic FMD system for underwater tubular
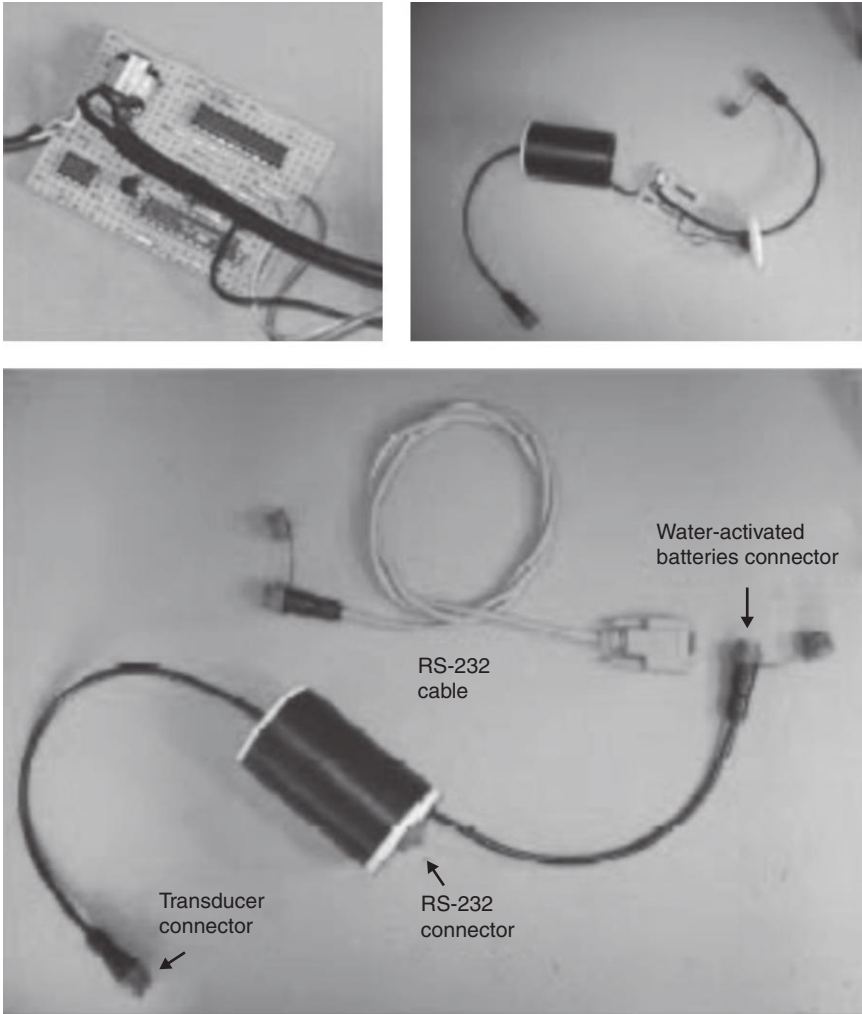
17.4 Waterproof sensing system for flood detection method (*Source*: From Mijarez, 2006, page 153).

steel structures was described by Hayward (1993). The system targeted angled, horizontal, and vertical members of steel jacket offshore installations and it required only removal of external marine growth over the operational diameter of the sensor head. Ultrasonic-based FMD requires divers and an accurate placement of transducers (Mijarez, 2006). The ultrasonic-based FMD was evaluated in the framework of the Inter-Calibration of Offshore NDE (ICON) project, which aimed at testing NDE methods and equipment. The FMD method was evaluated using clean 0.4 m and 0.5 m diameter tubes, randomly filled with water, with and without computer assistance. It was reported that at 50% or higher of water fill of the specimen, 100% probability of detection (POD) was obtained, and 70% POD at 10% water fill (Visser, 2002; Mijarez, 2006).

The transition of the FMD from a diver-based or ROV-based inspection to SHM has been proposed by researchers at the University of Manchester. A photo of the sensing system is presented in Fig. 17.4. The system uses guided waves and is based on sensors designed to be permanently attached to the inner wall of every tube in the subsea structure and powered by a normally inert seawater battery. Upon activation, the sensor transmits ultrasonic encoded information to an online monitoring system at deck level for decoding and identification of the transmitting sensor member. The transmitted signal exploits two communication channels, either through seawater or along the structure using the waveguide effect provided by the steel jacket (Mijarez *et al.*, 2005a, b, 2006, 2007).

17.5   Conclusion

Underwater structures play a vital role in modern society. Owing to the environmental conditions under which these structures operate, it is important to guarantee proper maintenance and avoid any failure that might compromise the marine environment and lead to economic losses.

In this chapter a comprehensive review of the most common method used for the NDE and SHM of underwater structures was given. Methods based on electromagnetic waves, stress waves, visual inspection, and radiography were discussed. Advantages and drawbacks of each methodology as it applies to underwater systems were reported. Table 17.1 lists the methodologies discussed here and associates them to the structures to which they can be applied. It must be stated that in their applicability to underwater structures, some of the sensing technologies listed in Table 17.1 are still in their infancy, and therefore only proof of concept has been shown in the laboratory or in pilot tests, or they have the potential to be used but have never been explored for underwater applications. For example, fiber optics belongs to the first category, while AE belongs to those sensing methods that have potential but have never been explored.

Table 17.1 Inspection and sensing methods for underwater structures. The methods are listed in alphabetical order

NDE methods	Offshore platforms	Piles and piers	Submerged pipelines	Naval vessels	Damage type	Size of equipment	Simplicity	*In situ* implementation	Cost
AE	x	x			Surface, interior	M	M/H	L	L/M
Acoustic leak detection			x		Full depth	M	M	M	M
ECs	x			x	Surface, underneath	L/M	M	L	M
Fiber optics	x		x	x	Surface	M	L/M	L	M
FMD	x				Full depth	L/M	M/H	M	L/M
Magnetic particles	x			x	Surface, underneath	L	H	M	L
Radiography	x	x		x	Surface, interior	H	M	M	H
Sonar	x	x	x	x	Surface	M	M	H	M/H
UT	x	x	x	x	Surface, interior	M	M	L/M	M
Vibration method	x	x		x	Interior, full depth	M	H	L/M	L/M
Visual inspection (by divers)	x	x	x	x	Surface	L/M	H	H	L
Visual inspection (with cameras)	x	x	x	x	Surface	M	H	H	M
Visual inspection (with dyes)	x		x		Full depth	M	M/H	M	M

17.6   References

Abu-Zeid M E, Bhatia K S, Marafi M A, Makdisi Y Y and Amer M F (1987), 'Measurement of fluorescence decay of crude oil: a potential technique to identify oil slicks', *Environmental Pollution*, **46**, 197–207.

Aghakouchak A A and Stiemer S F (2001), 'Fatigue reliability assessment of tubular joints of existing offshore structures', *Canadian Journal of Civil Engineering*, **28**, 691–698.

API 5L, *Specification for Line Pipe*, American Petroleum Institute, Washington, DC, USA, 2007.

Asokan T, Seet G and Angeles J (2003), 'The optimum dimensioning of an underwater manipulator for weld inspection', *Proc. of the Institution of Mechanical Engineers, Part M: J. of Engineering for the Maritime Environment*, **217**(4), 177–184.

Baker M J and Descamps B (1999), 'Reliability-based methods in the inspection planning of fixed offshore steel structures', *Journal of Constructional Steel Research*, **52**, 117–131.

Balaskó M, Sváb E, Kuba A, Kiss Z, Rodek L and Nagy A (2005), 'Pipe corrosion and deposit study using neutron- and gamma- radiation sources', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **542**(1–3), 302–308.

Barbagelata A and Barbagelata L (2004), 'New acoustic leak system saves money', *Offshore Pipeline Gas Journal*, **231**, 42–44.

Barbagelata A and Barbagelata L (2002), 'Acoustic leak detection for underwater oil and gas pipelines', *Sea Technology*, **43**(11), 39–44.

Bartz R, Spinrad R W and Kitchen J C (1988), 'A low power high resolution in-situ fluorometer for profiling and moored applications in water', *Proc SPIE 925, Ocean Optics IX*, 157–170.

Bayliss M, Short D and Bax M (1988), *Underwater Inspection*, E. & F. N. Spon, New York.

Belcher E, Dinh H, Lynn D and Laughlin T (1999), 'Beamforming and imaging with acoustic lenses in small, high frequency sonars', Proc. IEEE OCEANS Conf., Seattle, 1495–1499.

Burdekin F M, Talai-faz B, Brennan F P and Dover W D. (2002), Experimental validation of the ultimate strength of brace members with circumferential cracks, UMIST and University College London for the Health and Safety Executive. Offshore technology report.

Busby R F (1978), 'Underwater inspection/testing/monitoring of offshore structures', *Ocean Engineering*, **6**(4), 355–491.

Camagni P, Colombo G, Koechler C, Pedrini A, Omenetto N and Rossi G (1988) 'Diagnostics of oil pollution by laser-induced fluorescence', *IEEE Transactions on Geoscience And Remote Sensing*, **26**(1), 22–26.

Chan T H T, Yu L, Tam H Y, Ni Y Q, Liu S Y, Chung W H and Cheng L K (2006), 'Fiber Bragg grating sensors for structural health monitoring of Tsing Ma bridge. Background and experimental observation', *Engineering Structures*, **28**, 648–659.

de Raad J A (1995), 'Novel techniques for outside inspection of plant pipework', *Insight*, **37**, 409–412.

E&P (2001) http://www.epmag.com/archives/print/3581.htm.

El-Reedy M A (2002), 'Optimization study for the offshore platform inspection strategy', Society of Petroleum Engineers Inc.

Englot B, Johannsson H and Hover F (2009) ' Perception, stability analysis, and motion planning for autonomous ship hull inspection', *Proc. 15th Symposium on Unmanned Untethered Submersible Technology (UUST)*, http://web.mit.edu/hovergroup/pub/mit_hauv_uust.pdf.

Farrar C F and Worden K (2007), 'An introduction to structural health monitoring', *Philosophical Transactions of Royal Society A,* **365**, 303–315.

Fritzen C-P, Kraemer P and Klinkov M (2008), 'Structural health monitoring of offshore wind energy plants', *Proc. 4th EWSHM*, Cracow, Poland, pp. 3–21.

Fritzen C-P, Kraemer P and Klinkov M (2010), 'Vibration-based health monitoring for offshore wind energy plants', *Proc. 5th EWSHM*, Sorrento, Italy.

Goldberg L (1996), 'Diversity in Underwater Inspection', *Materials Evaluation*, **54**(3), 401–403.

GUL (2011) http://www.guided-ultrasonics.com/en/wp-content/uploads/2008/02/gul-subsea.pdf.

Hayward G, Pearson J and Stirling G (1993), 'An intelligent ultrasonic inspection system for flooded member detection in offshore structures', *IEEE TUFFC*, **40**(5), 512–521.

Hover F S, Vaganay J, Elkins M, Willcox S, Polidoro V, Morash J, Damus R and Desset S (2007), 'A vehicle system for autonomous relative survey of in-water ships', *Marine Technology Society Journal*, **41**(2), 44–55.

Huang S-C, Lin W-W, Tsai M-T and Chen M-H (2007), 'Fiber optic in-line distributed sensor for detection and localization of the pipeline leaks', *Sensors and Actuators A*, **135**(2), 570–579.

ISO 3183 (2007), *Petroleum and Natural Gas Industries – Steel Pipes for Pipeline Transportation Systems*, Geneva, Switzerland.

ITP (2011) http://www.itp-interpipe.com/products/subsea-cryogenic-pipelines/subsea-cryogenic-pipelines.php.

Kelly S W (1999), Underwater inspection criteria, Report prepared for the California State Lands Commission.

Kirkendall C K and Dandridge A (2004), 'Overview of high performance fibre-optic sensing', *Journal of Physics D: Applied Physics*, **37**(18), R197–R216.

Kondrat M, Szustakowski M, Pałka N, Ciurapiński W and Życzkowski M (2007), 'A Sagnac-Michelson fibre optic interferometer: signal processing for disturbance localization', *Opto-Electronics Review*, **15**(3), 127–132.

Laurendine T (2002), 'Inspection of offshore composite applications', NDE Evaluation Methods for Inspecting Offshore Composite Structures, Proceedings of Workshop held November 21, 2002 Houston, TX

Lopez-Anido R, Michael A P, Goodell B and Sandford T C (2004), 'Assessment of Wood Pile Deterioration due to Marine Organisms', *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **130**(2), 70, 7 pages.

Lozev M (2001), 'Phased-array technology promises increased ultrasonic inspection productivity, accuracy and flexibility', *EWI Insights*, **14**(1), 1–3.

Lozev M G, Smith R W and Grimmett B B (2005), 'Evaluation methods for detecting and monitoring of corrosion damage in risers', *Journal of Pressure Vessel Technology*, **127**, 244–254.

Lugg M and Topp D (2006), 'Recent developments and applications of the ACFM inspection method and ACSM stress measurement method', *Proc. ECNDT*, Berlin, Germany.

Manuel R M, Shlyagin M G and Miridonov S V (2008), 'Location of a time-varying disturbance using an array of identical fiber-optic interferometers interrogated by CW DFB laser', *Optics Express*, **16**(25), 20668–20675.

Marques F C R and Martins M V M (2000), 'Experiences in the use of ACFM for offshore platform inspection in Brazil', 15th WCNDT, Rome, Italy. http://www.ndt.net/article/wcndt00/papers/idn233/idn233.htm.

McStay D, McIlroy J, Forte A, Lunney F, Greenway T, Thabeth K and Dean G (2005), 'A new tool for the rapid remote detection of leaks from subsea pipelines during remotely operated vehicle inspections', *Journal of Optics A: Pure Applied Optics*, **7**, S346–S351.

McStay D, Milne R, Pollard P and Dunn J (1995) 'Sea trials of an optical fibre marine fluorosensor', *Measurement Science and Technology*., **6**, 1309–1316.

McStay D, Kerlin J and Acheson R (2007), 'An optical sensor for the detection of leaks from subsea pipelines and risers', *Journal of Physics: Conference Series* **76**(2007) 012009, (6 pages).

Mijarez R, Gaydecki P and Burdekin M (2005a), 'An axisymmetric guided wave encoded system for flood detection of oil rig cross-beams', *Measurement Science and Technology*, **16**(11), 2265–2274.

Mijarez R, Gaydecki P and Burdekin M (2005b), 'Continuous monitoring guided wave encoded sensor for oil rig flooded member detection', *Insight*, **47**(12), 748–750.

Mijarez-Castro R (2006), A remote and autonomous continuous monitoring ultrasonic system for flood detection in sub-sea members of offshore steel oilrigs, Ph.D. Thesis, University of Manchester, UK.

Mijarez R, Gaydecki P and Burdekin M (2007), 'Flood member detection for real-time structural health monitoring of sub-sea structures of offshore steel oilrigs,' *Smart Materials and Structures*, **16**(5), 1857.

MIT (2011) http://groups.csail.mit.edu/marine/pub/Kaess10istmp.pdf.

Morimoto R (2010), 'Estimating the benefits of effectively and proactively maintaining infrastructure with the innovative Smart Infrastructure sensor system', *Socio-Economic Planning Sciences*, **44**, 247–257.

Murawski L, Opoka S and Ostachowicz W (2010), 'Measurement and calculation errors estimation and damage detection possibility for SHM system of offshore structure', *Proc. Europ. Conf. Struct. Health Monitoring,* Sorrento, Italy.

Na W-B and Kundu T (2002), 'Underwater pipeline inspection using guided waves', *J. of Pressure Vessel Technology*, **124**, 196–200.

Papaelias M Ph, Lugg M, Smith M, Roberts C and Davis C L (2007), 'Detection and quantification of rolling contact fatigue cracks in rails using ACFM technology', *Proc. BINDT Conference*, Glasgow, UK, 2007.

MPh Papaelias, C Roberts and C L Davis (2008). 'A review on non-destructive evaluation of rails: state-of-the-art and future development,' *Proc. IMechE* Vol. 222 Part F: J. Rail and Rapid Transit, 367–384.

Pincu R and Kleinberger O (2010), 'Portable X-ray in the service of art,' *Materials Evaluation*, **68**(3), 311–318.

Raine, A (2002) 'A review of the development of the alternating current field measurement technique for subsea inspection' *Insight*, **44**(12) December 2002, 748–752.

Rizzo, P (2010). 'Water and wastewater pipe nondestructive evaluation and health monitoring: a review', *Advances in Civil Engineering*, Article ID 818597, 13 pages, doi:10.1155/2010/818597.

Rizzo P and Zhu, P (2010). Sensing Technology for Damage Assessment of Sign Supports and Cantilever Poles – Final Report No., FHWA – PA – 2010 -10 – PIT008 University of Pittsburgh.

Russell S J, Brady K R C, Dakin J P (2010), Real-time location of multiple time-varying strain disturbances, acting over a 40-km fiber section, using a novel dual-Sagnac interferometer, Lightwave Technology, Vol. 19, 205–213, 2001.

Saebjornsen F and Forli O (1983) Underwater NDE Techniques. Materials Evaluation/41/ April 1983.

Shull P J (2002), *Nondestructive Evaluation: Theory, Techniques, and Applications*, New York, Marcel Dekker.

So, P T and Dong, C Y (2001). Fluorescence Spectrophotometry. eLS.

Theakston J (2004) Good specifications can result in useful software-based leak detection, Pipeline Gas J. (March).

Tian, W-M (2008), Integrated method for the detection and location of underwater pipelines, *Applied Acoustics*, **69**(5), 387–398.

TSC 2011 http://www.tscinspectionsystems.co.uk/ (Accessed 2 July 2011).

Visser, W (2002) 'POD/POS curves for NDE examination' Visser Consultancy Limited for the Health and Safety Executive. Offshore technology report, Crown Copyright 2002.

Wang G, Pran K, Sagvolden G, Havsgard G B, Jensen A E, Johnson G A and Vohra S T (2001). 'Ship hull structure monitoring using fibre optic sensors,' *Smart Materials and Structures*, **10**, 472–478.

Wang Q and Wang X, (2010). 'Underwater natural gas pipeline leakage detection based on interferometric fiber optic sensor in experiment-scale,' 2010 Chinese Control and Decision Conference, 257–260.

Wassink, C H P, Roberts, M A, de Raad, J A and Bouma, T (October 2000), 'Condition Monitoring of Inaccessible Piping,' 15th WCNDT, Rome.

Wu D F, Zhang T Z and Jia B (2008), Modified Sagnac interferometer for distributed disturbance detection, *Microwave and Optical Technology Letters*, **50**(6), 1608–1610.

Xun Y, Sun X, Wu W, Liu N, Ye X and Zhu X (1990), In situ determination of oil spills using a laser fluorosensor, In *Proceedings of SPIE – The International Society for Optical Engineering*, Vol. **1230**, pp. 796–797.

Zingaretti, P S and Zanoli M (1998), Robust real-time detection of an underwater pipeline, *Engineering Applications of Artificial Intelligence*, **11**(2), 257–268.

18

Sensing solutions for assessing and monitoring offshore structures

M. H. KIM and J. M. LEE, Pusan National University, South Korea

Abstract: Structural health monitoring techniques for ship and offshore structures including recent literature survey in the related fields are summarized in this chapter. Typical hull monitoring systems (HMS) for ship structures are introduced with common sensor locations, as well as measurement parameters to be monitored. Long-based strain gauges (LBSG), accelerometers, and other types of sensors such as motion sensors and pressure sensors, which are employed for HMS, are introduced. Particular requirements for HMS in the context of the regulations and the codes by principal organization sectors are discussed. Moreover, emerging new technologies applicable for ship and offshore structures are introduced. The basic principles and the application of fiber optic sensors (FOS), acoustic emission (AE) sensors and crack detection (CD) sensors are reviewed in the later part of this chapter. Finally, new requirements and trends in structural health monitoring in marine industries are introduced. In particular, the importance of structural health monitoring technology applicable in a cryogenic environment is presented.

Key words: ships, offshore structures, structural health monitoring, fiber optic sensors (FOS), long-based strain gauges (LBSG), fatigue strength, hull stress monitoring system (HSMS).

18.1 Introduction

18.1.1 History of offshore structures

The rapid increase of oil and gas prices has resulted in the installation of a significant number of offshore structures for energy exploration in the ocean. The exploration depth is continuously increasing, from continental shelf to deep sea beyond 1000 m. Offshore structures include fixed jackets and concrete platforms, semi-submersibles, tension leg platforms (TLPs), spars, jack-ups, and floating production storage and offloading (FPSO). Common offshore structures are illustrated in Fig. 18.1. Fixed offshore structures are by far the most common kind of these structures.[1]
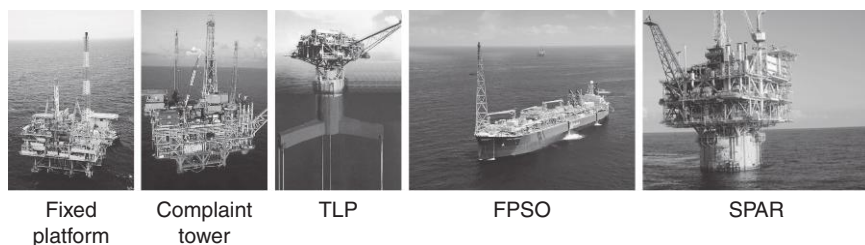
18.1 Typical offshore structures (left to right): Fixed platform, complaint tower, TLP, FPSO and SPAR.

| Fixed platform | Complaint tower | TLP | FPSO | SPAR |

Steel jacket-type platforms have been widely developed for various purposes, including offshore drilling, processing, and support of offshore operations.

It has been reported that about 5600 offshore structures have been installed and operated by oil and gas companies in the United States. Offshore structures in the Gulf of Mexico are getting old as well, and the average age is known to be 35 years. It has been reported that the age distribution of offshore structures installed in the UK continental shelf (UKCS) and the Norwegian continental shelf (NCS) shows that a relatively large number of installations have crossed of 20 years.[2,3]

Offshore structures used in petroleum activity are normally designed for various phases of their life. These phases include construction, float out from construction site to transport barge, transport to field, lifting from barge to site, operation on site and, finally, removal. The design codes normally include three limit states. These limit states are ultimate limit state (ULS), fatigue limit state (FLS), and accidental limit state (ALS).[4]

Structural integrity assessment of existing structures is performed to extend the service life of the facility, as new methods of production and new discoveries may result in a request for life extension.[5,6,7] The most generally accepted standard for offshore structures is ISO 19900 'Petroleum and natural gas industries – Offshore Structures – Part 1: General Requirements' ISO (2002). This standard provides general design rules and general rules for assessment of existing structures. The Norwegian regulations (PSA 2004) refer to NORSOK N-001 (NORSOK 2004) for structural design, which again refer to ISO 19900 (ISO 2002) for assessing existing structures. Other standards, such as API RP2A-WSD (API 2000) and ISO/DIS 13822 (ISO 2000), are also available for detailed procedures for integrity assessment of the existing structures.[2,8]

Although the removal of offshore structures is mainly an economic decision, the old structures should be properly maintained or removed to prevent any catastrophy. The Deepwater Horizon oil spill, or BP oil spill, is the largest accidental marine oil spill in the history of the petroleum industry, as shown

18.2 Deepwater horizon offshore drilling unit on fire in 2010. (*Source*: Courtesy of the United States Coast Guard.)

in Fig. 18.2. The spill caused extensive damage to marine wildlife habitats and the local community. This kind of catastrophic failure in offshore structures can be effectively avoided by proper regulation, inspection, and maintenance, as well as structural health monitoring.

18.1.2   Aims and scope

Structural health monitoring of the structures is essential for ensuring their safe operation. In this regard, the safety issues, including inspection, maintenance, and repair processes for offshore structures, have become of crucial importance in the offshore community.

   The design of a structural health monitoring system essentially requires consideration of the trade-off between initial costs and long-term operating costs. In particular, the difficult issues encountered in designing structural health monitoring systems for ships and offshore structures are:

- Operation in hostile environment of salt water, storm waves, and cargoes.
- Having a huge area of steel surface, and difficulties in access.

- Use of instrumentation being more restricted. Instrumentation transducers and connecting wires have very limited durability in a hostile environment.
- Instrumentation and monitoring system are required to determine loadings, response, and performance characteristics of critical structural elements while the ship is in service.
- Uncertainties in loadings (environmental, operating) significantly influencing the monitoring system for validating structural response and performance analysis method.

The classes of damage considered for ships and offshore structures include ship collision, slamming induced by storm, ice impact, dropped objects, fatigue fracture, fire and blast, corrosion/material degradation to name a few. Unforeseen damage to platforms may occur due to accidents, storm damage, and various geological problems.

Salvino *et al*. defined the main goals of structural health monitoring systems as:[9]

- Assessment of structural degradation and gradually degrading conditions
- Verification of design assumptions associated with design loadings and structural responses
- Assessment of potential failures due to errors in the design, fabrication and operation
- Assessment of the operational utilization of the structure.

The benefits of structural health monitoring (SHM) include reduced inspection costs, minimized preventive maintenance, increased asset availability, and extension of the remaining useful life of structures. In particular, future benefits of using SHM for offshore structures can be summarized as:

- Design validation through sea trials
- Through-life load and usage monitoring
- Damage detection and diagnostic systems
- Prognosis, mainly for fatigue life predictions.

A special point that needs to be noted is that a structural health monitoring system for ships and offshore structures in marine operation always needs to consider the regulations and codes of principal organization sectors involved in the development such as:

- Regulatory agencies
- Classification societies

- Manufacturers, designers, builders and repairers
- Owners and operators.

18.1.3   Current SHM techniques for ships and offshore structures

The basic rule requirements of a hull stress monitoring system (HSMS) are in accordance with International Maritime Organization (IMO) recommendations for 'fitting of HSMS (MSC/Circ.646 6th June 1994, Maritime Safety Committee)'. The major classification societies, such as LR (Lloyd's Register), Det Norske Veritas (DNV), and American Bureau of Shipping (ABS), and the International Association of Classification Societies (IACS) not only provide similar monitoring guidance for minimum parameters to be monitored but also recognize the fitting of enhanced and more comprehensive monitoring systems through various notations. The notations and requirements of classification societies concerning the provision of the typical monitoring for bulk carrier and tankers are summarized in Table 18.1.

18.2   Hull response monitoring systems

The current state-of-the-art structural health monitoring system for ships is Hull Response Monitoring Systems (HRMS).[10] HRMS is a system that measures and displays key ship motions and hull structural responses.[11] Real-time ship motions and stresses monitored by HRMS provide the onset and severity of hull structural responses to the sea and ice to the operators. IMO, IACS, and other individual Classification Societies (LR; DNV; and ABS) have recommended the use of HRMS.

HRMS consists of an affordable number of strain, temperature, and acceleration sensors placed in predetermined locations within ships' structures. Strain sensors are typically placed in the vicinity of possible crack locations,

Table 18.1 Notations and requirement of the classification societies

Class	Notation	Gauges
LR	SEA (HSS-4)	4 LBSG, 1 bow vertical Accelerometer
DNV	HMON-1	4 LBSG, 1 bow vertical Accelerometer, 1 midship accelerometer
ABS	HM2 + R	4 LBSG, 1 bow vertical Accelerometer
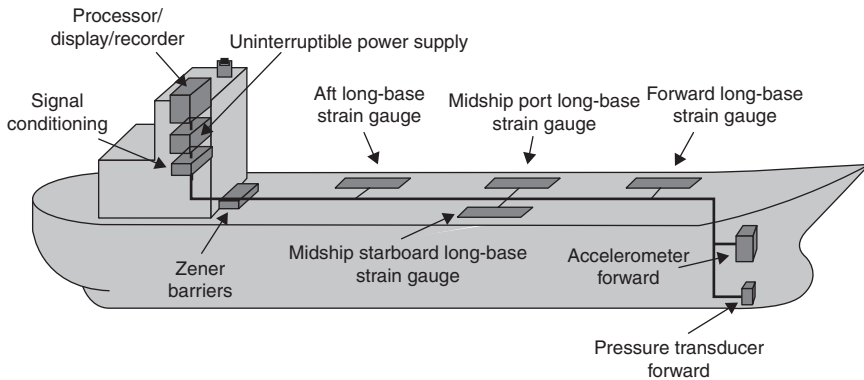
18.3 Typical hull monitoring system for a bulk carrier.

including various weld joints, as shown in Fig. 18.3. Additional sensors including global positioning system (GPS), hull hydrostatic pressure (external and in-tank), weather and motion prediction, and linkage to other ship instruments are provided.

A hull monitoring system enables the operator of the vessel to monitor all relevant responses (motions, accelerations, loads, bending moments, stresses, etc.) and provides rational guidance on preventive measures in heavy weather conditions. HSMS is a system that provides real-time information, such as motions and global stress experienced by the ship, to the crew of the ship while navigating as well as during loading and unloading operations.

While HRMS applications have matured as an industry, future applications on ice-class ships need more attention. Also, further research effort is required to relate the current HRMS system to at-sea operational guidance as well as route and schedule planning.

18.3   Fatigue monitoring sensors

Fatigue analysis consists of the characterization of short- and long-term cyclic conditions (loading and unloading of cargoes, hydrostatic pressure, hydrodynamic loadings, and machinery and equipment vibrations), the determination of the cyclic forces and strains in structural elements, and the determination of potential degradation in strength and stiffness degradation in structural elements. Fatigue damage to structural components arises because of the cyclic nature of wave or wind loadings.

The quality of fatigue strength assessment depends on the accuracy of stress concentration evaluations, ductile and fatigue resistance materials, determination of cyclic loading history, robust (damage tolerant) design, and construction/quality assurance and control.
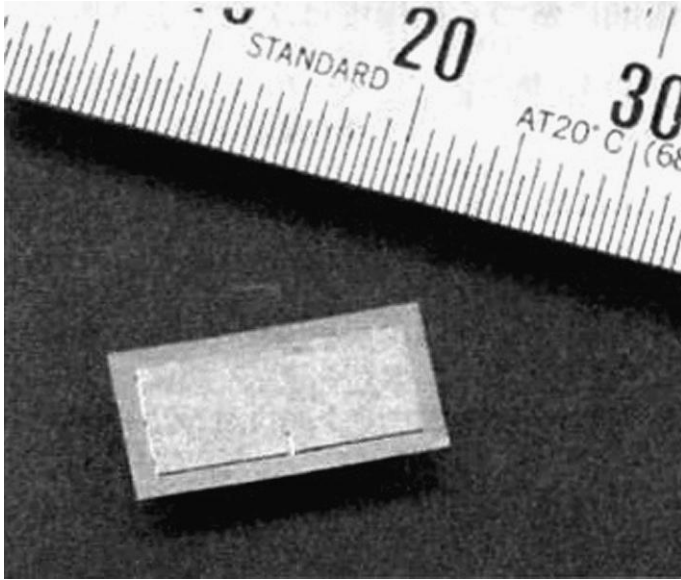
18.4 Fatigue crack monitoring sensor.

Crack growth due to fatigue and stress corrosion is normally a slow degradation process up to a point, beyond which failure may be sudden and catastrophic. Detection as early as possible during this initial period of crack growth is essential if the consequences of unexpected failure are to be avoided.

The fatigue gauge is commonly employed for the crack-growth measurement proportional to the cumulative fatigue damage for welded joints. These sensors are made of thin metal pieces and can be placed in front of stress-concentrated areas within structures for detecting fatigue cracks. These sensors are cheap, small, and made of thin metal pieces as shown in Fig. 18.4. When a fatigue crack propagates through the surface of the sensor, due to repeated fatigue loadings, the sensor can be used for measuring fatigue-crack length. Further development is required in offshore industry for easier installation of the sensors to the structure and robust sensors to offshore environmental circumstances.

18.4   Air gap sensing system

Air gap is defined as the positive difference between the highest crest elevation for the design return period and the underside of the lowest deck level.[12]

ISO 19902 standard on fixed offshore installations describes the necessity for measuring the air gap as:

> Where air gap measurement devices are correctly set up, calibrated and maintained, continuous records of wave heights and tide can provide very useful information on environmental conditions. Where this can be combined with directionality data and ideally some method of estimating actions, the data can be used in analyses and assessment of defects and of remaining life, possibly reducing conservatism.

Simple methods, such as a tape measure from the cellar deck, have also been widely used, which have recently been replaced by radar measurement. Measured values are typically recorded and averaged over a given time period, for example, every three years.

While periodic survey of air gap is sufficient most of the time, real-time continuous monitoring of the air gap is important in seismically active regions or in harsh local conditions.

18.5   Corrosion monitoring system

Ballast tanks and crude cargo tanks are continuously exposed to corrosive environment. Typically, ballast condition is worst in terms of corrosion damage, particularly in empty or partially filled conditions. Cathodic protection or coatings are commonly adopted for protecting structures from corrosive environment.

Problems are due to improperly designed, applied, and maintained corrosion systems, and incompatibilities between structural and corrosion protection systems, such as flexible bulkhead covered with stiff coatings and corrosion cells set up between the parent materials and the weld heat affected zone (HAZ) region resulting in grooving corrosion.

18.6   Acoustic emissions monitoring sensors

Acoustic emission (AE) techniques can be used for monitoring corrosion in onshore tanks, but it is known that these would be much more difficult to apply on an FPSO, because the emission levels are too low and are easily masked by other noises on the installation.

AE monitoring uses the transient acoustic stress-wave for detecting damage mechanisms such as cracks by characterizing the sound patterns due to structural anomalies induced locally in structures. The system has been used in areas with high risk of fatigue cracking and with difficult access for inspection.[12]

AE provides real-time information on fatigue crack initiation at the early stages of propagation and growth, and can be used with strain gauges to correlate the structural stress levels. AE systems have been applied extensively
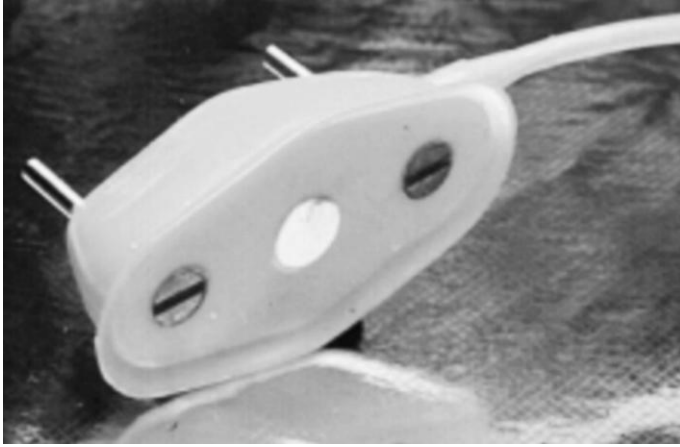
18.5 AE sensor for use on underwater marine structure.

for the structural monitoring of critical structural members over recent decades in the offshore industry.[13]

AE monitoring can overcome various difficulties with crack detection (CD) in service, and is potentially a very promising method of inspection. It is sensitive to the propagating cracks, that is, the structurally significant defects, and provides information on growth rate under service loading conditions, guiding inspection and repair work for maximum cost effective maintenance. Cost effective structural monitoring based on an acoustic method is summarized in References [14,15].

This monitoring technology can provide improved assurance of overall integrity, justifying the additional work and cost involved. By continuously monitoring a vessel or offshore structure over a period of time, usually several weeks or months depending on the minimum acceptable defect size, enhanced assurance of structural integrity can be obtained. The design of an underwater sensor is shown in Fig. 18.5, attached to a marine structure (ballast tank in vessel or jack-up platform) where it was being used to evaluate the most effective sensor for monitoring.

18.7   Vibration-based damage assessment approaches

Vibration-based damage monitoring methods have been extensively used for structural monitoring of offshore platforms. The vibration-based damage monitoring system has been in operation since the late 1970s on a number of offshore jackets.

Vibration parameters are monitored by on-line damage assessment techniques. They comprise techniques that are based on examination of changes

in natural frequencies, mode shapes, or mode shape curvatures (frequency domain methods). Doebling *et al.*[16] and Sohn *et al.*[17] published a state-of-the-art review on vibration-based damage identification methods.

Typically, accelerometers are placed on the topside of an installation with additional subsea accelerometers. The vibration response of the jacket to wave loading is continuously monitored. Any major structural damage to the platform is reflected by a change in structural response.[18] The method is known to be sensitive enough to detect a frequency change of 0.5%. However, the method cannot detect minor damage, such as small defects or local fatigue cracks.

