Politecnico di Milano
Department of Civil and Environmental Engineering
Doctoral Programme In Structural, Seismic and Geotechnical Engineering

---

Optimal design of sensor networks for structural health monitoring

Doctoral Dissertation of:
Giovanni Capellari

Supervisor:
Prof. Stefano Mariani

Co-supervisor:
Prof. Eleni Chatzi

Tutor:
Prof. Claudia Comi

The Chair of the Doctoral Program:
Prof. Roberto Paolucci

Year 2017 – Cycle XXX

Abstract

The objective of the work presented in this thesis is the development of techniques for the optimal design of sensor networks for Structural Health Monitoring (SHM). Two methods are here proposed, a deterministic and a stochastic one.
In the first one, the uncertainties associated with both the measurements and the mechanical parameters to be estimated (e.g. stiffness, Young's modulus or damage index) are disregarded. The optimal sensor placement is obtained by maximizing the sensitivity of the structural response with respect to a variation of the mechanical properties to be estimated. In order to guarantee a low computational cost, even for high numbers of problem unknowns (number of sensors), a topology optimization scheme is adopted. Moreover, in order to account for the different length-scales of the problem, i.e., the dimensions of the structure, of the damaged zones and of the sensor boards, a multi-scale optimization approach is introduced. The procedure allows to both reduce the computational cost of the optimization problem and appropriately tune the spatial resolution of the solution. The strategy is applied both to a benchmark problem, a clamped square plate, and to a section of stiffened fuselage.
The second method here proposed is based on Bayesian experimental design: the optimal sensor placement is obtained by maximizing the expected Shannon information gain between the prior and the posterior probability distributions of the parameters to be estimated. In order to numerically solve the optimization problem, the unbearable computational cost of the employed Monte Carlo estimator is greatly reduced by exploiting surrogate

modeling techniques based on Polynomial Chaos Expansion (PCE), which allow to efficiently reproduce the input-output relations of the physics-based models. Two surrogate modeling strategies are introduced and compared: these are based either on the definition of a joint input variable, which takes into account both the parameters and the design variable, or on the combination of model order reduction methods, i.e., Principal Component Analysis (PCA), and PCE. In order to handle the noisy objective function, the adoption of a stochastic optimization method, namely the Covariance Matrix Adaptation Evolutionary Strategy, is introduced. Since the presented framework allows to take into account several experimental settings, i.e., sensor spatial configuration, number of sensors and measurement noise, a comprehensive method to optimally design the SHM sensor network is proposed. Moreover, different sensor network designs can be compared, taking into account both their cost and effectiveness, through a cost-benefit optimization approach, by adopting a Pareto frontier multi-objective optimization. The procedure is applied to the benchmark problem already considered for the deterministic approach, and on a large-scale numerical application, i.e. the Pirelli tower in Milan.
Since the capability of any monitoring system in estimating the mechanical parameters can be prevented if the parameters result to be practically non-identifiable, the use of information theory based indices is proposed in order to measure the occurrence of two sources of practical non-identifiability: the compensation of the effects of the parameters on the measurements is quantified through the conditional mutual information; the lack of sensitivity of the measured quantities with respect to each parameter is measured through the mutual information. The effectiveness of these indices is validated on a non-linear structural problem, i.e., an 8-storey shear-type building.

KEYWORDS: structural health monitoring; optimal sensor placement; Bayesian inference; sensor networks; damage detection; uncertainty quantification; identifiability; information theory; topology optimization

Sommario

L'obiettivo del lavoro di tesi qui presentato riguarda lo sviluppo di metodi per la progettazione ottimale di reti di sensori per il monitoraggio strutturale. Vengono qui proposte sia una metodologia deterministica che una stocastica.
Nella prima, si suppone di non considerare nè le incertezze associate alle misurazioni nè quelle ai parametri meccanici da stimare (ad esempio il modulo di Young, la rigidezza o l'indice di danno). Il posizionamento ottimo dei sensori si ottiene massimizzando la sensitività della risposta strutturale rispetto alla variazione delle proprietà meccanica da stimare. Al fine di garantire un basso costo computazionale, anche per un elevato numero di incognite del problema (associato al numero di sensori), viene adottata una procedura di ottimizzazione topologica. Inoltre, al fine ti considerare le differenti scale del problema, ovvero le dimensioni della struttura, della zona danneggiata e dei sensori, viene introdotto un metodo di ottimizzazione multi-scala. La metodologia consente di ridurre il costo computazionale del problema di ottimizzazione e calibrare opportunamente la risoluzione spaziale della soluzione. La strategia viene applicata sia ad un problema di riferimento, una piastra quadrata incastrata agli estremi, sia ad un modello strutturale reale, ovvero una sezione di fusoliera.
Il secondo metodo qui proposto si basa sulla teoria chiamata "Bayesian experimental design": il posizionamento ottimo dei sensori viene ottenuto massimizzando il valore atteso della differenza di informazione di Shannon tra le distribuzioni di probabilità a priori e a posteriori dei parametri da stimare. Al fine di risolvere numericamente il problema di ottimizzazione,

l'elevato costo computazionale dello stimatore della funzione obiettivo, basato sul metodo Monte Carlo, viene ridotto utilizzando modelli surrogati basati su Polynomial Chaos Expansion (PCE), che consentono di riprodurre le relazioni tra inputs e outputs del modello agli elementi finiti. Vengono introdotte e comparate due strategie di modelli surrogati: queste sono basate alternativamente o sulla definizione di una variabile di input che permette di considerare congiuntamente sia i parametri che la variabile di progetto, o sull'applicazione combinata di un metodo di riduzione dell'ordine del modello (analisi delle componenti principali) e PCE. Al fine di calcolare i massimi della funzione obiettivo anche nel caso di un elevato rumore numerico, viene altresì utilizzato un metodo di ottimizzazione per problemi stocastici (Covariance Matrix Adaptation Evolutionary Strategy). Dal momento che il metodo presentato permette di considerare molteplici parametri sperimentali, ovvero la configurazione spaziale, il numero di sensori e l'errore associato alle misurazioni, è quindi possibile progettare in modo onnicomprensivo il sistema di monitoraggio. Inoltre, differenti soluzioni progettuali possono essere comparate, prendendo in considerazione sia il costo che l'efficacia associati alla rete di sensori, attraverso un'analisi costi-benefici, seguendo un approccio di ottimizzazione multi obiettivo. La procedura viene applicata sia al problema di riferimento introdotto per il caso deterministico che ad un modello strutturale di grande scala, ovvero il grattacielo Pirelli a Milano.

Dal momento che l'efficacia di ogni sistema di monitoraggio può essere compromessa se i parametri da stimare risultano essere non identificabili, si propone l'utilizzo di indici basati sulla teoria dell'informazione, al fine di individuare le due cause di non identificabilità: la compensazione degli effetti dei parametri sulle misurazioni viene quantificata attraverso l'informazione mutua condizionale; la mancanza di sensitività delle quantità misurate rispetto ad ogni parametro viene misurato attraverso l'informazione mutua. L'efficacia di questi indici viene verificata considerando un problema strutturale non-lineare, ovvero un edificio a taglio di 8 piani.

Acknowledgements

This few lines are not adequate to express my sincere gratitude to everyone who made this thesis possible.

First of all, I would like to thank my supervisor Professor Stefano Mariani, for his continuous support during my Ph.D. studies, for his valuable suggestions and encouragements, and for letting me pursuing my interests and inclinations. Moreover, I am grateful for supporting me during the hard times and because he helped me to grow both as a person and as a researcher.
I would like to thank my co-supervisor Professor Eleni Chatzi, for her insightful tips and suggestions. I am also grateful for letting me know a different academic environment, which has been truly stimulating.
Besides my supervisors, I would like to thank Francesco Caimmi for his support in the experimental activities and Alessandro Frigerio for his knowledge of sensor electronics.
I would like to thank the reviewers of the journal papers for their valuable comments and all the researchers, PhD students and Professors which I met in conferences and courses, which amazed me and fed my scientific curiosity.
I thank my fellow doctoral students for their friendship and for having shared this long adventure together, both during fun moments and hard times.
A really special thank to Mariagrazia, who has began to be part of my life during this journey and fill up my days with joy: she supported and tolerated me and my mood swings. Thank you for being in my life.
Thank to my family: they encouraged me during this hard path and supported spiritually. Last but not least, I would like to thank all my friends

and flatmates with which I shared funny and relaxing moments together.

Contents

List of Figures

List of Tables

List of Algorithms

CHAPTER 1

Introduction

1.1 Background and motivation

According to the Buildings Performance Institute of Europe (BPIE), more than 40% of the residential buildings in the European Union have been constructed before the 1960s [Buildings Performance Institute Europe, 2011]. Moreover, more than 50% of the European infrastructures, i.e., railways and motorways, have been built in the 1950s and 1960s [European Commission, 2014]. Similarly, over 50% of the bridges in the USA where erected before the 1940s [Stallings et al., 2000].
Since the majority of the structures and infrastructures in the developed countries are reaching the end of their expected service life (defined as the time in service until the structure no longer fulfills the functional requirements), the public and private investments in the construction sector are progressively shifting from new projects towards restoration, maintenance or rehabilitation of existing structures [Costa et al., 2013,Furuta et al., 2014]. In order to support the decision makers and the designers in optimizing the required interventions, it is fundamental to obtain information about the actual state of the mechanical or structural systems. For this reason, Structural Health Monitoring (SHM), which refers to any method for estimating

the structural properties (or their variation in time) of a certain system, is becoming crucial in the civil engineering sector.
The main goals of SHM can be summarized as follows [Balageas et al., 2006]:

- load detection;
- damage detection;
- damage localization;
- damage identification;
- remaining lifetime prognosis.

Damage here reads as any variation in time of the mechanical properties of the system, due to external conditions to which the structure is subjected.
In order to attain the mentioned goals, any SHM strategy can be conceptually divided into the following stages [Farrar and Worden, 2007]:

1. the system state is observed through periodic or continuous measurements, collected for a certain period of time. It is possible to refer to long-time observations when the user is interested in slow variations of the system, such as those induced by corrosion phenomena or short-time observations when impacts or high-frequency events have to be investigated;
2. representatives of features or indices are extracted, in order to appropriately describe the phenomena, which have to be investigated;
3. the measured data is analyzed and processed in order to achieve the aforementioned objectives, either in an on-line or off-line fashion.

In Figure 1.1 a conceptual block diagram representation of the SHM paradigm is shown.
The first SHM systems were developed in the oil industry during the 1980s for the monitoring of off-shore structures subjected to extreme environmental conditions. The first large-scale applications in civil engineering were strategic infrastructures (WASHMS Hong Kong [Wong et al., 2000]) and bridges [Brownjohn and Moyo, 2001, Lynch et al., 2003, U.S. Department of Transportation, 2001] in developed countries, where the additional investments required for establishing the monitoring system were justified by the relevant social and economic losses due to potential structural failures. Thanks to the development of off-the-shelf electronic devices, the

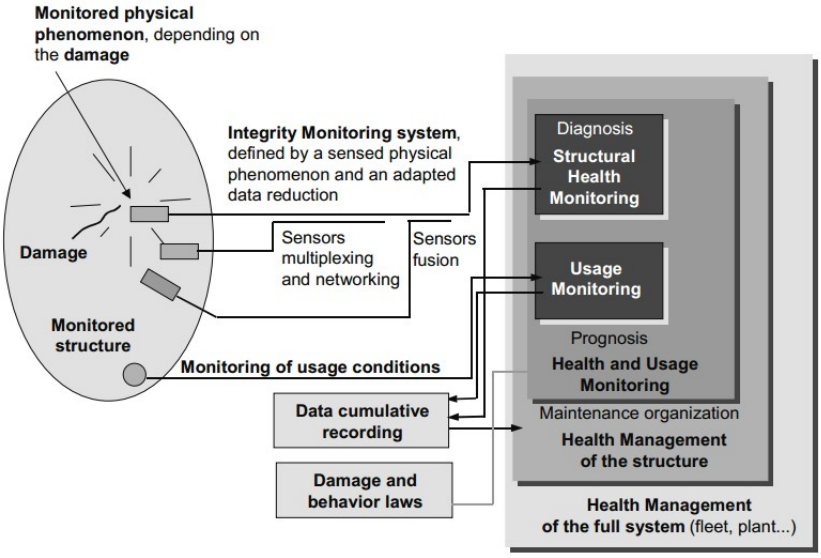


Figure 1.1. *Block diagram representation of a SHM system [Balageas et al., 2006]*

reduction of cost of both the sensors [National Concrete Pavement Technology Center, 2011, Sabato et al., 2017], the data acquisition systems [Basto et al., 2017, Girolami et al., 2017] and the communication protocols [Cho et al., 2008, Noel et al., 2017], and the increase in the computational resources of the processing devices, smart and cheap SHM systems can now be adopted for a wider range of structures and problems, such as aeronautical or low-scale civil applications (see [Bimpas et al., 2011, Moser and Moaveni, 2013, Ubertini et al., 2016, Ferrari et al., 2016, Yin et al., 2016, Liu and Cao, 2017, Lo Iacono et al., 2017]). Furthermore, SHM applications in emerging economies are becoming progressively affordable (see [Yang et al., 2015, Annamdas et al., 2017, Yu et al., 2017]). Due to the aforementioned reasons, the associated SHM market is expected to expand globally at an annual growth rate of 25% in the next 5 years, reaching a global market size of about USD 3.4 billions in 2022 [MarketsandMarkets, 2017].
A new emerging trend in SHM concerns the coupling of SHM technologies with innovative device-to-cloud connectivity platforms, which allow to manage, store and process the measured data remotely and automatically. Some recent IoT-based (Internet of Things) SHM solutions are described in [Abdelgawad and Yelamarthi, 2016, Abdelgawad and Yelamarthi, 2017, Arcadius et al., 2017].
It is important to underline that SHM may be implemented either as a one-

time application, or in a continuous fashion. In the first case, the actual structural conditions of the system are estimated and, therefore, the structural reliability is quantified at a specific time. This allows, e.g., to evaluate the associated risk of system failure [Kurata et al., 2005, Wenzel, 2009] and provides useful data for the insurance and risk management sector [Jia, 2017]. On the other hand, through a continuous monitoring scheme, it is possible to predict the evolution of the system state and, hence, plan and optimize the maintenance interventions (Predictive Maintenance, PM) or performing maintenance when this is needed (Condition-Based Maintenance, CBM) [Hoffman, 2007, Ni and Wong, 2012]. These approaches guarantee an effective decrease in the overall lifetime costs with respect to the traditional time-based PM, as actions are performed only when warranted.

Over the last years, several procedures and methods [Sohn et al., 2004] have been developed in order to detect and estimate structural damage. Some of these [Fan and Qiao, 2011] are natural frequency-based methods [Salawu, 1997], mode shape-based methods [Farrar and Doebling, 1997], curvature/strain mode shape-based methods [Zhang and Aktan, 1998] or other modal parameters based methods [Farrar and Jauregui, 1998]. System identification methods, referred to as procedures to estimate the properties of a structural system defined by means of the parameters of the associated mathematical model, can be classified into two main groups [Moaveni and Conte, 2007]: output-only methods [He et al., 2006, Ubertini et al., 2013, Pioldi et al., 2016, Pioldi and Rizzi, 2016, Pioldi et al., 2017], where only the response of the mechanical system is measured; and input-output methods [Moaveni and Asgarieh, 2012, Asgarieh et al., 2014], where both the inputs and the responses of the system are observed. Damage identification can be then performed by detecting possible changes in the properties of the structural system over time.

In order to estimate both the structural properties and their associated uncertainties, the application of the Bayesian framework to structural health monitoring has been first introduced in [Beck and Katafygiotis, 1998]. Some researchers have focused on the estimation of the parameters of simple mechanical systems or non-linear constitutive laws of materials. The identification of non-linear mechanical models has been performed either through the unscented Kalman filter [Mariani and Ghisi, 2007, Chatzi et al., 2010], the extended Kalman filter [Corigliano and Mariani, 2004, Yang et al., 2006] or the particle filter [Eftekhar Azam et al., 2012]. In order to speed up the estimation, methods based on the combination of model order reduction strategies and Bayesian recursive filtering have been pro-

posed in [Eftekhar Azam, 2014, Capellari et al., 2015a, Capellari et al., 2016d]. Other works regard the estimation of the most likely structural model that guarantees the best match with the experimental data, following the Bayesian frameworks [Beck and Au, 2002, Beck and Yuen, 2004].

1.2 Objectives and scope

The capability of any SHM procedure to detect and estimate structural damages depends on both the employed estimation method used to process the measured data and on the SHM system itself, in terms of all the hardware components needed to obtain the observations, which are:

- a set of sensing nodes, which can include several sensors measuring different physical quantities; these can be powered through a wired network, batteries or energy harvesting modules;
- a data communication system, which could be either wired or wireless-based;
- a post-processing unit, which allows to collect, store and process the measured data.

Any SHM strategy can be, therefore, interpreted as an experimental procedure, where some quantities are measured and some others have to be estimated. The experimental setup in SHM corresponds to the definition of all the characteristics of the sensor network which can affect the measurements, such as:

- the position of the sensors on the structure;
- the physical quantities to be measured;
- the number of sensors;
- the type of sensors.