Figure 18.6 illustrates an example of damage detection based on change in natural frequencies. By comparing the natural frequency data between the damaged and the undamaged platforms, the occurrence of damage in the platform can be easily determined.[19]

Loland and Dodds (1976) discuss practical experiences learned by monitoring three North Sea platforms over 6–9 months.[20] Discussions of platform geometry, instrumentation, environmental conditions during measurements, and system cost are presented. In the paper, the following five requirements for the vibration-based damage monitoring systems are identified.
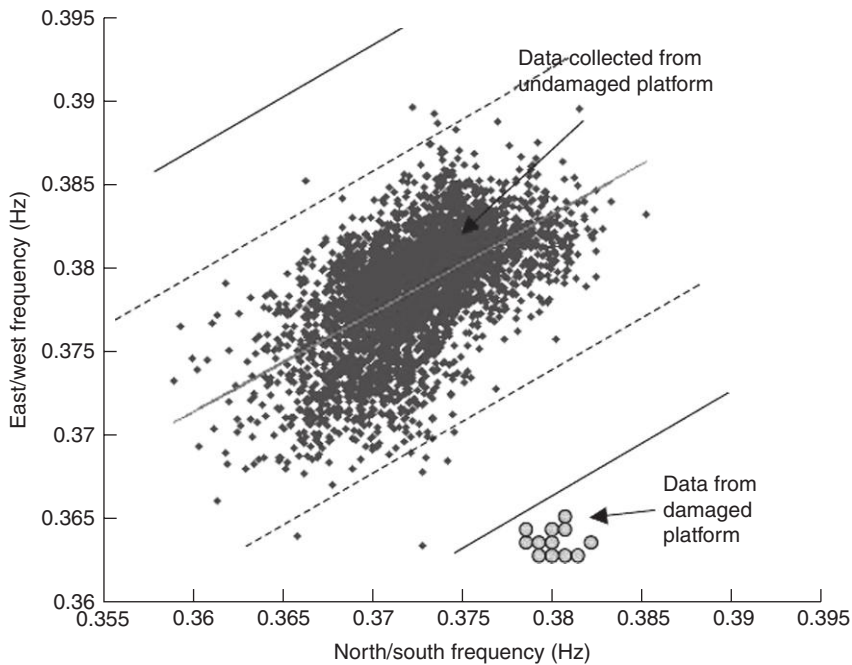
18.6 An example of damage detection based on the change in natural frequencies.[6]

- Ambient (sea and wind) excitation must be used to extract the resonant frequencies.
- Vibration spectra must remain stable over long periods of time.
- Instruments must withstand environmental challenges.
- Mode shapes must be identifiable from above-water measurements
- The system must offer financial advantages over the use of divers.

18.8   Fiber optic sensors (FOS)

FOS have recently been employed in place of conventional strain gauges to determine the stress or loading regime in part of a structure. The strain measuring techniques have been widely applied across many industries, including the offshore industry.

A typical Bragg grating type fiber optic sensor is shown in Fig. 18.7. FOSs exhibit advantages such as flexibility, embeddability, multiplexity, and electromagnetic immunity (EMI) immunity compared to traditional sensors. However, fiber optic strain sensors are limited in application for stress variation only, not absolute stress levels. Cabling, especially for conventional strain gauges in a humid environment, is another important issue to be considered. Among various fiber optic sensor methods, fiber Bragg grating (FBG) sensors are the most promising candidate for structural health monitoring application for offshore platforms.[21]

FBG sensors have been employed on offshore platforms for monitoring ship collisions and ocean wave loads.[22] It is reported that the FBG sensors installed at the bottom of the central pillar of the jacket performed well and did not show any significant reduction of sensing performance, as shown in Fig. 18.8. However, conventional strain gauges failed to operate, due to the detrimental corrosion of sea water.[22]
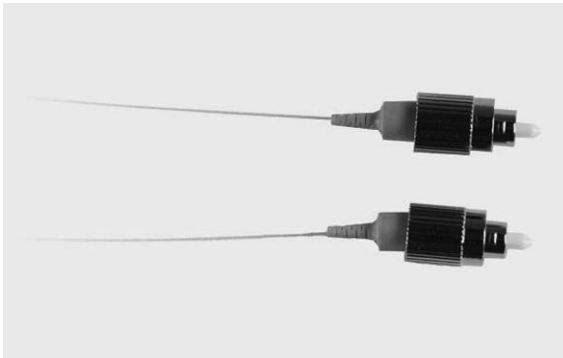
18.7 Typical FOS-based on Bragg gratings (FBGs).
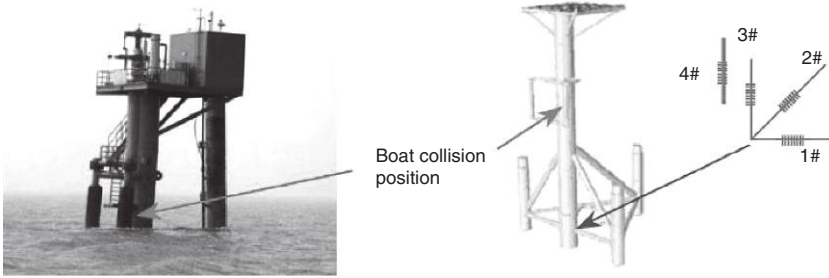
18.8 Platform, model and location of FOS.[8]
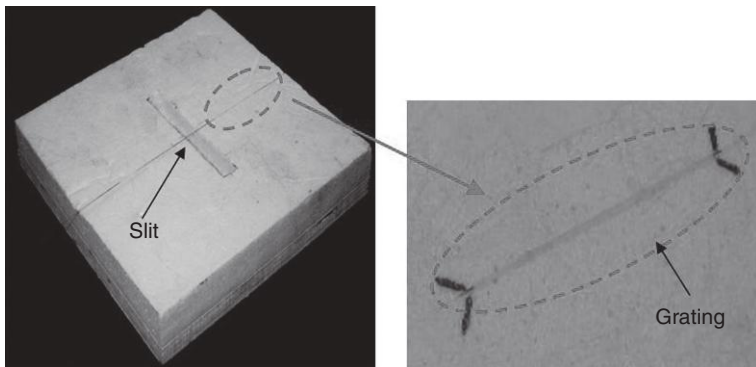
18.9 Installation of FBG sensors in LNG insulation system.[9]

Another example of fiber optic sensor application in ships and offshore structures is in the structural health monitoring of the failure modes of insulation panels of liquefied natural gas (LNG) carriers. Insulation panels of LNG cargo tanks may include mechanical failures such as cracks, as well as delamination within the layers due to impact sloshing loads and fatigue loadings, and these failures cause a significant decrease of structural safety. Figure 18.9 shows a fiber optic sensor embedded within the insulation system of an LNG vessel. A structural health monitoring system has been reported that employs FOS for monitoring the various failures that can occur in LNG insulation panels.[23]

18.9   Riser and anchor chain monitoring

The riser and anchor chain monitoring system consists of an array of sonar positioned beneath the platform, which emits signals in and around the horizontal plant.[12] Plate XVII in the color section between pages 374

and 375 shows a typical riser and anchor chain monitoring system. The purpose of this technique is to monitor the position of risers and anchor chains, and alert the operator to any displacements outside of predefined limits. In addition, failure analysis can be performed using the resulting computerized data. The measured data can be used to determine the precise location of riser and anchor chains and to continuously monitor their position.

18.10 Conclusion and future trends

In this article, various techniques for monitoring the structural integrity of ship and offshore structures are discussed. Typical HMSs are introduced with common sensor locations as well as measurement parameters to be monitored. The long based strain gauge (LBSG) is typically installed at the midship and quarterdeck of vessels for monitoring hull girder strength. Ship motion and slamming pressures are measured by accelerometers and pressure sensors located at the bow of the vessel. Different requirements for the hull monitoring system are summarized for each classification society. Then a review of the introduction of emerging new technologies that are applicable for ship and offshore structures follows. The basic principles and application of FOS, AE sensors, fatigue CD sensors, vibration-based damage monitoring, and riser and anchor chain monitoring systems are reviewed. Finally, new requirements and future trends in structural health monitoring in marine industries are introduced. In particular, the importance of structural health monitoring technology that is applicable in cryogenic environment is indicated. Another important issue in structural monitoring is subsea system integrity monitoring. The integrity and safety assessment, as well as remaining life assessment, of deep sea pipelines and the related subsea systems for oil and gas exploration are of critical importance and need special attention. In conclusion, the structural health monitoring technique is an integral system for ensuring the safety of ships and offshore structures, as well as for protecting environment.

18.11 References

1. Haritos, H (2007), 'Introduction to the analysis and design of offshore structures– an overview', *Electronic Journal of Structural Engineering,* Special Issue: Loading on Structures. pp. 55–65.
2. Ersdal, G (2005), 'Assessment of existing offshore structures for life extension', Doctoral thesis, University of Stavanger, Stavanger, Norway.
3. Minerals Management Service (2004), *History of the Offshore Oil and Gas Industry in Southern Louisiana,* Outer Continental Shelf Study – MMS 2004–050, Minerals Management Service, Louisiana, USA.

4. Det Norske Veritas (2007), *Rules for Classification of Offshore Drilling and Support Units, Special Provisions for Ageing Mobile Offshore and Self-Elevating Structures,* OSS-101, Det Norske Veritas, Høvik, Norway.

5. NORSOK (1997), Condition Monitoring of Load Bearing Structures, NORSOK N-005, NORSOK, Oslo, Norway.

6. International Standards for Organization (2004), *Performance Parameters for Condition Monitoring of Structures*, ISO 16587:2004, International Standards for Organization, Geneva, Switzerland.

7. American Petroleum Institute (2007), *Structural Integrity Management of Fixed Offshore Structures,* ANSI/API recommended practice 2SIM (Draft), American Petroleum Institute, Washington, DC, USA.

8. American Petroleum Institute (2000), *Offshore Structures Design Practice*, API RP 2A, 21st edition. American Petroleum Institute, Washington, DC, USA.

9. Salvino, LW and Collette, MD (2009), 'Monitoring marine structures', in Boller, C., Fou-Kuo Chang and Fujino, Y. (eds), *Encyclopedia of Structural Health Monitoring*, John Wiley and Sons, Ltd.

10. Kim, M and Kim, D (2009), 'Ship and offshore structures', in Boller, C., Fou-Kuo Chang and Fujino, Y. (eds), *Encyclopedia of Structural Health Monitoring*, John Wiley and Sons, Ltd.

11. Ship Structure Committee (1997), *State of the Art in Hull Response Monitoring System*, SSC-401, Ship Structure Committee.

12. Health and Safety Executive (2009), *Structural Integrity Monitoring – Review and Appraisal of Current Technology for Offshore Applications,* RR685, Health and Safety Executive, Norwich, UK.

13. Gorman, M (2009), 'Acoustic emission', in Boller, C., Fou-Kuo Chang and Fujino, Y. (eds), *Encyclopedia of Structural Health Monitoring*, John Wiley and Sons, Ltd.

14. Mecon Limited (2005), *Cost Effective Structural Monitoring – An Acoustic Method, Phase II*, RR325, Health and Safety Executive, Norwich, UK.

15. Mecon Limited (2005), *Cost Effective Structural Monitoring – An Acoustic Method, Phase IIb*, RR326, Health and Safety Executive, Norwich, UK.

16. Doebling, SW, Farrar, CR, Prime MB and Shevitz, DW (1996), *Damage Identification and Health Monitoring of Structural and Mechanical Systems from Changes in Their Vibration Characteristics: A Literature Review*, LA-13070-MS, Los Alamos National Laboratory, New Mexico, USA.

17. Sohn, H, Farrar, CR, Hemez, FM, Shunk, DD, Stinemates, DW, Nadler, BR and Czarnecki, JJ (2004), *A Review of Structural Health Monitoring Literature 1996 – 2001*, LA-13976-MS, Los Alamos National Laboratory, New Mexico, USA.

18. Mojtahedi, A, Yaghin, MA, Hassanzadeh, Y, Ettefagh, MM, Aminfar, MH and Aghdam, AB (2011), 'Developing a robust SHM method for offshore jacket platform using model updating and fuzzy logic system', *Applied Ocean Research,* **33**, pp. 398–411.

19. Farrar, CR and Doebling, SW (1997), 'An overview of modal-based damage identification methods', *Proceedings of the International Conference on Damage Assessment of Structures*, International Conference on Damage Assessment of Structures, Sheffield, UK.

20. Loland, O and Dodds, JC (1976), 'Experience in developing and operating integrity monitoring system in North Sea', *Proceedings of the 8th Annual Offshore Technology Conference, Offshore Technology Conference*, Houston, Texas, USA, pp. 313–319.

21. Peters, K (2009), 'Fiber Bragg grating sensors', in Boller, C., Fou-Kuo Chang and Fujino, Y. (eds), *Encyclopedia of Structural Health Monitoring*, John Wiley and Sons, Ltd.

22. Ren, L, Li, HN, Zhou, J, Li, DS and Sun, L (2006), 'Health monitoring system for offshore platform with fiber Bragg grating sensors', *Optical Engineering*, **45**, 8, 084401-084401-9.

23. Kim, MH, Kim, DH, Kang, SW and Lee, JM (2006), 'An interlaminar strain measurement for insulation panels of LNG carriers', *Strain – An International Journal for Experimental Mechanics,* **42**, 2, pp. 97–106.

19

Sensing solutions for assessing and monitoring wind turbines

R. ROLFES, S. TSIAPOKI and M.W. HÄCKELL, Gottfried
Wilhelm Leibniz Universität Hannover, Germany

Abstract: This chapter focuses on structural health monitoring (SHM) for wind turbines. Repair and downtime costs can potentially be reduced with the continuous monitoring of structure and environment. Available SHM approaches for rotor blades and support structure are presented extensively, with a special focus on the sensor technology and the methods used. Important monitoring methods, such as modal-based methods, acoustic emission, and ultrasound wave propagation are presented. For monitoring of support structures, different foundation concepts are introduced with specific neuralgic spots. Subsequently, some of the global and local SHM approaches, as well as case studies are presented.

Key words: structural health monitoring (SHM), condition monitoring (CM), offshore wind turbines (OWT), blades, support structures, system identification, operational modal analysis.

19.1 Introduction

In the future, the demand for renewable energy will increase strongly, due to limited fossil resources. Beneath solar and hydraulic energy production, wind energy plays an important role within this 'green' energy sector. Wind turbines (WT) have been built for energy production over several decades onshore. Recently, offshore installations have increased greatly, representing a new and vital field of application. With the large number of onshore plants and a rapidly growing number of offshore installations, with weather dependent accessibility, the need for (remote) monitoring of these structures is evident.

On- and off-shore turbines have in common the ability to cover a large variety of engineering professions: beginning with the footing as piles or gravity foundations over large steel constructions, turbines, and generators, to the aerodynamic design and high performance materials of blades. A result is the variety of different monitoring fields. Figure 19.1 gives an overview of the monitoring nomenclature used for (offshore) WTs, differentiating

565
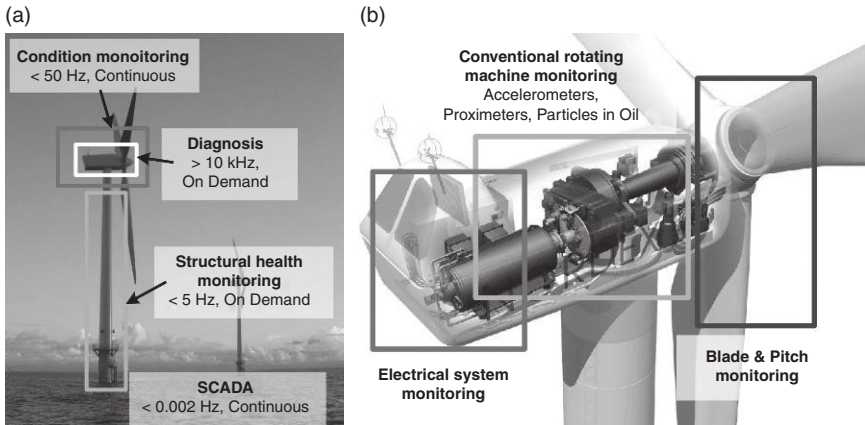
19.1 Nomenclature for (O)WEC monitoring with according frequency ranges (a) and subcategories for monitoring categories within the nacelle and rotor (b). (*Source*: Crabtree *et al.*, 2010) In some concepts, the electrical system is located in the lower part of the tower around the entrance level.

between condition monitoring (CM) for rotating machinery and electrical components, structural health monitoring (SHM) for the support structure and blades, and monitoring of environmental and operational conditions (EOC), also known as supervisory control and data acquisition (SCADA) data. In this case, the support structure is understood as the complete structure below the nacelle, independent of the type of structure being realized. Further, monitoring within the nacelle is separated into electrical system monitoring, rotating machine monitoring, and blade and pitch monitoring. Beneath this, SHM can be divided into global (vibration-based) monitoring and local monitoring of specific OWEC parts, e.g. scour monitoring or monitoring of grouted joint connections.

Remote monitoring of all the above-mentioned parts gains constantly in importance, since the costs of repair and standstill time are very high. In the domain of offshore WTs the financial losses can be even higher, as the turbines are accessible for repair only via an equipped ship. In Fig. 19.2 the annual failure rates and the corresponding down time in days are presented. Even for parts with a low annual failure rate, the down time may reach up to almost 8 days. Therefore, it is very important that damage is detected at a very early stage through constant remote monitoring. Concerning the rotor blades, visual inspections are still necessary for decisions about maintenance, repair, and shutdown, despite the existence of many different methods. Currently, no method has reached a high level of reliability and market readiness to reduce the need for visual inspections, which are potentially dangerous and require a shutdown of the WT.
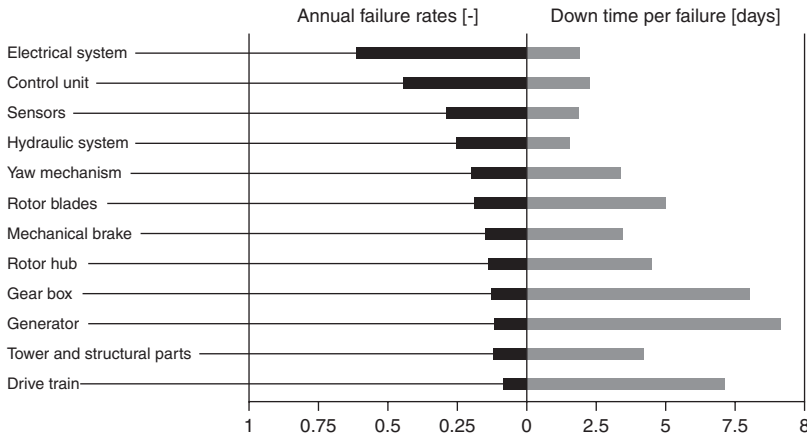
19.2 Annual failure rates of WT particles and corresponding down time in days. (*Source*: ISET, Germany, http://www.iset.uni-kassel.de/pls/ w3isetdad/www_iset_new.main_page.)

As shown in Fig. 19.1, monitoring rotating machinery (CM) is undertaken from 'conventional' fields of industry, as is monitoring of the electrical system where different systems have been introduced to the market. SHM at WT support structures, blades and pitch control in contrast, constitute rather new scientific fields. Hence, the following section will present recent developments for SHM and blade monitoring (see Sections 19.3 and 19.4) beneath a discussion of existing literature for CM/SHM and European research alliances (Section 19.2) before closing with a summary.

19.2   Review of offshore wind turbine (OWT) monitoring

The following section will give a short introduction to some of the standard literature in the field of WT engineering in general, and the monitoring of WTs in particular. Overall, four steps are distinguished in SHM/CM: damage detection, damage localization, damage quantification, and damage prediction (Rytter, 1993). It is also important to note that a 'sensor cannot measure damage' (Worden *et al.*, 2007),[1] hence damage-sensitive features have to be extracted. Stochastic environment and loading due to wind and waves influence system identification. Large variations in EOCs[2] make EOC monitoring almost as important as structural monitoring itself. For

---

[1] Basic knowledge on SHM over several decades is formulated in seven general axioms, accompanied by examples, building an essential paper in SHM/CM.

[2] Typical EOCs are: wind speed and direction, wave direction, height and spectra, temperature, nacelle position, rotor speed, pitch angle, and electrical output.

offshore turbines, each labor-based monitoring issue becomes even more critical, since distances between turbines have to be covered by ship, and finding weather windows with acceptable wind and wave conditions can be a very limiting factor. Hence, one overall goal in WT monitoring is condition-based maintenance, rather than preventive or corrective maintenance, to omit unforeseen damage.

19.2.1   Structural health monitoring (SHM)/Condition monitoring (CM) literature

The available literature concerning WTs in general takes into account many different aspects of the relevant technology. Tong (2010) focuses on power generation from WTs, and the design of all kinds of WTs as well as their components. Design techniques and typical approaches are presented. Gasch and Twele (2012) offer a more general approach, in addition to the classical design and analysis approaches, covering fields such as the historical development of windmills and the planning, operation, and economics of wind-farm projects. Burton *et al.* (2011) refer in addition to the component design and electric power system, especially to the aerodynamic parameters and the offshore WTs. The aim is to reflect the evolution of design rules and the principal innovations in technology. Finally, Harrison *et al.* (2000) follow a different approach from the above-mentioned literature sources and focus on the cost differences between WTs of different sizes and configurations, and on the techno-economics of this type of renewable energy.

General reviews of SHM were carried out by e.g. Chang *et al.* (2003) and Doebling *et al.* (1996, 1998), providing an overview of different methodologies. Sohn *et al.* (2007) give a broad overview of the influence of EOC-variability on SHM techniques, providing many examples. Major variations are temperature, boundary conditions, mass loading (e.g. traffic on bridges), and wind-induced vibrations. An important point presented in this paper, which also applies strongly for offshore WTs, is data normalization to compensate for variations in recorded data.

A comprehensive literature review describing the methodology and algorithms of CM and fault detection systems (condition monitoring systems, CMS and fault detection systems, FDS) can be found in (Hameed *et al.*, 2009). Here, FDSs are composed of measurement, signal processing, classification and action, seen as subcomponents of a CMS. Extensive examples are given for CM of local parts such as rotors, drive trains, pitch mechanisms, yaw systems, and electrical systems. Global FDSs are briefly described in connection with artificial intelligence (AI). Márquez *et al.* (2012) provide one of the most extensive literature reviews on CM, giving a compact overview of available CM types and signal-processing methods. The vast majority of reviewed literature concentrates on blades, rotor, gearbox, generator,

and bearings, not on the support structure. Nevertheless, this review marks a good starting point for a broad overview of the topic. Growth of wind energy and discussion of SHM/CM costs are found in (Butterfield *et al.*, 2009), and major challenges and demands are also pointed out. Liu (2010) analyses SHM for (offshore) wind-energy systems from the Chinese point of view, providing also the status of wind energy in China (the second largest market for offshore wind after Europe).

A survey of available CM systems on the market was prepared within the Supergen Initiative by Crabtree *et al.* (see (Crabtree *et al.*, 2010)), pointing out 'a clear trend toward vibration monitoring.' Discussion with the particular companies is included as well. Especially, company information and product details provide a good overview. A further review of available CM methods was carried out by Yang *et al.* (2012). In both cases, the support structure is addressed only marginally. Since both blade monitoring and, especially, SHM for support structures must be considered in a pre-commercial development stage, these two types are addressed in the two following sections. For SHM of support structures, local and global monitoring approaches are distinguished, since the methodologies are distinctly different.

19.2.2   European research networks and projects

Many recent research activities in Europe have focused on offshore wind energy. This section gives a short overview of some selected networks, providing also their web pages for further information. In general, research focuses on floating or bottom-fixed structures. Obviously, countries neighboring the North Sea's shallow parts focus on bottom-fixed structures, whereas countries with steeper coast lines, e.g. Norway, prefer floating solutions, even if their readiness level is distinctively lower. In total, major research fields can be split into the following categories:

- Environment and ecological impacts
- Support structures and design
- Operation and maintenance/reliability
- Logistics and installation
- Grid integration and electrical systems
- Aerodynamics and wake effects.

In many cases, research activities are bundled into research networks with different industrial and scientific partners. In Germany, the *Research at alpha ventus* [3] initiative uses a test field with 12 plants (six on tripods and

---

[3] REAVE, Germany, http://rave.iwes.fraunhofer.de/rave/pages/welcome.

six on jacket foundations) for data collection. The initiative is managed centrally by a coordination project, hosts a variety of research fields, and new national research partners can join the initiative. The Netherlands aims also for an offshore research wind park within the *Far and Large Offshore Wind energy*[4] project, and a second consortium is We@sea[5]. Floating offshore wind-energy plants are the focus of two Norwegian research networks, the *Norwegian Centre for Offshore Wind Energy*[6] and the *Norwegian Research Centre for Offshore Wind Technology.*[7] The Spanish *Centro Nacional de Energías Renovables*[8] focuses on renewable energies such as wind, solar, and biomass, and a floating offshore turbine is expected in 2013/14. Within *The Swedish Wind Power Technology Centre*,[9] which started in 2010, the complete design of WTs is addressed, while Vindval[10] focuses on environmental impacts. Two doctoral programs closely linked to the industry are running in the United Kingdom at the *Centre for Doctoral Training in Wind Energy Systems*[11] and *Industrial Doctorate Centre in Offshore Renewable Energy.*[12] *The Offshore Wind Accelerator*[13] is another large project, with a volume of about 49 mio.€; it aims at market-pull innovations to reduce energy costs. Finally, seven academic and 19 industrial partners form the *SUPERGEN Wind Energy Technologies Consortium*,[14] which is supposed to last from 2009 to 2018.

19.3   Structural health monitoring (SHM) for blades

In the current chapter the structural design of rotor blades, the materials used for their construction, as well as the manufacturing process are described, and finally some of the most commonly occurring damages are mentioned. Subsequently, structural health monitoring methods used in rotor blades, such as modal approaches and acoustic emission, in combination with the corresponding sensors required for each method, are

---

[4] FLOW, Netherlands, http://tudelft.nl/en/research/knowledge-centres/duwind/flow/.

[5] We@Sea. Netherlands, http://www.we-at-sea.org/.

[6] NORCOWE, Norway, http://www.norcowe.no/.[7]NOWITEC, Norway, http://www.sintef.no/projectweb/nowitech/.

[8] CENER, Spain, http://www.cener.com/es/index.asp.

[9] SWPTC, Sweden, http://www.chalmers.se/ee/swptc-en.

[10] Vindval, Sweden, http://www.naturvardsverket.se/SamarbetswebbplatserSv/Vindval/About-Vindval/.

[11] CDT, United Kingdom, http://www.strath.ac.uk/windenergy/exampleprojects/.

[12] IDCORE, United Kingdom.

[13] OWA, United Kingdom, http://www.carbontrust.com/our-clients/o/offshore-wind-accelerator.

[14] SUPERGEN Wind, United Kingdom, http://www.supergen-wind.org.uk/.

presented. Additionally, for the sake of completeness some non-destructive testing methods are mentioned. There follows a description of some of the available rotor blade monitoring systems in market and their function principles. Finally, the last part of the chapter focuses on the research performed in this field.

19.3.1   Blade structure

Within the past few years the design of WT blades has developed very rapidly, having now achieved a maximum length of 75 m on operating WTs, while an 83.5 m rotor blade has already been manufactured. The structure of a rotor blade can be separated into two components: the shell and the beam. The shell, which consists of the upper and lower shell, has an aerodynamic shape, and the beam may have the form either of a box girder or of two laminated webs that offer shear stiffness and buckling resistance (Tong, 2010). The aerodynamic shape of the rotor blade should be optimal in order to efficiently convert the aerodynamic loads to motion. The box girder or the webs on the other side serve as a load-carrying beam that transmits the aerodynamic forces to the WT hub. In the longitudinal direction of the blade, the shape changes from an aerodynamic shape at the outer part of the blade to a cylindrical shape at the root.

Nowadays, most blade manufacturers use E-Glass fibers for the rotor blades, and more specifically for the upper and lower shell. The need for lighter constructions leads to an increasing use of carbon fiber, which is characterized by very high stiffness. However, due to the high cost, carbon fiber is used in hybrid carbon/glass laminates for highly stressed areas, for example in the spar flanges reinforced with carbon fiber in the principal stress direction. In addition, wood or other natural materials are used for the shells or the internal sandwich structures (Hau, 2006). As far as the resins are concerned, epoxy has lately gained popularity, as against polymer resins, due to its better strength properties. Finally, gelcoats are used on the outer surfaces of the blades to protect them from exposure to environmental conditions, such as humidity or UV-radiation.

Processing methods used for composite materials, such as lamination of pre-pregs and vacuum-assisted resin transfer molding (VARTM), are required during the manufacturing of a rotor blade. Initially, the shells, which are light-weight sandwich structures, are manufactured. The fibers or pre-pregs are manually placed on an open mold. Before curing, the mounds are closed and a pump is used to create vacuum conditions. Thus, air inclusions are significantly avoided and the dangerous exhalations of the resin reduced. Afterwards, the mounds are separated so that the webs or box girder is placed between the upper and lower shell. Finally, the upper and lower shells are bonded with the webs or box girder, and the two shells are

bonded to one another. The bondlines that occur from this bonding procedure are positions with a high possibility of damage (debonding).

Damage that derives from imperfections during manufacturing processes, such as bondline failure and web debonding, may severely affect the structural integrity of the blade. In addition, delamination or damage due to overload and fatigue are also among the most often encountered cases. Therefore, CM is absolutely necessary to prevent such incidents and reduce the costs of repair and standstill time.

19.3.2   Methods and sensors for SHM on rotor blades

Passive systems do not require addition of energy for their operation. Passive sensors are used to 'listen' to the structure but do not interact with it (Giurgiutiu, 2005). Modal approaches, acoustic emission (AE), strain measurement methods, and impedance tomography are some CM methods that belong to this category. On the other hand, active methods, such as ultrasonic wave propagation, interact with the structure. Additionally, some nondestructive testing methods, such as thermography and laser vibrometry, are mentioned. The main difference between SHM methods and nondestructive methods is that in the former the sensors are integrated into the structure.

Modal approaches

Modal approaches are among the most common and conventional methods for monitoring rotor blades. All elastic bodies, when stimulated, vibrate at their natural frequencies, so that any changes of characteristics or of geometry cause changes to the eigenfrequencies of the blade. By observing the vibration of the blade, the natural frequencies are determined and any changes in the blade's spectra monitored indicate damage. The sensors required are accelerometers, piezoelectric sensors, or microelectromechanical systems (MEMS) (Boller *et al.*, 2009). Modal parameters characterize the global behavior of the structure, and it is therefore difficult to extract information about small damages that are a local incident. Often the modal-based methods have to be supplemented by finite element analyses to locate and quantify the damage (Wang and Deng, 1999; Mattson and Pandit, 2006).

Acoustic emission (AE)

The AE method is also one the most commonly used for monitoring rotor blades. Any kind of damage, such as cracking or debonding, causes elastic waves in the material, which may be captured with piezoelectric sensors.

The AE signals obtained, when classified according to amplitude and energy content, may give information about the kind of damage (Dutton, 2004). AE sensors have high sensitivity, and many types of damage can be detected and located even before they become visible (<1 cm) (Sørensen, 2002). Nevertheless, a large number of sensors are required to localize damage in a big structure, such as a rotor blade. As a result, there are a high number of outputs and cables. Additionally, the background noise caused by a WT in operation introduces difficulties to the signal processing and the association of the signals to damage types.

Strain measurement

The measurement of strain is also used in rotor blades for monitoring. The sensors used are typically foil strain gauges or fiber optical strain (FOS) gauges that are placed in critical areas of the blade. The insensitivity of FOS gauges to lightning has led lately to an increase in their use. Fiber optic Bragg gratings, which can be used as sensing elements in optical fiber sensors, may provide a large number of low-bandwidth strain measurements (Boller *et al.*, 2009). Also in this case, a high number of sensors are required to detect small damages.

Impedance tomography

Impedance tomography is a method that uses the conductive properties of carbon nanotubes, or other conductive particles such as CFRP, for detecting damage. The impedance principle in particular may be used for detecting damage caused by fatigue. Fatigue damage causes the reduction of stiffness, and consequently the electrical resistance increases (Seo and Lee, 1999). In this case, the sensor used is actually the carbon nanotube or any other conductive particle. A disadvantage of this method is its inapplicability to the rotor blades of an operating WT.

Ultrasonic wave propagation

One of the most commonly used active systems for monitoring WT rotor blades is the ultrasonic wave propagation method, which uses piezoelectric sensors and, more specifically, actuators or transmitters and receivers. The actuators transfer ultrasound waves into the material, which are then captured from the receiver. The existence of damage between the transmitter and the receiver will cause change to the waves' characteristics. Ultrasound wave propagation is capable of revealing planar cracks, such as delaminations, which are oriented perpendicularly to the direction of sound wave propagation (Ciang *et al.*, 2008). The localization of damage, taking into account the time delay of the returned signal, cannot be accurate for

complex structures, such as rotor blades. However, ultrasonic wave propagation is a promising method for damage detection.

Thermography

Thermography is a method to map the temperature distribution within the rotor blade with the assistance of an infrared sensor or camera (Meinlschmidt and Aderhold, 2006). The frequent change of loads produces temperature differences in the blades, indicating stress redistribution. In addition, small cracks produce friction due to cyclic excitation. Nevertheless, it is difficult to use this method in operation, i.e. on rotating rotor blades, so that it can detect interior damage.

Laser ultrasound or laser vibrometry

This method requires no contact with the structure. The sensor used is a laser, while the operating principle is based on the excitation of the structure by a laser beam pulse. The laser also measures the response of the structure. However, the disadvantages of this method are the difficulty of use in the field, and the fact that the laser excitation may cause minor damage to the surface of the rotor blade (Ciang *et al*., 2008; Boller *et al*., 2009).

19.3.3   Monitoring systems in market

In the following section some selected monitoring systems for rotor blades are presented. In every case the operational principle, the required sensors, and their positions are described. The monitoring systems used in WT blades under operating conditions are mostly a combination of the above-mentioned methods.

BLADEcontrol (Bosch Rexroth)

BLADEcontrol is an online health monitoring system that uses global parameters. The operating principle is based on modal approaches. Any change of the frequency characteristics from the initial recorded state can be interpreted as damage (Bosch Rexroth[15]). Rotational frequency and temperature are both taken into account, as a description of the operational state of the blade and the environmental conditions. BLADEcontrol also detects ice-formation on the blades, by the decrease of eigenfrequencies, with a limit of detection of 0.04% of the blade's mass. BLADEcontrol uses a two-dimensional piezoelectric accelerometer inside the blade located one-third of the blade's length from the root. This is supported by a one-dimensional accelerometer located at the shaft, and a rotational trigger sensor and

---

[15] Bosch Rexroth Monitoring Systems GmbH, Germany, http://www.boschrexroth. de/windenergie.

a temperature sensor element placed inside the blade. The measurement frequency is 1000 Hz, and the system measures constantly.

MOOG Blade Sensing System

The second system available on the market is MOOG Blade Sensing System, which currently is focused on load control. Active load control or individual pitch control (IPC) uses real-time data to adjust the pitch of the turbine blades. Fatigue loads or peak loads are measured at the roots of the blades, and are combined into a resultant force that measures the imbalance. This measurement is used to adjust the pitch angle of the blades to remove the load imbalance, reducing the need for load monitoring. A sensor array containing four fiber optic strain sensors, with four collocated fiber optic temperature compensation sensors, is installed in each blade at the circumference of the root area, measuring in the longitudinal direction (Fig. 19.3). Additionally, optical interconnection cables are used (one per array) to connect the sensor arrays to the measurement unit located in the turbine hub.

SKF Blade Monitoring System

SKF Blade Monitoring System can be classified with health monitoring systems using global approaches. The operation is based on a laser placed

Fiber sensor interrogator unit in turbine hub

Fiber optic root sensor array
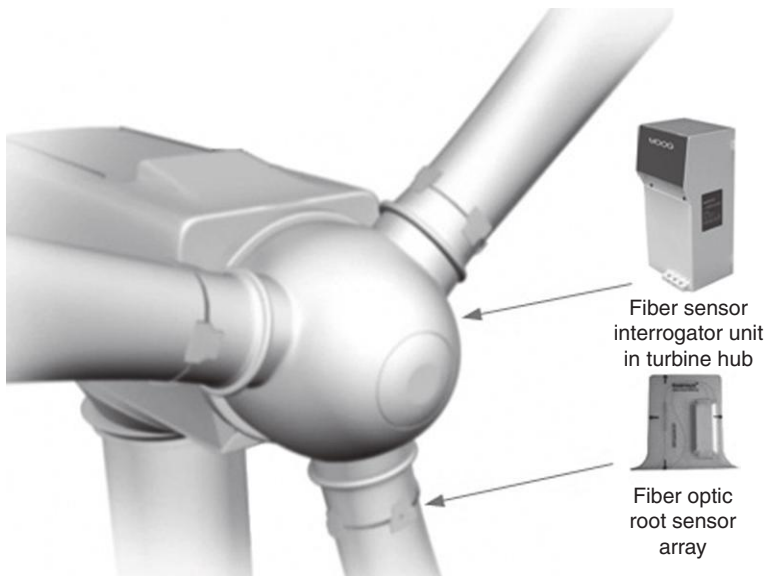
19.3 Instrumentation of the MOOG rotor sensing system. (*Source*: MOOG.)

on the top of the nacelle, beside the onboard meteorological station, which measures the deflection amplitude of each blade's tip at the same fixed position during rotation. Each pass of the blades is recorded and is related to the measurement data obtained from the meteorological and turbine control regarding the parameters of wind speed, rotational speed, and pitch angle. Thereby, the current behavior of the blade is compared to its own historical behavior once stored, or to design amplitude, as well as to the behavior of the other rotor blades. The outdoor laser has a range of 1–300 m, a sampling frequency of 1 kHz and a measuring accuracy of 0.1%, which is highly dependent on the stiffness of the installation frame and nacelle top.

LM Wind Power–LM blade monitoring

LM Blade Monitoring may be classified in the field of load monitoring systems (Zerbst, 2011). This monitoring system is a crack detection system based on a set of optical fibers embedded in the rotor blade. More specifically, three optical fibers are embedded in the composite laminate at strategic areas, such as the trailing edge (Söker[16]). The light flow inside the fibers is permanently controlled, giving information about the health status of the rotor blade. When a crack is initiated and propagates through the trailing edge, it will result in cutting the fiber located closer to the trailing edge. A disadvantage of this system is that when the blade is damaged, the fibers are broken and consequently the sensor is lost, while there is no possibility of reinstallation.

19.3.4   Research review

Although some monitoring systems are available in the market, visual inspections are still necessary for monitoring the rotor blades, since the existing systems have not yet reached a high state of reliability. BLADEcontrol system is efficient in using only two accelerometers, but the use of electric sensor technology endangers the structure from lightning strike. Additionally, it is disputable whether frequency-based methods are sensitive enough for damage detection. The MOOG Blade Sensing System, on the other hand, is focused more on load control than on damage detection. As far as the SKF Blade Monitoring System is concerned, the measurement accuracy is highly dependent on the quality of the fixing of the system on the nacelle. Measurement errors may occur if there is relative movement between the

---

[16] Söker H., Berg-Pollack Antje, Kensche C., Rotor Blade Monitoring-The Technical Essentials, http://www.dewi.de/dewi/fileadmin/pdf/publications/Publikations/11_1_Soeker.pdf.

laser and the nacelle. At the same time, comparable operational conditions for a certain pitch control angle should be available. Finally, the LM Blade Monitoring System does not allow any reinstallation after damage, and may detect damage only once.

Hence, the need for further research on monitoring of rotor blades. The development of methods that are more sensitive to damage than the modal-based approaches is a major challenge. Additionally, components that endanger the rotor blades from lightning strike, such us cabling or metallic parts, should be avoided. The use of optical sensors is a solution to this problem. Finally, another parameter that should be taken into account is the number of outputs. In this section, some selected research projects are presented.

ISD Proportionality method approach for SHM of rotor blades

A new method for SHM on WT blades has been developed at the Institute of Structural Analysis (ISD[17]), Leibniz Universität Hannover, Germany. The advantages of this system are more sensitive and earlier reaction to damage in comparison to eigenfrequency-based methods, as well as insensitivity to lightning strike, since no metallic components or wiring is included.

The new approach is based on the extension of Gasch's proportionality method, according to which, under harmonic excitation maximum oscillation velocity and maximum dynamic stress are proportional by a factor (Zerbst *et al.*, 2011). This factor is referred to as the proportionality factor, and it describes the structure's dynamic behavior (Equations [19.1] and [19.2]). Hence, a change in the proportionality factor can be used as damage indication.

$$\sigma_{dyn,\max} = p_{\text{system}} \cdot \left| v_{\max} \right| \qquad [19.1]$$

$$\text{where} \quad p_{\text{system}} = \sqrt{E\rho} \cdot p_{\text{profile}} \cdot f_{\text{profile}} \cdot p_{\text{boundary}} \cdot p_{\text{excitation}} \qquad [19.2]$$

Maximum stress values occur at the root of a rotor blade, while the maximum oscillation velocity is at the tip. Nevertheless, it has been proven that the proportionality is also valid for velocities measured in the middle of the rotor blade.

For the application of the proportionality method to rotor blades, a new sensing concept was developed. The sensors used are a four-sensor chain of FOS gauges, placed at the root area of the blade and oriented

---

[17] Institut für Statik und Dynamik, Leibniz Universität Hannover, Germany, http://www.isd.uni-hannover.de/.

in the longitudinal direction, in combination with two novel sensor prototypes measuring deflections at parallel positions, on the webs of the blade (Zerbst, 2011). The locations of the sensors in the rotor blade are schematically presented in Fig. 19.4. The deflection sensor is specifically designed for use in WT rotor blades. The strain measurements at the root occurring from the fiber optic sensors are used for determining the maximum stress values, while the measured deflections on the inside of the blade are differentiated and used for determining the maximum oscillation velocity.

More specifically, as far as the new deflection sensors are concerned, a glass fiber string is connected to the web attached to a single fixation hook, far inside the blade (sensing location). This string is guided along the web to the root section, where the active sensor unit is also fixed to the web (sensor location). Both active and passive units are shown in Fig. 19.5. Deflection of the blade in the sensing location is possible in both the edgewise and flapwise directions, and this results in a change of the string angle at the sensor location, which is located close to the root section. This angle can be related to the deflection amplitude, since the string tension is constant after the initial calibration process. The basic sensing system is fixed to the active unit. Directly behind the location of string fixation to the web, two orthogonally oriented load cells are carefully connected to the string. A change of string angle is recorded as voltage per millimeter (V/mm). The deflection sensor installed inside a rotor blade, and the load cell connection to the string, are shown in Figs 19.6 and 19.7, respectively.
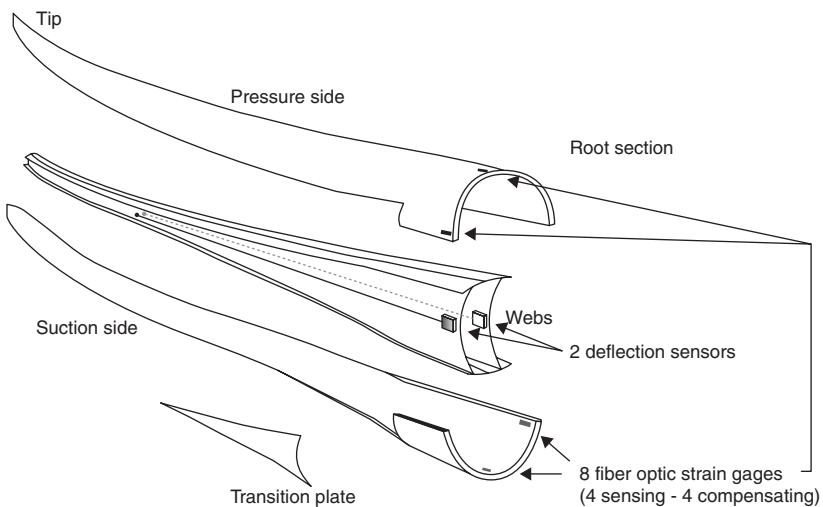
19.4 Schematic sketch of the sensor locations used in the ISD proportionality method approach (Zerbst, 2011).
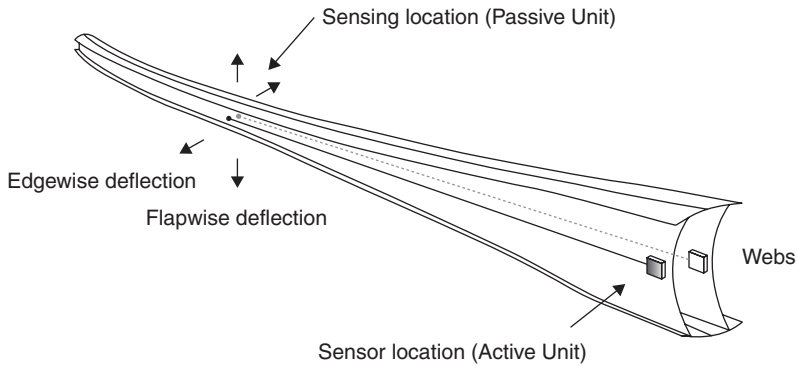
19.5 Locations of active and passive units of the deflection sensor inside the rotor blade (Zerbst 2011).

19.6 Active sensor unit of deflection sensor fixed to fiber patches inside a rotor blade (ISD).

Nevertheless, it appears that an extension of the proportionality method for the quantity of acceleration will also allow its use for the application of fiber optical accelerometers, instead of the new deflection sensor. Thus, the method may be applied to data deriving from a monitoring system equipped with widely used sensors, such as FOS gauges and fiber optical accelerometers.

The method, along with the novel sensing concept, was applied to a real 50.8 m blade during a fatigue test in the edgewise direction in a test facility in Aalborg, Denmark. During the test, damage 1.5 m in length was induced on the upper trailing edge bondline, which further propagated to reach a length of 2 m. Both the initial damage and its propagation were successfully
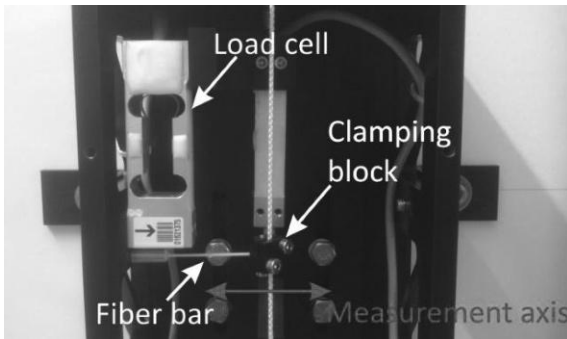
19.7 Load cell connection to the string (ISD).

detected by decrease of the proportionality factor. In Fig. 19.8, the decrease of the proportionality factor due to damage at the trailing edge is presented. After the propagation of damage to 2 m, the fatigue test was stopped to perform repairs, and then restarted. This decrease coincided significantly with the decrease of the factor calculated from numerical analyses, which were simultaneously conducted for several damage scenarios. The results of the numerical analyses conducted on a rotor blade have shown that the proportionality method is more sensitive to damage than the modal-based methods, since the decrease of the proportionality factor was much more significant than the decrease of eigenfrequencies. Thus, the application of the proportionality method and of the novel sensing concept is more sensitive to damage than the modal-based methods. Finally, the sensing concept and the method were applied on a 50.8 m rotor blade of a 3.3 MW WT under operation for almost 4 months. During this period, no damage occurred. However, the recorded signals indicate that the method can be applied to some operational cases.

The new approach is being further examined and evolved. The improvement in sensitivity of the method is being studied and, at the same time, appropriate signal processing is being pursued, in order to extend the application field of the method. A very important addition to the system is the application of optical microphones to verify and reinforce the structural monitoring system and help prevent false alarms.

Wölfel SHM.Blade

SHM.Blade System is developed by Wölfel[18] and consists of the combination of three different, independent measurement methods: AE, acousto-ultrasonics,

_____

18 Wölfel, SHM.Blade, Germany, http://www.woelfel.de/produkte/windenergie/shm-blade.html.
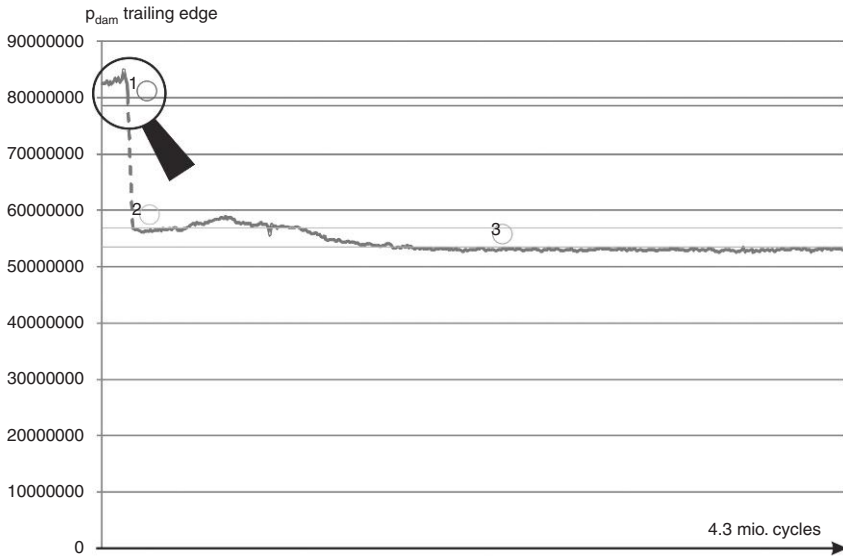
19.8 Proportionality factor time history during edgewise fatigue test (Zerbst, 2011).

and operational modal analysis. In the first approach, the piezoelectric ceramic sensors, which capture the elastic surface waves caused by the cracking of the glass fibers in the form of strain, are placed between the lower laminate layers inside the rotor blade. Alternatively, they may be placed on the walls on the inside of the blade. The second subsystem uses ultrasound waves that are actively sent from the rotor blade to the ceramic sensors. This method is appropriate for the monitoring of the bondlines and the aerodynamic shells, where smaller cracks occur. The third subsystem uses operational modal analysis approaches, with conventional acceleration sensors. The combination of the three methods results in a large amount of sensors and cabling.

Structural neural system (SNS)

SNS is inspired from the human neural system (Boller, Chang and Fujino, 2009). It uses AE monitoring for detecting damage, such as fiber breaking or delamination in composite materials. The functional principle is based on a sensor network, consisting of continuous sensors that form a neuron that are subsequently connected to an analog processor. In Fig. 19.9 each row and column forms a different neuron, while the neurons are not connected to one another. Thus, there is only one output per neuron. The output of the neurons is connected to the SNS analog processor (SNSAP), reducing the number of data acquisition channels required. Afterwards, two channels
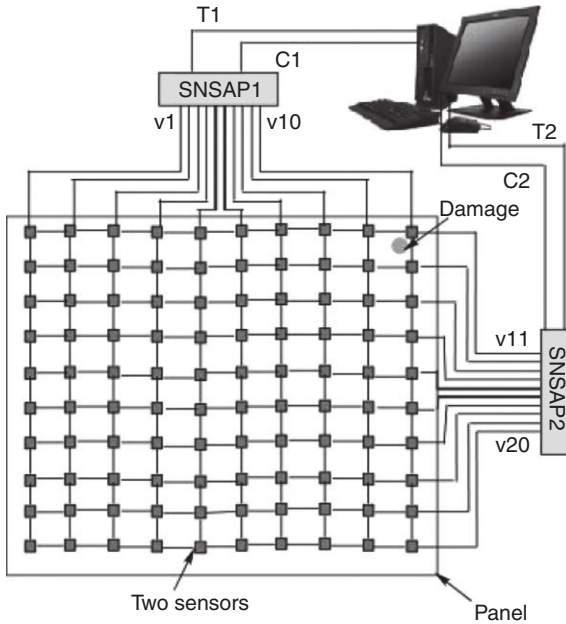
19.9 A sensor array example of the SNS system (Kirikera).

are created for predicting the location and severity of the damage. Kirikera *et al.* implemented the SNS on a 9 m long WT blade to identify the location of propagating cracks during a quasi-static test (Fig. 19.10) (Kirikera, 2006). This system may be applied only close to the root of the rotor blade, to protect against lightning strike.

Inspection robot

An alternative approach for monitoring rotor blades is the Rope-Climbing WT Inspection Robot (RIWEA), which was developed by Fraunhofer IFF.[19] The concept is especially suitable for offshore WT, where optical inspection is quite dangerous. The robot is made of glass fiber reinforced plastic, and is equipped with an infrared radiator, a high-resolution thermal camera, an ultrasonic system, and a high-resolution camera. The infrared radiator conducts heat to the surface of the rotor blades, while the high-resolution thermal camera records the temperature pattern. The inspection robot can replace visual inspection, but does not offer continuous monitoring of the rotor blades, since it cannot be used during operation.

---

[19] Fraunhofer IFF, Germany, http://www.fraunhofer.de/de/fraunhofer-forschungsthemen/energie-wohnen/windenergie/Roboter-RIWEA.html.

Table 19.1 Methods for SHM of wind turbine rotor blades

Method	Sensors	Description
Frequency-based method-vibration	• Accelerometers • Piezoelectric sensors • Microelectromechanical systems (MEMS)	• Global monitoring – passive method • Detection of small damage is difficult • Supplemented by FE models for damage localization and quantification • Sensitivity is disputable
AE	Piezoelectric sensors	• Local monitoring – passive method • Large number of sensors required for the localization of damage in a big structure • Big number of outputs and cabling • Appropriate for smaller structures
Ultrasonic wave propagation	• Piezoelectric wafer • Actuators/transmitters and receivers	• Local monitoring-active method • Damage localization is difficult
Strain measurement	• Typical foil strain gauges • FOS	• Global approach-passive method • Placed in critical areas of the blade • Detection of small damage is difficult
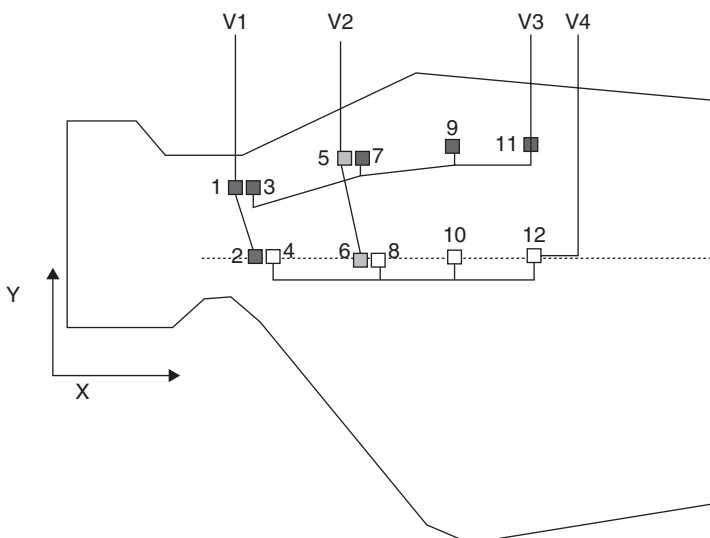
19.10 Sensor locations on the blade for the experimental setup (Kirikera).

Table 19.2 Systems for SHM of wind turbine rotor blades

System	Method	Sensor	Description
BLADEcontrol	Frequency-based method	Two accelerometers	• Damage and ice detection • Available in market
MOOG RMS	Load control	Four FOS gauges per blade	• Load monitoring and load control • Available in market
SKF Blade Monitoring System	Blade deflection	Laser	• Highly dependent on environmental conditions • Not widely used
LM Blade Monitoring	Breakage of embedded fiber	Embedded fiber	• Sensor can be used only once-cannot be replaced • Low cost • Not widely used
ISD SHM Proportionality Method Approach	Proportionality method	• Four strain gauges • One deflection sensor or fiber optical accelerometers	• More sensitive than frequency-based methods • Insensitive to lightning strike • At optimization stage
Wölfel SHM.Blade	• AE • Acousto-ultrasonics • Operational modal analysis	• Piezoelectric sensors • Compatible strain gauges	• Big number of sensors required • Available in market
SNS	AE	• Piezoelectric sensors	• Can only detect damage close to the blade root • At testing stage
Inspection robot	• Thermography • Ultrasonic wave propagation	• Infrared radiator • High resolution thermal camera • Ultrasonic system • High resolution camera camera	• Non-applicable during operation • At design stage

The aforementioned methods and sensors are briefly presented in the following tables. The most important methods used for the SHM of rotor blades are presented in Table 19.1, while Table 19.2 includes some of the SHM systems for rotor blades that are available or under research and optimization.

19.4   SHM for WT support structures

Support structures represent an important part of a WT; especially for offshore turbines, investigations into structures and installation comprise a large part of total costs. This chapter will present some selected local monitoring approaches, as well as fundamentals and current research efforts on SHM for WT support structures. In contrast to Chapter 3, no commercial solutions for support structures are presented. First, for SHM it is important to understand the development of WT structures during the last decades to get an impression of their dimensions and characteristics.

19.4.1   Wind energy development

Onshore wind-energy plants have been subject to a constant transformation in size and power output. Engine-power classes shifted continuously from below 500 kW in the early 1990s, from 2–3 MW through the last decade to above 3 MW in recent years. Accordingly, average rotor diameters have grown from 20–30 m in 1990 to 80–90 m in 2010, while average nacelle heights have grown from 60 m to 140 m today. As a result of the long development span, most plants below 2 MW have operated for 10 years or longer, providing a large demand for monitoring toward the end of their assumed life-time of 20–25 years.

For offshore wind-energy plants, development history is briefer. Today, ten European nations and China produce power with offshore wind energy. There are 62 offshore parks operating worldwide; 53 are located in Europe and nine in China. In total, 1579 plants achieve an output of about 4000 MW, 3392 MW of which is in Europe. In 2011, the year with most offshore installations so far, 126 farshore (more than 5.5 km from coast line) and 129 nearshore plants, with output of 700 MW, were erected. The 2–5 MW plants are in the majority, while plants larger than 5 MW have only recently occurred. In 2001 the average power output was 1 MW, while today it is 2.5 MW. The mean nacelle height is about 80 m (from sea level), which is smaller than for onshore plants due to lower surface roughness and a more constant wind profile. In 2001, the mean distance to shore was already 10 km with a mean water depth of 8.4 m. In 2011 the mean distance increased to 13.2 km and the mean water depth was 14.5 m. Germany has the most

remote plants, with a mean of 51 km distance to shore and 28 m water depth.[20]

The following sub-section will give a short overview of different support structure concepts for on- and off-shore plants, pointing out hot spots and neuralgic points for different structures. It is followed by sub-sections addressing local SHM approaches and global monitoring of the support structure.

19.4.2   On and offshore concepts for WT support structures

In general, WTs have a rather slender and flexible structure, with a high head mass due to nacelle and blades, while the fixed-end degree depends strongly on the foundation type. Large uncertainty in soil parameters, especially under drained conditions, and the aerodynamic behavior of the rotor still complicate numerical modeling, in particular for dynamic purposes addressed in monitoring tasks. Onshore, most turbines are based on lattice towers or monopiles on concrete foundations with ground anchors. Some older turbines with smaller nacelle heights are based on guyed towers. For offshore turbines, a variety of support structures exist, see e.g. Fig. 19.11 – while gravity foundations and monopiles are used for nearshore parks with less water depth, jacket foundations are used for deeper seas as well as tripod and tripile solutions. Floating structure, as tension leg platforms (TLPs), have only been built for research reasons. Each support structure type comprises its own characteristics and damage scenarios. Accordingly, local monitoring tasks will strongly depend on the investigated structure, while global approaches can possibly be applied to different foundation types.

Grouted joints, serving as connections for driven piles for offshore structures, were adapted from the oil and gas industry. Monopiles and tripiles are connected to the tower above sea level, while jackets and tripods use grouted joints right above the seabed to connect the piles to the support structure. Since the loading behavior differs strongly between large oil and gas platforms, the long-term behavior of these connections is of high interest.

Welded connections occur at many different locations at on- and off-shore structures. Due to the many welded nodes, jacket structures play an outstanding role here. Welding seams are considered to be potential starting points for cracks; fatigue could also be a problem here. Depending on the specific design, so-called hot spots within the structure with maximal

---

[20] More detailed information and extensive illustrations on the development of on- and offshore wind energy can be found in (Pfaffel *et al.*, 2012).

| Monopile | Gravity | Jacket | Monopile | Gravity | Jacket |

| Suction Bucket | Tripod | Tripile | Spar | Barge | TLP |
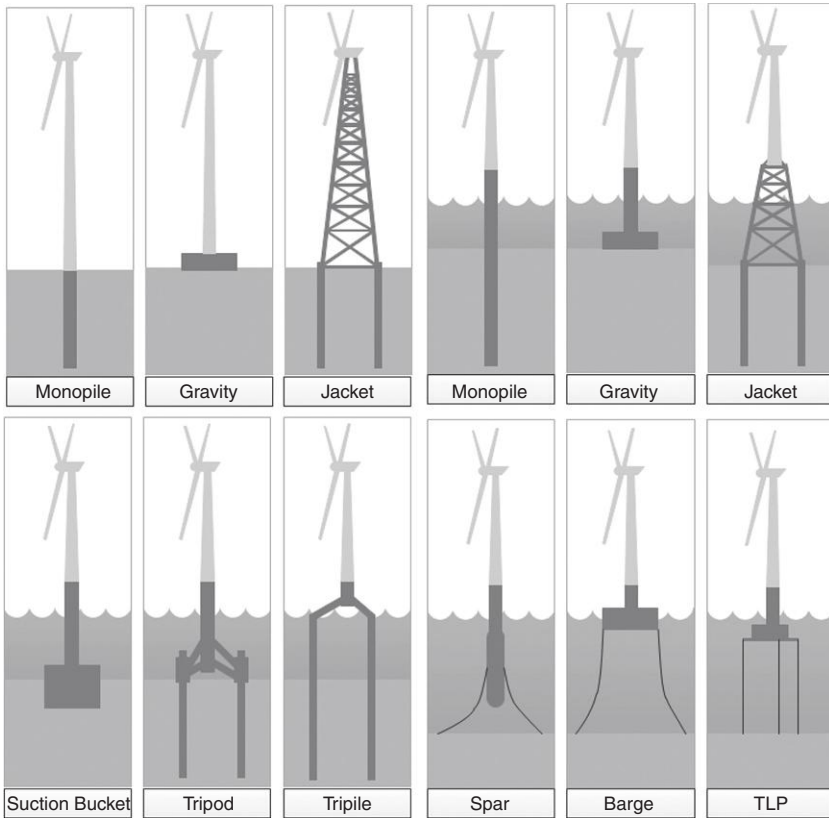
19.11 Different support structure types for on- and off-shore turbines.

strain amplitudes form areas of special interest, where damage could have potentially more serious consequences. For example, each steel tower typically consists of multiple sections, connected with a large number of bolts. Concrete towers, so far built only onshore, also consist of different sections of precast or *in situ* concrete. These connections must be considered critical points and are of high interest for monitoring approaches.

19.4.3   Local SHM approaches for support structures

Fatigue monitoring

As mentioned above, fatigue can be a critical issue at many locations of the support structure, since dynamic loading plays a major role for WTs. Commonly, strain gauges or acceleration sensors are used to collect structural responses. Sampling rates must be high enough to cover global and

local mode shapes that dominate motions. Stress cycles are analyzed by rain flow counting and linear damage accumulation (Palmgren-Miner).

If EOCs are collected and classified, fatigue can be assigned to these groups. Rabe (in Rolfes *et al.*) uses strain measurements from the offshore test field *alpha ventus* (Research at *alpha ventus* (RAVE), Germany n.d.) to calculate fatigue loadings per EOC class and forming a distribution of fatigue. Doing so, only by measuring the EOCs, fatigue can be simulated for future datasets. A Markov-chain–Monte-Carlo method is used in addition, to predict future EOCs from measured past EOCs. These simulated EOCs are again taken to simulate future fatigue loading. Results for simulations over a period of 1 month showed deviation of less than 10% compared to measured fatigue.

Scour monitoring

For offshore structures, development of scours at the seabed is a very important phenomenon, which is still a subject of research. Many studies have been carried out analyzing monopiles under currents and waves, an extensive review of which can be found in (Zanke *et al.*, 2011). However, interaction within pile groups, as present for triples, jackets, and tripods, remains a challenging task for engineers. Depending on the pile diameter, scours of several meters can occur and affect the structural behavior of an offshore WT.

Monitoring is realized with beam echo sounders, which are installed directly on the support structure or are vessel-based. Figure 19.12 shows a beam echo sounder installed at a tripod foundation within *alpha ventus*. Here, extensive monitoring was performed for research purposes. For pile groups, the development of local and global scours could be monitored. There are strong scour dynamics caused by storms and extreme environmental conditions.

Monitoring of grouted joint

Grouted joints form a neuralgic point in such different support structure concepts as mono- and triples or jackets and tripods. Generally, piles are driven into ground, and the sleeves of the support structures (tripods, jackets) are connected with an injected high performance compound. Through this, possible skew positions of the piles can be balanced. Accordingly, all loads must be transferred through the grouted joints. For monopiles, the single pile is connected to the tower with a grout above sea level. The connection technique was adapted from the oil and gas industry, where it is used for large offshore platforms. Nevertheless, loading behavior for WTs differs strongly from large offshore platforms, since the dynamic loading is very high.

General and longtime behavior of these connections is of great interest for the offshore industry. Within GIGAWIND *alpha ventus* (Rolfes *et al.*,

(a)

(b)
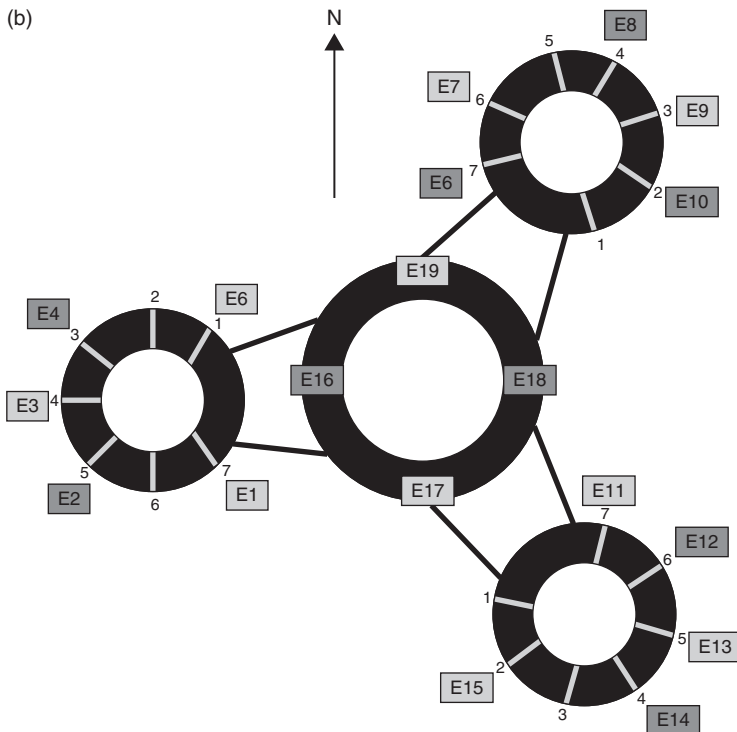
19.12 (a) Installation of beam echo sounder at tripod foundation in *alpha ventus*; (b) position of echo sounder. (*Source:* Lambers-Huesmann and Zeiler, 2010)

2012), a German research project, a prototype of a displacement measuring system was developed and mounted to a tripod foundation (see Fig. 19.13). The system is mounted onto pile and sleeve via magnets and is suitable to be adapted to varying geometries. Horizontal displacement transducers can collect unique data and verify movements in correlation with strain measurements and waves. In addition, temperature sensors and a leakage sensor were installed within the oil-filled box, which is protected by two membranes.

Splash zone monitoring

The splash zone is understood as the region between water and air, where tidal changes occur and waves hit the structure. This zone is interesting for monitoring for two reasons: firstly, this area provides an optimal environment for corrosion and secondly, (breaking) waves introduce a major loading component. Both aspects have also been covered within GIGAWIND *alpha ventus* (Rolfes *et al.*, 2012).

Corrosion monitoring was realized through test specimens mounted offshore over several years (See Fig. 19.14c). They were located at different elevation levels within and above the tidal influence. A major risk for corrosion protection for WTs, after corrosion itself, is collision with supply vessels.

Wave monitoring for breaking waves and diffraction effects was realized through 30 pressure cells mounted at three different levels (See Fig. 19.14a
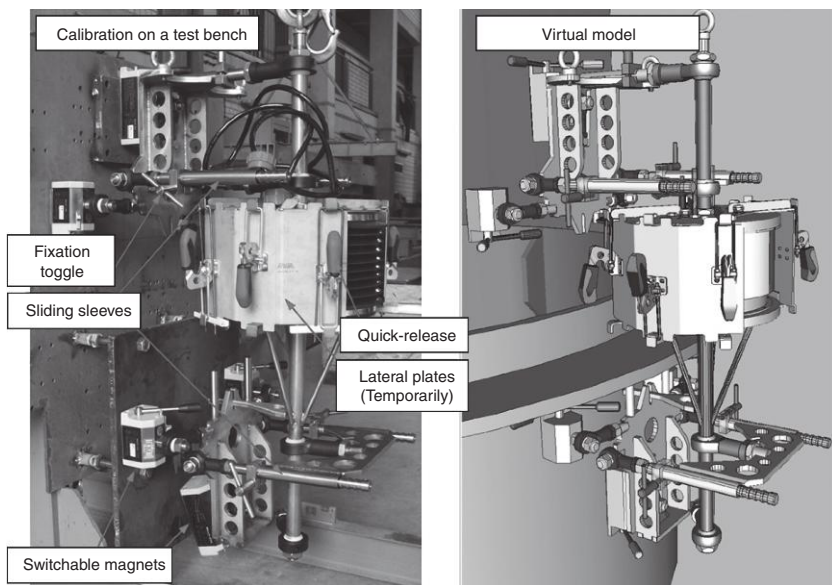
19.13 Prototype of sensor system for relative displacements between pile and sleeve at tripod foundation. (*Source*: Scholle *et al.*, 2012.)

and 19.14b). To cover both a vertical and a horizontal profile, 23 pressure cells were mounted around the circumference at the lower ring, and 4 and 3 cells were mounted at the upper rings, respectively. With this set-up, wave motion around the structure could be monitored and new design parameters extracted (Hildebrandt *et al.*, 2009; Hildebrandt, 2010).

19.4.4 Global SHM for support structures

Even though the major focus of monitoring at WTs is on CM, SHM approaches also play an important role for older structures and remote plants with limited accessibility. Inspection of support structures is mandatory, but its realization remains unspecified. Instead of a visual inspection of kilometers of welded seams, or local monitoring with hundreds of sensors, looking at the structure's behavior during operation provides a promising alternative. Most commonly, acceleration sensors and strain gauges are used to monitor WT support structures, along with EOC- (SCADA-) data (wind speed/direction, nacelle position, rotor speed, power output, temperature, and turbulence intensity). Desired values are modal properties or condition parameters based on system identification procedures. Within the four general steps of SHM (see Rytter, 1993) damage detection is purely sensor-based, while localization, quantification, and prediction might or do necessitate a numerical model of some kind.

Modal parameters

Since modal properties, e.g. the eigenfrequency, are directly linked to physical values such as mass and stiffness, it is self-evident to use these values for

(a)   (b)   (c)
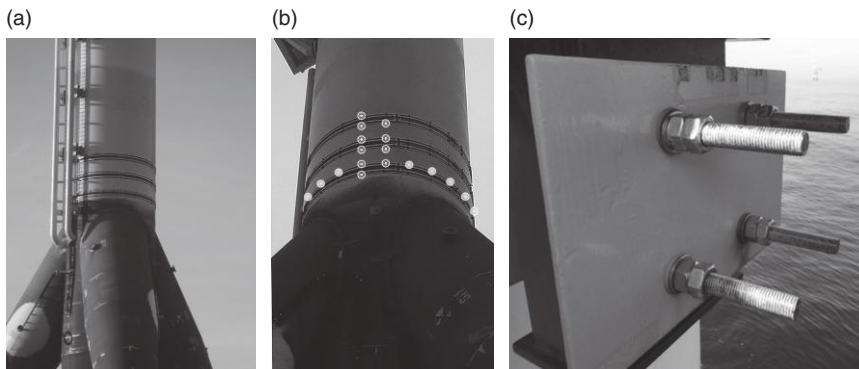
19.14 Installed pressure cells at three levels above upper braces (a, b) and example sheet for corrosion monitoring in splash zone (c). (*Source*: Rolfes *et al.*, 2012.)

idenfying structural changes. Note that for damage detection, damage dimensions are inversely proportional to the required sampling frequency. Small damage affects high frequent local modes, while large damage and structural change affect global modes. Common damage features are eigenfrequency, mode shape, modal curvature (second deviation of mode shape), and modal contribution of mode shapes. The reader is referred to Ewins (2000), Da Maia *et al*. (1997), and Franklin and Powell (1980) for the basic theory of experimental and operational modal analysis (EMA, OMA). Modal properties provide essential information for numerical model updating, which is in turn needed for some damage quantification and localization techniques.

It is essential that the damage scenario being addressed affects the monitored mode shape/modal parameter. Natural frequencies – especially the often monitored first global modes and frequencies – might not have sufficient sensitivity to damage. Variations due to EOCs deteriorate this problem. Mode shapes tend to be more sensitive to local damage, but many sensors are needed, especially when curvatures are addressed (Kim *et al*., 2003). Modal curvature was investigated by Pandey *et al*. (1991). Adams *et al*. (2011) uses modal contribution as the damage feature for a Micon 65/13 turbine in combination with a numerical model, which was validated by measurement. Damage locations are the blade root, the low speed shaft, and the yaw joint.

Automated extraction

In general, all monitoring approaches that concentrate on modal parameters rely on automated procedures for their extraction since, for a normal monitoring time, several thousand data sets must be analyzed. Just recently, several papers have dealt with automated extraction of modal parameters, such as Allemang *et al* (2010), Magalhães *et al*. (2009), and Kraemer (Fritzen and Kraemer, 2011a), and also utilize fuzzy clustering (Carden and Brownjohn, 2008), classification algorithms (Reynders *et al*., 2012), or triangulation (Rolfes *et al*., 2012; Häckell, 2013).