The objective of the present work is to investigate the problem of optimally designing sensor networks for SHM applications, such that the monitoring system capability for estimating the properties of the structural system is maximized. In other words, the aim of the research pertains to a step which precedes the conceptual phases detailed in the previous paragraph, but, nevertheless, is crucial for the effectiveness of the SHM application. It is interesting to underline that, so far, more attention has been paid in the

research community on the development of mathematical methods for SHM (as summarized in the previous paragraph), rather than on the optimization of the sensor networks. With the increasing use of low-cost off-the-shelf sensors and the continuous spreading of SHM in the civil engineering sector, it is crucial to develop SHM sensor network design strategies in order to reduce the overall cost and the uncertainties associated with the estimated quantities. A review on the existing methods for sensor networks optimization will be given in Chapters 2 and 4.
Assuming that the behaviour of the considered structure can be predicted through a suitable numerical model, all the previously listed experimental setup parameters can be defined in the model and therefore considered as unknown variables in the optimization strategy. For instance, supposing to exploit the finite element method, the position of the sensors can be defined by taking advantage of the nodal discretization of the structural model and, therefore, switching from a continuous to a discretized optimization formulation. Regarding the type of sensors, any technological characteristic, such as sensitivity, signal-to-noise ratio, resolution, etc., can be in principle considered as an optimization variable. Therefore, the optimal type of sensor can be chosen among those available on the market, in order to fulfill the specifications obtained through the optimization procedure.
It should be underlined that the optimization of the sensor network is beneficial from two points of view. Let one first of all assume that, thanks to the optimization of the sensor network characteristics, the number of sensors can be reduced. This results in a corresponding cost reduction of the overall SHM system, a simplification of the data acquisition system (e.g. communication system and conditioning unit) and of the system assembly phase (with a consequent reduction of the technicians cost).
An additional benefit, which can be achieved concerns the reduction of the amount of data that has to be handled: in other words, if the SHM system is optimized, the resulting measurements are more "informative" and, thus, less sensors are necessary to guarantee the same estimation accuracy. Therefore, both the cost and complexity of the data storage system and the required computational resources can be effectively reduced. Moreover, as the computational cost of the post-processing phase decreases, the applicability of real-time estimation methods is enhanced.
If, conversely, the number and type of sensors are supposed to be kept constant, the optimization of the sensor network guarantees the increase of information provided by the monitoring system and, therefore, a consequent reduction of the estimate uncertainties.
Two strategies are here proposed to optimally design a SHM sensor network,

namely a deterministic and a stochastic one.
In the first case, the uncertainties related to the measurement process, i.e., the measurement noise, and the uncertainties associated with the parameters to be estimated, are disregarded and, thus, only the model response is supposed to affect the optimal design. The proposed method relies on the maximization of the measurements sensitivity with respect to the quantities to be estimated (e.g. damage indices, stiffness, Young's modulus). In this case, the employment of a topology optimization strategy guarantees a very low computational cost. The formulation allows to handle complex geometries by adopting a flexible multi-scale approach.
The second method here proposed is based on the application of Bayesian stochastic framework, which allows to naturally account for all the uncertainties in the experimental measurement process. The usefulness of the sensor network is quantified through an index based on information theory, originally developed within the computer science research community for the quantification of uncertainty relating to random variables. The SHM sensor network is therefore optimized, in terms of number, position and type of sensors, by maximizing the relevant expected Shannon information gain, which is a measure of the utility of the measurements with respect to the quantities to be estimated. As in most of the stochastic approaches, large computational resources would be needed to take into account the uncertainties, preventing the applicability of the method to large structural models. The coupling with surrogate models, which aim at replacing the original, computationally expensive, numerical model by reproducing the relation between inputs and outputs, model order reduction strategies and stochastic optimization methods allow to overcome the latter problem.
The two methods for the optimization of sensor networks here developed are applicable to both static and dynamic monitoring applications. Dynamic systems are referred to structural systems whose properties can vary in time and their evolution has to be tracked; nevertheless, these methods are only suitable for applications where the inertial effects on the system responses can be supposed to be negligible.
It is important to underline that the sensor configurations provided by the two methods may be sub-optimal solutions of the optimization problem, as they are obtained using numerical approaches. Moreover, in the stochastic method, the sub-optimality can be due to the objective function noise, as it will be discussed in Chapter 4.
In the end, in order to study the practical identifiability of the model parameters, i.e., to check for uniqueness of solution, the use of information theory-based indices are proposed.

The main contributions of the work presented in the thesis are:

- the development of a deterministic multi-scale optimal sensor placement method for SHM applications, based on the maximization of the measurements sensitivity to damage; the effectiveness of the method is verified via both a benchmark and a real-case structural model;
- the development of a stochastic optimal sensor placement method based on information theory, by exploiting the combination of surrogate modeling and optimization algorithms for stochastic problems; the efficiency of the proposed procedures is checked both through the application to a benchmark and to a large-scale structural model;
- the development of a strategy for the cost-benefit optimization of sensor networks, both in terms of number, placement and type of sensors;
- the development of a strategy for studying practical identifiability through information theory-based indices.

1.3 Organization of the thesis

The thesis is organized as follows.
First, the description of the deterministic optimal sensor placement method based on the maximization of the measurement sensitivity is theoretically described in Chapter 2. A multi-scale approach based on topology optimization is here introduced, in order to reduce the overall computational cost and take into account different scales of the problem.
Then, in Chapter 3, the application of the method to two cases, i.e., a simple benchmark problem (clamped plate) and a real-case structure (aircraft fuselage) are presented and the relevant results are discussed.
A stochastic approach, based on the maximization of the information provided by the sensor network, is introduced in Chapter 4. The concepts of Bayesian experimental design and Bayesian inference are here introduced and their application to the SHM problem is proposed and discussed. Two strategies for the evaluation of the objective function are introduced: the first one is based on a particular implementation of a surrogate modeling strategy, the second one to the combined use of model order reduction techniques and surrogate modeling; the relevant advantages and disadvantages are highlighted. Then, an optimization algorithm for stochastic problems is introduced and its benefits for the problem at hand are examined.

The application of the stochastic method to both a benchmark and a real-size structure is discussed in Chapter 5. Moreover, a scheme for optimizing the overall SHM system, both in terms of number, type and position of sensors, is introduced.
Next, the problem of practical identifiability of the parameters is analyzed and a method based on the information theory is proposed in Chapter 6.
Chapter 7 of the thesis is dedicated to the concluding remarks and suggestions for future work on the topics presented.

CHAPTER 2

Deterministic optimal sensor placement

2.1 Introduction

In this chapter, a method to optimally place sensors on a structure is presented. The appropriate choice of the sensor positions allows to obtain an effective sensor network and to reduce its cost. A basic main hypothesis is put in place: the uncertainties related to the measurement process, such as the measurement noise, are not considered in the optimization algorithm and, in this sense, the method is purely deterministic. Moreover, the uncertainties related with the quantities to be estimated are disregarded, and these are therefore treated as deterministic variables rather than random variables. Unlike stochastic approaches, which allow to take into account the two aforementioned sources of uncertainty, a deterministic strategy is characterized by a lower computational cost and, thus, it can provide a quick indication about the optimal sensor placement configuration.
Some thorough reviews of the most used Optimal Sensor Placement (OSP) methods using deterministic approaches, can be found in [Meo and Zumpano, 2005, Yi and Li, 2012, Mallardo and Aliabadi, 2013, Leyder et al., 2015]. These are:

- the Effective Independence Method (EFI) [Kammer, 1996, Yang and

Lu, 2017], which is based on the maximization of the determinant of the Fisher information matrix (which is a measure of the information content of the measured signal with respect to the parameters to be estimated);

- the Effective Independence Method-Driving-Point Residue (EFI-DPR) [Imamovic, 1998], was proposed to overcome a typical limitation of the EFI method, i.e., the selection of sensor locations of low energy contents; the candidate solutions obtained with the EFI method are then modified through a weighting coefficient, which takes into account the energy content at each position.

- the Kinetic Energy Method (KEM) [Heo et al., 1997, Heo and Jeon, 2016], which is based on the maximization of the kinetic energy of the observations; the method is a modification of the EFI algorithm, where the Fisher information is weighted through the mass matrix of the associated finite element model;

- the Variance Method (VM), which is an evolution of the Most Informative Subset (MIS) technique introduced in [Fedorov and Hackl, 1994] (that is a strategy developed for the approximation of phenomena, through a finite number of observations);

- the Eigenvalue Vector Product (EVP) [Doebling, 1995], which is based on the maximization of the vibration energy of the mode shapes;

- the Non-optimal Driving Point (NODP) [Imamovic, 1998], which consists in deselecting the sensor locations characterized by the smallest target mode shape displacements.

The goal of most of these methods is to obtain the optimal sensor placement of the network configuration on a structure in order to estimate the targeted mode shapes of the considered system, relying on the maximization of different objective functions.
Let one assume now that the SHM system is placed in order to detect damage, which could in principle occur in any position on the structure. The sensors' spatial configuration can be optimized by maximizing the sensitivity of the SHM network to the damage to be detected. As shown in [Mariani et al., 2013a, Mariani et al., 2013b, Caimmi et al., 2014], the sensor placement optimization problem can be formulated as a topology optimization problem; therefore, well-established algorithms suitable for this kind of problems [Bendsøe and Sigmund, 2003] can be adopted for

obtaining the optimal solution. A multi-scale optimization strategy, based on the same rationale, is herein presented. Its purpose is twofold [Capellari et al., 2017a]:

1. the problem of optimal sensor placement can be split into two stages: the first one aims at identifying the regions of higher sensitivity to damage, while the second one allows to optimize the sensor position within the selected zones. In this way, the difference in the length-scales related with the dimensions of the structure, with the damaged zone and with the sensor itself are, thus, taken into account;

2. the overall computational cost of the optimization procedure can be reduced, enabling consideration of the unknown damage position, even for large and complex structures.

The chapter is organized as follows: in Section 2.2, the theoretical formulation is introduced and discussed, then the multi-scale approach is explained in Section 2.3.

2.2 Theoretical formulation

The aim of the optimal sensor deployment strategy herein proposed is to place the sensors on the structure so that the sensitivity of the monitoring system with respect to the variation of mechanical properties due to damage is maximized. Mechanical damage is the set of all micro and macro phenomena which induce changes in the mechanical properties of a certain material. A broad discussion on mathematical models for Damage Mechanics in different materials can be found in [Krajcinovic, 1996] and [Lemaitre, 1996]. The degradation of the material stiffness is here assumed to be modeled as a reduction of the Young's modulus, for isotropic materials (see [Krajcinovic, 1996]).
The sensitivity of the structural response with respect to changes in the material properties can be directly obtained through a numerical model of the structure, by comparing its response to the external loads between the undamaged original structure to be analyzed and its damaged counterpart, featuring a local stiffness reduction. In order to take that into account, in principle, damage can be located anywhere within the system, the structure is conceptually divided into n subdomains, thereby assuming that the mechanical properties vary homogeneously and simultaneously over time in each subdomain. n auxiliary damaged structures are considered, each one characterized by having an undamaged material in all the regions apart

from the j-th region, in which a damage is fictitiously assumed. The choice of the number of regions depends on the resolution required for damage localization in the problem at hand. For instance, if point-wise or very small damages are supposed to occur, then a very fine resolution is needed and, thus, n is assumed to be equal to the number of elements in the numerical discretization (e.g. finite elements). If, otherwise, the stiffness degradation is supposed to affect a large portion of the structure (e.g. material aging), n would correspond to the number of sub-structures in which the system can be divided to account for the different structural components (see [Corigliano et al., 2013, Corigliano et al., 2015]).
As previously mentioned, damage is here assumed to be modeled as a reduction in the Young's modulus: therefore, calling E the mechanical property of the original healthy material, the Young's modulus of the damaged material is defined as $E_j = E(1 - d_j)$, where $j = 1, ..., n$ identifies the region to which this reduction is associated. d_j is a scalar damage index $0 \leq d_j < 1$, which allows to treat the stiffness reduction as a dimensionless quantity. For instance, if $d_j = 0$, the Young's modulus in the j-th region is $E_j = E$ (the material is undamaged). Conversely, assuming $d_j \to 1$, the associated Young's modulus is $E_j \to 0$: this represents the upper-bound of the damage index, as it would correspond to a zero-stiffness material.
As previously stated, the goal of the method is to position the sensors so that the sensitivity of the measurements with respect to the damage index (or to the Young's modulus) is maximized. The objective function of the associated optimization problem would then be a function of the difference between the responses (in terms of, e.g., displacements and rotations, or accelerations in case of dynamics) of the undamaged and all the damaged structures.
Unlike standard optimization methods used for deterministic optimal sensor placement, such as genetic algorithms (employed in [Swann and Chattopadhyay, 2006]) or greedy algorithms (as in [Kammer, 1996]), the problem is here formulated as a topology optimization procedure [Mariani et al., 2013a, Mariani et al., 2013b], in order to boost the computational efficiency. The problem can then be treated by following a standard strategy in topology optimization, i.e., by defining an appropriate field variable $0 \leq \xi(x_1, x_2, x_3) \leq 1$, where x_1, x_2 and x_3 are the relevant reference axes. The higher is ξ, the larger is the sensitivity of the structural response, at the coordinate $\{x_1, x_2, x_3\}$, with respect to damage. Moving from the continuous to the discrete space, the variable $\xi(x_1, x_2, x_3)$ is re-formulated in the corresponding vectorial form $\boldsymbol{\xi} \in \mathbb{R}^n$, with n denoting the number of aforementioned regions. Each component of the unknown variable of the

optimization problem ξ_i (with $i = 1, ..., n$) is then associated with the i-th region. While the previously defined index j is referred to the damage locations (and, therefore, to the auxiliary structures), the index i is related with the sensors locations.
By formulating the topology optimization problem as described in [Bendsøe and Kikuchi, 1988], the resulting objective function to be maximized is defined as follows:

$$\psi = \sum_{j=1}^{n} \sum_{i=1}^{n} \xi_i^p \|\mathbf{v}_{ji} - \mathbf{v}_i\| \tag{2.1}$$

In Eq. (2.1), the symbol $\|\square\|$ is an appropriate norm (e.g. L^2 norm) of the argument $\square$. $\mathbf{v}_i$ is the response (e.g., in terms of displacements, rotations, etc.) associated with the i-th element of the undamaged structure, while $\mathbf{v}_{ij}$ is the response associated with the i-th element of the j-th damaged structure, i.e., with a damage located at the j-th element, ξ_i is the i-th component of the unknown variable $\boldsymbol{\xi}$, p is a penalty coefficient.
The optimal sensor configuration is defined in the following way: if $\xi_i = 0$, no sensors have to be placed in the i-th region, if $\xi_i = 1$ one sensor should be placed at the i-th region. Otherwise, if $0 < \xi_i < 1$ for some regions, it implies that an optimal solution, compliant with the assigned number of sensors, cannot be obtained. In other words, as ξ_i is defined in the domain of the real numbers, it may happen that the number of regions with non-zero values are greater than the chosen number of sensors. Typically, this problem can arise for symmetric geometries, as shown in [Bruggi and Mariani, 2013, Mariani et al., 2014]: in these cases, a sub-optimal solution, in fulfillment with the constraint on the number of sensors and physically relevant to the optimal solution, should be chosen. In order to mitigate the latter problem, a usual strategy in SIMP-based (Solid Isotropic Material with Penalization) [Bendsøe, 1989] approaches of topology optimization is followed: the exponent $p \geq 1$ is introduced in order to penalize intermediate densities (i.e., between 0 and 1) and attain pure $0 - 1$ distributions of the unknown variables.
It should be underlined that, unlike other types of structural optimization, such as shape or size optimization [Christensen and Klarbring, 2008], topology optimization allows to explore all possible spatial configurations, as the candidate solutions are not limited only to a subset of allowable shapes. Therefore, both adjacent and nonadjacent optimal regions can be selected.
According to the formulation in Eq. (2.1), the effects of damages of similar magnitude but placed in different locations could produce consistent changes in the structural response in some areas and small changes in some

others. Therefore the variation of response in the latter case would be hidden by the other one in the sum, driving accordingly the sensor placement. This formulation can lead to ineffective sensor networks if the damages which provoke small sensitivities are located in structurally unsafe locations. In other words, by applying this approach, the sensors tend to be placed where large variations of the structural response are observed, independently on the damage location; since damage identification is the goal of the monitoring system, this would result in an inefficient sensor network. In order to counter-balance these unwanted effects, the formulation can be modified by weighting all the contributions in the sum, through their maximum values, as follows:

$$\psi = \sum_{j=1}^{n} \frac{\sum_{i=1}^{n} \xi_i^p \|\mathbf{v}_{ji} - \mathbf{v}_i\|}{\max_i \left[\xi_i^p \|\mathbf{v}_{ji} - \mathbf{v}_i\|\right]} \tag{2.2}$$

The resulting optimization statement is defined as:

$$\begin{cases} \boldsymbol{\xi}^* = \arg \max\limits_{\xi_i \in [0,1]} \left\{ \sum\limits_{j=1}^{n} \dfrac{\sum_{i=1}^{n} \xi_i^p \|\mathbf{v}_{ji} - \mathbf{v}_i\|}{\max_i \left[\xi_i^p \|\mathbf{v}_{ji} - \mathbf{v}_i\|\right]} \right\} \\ \text{subject to} \quad \sum\limits_{i=1}^{n} \xi_i \leq n_s \end{cases} \tag{2.3}$$

where one constraint is taken into account, namely the total number of sensors n_s to be located.
The formulation here employed for the problem of optimal sensor placement is analogous to the topology optimization problem for volume-constrained minimum compliance (see [Bruns and Tortorelli, 2001, Bruggi and Venini, 2008]), which aims at optimizing the topology of a structure such that its compliance is minimized for a fixed constrained volume. In the optimal sensor placement problem, the field ξ would define the topology of the optimized structure and the constraint on the total number of sensors n_s would correspond to the total volume of material to be employed.
The optimization is performed through the well-known Method of Moving Asymptotes (MMA) introduced in [Svanberg, 1987, Bendsøe and Sigmund, 2004]. The procedure is based on the convex linearization of the objective function ψ with respect to the density field $\boldsymbol{\xi}$. A well-established general approach for the solution of constrained optimization problems, relies on the solution of a sequence of linearized sub-problems, according to the following algorithm steps:

1. choice of the initial value of the first iteration $\boldsymbol{\xi}^0$;

2. computation of $\psi(\boldsymbol{\xi}^k)$ and $\nabla\psi(\boldsymbol{\xi}^k)$, where the nabla symbol ∇ is referred to the gradient operator and the index k is related with the procedure iteration;

3. generation of an optimization sub-problem, by replacing the original function ψ with its approximated linearized version, as computed in step 2;

4. setting of the initial point of the next iteration $k+1$ as the solution of the sub-problem in step 3.

The MMA is a generalization of the latter algorithm, where the linearization is performed with respect to variables of type $\frac{1}{U_i-\xi_i}$ and $\frac{1}{\xi_i-L_i}$, where U_i and L_i are called "moving asymptotes". It is demonstrated in [Svanberg, 1987] that, by allowing U_i and L_i to change during each iteration, depending on $\psi(\boldsymbol{\xi}^k)$ and $\nabla\psi(\boldsymbol{\xi}^k)$, a more efficient and stable method can be obtained. The sub-problem of each iteration is solved through a dual method for convex programming, as described in [Fleury, 1979, Svanberg, 1982]. In order to perform the MMA algorithm, the sensitivity of the objective function ψ to the density unknowns ξ_i are required. In the problem at hand, these are computed from Eq. (2.2) as:

$$\frac{\partial\psi}{\partial\xi_i} = \sum_{j=1}^{n} \frac{\sum_{i=1}^{n} p\xi_i^{p-1}\|\mathbf{v}_{ji}-\mathbf{v}_i\|}{\max_i\left[\xi_i^p\|\mathbf{v}_{ji}-\mathbf{v}_i\|\right]} \tag{2.4}$$

The optimization procedure for solving the optimal sensor placement defined in Eq. (2.3), is summarized in Algorithm 1.

Algorithm 1 Algorithm for the deterministic optimization of SHM sensor networks through topology optimization.

Compute response of the undamaged structure: $\mathbf{v}$
for $j = 1 : n$ **do**
  Compute response of the j-th auxiliary structure: $\mathbf{v}_j$
end
Compute sensitivities according to Eq. (2.4)
Solve Eq. (2.3) through MMA algorithm

2.3 Multi-scale optimization

The MMA algorithm allows to efficiently solve the constrained optimization problem defined in Eq. (2.3), even for high numbers of sensors n_s. As

stated in Section 2.2, in order to allow considering any possible damage position, the undamaged structure and all the n auxiliary structures have to be examined, resulting in $n + 1$ computations of the corresponding FE model responses.
Let one assume that the optimization problem discretization coincides with the FE discretization and, thus, the number of regions n corresponds with the number of finite elements. Therefore, calling t_{FE} the computational time required for the response evaluation of one auxiliary structure, the overall computational time of the pre-processing phase, i.e., related with the calculation of $\mathbf{v}_{ij}$ and $\mathbf{v}_i$, with $i = 1, ..., n$ and $j = 1, ..., n$, is $(n + 1)t_{FE}$. Thus, the applicability of the method to large-scale problems, so of high number of finite elements n, is troublesome since the overall computational cost increases due to the increment of both the number of FE analyses to be solved and the computational cost of each of them.
A possible solution would be to discretize the structure, such that $n < n_{el}$, where n_{el} is the number of FE elements: in other words, each region of the optimal sensor placement problem would include several finite elements, thus obtaining a coarser discretization. In this way, the number of FE model responses would be reduced, while keeping t_{FE} constant. This, of course, would result in a trade-off between the accuracy of the optimal sensor placement solution and the computational cost.
If geometrically complex or large-scale structures have to be analyzed, a multi-scale optimization approach allows to solve the aforementioned problems, without loosing accuracy [Capellari et al., 2016a].
The procedure consists of two phases:

1. macro-scale: the structural model is discretized in n_M parts and the N_M regions with higher sensitivity to damage are selected;

2. meso-scale: each optimal region selected from the macro-scale phase is considered as a new optimization problem, by imposing as boundary conditions the relevant responses obtained at the macro-scale; the N_m optimal elements at the meso-scale are selected among the n_m FE elements of the associated sub-problem.

In this way it is then possible to first define a certain number of regions, which are more sensitive to damage, and subsequently to select the precise position of the sensors on each of them.
The procedure is beneficial not only from the computational point of view, but also for handling complex problems, e.g. structural regions or unimportant details at the macro-scale, but still dangerous for the overall structural

safety. For instance, local stress amplifications, due to geometric, loading or material discontinuities, can cause inception and subsequent growth of damage.
The parameters n and N can be chosen separately for each length-scale: the first one defines the space discretization of the optimization problem; the second one determines the number of regions with higher sensitivity to be selected among the n possible ones. Therefore, the discretization at each scale can be conveniently tuned in order to account for the different length-scales which characterize the problem. For instance, considering complex geometries characterized by several structural components, the system can be discretized at the macro-scale in order to select the components with higher sensitivity. Then, each selected component can be discretized at the micro-scale in order to finely tune the sensor position. This approach is particularly useful when micro-sensors, such as MEMS-based (Micro Electro-Mechanical Systems) devices, are employed: the exact spatial sensor placement can be obtained by setting the mesh size as equal to the sensor packaging dimension only at the meso-scale, while retaining a coarser mesh at the macro-scale. Such very fine discretization would be computationally unbearable, if the single-scale approach described in the previous paragraph was adopted.
The whole procedure is graphically summarized in Fig. 2.1, where $\mathbf{\Omega}$ and $\mathbf{f}$ are respectively the domain where the boundary conditions are defined and the loads applied to the model to be analyzed. It is important to underline that the optimal sensor placement does depend on the loading and boundary conditions; in other words, the solution of the optimization problem is computed for a specific type and magnitude of loads. Therefore, in order to obtain an optimal sensor placement consistent with a multiplicity of different loading scenarios, the optimization scheme here described has to be performed by considering the model responses $\mathbf{v}_{ij}$ for all the possible conditions. Moreover, if the loads are supposed to be not known, the optimal solutions can be obtained by solving several optimization problems with a combination of different input loads, generated by assuming a certain probability density function, following a Monte Carlo approach. The optimal solution would then be the one that guarantees the maximum value of ψ, among all the considered ones. It is important to highlight that while the n_M damaged auxiliary structures and the undamaged structure of the macro-scale problem are subjected to $\mathbf{f}$, the n_m damaged auxiliary structures and the undamaged structure of the meso-scale problem are subjected to both the boundary conditions obtained by the relevant structure of the macro-scale problem and to the external loads acting only on those regions.

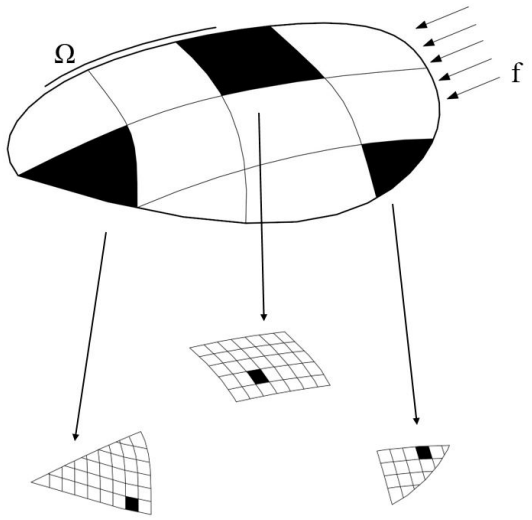


Figure 2.1. *Graphical representation of the proposed multi-scale optimization strategy.*

From Eq. (2.3), the optimization formulation at the macro-scale is defined as:

$$\begin{cases} \boldsymbol{\xi}^M = \arg \max\limits_{\xi_i{}^M \in [0,1]} \psi_M = \arg \max\limits_{\xi_i{}^M \in [0,1]} \left\{ \sum\limits_{j=1}^{n_M} \dfrac{\sum_{i=1}^{n_M} (\xi_i{}^M)^p \| \mathbf{v}_{ji}{}^M - \mathbf{v}_i{}^M \|}{\max_i \left[(\xi_i{}^M)^p \| \mathbf{v}_{ji}{}^M - \mathbf{v}_i{}^M \| \right]} \right\} \\ \text{subject to } \sum\limits_{i=1}^{n_M} \xi_i{}^M \leq N_M \end{cases} \tag{2.5}$$

where ψ_M is the objective function of the associate optimization problem, $\boldsymbol{\xi}^M$ is the unknown function which defines the optimal regions, $\mathbf{v}_{ji}^M$ is the response of the i-th element of the auxiliary structure with damage at the j-th element, and $\mathbf{v}_j^M$ is the response of the i-th element of the undamaged structure. Once the solution at the macro-scale $\boldsymbol{\xi}^M$ is obtained and the N_M optimal regions are selected, the respective problems at the meso-scale are

Algorithm 2 Algorithm for the deterministic multi-scale optimization of SHM sensor networks through topology optimization.

Macro scale optimization
begin
  Compute response of the undamaged structure: $\mathbf{v}^M$
  for $j = 1 : n^M$ **do**
    Compute response of the j-th auxiliary structure: $\mathbf{v}_j^M$
  end
  Compute sensitivities according to Eq. (2.4)
  Solve Eq. (2.5) through MMA algorithm
end
Meso scale optimization
begin
  for $k = 1 : N^M$ **do**
    Select k-th region of macroscale problem
    Enforce boundary conditions from $\mathbf{v}^M$ and $\mathbf{v}_j^M$
    Compute response of the undamaged structure: ${\mathbf{v}_k}^m$
    for $j = 1 : n^m$ **do**
      Compute response of the j-th auxiliary structure: $\mathbf{v}_{jk}^m$
    end
    Compute sensitivities according to Eq. (2.4)
    Solve Eq. (2.6) through MMA algorithm
  end
end

considered. For each region, the following optimization is then performed:

$$\begin{cases} \boldsymbol{\xi}^m = \arg \max\limits_{{\xi_i}^m \in [0,1]} \psi_m = \arg \max\limits_{{\xi_i}^m \in [0,1]} \left\{ \sum\limits_{j=1}^{n_m} \dfrac{\sum_{i=1}^{n_m} ({\xi_i}^m)^p \|{\mathbf{v}_{ji}}^m - {\mathbf{v}_i}^m\|}{\max_i \left[({\xi_i}^m)^p \|{\mathbf{v}_{ji}}^m - {\mathbf{v}_i}^m\| \right]} \right\} \\ \text{subject to} \quad \sum\limits_{i=1}^{n_m} {\xi_i}^m \leq N_m \end{cases} \tag{2.6}$$

where ψ_m is the objective function of the associated optimization problem, $\boldsymbol{\xi}^m$ is the unknown function which defines the optimal sensors positions, $\mathbf{v}_{ji}^m$ is the response of the i-th element of the auxiliary structure with damage at the j-th element, and $\mathbf{v}_i^m$ is the response of the i-th element of the undamaged structure. Analogously to the one-scale problem defined in Eq. (2.3), a constraint on the number of sensors is defined by choosing the parameter N_m, which corresponds to the number of sensors to be deployed for each

macro region.
The whole multi-scale optimization procedure is summarized in Algorithm 2.
In order to investigate the computational benefits of adopting a multi-scale optimization scheme, the computational complexity is here studied in terms of the order of required floating-point operations. The single-scale procedure described in Algorithm 1 is characterized by two main phases: the computation of the model response for each auxiliary structure and the optimization of the SHM sensor network. Assuming a FE solver for linear problems (Gauss-Jordan procedure), its computational complexity is $O[n_{dof}^3]$ [Farmaga et al., 2011, Sharma et al., 2013], where n_{dof} is the number of degrees of freedom of the FE model. As the optimization procedure scales at least quadratically with the number n_{el} of model elements ($O[n_{el}^2]$) [Svanberg, 1987], the overall computational complexity of the single-scale approach is:

$$O[(n+1)n_{dof}^3 + n_{el}^2] \tag{2.7}$$

The same considerations introduced for the single-scale method can be adopted for analyzing the computational complexity of the multi-scale procedure (as described in Algorithm 4). Assuming two scales, the overall computational burden is due to both the macro and the meso scale optimizations. The first phase is characterized by a $O[(n^M+1)(n_{dof}^M)^3 + (n_{el}^M)^2]$ complexity, where n_{el}^M is the number of elements of the FE model at the macro scale and n_{dof}^M the associated number of degrees of freedom. Then, for each selected optimal region, the associated optimization problem at the meso-scale is endowed with a $O[(n^m+1)(n_{dof}^m)^3 + (n_{el}^m)^2)]$ complexity, where n_{el}^m is the number of elements of the FE model at the meso scale and n_{dof}^m the associated number of degrees of freedom. Therefore, the overall computational complexity of the multi-scale procedure is:

$$O\{(n^M+1)(n_{dof}^M)^3 + (n_{el}^M)^2 + N^M[(n^m+1)(n_{dof}^m)^3 + (n_{el}^m)^2]\} \tag{2.8}$$

One can point out that the computational complexity scales linearly with the number of FEM analyses to be performed, quadratically with the number of elements and cubically with the number of DOFs. Thus, one can conclude that the multi-scale approach is computationally more efficient than the single-scale method, as, despite an associated increase in the number of FEM analyses, it allows to reduce both the number of finite elements and the number of DOFs in each model.

2.4 Conclusions

In the present chapter, a multi-scale optimization strategy for the optimal placement of sensors for damage detection (where damage is considered as a local reduction of stiffness) has been introduced.
The method rationale is the maximization of the sensitivity of the response (e.g. in terms displacements or rotations) with respect to the variation of mechanical properties due to damage to be detected. Since the measurement errors and the uncertainty related with the quantities to be estimated are not taken into account, the method is purely deterministic. The employed formulation allows to account for damages located anywhere in the structure.
The multi-scale approach consists in splitting the problem in multiple phases: first, a certain number of optimal regions, whose response is mainly due to the presence of damage of any value and location, are selected, then the location of sensors is finely optimized at each of them. The advantage of this top-down technique is twofold: the computational cost for complex problems is reduced; the discretization size of the associated regions can be appropriately tuned at each scale, in order to preserve accuracy and take into account the relevant length-scales, such as structural details and micro-sensors size.
In Chapter 3 the application of the single-scale and multi-scale approaches to two numerical examples will be discussed.

CHAPTER 3

Deterministic optimal sensor placement: numerical experiments

3.1 Introduction

In this chapter, the optimization methods introduced in Chapter 2 are applied to two numerical examples and the resulting optimal sensor placement solutions are discussed.
First, both the single-scale (see Algorithm 1) and the multi-scale strategies (see Algorithm 2) are adopted on a simple 2-dimensional benchmark problem, i.e., a clamped square plate. Two cases are considered: a uniform material (with only 1 possible type of damage), a composite material (with different damage types).
Then, the multi-scale algorithm is applied to a real structure, a fuselage section model, characterized by a complex geometry and different damage scenarios. The benefits of the proposed algorithm, in terms of required computational time, are highlighted.

3.1.1 Optimal sensor placement on a thin square plate

In order to appropriately assess the deterministic optimal sensor placement method introduced in Chapter 2, the application of the single-scale (Algorithm 1) and multi-scale (Algorithm 2) strategies to a thin square plate clamped along the edges and subject to a force applied at its center, directed orthogonally to its mid-plane is considered. This application case has been already considered in [Capellari et al., 2015b, Capellari et al., 2016d] as a benchmark problem for damage detection. The load is supposed to be static or varying quasi-statically, such that the inertial effects can be disregarded. The in-plane dimensions of the plate are 200 mm $\times$ 200 mm and its thickness is 5 mm. It is assumed to be uniformly made of Aluminium $6061 - T6$ [Association, 2000, ASTM International, 2014]. The Young's modulus is $E = 68.9$ MPa and the density is $\rho = 2.5 \cdot 10^3$ kg/m^3. The boundary conditions, the external load and the reference system are depicted in Fig. 3.1.