Condition parameters

More abstract condition parameters are based on stochastic subspace identification (SSI), (Van Overschee and De Moor, 1991, 1996; Brincker and Andersen, 2006)) and vector autoregressive models (VAR), (Neumaier and Schneider, 2001; Lütkepohl, 2006). Models have been applied to a 5 MW onshore plant (Fritzen and Kraemer, 2011b) (Kraemer, 2011) as well as to a 5 MW offshore plant within the test field *alpha ventus* (Rolfes *et al*., 2012; Häckell, 2013). These CPs take many different structural modes into account rather than focusing on single frequencies or mode shapes. Hence, they give information about the general dynamic behavior of the (O) wind

energy converters (WECs). To overcome the difficulty of changing EOCs, which influences CPs-values strongly, classification of data sets, e.g. through fuzzy clustering (Fritzen and Kraemer, 2011b) (Kraemer, 2011) or affinity propagation (AP), (Givoni and Frey, 2009; Frey and Dueck, 2007; Rolfes *et al.*, 2012; Häckell, 2013) is necessary. These techniques can also improve estimation of modal parameters, since variations can be reduced.

19.4.5   Case studies

During the last years, several research activities have focused on SHM for support structures of WTs, comprising different combinations of SHM techniques, sensors, and data acquisition and evaluation.

Benedetti *et al.* monitored an onshore WT with 13 m rotor diameter and 18 m hub height through strain measurement (Benedetti, Fontanari and Zonta, 2011). Measurements were taken with fiber Bragg gratings (FBG), a relatively new optical method for strain measurement, from the tower root right above the foundation to estimate root moments. Numerical simulation of the WT tower provided eigenfrequencies and tower base strain. Strain abnormalities for different crack lengths above the tower root were analyzed and the number of required sensors and their placement are discussed.

In (Smarsly *et al.*, 2011, 2012), a monitoring system for a 500 kW Enercon E-40 onshore WT in Dortmund, Germany, is presented. The sensor system is composed of displacement, strain, and acceleration sensors and combined with a self-managing multi-agent software system that continuously collects data and runs analyses. Faulty signals (e.g. a damaged temp. sensor) can be detected, and users are automatically notified by mail.

In Lachmann *et al.* (2011), acceleration data from an onshore WT, which has been monitored over 17 years, is used for system identification. Wind loads, inertia forces at the rotor, and dynamic loading in the support structure are identified as fatigue-relevant loads. Modal parameters are extracted by SSI and enhanced frequency domain decomposition (EFDD), (see Jacobsen *et al.* (2006)). Based on the modal parameters, modal validation is done with MOPAC (Nguyen *et al.*, 2010), a Java-based software package. Artificial damage is applied within the numerical model for four spring elements, representing tower section connections. Damage detection is then performed with wavelet transformation (Chang and Sun, 2005).

Recently, system identification through VAR models and vibration-based damage detection through condition parameters and numerical model updating at a 5 MW onshore structure under consideration of changing EOCs have been performed by Fritzen and Kreamer (2011b).

Rolfes *et al.* present an SHM approach that addresses damage identification, quantification, and localization (Rolfes *et al.*, 2006, 2007, 2010). Damage

identification is realized through monitoring the proportionality between accelerations below the nacelle and strain at the tower root, including also wireless sensing (Rolfes *et al.*, 2007). Here, data were acquired at a 14 year old NEG-Micon 250 (250 kW), and a 6 year old Vestas V-80 (2 MW) turbine. Numerical model validation was also performed.

An array of wireless sensors is used with a NEG-Micon and Vestas with 40 and 78 m hub height, respectively (Swartz *et al.*, 2010). Modal frequencies were identified by the eigenfrequency realization algorithm (ERA), a predecessor of SSI. A detailed overview and introduction on all technical components of wireless sensing for civil infrastructure, as one of the key features for a cost-efficient SHM approach, can be found in (Lynch, 2007). The offshore test mast *Amrumbank West* was analyzed in (Rolfes *et al.*, 2006), where a numerical model is validated and proportionalities between acceleration and strain are monitored. For damage localization, the multiparameter eigenvalue problem (MEP), (Cottin 2001; Cottin and Reetz 2006)) is applied to physical models. This method uses frequency changes in combination with the decomposed stiffness matrix of a numerical model of the structure to locate and quantify structural changes. This method was also applied to two onshore WTs with artificial damage (Rolfes *et al.*, 2010). One is a Südwind S70 (1.5 MW) with a 144 m lattice tower where bolts were loosened at the highest section, the other a guyed Südwind S1200 where tension was reduced in one cable. In both cases the MEP could localize and quantify structural damage correctly.

SHM, applied to an AREVA Wind M5000 (5 MW) offshore turbine, was recently presented in the final report of the research project GIGAWIND *alpha ventus* (Rolfes *et al.*, 2012). The 118 m high WT is located in the German Bight within the test field *alpha ventus*. In addition to acceleration responses and strain gauges at different levels, many EOCs were recorded.

Figure 19.15 and Plate XVIII (see the color section between pages 374 and 375) display some basic EOCs, including wind speed and direction over 17 months. It can be seen that strong wind scenarios typically come from the west. Nacelle positions accordingly have a main orientation to the west too. The center plot gives an idea of the interdependency between rotor speed and wind speed/direction, forming a tornado with a plateau of constant rotor speed for high wind speeds (pitch controlled operation). Modal parameters were extracted from solutions of data-driven SSI data for 24 channels at six elevation levels through stability diagrams over several model orders (see e.g. Plate XIX in the color section between pages 374 and 375). This automated identification procedure allowed the analysis of several thousand data sets under many different EOC combinations – detail on system identification and data classification can be found in (Häckell, 2013).
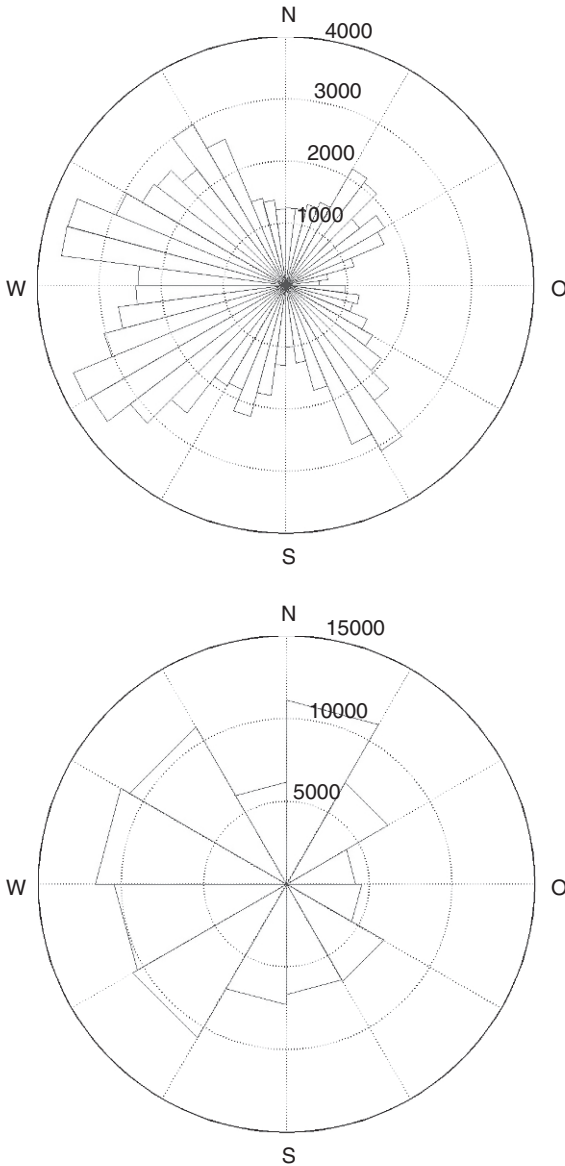
19.15 (a) Wind speed during 17 months of observation. (*Source*: Rolfes *et al.*, 2012.) (b) nacelle position during 17 months of observation. (*Source*: Rolfes *et al.*, 2012.)

The first, global eigenvectors of the structure, extracted from stable paths in Plate XIX, are given in Plate XX in the color section between pages 374 and 375. Since, for each level, two sensors with two orthogonal measuring

directions were installed, three-dimensional eigenvectors can be drawn. The first two orthogonal eigenvectors represent the first bending mode of the tower. Through the indicated nacelle position (rotated coordinate system), it can be seen that, indeed, the first mode moves in the rotor plane while the second moves out of the plane.

The third and the fifth modes both move predominantly in the rotor plane. The fourth mode is a torsion mode and accordingly has participation in both directions. This procedure was repeated for 30 000 data sets to analyze variances of modal parameters due to different EOCs. A rough overview can be seen in Fig. 19.16. The bottom plot displays the identified frequencies over the different data sets. While this plot seems to indicate quite noisy frequency bands, the upper plot shows the occurrence of the different estimates within frequency bins. Here, more clear peaks can be seen at the single eigenfrequencies. Subsequently, the identified modal parameters can be clustered either manually by certain EOCs, to draw dependencies to a single EOC under fixed conditions, or by clustering methods (as k-means, fuzzy-k-means or affinity propagation) to achieve abstract clusters of datasets by all EOCs investigated. Clustering reduces eigenfrequency variance, grouping similar system states.

On the one hand, this provides expectancy values for future data sets, while on the other hand numerical models can be validated in different
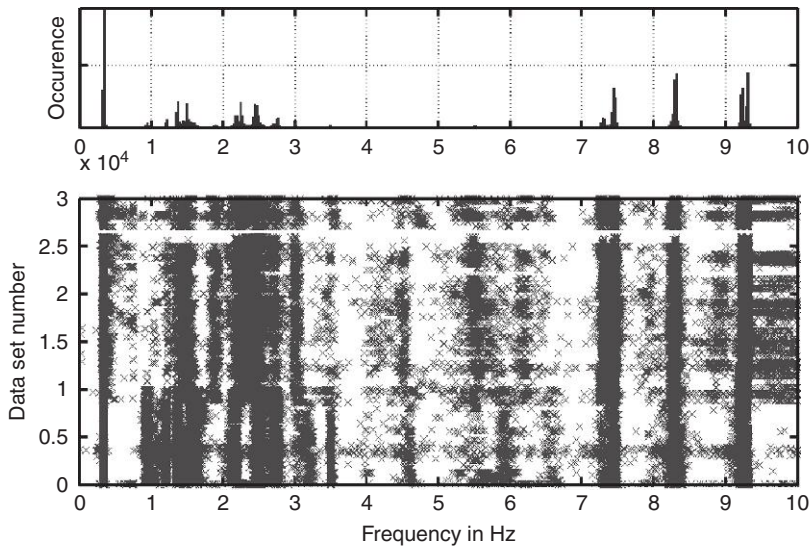
19.16 Results of automated extraction of modal parameters between 0 and 10 Hz for 30 000 data sets, each holding responses for 10 min (bottom), and relative occurrence of frequencies (top). (*Source*: Rolfes *et al.*, 2012.)
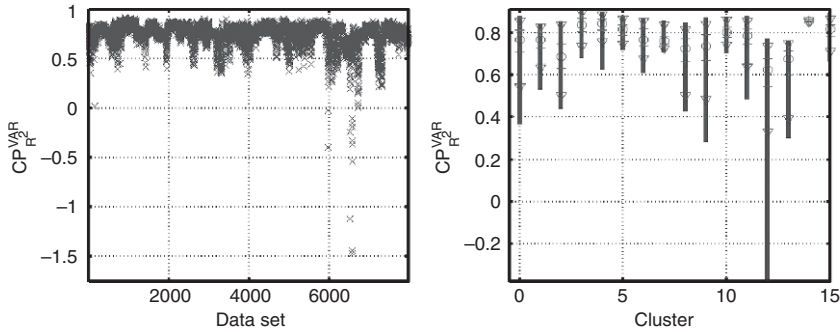
19.17 Condition parameter $[CP]-(R \wedge 2) \wedge VAR$ for a period of 16 weeks (8064 data sets) and 15 clusters. The left plot shows the parameter in chronological order with variations between 0.9 (one is the maximum value) −1.5. The right plot shows the parameter variations within the different clusters ('o'-Median, '+'-25/75% percentile und '∇'-5/95%-percentile, blue bar 1–99%-percentile). Cluster '0' indicates the complete set of parameters. (*Source*: Rolfes *et al.*, 2012.)

operational conditions for a more sophisticated SHM system (damage localization). Besides insight to modal properties, clustering of data sets allows calculation of condition parameters between different data sets, to compare the dynamic structural behavior for different time instances. Figure 19.17 shows such a parameter for 16 weeks, or 8064 data sets.

By way of comparison, vector autoregressive (VAR) models are used. In a VAR model, time series of n sensors are represented through a number of coefficients that were fitted by least squares onto the time series.[21] These parameters, in the case of VAR models a parameter matrix, reproduces the time series by multiplying them onto initial values. Basically, a parameter matrix from a reference time series (the cluster center) is taken to reproduce a second time series. The differences between estimated and actual time series are then transformed to calculate $CP_{R^2}^{VAR}$. If a good agreement exists, $CP_{R^2}^{VAR}$ tends to one; if not, small values result, which can even become negative. In Fig. 19.17 the analyzed time series were divided into 15 clusters. The left plot shows $CP_{R^2}^{VAR}$ for each of the 8064 data sets, most of the values ranging between 0.6 and 0.9. Distinctive drops can be seen repeatedly. The right hand side of Fig. 19.17 shows the statistics of the 15 clusters. Cluster number '0' denotes the complete group of 8064 data sets. It is obvious that variances in $CP_{R^2}^{VAR}$ within most of the single clusters are smaller than for the complete set. Hence, better probabilistic models for the distribution of $CP_{R^2}^{VAR}$ are built within each cluster. New incoming data sets can be assigned

---

[21] The reader is referred to Lütkepohl (Lütkepohl 2006) for extensive descriptions of VAR-Models and System identification procedures.

to a cluster, and $\text{CP}^{\text{VAR}}_{R^2}$ is calculated between the new data set and the cluster center and afterwards compared to the probabilistic model. Finally, the hypothesis of a healthy structure can be accepted or withdrawn for a certain confidence interval. For a complete SHM system, different condition parameters should be used in parallel.

19.5   Conclusion

Monitoring of onshore and, particularly, offshore WTs is a vital field, while the development of robust methods and systems is pursued from industry and research. Due to the many different system parts (rotor, turbine, towers, blades, grouted joint, …), applied sensor types strongly depend on the addressed task. Global monitoring of WTs and their individual parts, such as the rotor blades, is accomplished mainly by conventional sensor systems such as strain gauges and accelerometers. Modal approaches are among the most commonly used global monitoring methods. Nevertheless, it is disputable whether the modal approaches offer sufficient sensitivity to damage. Residual-based damage parameters might lead to more confidence here. Local monitoring necessitates specialized solutions, e.g. for pressure measurement of wave loads, while fatigue is monitored with standard strain gauges. More specifically for the blades, AE and wave propagation methods offer the possibility of damage detection locally, on condition that they are placed close to the occurring damage. Measurement of EOCs also plays an important role for all CM and SHM approaches; here LIDAR (Light Detection And Ranging) systems have recently received a lot of attention for wind field measurements. Optical strain and acceleration sensors with the possibility of a high number of sensor locations could be an alternative in future. Another strongly growing sector in SHM during the last years has been the wireless sensing that can be low cost, and energy harvesting.

There is a need for the development of methods that can detect damage at an early stage and offer higher sensitivity than the widely used modal-based methods. At the same time, the ability to localize and quantify damage without the restrictions of the available methods, such as AE, is of great importance. Such restrictions are the locations of the structure, where the sensors may be applied, and the size of the structure. The above-mentioned requirements can be fulfilled with the combination of different sensors and methods to achieve a complete and robust SHM system.

The future of SHM is promising, since many different fields are being developed. Vibration monitoring, supported through such new time-domain-based system identification procedures as SSI, is an example. On the other hand, the development of data processing methods could improve the sensitivity of the existing methods by efficiently removing background

noise caused by the operation of the turbine and the overall environmental conditions. Data normalization, a very important part aspect of SHM and CM on WTs, which can be achieved through measurement of EOCs in combination with classification algorithms, may also contribute to the optimization of data processing. Additionally, the numerical models that are potentially needed for damage localization, quantification, and prediction confront the challenge of adapting the many different states of the systems. Finally, there should be focus on developing methods that are robust and applicable during operating conditions. For instance, concerning the rotor blades, systems that offer insensitivity to lightning and least possible cabling are required.

19.6   References

Adams, D., White, J., Rumsey, M. and Farrar, C. (2011). Structural health monitoring of wind turbines: method and application to a HAWT. *Wind Energy*, **14**(4), pp. 603–623.

Allemang, R., Brown, D. and Phillips, A. (2010). Survey of modal techniques applicable to autonomous/semi-autonomous parameter identification. In *Proceedings, International Conference on Noise and Vibration Engineering (ISMA)*.

Benedetti, M., Fontanari, V. and Zonta, D. (2011). Structural health monitoring of wind towers: remote damage detection using strain sensors. *Smart Materials and Structures*, **20**(5), p. 055009.

Boller, C., Chang, F.-K. and Fujino, Y. (2009). *Encyclopedia of Structural Health Monitoring*, Chichester, West Sussex, U.K.: John Wiley.

Brincker, R. and Andersen, P. (2006). Understanding stochastic subspace identification. In *Proceedings of the 24th International Modal Analysis Conference (IMAC)*. International Modal Analysis Conference (IMAC). St. Louis, Missouri.

Burton, T., Sharpe, D., Jenkins, N. and Bossanyi, E. (2011). *Wind Energy Handbook*, Chichester: Wiley.

Butterfield, S., Sheng, S. and Oyague, F. (2009). Wind energy's new role in supplying the world's energy — what role will structural health monitoring play. In *Proceedings of the 8th International Workshop on Structural Health Monitoring*. 7th International Workshop on Structural Health Monitoring. Stanford, CA, USA: Chang, F.-K. (Ed.), DEStech Publications Inc.

Carden, E.P. and Brownjohn, J.M.W. (2008). Fuzzy clustering of stability diagrams for vibration-based structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering*, **23**(5), pp. 360–372.

Chang, C.C. and Sun, Z. (2005). Structural damage localization using spatial wavelet packet signature. *Smart Structures and Systems*, **1**(1), pp. 29–46.

Chang, P.C., Flatau, A. and Lui, S.C. (2003). Review paper: health monitoring of civil infrastructure. *Structural Health Monitoring*, **2**, pp. 341–358.

Ciang, C.C., Lee, J.-R. and Bang, H.-J. (2008). Structural health monitoring for a wind turbine system: a review of damage detection methods. *Measurement Science and Technology*, **19**(12), p. 122001.

Cottin, N. (2001). Dynamic model updating – a multiparameter eigenvalue problem. *Mechanical Systems and Signal Processing*, **15**(4), pp. 649–665.

Cottin, N. and Reetz, J. (2006). Accuracy of multiparameter eigenvalues used for dynamic model updating with measured natural frequencies only. *Mechanical Systems and Signal Processing*, **20**(1), pp. 65–77.

Crabtree, C.J., Tavner, P.J., Yang, W., Booth, C. and Watson, S. (2010). *Survey of Commercially Available Condition Monitoring Systems for Wind Turbines*, Durham University of Engineering and Computing Science for UK EPSRC Supergen Wind Energy Technologies Consortium.

Doebling, C.W., Farrar, C.R. and Prime, M.B. (1998). A summary review of vibration-based damage identification methods. *The Shock and Vibration Digest*, **30**(2), pp. 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.

Dutton, A.G. (2004). Thermoelastic stress measurement and acoustic emission monitoring in wind turbine blade testing. In *2004 European Wind Energy Conference and Exhibition*. 2004 European Wind Energy Conference and Exhibition. London, UK: Energy Research Unit, CCLRC Rutherford Appleton Laboratory, p. 9. Available at: http://www.2004ewec.info/files/23_1400_andrewdutton_01.pdf (Accessed 5 November, 2012).

Ewins, D. (2000). *Modal Testing : Theory, Practice and Application* 2nd edn., Baldock: Research Studies Press.

Franklin, G. and Powell, D. (1980). *Digital Control of Dynamic Systems*, Reading Mass.: Addison-Wesley Pub. Co.

Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *Science*, **315**(5814), pp. 972–976.

Fritzen, C.-P. and Kraemer, P. (2011a). Automated system identification and validation of numerical models of offshore wind turbines as basis for SHM-analysis. In *Proceedings of the 8th International Workshop on Structural Health Monitoring*. 8th International Workshop on Structural Health Monitoring. Stanford, CA, USA: Chang, F.-K. (Ed.), DEStech Publications Inc., pp. 1656–1663.

Fritzen, C.-P. and Kraemer, Peter (2011b). Vibration based damage detection for structures of offshore wind energy plants. In *Proceedings of the 8th International Workshop on Structural Health Monitoring*. 8th International Workshop on Structural Health Monitoring. Stanford, CA, USA: Chang, F.-K. (Ed.), DEStech Publications Inc., pp. 1656–1663.

García Márquez, F.P., Tobias, A.M., Pinar Pérez, J.M. and Papaelias, M. (2012). Condition monitoring of wind turbines: techniques and methods. *Renewable Energy*, **46**, pp. 169–178.

Gasch, R. and Twele, J. (2012). *Wind Power Plants Fundamentals, Design, Construction and Operation*, Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: http://proxy.uqtr.ca/login.cgi?action=loginandu=uqtranddb=springer-ebandurl=http://dx.doi.org/10.1007/978–3-642–22938–1 (Accessed 23 October, 2012).

Giurgiutiu, V. (2005). Tuned lamb wave excitation and detection with piezoelectric wafer active sensors for structural health monitoring. *Journal of Intelligent Material Systems and Structures*, **16**(4), pp. 291–305.

Givoni, I.E. and Frey, B. J (2009). A binary variable model for affinity propagation. *Neural Computation*, **21**(6), pp. 1589–1600.

Hameed, Z., Hong, Y.S., Cho, Y.M., Ahn, S.H. and Song, C.K. (2009). Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews*, **13**(1), pp. 1–39.

Harrison, R., Hau, E. and Snel, H. (2000). *Large Wind Turbines: Design and Economics*, Chichester; New York: Wiley.

Hau, E. (2006). *Wind Turbines Fundamentals, Technologies, Application, Economics*, Berlin; New York: Springer.

Häckell, M.W. and Rolfes, R. (2013). Monitoring a 5MW offshore wind energy converter – Condition parameters and triangulation based extraction of modal parameters. *Mechanical Systems and Signal Processing*, **40**(1), pp. 322–343.

Hildebrandt, A. (2010). *Physical Modeling and CFD Simulation of Wave Slamming on Offshore Wind Turbine Structures*, Aachen: ANSYS Conference and 28. CADFEM user's meeting. Available at: http://www.fi.uni-hannover.de/index.php?eID=tx_nawsecuredlandu=0andfile=uploads/tx_tkpublikationen/Hildebrandt_AnsysCADFEM.pdfandt=1302015925andhash=afc12795d4b81e e7adebb908fdaa1dae.

Hildebrandt, A., Stahlmann, A. and Schlurmann, T. (2009). Field data derived from Offshore Wind Energy Converters – Assessment and correlation of dynamic wave loads, Vancouver BC: *Proceedings of the 33rd International Association of Hydraulic Engineering and Research (IAHR) Biennial Congress*. Available at: http://www.fi.uni-hannover.de/index.php?eID=tx_nawsecuredlandu=0andfile=uploads/tx_tkpublikationen/IAHR_2009_FullPaper_Hildebrandt_01.pdfandt=1302015864andhash=23736572196d5487a46495e3b9fb6671.

Jacobsen, N.J., Andersen, P. and Brincker, R. (2006). Using enhanced frequency domain decomposition as a robust technique to harmonic excitation in operational modal analysis. In *Proceedings of ISMA2006: International Conference on Noise and Vibration Engineering*. pp. 18–20. Available at: ftp://ftp.svibs.com/Download/Literature/Papers/2006/2006_6.pdf (Accessed 23 October 2012).

Kim, J.-T., Ryu, Y.-S., Cho, H.-M. and Stubbs, N. (2003). Damage identification in beam-type structures: frequency-based method vs mode-shape-based method. *Engineering Structures*, **25**(1), pp. 57–67.

Kirikera, G.R. (2006). *A Structural Neural System for Structural Health Monitoring of Structures*. Dissertation. University of Cincinnati.

Kraemer, P. (2011). *Schadensdiagnoseverfahren für die {Z}ustandsüberwachung von {O}ffshore-{W}indenergieanlagen*. Dissertation. Siegen: Universität Siegen, Fakultät IV: Natur- und Ingenieurwissenschaftliche Fakultät, Institut für Mechanik und Regelungstechnik – Mechatronik, Schriftenreihe der Arbeitsgruppe für Technische Mechanik. Available at: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2011/532/.

Lachmann, S., Liu, X., Leimbach, K.-R., Höffer, R. and Hartmann, D. (2011). Monitoring und Kontinuierliche Systemidentifikation der Tragstrukturen einer Windenergieanlage zur Schädigungsverfolgung. In D-A-CH Tagung – Erdbeben und Baudynamik. Hannover, Germany, pp. 237–248.

Lambers-Huesmann, M. and Zeiler, M. (2010). Personal communication on scours measured in the German offshore test site alpha ventus.

Liu, W., Tang, B. and Jiang, Y. (2010). Status and problems of wind turbine structural health monitoring techniques in China. *Renewable Energy*, **35**(7), pp. 1414–1418.

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis.* Corrected 2nd printing., Berlin: Springer.

Lynch, J. P. (2007). An overview of wireless structural health monitoring for civil structures. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365**(1851), pp. 345–372.

Magalhães, F., Cunha, Á. and Caetano, E. (2009). Online automatic identification of the modal parameters of a long span arch bridge. *Mechanical Systems and Signal Processing*, **23**(2), pp. 316–329.

Maia, N., Silva, J. and He, J. (1997). *Theoretical and Experimental Modal Analysis*, Taunton Somerset England, New York: Research Studies Press; Wiley.

Mattson, S.G. and Pandit, S.M. (2006). Statistical moments of autoregressive model residuals for damage localisation. *Mechanical Systems and Signal Processing*, **20**(3), pp. 627–645.

Meinlschmidt, P. and Aderhold, J. (2006). Thermographic inspection of rotor blades. In *9th European Conference on NDT*. 9th European Conference on NDT. Braunschweig, Germany: Fraunhofer-Institute for Wood Research (WKI), p. 9. Available at: http://www.ndt.net/article/ecndt2006/doc/Tu.1.5.3.pdf.

Neumaier, A. and Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, **27**(1), pp. 27–57.

Nguyen, V.V., Hartmann, D., Baitsch, M. and König, M. (2010). A distributed agent-based approach for robust optimization. In *Proceedings of the 2nd International Conference on Engineering Optimization*. 2nd International Conference on Engineering Optimization. Lisbon, Portugal, p. 16. Available at: http://www1.dem.ist.utl.pt/engopt2010/Book_and_CD/Papers_CD_Final_Version/pdf/01/01273–01.pdf.

van Overschee, P. and De Moor, B. (1991). Subspace algorithms for the stochastic identification problem. *Decision and Control*, **2**, pp. 1321–1326.

van Overschee, P. and De Moor, B. (1996). *Subspace Identification for Linear Systems: Theory, Implementation, Applications*, Dordrecht, Netherlands: Kluwer Academic Publishers.

Pandey, A.K., Biswas, M. and Samman, M.M. (1991). Damage detection from changes in curvature mode shapes. *Sound and Vibration*, **145**(2), pp. 321–332.

Pfaffel, S., Berkhout, V., Faulstich, S., Kühn, P., Linke, K., Lyding, P. and Rothkegel, R. (2012). *Windenergie Report Deutschland 2011*, Kassel, Germany: Fraunhofer-Institut für Windenergie und Energiesystemtechnik. Available at: http://wind-monitor.iwes.fraunhofer.de/bilder/upload/Windreport_2011_de.pdf.

Research at alpha ventus (RAVE), Germany, Available at: http://rave.iwes.fraunhofer.de/rave/pages/welcome (Accessed 28 September 2012).

Reynders, E., Houbrechts, J. and De Roeck, G. (2012). Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing*, **29**, pp. 228–250.

Rolfes, R., Häckell, M.W., Reetz, J. and Haake, G. (2010). Damage Identification at Offshore Wind Energy Converters Using the Multiparameter Eigenvalue Problem. In *Proceedings of the 10th German Wind Energy Conference*. 10th German Wind Energy Conference, Bremen.

Rolfes, R., Gerasch, W.-J., Haake, G., Reetz, J. and Zerbst, S. (2006). Early Damage Detection System for Tower and Rotor Blades of Offshore Wind Turbines. In *Proceedings of the Third European Workshop on Structural Health Monitoring*

2006. 3rd European Workshop on Structural Health Monitoring 2006. Granada, pp. 455–462.

Rolfes, R., Lynch, J.P., Gerasch, W.-J., Haake, G., Reetz, J. and Zerbst, S. (2012). *Ganzheitliches Dimensionierungskonzept für OWEA-Tragstrukturen anhand von Messungen im Offshore-Testfeld alpha ventus, Final Report, founded by the Federal Ministry for the Environmant, Nature Conservation and Nuclear Safety, support code 0325032*, Hannover: Leibniz Universität Hannover.

Rolfes, R., Huhn, H., Schaumann, P., Schlurmann, T., Lohaus, L. and Achmus, M. (2007). Integral SHM-system for offshore wind turbines using smart wireless sensors. In *Proceedings of the 6th International Workshop on Structural Health Monitoring*. 6th International Workshop on Structural Health Monitoring. Stanford, CA, USA: Chang, F.-K. (Ed.), DEStech Publications Inc.

Rytter, A. (1993). *Vibration Based Inspection of Civil Engineering Structures*. Denmark: Aalborg University.

Scholle, N. and Lohaus, L. (2012). Offshore Measurement System for Relative Displacements of Grouted Joints. In *DGZfP-Proceedings BB 137*. Civil Structural Health Monitoring Workshop (CSHM-4) 'SHM systems supporting extension of the structures' service life'. Berlin .

Seo, D.-C. and Lee, J.-J. (1999). Damage detection of CFRP laminates using electrical resistance measurement and neural network. *Composite Structures*, **47**(1–4), pp. 525–530.

Smarsly, K., Law, K. H. and Hartmann, D. (2011). Implementing a multiagent-based self-managing structural health monitoring system on a wind turbine. In *NSF Engineering Research and Innovation Conference* .

Smarsly, K., Law, K. H and Hartmann, D. (2012). Multiagent-based collaborative framework for a self-managing structural health monitoring system. *Journal of Computing in Civil Engineering*, **26**(1), pp. 76–89.

Sohn, H. (2007). Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365**, pp. 539–560.

Sørensen, B.F. (2002). *Fundamentals for Remote Structural Health Monitoring of Wind Turbine Blades – A Preproject*, Roskilde: Risø National Laboratory : available from: Risø National Laboratory, Information Service Department.

Swartz, R.A. *et al*. (2010). Structural monitoring of wind turbines using wireless sensor networks. *Smart Structures and Systems*, **6**(3, SI), pp. 183–196.

Tong, W. (2010). *Wind Power Generation and Wind Turbine Design*, Southampton; Boston: WIT Press.

Wang, Q. and Deng, X. (1999). Damage detection with spatial wavelets. *International Journal of Solids and Structures*, **36**(23), pp. 3443–3468.

Worden, K. *et al*. (2007). The fundamental axioms of structural health monitoring. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, **463**(2082), pp. 1639–1664.

Yang, W. *et al*. (2012). Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, p.n/a–n/a.

Zanke, U.C.E. *et al*. (2011). Equilibrium scour depths around piles in noncohesive sediments under currents and waves. *Coastal Engineering*, **58**(10), pp. 986–991.

Zerbst, S. (2011). *Global Approach for Early Damage Detection on Rotor Blades of Wind Energy Converters*. Dissertation. Hannover: Gottfried Wilhelm Leibniz Universität Hannover, Fakultät für Bauingenieurwesen, Institut für Statik und Dynamik.

Zerbst, S. *et al.* (2011). Novel sensor concept for monitoring of wind turbine blades. In *Proceedings of the 8th International Workshop on Structural Health Monitoring 2011*. IWSHM. San Fransisco.

Sensing solutions for assessing and monitoring of nuclear power plants (NPPs)

H. SOHN, J. Y. YANG, H. S. LEE and B. J. PARK, Korea
Advanced Institute of Science and Technology (KAIST), South Korea

Abstract: The goal of this chapter is to keep readers abreast of available sensing and structural health monitoring (SHM) techniques for NPP structures. In particular, it focuses on the monitoring of pipelines inside NPPs. Since most NPP structures consist of a vast pipeline network system, the chapter first describes the overall piping systems in NPPs, such as a pressurized water reactor type, a boiling water reactor type, and underground pipelines, focusing on their structural characteristics as well as on various types of structural defects. The principles and mechanisms of various SHM techniques are also explained in detail, to help readers select the optimal SHM solution for their applications.

Key words: nuclear power plants (NPP), structural health monitoring (SHM), accelerometers, optical fiber sensors, piezoelectric transducers, magnetostrictive sensors (MSS), electromagnetic transducers, laser ultrasonic devices.

20.1 Introduction

World energy consumption has been increasing rapidly and is expected to almost double from 2005 to 2030. To meet this fast-growing future power demand, nuclear energy has attracted attention as one of the strongest alternatives for large-scale electricity generation in comparison to other energy sources (US Energy Information Administration, 2008). However, since the first launch of the commercial nuclear facilities in the 1950s, NPPs have been exposed to quite a few incidents. To classify and analyze information on NPP accidents, the International Atomic Energy Agency (IAEA) has introduced the International Nuclear and Radiological Event Scale (INES). According to past INES records, 11 events out of more than 100 incidents have been classified as severe accidents caused by the NPP system failures (Webb *et al.*, 2006; Spiegelberg, 2009). The danger of NPP accidents is that even a small defect in the critical members of NPPs can result in

605

catastrophic consequences. For example, people who have been exposed to radioactive material leaked by NPP accidents suffer from physical disabilities, and some have even died. In addition, vast areas near the NPP site become permanently contaminated, and therefore uninhabitable for any living creatures. Notable NPP accidents include those at Three Mile Island (United States, 1979; economic loss: USD 2400M), Chernobyl (Soviet Union, 1986; economic and human loss: USD 6700M and 4056 deaths), and, more recently, Fukushima (Japan, 2011; economic and human loss: USD 2400M and three deaths) (Fig. 20.1).

To add to these past NPP accidents, it has been reported that many NPPs also continue to deteriorate due to their structural aging. Among all operational NPPs worldwide, 78% have been in operation for over 20 years, and 30% for over 30 years (IAEA, 2007). Furthermore, public concern over the safety of NPPs is continuously growing. Many countries are facing growing public opposition to the construction and operation of NPPs and are being

(a)

(b)
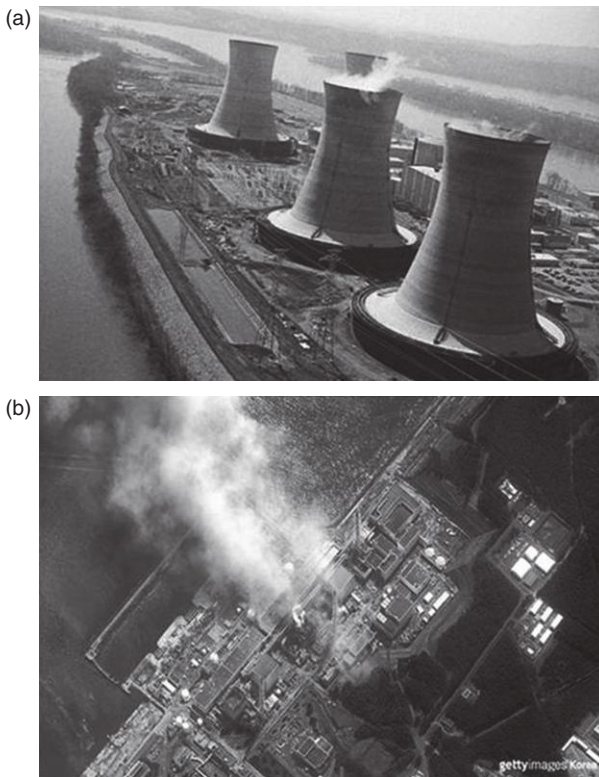
20.1 Previous catastrophic pipeline failures: (a) Three Mile Island NPP (8 November 2005, USA); and (b) Fukushima NPP (3 November 2011, Japan).

pressurized to shut down the existing facilities. In response to the public concern, nuclear regulatory authorities in many countries have tightened their safety measures and established strict inspection criteria for operational NPPs.

Conventional approaches for pipeline inspections in NPPs have been based on nondestructive testing (NDT) (Fig. 20.2). NDT techniques are now well established and routinely employed for NPP inspection so that structural damage can be detected at an early stage (Cartz, 1995; Blitz and Simpson, 1996).

However, despite the merits of NDT techniques, there are several critical limitations. First, NDT, not always but often, requires a periodic overhaul of the whole NPP facility, reducing its power production efficiency. Furthermore, NDT inspection of NPPs needs to be performed by trained, certified engineers, and the inspection typically takes a long time. Most importantly, many critical regions cannot be easily accessed by the NDT personnel (Adams, 2007). For example, pipelines covered with insulators or buried underground cannot be monitored unless the NDT personnel first remove the insulators or excavate the nearby area (Fig. 20.3).