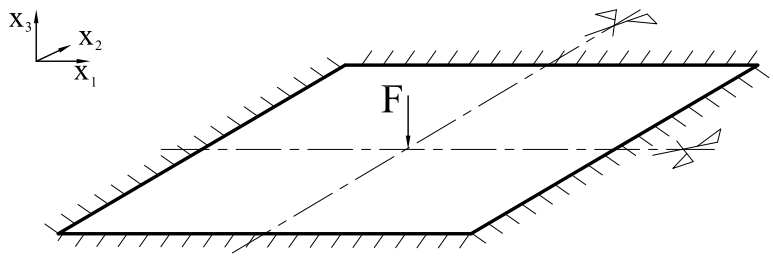


Figure 3.1. *Boundary conditions and loads of the considered square plate.*

In order to demonstrate that the method is completely non-intrusive and it does not require custom manipulations of the solver, a commercial software has been exploited. Thus, the plate is modeled using the commercial FE code Abaqus (Abaqus/CAE 6.10-1 ©Dassault Systemes, 2010), and discretized through a 20×20 mesh with S4R elements, which are 4-node general-purpose shell elements, able to take into account transverse shear deformations. Each node is characterized by 6 DOFs, namely 3 displacements (u_{x_1}, u_{x_2} and u_{x_3}) and 3 rotations (φ_{x_1}, φ_{x_2} and φ_{x_3}) about the 3 orthogonal axes x_1, x_2 and x_3 of the reference system. The chosen discretization corresponds to the resolution of the damage size which is supposed to be investigated, i.e., approximately 10 mm. Furthermore, it guarantees a good accuracy in the computation of the plate response.
The single-scale procedure (Algorithm 1) is first applied: the resulting opti-

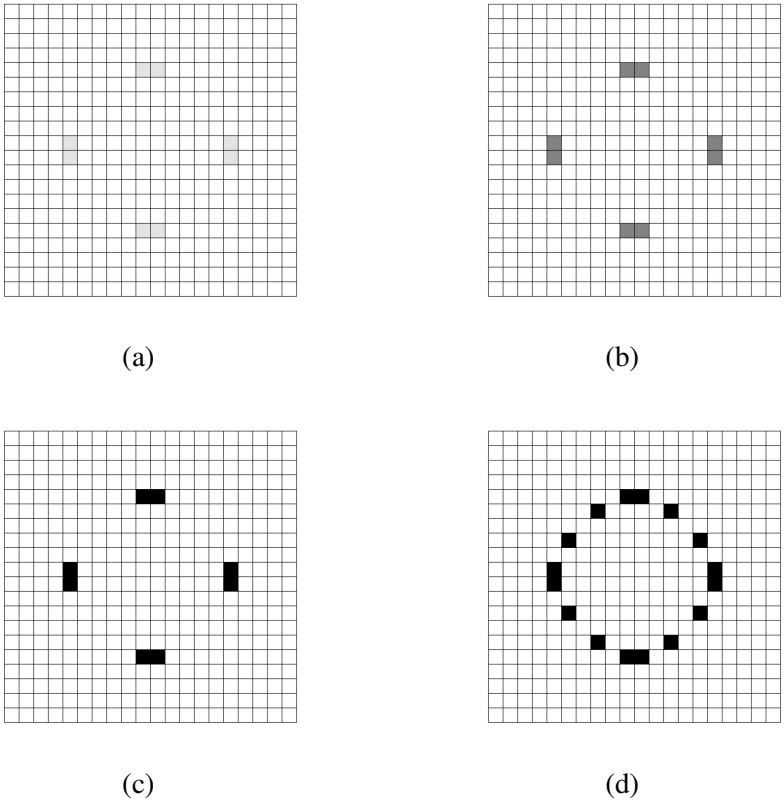


Figure 3.2. *Clamped plate problem for homogeneous material: plane (top) view of the optimal sensor placement corresponding to (a)* $n_s = 1$*, (b)* $n_s = 4$*, (c)* $n_s = 8$ *and (d)* $n_s = 16$*.*

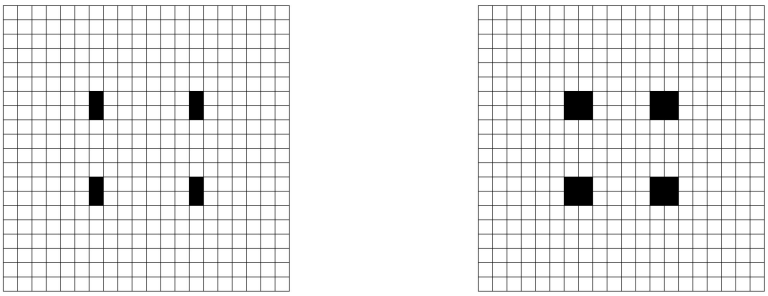
mal configurations for different numbers of sensors $n_s = 1, 4, 8, 16$ are shown in Fig. 3.2. In this example, rotations $\varphi = \sqrt{\varphi_{x_1}^2 + \varphi_{x_2}^2}$ are supposed to be measured by the sensors to be deployed, as their variations can be efficiently used for damage identification (as shown in [Capellari et al., 2015b, Capellari et al., 2016d]). These quantities could be measured through a set of 3-axis MEMS micro-accelerometers [Capellari et al., 2016b]: as discussed in [Łuczak, 2011, STMicroelectronics, 2014], any rotation of the mid-plane of the plate can be observed using the capability of the accelerometers of sensing the gravity acceleration g. Alternatively, low-cost tiltmeters can be employed [Kim, 2008] in order to measure rotations.
Owing to the problem geometry, the resulting optimal regions are located

symmetrically with respect to the two symmetry axes passing through the center of the plate. For $n_s = 1$ (Fig. 3.2a) and $n_s = 2$ (Fig. 3.2b), the density variable ξ_i assumes no-pure $0 - 1$ values: as discussed in Chapter 2, this is due to the geometrical symmetry of the problem, since the highlighted "grey" elements present the same sensitivity to damage.
It should be underlined that the optimal configuration depends on the type of damage which is supposed to be detected. In Fig. 3.2, the material is supposed to be homogeneous and, therefore, the damage is supposed to be accounted for only as a reduction of the Young's modulus. The plate in Fig. 3.1 is supposed to be made of a cross-ply carbon-epoxy composite material, featuring 9 layers [Bresciani, 2013]. The relevant mechanical properties of the transverse isotropic material are listed in Table 3.1, where E_{ii} are the Young's moduli, ν_{ij} is the Poisson's ratio and G_{ij} are the shear moduli. The indexes i and j define the relevant directions of the reference system, as specified in Fig. 3.1. Each ply of the laminate is assumed to be 0.3125 mm thick and the stacking sequence design is $[0/90]_s$.

Property	Value
E_{11} [MPa]	$1.61 \cdot 10^5$
E_{22} [MPa]	$1.14 \cdot 10^4$
ν_{12}	0.32
G_{12} [MPa]	$5.17 \cdot 10^3$
G_{13} [MPa]	$3.98 \cdot 10^3$

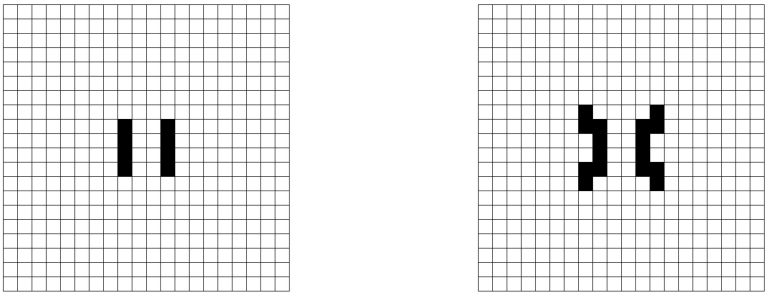
Table 3.1. *Mechanical properties of the composite material [Bresciani, 2013].*

Two cases are considered: a damage throughout all layers (Fig. 3.3), to simulate aging [Monnier et al., 2000, Barbero and Damiani, 2003, Mouzakis et al., 2014]; a damage at the central layer (Fig. 3.4), to simulate delamination [Geubelle and Baylor, 1998, Mariani and Corigliano, 2005]. It is then shown that the resulting optimal configurations are dependent on the type of damage and, if multiple damage types have to be considered, the outcomes of the algorithms should be fused, in order to obtain a sensor deployment which is robust with respect to the damage type. In both damage scenarios in Fig. 3.3 and 3.4, the optimal solutions are not double symmetric as in the homogeneous material case (Fig. 3.2). As concerning the first case, since 5 out of 9 layers are characterized by fibers parallel to the x_2 axis (vertical symmetric axis in Fig. 3.2), the damage-induced stiffness reduction causes larger rotations about the axis x_1, rather than about the axis x_2.



(a) (b)

Figure 3.3. *Clamped plate problem with composite material and damage throughout all layers (see Fig. 3.1): plane (top) view of the optimal sensor placement corresponding to (a) $n_s = 8$, (b) $n_s = 16$.*



(a) (b)

Figure 3.4. *Clamped plate problem with composite material and damage at the central layer (see Fig. 3.1): plane (top) view of the optimal sensor placement corresponding to (a) $n_s = 8$, (b) $n_s = 16$.*

Therefore, the resulting optimal sensor positions tend to be distributed more along the axis x_2 rather than along the axis x_1. A similar result can be observed for the second case (Fig. 3.4), as the fibers of the central layer are parallel to the x_2 axis. It is interesting to underline that, in this case, the sensors tend to be located close to each other, since the measurements correlation is not taken into account by the algorithm. In order to overcome this problem, an additional constraint on the minimum distance between each sensor could be embedded in the algorithm.

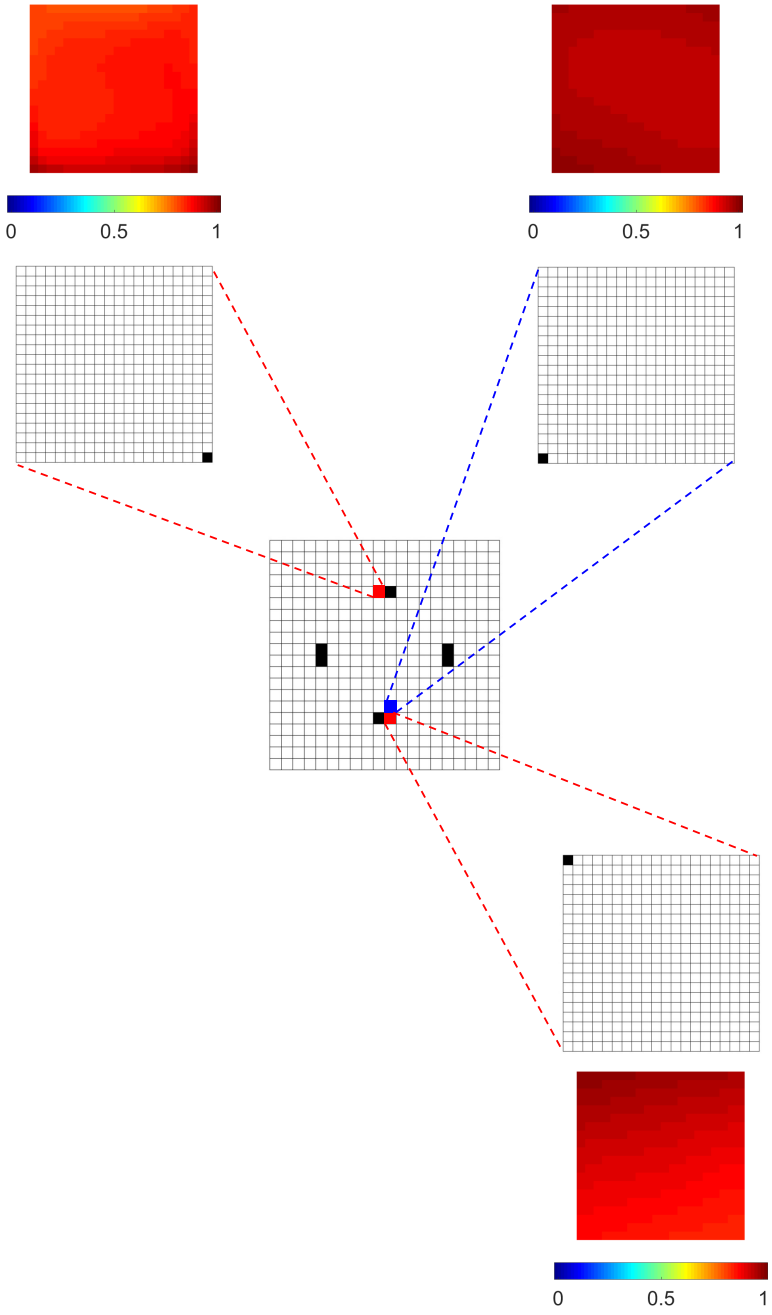


Figure 3.5. *Clamped plate problem: plane (top) view of the optimal sensor placement at the meso-scale (with $N_m = n_s = 1$) and sensitivity maps for selected meso-scale regions.*

It can be underlined that both in the homogeneous material case (Fig. 3.2) and in the composite material case (Figs. 3.3 and 3.4), the computational cost can be reduced by taking advantage of the symmetry about the vertical and horizontal axes.
Let us consider again the homogeneous material case, and assume that the single-scale optimization is considered as the macro-scale phase, i.e., the first step of the multi-scale approach. Once the most sensitive regions have been selected at the macro-scale (Fig. 3.2), the second step of the multi-scale optimization can be performed. The number of regions N_M to be optimally selected should be set equal to the number of sensors n_s, if only one sensor is supposed to be placed at each of them. Each previously defined macro element is discretized through a new 20×20 mesh, in order to obtain a very fine resolution on the position of the sensor.
In Fig. 3.5, the optimal sensor placement with $N_m = n_s = 1$, for three different optimal regions in the macro-scale, is shown. The elements which are zoomed in red correspond to the optimal regions obtained through the macro-scale optimization, while the blue one corresponds to the non-optimal position. The latter one has been chosen in order to check the capability of the optimizer to move the optimal sensor location towards the highest sensitivity regions (see the sensitivity maps of the meso-scale regions in Fig. 3.5). The corresponding colored images represent the values of the objective function ψ, which determines the optimal solutions at the meso-scale. As expected, the resulting solutions reflects the geometrical symmetry.
It is important to highlight that a crucial aspect for obtaining consistent results in a multi-scale approach is the link between each phase of the procedure, which in this case corresponds to the application of the appropriate constraints at the meso-scale regions. As previously specified, the plate is modeled through shell elements, characterized by 6 degrees of freedom per each node, namely 3 displacements and 3 rotations. It can be pointed out that using 4-node (S4R in [SIMULIA, 2013]) shell elements, the reconstruction of the correct rotation field from the macro-scale to the meso-scale problems cannot be guaranteed, leading to a wrong response field. On the other hand, the additional points at the mid-edges of the 8-node (S8R in [SIMULIA, 2013]) elements allows to enforce the necessary constraints to the shell.

3.1.1.1 Computational time

The benefit of adopting the multi-scale approach can be highlighted by comparing the computational time of the single-scale and the multi-scale strategies, for the isotropic homogeneous material case. It is possible to

easily assume that the computational time for finding the optimal configuration (i.e. running the MMA algorithm) is negligible with respect to the one associated with the response computation of the auxiliary structures. Therefore, the two-scale approach requires $20\times20+1 = 401$ analyses at the macro-scale phase and $20\times20+1 = 401$ analyses for each region to be considered at the meso-scale step. This results in a total of 802 analyses, if one sensor has to be placed (see Table 3.2). In this particular application where an isotropic homogeneous material and a symmetric geometry are considered, no additional analyses are required if $n_s = 4$ and $n_s = 8$, as optimal solutions can be easily obtained by appropriately exploiting the symmetry. On the other hand, for $n_s = 16$, at least 1203 analyses are required.
If the single-scale formulation defined in Eq. (2.3) is assumed to have the same resolution employed in the multi-scale approach, namely applying the mesh of the meso-scale level at all the elements of the structure, the total number of auxiliary structures would be $400 \times 400 + 1 = 160,001$ (see Table 3.2). Furthermore, each FE model of the auxiliary structures in the single-scale approach would be characterized by $400\times400 = 160,000$ finite elements in the mesh, resulting in an increased computational time for each analysis from about 0.2 s to about 49 s using an Intel Core i7-4790 CPU @ 3.60 GHz processor, 16 GB RAM and running on Windows 10 64 bit. The overall speedup between the single and the multi-scale approach can be therefore estimated as $\frac{160,001\times49}{401\times2\times0.2} \approx 50,000$ (see Table 3.3).

numerical model	Number of analyses
single-scale model	160,000+1
macroscale model	400+1
mesoscale model	400+1

Table 3.2. *Clamped plate problem: approach-dependent number of analyses.*

Approach	CPU time (s)
single-scale	7,480,049
multiscale	160.4

Table 3.3. *Clamped plate problem: approach-dependent computational times of the overall model analyses.*

3.1.2 Optimal sensor placement on a fuselage

In order to assess the capability of the proposed approach not only for the benchmark problem described in the previous section, but also for a real-size structure, the stiffened aircraft fuselage described in [Zak et al., 2012] is considered. The fuselage consists of a cylinder section, made of a composite curved shell, and a set of equally spaced aluminum rings. By taking advantage of the geometrical symmetries of the problem, only one quarter of the whole section is sufficient to be modeled and symmetry boundary conditions are applied to the vertical edges (namely only the radial displacements and the rotations about directions orthogonal to the symmetry axis are allowed). Moreover, in order to slightly simplify the problem, the window holes, which have been explicitly taken into account in [Zak et al., 2012], have not been considered; nevertheless, their structural effect could be taken into account introducing a pass-through damage. The geometry and the acting load of the problem are shown in Fig. 3.6 and the respective dimensions are detailed in Table 3.4.
While in the plate problem considered in Section 3.1.1 the material was assumed isotropic and homogeneous, and the resulting optimal sensor configuration was not affected by the employed elastic properties in the healthy state, in the case at hand, the solution can depend on the difference between the properties of the materials.

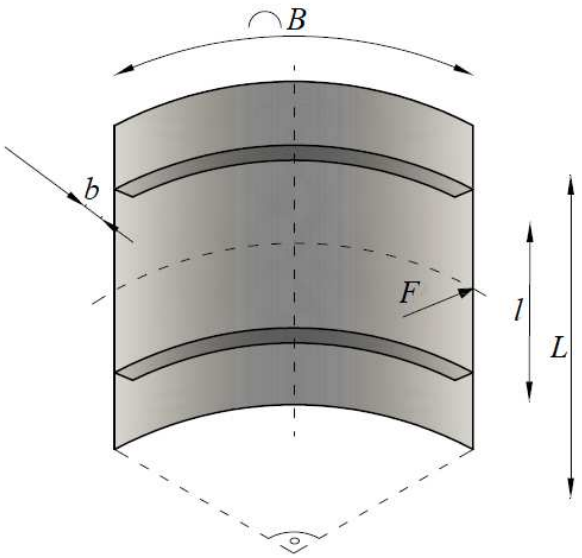


Figure 3.6. *Stiffened fuselage section problem: geometry and notation.*

Parameter	Value
L [mm]	1000
B [mm]	1000
l [mm]	600
b [mm]	30

Table 3.4. *Fuselage dimensions, see Fig. 3.6.*

The cylinder shell is made of a composite material, which is usually used in aeronautical applications (as described in [Guida and Marulo, 2014]), whose mechanical properties are listed in Table 3.5. The indices i and j define the relevant directions of the local reference system, where axis x_1 is aligned with the fiber direction and tangent to the cylinder, x_2 is transverse to it and parallel to the symmetry axis of the shell and x_3 is orthogonal to the shell-mid-plane (radial direction). Each ply of the laminate is assumed to be 0.625 mm thick and the stacking sequence design is $[90/0/\mp 45]_s$. The circular stiffener rings are supposed to be homogeneous and made of Aluminum 2024-O, whose properties are listed in Table 3.6.

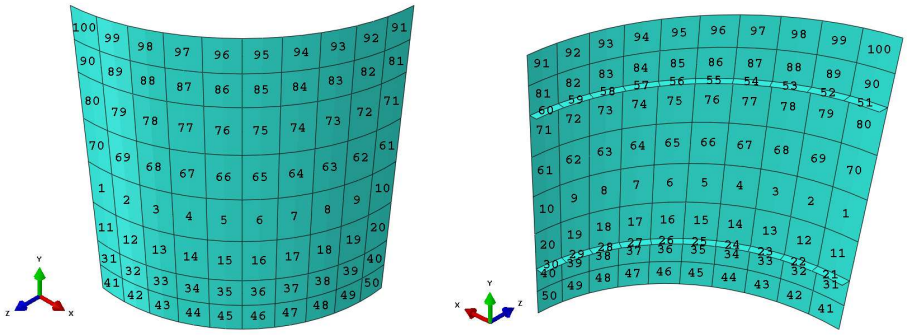
Property	Value
E_{11} [MPa]	$1.51 \cdot 10^5$
E_{22} [MPa]	$8.44 \cdot 10^3$
ν_{12}	0.018
G_{12} [MPa]	$4.20 \cdot 10^3$
G_{23} [MPa]	$2.71 \cdot 10^3$

Table 3.5. *Mechanical properties of the composite material.*

Property	Value
E [GPa]	73.1
ν	0.33

Table 3.6. *Mechanical properties of Aluminum 2024-O.*

The multi-scale approach defined in the previous section is applied to the fuselage structure. The discretization at the macro-scale is depicted in Fig. 3.7, with the relevant element labels (defined as index i in Section 2.3). A coarse 10×8 mesh has been used for the cylinder, while the stiffeners are discretized following the same mesh size, in order to match the nodes



(a) (b)

Figure 3.7. *Stiffened fuselage section problem: (a) external and (b) internal views of the discretization at the macro-scale.*

of different elements for both the types of structural components. The resulting discretization size (approximately 100 mm) sets the dimension of the damaged zones which aims to be investigated through the sensor network. Following the discussion regarding the choice of element types in Section 3, S8R shell elements are employed.
The optimization procedure allows to account for any type of damage of the structure to be analyzed, at any possible position. In this application, only two different scenarios are here considered:

1. a part-through damage on the composite cylinder, in order to account for aging; this damage is simulated by a reduction of the Young's moduli E_{11} and E_{22} of all the plies;

2. a delamination between plies 4 and 5, modeled by decreasing the Young's modulus E_{11} in the mid-plane plies of the composite, in the direction parallel to the composite fibers [Geubelle and Baylor, 1998].

The auxiliary damaged structures are defined featuring the Young's moduli reductions, nominally assumed to be 50% of the Young's moduli of the healthy virgin material. Since the material behavior is supposed to be linear, the choice of the reduced stiffness value for the evaluation of the sensitivity does not affect the resulting optimal sensor solution. As in Section 3.1.1, rotations are supposed to be measured through the sensor network.

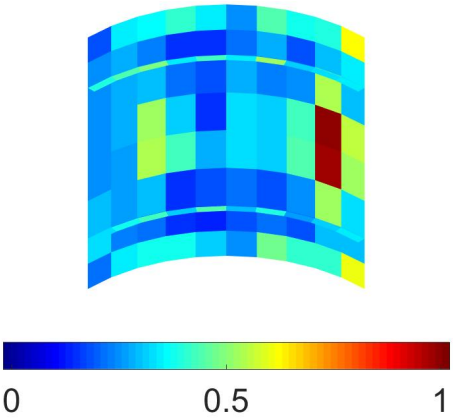


Figure 3.8. *Stiffened fuselage section problem, case 1, macroscale analysis: contour plot of the objective function* ψ_M.

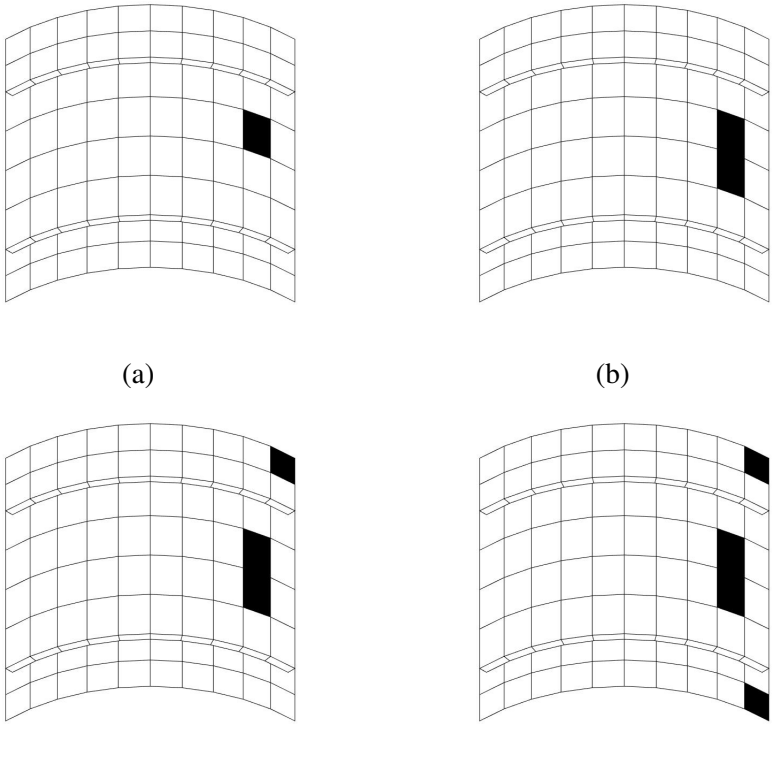


Figure 3.9. *Stiffened fuselage section problem, case 1, macroscale analysis: optimal sensor placement corresponding to: (a)* $N_M = 1$, *(b)* $N_M = 2$, *(c)* $N_M = 3$ *and (d)* $N_M = 4$.

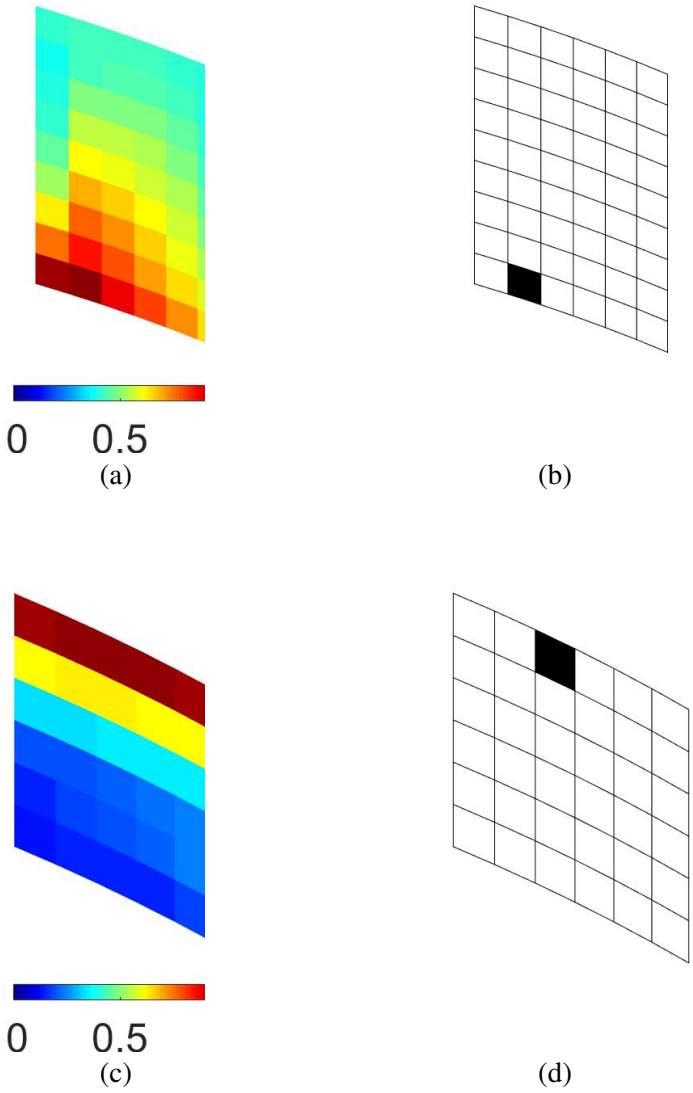


Figure 3.10. *Stiffened fuselage section problem, case 1, mesoscale analysis: (a) / (c), contour plots of the objective function ψ_m and relevant optimal sensor placement with $n_s = N_m = 1$ for (b) $i = 69$ and (d) $i = 100$.*

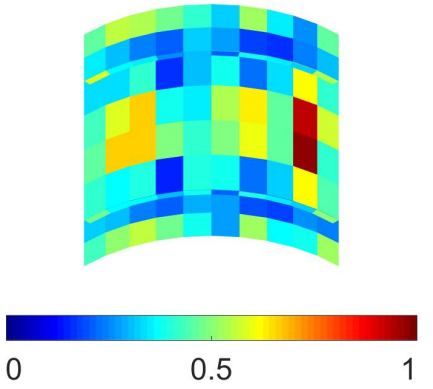


Figure 3.11. *Stiffened fuselage section problem, case 2, macroscale analysis: contour plot of the objective function* ψ_M.

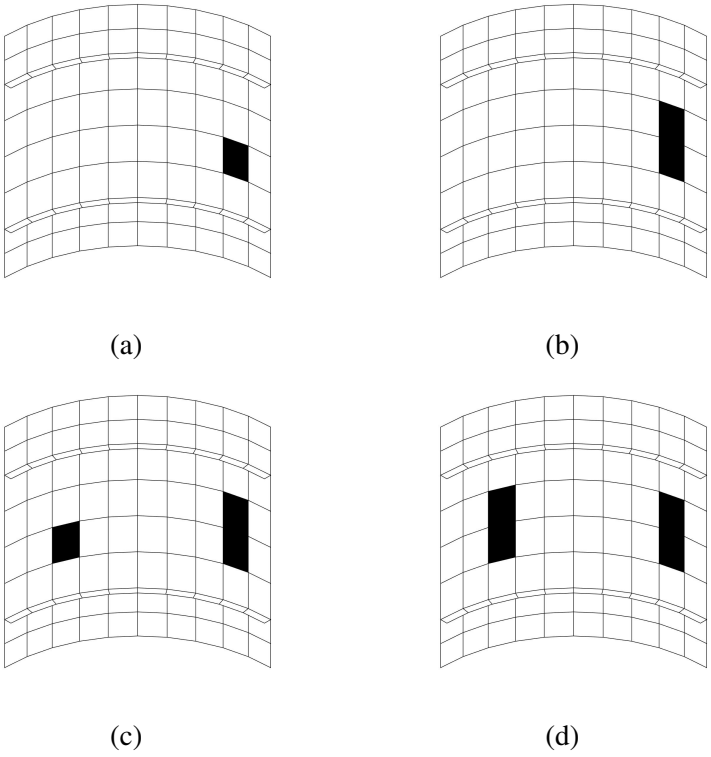


Figure 3.12. *Stiffened fuselage section problem, case 2, macroscale analysis: optimal sensor placement corresponding to: (a)* $N_M = 1$, *(b)* $N_M = 2$, *(c)* $N_M = 3$ *and (d)* $N_M = 4$.

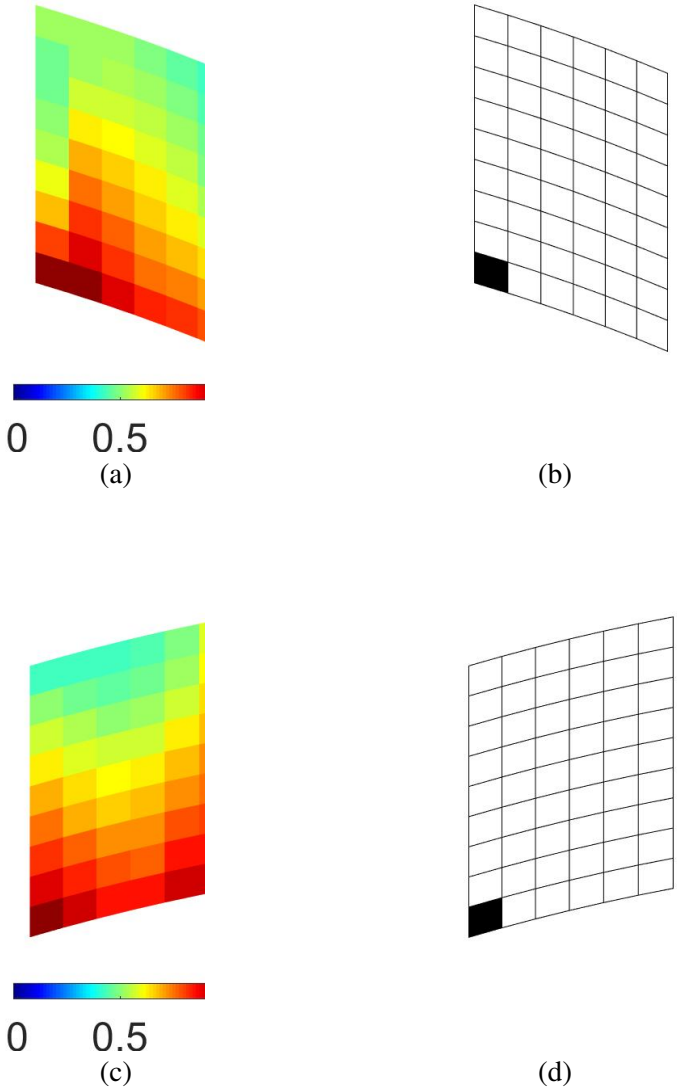


Figure 3.13. *Stiffened fuselage section problem, case 2, mesoscale analysis: (a) / (c) contour plots of the objective function ψ_m, and relevant optimal sensor placement with $n_s = N_m = 1$ for (b) $i = 69$ and (d) $i = 63$.*

Considering case 1, the objective function ψ_M at the macro-scale is shown in Fig. 3.8. As it can be expected, the regions with higher sensitivity lie near the point of application of the external load, where the change

in rotations due to a variation of the Young's modulus is maximum. The corresponding optimal regions selected through the optimization algorithm at the macro-scale are shown in Fig. 3.9, for $N_M = 1$, $N_M = 2$, $N_M = 3$ and $N_M = 4$.
At the meso-scale, the selected optimal regions have been discretized using a finer mesh, as shown in Fig. 3.10. The regions of the composite cylinder are discretized through a 6×9 mesh, while the regions of the stiffeners through a 1×6 mesh; in this way, the resulting discretization at the micro-scale (elements size approximately equal to 10 mm) reflects the micro-sensors dimensions to be deployed.
The objective function ψ_m values, for both the central and the external regions, are respectively shown in Fig. 3.10a and 3.10c. The relevant optimal sensor placement solutions for $n_s = N_m = 1$ are depicted in Fig. 3.10b and 3.10d. It turns out that two regions, i.e. internal and external to the stiffeners, are more sensitive to damage. The internal optimal location results to be on the circumference, where the load is applied (see Fig. 3.10b), while the external one is located near the model edge (see Fig. 3.10d).
Regarding case 2, the objective function ψ_M and the related optimal regions are respectively shown in Fig. 3.11 and 3.12, for $N_M = 1, 2, 3, 4$. The objective function ψ_m values at the meso-scale are shown in Figs. 3.13a and 3.13c, while the relevant optimal sensor placement solutions for $n_s = N_m = 1$ are displayed in Figs. 3.13b and 3.13d.
By analyzing the solutions provided by the optimization procedure for cases 1 and 2, it can be noted that the optimal configurations depend on the type of damage to be detected through the sensor network. Nevertheless, if no priority is given to the type of damage, the optimization could be performed by simply taking into account in the pre-processing phase all the auxiliary structures associated with the relevant damages.