To complement NDT techniques, the nuclear industry strives to adopt structural health monitoring (SHM) systems for NPP management. Compared with NDT, SHM techniques can provide the following potential benefits: (1) automated and continuous monitoring of NPP facilities; (2) reduction in overhaul frequency, cost, and labor necessary for NPP inspection; and (3) monitoring of hidden critical areas that have been difficult to inspect using conventional NDT techniques (Sohn *et al.*, 2003; Inman, 2005).

(a)

(b)

20.2 NDT techniques for NPP pipeline inspections: (a) Pipeline inspection using X-ray; and (b) P-scan inspection of a nuclear reactor.

(a)   (b)
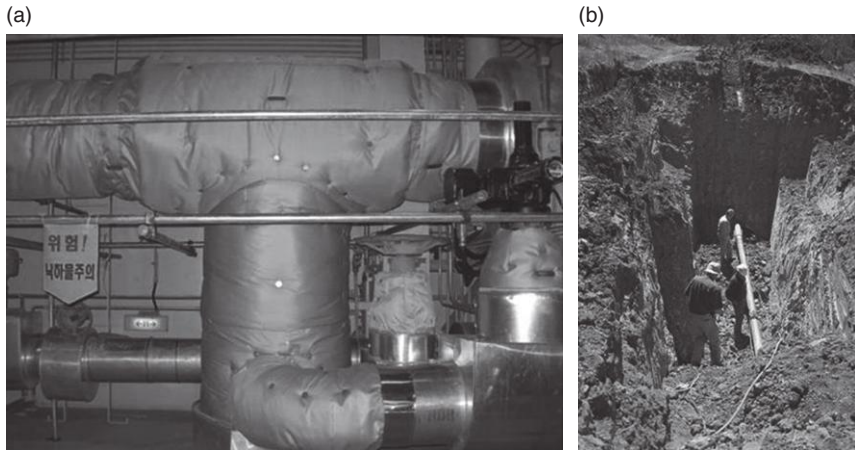
20.3 Examples of pipelines within NPPs that can be easily accessed and inspected: (a) Feed-water pipe with insulator (Wolsong NPP, Korea); and (b) Buried pipeline for service water supply (Nickel Institute).

However, several practical limitations have hampered the application of SHM to NPPs. First, it requires the permanent installation and embedment of distributed sensor networks that can survive the harsh operational condition of the target facilities, including the NPP facilities (Lynch and Loh, 2006). For example, sensors for SHM should be able to operate in a high-radiation ($10^3 \sim 10^7$ Gy/h) and high-temperature environment (up to 350°C) in NPPs. In addition, SHM techniques are normally less accurate than NDT techniques in quantifying the damage level, and they are vulnerable to false alarms induced by the operational variations.

The aim of this chapter is to enhance the reader's understanding of available sensing and SHM techniques for NPP structures. Since most NPP structures consist of pipelines, it first describes the overall piping systems in NPPs, focusing on their structural characteristics, as well as on various types of structural defects. The principles and mechanisms of various SHM techniques are also explained in detail to help readers apply SHM techniques in actual field applications.

20.2   Description of NPPs

The electricity production mechanism in NPPs is basically the same as that in conventional thermal power plants, since it follows a two-step energy conversion process. The heat generated from fuel consumption evaporates feed-water, and a strong steam flow produced by the heat rotates a turbine. Then, metal coils under strong magnetic fields in power generators are also rotated by the kinetic energy of the turbine, and this generates electric current in the

coils. The main difference of NPPs, compared with the conventional thermal power plants, is that the heat is produced by nuclear fission of the fuel in a nuclear reactor. Each type of reactor has its own mechanism of heat production and transmission, and has been selected by worldwide NPP operators according to their social needs and technical conditions.

20.2.1 Systems in a pressurized water reactor (PWR)-type NPP

Among the various kinds of nuclear reactors, PWRs constitute more than 70% of all NPPs worldwide (The New York Times, 2011). The main characteristic of PWR NPPs is that their entire structural components can be divided into two (primary and secondary) systems, according to their purposes. The primary system is generally contained in a reactor building related to the nuclear fission process, while the secondary system takes a main role in electricity production with the feed-water circulation (Fig. 20.4).

Primary system

The primary system is a flow circulation system that enables heat transfer from the reactor to the steam generator. It also controls the temperature and pressure of the piping systems in the containment building. The main components of the primary system include a nuclear reactor, control rods, pressurizer, steam generator, reactor coolant pump, and a reactor auxiliary system.

The nuclear reactor engages nuclear fuel in a fission chain reaction to produce heat, and this warms the water in the pipes of the primary system by thermal conduction through fuel cladding and convection from the bulk of
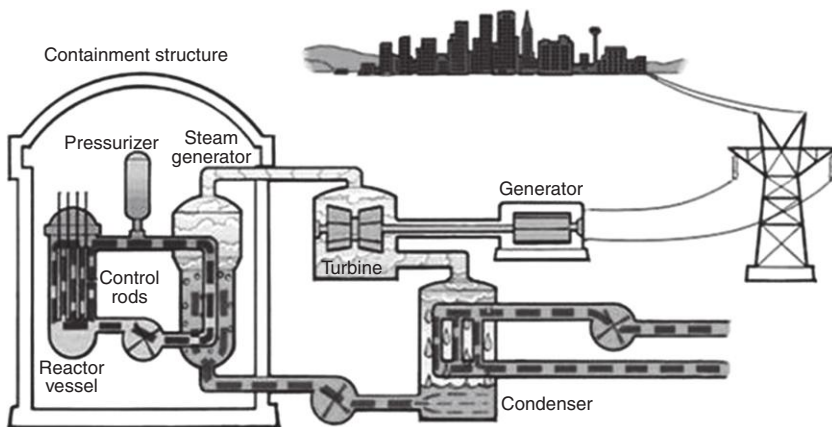
20.4 The PWR (United States NRC).

the coolant to the steam generator inner surface. Here, control rods made of different chemical elements are used to control the rate of the fission chain reaction of uranium and plutonium. The heated water is pumped by a reactor coolant pump and flows into a steam generator, in which it flows through hundreds of tubes and evaporates to pressurized steam. In this process, a pressurizer installed near the steam generator controls the pressure levels in the pipes.

Heat is transferred through the walls of these tubes to the lower pressure secondary system, located on the other side of the steam generator, and then flows to a turbine unit. The heat transfer is accomplished without mixing the two fluids, one in the primary system and the other in the secondary, which is necessary as the primary system is now contaminated with radioactive material. After the heat transfer at the steam generator, the water in the primary system passes through a reactor auxiliary system. This is used to control the chemical composition and volume of the reactor coolant, or to remove residual heat in the coolant during an emergency shutdown of the NPP.

The pipes used in the primary system are normally made of stainless steel for its durability in such a high-temperature and high-pressure environment. As stainless steel is quite expensive, some NPP operators use carbon steel pipes coated with a thin stainless steel layer. Due to the safety concern of the primary system, periodic inspections with conventional NDT techniques have been well established, and structural components have been designed in a highly conservative manner.

Secondary system

The secondary system is a flow circulation system that enables us to transmit the heat energy from the primary system and to generate electrical energy. The main components of the secondary system are a main feed-water system and turbine.

The main feed-water system needs to supply adequate, high quality water to the steam generator. The high-temperature steam is transported to a turbine, which converts the thermal energy from the high-pressure steam into mechanical energy, in the form of shaft rotation, so as to turn the generator. This conversion process is performed in two stages: first, in smaller high-pressure turbines, then, after passing through the moisture separator, in larger low-pressure turbines.

The exhausted steam is squeezed in a condenser, in which a feed-water pump increases the water pressure from a near-vacuum level by a factor of 4. The feed-water pump also drives the condensed water to circulate through the entire secondary system. Using residual heat from the turbine, the water

temperature is gradually increased near the feed-water pump for effective vaporization at the steam generator.

The pipes used in this secondary system are normally made of carbon steel, and the pipes between the steam generator and the turbine are generally covered with insulators to prevent heat loss. As the secondary system is exposed to a steep temperature gradient, carbon steel pipes are quite vulnerable to thermally induced stresses. The secondary system also has many blind spots for conventional NDT-based inspection, as the piping network is larger and more complex than the primary system.

20.2.2 Systems in a boiling water reactor (BWR)-type NPP

Another commonly used reactor type for civil nuclear facilities is BWRs. The main feature of the BWRs is that the reactor core heats the water, creates steam, and then directly drives a turbine to produce electricity. In other words, a BWR does not require several core components, such as a steam generator, pressurizer, and reactor auxiliary systems. Moreover, it does not need to divide its components into two systems as in the PWR (Fig. 20.5).

One characteristic of the BWR is that, compared to the PWR, the structural components near the reactor vessel operate at a substantially lower pressure. Also, the BWR guarantees a lower risk of a rupture causing loss of coolant, compared to the PWR, and a lower risk of core damage that induces such a rupture. This is due to fewer large diameter pipes, fewer welds, and
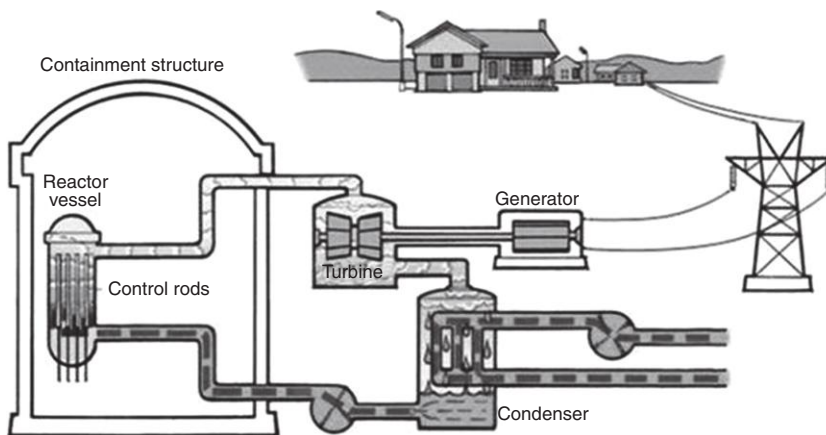
20.5 The BWR (United States NRC).

the absence of steam generators in the BWR. Therefore, the piping systems are less vulnerable to structural damage induced by the high-pressure environment or structural complexity. However, the feed-water system and the turbine are contaminated by radioactive materials as the water flows directly from the core reactor to the entire sub-system of the BWR. This means that even a small structural failure can result in massive radiation leakage, and the entire piping system requires shielding afterwards.

20.2.3   Underground pipelines, particularly for the emergency coolant system

Among various kinds of components in NPPs, underground pipelines might be best suited for continuous SHM due to their inaccessibility and vulnerability to structural defects. There are several types of underground pipelines in NPPs: service water, diesel fuel oil, fire protection, emergency feed-water, and condenser recirculating water pipelines (Braverman *et al.*, 2005).

The underground pipes are normally made of carbon steel and are exposed to a corrosive environment due to permeable soil conditions. Structural integrity of these underground pipeline systems has not been of highest interest for the nuclear industry, as the pipes are regarded as less important than the primary and secondary systems, and cannot be easily inspected by the conventional techniques. Even a small failure of these systems, however, can initiate the emergency safety system of the NPP and cause unintended shutdown, which results in an unexpected and detrimental malfunction of the entire facility. The Nuclear Regulatory Commission (NRC) of the US government recently recognized the inadequacy of safety measures for underground pipelines and requested researchers worldwide to propose proper inspection techniques for ensuring the integrity of underground pipelines in NPPs (Nuclear Energy Institute, 2009).

20.3   Types of damage in pipelines and their failure mechanisms

Pipelines in NPPs have been affected by several types of damage, namely erosion–corrosion wear (ECW), stress corrosion cracking (SCC), and thermally induced damage. This damage frequently develops due to environmental effects (temperature, pressure, radiation, chemical action, etc.) of NPPs. ECW has been observed frequently in the secondary systems of both BWRs and PWRs. SCC in BWRs has required inspection through design lives of NPPs. Stainless steel pipes used for the primary system are affected by thermal aging at the reactor operating temperature. In this chapter, the various types of damage to NPP pipeline are introduced.

(a)   (b)
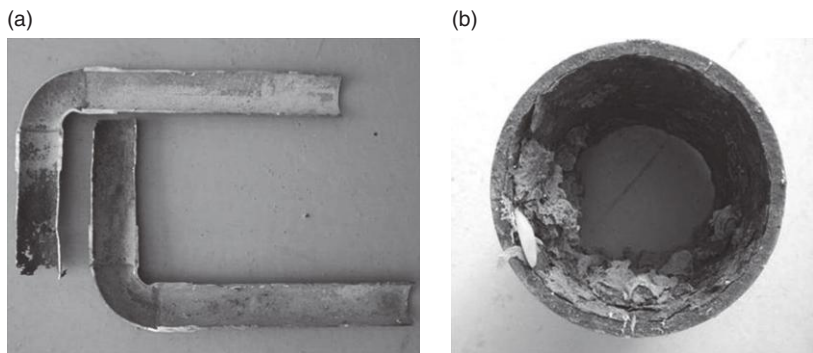
20.6 Erosion–corrosion wear in pipes: (a) Wear inside pipes (in which thickness has decreased due to ECW); and (b) Deposits in pipes (in which thickness has increased due to ECW).

20.3.1   Erosion–corrosion wear (ECW)

ECW is defined as a change in the wall thickness of a pipe due to the inter-action between erosion and corrosion of materials (Fig. 20.6). Erosion is the disturbance of a material surface coating caused by high velocity flows in pipelines, and corrosion is the removal of the protective surface coating due to electrochemical oxidation of metals in reaction with an oxidant. Flow accelerated corrosion (FAC), which can be observed frequently in NPPs, is a phenomenon whereby the thickness of the pipe wall decreases (referred to as wall-thinning), and can be considered one kind of ECW. Recently, the SHM/NDT society has been taking great interest in detecting ECW, because this type of damage frequently develops in NPP pipes.

ECW results possibly from the mechanochemical effect, which is driven by the coupling of mechanical and chemical phenomena on a molecular scale. There are several factors that influence ECW, including (1) localized distribution of flow; (2) water chemistry, such as dissolved oxygen content, pH, corrosiveness of flowing corrodant; (3) high temperature and envi-ronmental pressure; and (4) suspended solids in the water. ECW is nor-mally formed by the interaction of these factors (Brookhaven National Laboratory, 2011).

Research indicates that the exact formation mechanism of ECW has not yet been identified, but many models have been suggested to explain it. Among the various models, the platelet model is considered one of the more effective ways of explaining the ECW formation mechanism (shown in Fig. 20.7). According to this model, ECW starts with the initiation of a divot, which is a small dip in the passive coating of the material, and small turbulences in the fluid flow develop the divot (Fig. 20.7a). When the fluid turbulence is magnified, the fluid itself penetrates into the protective oxide film on the surface removing the surface of the material (Fig. 20.7b).

(a)

Divot

(b)

(c)

(d)
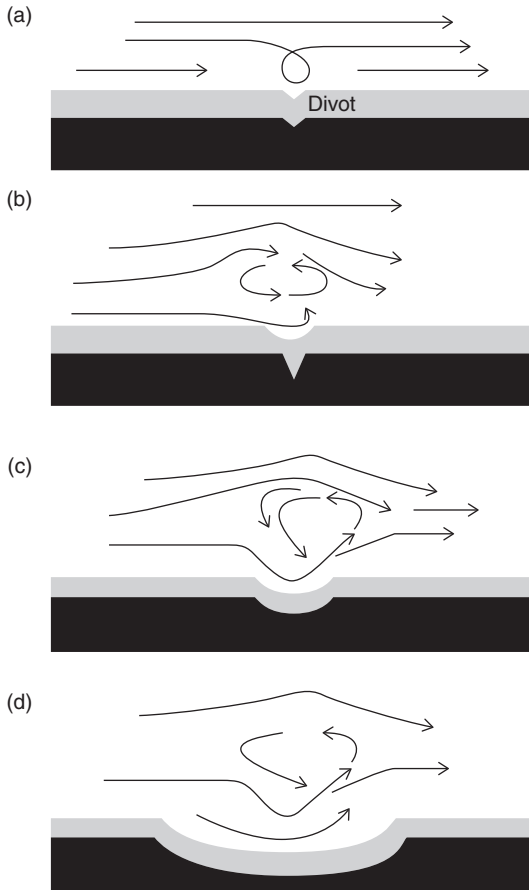
20.7 ECW process model: Platelet model.

Figure 20.7c and 20.7d show the progression of erosion–corrosion, in which the increased turbulence accelerates the formation of the divot. The rate of erosion–corrosion is faster under this turbulence condition than under stagnant and regular erosion conditions (Jones, 1996).

ECW may also result in an increase in the wall thickness of pipes (see Fig. 20.6). This phenomenon occurs from the deposition of various corrosion products that grow on the inner surface of the pipe. The continual change in the pipe wall thickness will lead to structural failures, such as grooves, wave gullies, teardrop-shaped pits, and, ultimately, destruction of the pipes.

ECW is generally found at locations where there are rapid increases in turbulence and alterations to the fluid flow. The pipe types that are particularly vulnerable to ECW include pipe elbows, pipe reducer–expanders, and T-junction parts. These types of pipe are normally located near the

condensate feed-water lines, wet-steam paths of the NPPs secondary coolant system, the steam generator, etc. (Baranenko and Yanchenko, 2007).

20.3.2   Stress corrosion cracking (SCC)

SCC is the environmental degradation of structural components caused by the external force that induces stress on pipes in a corrosive environment. The stress may eventually lead to the sudden and unexpected failure of normally ductile metals.

A notable difference between SCC and ECW is that the wall thickness of the pipe does not change during the SCC process, while ECW causes wall-thinning. Even a small loss of metal in SCC can result in a significant decrease in mechanical strength, and this can trigger sudden and rapid mechanical fracture and failure of the components and structures. A more serious problem is that SCC is hardly observable using common inspection techniques, as it is initiated at a microscopic scale. Figure 20.8 shows an SCC grown on a pipe surface.

The main factors of SCC include (1) a susceptible material (chemical composition of the metal, welding procedures, etc.); (2) corresponding environment (high pressure, high temperature, chemical characteristics of water, corrosive environment, etc.); and (3) tensile stress. SCC is normally due to the interaction of these factors. Although extensive research has been performed to investigate the formation of SCC, its mechanism has not yet been fully understood due to the complex interplay of metal, interface, and environmental properties. However, it is well known that the initiation of SCC is from microscopic cracks, normally perpendicular to the tensile stress, as shown in Fig. 20.9 (azgovernor.gov, 2011).

There are several mechanisms that explain how microscopic cracks grow to become SCC. For instance, active path dissolution is induced by the

(a)   (b)
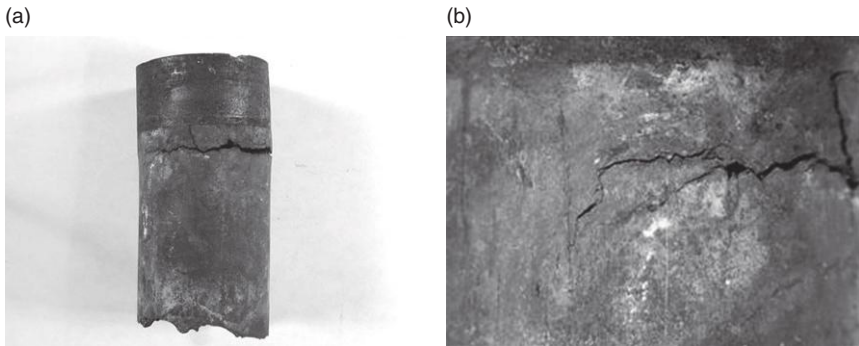
20.8 SCC appearances on the pipeline surface.
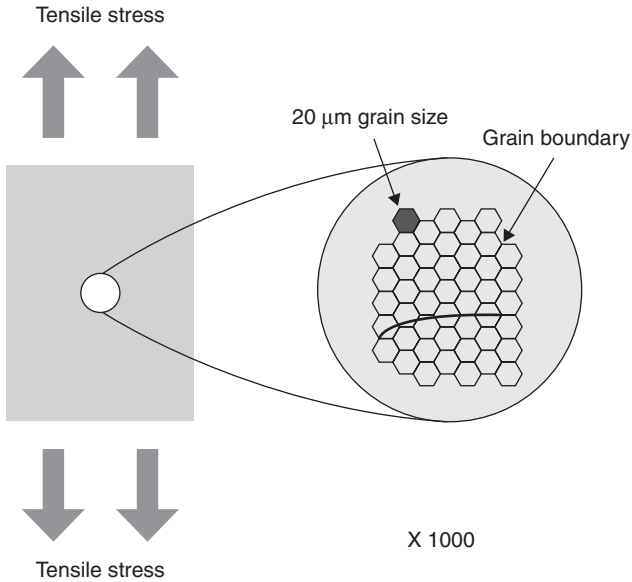
20.9 Illustration of formation associated with the SCC.

intergranular cracking along an electrochemically active grain boundary path. It is generated in the absence of stress, giving rise to intergranular corrosion that is uniformly distributed over the surface. The cracks open up if stress is applied, resulting in easier diffusion of corrosion products away from the cracking top; this allows the crack tip to propagate faster. Another mechanism is hydrogen embrittlement, which occurs from the interactions between hydrogen and metal. It is related to SCC in the sense that it may result in either an intergranular or a transgranular crack path. Hydrogen tends to be attracted to regions of high, tri-axial tensile stress in which the metal structure is dilated. Thus, it is drawn into the regions ahead of cracks or notches that are under stress. The dissolved hydrogen then speeds up the fracture of metal, possibly by making cleavage easier or by facilitating the development of intense local plastic deformation, which leads to embrittlement of the metal. This differs from the first mechanism, in that metal dissolution is not required for the crack to advance (National Physical Laboratory, 2010).

Pipes are vulnerable to SCC in places where the material is susceptible to, for instance, welding procedures, provided that both the corrosive environment and appropriate level of tensile stress are present. There are several locations in NPPs, such as feed-water nozzles, the jet pump, recirculation inlet and outlet nozzles, buried steel pipes, and the steam generator (Jones *et al.*, 1993; Ting, 1999; Karzov and Timofeev, 2006; Janulionis *et al.*, 2009).

These vulnerable spots include mechanically connected parts, such as pipe joints and damp locations. The former can develop highly concentrated stress, and the latter often develop corrosion.

20.3.3 Thermally induced damage

The high-temperature environment of NPPs can generate thermally induced damage in pipes. Thermally induced damage is mainly caused by thermal stratification, thermal striping, and thermal cycling phenomena. It typically occurs in pipe systems where the fluid flows at a low velocity with large temperature variance. Thermal stratification occurs when two types of steam with different temperatures come into contact. Their temperature difference causes the colder and heavier water to settle at the bottom of the pipe while allowing the warmer and lighter water to float over the colder water. When this thermal stratification phenomenon occurs, the pipe is submitted to loads due to the temperature difference between the upper and lower regions of its cross-section. The upper region of the pipe tends to expand; meanwhile, its lower region opposes this expansion. This phenomenon of simultaneous expansion and contraction causes longitudinal loads, which then bend the pipe as shown in Fig. 20.10 (the banana effect). At the same time, the lower part of the pipe (colder region) retains tension, while the upper part (hotter region) retains compression, at the separation interface of the fluids to sustain the geometrical orientation (or continuity). This in turn causes circumferential stresses that may deform the cross-section of the pipe, as can be seen in Fig. 20.11.

Another observation during thermal stratification is the local variation of temperature in the fluid interface known as thermal striping. Thermal striping causes high cycle thermal fatigue and cracks on the internal surface of the pipe. The thermal striping phenomenon is characterized by an oscillation frequency and amplitude (Fig. 20.12).

The third cause of damage is known as thermal cycling. This can occur when the reactor system undergoes an operational transient event, or when turbulence in the main pipe interacts with the thermally stratified layer in a branch pipe. It causes a boundary between the two regions to fluctuate.
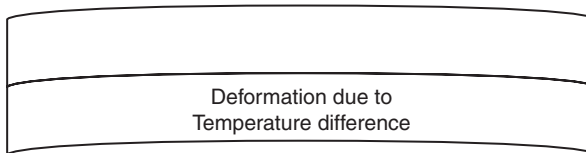
20.10 Longitudinal deformation due to the difference of cross-section temperature.
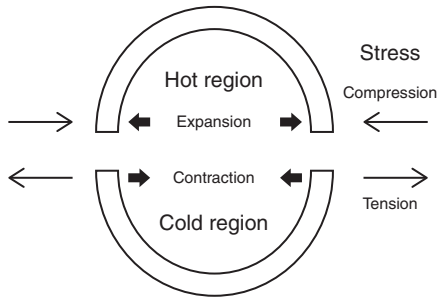
20.11 Stresses of the cross-section of the pipe.
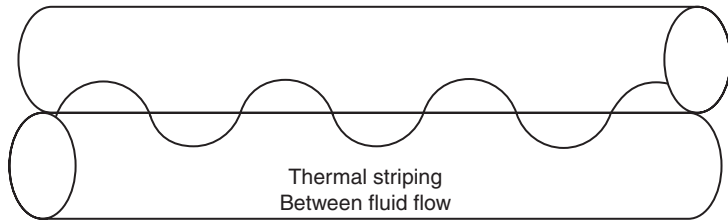
20.12 Thermal striping in the fluid interface.

(a)

(b)
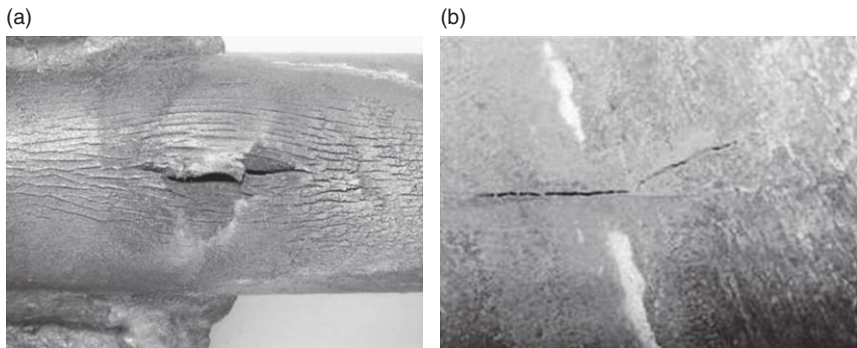
20.13 Thermal fatigue crack due to thermal cycling and thermal striping.

Thermal cycling and thermal striping alternate thermal stress through large temperature changes and cause fatigue cracks in pipes (Fig. 20.13).

As previously mentioned, thermally induced damage occurs frequently in places where there is contact between hot and cold fluids. In NPPs, pressurizer surge lines, emergency core cooling injection lines, residual heat removal lines, and feed-water lines are especially vulnerable to this type of damage (Kamaya and Taheri, 2008).

20.4   Sensor development for NPPs SHM

In the previous section, possible damage types in NPPs are discussed. For continuous detection and monitoring of such damage types, various types of sensors have been developed for NPP applications. In this section, we shall (1) introduce various types of sensors that are potentially applicable to NPP monitoring, (2) explain their basic working principles with advantages and limitations, and (3) review specific sensor applications for NPP monitoring. Particularly, six types of sensors are introduced here: accelerometers, optical sensors, piezoelectric transducers, magnetostrictive sensors (MSS), electromagnetic transducers, and laser ultrasonic devices.

20.4.1   Accelerometers

Acceleration is one of the physical quantities most frequently measured for characterizing dynamic systems. There are various types of accelerometers, such as piezoelectric, optical, laser, capacitive, and servo type accelerometers, and the selection of an accelerometer for a specific application depends among other things on the amplitude and frequency range of the response, sensitivity, and resolution (Fig. 20.14). Furthermore, several rugged accelerometers can function in harsh operational conditions, such as high temperature, pressure, and strain conditions, as well as cryogenic, radioactive, and pyroelectric environments.

The application of accelerometers to SHM of NPPs has been proposed in many studies. Choi *et al.* (2011) proposed an algorithm to find an external impact location on a pipeline structure using embedded accelerometers, since the external impact can induce structural damage with changes in a structural vibration characteristic. Hashemian (2010) used accelerometers to collect vibration data for the aging management of NPPs. By measuring the vibrations of operating machinery and comparing them with the normal historical behavior, abnormal states of the NPP were detected. In particular, accelerometers are often used to detect wall-thinning (one class of ECW) in pipelines (Seong *et al.*, 2004; Hur *et al.*, 2004; Lee *et al.*, 2004). As the thickness reduction of a pipe structure results in changes of the flow-induced vibration characteristics inside the pipe, accelerometers attached on the pipe surface can be used to detect wall-thinning. A similar technique is applied to the monitoring of a steam generator tube (Johnsen *et al.*, 1981; Curlee *et al.*, 1985).

Continuous monitoring of an entire NPP structure might be possible with a relatively small number of accelerometers. However, the major drawback of accelerometer-based SHM is its insensitivity to local damage, which does not necessarily affect the global behavior. Therefore, the damage detection process is challenging at an early stage using accelerometers.

(a)   (b)

(c)
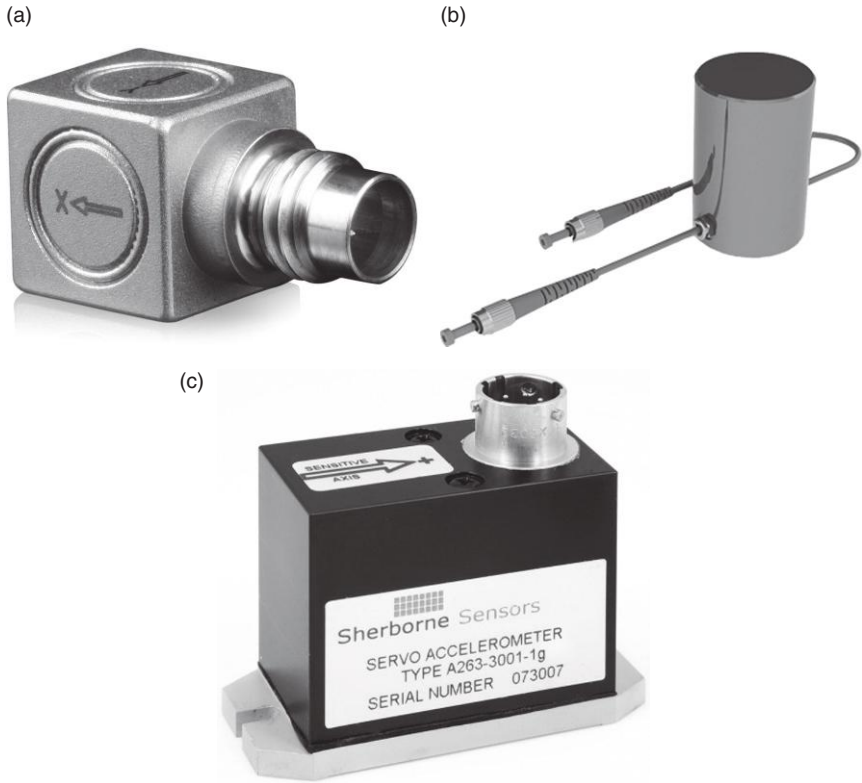
20.14 Various types of accelerometers: (a) Piezoelectric accelerometer (PCB Piezoelectronics); (b) Optical accelerometer (BestechAustrailia); and (c) Servo accelerometer (Tokyo Sokushin).

20.4.2   Optical fiber sensors

An optical fiber sensor is essentially an optical fiber (core) surrounded by a cladding with a lower refractive index. Light travels through the core with little energy loss, because of the total reflection of the light inside the core (Fig. 20.15). External stimulus, such as heat and strain, will affect the properties of the fiber, alter the characteristics of the light traveling inside the fiber core, and eventually modulate the intensity, wavelength, or phase of the light. These changes can be measured by detecting the reflected and scattered light from the optical fiber. The devices using an optical fiber and its characteristics to measure various physical parameters are collectively called optical fiber sensors. See Chapter 5 for more details on optical sensing.

As they are thin, flexible, highly resistant to high temperature (Shimada *et al.*, 2010), and applicable in a radiation environment (Fernandez *et al.*,

(a)

Polymer coating

Silica cladding

Silica core

(b)   Total internal reflection   Cladding (low refractive index)

Core (high refractive index)
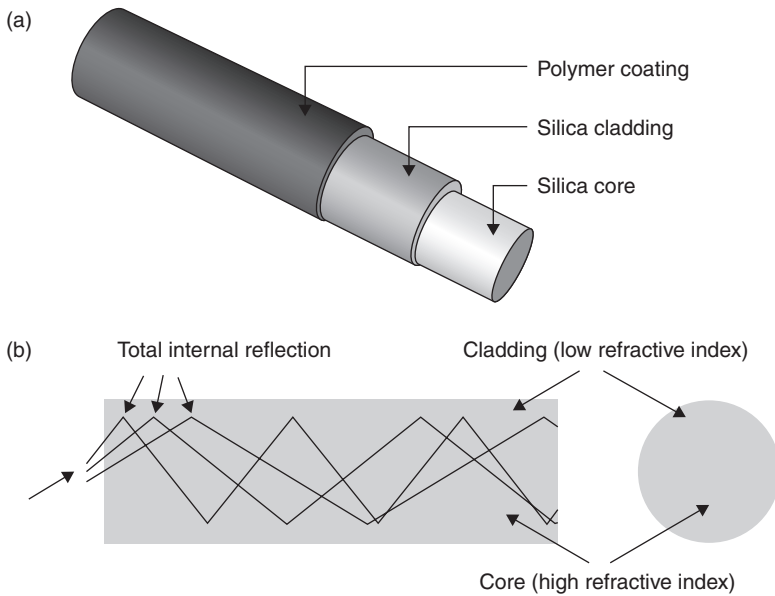
20.15 Optical fiber: (a) Composition; and (b) basic working principle of a typical optical fiber.

2006), optical fiber sensors are attractive for NPP monitoring. There are distributed optical fiber sensors that can measure temperature along the entire length of the fiber. With these sensors, temperature can be measured with a spatial resolution down to 2 cm, and hotspots on pipelines of NPPs are detected as they are heated up (Meunier *et al.*, 1995). It is also possible to measure stress and strain at the operating temperature of NPPs using embedded optical fiber sensors. This makes it easier to determine the life expectancy of critical components such as headers and steam pipelines, which pose a serious hazard in the event of failure. Augé *et al.* (2006) estimated the global level of corrosion by measuring the circumferential dilatation of the pipe to a given pressure level. A strain optical sensor with a hollow inside core was developed by Niewczas and McDonald (2007) to prevent radiation-induced attenuation. Optical fiber sensors are also used to measure chemical parameters such as pH (Deboux *et al.*, 1995), which is a key factor in controlling corrosion in the secondary system of the PWR.

Even with these advantages, the low sampling frequency of optical sensors limits their application for SHM with ultrasonic technologies. For example, the sampling rates of most commercial fiber Bragg grating (FBG) interrogators are less than 100 kHz (Todd *et al.*, 2001), even though several recent studies have achieved sampling rates of more than 100 kHz (Bentell *et al.*,

2009; Vella *et al.*, 2010). FBG sensors, just like accelerometers, are mainly used for sensing. Therefore, the subsequent signal processing remains a challenge, with limited control over excitation.

20.4.3   Piezoelectric transducers

The piezoelectric effect of piezoelectric materials allows the electromechanical interaction between mechanical strain and electrical charge. In other words, when an input voltage is applied to a piezoelectric material, a mechanical strain is created, which then excites the host structure. On the other hand, when the material is coupled to the host structure, a mechanical strain from the host structure is transferred to the piezoelectric materials, and it generates a voltage output, acting as a sensor. Because of this unique piezoelectric nature, the piezoelectric materials are commonly in use for both actuation and sensing applications. See Chapter 4 for more details on piezoelectric sensing and actuation.

Lead zirconate titanate (PZT) is one of the piezoelectric materials most widely used for SHM applications. The small size and non-intrusive nature of PZT are also attractive for SHM monitoring. However, its low operational (Curie) temperature, weak resistance to radiation, and electromagnetic interferences can be problematic for NPP applications. Recently Giurgiutiu *et al.* (2010) developed a piezoelectric sensor that can survive up to 700°C. Parks and Tittmann (2011) have investigated the irradiation effects on piezoelectric transducers, and have developed a radiation resisting device using aluminum nitride (AlN) crystals.

Conventional piezoelectric transducers have been hard to install on curved surfaces due to their brittleness. This problem can be overcome by using flexible piezoelectric transducers, such as polyvinylidene fluoride (PVDF) or macro–fiber composite (MFC). Though the early products of PVDF and the MFC are not appropriate for high-temperature applications, Shin *et al.* (2011) recently developed a film-type flexible PZT transducer that can be used at temperatures above 550°C.