3.1.2.1 Computational time

As in Section 3.1.1, the overall computational cost is mainly attributed to the computation of the responses associated with the auxiliary structures. In the single-scale case, the computational burden is larger for two reasons: first of all, the required number of auxiliary structures is higher than in the meso-scale case; moreover, each FE model features a larger number of degrees of freedom and, hence, a larger dimension of the associated problem to be solved. In Table 3.7, the number of required analyses for each approach, e.g. single-scale and multi-scale, are reported. The CPU computational time required for each simulation is 3.3 s in the single-scale

analysis, 0.2 s for each macro-scale analysis and 0.1 s for each meso-scale analysis. Therefore, the resulting total computational time is calculated in Table 3.8: an overall speed-up of approximately 200 is reached.

numerical model	Number of analyses	Number of model DOFs
single-scale model	3720+1	68,430
macroscale model	100+1	2046
mesoscale model	100+1	1158

Table 3.7. *Stiffened fuselage section problem: approach–dependent number of analyses, and relevant number of degrees of freedom (DOFs) for each analysis.*

Approach	CPU time (s)
single-scale	12276
multiscale	60.6

Table 3.8. *Stiffened fuselage section problem: approach-dependent computational times of the overall model analyses.*

3.2 Conclusions

In the present chapter, the applications of the single-scale and multi-scale optimal sensor placement methods presented in Chapter 2 have been discussed.
Two problems have been considered: a clamped square plate and a fuselage section. The first benchmark problem is characterized by a homogeneous material and symmetric boundary conditions, while the second one is characterized by multiple structural components with different mechanical properties.
It can be highlighted that the resulting optimal configurations depend on the type of damage (e.g. impact, delamination, etc.) which can occur, but its unknown position on the structure is naturally taken into account by the algorithm. In both cases, the benefit resulting by the adoption of the multi-scale procedure are underlined in terms of computational cost, with respect to the single-scale approach. Moreover, as the problem is split in steps, with increasing mesh resolution, it is possible to account for the different scales of the SHM system, i.e. the macroscopic one, associated to the dimensions of the structural system to be monitored, a mesoscopic one, related to the

characteristic size of the damaged region, and a microscopic one, linked to the size of the micro-sensors to be deployed.

CHAPTER 4

Stochastic optimal sensor placement

4.1 Introduction

In Chapter 2, a deterministic approach for optimally positioning sensors on a structure has been introduced. The main assumption at the basis of this method is that no uncertainties are taken into account and a purely deterministic relation between the physical quantities to be measured and those to be estimated (e.g. Young's modulus, damage index, etc.) is exploited. Despite the adoption of these assumptions guarantees a high computational efficiency, it is not possible to consider in the optimization process neither the uncertainties relating to the measurement process (e.g. related with the sensors accuracy), nor those associated with the quantities to be estimated.
Any SHM procedure can be viewed as an experiment, where some physical quantities have to be measured in order to obtain some insights on the mechanical properties of the structure. In practice, any experimental process is always affected by a large variety of physical phenomena that cannot be controlled and predicted, which leads to a discrepancy between the structural response and the observed measurements. Some of them are [Taylor, 1997]:

- measurement noise: it is mainly due to the mode of operation of the

sensors and to the measurement chain, i.e. all the components of the monitoring system which allow to handle and manage the raw data (e.g. data acquisition system, data storage);

- imperfect realization of the measurement setup;
- effects of the environmental conditions;

Moreover, once the measurements are acquired, another source of error is associated with the post-processing phase, when the structural conditions have to be estimated (e.g. through some damage indexes), due to the numerical error of the associated mathematical model. For instance, using a FE model for predicting the structural response, sources of uncertainties can be due to the model discretization (mesh size) or the employment of linearization approaches for non-linear structural problems.
In principle, all these effects could be taken into account from a deterministic point of view, if it would be possible to model and, therefore, predict all the physical phenomena inducing uncertainty. In practice, it is possible to handle these uncertainties by treating the measurements and the parameters to be estimated as random variables. Within this stochastic approach, the Bayesian statistical framework allows to take into consideration these uncertainties and back-propagate them on the quantities to be estimated, i.e., the mechanical properties of the structural system to be investigated, such as Young's modulus or structural stiffness. The variation of these parameters can then be used as a possible indication for damage growth. It is straightforward to assume that the higher the effectiveness of the SHM system is, the lower the uncertainties associated with the estimated parameters are.
It is important to underline that the uncertainties relating to the parameter estimates depend not only on the measurement noise (related with the type of sensors employed), but also on the "amount of information" provided by the sensor network, i.e., the number and the spatial configuration of the sensors. Therefore, these experimental settings can be treated as unknown variables to be optimized, in order to obtain an effective SHM system, which guarantees a low uncertainty related with the quantities to be estimated.
A stochastic approach for the optimal sensor placement was introduced in [Heredia-Zavoni and Esteva, 1998], relying on the minimization of a Bayesian loss function: the instrumentation settings are chosen such that the expectation of the squared error loss function between the estimated and the target values of the quantities to be estimated is minimized. The idea of using the concepts of entropy and mutual information for evaluating the amount of information which can be inferred from an experiment and,

hence, in the design of sensor networks, was first introduced in [Sobczyk, 1987].
In [Papadimitriou et al., 2000,Papadimitriou, 2004], a strategy for designing the SHM system, in terms of optimal spatial configuration, was proposed: the sensors positions are selected such that the information entropy is minimized. The information entropy is defined as follows [Shannon, 1948]:

$$H(D) = -\int p(\boldsymbol{\theta}|D)\ln p(\boldsymbol{\theta}|D)d\boldsymbol{\theta} \tag{4.1}$$

where $\boldsymbol{\theta}$ represents the parameters to be estimated, D the measured data and $p(\boldsymbol{\theta}|D)$ the conditional probability of observing $\boldsymbol{\theta}$, given D. In order to numerically evaluate the objective function defined in Eq. (4.1), the integral terms araising from an analytical manipulation of the information entropy, were approximated through the Laplace method of asymptotic expansions [Bleistein and Handelsman, 1975], which allows to obtain an associated algebraic formulation. In practice, the objective function is approximated through a smooth function centered on a nominal value of the parameters, which has to be a priori chosen.
To what it concerns the optimization methods for inferring the optimal configuration, in most of the studies proposed in literature, the problem has been treated as a discrete optimization problem and, therefore, Genetic Algorithms (GAs) have been exploited in [Yao et al., 1993, Chisari et al., 2017]. According to an alternative greedy approach proposed in [Kammer, 1991, Papadimitriou, 2004], the optimal configuration can be obtained by splitting the optimization problem into a number of sub-problems, where only one sensor is added at each step, so that the increase in the objective function value is maximized.
In the present work, the optimal SHM design is treated as an optimal Bayesian experimental design, where the experimental settings have to be determined so that the information on the quantities to be estimated, as provided by the measurements, is maximum with respect to the parameters to be estimated. From a Bayesian viewpoint, the prior knowledge (defined through an appropriate probability density function) of the parameters is updated through the measurements and the resulting posterior probability density function is obtained. The optimal sensor configuration corresponds, then, to the maximization of the expected Shannon information gain between the prior and the posterior [Lindley, 1956], i.e., the difference between the information entropy associated to the assumed probability distributions of the parameters to be estimated, before performing the measurements (prior), and the updated ones, through the observed data (posterior).

The numerical computation of the aforementioned objective function is performed through a Monte Carlo sampling approach [Ryan, 2003]. Compared to the aforementioned Laplacian approach [Papadimitriou et al., 2000], there is no need of approximating the objective function in correspondence with a certain value of the parameters: therefore, the whole variability of the parameters to be estimated (defined through the associated prior probability distribution) can be exploited. In order to implement a Monte Carlo-based approach, the repeated evaluation of the system response is needed, leading to a high computational cost. Since this approach would be unfeasible, especially for large structures or non-linear problems, surrogate modeling allows to replace the original computationally demanding model with a cheaper one, by mimicking the relations between the mechanical parameters and the model response. Two strategies for building the surrogate models are here presented and compared: one, which is based on the Polynomial Chaos Expansion (PCE) [Ghanem and Spanos, 1990, Blatman and Sudret, 2010, Spiridonakos et al., 2016], is suitable for 1-dimensional problems [Huan and Marzouk, 2013], the second one, which relies on the synergy of surrogate models and model order reduction methods, is suitable for complex high-dimensional systems. The PCE-based surrogate models allow to naturally take into account the uncertainties related with the input variables of the model, by assigning the relevant probability distributions, and to efficiently compute the associated outputs.
The resulting stochastic objective function can be maximized by exploiting some suitable optimization evolutionary strategies [Hansen, 2006].
If not only the sensor configuration, but also other experimental settings, i.e., the number and type of sensors, have to be optimized, the theoretical framework can be extended, according to a cost-benefit approach.
The remainder of the chapter is organized as follows: first, the theoretical framework of optimal Bayesian experimental design, for the OSP problem in SHM, is presented in Section 4.2. Then, the strategies for numerically approximating the objective function are presented in Section 4.3 and the optimization procedure is described in Section 4.4. In the end, a comprehensive method for the optimal design of the SHM system is described in Section 4.6.

4.2 Theoretical framework

4.2.1 Bayesian inference

Two events H and D are considered. The conditional probability of occurrence of event H, given that D takes place, is defined as follows:

$$P(H|D) = \frac{P(H \cap D)}{P(D)} \tag{4.2}$$

where the symbol $P(\square)$ refers to the probability of $\square$ and $\cap$ indicates the intersection of the two events H and D.
Considering an experiment, suppose that H is an experimental hypothesis and D represents the data which can affect the hypothesis. Bayes' rule allows to compute the conditional probability $P(H|D)$ as follows [Jeffreys, 1973]:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \tag{4.3}$$

where $P(H|D)$ is the posterior probability and $P(H)$ is the prior probability of H, which represents the probability of occurrence of H, before D is available, i.e. before performing the measurements. $P(D|H)$ is called likelihood and it represents the plausibility of outcome D, given the hypothesis H; $P(D)$ is named as model evidence or marginal likelihood and it is a neutral factor with respect to the values of H.
Bayes' rule, here defined for the discrete events H and D, can be applied to continuous-valued parameters, i.e., variables characterized by an uncountably infinite number of possible values. The following random vectors are defined: the parameter vector $\boldsymbol{\theta} = [\theta_1\ \theta_2 \cdots\ \theta_{n_{\boldsymbol{\theta}}}] \in \mathbb{R}^{n_{\boldsymbol{\theta}}}$ to be estimated, and the data $\mathbf{y} \in \mathbb{R}^{n_{\mathbf{y}}}$ obtained through a set of sensors, where $n_{\boldsymbol{\theta}}$ is the number of parameters and $n_{\mathbf{y}}$ is the number of measurements. A prior pdf $p(\boldsymbol{\theta})$ is assumed to represent the prior knowledge on $\boldsymbol{\theta}$. The prior pdf can be suitably chosen in order to take into account past information, such as previous experiments or the subjective belief of an experienced expert. If no previous information is available on the values of the parameters, an uninformative distribution can be considered. The pdf associated with the parameters can be updated, taking into account the data $\mathbf{y}$, through the following:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \tag{4.4}$$

The expression $p(\cdot|\cdot)$ represents the conditional pdf of the first term with respect to the second one. Thus, $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior pdf, i.e. the probability density function of $\boldsymbol{\theta}$, given $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood (also called

sampling distribution), and $p(\mathbf{y})$ is the evidence, i.e., the distribution of the observed data marginalized over $\boldsymbol{\theta}$.
The Bayes' theorem is particularly suitable for structural health monitoring problems: given a structure, a class of mathematical models can be defined to describe the structural behaviour of the system . In this view, the aforementioned parameters $\boldsymbol{\theta}$ are the model parameters, i.e., geometrical or mechanical properties (such as Young's modulus, bending stiffness, etc.), and $\mathbf{y}$ are the data which are provided through the monitoring system. Bayesian model updating was first introduced for structural applications in [Beck and Katafygiotis, 1998, Katafygiotis and Beck, 1998] and it allows to update the prior knowledge of $\boldsymbol{\theta}$, which is represented by $p(\boldsymbol{\theta})$, and compute the posterior probability density function $p(\boldsymbol{\theta}|\mathbf{y})$, i.e., the maximum a posteriori estimate $\boldsymbol{\theta}^* = \arg\ \max_{\boldsymbol{\theta}}[p(\boldsymbol{\theta}|\mathbf{y})]$ and the related uncertainty.

4.2.2 Bayesian experimental design

The ability of estimating the uncertainties relating to the parameter vector $\boldsymbol{\theta}$ through Bayesian model updating can be affected by two factors:

- the model, which relates parameters $\boldsymbol{\theta}$ and measurements $\mathbf{y}$;
- the experimental settings, such as the position and type of sensors of the SHM network, or, in other words, the information provided by the measurements $\mathbf{y}$.

The first issue, i.e., the problem of selecting the optimal class of models, has been addressed in [Beck and Yuen, 2004]. In other words, the goal is to choose the type of model which is more compatible with the data, i.e., that provides a better fit, rather than finding the associated model parameters. Since in the present work we are interested in optimally designing the sensor network, we assume that the model class is already specified and, therefore, its optimal selection will not be here treated.
The second problem can be tackled by following the general-purpose decision-theoretic approach introduced in [Lindley, 1956, Raiffa and Schlaifer, 1961]: in this case the decisions to be taken before performing the measurements are basically the choice of the experimental setting, in terms, e.g., of the spatial sensor configuration, the number of sensors, the type of sensors, etc.. To this end, an additional term, the design variable $\mathbf{d} \in \mathbb{R}^{n_d}$, is introduced in the formulation, where n_d is the dimension of the design variable vector.

The Bayes' theorem in Eq. (4.4) is then slightly modified as follows:

$$p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) = \frac{p(\mathbf{y}|\boldsymbol{\theta},\mathbf{d})p(\boldsymbol{\theta}|\mathbf{d})}{p(\mathbf{y}|\mathbf{d})} \tag{4.5}$$

Here, all the previously introduced probability distributions are conditioned with respect to the design variable $\mathbf{d}$, as both the measurements and the parameters to be estimated depend on the experimental settings.
According to [Lindley, 1956], the expected utility of the associated experiment is a function of the design settings $\mathbf{d}$ and it can be quantified through the following expression:

$$U(\mathbf{d}) = \int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} u(\mathbf{d},\mathbf{y},\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \tag{4.6}$$

In Eq. (4.6), $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\Theta}$ respectively represent the domains of the measurements $\mathbf{y}$ and of the parameters $\boldsymbol{\theta}$. Therefore, $p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})$ is the conditional pdf of $\boldsymbol{\theta}$, given $\mathbf{y}$ and $\mathbf{d}$, while $p(\mathbf{y}|\mathbf{d})$ the conditional pdf of $\mathbf{y}$, given $\mathbf{d}$. From a decision-theory perspective, the expected utility allows to choose which action should be performed in order to achieve a certain goal, when the consequences of the chosen act are uncertain. Therefore, it can be described as a weighted average of the utilities of each possible outcomes of a certain action. The weights reflect the probabilities that an action would lead to a certain outcome. In the case at hand, the action is represented by the design of the monitoring system and the goal is the estimation of the parameters. The function $u(\mathbf{d},\mathbf{y},\boldsymbol{\theta})$ is called utility function and it is a scalar measure of the usefulness of the experiment, i.e. it quantifies the extent to which certain measurement values are preferable with respect to the goal of the monitoring system.
The optimal experimental design $\mathbf{d}^*$ defines the experimental settings for which the utility of the experiment is maximized:

$$\mathbf{d}^* = \arg\max_{\mathbf{d}\in\boldsymbol{\mathcal{D}}} \left[\int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} u(\mathbf{d},\mathbf{y},\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \right] = \arg\max_{\mathbf{d}\in\boldsymbol{\mathcal{D}}} [U(\mathbf{d})] \tag{4.7}$$

where $\boldsymbol{\mathcal{D}}$ is the design space, which is the domain of all the possible experimental settings (e.g. the locations where the sensors could be placed).
The rationale of the optimization problem is to find the experimental settings $\mathbf{d}$ for which the usefulness of the experiment is maximum in regards with its purpose. Since the design of the experiment has to be put in place before performing the measurements, i.e., under no knowledge of $\mathbf{y}$ and $\boldsymbol{\theta}$, the objective function of the optimization problem is not directly defined as

$u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$, but its expectation $U(\mathbf{d})$ with respect to $\mathbf{y}$ and $\boldsymbol{\theta}$. In other words, the optimal point $\mathbf{d}^*$ cannot be found by simply maximizing $u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$ for specific values of $\mathbf{y}$ and $\boldsymbol{\theta}$ since these are random variables. Therefore, the optimal point should be looked for in the design space $\mathcal{D}$, by exploring the probability distributions $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$ and $p(\mathbf{y}|\mathbf{d})$ in the domains $\mathcal{Y}$ and $\boldsymbol{\Theta}$.
In the problem here considered, i.e., the Bayesian model updating for SHM, the design variable $\mathbf{d}$ can in principle take into account all the settings of a sensor network, which can affect the accuracy of the parameter estimation, such as the position of the sensors on the structure, the number of measurement channels or the measurement error. In the remainder of the chapter, the number and type of sensors will be assumed to be characteristics of the sensor network and, therefore, they will be treated as constants in the optimization problem. Thus, the design variable $\mathbf{d}$ will be a function of the sensor spatial configuration only and, therefore, only the OSP problem will be considered. Nevertheless, in Section 4.6, the comprehensive optimization problem will be taken into consideration and the design of the sensor network, both in terms of spatial configuration, number and type of sensors, will be discussed.
It can be underlined that the approach based on the Bayesian experimental design can be employed also to optimize the excitation [Green and Worden, 2014] such that the accuracy of the parameter estimation is maximized. This can be obtained by parametrizing the excitation and include its characteristics, such as location and amplitude of the acting loads, in the design variable $\mathbf{d}$. Although the input-output optimization is of interest for obtaining an effective SHM system, the excitation optimization is beyond the scope of this thesis.
The choice of $u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$ depends on the purpose of the experiment, e.g. either estimation of the parameters or prediction of the future states of the mechanical system. In [Chaloner and Verdinelli, 1995] a thorough review of utility functions is discussed. In the case at hand, the aim of the experiment is the inference (estimation) of the parameters $\boldsymbol{\theta}$: according to [De Groot, 1962], a suitable utility function is then the Kullback-Leibler Divergence (KLD) [Kullback and Leibler, 1951, Kullback, 1959], also called relative entropy. In the simplest case where the model is supposed to be linear and the posterior pdfs can be assumed as Gaussian, the optimization problem results in the so-called Bayesian D-optimality [Bernardo, 1979], i.e. the maximization of the determinant of the information matrix of the measurements.
Let us now recall the definition of the KLD from a theoretical point of view. Considering two generic probability distributions P and Q of a random

variable $\mathbf{x}$, the KLD from P to Q is defined as:

$$D_{KL}(P||Q) = \int_{\mathcal{X}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \tag{4.8}$$

where $\mathcal{X}$ is the domain of $\mathbf{x}$, $p(\mathbf{x})$ and $q(\mathbf{x})$ are the related probability densities. The KLD represents the measure of information increase from Q to P. If the two distributions are identical, i.e. $P = Q$, then $D_{KL}(P||Q) = 0$. For the Bayesian inference problem, the goal of the optimization is to design the sensor network such that the highest possible information is provided by the measurements $\mathbf{y}$ with respect to the parameters $\boldsymbol{\theta}$ to be estimated. In other words, the design variable $\mathbf{d}$ has to be chosen such that the information gain between the prior pdf $p(\boldsymbol{\theta}|\mathbf{d})$ and the posterior pdf $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$ is maximum. The resulting utility function is:

$$u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta}) = D_{KL}\left[p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})||p(\boldsymbol{\theta}|\mathbf{d})\right] = \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})}{p(\boldsymbol{\theta}|\mathbf{d})} d\boldsymbol{\theta} \tag{4.9}$$

It is important to underline that in Eq. (4.9) the parameter vector $\boldsymbol{\theta}$ represents a dummy variable in the integral: thus, the utility function is not a function of $\boldsymbol{\theta}$. Since $u(\mathbf{d}, \mathbf{y}) = u(\mathbf{d}, \mathbf{y}, \boldsymbol{\theta})$ and, using Eq. (4.9), the objective function in Eq. (4.6) becomes:

$$\begin{aligned} U(\mathbf{d}) &= \int_{\mathcal{Y}} \int_{\boldsymbol{\Theta}} u(\mathbf{d}, \mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}) d\boldsymbol{\theta} p(\mathbf{y}|\mathbf{d}) d\mathbf{y} = \\ &= \int_{\mathcal{Y}} u(\mathbf{d}, \mathbf{y}) p(\mathbf{y}|\mathbf{d}) d\mathbf{y} = \\ &= \int_{\mathcal{Y}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})}{p(\boldsymbol{\theta}|\mathbf{d})} p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \end{aligned} \tag{4.10}$$

The optimization statement in Eq. (4.7) is then re-written as:

$$\mathbf{d}^* = \arg\max_{\mathbf{d} \in \mathcal{D}} \left[\int_{\mathcal{Y}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})}{p(\boldsymbol{\theta}|\mathbf{d})} p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \right] = \arg\max_{\mathbf{d} \in \mathcal{D}} \left[U(\mathbf{d}) \right] \tag{4.11}$$

With this choice of utility function, the resulting objective function $U(\mathbf{d})$ is called expected gain in Shannon information [Shannon, 1948] or Lindley information measure [Lindley, 1956]. The rationale of the optimization formulation is to design the monitoring system such that the decrease of entropy between the prior and posterior distributions, due to the measured data, is maximized; in other words, the optimal sensor configuration guarantees that the associated observations are more informative, for the parameters

inference, than those obtained from any alternative configuration.
In the optimization approach herein described, the model class is assumed to be known, i.e., it is implicitly chosen a priori. If one would like to guarantee the robustness of the optimal sensor placement with respect to the model class, a new objective function could be defined as the weighted sum of the expected Shannon information gain associated to the different model classes. On the other hand, if the scope of the SHM optimization would be the selection of the optimal model class, instead of the prediction of $\boldsymbol{\theta}$, different metrics, such as the Jensen Shannon Divergence (JSD) should be used, as shown in [Vanlier et al., 2014].
As shown in Appendix A, the expected gain in Shannon information turns out to be equal to the Mutual Information (MI) $I(\boldsymbol{\theta}, \mathbf{y})$ between the parameters $\boldsymbol{\theta}$ and the measurements $\mathbf{y}$ [Cover and Thomas, 2012]. Since the mutual information is a measure of "similarity" or, in other words, the amount of shared information between two random variables, a different interpretation of the optimization problem can be highlighted: the optimal design $\mathbf{d}^*$ guarantees that the reduction of uncertainty related to $\boldsymbol{\theta}$ due to the knowledge of $\mathbf{y}$ is maximized. The concept of mutual information will be further defined and discussed in Chapter 6.

4.3 Numerical approximation of the objective function

In order to solve the optimization problem, a strategy to compute the objective function $U(\mathbf{d})$ defined in Eq. (4.11) is needed. Since, for general cases, the double integration cannot be performed analitically, a numerical procedure has to be followed. Using Eq. (4.5) and applying the quotient rule for logarithms, the objective function can be re-written as (see [Ryan, 2003, Huan and Marzouk, 2013]):

$$\begin{aligned} U(\mathbf{d}) &= \int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})}{p(\boldsymbol{\theta}|\mathbf{d})} p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = \\ &= \int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} \{\ln[p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})] - \ln[p(\mathbf{y}|\mathbf{d})]\} p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y} \end{aligned} \tag{4.12}$$

Without loss of generality, the prior distribution is supposed to be independent of the design variable $\mathbf{d}$, having therefore: $p(\boldsymbol{\theta}|\mathbf{d}) = p(\boldsymbol{\theta})$. According to this assumption, the prior knowledge of the parameters is not affected by the experimental settings.
The double integration in Eq. (4.12) can be approximated using the associ-

ated Monte Carlo (MC) estimator [Huan and Marzouk, 2013]:

$$\hat{U}(\mathbf{d}) = \frac{1}{N_{out}} \sum_{i=1}^{N_{out}} \left\{ \ln \left[p(\mathbf{y}^i|\boldsymbol{\theta}^i, \mathbf{d}) \right] - \ln \left[p(\mathbf{y}^i|\mathbf{d}) \right] \right\} \tag{4.13}$$

where N_{out} is the number of samples $\boldsymbol{\theta}^i$ and $\mathbf{y}^i$ to be drawn respectively from the prior pdf $p(\boldsymbol{\theta})$ and the posterior pdf $p(\mathbf{y}|\boldsymbol{\theta} = \boldsymbol{\theta}^i, \mathbf{d})$. The prior pdf $p(\boldsymbol{\theta})$ is supposed to be chosen by the designer and it reflects the prior knowledge on the parameters distributions, before the measurements are performed.
The term $p(\mathbf{y}^i|\mathbf{d})$ can be computed through an analogous MC estimator as:

$$p(\mathbf{y}^i|\mathbf{d}) \simeq \frac{1}{N_{in}} \sum_{j=1}^{N_{in}} p(\mathbf{y}^i|\boldsymbol{\theta}^j, \mathbf{d}) \tag{4.14}$$

where N_{in} is the number of samples $\boldsymbol{\theta}^j$ to be drawn from the prior pdf $p(\boldsymbol{\theta})$. The computational cost of the MC approach can be reduced by using the same batch of samples $\boldsymbol{\theta}^i = \boldsymbol{\theta}^j$, both for Eqs. (4.13) and (4.14); the resulting required number of likelihood function evaluations decreases from $N_{in} \times N_{out}$ to $N = N_{in} = N_{out}$. On the other hand, the associated estimator bias increases proportionally to $\frac{1}{N_{in}}$ (see [Ryan, 2003]).
The resulting MC estimator $\hat{U}(\mathbf{d})$ is, then:

$$\hat{U}(\mathbf{d}) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \ln \left[p(\mathbf{y}^i|\boldsymbol{\theta}^i, \mathbf{d}) \right] - \ln \left[\frac{1}{N} \sum_{j=1}^{N} p(\mathbf{y}^i|\boldsymbol{\theta}^j, \mathbf{d}) \right] \right\} \tag{4.15}$$

Considering Eq. (4.15), the estimation of $U(\mathbf{d})$ basically depends on the evaluation of the likelihood function $p(\mathbf{y}^i|\boldsymbol{\theta}^j, \mathbf{d})$. A benefit of this approach is that no evaluations of the posterior pdf $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$, which is unknown prior to executing the measurements, are required.
Following Appendix A, a possible alternative to the MC estimator here presented would be to exploit the numerical approaches which are usually used for evaluating the MI, as reviewed in [Walters-Williams and Li, 2009]. The most common MI estimator is the Kraskov estimator, which was introduced in [Kraskov et al., 2004]. It is based on the k-nearest neighbor method, which a classification algorithm for pattern recognition [Fix and Hodges Jr, 1951], and it is defined as follows:

$$\hat{U} = -\psi(k) + \psi(N) + \log(c_{n_{\boldsymbol{\theta}}} + c_{n_{\mathbf{y}}}) + \frac{n_{\boldsymbol{\theta}} + n_{\mathbf{y}}}{N} \sum_{i=1}^{N} \log \epsilon(i) \tag{4.16}$$

In Eq. (4.16) ψ is the Digamma function, N is the number of samples, $n_{\boldsymbol{\theta}}$ and $n_{\mathbf{y}}$ are the dimensions respectively of $\boldsymbol{\theta}$ and $\mathbf{y}$, $c_{n_{\boldsymbol{\theta}}} = \frac{\pi^{n_{\boldsymbol{\theta}}/2}}{\Gamma(n_{\boldsymbol{\theta}}+1)}$ and $c_{n_{\mathbf{y}}} = \frac{\pi^{n_{\mathbf{y}}/2}}{\Gamma(n_{\mathbf{y}}+1)}$ are the volumes of the associated dimensions unit balls (i.e. the regions enclosed in associated multi-dimensional hyperspheres), Γ is the Gamma function, $\epsilon(i)$ is twice the distance between the generic point $(\boldsymbol{\theta}\ \mathbf{y})$ in the joint domain $\boldsymbol{\Theta} \cup \boldsymbol{\mathcal{Y}}$ and its k-th nearest neighbors. The reader may refer to [Altman, 1992]) for more details on the k-th nearest neighbors algorithm.
As it will be highlighted in Section 4.3.1, in the MC estimator the knowledge of the likelihood function is naturally taken into account. On the other hand, the MI estimators depend only on the sampled values of $\boldsymbol{\theta}$ and $\mathbf{y}$, as the MI has been developed in the context of machine learning applications [Vergara and Estévez, 2014], as a purely data-driven method. Therefore, although the MI estimators are computationally faster, they are intrinsically less accurate, therefore resulting in a noisier objective function. The two different numerical approaches will be compared in Section 5.3.

4.3.1 Model response

The evaluation of the likelihood function $p(\mathbf{y}^i|\boldsymbol{\theta}^j, \mathbf{d})$ in Eq. (4.15), for the SHM problem, will be now discussed.
Let us consider a structural system, whose SHM sensor network has to be designed, and assume it is subjected to a set of forces and constraints. Let us denote as $\mathbf{v} \in \mathbb{R}^{n_{dof}}$ the response of the associated structural model, obtained through a numerical method, e.g. the FE method. The number of degrees of freedom $n_{dof} = n_{nodes} n_{dof/node}$ of the FE model depends both on the number of nodes n_{nodes} of the structural model and on the number of degrees of freedom per each node $n_{dof/node}$. Having defined an orthogonal reference system and the associated axes x_1, x_2, x_3, each node can feature 3 displacements u_{x_1}, u_{x_2}, u_{x_3} along the axes and 3 rotations φ_{x_1}, φ_{x_2}, φ_{x_3} about the same axes.
The measurements $\mathbf{y}$ can be related with the structural response through the following model:

$$\mathbf{y} = \mathbf{H}(\mathbf{v}, \boldsymbol{\epsilon}) \tag{4.17}$$

where $\mathbf{H}$ is a functional which relates responses and observations, $\boldsymbol{\epsilon} \in \mathbb{R}^{n_{\mathbf{y}}}$ represents the prediction error, which accounts for both the measurement error, associated with the sensors characteristics, and the model error, associated with the numerical approximation introduced by the FE method. The

associated pdf $p_{\boldsymbol{\epsilon}} = p(\boldsymbol{\epsilon})$ is commonly assumed to be a Gaussian white noise $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\mathbf{0} \in \mathbb{R}^{n_\mathbf{y}}$ is the mean vector and $\boldsymbol{\Sigma} \in \mathbb{R}^{n_\mathbf{y} \times n_\mathbf{y}}$ the covariance matrix. This assumption is required for the applicability of many Bayesian inference methods (e.g. Kalman filter [Kalman, 1960]). Nevertheless, other types of prediction errors, such as coloured noise, should be assumed for specific applications [Kuhlmann, 2003, Chang, 2014].
Assuming a linear operator $\mathbf{H}$, Eq. (4.17) becomes:

$$\mathbf{y} = \mathbf{L}\mathbf{v} + \boldsymbol{\epsilon} \tag{4.18}$$

The matrix $\mathbf{L} \in \mathbb{R}^{n_\mathbf{y} \times n_{dof}}$ is an operator which aims at selecting the measured response components. The generic element of this Boolean matrix is defined as follows:

$$L_{ij} = \begin{cases} 1 & \text{if the } i\text{-th observation corresponds to the } j\text{-th DOF} \\ 0 & \text{if the } i\text{-th observation does not correspond to the } j\text{-th DOF} \end{cases} \tag{4.19}$$

According to the definition in Eq. (4.19), $\mathbf{L}$ depends on the position of the sensors on the structure, i.e., on the model nodes where the sensors are placed, as well as on the orientation and physical quantity to be measured, either displacements or rotations about the directions x_1, x_2 and x_3. In other words, the operator $\mathbf{L} = \mathbf{L}(\mathbf{d})$ is a function of the design variable $\mathbf{d}$.
Since the model response $\mathbf{v}$ depends on the model parameters $\boldsymbol{\theta}$, then the measurement model in Eq. (4.19) can be re-written as follows:

$$\mathbf{y} = \mathbf{L}(\mathbf{d})\mathbf{v}(\boldsymbol{\theta}) + \boldsymbol{\epsilon} = \boldsymbol{\mathcal{M}}(\mathbf{d}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} \tag{4.20}$$

where $\boldsymbol{\mathcal{M}}(\mathbf{d}, \boldsymbol{\theta}) : \mathbb{R}^{n_\mathbf{d}} \times \mathbb{R}^{n_{\boldsymbol{\theta}}} \rightarrow \mathbb{R}^{n_\mathbf{y}}$ is the forward model operator which relates the model inputs, i.e. the design variable $\mathbf{d}$ and $\boldsymbol{\theta}$, with the measurements $\mathbf{y}$.
Following [Yuen, 2010], the likelihood function can then be expressed as:

$$p(\mathbf{y}^i | \boldsymbol{\theta}^j, \mathbf{d}) = p_{\boldsymbol{\epsilon}}\Big(\mathbf{y}^i - \boldsymbol{\mathcal{M}}(\mathbf{d}, \boldsymbol{\theta}^j)\Big) \tag{4.21}$$

By recalling the MC estimator in Eq. (4.15) and the likelihood function in Eq. (4.21), and knowing that $\mathbf{y}^i = \boldsymbol{\mathcal{M}}(\mathbf{d}, \boldsymbol{\theta}^i) + \boldsymbol{\epsilon}^i$ (from Eq. (4.20)), the

resulting estimator of the expected Shannon information gain is:

$$\begin{aligned}\hat{U}(\mathbf{d}) &= \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p_{\boldsymbol{\epsilon}}\left(\mathbf{y}^i - \boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^i)\right)\right] - \ln\left[\frac{1}{N}\sum_{j=1}^{N} p_{\boldsymbol{\epsilon}}\left(\mathbf{y}^i - \boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^j)\right)\right]\right\} = \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}^i)\right] - \ln\left[\frac{1}{N}\sum_{j=1}^{N} p_{\boldsymbol{\epsilon}}\left(\boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^i) + \boldsymbol{\epsilon}^i - \boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^j)\right)\right]\right\} = \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}^i)\right]\right\} - \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[\frac{1}{N}\sum_{j=1}^{N} p_{\boldsymbol{\epsilon}}\left(\boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^i) + \boldsymbol{\epsilon}^i - \boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^j)\right)\right]\right\}\end{aligned} \tag{4.22}$$

According to Eq. (4.22), the MC estimator is obtained through the sum of two terms: the first one depends on the prediction error only, the second one also on the design variable and on the parameters.
If the prediction error $\boldsymbol{\epsilon}$ is independent of the design variable $\mathbf{d}$, i.e. $p(\boldsymbol{\theta}|\mathbf{d}) = p(\boldsymbol{\theta})$, the first term turns out to be independent on $\mathbf{d}$ and, therefore, it can be dropped from the objective function, as we are not interested in the precise value of the expected Shannon information gain, but only in the optimal values $\mathbf{d}^*$. In this case, the resulting estimator would be biased by a constant term, but the computational time required for the evaluation of the objective function would be significantly reduced. This case occurs, for instance, when a unique type of sensor is planned to be installed at all the sensor nodes. The related prediction error has a constant variance across all the measurement channels: the covariance matrix results to be $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, where σ is the standard deviation and $\mathbf{I} \in \mathbb{R}^{n_\mathbf{y}\times n_\mathbf{y}}$ is the identity matrix.
On the other hand, if the standard deviation $\sigma = \sigma(\mathbf{d})$ depends on the design variable, the first term has to be kept in the objective function as it affects the optimal solution. Such case could occur when, for instance, the variance depends on the amplitude of the response: the variance of the c-th response component is $\sigma_c = r\mathbf{G}_c(\boldsymbol{\theta},\mathbf{d})$, where r is a proportionality ratio. Another example of design-dependent variance occurs when the prediction errors at different positions are spatially correlated due the structural response, as it has been considered in [Papadimitriou and Lombaert, 2012] by expressing the covariance matrix as a function of the distance between each pair of sensors.