Many researchers have used piezoelectric transducers to measure propagating ultrasonic waves and detect the presence of wave components scattered by damage in pipelines. Lowe *et al.* (1998) and Alleyne *et al.* (2001) installed a circumferential array of piezoelectric transducers to generate torsional waves and record backscattered flexural waves from a crack. Davies and Cawley (2009) used these backscattered flexural waves information to visualize the crack location. Cui *et al.* (2011) generated a torsional guided wave by placing diagonal MFCs at 45° with respect to the longitudinal direction of the pipe and compared the scattered torsional waves before and after crack introduction. Wall-thinning or pipe thickness change due

to ECW can also be detected by wave propagation measurements using piezoelectric transducers. The pipe thickness change can be detected by (1) generating pressure waves in the thickness direction and measuring back-scattered reflections from the wall-thinning region (Yu *et al.*, 2007); (2) generating lamb waves in the axial direction and measuring reflected waves from the wall-thinning region (Isa and Rajkumar, 2009); and (3) generating lamb waves in the axial direction and measuring transmitted waves through the wall-thinning region (Yu *et al.*, 2007; Yu *et al.*, 2008). Piezoelectric transducers are also used to monitor acoustic emission (AE) signals generated by crack initiation for NPP leak detection (Rodriguez and Raj, 1997; Runow, 1985). Guided wave-based techniques with piezoelectric sensors are also widely used because guided waves can travel a long distance with little attenuation, allowing long range monitoring (Meyer *et al.*, 2011). This technique is applied for inspecting baffle-former bolts in PWR pressure vessel assemblies (IAEA 2007), buried and underground pipes (EPRI, 2008), and steam generator tubes (Rose *et al.*, 1994). Using piezoelectric sensors, Umeadi *et al.* (2008), O'Keefe *et al.* (2009), and Catania and Ferrari (2009) recorded changes in flow characteristics to identify structural damage, while Dezfouli and Zabihollah (2010) measured the deformation and deflection of pipes due to external loading.

20.4.4   Magnetostrictive sensors (MSS)

MSSs generate and measure elastic waves using the interaction between mechanical strain and magnetic field. Certain materials called ferromagnetic materials behave as natural magnets. In these materials, strain can be induced by applying an external magnetic field, and this property is known as magnetostriction (Fig. 20.16).

When an electrical current is applied through the solenoid coil around the ferromagnetic rod, a magnetic field is created in the longitudinal direction. This induced magnetic field causes the magnetostrictive material inside the coil to either contract or elongate (Joule effect) in the longitudinal direction and subsequently generates ultrasonic waves on the target structure where the magnetostrictive material is attached. Conversely, induced waves can be measured as changes in the magnetic field when a stress is applied (Villari effect). This effect is used for the compressive behavior, but the relationship between torsional behavior and magnetostriction has also been studied. Similar to the Joule and Villari effect, a magnetic field can be applied when the ferromagnetic rod is subjected to torque (Matteucci effect or inverse-Wiedemann effect) and vice versa (Wiedemann effect). For a detailed explanation of the working principle of the Wiedemann effect, assume that an electric current is flowing through the ferromagnetic rod. Then the electric
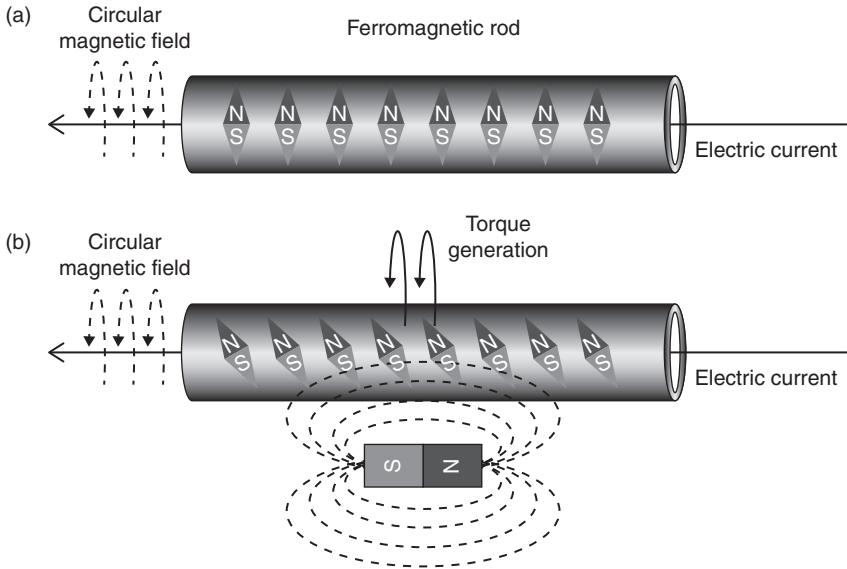
20.16 Principle of magnetostrictive sensor (Wiedemann effect): (a) When an electric current is flowing through the ferromagnetic rod, the current generates a circular magnetic field around the rod; and (b) When the rod is placed in a longitudinal magnetic field, a helical magnetic field is generated as a result of the circular and longitudinal magnetic field. This changes the alignment of the ferromagnetic field inside the rod, which leads to the twisting of the rod and the generation of torque.

current generates a circular magnetic field around the rod. If this rod is placed in a longitudinal magnetic field, a helical magnetic field is generated as a resultant combination of the circular and longitudinal magnetic field. This, in turn, changes the alignment of the ferromagnetic field inside the rod, which leads to the twisting of the rod and the generation of torque. It is also possible to measure waves using a similar principle (inverse-Wiedemann effect). MSSs, which appear in Fig. 20.17, can measure and generate strain/stress/torque using the characteristics explained above. A more complete discussion of MSSs is provided by Kwun and Bartels (1998).

The applicability of MSSs to NPP pipeline monitoring has been studied since 1970s (Hans and Podgorski, 1976, 1977; Kwun and Dynes, 1998; Lu, 1999), especially for crack and notch inspections. The biggest advantage of MSSs is that they can generate pure torsional (T), shear (S), or longitudinal (L) wave modes by continuously exciting ultrasonic waves along the circumference. In particular, MSSs are mainly used to generate a pure torsional wave mode, because of their ability to travel a long distance with little attenuation, non-dispersive character, and high sensitivity

20.17 Typical MSSs to measure and create strain/stress/torque.

to longitudinal cracks. Cheong *et al*. (2006) and Kim *et al*. (2011) gener-ated torsional waves (single T(0,1) mode) with MSSs and measured the reflected torsional and flexural waves from notches to determine the loca-tion and size of damage. Kwun *et al*. (2010) observed that the high fre-quency torsional waves generated by MSSs are insensitive to mechanical features of pipelines, such as clamps, hangers, and supports so that these high frequency waves are suitable for inspecting pipes with additional structural complexity. While traditional MSSs generate guided waves up to several kHz, a recently developed magnetostrictive patch transducer can generate pure torsional modes up to 2 MHz. This is achieved by using a meander coil instead of a solenoid coil. Shear horizontal (SH) modes are also used, as the SH modes have wave characteristics similar to those of the torsional modes (Luo *et al*., 2004).

Since they have a good stability in temperature and in radiation (Shimada, 1993), MSSs are applicable to the harsh environment of the NPPs. However, MSSs are hard to install in real structures, since they are applicable only to inductive materials, and their performance is highly influenced by exter-nal magnetic fields. Their cost compared to that of other sensors is another disadvantage. Also, since they are non-contact sensors, it is required to pre-cisely control the gap between the sensor and the target object. The typical gap is about 1 cm for applications.

20.4.5   Electromagnetic transducers

An electromagnetic acoustic transducer (EMAT) is a transducer for non-contact wave generation and reception using electromagnetic mechanisms (Fig. 20.18a). EMATs can generate SH bulk waves, surface waves, and lamb waves in conductive and ferromagnetic materials. The operational principle of EMATs is as follows (Fig. 20.18b): (1) a current i in the coil, which is par-allel to the surface of a conductive sample, generates a magnetic field; (2) an alternating magnetic field in the sample surface induces eddy currents;

(a)

(b)
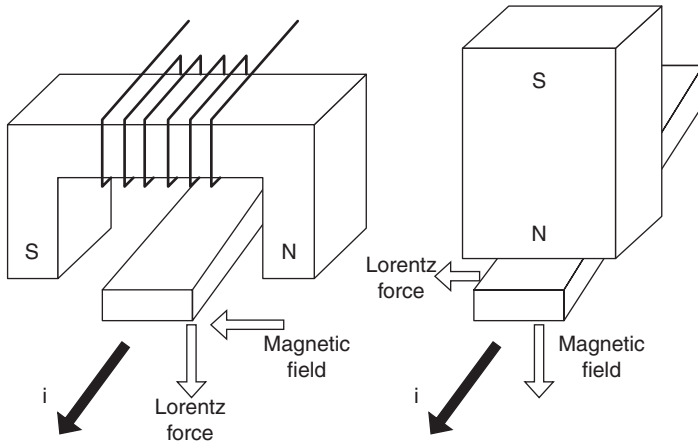
20.18 EMAT: (a) Typical EMAT; and (b) its working principle.

(3) the eddy currents in the magnetic field experience a Lorentz force; (4) a Lorentz force is applied on the surface region of the material due to interaction between electrons and atoms; and (5) the distribution of Lorentz force is controlled by the design of the magnet, the design of the electric coil, the properties of the test material, the relative position between the transducer and the test part, and the excitation signal for the transducer. SH waves are widely used, due to their low attenuation and high sensitivity to longitudinal cracks, and their characteristics have been examined by Murayama *et al*. (2003, 2004).

The biggest advantage of EMATs is their ability to generate a specific wave mode. By taking advantage of the ability to control the generated magnetic fields, electric current and electromagnetic force, EMATs can selectively generate shear, horizontal, and guided waves, as shown in Fig. 20.19. Another advantage of EMATs is that they are non-contact devices, but the spacing between the sensor and the structure needs to be precisely controlled as it has been for MSSs. Additionally, their application is limited to metallic or magnetic products.

EMATs have been used for many SHM applications (Salzburger, 2009). Zhao *et al.* (2005, 2007) identified dents by generating an L_1 mode SH wave with EMATs and by detecting changes of the waveforms as they passed across the dents. Sato *et al.* (1996) measured the thickness changes of pipes using EMAT-generated waves. Other researchers have used EMATs to detect flaws or SCC (Morimoto *et al.*, 1996; Hirao and Ogi, 1998; Kercel *et al.*, 2003; Uribe *et al.*, 2009; Tappert *et al.*, 2008, 2009; Al-Fahad, 2009; Bucker *et al.*, 2009; Chang *et al.*, 2011; Marr *et al.*, 2011). Similar to damage detection
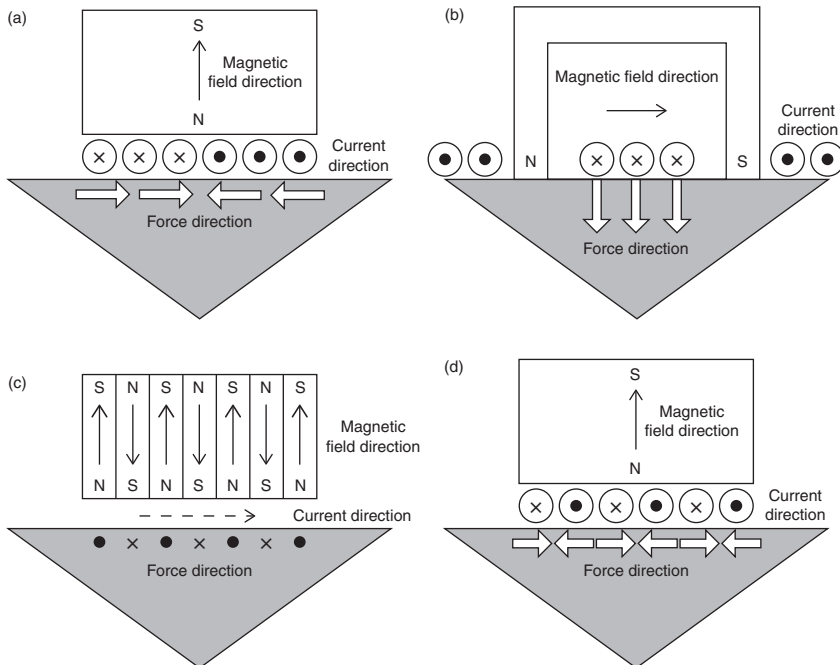
20.19 Types of ultrasonic waves generated by different EMAT configurations: (a) A spiral coil EMAT generates radial polarized shear waves propagating; (b) A tangential field EMAT generates polarized waves; (c) a periodic permanent magnet EMAT generates SH waves; (d) a meander coil EMAT excites longitudinal/shear vertical/Rayleigh / Lamb waves. All these waves are propagating normal to the surface.

techniques using other types of sensors, damage can be identified by generating axisymmetric waves and measuring the reflected non-axisymmetric waves from the damage in pulse-echo mode.

20.4.6   Laser ultrasonic devices

Another alternative for ultrasonic generation and sensing is laser ultrasonics (Scruby and Drain, 1990). This is a non-contact, non-destructive technique for performing ultrasonic evaluations, such as thickness measurement, flaw detection, and material characterization, where conventional techniques cannot easily be applied due to environmental constraints. Laser ultrasonic devices can work even with high temperature surfaces (Nakano and Nagai, 1993), and their scanning capability makes it possible to generate and measure ultrasonic waves over a target surface with high spatial resolution.

Laser ultrasonic devices (Fig. 20.20) can be divided into two categories: one for ultrasonic generation and the other for sensing. For ultrasonic generation, high-power short-pulse lasers, such as solid state Q-switched Nd:Yag and CO_2 gas lasers, are commonly used. When a solid surface is illuminated by a high-power laser, a localized heating and corresponding elastic expansion is produced for an infinitesimal area of the surface, acting as an ultrasonic wave source. Since the ultrasonic generation relies on an instant thermal gradient over a small area rather than an absolute temperature increase, the pulse laser can also be used to generate ultrasonic waves in high temperature specimens. As the generation of ultrasonic waves requires an instant exertion of a highly focused laser beam, the power level, laser duration, and laser beam size need to be carefully tailored to generate elastic waves in a predetermined direction. If the laser power density is increased above a certain threshold, surface melting,

(a)   (b)
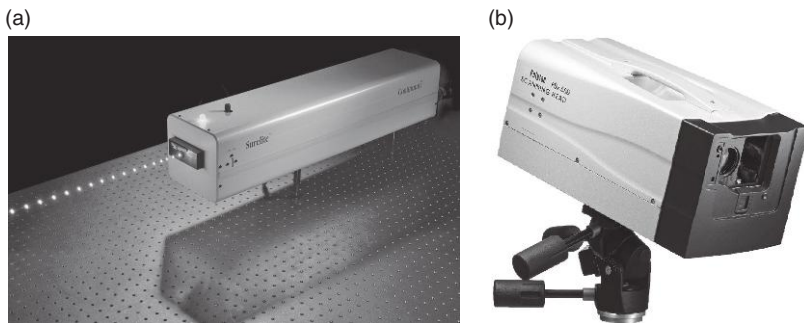
20.20 Commercial laser ultrasonic devices: (a) Nd:Yag laser for ultrasonic generation; (b) Laser scanning vibrometer for ultrasonic sensing.

vaporization, ablation, and plasma formation may occur. (Scuby and Drain, 1990)

As for ultrasonic sensing, there are various means such as two-beam homodyne, two-beam heterodyne, time-delay, Fabry–Pérot, dynamic holographic, multibeam, fiber interferometry, optical beam deflection, and knife edge detection. Among these, laser interferometry using the Doppler effect, which is one type of the two-beam heterodyne methods, is the most widely used because of its sensitivity and stability compared to other intensity modulation interferometers. This technique measures the out-of-plane velocity of the target surface by detecting the frequency shift of the reflected laser beam. The used laser beam should be monochromatic, linearly polarized for measurement sensitivity improvement.

This non-contact nature is the laser ultrasonic device's most powerful advantage, making it possible to generate and measure ultrasonic waves from a distant location without any electrical wires or sensor installation. The laser ultrasonic technique can even be used for monitoring hidden areas by guiding the laser beams through optical fibers (Fomitchov et al., 2002). Another big advantage is that it can achieve a high spatial resolution in ultrasonic wave sensing due to the small laser beam size compared to the sizes of other types of sensors.

Ultrasonic waves, either generated (Park et al., 2006; Sato et al., 2006) or measured (Teufel et al., 1992; Kania and Carrol, 1998) by laser ultrasonic devices, have been used to detect damage such as cracks and wall-thinning in pipes. The excitation and sensing lasers can be incorporated with scanning mirrors so that guided wavefield images with a high spatial resolution can be constructed. For example, Dixon et al. (2010) performed an A-scan by scanning the excitation laser over a target surface and measured ultrasonic signals with an EMAT to detect cracks in metal sheets. Yashiro et al. (2008) and Takatsubo et al. (2009) used a pulsed laser excitation scanning system to detect the interaction between ultrasonic wave propagation and defects. Wave reflection from the defect, and thus the presence of the defect, was identified by creating ultrasonic propagation images. Similar ultrasonic wavefield images can also be constructed using a sensing laser interferometer or laser Doppler vibrometer (Staszewski et al., 2007; Hayashi et al., 2008; Salim et al., 2009; Sohn et al., 2011).

Even with these advantages, there are several limitations to the laser ultrasonic methods. First of all, since the laser ray contains high energy, especially for the laser ultrasonic generation devices, precaution should be taken for eye safety. If the energy level of the laser increases too much, it can also damage the target specimens. Also, because the signal-to-noise ratio of laser measured signals is typically much lower than that of conventional sensors, time averaging of the signals is often necessary, increasing the data collection time. The ultrasonic wave generation efficiency and measurement

sensitivity highly depend on the surface condition, so a special surface treatment is often required for effective excitation and sensing. Last but not the least, the high cost of the laser ultrasonic device makes it less attractive when it comes to real applications.

20.5   Conclusion and future trends

This chapter provides an in-depth discussion of NPP SHM. In particular, among the various structural components in NPPs, this chapter focuses on the monitoring of pipe components, since they are one of the most critical components in NPPs. PWR and BWR are introduced as the two main NPP reactor types. A detailed description of the PWR's primary and secondary systems is provided, and a comparison of the advantages and disadvantages of both reactor types is also given. The effects of critical damage types, such as erosion and corrosion wear, stress corrosion crack, and thermally induced damage, on NPP pipes are described with their formation mechanisms. Then, critical hot spots vulnerable to damage are identified, and sensors and monitoring techniques used to detect them are introduced.

NPP SHM is still in its infancy. To make a transition to real practice, many technical challenges need to be addressed, besides nontechnical issues such as regulations and standardization. Here, there major technical hurdles related to the current state of NPP SHM are identified. First, the ruggedness of many sensors under realistic NPP operational conditions is still not fully verified. Considering the expected lifespan of these sensors, this issue should be addressed before any SHM can be permanently installed for actual NPPs. Second, most of the current SHM studies are limited to simple structures with relative easy access. However, in reality, SHM systems are most needed for damage prone complex components open in hidden or inaccessible areas. For example, connections of underground pipes are particularly prone to structural damage, and these components would benefit most from online embedded SHM systems. Finally, current SHM studies mainly focus on damage identification, and less work is done for damage quantification, although that is essential for SHM systems to be accepted by the NPP NDT community.

20.6   Acknowledgment

20.7   References

Adams D (2007), *Health Monitoring of Structural Materials and Components*, John Wiley & Sons Ltd., Chichester.

Al-Fahad D (2009), 'Detection of pipeline coating failures and stress corrosion cracking (SCC) using electro-magnetic acoustic transducer (EMAT) technology', *Saudi Aramco Journal of Technology*, 49–52.

Alleyne D, Pavlakovic B, Lowe M and Cawley P (2001), 'Rapid, long range inspection of chemical plant pipework using guided waves,' *Insight*, **43**, 93–96.

Anon. (2008), International Energy Outlook, Washington D.C., U.S. Energy Information Administration Department of Energy.

Augé L, Capra B, Lasne M, Bernard O, Benefice P and Comby R (2006), 'Risk management and maintenance optimization of nuclear reactor cooling piping system', *Journal De Physique*, **136**, 263–271.

Azgovernor. Gov (2011), Available from: http://www.azgovernor.gov/ (Accessed 21 August 2011).

Baranenko V and Yanchenko Y (2007), 'Solving the problems of erosion-corrosion wear of equipment and pipelines in nuclear power plants: domestic and international experience', *Thermal Engineering*, **54**, 348–355.

Bentell J, Uwaerts D, Cloots J, Bocquet T and Neys J. (2009), '250 kHz sampling rate FBG interrogator with strong anti-aliasing signal processing', *Proceedings of SPIE*, **7503**(1), 75034S-75034S-4.

Blitz J and Simpson G (1996), *Ultrasonic Methods of Non-Destructive Testing*, Chapman & Hall, Cambridge.

Braverman J, DeGrassi G, Martinez-Guridi G, Morante R and Hofmayer C (2005), *Risk-Informed Assessment of Degraded Buried Piping Systems in Nuclear Power Plants*, Brookhaven National Laboratory, Washington D.C.

Brookhaven National Laboratory (2011), Evaluation of Erosion/Corrosion Damage at Two Nuclear Power Plants, Available from: http://www.iasmirt.org/iasmirt-3/SMiRT10/DC_250715 (Accessed 20 August 2011).

Bucker T, Doescher C and Brown B (2009), 'Acceptance of emat based in-line inspection technology for the assessment of stress corrosion cracking and other forms of cracking in pipelines', *NACE – International Corrosion Conference Series*, Article # 09108.

Cartz L (1995), Nondestructive testing: radiography, ultrasonics, liquid penetrant, magnetic particle, eddy current, ASM International, Materials Park.

Catania A and Ferrari A (2009), 'Development and assessment of a new operating principle for the measurement of unsteady flow rates in high-pressure pipelines', *Flow Measurement and Instrumentation*, **20**, 230–240.

Chang M, Lin Z and Jeng J (2011), 'The study on defects of 6-inch cylinder by electromagnetic acoustic transducer (EMAT)', *Advanced Science Letters*, **4**, 1062–1066.

Cheong Y, Jung H and Kim Y (2006), 'Comparison of an array of EMATs technique and a magnetostrictive sensor technique for a guided wave inspection of a pipe', *Key Engineering Materials*, **321–323**, 780–783.

Choi Y, Park J and Choi K (2011), 'An impact localization technique for a nuclear power plant by using sensors of different types', *ISA Transactions*, **50**, 111–118.

Cui L, Liu Y and Soh C (2011), 'Health monitoring of cylindrical structures using torsional wave generated by piezoelectric macro-fiber composite', *Proceeding of SPIE-The International Society for Optical Engineering*, 7984, Article # 79840G.

Curlee Jr. N, Frick T and Mabon I (1985), 'Tube vibration measurements on a feedring nuclear steam generator', *American Society of Mechanical Engineers Heat Transfer Division*, **51**, 43–55.

Davies J and Cawley P (2009), 'The application of synthetic focusing for imaging crack-like defects in pipelines using guided waves', *IEEE Transactions on Ultrasonics*, *Ferroelectrics, and Frequency Control*, **56**, 759–770.

Deboux B, Lewis C, Scully P and Edwards R (1995), 'A novel technique for optical fiber pH sensing based on methylene blue absorption', *Journal of Lightwave Technology*, **13**(7), 1407–1414.

Dezfouli S and Zabihollah A (2010), 'Structural health monitoring of buried pipelines under static dislocation and vibration', *Proceedings of 2010 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications*, Article # 69322H, 325–329.

Dixon S, Burrows S E, Dutton B and Fan Y (2010), 'Detection of cracks in metal sheets using pulsed laser generated ultrasound and EMAT detection', *Ultrasonics*, **51**(1), 7–16.

EPRI (2008), 'Recommendations for an effective program to control the degradation of buried piping', Electric Power Research Institute (EPRI), Charlotte, NC, EPRI 1016456.

Fomitchov P, Kim Y, Kromine A and Krishnaswamy S (2002), 'Laser ultrasonic array system for real-time cure monitoring of polymer-matrix composite', *Journal of Composite Materials*, **36**(15), 1889–1901.

Fernandez Fernandez A, Brichard B, Berghmans F, El Rabii H, Fokine M and Popov M (2006), 'Chemical composition fiber gratings in a high mixed gamma neutron radiation field', *IEEE Transactions on Nuclear Science*, **53**, 1607–1613.

Giurgiutiu V, Xu B and Liu W (2010), 'Development and testing of high-temperature piezoelectric wafer active sensors for extreme environments', *Structural Health Monitoring*, **9**(6), 513–525.

Hans R and Podgorski J (1976), 'Magnetostrictive high-temperature sound and vibration measurement device for nuclear power plants', *Siemens Z*, **50**, 527–532.

Hans R and Podgorski J (1977), 'Magnetostrictive device for high-temperature sound and vibration measurement in nuclear power stations', *Siemens Reviews*, **44**, 171–175.

Hashemian H (2010), 'Aging management of instrumentation & control sensors in nuclear power plants', *Nuclear Engineering and Design*, **240**, 3781–3790.

Hayashi T, Kojika Y, Kataoka K and Takikawa M (2008), 'Visualization of guided wave propagation with laser Doppler vibrometer scanning on curved surfaces', *AIP Conference Proceedings*, **975**, 178–184.

Hur S, Nam S, Cha D, Seong S, Park W, Kim J, Nam J and Lee S (2004), 'Measurement of the piping elbow integrity using the fiber-optics technique, *Key Engineering Materials*, **270–273**, 750–755.

Hirao M and Ogi H (1999), 'An SH-wave EMAT technique for gas pipeline inspection', *NDT&E International*, **32**, 127–132.

IAEA (2007), 'Assessment and management of ageing of major nuclear power plant components important to safety: PWR vessel internals – 2007 update', International Atomic Energy Agency (IAEA), Vienna, Austria, IAEA-TECDOC-1557.

Inman D J (2005), *Damage Prognosis for Aerospace, Civil and Mechanical Systems*, Wiley, Chichester.

Isa D and Rajkumar R (2009), 'Pipeline defect prediction using support vector machines', *Applied Artificial Intelligence*, **23**, 758–771.

Janulionis R, Dundulis G and Karalevicius R (2009), 'Evaluation of the inter granular stress corrosion cracking defects in austenitic stainless steel piping', *Proceedings of International Conference on Nuclear Engineering*, **4**, 443–448.

Jones D (1996), *Principles and Prevention of Corrosion*, 2nd Edition, New Jersey, Prentice Hall, Inc.

Jones R, Gilman J and Nelson J (1993), 'Controlling stress corrosion cracking in boiling water reactors', *Nuclear Engineering and Design*, **143**, 111–123.

Johnsen J. R, Harris C, McGuinn E, Simonis J, Thoren D and Brown J (1981), 'Flow-induced vibration analysis of Three Mile Island Unit-2 once through steam generator tubes', EPRI Report.

Kamaya M and Taheri S (2008), 'A study on the evolution of crack networks under thermal fatigue loading', *Nuclear Engineering and Design*, **238**, 2147–2154.

Kania R and Carrol B (1998), 'Non-destructive techniques for measurement and assessment of corrosion damage on pipelines', *Proceedings of the International Pipeline Conference*, **1**, 309–313.

Karzov G and Timofeev B (2006), 'Analysis of the events of failures of pipelines made of austenitic steel in nuclear power generating industry', *Strength of Materials*, **38**, 359–366.

Kercel S, Tucker J and Varma V (2003), 'Pipeline flaw detection with wavelet packets and gas', *Proceedings of SPIE – The International Society for Optical Engineering*, **5103**, 217–226.

Kim Y, Moon H, Park K and Lee J (2011), 'Generating and detecting torsional guided waves using magnetostrictive sensors of crossed coils', *NDT & E International*, **44**, 145–151.

Kwun H and Bartels K (1998), 'Magnetostrictive sensor technology and its applications', *Ultrasonics,* **36**(1–5), 171–178.

Kwun H and Dynes C (1998), 'Long-range guided wave inspection of pipe using the magnetostrictive sensor technology – Feasibility of defect characterization', *Proceedings of SPIE – The international Society for Optical Engineering*, **3398**, 28–34.

Kwun H, Kim S and Light G (2010), 'Improving guided wave testing of pipelines with mechanical attachments', *Materials Evaluation*, **68**, 927,932.

Lee N, Bahn C, Lee S, Kim J, Hwang I, Lee J, Kim J and Luk V (2004), 'Development of an on-line ultrasonic system to monitor flow-accelerated corrosion of piping in nuclear power plants', *Key Engineering Materials*, **270–273**, 2232–2238.

Liang C, Sun F and Rogers C A (1996), 'Electro-mechanical impedance modeling of active material systems', *Smart Materials and Structures*, **5**, 171–186.

Lowe M, Alleyne D and Cawley P (1998), 'The mode conversion of a guided wave by a part-circumferential notch in a pipe', *Journal of Applied Mechanics*, **65**(3), 649–656.

Lu Y (1999), 'Magnetostrictive characterization of notches in steel pipes', *Proceedings of SPIE – The International Society for Optical Engineering*, **3588**, 48–56.

Luo W, Rose J and Kwun H (2004), 'Circumferential shear horizontal wave axial-crack sizing in pipes', *Research in Nondestructive Evaluation*, **15**, 149–171.

Lynch J and Loh K (2006), 'A summary review of wireless sensors and sensor networks for structural health monitoring', *The Shock and Vibration Digest*, **38**, 91–128.

Marr J, Sanjuan E, Rosca G, Sutherland J and Mann A (2011), 'Validation of the latest generation EMAT ILI technology for SCC management', *23rd Pipeline Pigging and Integrity Management*, 1, 19p.

Meunier C, Guerin J, Lequime M, Rioual M, Noel E, Eguiazabal D, Fleury D, Maurin J and Mongin R (1995), 'Industrial prototype of a fiber-optic sensor network for the thermal monitoring of the turbogenerator of a nuclear power plant-design, qualification, and settlement', *Journal of Lightwave Technology*, **13**, 1354–1361.

Meyer R M, Ramuhalli P, Bond L J and Cumblidge S E (2011), 'Developing effective continuous on-line monitoring technologies to manage service degradation of nuclear power plants', *Proceedings of 2011 IEEE Conference on Prognostics and Health Management*, 1–7.

Morimoto K, Urabe Y, Sugiyama M, Kitahara H and Takata S (1996), 'Development of in-operation-inspection technique to detect crack of vessels in nuclear power plant by ultrasonic testing method', *Proceedings of the International Conference on Pressure Vessel Technology*, **1**, 427–432.

Murayama R, Hoshihara H and Fukushige T (2003), 'Development of an Electromagnetic Acoustic Transducer that can Alternately Drive the Lamb Wave and Shear Horizontal Plate Wave', *Japanese Journal of Applied Physics*, **42**, 3180–3183.

Murayama R, Makiyama S, Kodama M and Taniguchi Y (2004), 'Development of an ultrasonic inspection robot using an electromagnetic acoustic transducer for a Lamb wave and an SH-plate wave', *Ultrasonics*, **42**, 825–829.

Nakano H and Nagai S (1993), 'Crack measurements by laser ultrasonic at high temperatures', *Japanese Journal of Applied Physics*, **32**, 2540–2542.

National Physical Laboratory (2000), Stress Corrosion Cracking: guides to good practice in corrosion control, HMSO.

Niewczas P and McDonald J (2007), 'Advanced optical sensors for power and energy systems applications', *IEEE Instrumentation and Measurement Magazine*, **10**, 18–28.

Nuclear Energy Institute (2009), *Guideline for the Management of Underground Piping and Tank Integrity*, Nuclear Energy Institute, Washington D.C.

O'Keefe C, Maron R, Fernald M, Bailey T, Van Der Spek A and Davis M. (2009), 'New developments in flow and pipe management capabilities through new velocity profile measurement and pipe wall wear monitoring instrumentation', *SME Annual Meeting and Exhibit and CMA's 111th National Western Mining Conference*, 1, 26–33.

Parks D and Tittmann B (2011), 'Ultrasonic NDE in a reactor core', *Proceedings of 2011 IEEE Sensors*, 618–622

Rodriguez P and Raj B (1997), 'Development of in-service inspection techniques for nuclear power plants in India', *International Journal of Pressure Vessels and Piping*, **73**(1), 59–68.

Rose J L, Rajana K M and Carr F T (1994), 'Ultrasonic guided wave inspection concepts for steam generator tubing', *Material Evaluation*, **52**, 307–311.

Runow P (1985), 'The use of acoustic emission methods as aids to the structural integrity assessment of nuclear power plants', *International Journal of Pressure Vessels and Piping*, **21**(3), 157–207.

Salim M, Hayashi T, Murase M and Kamiya S (2009), 'Visualization and modal analysis of guided waves from a defect in a pipe', *Japanese Journal of Applied Physics*, **48**(7 PART2), Article # 07GD06.

Salzburger H (2009), 'EMAT's and its potential for modern NDE – State of the art and latest applications', *Proceedings of 2009 IEEE International Ultrasonics Symposium*, 621–628.

Sato M, Sasajima H, Tsuneoka O, Hasegawa I and O'Shima E (1996), 'Development of in-operating pipe wall thickness monitoring system using heat resistant ultrasonic transducer in nuclear power plant', *Proceedings of the International Conference on Pressure Vessel Technology*, **1**, 443–448.

Sato H, Lebedev M and Akedo J (2006), 'Theoretical and experimental investigation of propagation of guide waves in cylindrical pipe filled with fluid', *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, **45**, 4573–4576.

Scruby C and Drain L (1990), Laser Ultrasonics: Techniques and Applications, Taylor & Francis

Seong S, Kim J, Hur S, Kim J, Park W and Cha D (2004), 'The development of fusion sensor techniques for condition monitoring of a check valve, *Key Engineering Materials*, **270–273**, 2220–2225.

Shimada M (1993), 'Magnetostrictive torque sensor and its output characteristics', *Journal of Applied Physics*, **73**(10), 6872–6874.

Shimada Y, Nishimura A, Yoshikawa M and Kobayashi T (2010), 'Design of monitoring system of high temperature piping system by heat resistant fiber Bragg grating', *Journal of Laser Micro Nanoengineering*, **5**, 99–102.

Shin J, Cheung C, Kruger S and Jen C (2011), 'Electrical and optical properties of PZT thick films in high temperature flexible ultrasonic transducers for structural health monitoring and NDT', *International Workshop Smart Materials, Structures & NDT in Aerospace*, 1–9.

Sohn H, Farrar C, Hemez F, Czarnecki J, Shunk D, Stinemates D and Nadler B (2003), A review of structural health monitoring literature: 1996–2001, Los Alamos National Laboratory, Los Alamos.

Sohn H, Dutta D, Yang J, DeSimio M, Olson S and Swenson E (2011), 'Automated detection of delamination and disband from wavefield images obtained using a scanning laser vibrometer,' *Smart Materials and Structure,* **20**(4), Article # 045017 (pp. 1–10).

Spiegelberg R (2009), 'A matter of degree', *The IAEA Bulletin* 51–1. Available from: http://www.iaea.org/Publications/Magazines/Bulletin/Bull511/51102744649. html (Accessed 30 August 2011).

Staszewski, W, Lee B and Traynor R (2007), 'Fatigue crack detection in metallic structure with Lamb waves and 3D laser vibrometry', *Measurement Science and Technology,* **18**, 727–739.

Takatsubo J, Wang B, Miyauchi H, Toyama N and Urabe K (2009), 'Visualization of ultrasonic waves scattered from rear defects by using a laser-based imaging technique', *AIP Conference Proceedings*, **1096**, 666–673.

Tappert S, Mann A, Van Boven G, Allen D and Balzer M (2008), 'Inline Inspection for cracks in gas pipelines-enhancements derived from 5 years' operational experience', *Proceedings of the Biennial International Pipeline Conference*, **2**, 161–167.

Tappert S, Allen D, Mann A, Balzer M and Van Boven G (2009), 'Third-generation EMAT tool enhanced for finding SCC and disbonded coating in dry gas pipeline', *Pipeline and Gas Journal*, **236**, 26–34.

Teufel M, Trimis D, Lohuller A, Takeda Y and Durst F (1992), 'Determination of velocity profiles in oscillating pipe-flow by using laser Doppler velocimetry and ultrasonic measuring devices', *Flow Measurement and Instrument*, **3**, 95–101.

The New York Times (2011), U.S. Nuclear plant have same risks, and backups, as Japan Counterparts, Available from: http://www.nytimes.com/2011/03/14/world/asia/14industry.html (Accessed 15 August 2011).

Ting K (1999), 'The evaluation of intergranular stress corrosion cracking problems of stainless steel piping in Taiwan BWR-6 nuclear power plant', *Nuclear Engineering and Design*, **191**, 245–254.

Todd M, Johnson G and Althouse B (2001), 'A novel Bragg grating sensor interrogation system utilizing a scanning filter, a Mach-Zehnder interferometer and a 3×3 coupler', *Journal Measurement Science and Technology*, **12**(7), 771–777.

Umeadi B and Jones K (2008), 'The development of an intelligent sensor for the monitoring of pipeline system integrity', *Society of Petroleum Engineers – SPE Russian Oil and Gas Technical Conference and Exhibition*, **1**, 581–590.

Uribe S, Nakamura N, Ogi H and Hirao M (2009), 'Mode conversion of SH guided waves at defects for pipeline inspection', *AIP Conference Proceedings*, **1096**, 1550–1557.

U.S. Energy Information Administration (2008), International Energy Outlook, Department of Energy: Washington D.C.

Vella T, Chadderdon S, Selfridge R, Schultz S, Webb S, Park C, Peters K and Zikry M. (2010), 'Full-spectrum interrogation of fiber Bragg gratings at 100 kHz for detection of impact loading', *Measurement Science and Technology*, **21**, 094009.

Webb G, Anderson R and Gaffney M (2006), 'Classification of events with an off-site radiological impact at the Sellafield site between 1950 and 2000, using the International Nuclear Event Scale', *Journal of Radiological Protection*, **26**(1), 33–49.

Yashiro S, Takatsubo J, Miyauchi H and Toyama N (2008), 'A novel technique for visualizing ultrasonic waves in general solid media by pulsed laser scan', *NDT&E International*, **41**, 137–144.

Yu L, Giurgiutiu V, Chao Y and Pollock P (2007), 'In-situ multi-mode sensing with embedded piezoelectric wafer active sensors for critical pipeline health monitoring', *Proceeding of ASME International Mechanical Engineering Congress and Exposition*, **10A**, 687–696.

Yu L, Giurgiutiu V and Pollock P (2008), 'A multi-mode sensing system for corrosion detection using piezoelectric wafer active sensors', *Proceedings of SPIE-The International Society for Optical Engineering*, **6932**, Article # 69322H.

Zhao X, Varma V, Mei G, Ayhan B and Kwan C (2005), 'In-line nondestructive inspection of mechanical dents on pipelines with guided Shear Horizontal wave

Electromagnetic Acoustic Transducers', *Journal of Pressure Vessel Technology, Transactions of the ASME*, **127**(3), 304–309.

Zhao X, Varma V, Mei G and Chen H (2007), 'In-line nondestructive inspection and classification of mechanical dents in a pipeline with SH wave EMATS', *AIP Conference Proceedings*, **894**, 144–151.

21
Sensing solutions for assessing and monitoring power systems

I. A. HISKENS, University of Michigan, USA

Abstract: The chapter describes sensing systems that underpin the safe and reliable operation of large-scale power systems. It provides an overview of power system structure and operation, identifying the important characteristics of generation, transmission, and distribution. A discussion of substations highlights their vital role in establishing network interconnections and providing measurements of network quantities. These functions are fundamental to the supervisory control and data acquisition (SCADA) systems that facilitate power system monitoring and operation. A discussion of sensing equipment and systems considers current and voltage transformers, distribution network sensing, phasor measurement units, and electromechanical mode monitoring. Operational use of SCADA measurements is considered, and an overview of state estimation is provided. The chapter concludes with a discussion of techniques for assessing power system security and reliability.

Key words: power systems, measurement systems, security, state estimation.

21.1   Introduction

Reliable electricity supply underpins modern society. In fact, the US National Academy of Engineering named electrification as the greatest engineering achievement of the twentieth century.[1] Yet electricity supply is taken for granted, with few people having thought about the vast system that is in place to ensure that electricity is available whenever it is required.

Electrical power systems span continental-scale distances, interconnecting extensive generation facilities and ultimately supplying the electricity needs of a myriad of consumers. Consider, for example, the eastern interconnected power system of the United States. This one power system stretches from the Dakotas in the west, across to New England in the north-east, and as far south as the tip of Florida. Control of this system must ensure that, at all times, the operational limits of components such as generators, transformers, and transmission lines are enforced, and that voltages across the system

638

satisfy statutory requirements.[2,3] This is achieved through a combination of closed-loop controls and extensive monitoring and assessment. Furthermore, fast protection schemes are required to minimize the physical damage caused by faults, such as lightning strikes or equipment failures, and must prevent situations that might endanger personnel.[4]

This chapter discusses sensing, monitoring, and assessment technologies that underpin reliable and economic operation of power systems. To set the scene, Section 21.2 provides an overview of power system composition and operation. Sensing equipment and systems are discussed in Section 21.3. Power system operation is considered in Section 21.4, with particular attention given to the monitoring and assessment techniques that underpin real-time operation. Section 21.5 provides conclusions. An Appendix summarizes the basic concepts of alternating current (AC) power supply.

21.2   Power system overview

Power systems typically consist of numerous large generating stations that supply consumers through extensive transmission and distribution networks. An illustration of a basic power system is given in Fig. 21.1.

21.2.1   Generation

The bulk of electricity is produced through electromechanical energy conversion using synchronous generators. In the case of coal and nuclear power generation, the fuel is used to produce steam, which drives large high-pressure turbines. The turbines spin the rotor of the generator, producing three-phase AC electricity.[5,6] In hydro-power production, the downhill flow of water provides the energy source for driving turbines, while in

Generation	Transmission		Distribution			Customers
Power Plant	345 000 Volts	**Station**	**Sub-Transmission**	**Sub-**	13 200 Volts	2400 Volts
15 000 Volts	230 000 Volts		120 000 Volts	**station**	8 320 Volts	120 Volts
	120 000 Volts		40 000 Volts		4 800 Volts	
			24 000 Volts			
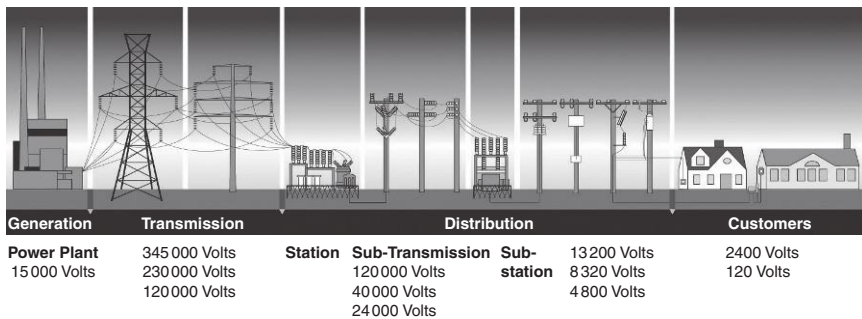
21.1 Power system overview. (*Source*: © 2012 DTE Energy. Reprinted with permission.)

gas-fired generation, combustion within a jet engine provides the driving torque.

Wind generation supplies an increasing proportion of electrical energy, though its contribution is still relatively small.[7] In this case, the lift established by the aerodynamics of the blades rotates the shaft and drives the rotor of a generator. To cater for the variability inherent in wind, the electromechanical energy conversion process tends to be more complicated, often involving power electronic converters.[8]

Photovoltaic (PV) solar sources are also becoming popular, though their contribution is quite small. PV cells form the building blocks for larger conversion systems.[9,10] The cells are made from semiconductor material that converts solar radiation into an electrical current. A power electronic inverter is required to transform the direct current (DC) electricity produced by PV cells into AC for connection to the grid.

21.2.2   Transmission

Generation facilities are often located some distance from consumers, requiring transportation of large amounts of electricity. As electrical current I flows from source to sink through the resistance R of the conductors, some of the electrical energy is converted to heat in the form of I^2R losses. Because losses are proportional to the square of the current, reducing current is hugely beneficial. The electrical power produced by a generator is given by $|S| = VI$, as explained in the Appendix. Therefore, for a given electrical power S, the current I can be reduced by increasing the voltage V.

Accordingly, the transmission network is composed of high-voltage transmission lines that carry bulk electricity over long distances. Transmission voltages are tailored to the requirements of particular connections, and are also governed by utility standardization. Typical transmission voltages tend to lie in the range 275–765 kV, though some lower-voltage lines may play a transmission role, and some higher-voltage lines also exist. Because of the high voltage, conductors must be separated from each other and from the ground by a large distance. This implies the need for high towers and large insulator strings. More importantly, from a sensing point of view, transducers must be capable of withstanding very high voltages and currents. This is considered further in Section 21.3.

Transmission line design establishes the maximum voltage and current to which a line may be subjected. If voltage exceeds the design upper limit, the probability of a catastrophic flash-over (arcing) becomes unacceptable. On the other hand, low voltage causes increased current flow, and therefore incurs higher losses and reduced efficiency. The heating effect of I^2R

losses causes the conductors (wires) to expand, leading to increased conductor sag[1] and reduced clearances to ground.[11] Line design must ensure that conductors always have a certain safe clearance from the ground. This limits the maximum permissible sag, and hence places a limit on the current that a transmission line may carry.

The transmission system usually has an interconnected (meshed) structure, allowing electricity to flow over multiple paths between source and sink nodes within the network. This provides high reliability, as electricity flow is not interrupted by the disconnection (outage) of any single transmission line. Rather, such a disconnection will cause electrical power to redistribute across the remaining lines. Power system operators must continually monitor line flows, however, to ensure that redistributed flows do not exceed line capabilities.

21.2.3 Transformers

Electricity is generated at voltage levels in the range 15–25 kV, transmitted at much higher voltages, and consumed at voltages below 1000 V. Interconnection between different voltage levels is achieved using transformers.[13] As shown schematically in Fig. 21.2, transformers consist of primary and secondary windings (coils of conductor) wrapped around a ferromagnetic core. From Ampère's circuital law, current flow in a winding will produce magnetic flux in the core, with that flux linking both windings. If the current is time-varying, then the magnetic flux will also vary with time. Faraday's law states that this time-varying magnetic flux will induce a voltage difference across the ends of the winding.

Together these laws establish the relationships between voltages and currents of the primary and secondary windings of an ideal transformer,

$$\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s} \qquad [21.1]$$

where V_p, I_p, N_p are the primary-side voltage, current and number of turns in the winding, and likewise for the secondary-side quantities. As an example, a generator transformer that steps voltage up from 20 to 354 kV would have a winding ratio of $N_s / N_p = 345 / 20 = 17.25$, i.e., the secondary winding would have 17.25 turns for every single turn on the primary winding.

Transformer windings utilize high quality, low resistance conductors. Nevertheless, some I^2R heating is inevitable under peak loading (high

---

[1] Sag is the vertical distance between the conductor and a straight line joining the two adjacent transmission towers.
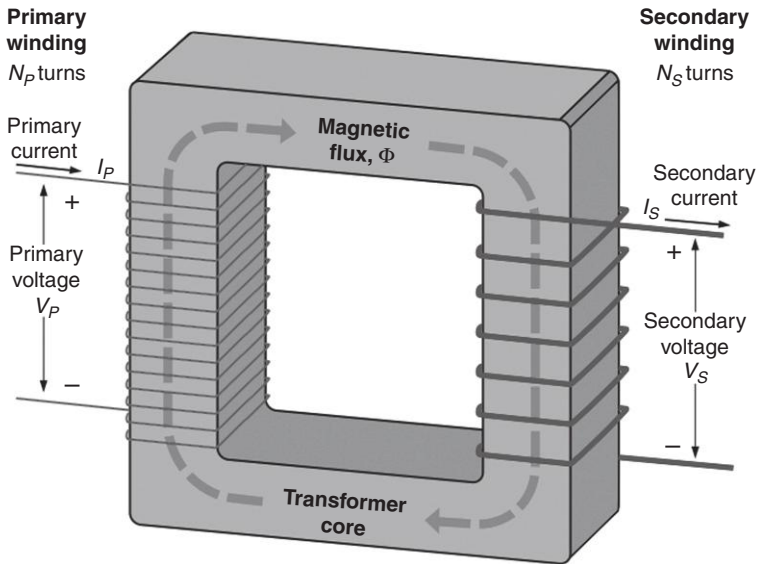
21.2 Transformer schematic. (*Source*: from Reference [12].)

current) conditions. Heat management is vitally important for ensuring reliable transformer operation. This is achieved through close monitoring of the internal temperature, with increasingly voracious heat dissipation strategies triggered as temperature rises.

Incidentally, this need to transform between voltage levels was recognized early in the development of power systems. Because transformers require AC voltages and currents, and cannot operate with DC, AC power systems triumphed over DC.[14] With the relatively recent advent of power electronics, however, transformation between DC voltage levels is now routine.

21.2.4   Sub-transmission and distribution

The physical size of transmission lines, their cost, and their power carrying capability make them unsuited for supplying electricity within localities. Transmission lines can be thought of as the freeways of the electricity supply system. At the regional level, sub-transmission and distribution lines play the role of the arterial roads and streets of the supply system. Sub-transmission lines operate in a voltage range around 40–120 kV. Their role is regional, taking electrical power from the transmission network to communities and large industrial loads. At those locations, the supply voltage is further reduced, typically to a level around 8–22 kV. This lower-voltage distribution network is able to span out along streets and throughout industrial sites to deliver electrical power to all consumers. The final step in that

delivery process is to transform the voltage down to the level expected by consumers, for example 120 V in North America or 230 V across Europe.

21.2.5   Substations

Substations establish interconnection points within power systems, with the switching status of circuit-breakers defining the network topology, and transformers enabling connections between voltage levels. Figure 21.3 illustrates a typical arrangement of substation equipment. In the center of the figure, the transformer (item 9) demarcates between higher-voltage equipment to the left and lower-voltage equipment to the right. At the points where the transmission lines connect to the substation, voltage transformers (item 4) measure the line voltages. Current transformers (CT) (item 7) measure the current flowing in the respective circuits. Circuit-breakers (item 6) are designed to provide switching capability under high current conditions, and are used to switch lines/transformers in and out of service. In contrast, disconnect switches (item 5) can open only if the current flow is near zero. Lightning arresters (item 8) prevent voltage spikes from damaging the transformer.

Figure 21.4a provides an alternative illustration of substation layout. This figure shows a double-busbar double-breaker arrangement, and identifies typical locations for the voltage and CTs that provide sensing capabilities within the substation. Further discussion of these sensors is provided in Section 21.3.

The example presented in Fig. 21.4 illustrates the role played by circuit-breakers in determining line connection arrangements. The substation layout of Fig. 21.4a shows four transmission lines, each terminating in two circuit-breakers. This arrangement allows each line to be connected to busbar A and/or B, depending on circuit-breaker status. Figure 21.4b shows two possible connection arrangements for the transmission lines. The upper case would apply if all circuit-breakers were closed. The lower case would be applicable if circuit-breakers CB11, CB21, CB32, and CB42 were closed, with the other four circuit-breakers open.

Almost all measurements of network voltages and currents are acquired from within substations. Likewise, most switching control actions involve substation-based circuit-breakers. Hence substations play a fundamentally
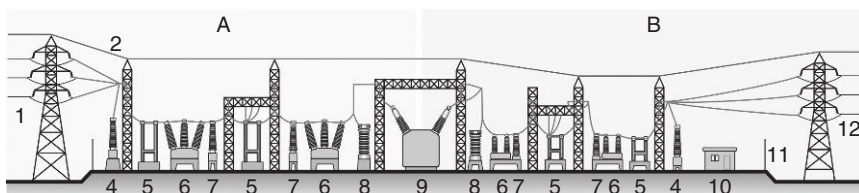
21.3 Substation layout. (*Source*: from Reference [15].)

(a)

Busbar A

CB11   CB21   CB31   CB41

CB12   CB22   CB32   CB42

Busbar B

L1   L3

L2   L4

Voltage tfmr

Current tfmr

(b)

L2   L4

L1   L3

L2   L4

L1   L3
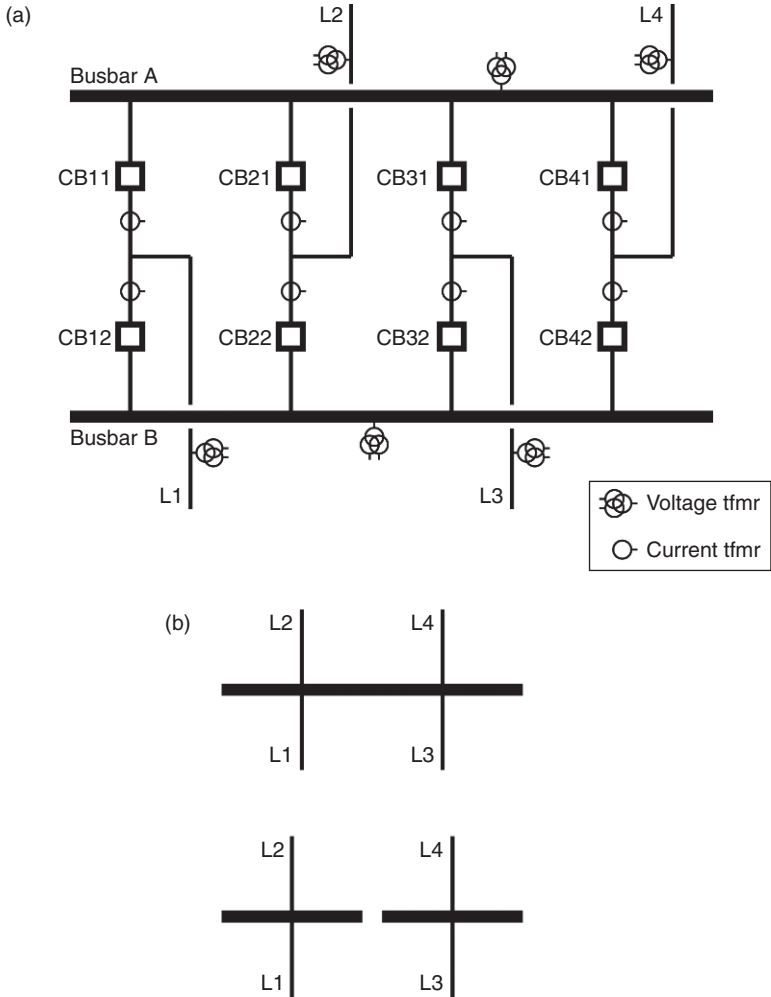
21.4 Circuit-breaker status dictates substation topology.

important role in the SCADA system that underpins power system monitoring and operations. Substation communications must therefore be reliable and secure. Within modern substations, interactions between intelligent electronic devices (IED), such as data acquisition devices, digital protection relays and phasor measurement units (PMU), occur over a high-speed local area network (LAN) and are governed by the IEC 61850 standard.[16] Data transfer between substations and the centralized SCADA system tends to be governed by IEC 60870 and DNP3 protocols.[17] A range of media are used for such communications, including power line carrier and fiber optic links.

SCADA systems for monitoring and controlling power systems may cover extensive geographical areas, and so integrity of long-distance communications is extremely important.

21.2.6   Power system operations

The physical characteristics of power systems dictate that if total generation exceeds total load plus losses, then the system frequency will increase. Likewise, a deficit of generation will result in frequency declining. To ensure the safety and integrity of large generators and motors, frequency must be kept close to its nominal value, for example 60 Hz in North America and 50 Hz in Europe. Therefore power system operations must continually maintain this generation/load balance. Unpredictable changes in loads occur continually, as electrical devices switch on and off. Closed-loop controls restore balance, though operators must ensure that sufficient generation is always available to respond to the control signals.

The total load of a power system typically exhibits a daily cycle, attaining a minimum overnight and rising to a maximum during the afternoon and into the early evening. The timing and shape of the peak depends upon location and season. Two example daily load cycles are provided in Fig. 21.5. Both are for the PJM market, with Fig. 21.5a showing a typical summer peak and Fig. 21.5b showing a winter peak. As the load increases from the morning trough, a large amount of generation must be progressively brought into service. Likewise, generation must be taken out of service as the total load reduces through the evening.

The large swings in generation and load give rise to large variations in power flow and voltage patterns across the transmission, sub-transmission, and distribution networks. Power system operations must ensure safety, reliability, and efficiency at all times. To do so requires extensive monitoring of system conditions and assessment of the impact of credible contingencies. All line flows and voltages must be within limits, and must remain so even with the loss of any single item of plant, such as a generator or transmission line.

21.2.7   Failures in monitoring and assessment

In August 2003, a major blackout affected a large section of the north-eastern United States and Canada.[19,20] While numerous events coincided to initiate this blackout, a breakdown in monitoring and assessment certainly played a crucial role. Physical events included the tripping of a heavily loaded 345 kV transmission line, and subsequent overloading and tripping of adjacent 345 kV lines. That left other 345 and 138 kV lines overloaded, and a cascade of line tripping quickly enveloped the region.

(a)

(b)
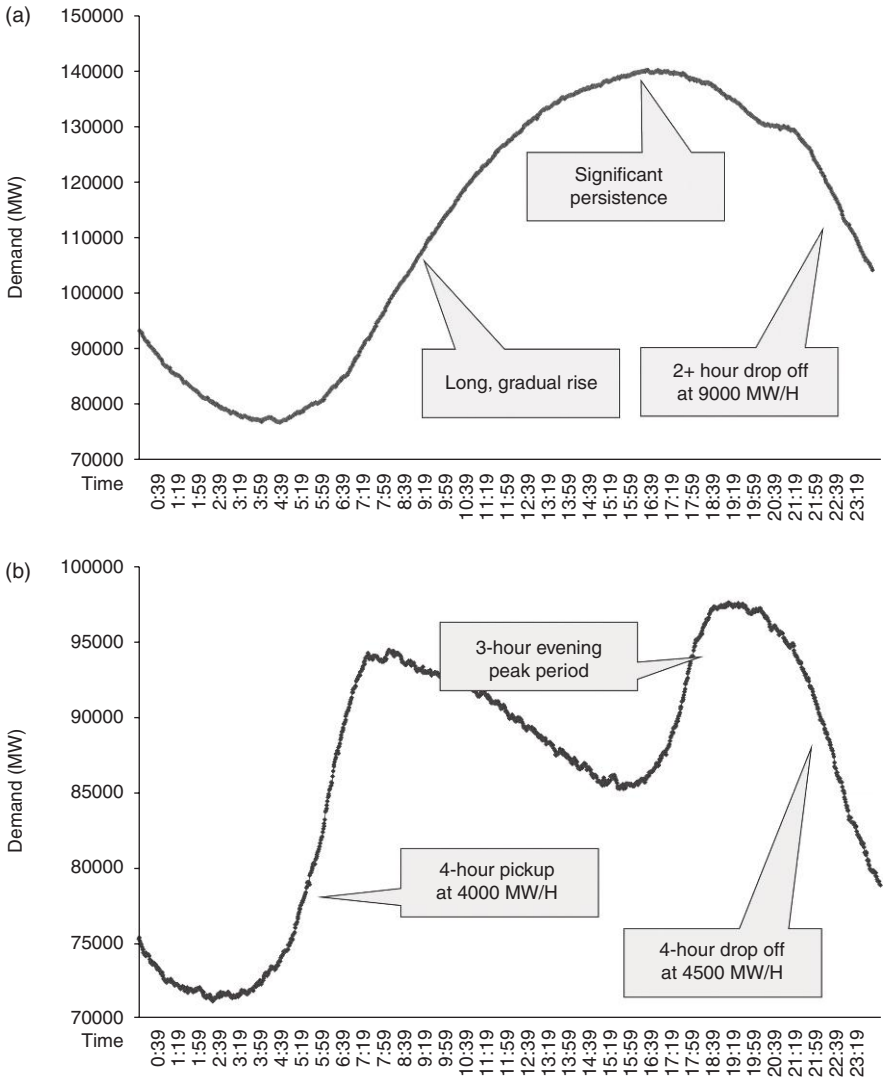
21.5 Daily load cycles for the PJM market. (*Source*: © 2012 PJM. Reprinted with permission from Reference [18].)

The events unfolded as they did because the SCADA system covering that section of the grid had suffered a software fault and was not providing an up-to-date view of grid conditions. Operators were therefore oblivious to the precarious state of the grid, and so did not act to correct the situation. It could be argued that if the SCADA system had been providing accurate information, operators would have acted quickly to reduce the flow on the

heavily overloaded lines, thereby preventing the ensuing cascade and massive loss of load.

21.3   Sensing equipment and systems

The high voltages and currents of power systems preclude direct measurement of power system quantities. Rather, transformers are required to establish scaled replicas of the actual voltage and current waveforms.[21] These scaled signals can then be interfaced with standard low-voltage measurement equipment.

21.3.1   Current transformers (CT)

CTs are conceptually like other transformers, with a primary (power system) winding and a secondary (measurement system) winding. Referring to Fig. 21.2, the secondary winding consists of N_s turns, which are evenly distributed around a magnetic core.

In a CT, that core encircles a single conductor, which forms the primary winding and carries the current to be measured. It follows from Equation [21.1] that with $N_p = 1$, the relationship between primary and secondary current is given by,

$$I_s = \frac{1}{N_s} I_p \qquad [21.2]$$

The two most common types of CTs are shown in Fig. 21.6. In both cases, a porcelain bushing provides separation between the high voltage at the top and the grounded base. Hairpin/tank CTs have their secondary cores located in the tank at the base of the bushing. This requires the primary conductor to have a hairpin shape so that it may connect to the external system at the top yet pass through the cores at the base. The primary conductor is at high voltage, but passes close to the grounded base. It must therefore be wrapped in sufficient insulation to prevent flash-over. Top-core CTs, on the other hand, have their secondary cores on the high-voltage side of the bushing. This has the advantage that the primary conductor is quite short. However, with the weight of the cores at the top of the structure, the supporting bushing must be much stronger than for tank CTs. The high center of mass is particularly troublesome in areas where seismic activity is common. Also, the secondary windings are in close proximity to high voltages, and so must be sufficiently well insulated.

Based on Equation [21.2], current must flow in the secondary circuit whenever current flows in the primary (power system) conductor. Because
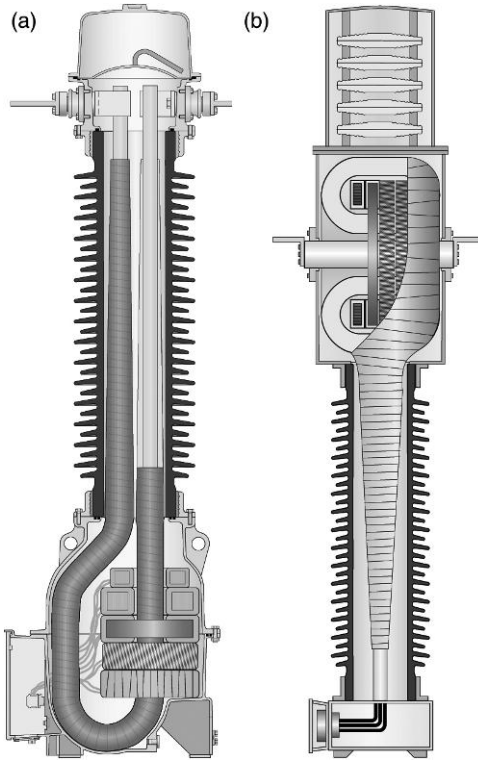
21.6 Common types of CT: (a) hair-pin/tank CT and (b) top-core CT. (*Source*: © 2009 ABB. Reprinted with permission from Reference [22].)

primary current is dictated by system conditions, the secondary circuit must always allow current flow. It should therefore never be open-circuited. If an open-circuit were to occur, current would be forced to flow through the (almost) infinite impedance of the open-circuit, producing a dangerously high voltage and destructive arcing.

21.3.2   Voltage transformers

The primary side of a voltage transformer (VT) is connected between one phase of the power system and ground. The secondary side is connected to the measurement system. For voltage levels below 100–115 kV, VTs use standard inductive coupling, as illustrated in Fig. 21.2. At higher voltage levels, the winding ratio N_p/N_s becomes excessive and is physically challenging to implement. In such cases, capacitor voltage transformers (CVT) are more economical. A CVT consists of a capacitive voltage divider connected to a standard inductive transformer. Figure 21.7a shows this arrangement diagrammatically.

(a)

(b)
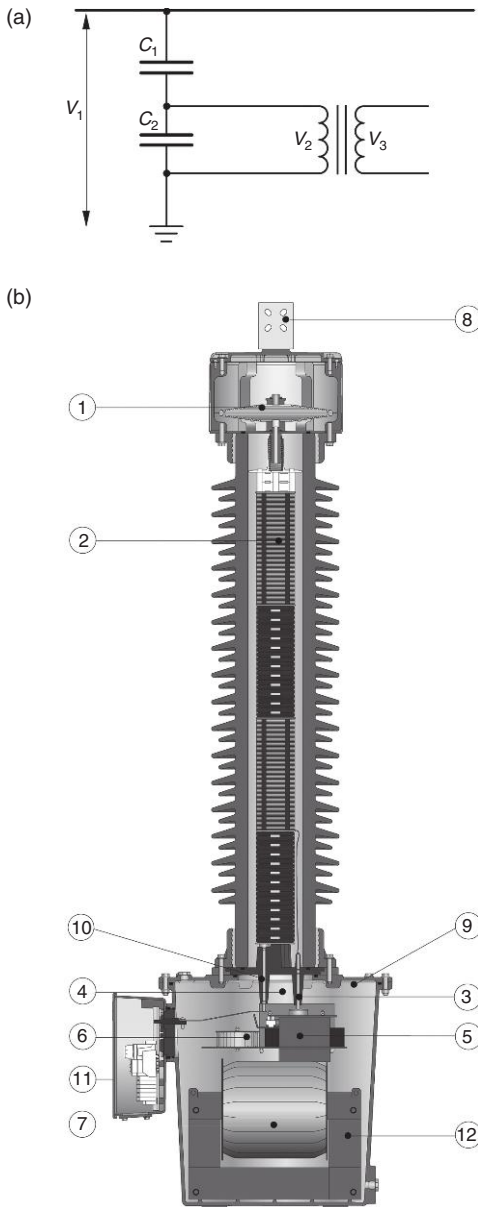
21.7 Capacitor voltage transformer. (a) Equivalent circuit. (b) Cutaway view. (*Source*: © 2009 ABB. Reprinted with permission from Reference [22].)

The voltage divider establishes the voltage relationship,

$$\frac{V_2}{V_1} = \frac{C_1}{C_1 \square C_2}$$

[21.3]

Typical CVT designs set the nominal value for voltage V_2 around $22/\sqrt{3}$ kV. Given the primary voltage V_1, the relationship Equation [21.3] establishes the required ratio of capacitors C_1 and C_2. Consider, for example, voltage measurement on the 345 kV system. As discussed in the Appendix, the corresponding phase-to-neutral voltage is $V_1 = 345/\sqrt{3}$ kV. From Equation [21.3], the required capacitor ratio is $C_2/C_1 = 14.68$. The output voltage V_3 is normally around $110/\sqrt{3}$ V, though that choice may vary with local utility standards. For the case where $V_2 = 22/\sqrt{3}$ kV and $V_3 = 110/\sqrt{3}$ V, the required turns ratio for the output transformer is $N_p/N_s = 200$. Apart from providing voltage scaling, the output transformer also serves to isolate the measurement system from the primary power system.

Figure 21.7b shows a cutaway view of a CVT. The capacitive elements (item 2) are built into the bushing that separates the high-voltage terminal from the grounded base. The output transformer, consisting of its primary/secondary windings (item 7) and core (item 12), resides in the base.

21.3.3   Performance requirements

Voltage and current transformers must provide accurate replicas of the actual measured quantities under normal and fault conditions. Under normal operating conditions, voltages and currents are sinusoidal and transformation from the primary high voltage/current signals to secondary measurement-level signals is straightforward. Very high accuracy is required in certain applications, for example electricity markets. In that case, accurate monitoring of generation and load is vital, as financial transactions depend upon the monitored quantities.

Under fault conditions, currents may rise to levels that are a factor of 20 or more times normal load currents. CTs must be designed to minimize the effects of magnetic saturation at such high current levels; otherwise secondary-side signals will be highly distorted.[21] Saturation is also of concern during transients, such as when a transmission line is energized (switched into service) or at the onset of a fault. Under such conditions, primary currents experience a transient DC offset. This can lead to asymmetric saturation of the CT and significant distortion in the secondary output. As the DC component decays, the accuracy of the reproduction recovers.

21.3.4   Measurement systems

Voltage and current measurements are used for a variety of purposes, in particular protection, monitoring, and metering. Protection schemes, such as over-current, differential, and distance protection, use measurements to

identify abnormal conditions and determine the appropriate trip signals.[4] Extremely high currents are indicative of a fault condition and are detected by over-current protection. Differential protection monitors the net current flow into and out of a device or a section of a power system, with a non-zero net summation indicative of an internal fault. Distance protection monitors both the voltage and current of a transmission line, and computes the ratio *V/I*. This ratio takes a small value when a fault occurs on the line, and in fact provides a good indication of the distance to the fault along the line. For protection applications, CTs and VTs require fast transient response, and must provide a good quality representation of the actual system quantities under highly stressed conditions.

Monitoring and metering are concerned with tracking system conditions through normal daily variations. They differ in that metering is used for financial transactions, and therefore must have high accuracy. A wide range of power system quantities must be monitored in substations and telemetered via the SCADA system: voltages and currents; active and reactive power of generators, loads and line flows; protection and circuit-breaker status; and transformer temperature and tap position. Note, though, that active and reactive power are not directly measurable. Rather, they are computed from the available voltage and current measurements. Propagation of measurement errors through this calculation must be carefully considered.

Weather monitoring has always played an important role in power system operation. Electricity demand is closely correlated with weather, and so knowledge of weather patterns is useful for assessing generation and transmission needs. Storm tracking is also very helpful for determining areas where the grid may be vulnerable to lightning strikes or strong winds. Weather monitoring is also very important for adaptive rating of transmission lines.[23] Usually the rating of a line, i.e., its maximum current carrying capability, is dependent upon its ability to dissipate heat arising from I^2R losses. Accordingly, a line's rating will be much higher on a windy winter day than on a hot, still summer day. Without up-to-date weather information, line rating must be based on a conservative estimate of weather conditions. Therefore, weather monitoring facilitates more effective use of transmission assets.

21.3.5   Distribution network sensing

Unlike transmission networks, sensing capabilities across lower-voltage distribution networks have traditionally been rather limited. This is changing though, as utilities move to implement 'advanced metering infrastructure' (AMI).[24] This infrastructure takes various forms, but typically consists of 'smart meters' that monitor the real-time energy use of customers and transmit that data over a LAN to a point where the data are consolidated. It

is then forwarded over broadband to the utility's data management facility. Some AMI implementations allow utilities to pass information back to customers, enabling active demand control. Communication latency may be significant though, depending on the particular form of implementation.

In numerous installations, AMI also collects voltage and current measurements from across the distribution network, enabling utilities to assess network conditions in near real-time. Based on this information, the network can be optimized to reduce losses and/or improve the voltage profile. Under outage conditions, the data available from AMI can be used to provide a fairly precise estimate of the location of the open-circuit, enabling faster restoration of service.

As mentioned in the Appendix, it is generally assumed that three-phase voltages and currents are balanced. At the transmission level, the influence of individual loads is negligible due to averaging effects, so balanced conditions apply. At the distribution level, however, customer loads may affect individual phase voltages and currents, resulting in unbalanced conditions. Utilities attempt to balance customer loads equally across the three phases, but perfect balancing is impossible. The resulting unbalance has a number of undesirable effects, including high neutral currents and motor overheating.[25] Where unbalance is particularly troublesome, 'negative phase sequence' meters can be installed to monitor the extent of the problem as system conditions vary.

21.3.6   Phasor measurement units (PMU)

As explained in the Appendix, AC voltages and currents are usually represented by complex numbers, $\square = V \angle \theta_v$ and $\square = I \angle \theta_i$ respectively, which are known as phasors. The magnitudes V and I can be provided by standard substation measurement equipment, as discussed earlier. The angles θ_v and θ_i describe the phase offset of the AC (time domain) voltage and current waveforms relative to a reference cosine that maintains nominal system frequency (60 Hz in North America, for example) and that is synchronized to Coordinated Universal Time (UTC). Synchronization to UTC is achieved by accessing timing signals from the global positioning satellite (GPS) system, and ensures that the reference is well defined across an entire grid. Standard measurement equipment does not have access to GPS timing, and so cannot provide phase information.

Synchronized phasor (synchrophasor) measurement systems have been developed to provide accurate values of both the magnitude and phase of measured voltages and currents, thereby providing full phasor information. The monitoring devices used to generate synchrophasors are referred to as PMUs, with IEEE standard C37.118.1–2011[26] establishing performance and compliance requirements. According to the standard, PMUs must also
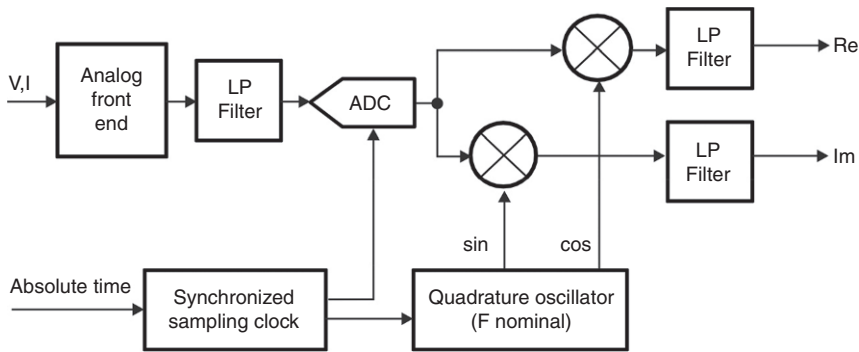
21.8 PMU signal processing model. (*Source*: © 2011 IEEE. Reprinted with permission from Reference [26].)

provide accurate estimates of frequency and rate-of-change of frequency, and must time-tag their outputs. Figure 21.8 provides a block diagram representation of a typical PMU implementation. The analog voltage and current signals are initially filtered to prevent aliasing, and are then sampled. The sampled signals undergo complex demodulation using an oscillator output that is synchronized with absolute time provided by a GPS. Low-pass filtering produces the real and imaginary components of the complex phasor, relative to the UTC reference. The standard does not specify implementation details, with other signal processing algorithms permitted. The PMU output must accurately time-stamp the input signals, and therefore must take into account the processing delay. PMUs typically provide an output rate of 30 samples per second.

The high accuracy and time synchronization of magnitude and phase measurements provided by PMUs, together with their rapid update rate, enable a range of monitoring and control applications.[27,28] Wide-area monitoring provides operators with enhanced visibility of large sections of the grid, enabling clearer assessment of the state of the system and faster appraisal of emerging problems. For example, as power system loading increases, voltage angles across the system tend to diverge, with heavily stressed conditions giving rise to large angle separation. The synchronized angle information provided by PMUs enables direct monitoring of this valuable indicator of system robustness.[29] Apart from real-time operational benefits, the high sample rate and time synchronization of PMU measurements make them particularly valuable for post-disturbance investigations that seek to clarify initiating conditions and the subsequent sequence of events. A number of wide-area monitoring schemes are currently in operation around the world.[30,31] Development of such schemes is on-going, with increasing deployment of PMUs and continual enhancement of data handling capabilities.

PMU measurements provide valuable inputs for state estimation, which is discussed in Section 21.4.1. State estimator convergence and solution quality benefit greatly from the high accuracy of PMU measurements and availability of angle difference measurements.[32] The high sampling rate and accuracy of PMU measurements are also well suited to closed-loop control applications that require inputs from remote locations. This capability has been explored extensively in applications ranging from stabilization[33] to the provision of reactive power support.[34] However, issues such as communication latency, sample loss and cyber security are currently an impediment to more expansive applications of PMU-based feedback.

21.3.7   Electromechanical mode monitoring

Power systems experience low-frequency electromechanical oscillations due to dynamic interactions between generators.[6] These oscillations are continually excited by small random load changes. Under normal operating conditions, these electromechanical modes are very well damped. However, as a power system becomes stressed, through excessive load growth or tripping of key generators and/or transmission lines, damping tends to reduce and the oscillations become more pronounced. This is illustrated in the system recording of Fig. 21.9, which shows power flow during the major event documented in Reference [35]. Prior to the tripping of a 500 kV transmission line at 15:42, oscillations were present but were insignificant. Line tripping weakened the system, as indicated by the poorly damped transient response. Over the following 5 min period, the background oscillations were generally larger than they were before the line outage, suggesting a reduction in damping. The second line trip at 15:47 actually led to system instability, as indicated by the subsequent growing oscillations.

For power systems that operate near their stability limit, damping is a good indicator of system resilience. With increasing availability of high quality measurements from PMUs, it is becoming common for utilities to install monitors that estimate modal frequency and damping in near real-time.[36] Various techniques are in use for determining this estimate.[37] Prony analysis[38] is well suited to identifying the modes present in measurements of system disturbances. Least-squares algorithms[39,40] have been developed for estimating the modal content of ambient (background) conditions, where excitation is provided by small, random load changes. Even though these two approaches are conceptually quite different, it is shown in Reference [39] that they are effectively equivalent.

Several forms of least-squares methods have been proposed, covering both block and recursive algorithms. In a general sense, these methods use
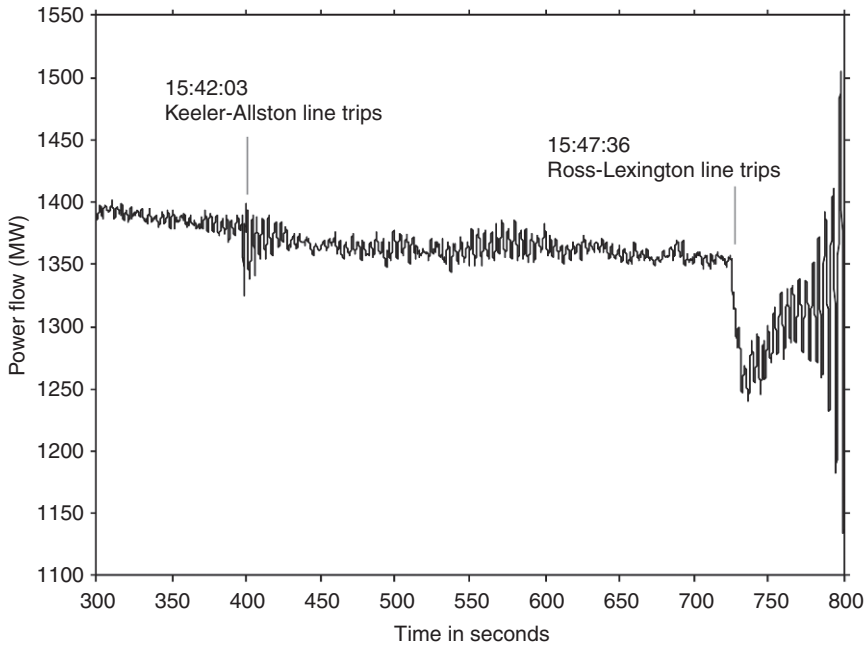
21.9 Oscillations recorded during WECC disturbance. (*Source*: © 1999 IEEE. Reprinted with permission from Reference [35].)

autoregressive (AR) type models, establish a linear prediction from past measurements, and minimize the prediction error by adjusting the model coefficients. The polynomial defined by these estimated coefficients approximates the characteristic polynomial of the system. Therefore, the desired system modes can be obtained from the roots of this polynomial.

21.4   Control center monitoring and assessment

Control centers have the task of monitoring geographically expansive power systems. This is achieved through a SCADA system that communicates with substation-based data acquisition systems. Typical SCADA systems obtain updated measurement data from across the power system every 2–4 s. These data are vital for following trends in system behavior, and are presented to operators through computer graphical user interfaces and wallboard displays, as shown in Fig. 21.10.

The usefulness of SCADA data is limited though, due to inherent measurement errors. To undertake more detailed evaluation of the system's operational condition, it is first necessary to filter the measurements to obtain a consistent and accurate estimate of all system quantities. This process is
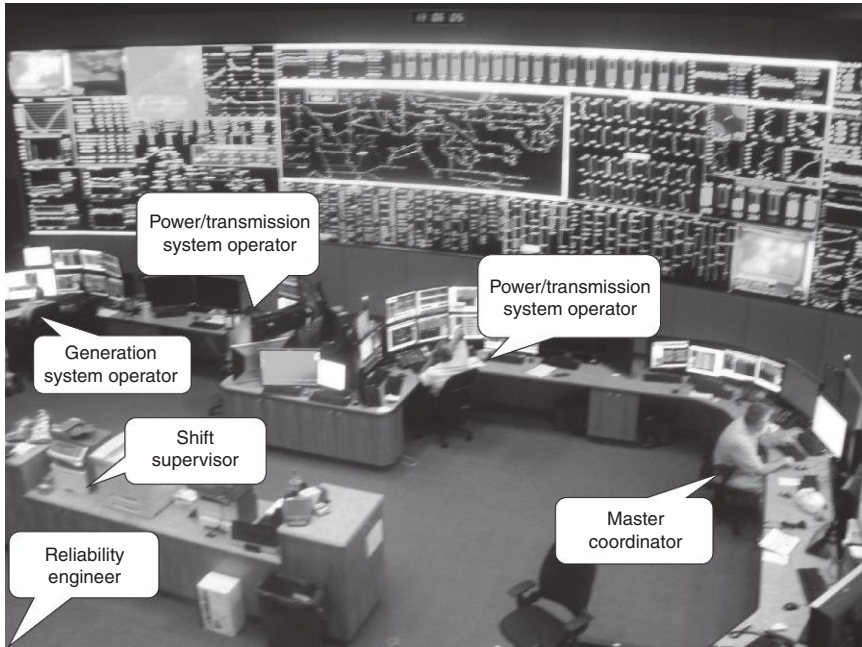
21.10 Presentation of SCADA measurement data. (*Source*: © 2012 PJM. Reprinted with permission from Reference [18].)

called state estimation. Once the current state of the system is known with confidence, it is then possible to undertake a thorough assessment of the system's security and resilience.

21.4.1   State estimation

Nonlinear weighted least-squares algorithm

The state of a system is defined as the smallest set of variables from which all other system quantities can be computed. In other words, knowledge of the system state implies complete knowledge of the system. In considering power systems, it is commonly assumed that the network topology and parameters such as line impedances are known. Then the set of all (complex) node voltages fulfills the requirements for the system state. For instance, knowing the voltages at the two end nodes of a line allows the active and reactive power flow on that line to be computed. The generation/load at a node can then be determined by summing all the line flows out of that node. Hence all line flows and node injections can be computed from the node voltages.

The first step in computing the state estimate is to build an up-to-date model of the network topology. This is achieved by using telemetered circuit-breaker status information, along with substation models of the form shown in Fig. 21.4a, to establish the prevailing node–branch model, for example Fig. 21.4b. The desired state estimate is then obtained by fitting the telemetered measurements to the underlying model.[41,42] Each measurement can be related to the estimated state through an equation of the form,

$$z_i = h_i(x) + \epsilon_i \qquad [21.4]$$

where z_i is the i-th measurement (a known quantity), x is the desired state vector, h_i is a nonlinear function relating the state vector to the i-th measured quantity, and ϵ_i is some (unknown) random error in the i-th measurement. The function h_i is given by the underlying 'power flow' equations.[41]

The general aim of state estimation is to determine the value of the state x that minimizes, in an overall sense, the errors ϵ_i across the entire measurement set. This minimization must, however, take into account the accuracy associated with each measurement. This can be achieved by normalizing the errors by a weighting factor that reflects measurement accuracy.

Measurement errors are random and are commonly assumed to satisfy a Gaussian (normal) distribution. The variance σ_i^2 associated with the i-th measurement provides an indication of its spread, with more accurate measurements having smaller variance. Furthermore, it follows from the properties of the normal distribution that if ϵ_i has variance σ_i^2 then ϵ_i/σ_i has a variance of 1. Normalized errors are therefore equivalent.

Given this equivalence of normalized errors, an indication of the overall error across all measurements can be expressed as $\sum_i (\epsilon_i/\sigma_i)^2$. Using Equation [21.4], the error minimization problem has the nonlinear weighted least-squares form,

$$\hat{x} = \arg\min J(x) = (z - h(x))^{\mathrm{T}} R_z^{-1} (z - h(x)) \qquad [21.5]$$

where $R_z \equiv \mathrm{diag}(\sigma_1^2 \cdots \sigma_m^2)$ is the covariance matrix and $\hat{x}$ is the desired estimate of the state. A minimum can be obtained using a Gauss-Newton iterative process,

$$\Delta x^k = \left(H(x^k)^{\mathrm{T}} R_z^{-1} H(x^k) \right)^{-1} H(x^k)^{\mathrm{T}} R_z^{-1} (z - h(x^k)) \qquad [21.6]$$

$$x^{k+1} = x^k - \alpha \Delta x^k \qquad [21.7]$$

where x^k is the estimate of the state at the k-th iteration, $H(x^k)$ is the Jacobian matrix $\delta h/\delta x$ evaluated at x^k, and α is an acceleration factor.

It should be noted that the inverse of the *gain matrix* $H^T R_z^{-1} H$ in Equation [21.6] is not actually computed. Rather, the equation is generally solved using Cholesky factorization,[42] with the desired solution Δx^k of Equation [21.6] obtained using forward and backward substitution. If the gain matrix is ill-conditioned, i.e., almost singular, Cholesky factorization may encounter numerical difficulties. In such cases, QR-factorization is preferred.[43] It is slower, but has better numerical properties.

A necessary condition for the gain matrix to be nonsingular is that the number of measurements must exceed the number of state variables, $\dim(z) \geq \dim(x)$. The difference $\dim(z) - \dim(x)$ is referred to as the measurement redundancy. The accuracy of the state estimate generally improves with increasing redundancy. Note, though, that having redundant measurements does not guarantee nonsingularity of the gain matrix. For example, even if redundancy is high across most of the system, singularity would occur if a group of substations were devoid of any measurements. Clearly, both the number and location of measurements are important. This is discussed in the following section.

Observability

Solvability of the state estimation problem Equation [21.5] through the use of Equations [21.6]–[21.7] ensures that the estimated value of the system state $\hat{x}$ is well defined[2] for the given set of measurements. In other words, the state is observable from that measurement set. Conversely, singularity of the gain matrix, and consequently insolvability of Equation [21.6], implies that the measurement set is deficient. In such a case, the factorization process employed to solve Equation [21.6] will fail, and hence the state cannot be estimated (observed) from the given measurements. It is therefore a standard practice for a check to be made of observability before to solving the state estimator. The role of observability analysis is firstly to determine whether the entire system state is observable. If it is not, the subsequent task is to establish the maximal observable subnetwork, and the minimum set of extra injection measurements (pseudo-measurements) that would restore global observability.

Observability analysis is usually undertaken using a topological tree-spanning algorithm. The basis for this approach is that if a bus voltage is measured together with the line flow from that bus to a neighbor, then the

---

[2] The minimization Equation [21.5] is non-convex due to the nonlinearity of the power flow equations $h(x)$. However, if Equation [21.6] is solvable then isolated solutions of Equation [21.5] are generally well defined.

voltage at the neighboring bus can be estimated. (This follows from basic circuit properties.) This process, together with other similar rules, can be used to grow trees that span the observable regions of the network. If the entire network is observable, then state estimation can proceed. If that is not the case, it is common practice for fictitious pseudo-measurements to be progressively added until full network observability is achieved. Often forecast values of loads are used for these extra 'measurements.' Their number and location must be carefully chosen to ensure that they do not corrupt the estimate attainable from the true measurements.

Bad measurements

It is not uncommon for measurements to occasionally drift or lose fidelity. Identifying such bad measurements is a standard function of state estimation. Two statistical tests are used for this purpose. Because the measurements are assumed to be normally distributed, the random variable J given by Equation [21.5] is distributed according to the χ^2 (chi-squared) distribution. Therefore the χ^2-test can be used to assess the likelihood that the measurement set is grossly in error. Such large errors generally occur only if incorrect circuit-breaker status causes a mismatch between the topology of the model and that of the actual system, or if a significant set of telemetered data becomes corrupted. The latter could occur as a result of a failure in the communications from one or more substations.

The quality of individual measurements can be assessed using the residual test. The measurement residual is the estimate of the error, $\epsilon_i = z - h_i\left(\hat{x}\right)$, where $\hat{x}$ is the estimated state. The residual is normally distributed with zero mean and a variance that is derived from σ_i^2. Typically, residuals that lie outside the 95% confidence interval are rejected as bad data. Such measurements are removed from the active measurement set, and flagged as requiring attention from field technicians.

21.4.2   Security assessment

SCADA data and state estimation results provide a comprehensive snapshot of the prevailing conditions across the power system. Operators continually monitor these conditions to ensure that system quantities remain within range and follow anticipated patterns. In addition, they must also regularly evaluate the consequences of credible contingencies, such as the outage of a major generator or transmission line. This process is referred to as contingency analysis, and uses power flow simulation to investigate the effects of postulated outages. The power flow solves the nonlinear, algebraic equations governing power system steady-state conditions. Usually the state estimator establishes the base case, and then an automated process solves the power

flow to evaluate system performance for each contingency in a (possibly large) set of cases. If any anomalies are identified, operators are notified and can assess the risk of the troublesome contingencies. If necessary, they will then determine and implement appropriate preventative actions to ensure that the system would continue to operate securely should those events actually occur. Examples of such actions include rescheduling generation, reconfiguring substation connections, and switching capacitors/reactors.

In an on-going drive to improve the efficiency of power systems, operators are increasingly making use of security-constrained optimal power flow (SCOPF) tools.[3,44] These tools seek a minimum cost solution, typically minimum losses, that satisfies the power flow constraints in the presence of a set of contingencies. The state estimator again provides the starting point for the solution algorithm. The solution to the SCOPF provides operators with the sensitivity information required to maintain optimal and secure power system conditions.

Not all insecure operating conditions can be identified using power-flow-based contingency analysis. Certain events may induce unacceptable dynamic behavior that ultimately leads to system degradation or even instability. Assessment of such dynamic phenomena typically requires repeated numerical integration, which is time consuming and therefore not suited to an on-line environment. In such cases, it is common for an extensive range of operating scenarios to be evaluated off-line, and the results summarized in the form of nomograms.[45] Each nomogram considers a pair of system quantities that are indicative of potentially troublesome operation, and provides a graphical representation of the boundary between acceptable and unacceptable conditions. The boundary describes a curve on a planar diagram, with the given pair of measured values defining a point on that diagram. The proximity of system conditions to each nomogram's boundary is continually monitored. If potentially insecure conditions are detected, operators will evaluate the risk and, if necessary, take action to restore the system to within the secure regions of the nomograms.

Off-line assessment can never duplicate actual system operation. Also, the boundary separating secure from insecure conditions is generally multi-dimensional, with the two-dimensional nomograms limited in their ability to provide an equivalent relationship. As a result, nomograms tend to give a conservative assessment of system security. Addressing this conservativeness requires the development of more sophisticated analysis tools that are suited to an on-line environment. Work in this area is on-going.[45,46]

21.5   Conclusion

Extensive sensing systems are required to ensure the safe and secure operation of large-scale power systems. The high voltages and currents of power

systems cannot be measured directly, but rather require VTs and CTs to interface between the high-voltage network and the low-voltage measurement system. The outputs of such instrument transformers are typically around 120 V and 5 A, and must accurately replicate high voltage quantities, even under extreme fault and transient conditions. The signals provided by instrument transformers are used directly for both protection and monitoring, and enable the computation of other non-measurable quantities such as active and reactive power. Recent developments in digital protection and PMUs allow measurements to be accurately time tagged. This forms the basis for computing phase angle differences between remote locations. For lower-voltage distribution networks, recent developments in AMI are enabling widespread measurement of voltages and currents at the consumer level.

Measurements from across power systems are communicated to control centers via SCADA systems. Substations play an important role in SCADA, as they establish the interconnection points within power systems, and therefore house the major sensing systems. At control centers, measurements from SCADA are filtered using a nonlinear weighted least-squares algorithm to establish a consistent estimate of the entire system state. State estimator results, together with PMU data, are used in a variety of analysis and optimization applications that assist operators in assessing system security and determining optimal corrective actions.

21.6   References

1. G. Constable and R. Somerville (2003), *A Century of Innovation: Twenty Engineering Achievements that Transformed Our Lives*. Joseph Henry Press, Washington DC.
2. R. Miller and J. Malinowski (1994), *Power System Operation*, 3rd edn. McGraw-Hill, New York, NY.
3. A. Wood and B. Wollenberg (1996), *Power Generation, Operation and Control*, 2nd edn. John Wiley and Sons, Inc., New York.
4. J. Blackburn (1998), *Protective Relaying Principles and Applications*, 2nd edn. Marcel Dekker, New York.
5. P. Sauer and M. Pai (1998), *Power System Dynamics and Stability*. Prentice Hall, Upper Saddle River, NJ.
6. P. Kundur (1994), *Power System Stability and Control. EPRI Power System Engineering Series*. McGraw Hill, New York, NY.
7. IEA Wind (2011) Annual Report. International Energy Agency, July 2012.
8. T. Ackermann (Editor) (2005), *Wind Power in Power Systems*. John Wiley and Sons, England.
9. P. Lynn (2010), *Electricity from Sunlight: An Introduction to Photovoltaics*. Wiley, Chichester, West Sussex, UK.
10. G. Masters (2004), *Renewable and Efficient Electric Power Systems*. Wiley-Interscience, Hoboken, NJ.

11. D. Douglas and R. Thrash (2007), 'Sag and tension of conductor', in *Electric Power Generation, Transmission, and Distribution*, 2nd edn. L.L. Grigsby (Editor), CRC Press, Boca Raton, FL, pp. 14.1–14.41.

12. 'Transformer', Wikipedia, The Free Encyclopedia. [Online]. Available: http: // en.wikipedia.org/wiki/Transformer.

13. M. Heathcote (2007), *J&P Transformer Book*, 13th edn. Elsevier, Oxford, UK.

14. J. Jonnes (2004), *Empires of Light: Edison, Tesla, Westinghouse, and the Race to Electrify the World*. Random House, New York, NY.

15. 'Electrical substation', Wikipedia, The Free Encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki/Electrical substation.

16. R. Mackiewicz (2006), 'Overview of IEC 61850 and benefits', in *Proceedings of the IEEE Power Engineering Society General Meeting*, Montreal, Quebec, July 2006.

17. A. West (2005), 'SCADA and substation control communications', in *Proceedings of the Southern African SCADA and MES Conference – IDC Technologies*, Johannesburg, South Africa.

18. 'PJM101: The basics (2012)', PJM Training Material, October 2012. [Online]. Available: http://www.pjm.com/training/training-material.aspx.

19. U.S.-Canada Power System Outage Task Force (2004), Final Report on the 14 August 2003 Blackout in the United States and Canada: Causes and Recommendations, April 2004.

20. G. Andersson, P. Donalek, R. Farmer, N. Hatziargyriou, I. Kamwa, P. Kundur, N. Martins, J. Paserba, P. Pourbeik, J. Sanchez-Gasca, R. Schulz, A. Stankovic, C. Taylor and V. Vittal (2005), 'Causes of the major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance', *IEEE Transactions on Power Systems*, vol. **20**, no. 4, pp. 1922–1928, November 2005.

21. R. Millikin (2012), 'Instrument transformers', in *Electric Power Transformer Engineering*, 3rd Edition, J.H. Harlow (Editor). CRC Press, Boca Raton, FL, pp. 7.1–7.35.

22. K. Sjövall and M. Findell (2009), 'Instrument transformers application guide', ABB AB High Voltage Products, Ludvika, Sweden.

23. S. Maslennikov and E. Litvinov (2009), 'Adaptive emergency transmission rates in power system and market operation', *IEEE Transactions on Power Systems*, vol. **24**, no. 2, pp. 923–929, May 2009.

24. 'Advanced metering infrastructure (2008),' National Energy Technology Laboratory (NETL) Modern Grid Strategy, February 2008.

25. V. Gosbell, S. Perera and V. Smith (2002), 'Voltage unbalance', Technical Note No. 6, Integral Energy Power Quality Centre, October 2002.

26. IEEE Std C37.118.1–2011 (2011), IEEE Standard for Synchrophasor Measurements for Power Systems. Institute of Electrical and Electronics Engineers, Inc., New York.

27. V. Terzija, G. Valverde, D. Cai, P. Regulski, V. Madani, J. Fitch, S. Skok, M. Begovic and A. Phadke (2011), 'Wide-area monitoring, protection, and control of future electric power networks', *Proceedings of the IEEE*, vol. **99**, no. 1, pp. 80–93, January 2011.

28. 'Real-time application of synchrophasors for improving reliability (2010),' *North American Electric Reliability Corporation (NERC)*, Princeton, NJ, October 2010.

29. J. Ballance, B. Bhargava and G. Rodriguez (2003), 'Monitoring power system dynamics using phasor measurement technology for power system dynamic security assessment', in *Proceedings of the IEEE Bologna PowerTech Conference*, Bologna, Italy, June 2003.

30. J. Cai, Z. Huang, J. Hauer and K. Martin (2005), 'Current status and experience of WAMS implementation in North America', in *Proceedings of the IEEE PES Transmission and Distribution Conference and Exhibition, Asia and Pacific,* Dalian, China, August 2005.

31. I. Kamwa, J. Béland, G. Trudel, R. Grondin, C. Lafond and D. McNabb (2006), 'Wide-area monitoring and control at Hydro-Québec: Past, present and future', in *Proceedings of the IEEE Power Engineering Society General Meeting*, Montreal, Quebec, July 2006.

32. A. Gómez-Expósito, A. Abur, P. Rousseaux, A. de la Villa Jaén and C. Gómez-Quiles (2011), 'On the use of PMUs in power system state estimation', in *Proceedings of the 17th Power Systems Computation Conference*, Stockholm, Sweden, August 2011.

33. I. Kamwa, R. Grondin and Y. Hébert (2001), 'Wide-area measurement based stabilizing control of large power systems – A decentralized/hierarchical approach', *IEEE Transactions on Power Systems*, vol. **16**, no. 1, pp. 136–153, February 2001.

34. A. Johnson, R. Tucker, T. Tran, J. Paserba, D. Sullivan, C. Anderson and D. Whitehead (2007), 'Static var compensation controlled via synchrophasors', in *Proceedings of the Western Protective Relay Conference*, Spokane, WA, October 2007.

35. D. Kosterev, C. Taylor and W. Mittelstadt (1999), 'Model validation for the August 10, 1996 WSCC system outage', *IEEE Transactions on Power Systems*, vol. **14**, no. 3, pp. 967–979, August 1999.

36. D. Wilson (2005), 'Continuous damping measurement for power system analysis', in *Proceedings of the IEEE Power Tech*, St. Petersburg, Russia, June 2005.

37. D. Trudnowski and J. Pierre (2009), 'Overview of algorithms for estimating swing modes from measured responses', in *Proceedings of the IEEE Power and Energy Society General Meeting*, Calgary, Alberta, Canada, July 2009.

38. J. Hauer (1991), 'Application of Prony analysis to the determination of modal content and equivalent models for measured power system response', *IEEE Transactions on Power Systems*, vol. **6**, no. 3, pp. 1062–1068, August 1991.

39. J. Pierre, D. Trudnowski and M. Donnelly (1997), 'Initial results in electromechanical mode identification from ambient data', *IEEE Transactions on Power Systems*, vol. **12**, no. 3, pp. 1245–1251, August 1997.

40. N. Zhou, J. Pierre, D. Trudnowski and R. Guttromson (2007), 'Robust RLS methods for online estimation of power system electromechanical modes', *IEEE Transactions on Power Systems*, vol. **22**, no. 3, pp. 1240–1249, August 2007.

41. A. Monticelli (1999), *State Estimation in Electric Power Systems*. Kluwer Academic Publishers, Norwell, MA.

42. A. Abur and A. Gomez-Expósito (2004), *Power System State Estimation*. Marcel Dekker, New York, NY.

43. G. Golub and C. Van Loan (1996), *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore, MD.

44. B. Stott and O. Alsaç (2012) 'Optimal power flow -basic requirements for real-life problems and their solutions', White paper, available from the authors, July 2012.
45. Y. Makarov, P. Du, S. Lu, T. Nguyen, X. Guo, J. Burns, J. Gronquist and M. Pai (2012), 'PMU-based wide-area security assessment: Concept, method, and implementation', *IEEE Transactions on Smart Grid*, vol. **3**, no. 3, pp. 1325–1331, September 2012.
46. H.-D. Chiang (2011), *Direct Methods for Stability Analysis of Electric Power Systems*. Wiley, Hoboken, NJ.
47. P. Mattavelli, G. Verghese and A. Stankovic (1997), 'Phasor dynamics of thyristor-controlled series capacitor systems', *IEEE Transactions on Power Systems*, vol. **12**, no. 3, pp. 1259–1267, August 1997.

21.7   Appendix: basic AC system concepts

AC voltages and currents vary sinusoidally with time, and can be described by:

$$v(t) = V_{peak} \cos(\omega t + \theta_v) \tag{21.8}$$

$$i(t) = I_{peak} \cos(\omega t + \theta_i) \tag{21.9}$$

where V_{peak} is the peak value of the voltage waveform, ω is the angular frequency of the power system ($2\pi60$ rad/s in North America, for example), θ_v is the phase angle of the voltage waveform relative to a system-wide reference waveform $\cos(\omega t)$, and likewise for current. In considering AC systems, it is common to use *root mean square* (RMS) quantities. RMS voltage is defined as:

$$V_{rms} = \sqrt{\frac{1}{T} \int_0^T v^2(t)\,dt}$$

where $T = 2\pi/\omega$ is the period of $v(t)$. This gives $V_{rms} = V_{peak}/\sqrt{2}$ when $v(t)$ has the sinusoidal form of Equation [21.8]. Likewise, for sinusoidal current given by Equation [21.9], the corresponding RMS value is $I_{rms} = I_{peak}/\sqrt{2}$. Because of the ubiquitous use of RMS quantities, it is common for the 'rms' subscript to be dropped.

Time domain voltages and currents, Equations [21.8]–[21.9], can be expressed as complex numbers called *phasors* through the use of Euler's identity, $e^{j\phi} = \cos\phi + j\sin\phi$. It follows that Equation [21.8] can be written:

$$v(t) = \mathrm{Re}\left\{ V_{\mathrm{peak}} e^{j(\omega t + \theta_v)} \right\}$$

$$= \mathrm{Re}\left\{ \sqrt{2} \underbrace{V e^{j\theta_v}}_{\mathbf{V}} e^{j\omega t} \right\}$$

where the complex quantity $\mathbf{V} = V e^{j\theta_v} = V \underline{/\theta_v}$, with $V = V_{\mathrm{peak}}/\sqrt{2}$, is referred to as a voltage phasor. Likewise, the current phasor corresponding to Equation [21.9] is $\mathbf{I} = Ie^{j\theta_i} = I \underline{/\theta_i}$, where $I = I_{\mathrm{peak}}/\sqrt{2}$.

Voltage-current relationships for resistors, inductors and capacitors, in both the time domain and phasor domain, are given in Table 21.1. To derive the phasor relationship for inductors, consider an inductor current $i_L(t)$ of the form Equation [21.9]. Then:

$$v_L(t) = L \frac{di_L(t)}{dt}$$

$$= -L\sqrt{2} I \omega \sin(\omega t + \theta_i)$$

$$= \omega L \times \sqrt{2} I \cos\left(\omega t + \theta_i + \frac{\pi}{2} \right)$$

$$= \mathrm{Re}\left\{ \sqrt{2} \underbrace{\omega L I e^{j\left(\theta_i + \frac{\pi}{2}\right)}}_{\mathbf{V}_L} e^{j\omega t} \right\}$$

where $\mathbf{V}_L$ is the phasor representation of $v_L(t)$. Because $e^{j\pi/2} = j$, this becomes:

$$\mathbf{V}_L = j\omega L I e^{j\theta_i} = j\omega L \mathbf{I}_L.$$

A similar derivation yields the phasor relationship for capacitors that is given in Table 21.1. It should be noted that these derivations assume that all voltages and currents are in sinusoidal steady-state. Phasor representations are therefore not applicable during fast transients, though a concept of dynamic phasors can be used to describe variations that are slow relative to the underlying oscillation frequency.[47]

The instantaneous power associated with a circuit element is given by:

Table 21.1 Voltage–current relationships for circuit elements

	Resistor	Inductor	Capacitor
Time domain	$v = Ri$	$v = L\, di/dt$	$i = C\, dv/dt$
Phasor domain	$\mathbf{V} = R\mathbf{I}$	$\mathbf{V} = j\omega L \mathbf{I}$	$\mathbf{I} = j\omega C \mathbf{V}$

$$p(t) = v(t)i(t)$$

where $v(t)$ is the voltage across the element and $i(t)$ is the current through it. With voltage and current described by Equations [21.8] and [21.9], some manipulation gives:

$$
\begin{aligned}
p(t) &= v(t)i(t) \\
&= P\big(1 + \cos\big(2(\omega t + \theta_v)\big)\big) + Q\sin\big(2(\omega t + \theta_v)\big)
\end{aligned}
\qquad [21.10]
$$

where

$$P = \frac{V_{peak}I_{peak}}{2}\cos(\theta_v - \theta_i) = VI\cos(\theta_v - \theta_i) \qquad [21.11]$$

$$Q = \frac{V_{peak}I_{peak}}{2}\sin(\theta_v - \theta_i) = VI\sin(\theta_v - \theta_i) \qquad [21.12]$$

The first term in Equation [21.10] describes the instantaneous power associated with resistive elements. It is a double-frequency sinusoid with average value P given by Equation [21.11]. This average power is also known as *real power* or *active power*. The second term describes the instantaneous power associated with *reactive* circuit elements, namely inductors and capacitors. It is also a double-frequency sinusoid, but with an average value of zero. The coefficient Q given by Equation [21.12] is referred to as *reactive power*.

Real and reactive power can be expressed as *complex power*, which is given by the produce of the voltage phasor times the conjugate of the current phasor:

$$S = \mathbf{VI}^* = VI\cos(\theta_v - \theta_i) + jVI\sin(\theta_v - \theta_i) = P + jQ$$

The magnitude of the complex power $|S| = VI$ is referred to as *apparent power*, and the multiplier $\cos(\theta_v - \theta_i)$ is called the *power factor*. Apparent power has units of volt-amps (VA), real power has units of watts (W) and the units of reactive power are volt-amp-reactive (VAr).

Most power systems are three-phase, as illustrated in Fig. 21.11. The three voltage sources V_a, V_b, V_c each energize a conductor of the three-phase line that connects the source with the loads. Each load consists of three separate elements, connected in a Y arrangement for Load 1 and a Δ arrangement for Load 2. In a balanced three-phase system, the voltage sources satisfy:

$$\mathbf{V}_{an} = V\underline{/\,\theta_v}, \ \mathbf{V}_{bn} = V\underline{/\,\theta_v - 120°}, \ \mathbf{V}_{cn} = V\underline{/\,\theta_v + 120°}$$

where $\mathbf{V}_{an}$ is the line-to-neutral voltage appearing between the a-phase conductor and the neutral point n, and likewise for $\mathbf{V}_{bn}$ and $\mathbf{V}_{cn}$. In other words, voltage magnitudes are all equal, while phase angles are evenly displaced around the circle. The loads are referred to as balanced if $Z_a = Z_b = Z_c$ and $Z_{ab} = Z_{bc} = Z_{ca}$. Under balanced loading conditions, the currents will also be balanced:

$$\mathbf{I}_a = I\ \underline{/\,\theta_i}, \ \mathbf{I}_b = I\underline{/\,\theta_i - 120°}, \ \mathbf{I}_c = I\ \underline{/\,\theta_i + 120°}.$$

It is easy to show that $\mathbf{I}_a + \mathbf{I}_b + \mathbf{I}_c = 0$ for balanced currents. Hence the current flowing in the neutrals, point n of the source and point n' of the Y-load, will be identically zero. Most power systems are very close to balanced, especially at higher voltage levels.

In a three-phase system, line-to-line voltage refers to the voltage difference between conductors. From Fig. 21.11, the voltage between phases a and b is given by:

$$\mathbf{V}_{ab} = \mathbf{V}_{an} - \mathbf{V}_{bn} = V\underline{/\,\theta_v} - V\underline{/\,\theta_v - 120°} = \sqrt{3}V\underline{/\,\theta_v + 30°}$$

Likewise:

$$\mathbf{V}_{bc} = \sqrt{3}V\underline{/\,\theta_v - 90°}, \ \ \mathbf{V}_{ca} = \sqrt{3}V\underline{/\,\theta_v + 150°}.$$

The line-to-line voltages are therefore also balanced, with magnitude $\sqrt{3}$ times the line-to-neutral voltage. It is common for grid voltages to be
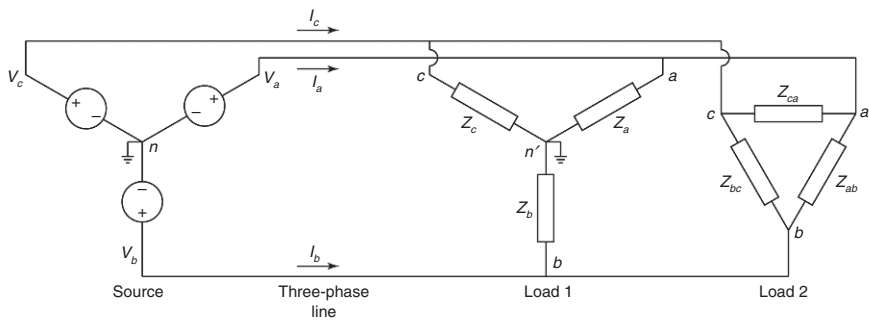
21.11 Three-phase power system.

specified as line-to-line quantities. For example, the 345 kV rating of a transmission line refers to its nominal line-to-line voltage. On the other hand, outlet voltages, for example 120 V in North America and 230 V in Europe, are line-to-neutral voltages.

For balanced three-phase systems, instantaneous power is given by:

$$p(t) = v_a(t)i_a(t) + v_b(t)i_b(t) + v_c(t)i_c(t)$$
$$= 3V_{an}I_a \cos(\theta_v - \theta_i).$$

Even though the power delivered by each phase oscillates at a frequency of 2ω, those time-varying components of the instantaneous power cancel across the three phases, leaving constant power. This is a major advantage of three-phase systems. Complex power for a three-phase system is given by:

$$S_{3\phi} = \mathbf{V}_{an}\mathbf{I}_a^* + \mathbf{V}_{bn}\mathbf{I}_b^* + \mathbf{V}_{cn}\mathbf{I}_c^*$$
$$= 3\mathbf{V}_{an}\mathbf{I}_a^* = 3S_a.$$

In other words, each phase carries exactly one third of the total three-phase complex power.

Index

This page intentionally left blank