4.3.2 Surrogate modeling

The computational cost of the MC estimator in Eq. (4.22) is mainly due to the repeated evaluations of the model response $\boldsymbol{\mathcal{M}}(\mathbf{d},\boldsymbol{\theta}^i)$, for each of the N samples $\boldsymbol{\theta}^i = \boldsymbol{\theta}^j$ drawn from the prior pdf $p(\boldsymbol{\theta})$. From a practical point

of view, the computation of $\hat{U}(\mathbf{d})$ can become unfeasible either because of the very high number of degrees of freedom (DOFs) in the FE models, e.g. for very large real structures, or because of the presence of model non-linearities, due to geometrical or mechanical reasons.
In order to reduce the overall computational cost of the evaluation of the model response, the exploitation of surrogate models has been proposed in [Huan and Marzouk, 2013]. A surrogate model (or metamodel) aims at rebuilding the relationship between input and output, as computed through the original model, by means of a more computationally efficient formulation. It is important to underline that these approaches rely on the input-output data only, in a sort of data-driven black-box approach: thus, the associated underlying physics of the problem is therefore lost. This means that, if the physical behavior of the problem changes, a new surrogate model should be built, using the input-output data of the relevant original model.
One of the most widely exploited types of surrogate models is based on the Polynomial Chaos Expansions (PCE). The PCE was first introduced in [Wiener, 1938] for standard Gaussian random variables and then generalized for various other probability distributions in [Ghanem and Spanos, 1990, Xiu and Karniadakis, 2002, Xiu et al., 2002, Xiu and Karniadakis, 2003].
The approach is first of all framed in a general context. The input vectorial random variable $\boldsymbol{\chi} \in \mathbb{R}^M$ is defined, with independent components and a joint probability density function $p_{\boldsymbol{\chi}}$ (with M being the dimension of the input vector). Let us consider the output vectorial random variable $\boldsymbol{\Upsilon} \in \mathbb{R}^{n_\Upsilon}$, which represents the model response (with n_Υ being the dimension of the input vector).
For the sake of simplicity, a scalar numerical model $\Upsilon = \mathcal{M}(\boldsymbol{\chi})$ is now considered, where $\Upsilon \in \mathbb{R}$ is the scalar output random variable. The model is assumed to have a finite variance, such that:

$$\mathbb{E}\left[\Upsilon^2\right] = \int_{\mathcal{D}_{\boldsymbol{\chi}}} \mathcal{M}^2(\boldsymbol{\chi}) p_{\boldsymbol{\chi}}(\boldsymbol{\chi}) d\boldsymbol{\chi} < \infty \tag{4.23}$$

where the symbol $\mathbb{E}[\square]$ stands for the expected value of $\square$.
The PCE of the scalar model $\mathcal{M}$ is:

$$\Upsilon = \mathcal{M}(\boldsymbol{\chi}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \phi_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\chi}) \tag{4.24}$$

where $\Psi_{\boldsymbol{\alpha}}$ are multivariate polynomials which are orthonormal with respect to $p_{\boldsymbol{\chi}}$, i.e.:

$$\mathbb{E}\left[\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\chi}) \Psi_{\boldsymbol{\beta}}(\boldsymbol{\chi})\right] = \delta_{\boldsymbol{\alpha},\boldsymbol{\beta}} \tag{4.25}$$

where $\delta_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 1$ if $\boldsymbol{\alpha} = \boldsymbol{\beta}$ and $\delta_{\boldsymbol{\alpha},\boldsymbol{\beta}} = 0$ otherwise. $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_M\} \in \mathbb{N}^M$ is a multi-index associated with the components of Ψ and $\phi_{\boldsymbol{\alpha}} \in \mathbb{R}$ are the related coefficients. Considering real applications, the sum is truncated by retaining only those polynomials whose total degree $|\boldsymbol{\alpha}|$ is less than a certain value p, as follows:

$$\Upsilon \simeq \mathcal{M}^{PCE}(\boldsymbol{\chi}) = \sum_{|\boldsymbol{\alpha}| \leq p} \phi_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\chi}) \tag{4.26}$$

where $|\boldsymbol{\alpha}| = \sum_{i=1}^{M} \alpha_i$, p is the degree of the retained polynomials, and $\mathcal{M}^{PCE}$ is the surrogate model.

For vectorial models $\boldsymbol{\mathcal{M}}$, the response $\boldsymbol{\Upsilon}$ can be approximated component-wise, by building a set of n_{Υ} PCE surrogate models.

In order to compute the unknown polynomial coefficients $\phi_{\boldsymbol{\alpha}}$, both intrusive and non-intrusive methods can be adopted [Le Maître and Knio, 2010]. Intrusive approaches rely on the projection of the original computational model onto the subspace spanned by the PCE, through the Galerkin projection [Ghanem and Spanos, 2003]. While these methods present some advantages, e.g. no multi-dimensional Gauss quadrature are needed and the increase of the computational cost is linear with the number of basis polynomials, they are not suitable for general cases, since they usually require the custom modification of the computational solver. According to these methods, the variables in the governing equations are replaced with their polynomial chaos expansions. For instance, for linear structural FE problems, the stiffness matrix and the response vector are approximated through a truncated PC expansion, leading to a linear system of equations to be solved [Herzog et al., 2008].

On the other hand, non-intrusive methods allow to compute the bases by simply processing a batch of sampled input variable $\boldsymbol{\chi}$ and the corresponding model evaluations $\boldsymbol{\Upsilon}$, which form the so-called experimental design. No manipulations of the FE solver are needed, making this approach particularly suitable for general purpose problems. Two methods can be used for computing the coefficients in a non-intrusive way: projection approach [Ghiocel and Ghanem, 2002, Le Maitre et al., 2002], where the computation of each coefficient is formulated as a multidimensional integral, and least-square minimization [Berveiller et al., 2006].

The main advantage of the second approach is that an arbitrary number of samples can be used in order to estimate the coefficients, while the first one requires the points to lie according to the chosen quadrature rule. Thus, the second method will be here employed.

The corresponding formulation of the least-square minimization problem

is:

$$\hat{\phi} = \arg\min_{\phi} \mathbb{E}\left\{\left[\phi^T \boldsymbol{\Psi}(\boldsymbol{\chi}) - \mathcal{M}(\boldsymbol{\chi})\right]^2\right\} \tag{4.27}$$

where $\hat{\phi}$ is the set of coefficients to be estimated. In order to reduce the computational cost of the least-square approach, a method based on least angle regression (LAR) and introduced in [Blatman, 2009, Blatman and Sudret, 2011] is here adopted. The method basically relies on the selection of the most significant coefficients of the PC expansion, allowing to reduce the number of model evaluations which are required to build the experimental design for the coefficient estimation.
The types of PCE bases to be employed depend on the type of pdfs of the input variables, as they have to be orthonormal with respect to them. In Appendix B, the types of bases and associated pdfs are listed.
The PCE surrogate modeling strategy is here applied to speed up the computation of the model response, needed in order to estimate the objective function $\hat{U}(\mathbf{d})$. Following the non-intrusive method previously described, a set of $N^{PCE} << N$ samples of the input random variable $\boldsymbol{\chi}$ have to be drawn from the prior pdf $p_{\boldsymbol{\chi}}$, then the corresponding model evaluations Υ are computed through the FE model $\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d})$. Once the metamodel is built, the N samples required for the estimation of $\hat{U}(\mathbf{d})$ can be computed through the computationally efficient surrogate model.
The number N^{PCE} of input-output samples, needed to build the surrogate model, should be chosen by taking into account the accuracy of the metamodel in predicting the original model response. More details about this topic will be given in Chapter 5.
The accuracy of the PCE surrogate model employed for computing the objective function (through the MC estimator) is assessed through the Leave-One-Out (LOO) cross-validation error, as defined in [Blatman, 2009]:

$$\epsilon_{LOO} = \frac{\sum_{i=1}^{N}\left(\mathcal{M}(\mathbf{x}^i) - \mathcal{M}^{PCE}(\mathbf{x}^i)\right)^2}{1 - h_i} \Bigg/ \left(\sum_{i=1}^{N}\left(\mathcal{M}(\mathbf{x}^i) - \hat{\mu}_Y\right)^2\right) \tag{4.28}$$

where $\hat{\mu}_Y = \frac{1}{N}\sum_{i=1}^{N}\mathcal{M}(\mathbf{x}^i)$ is the sample mean of the experimental design response and h_i is the i^{th} component of the vector defined as:

$$\mathbf{h} = \text{diag}\left(\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\right) \tag{4.29}$$

where $\mathbf{A}$ is the experimental matrix that contains the values of all the basis polynomials in the experimental design points, i.e. $A_{ij} = \Psi_j(\chi^i)$, with

$i = 1, \cdots, N^{PCE}$ and $j = 1, \cdots, p-1$.
Two possible different metamodeling settings are here introduced in order to build the surrogate model, as regards to the choice of the input and output variable in the problem at hand:

1. the input variable $\boldsymbol{\chi}$ includes both the parameters $\boldsymbol{\theta}$ and the design variable $\mathbf{d}$, while the output variable $\boldsymbol{\Upsilon}$ corresponds to the components of the model response $\mathbf{v}$ at the observed DOFs;

2. the input variable $\boldsymbol{\chi}$ is represented by the parameters $\boldsymbol{\theta}$, while the output $\boldsymbol{\Upsilon}$ is the reduced-order system response;

The two approaches are respectively explained in Sections 4.3.2.1 and 4.3.2.2.

4.3.2.1 Joint input PCE

A possible approach is to follow the procedure introduced in [Huan and Marzouk, 2013] and consider the input variable $\boldsymbol{\chi}$ as a joint random variable of both the parameters to be estimated and the design variable, as follows:

$$\boldsymbol{\chi} = \begin{Bmatrix} \boldsymbol{\theta} \\ \tilde{\mathbf{d}} \end{Bmatrix} \tag{4.30}$$

where the design variable $\tilde{\mathbf{d}}$ is assumed to identify only one observed DOF, such that:

$$\mathbf{d} = \begin{Bmatrix} \tilde{\mathbf{d}}_1 \\ \vdots \\ \tilde{\mathbf{d}}_{n_\mathrm{y}} \end{Bmatrix} \tag{4.31}$$

Each component of the joint input variable $\boldsymbol{\chi}$ is sampled from the corresponding pdf (which are assumed to be uncorrelated), i.e. the parameters $\boldsymbol{\theta}$ are sampled from the prior pdf $p(\boldsymbol{\theta})$ and the design variable $\tilde{\mathbf{d}}$ from the related pdf $p_{\tilde{\mathbf{d}}}$.
If the sensors are equally likely to be placed all over the structure, then a uniform distribution over the design space $p_{\tilde{\mathbf{d}}} = \mathcal{U}(\mathcal{D})$ has to be chosen. The design space $\mathcal{D}$ should be carefully chosen in order to reflect the regions of the structures where the sensors can be practically placed.
Each value of $\boldsymbol{\chi}$ uniquely identifies one observed DOF (defined by $\tilde{\mathbf{d}}$): the corresponding output variable is, therefore, the component of the original FE model response $\mathbf{v}$ associated to that DOF. With this choice of the input

variable, the resulting output variable is 1-dimensional and, therefore, only one PCE surrogate is needed:

$$v = \mathcal{M}^{PCE}(\boldsymbol{\chi}) \tag{4.32}$$

where $\mathcal{M}^{PCE}$ is the PCE surrogate model and v is the scalar model response, i.e., the component of the model response vector which corresponds with the DOF associated to the values of $\tilde{\mathbf{d}}$. The whole vectorial model response $\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d}) \in \mathbb{R}^{n_\mathbf{y}}$ can be then computed as follows:

$$\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d}) \simeq \begin{Bmatrix} \mathcal{M}^{PCE}(\boldsymbol{\chi}_1) \\ \vdots \\ \mathcal{M}^{PCE}(\boldsymbol{\chi}_{n_\mathbf{y}}) \end{Bmatrix} \tag{4.33}$$

with $\boldsymbol{\chi}_s = \{\boldsymbol{\theta} \;\; \tilde{\mathbf{d}}_s\}^T$ and $\mathbf{d}_s$ being the design variable which corresponds to the s-th measurement.
According to this formulation, the design variable $\tilde{\mathbf{d}}$, which defines the position and orientation of one sensor, is defined as follows:

$$\tilde{\mathbf{d}}_s = \{{x_1}^s \;\; {x_2}^s \;\; {x_3}^s \;\; \delta_s\}^T \tag{4.34}$$

where ${x_1}^s$, ${x_2}^s$, ${x_3}^s$ are the coordinates of the location where the s-th measurement is supposed to be taken; $\delta_s \in [1 \; n_{dof/node}] \subset \mathbb{N}$ is a scalar integer value which defines the measured DOF, with $n_{dof/node}$ being the number of DOFs which can be measured for each sensor node (either displacements or rotations). Therefore, δ_s can assume one of the values $(1, 2, 3, 4, 5, 6)$, which respectively correspond to the DOFs $(u_{x_1}, u_{x_2}, u_{x_3}, \varphi_{x_1}, \varphi_{x_2}, \varphi_{x_3})$.
If only one physical quantity for each sensor node can be measured, then Eq. (4.33) can be simplified as:

$$\tilde{\mathbf{d}}_s = \{{x_1}^s \;\; {x_2}^s \;\; {x_3}^s\}^T \tag{4.35}$$

It should be highlighted that the definition of the design variable $\mathbf{d}$ can affect the effectiveness of the surrogate modeling. A possible alternative formulation to those defined in Eqs. (4.34) and (4.35) relies on the nodal labeling of the FE model DOFs, as follows:

$$\tilde{\mathbf{d}}_s = \xi_s \tag{4.36}$$

where $\xi_s \in [1 \; n_{nodes}] \subset \mathbb{N}$ is the integer index of the associated DOF. Despite the beneficial dimension reduction that can be reached, the adoption of this formulation would be detrimental, as the spatial variation of the system

response would be destroyed by the node labeling. For instance, this would lead to consistent discontinuities of the objective function in the associated $n_{\mathbf{d}}$ multidimensional space, resulting in a very inefficient optima search. For all these reasons, in the remainder, the formulations in Eqs. (4.34) and (4.35), where the design variable is defined through the nodal spatial coordinates, will be chosen.
As it will be demonstrated in Section 5.2, the formulation proposed in this section is suitable for simple applications, e.g. when only the sensor positions have to be optimized and/or only one physical quantity can be measured for each FE model node.
In the most general cases, where the design variables account for both the position and the orientation as in Eq. (4.34), the PCE surrogate would have to represent different physical quantities, i.e., either displacements and rotations in the three directions x_1, x_2 and x_3. The applicability of this method would become very inefficient, as 6 different response fields should be aggregated in one single surrogate, therefore requiring a very high number of samples N^{PCE} in the experimental design in order to build a surrogate model with a high accuracy as compared to the original FE model and, thus, compromising the attempt of recurring to the surrogate model. One possible alternative formulation would be to build 6 different surrogate models, one for each physical quantity.
In addition to the disadvantages here mentioned, this formulation shows to be particularly troublesome when models representing real structures have to be considered, where large number of nodes or complex geometries have to be handled. In these cases, for obtaining an accurate surrogate model, a large number N^{PCE} of samples $\tilde{\mathbf{d}}$ would be required in order the population of design points to be sufficiently representative of the FE model node positions.

4.3.2.2 PCA-PCE

As discussed in Section 4.3.2.1, the surrogate model setting based on the choice of $\boldsymbol{\theta}$ and $\mathbf{d}$ as input variables is particularly suitable for simple applications, but its use becomes troublesome whenever complex structures have to be considered.
While the inclusion of the design variable $\mathbf{d}$ in the input variable allows to reduce the number of PCE surrogate models, it leads to the aforementioned drawbacks. On the other hand, it is possible to underline that in the original model the only role of $\mathbf{d}$ is the selection of the measured DOFs, through the

boolean matrix $\mathbf{L}(\mathbf{d})$. The computational cost associated with the selection is negligible with respect to the one due to the computation of the system response $\mathbf{v}(\boldsymbol{\theta})$ through the FE model. On the basis of this consideration, a possible alternative solution is to use the PCE surrogate model in order to reduce the computational cost of the FE model only. In other words, the resulting input variable is simply:

$$\boldsymbol{\chi} = \boldsymbol{\theta} \tag{4.37}$$

as the design variable $\mathbf{d}$ is not considered in the surrogate model.
An additional advantage of such a choice lies in the associated dimensionality reduction of $\boldsymbol{\chi}$, from $n_{\boldsymbol{\theta}}+4$ (Eq. (4.30)) to $n_{\boldsymbol{\theta}}$ (Eq. (4.37)). This beneficial side effect proves particularly important for PCE-based surrogates, which are affected by the well-known "curse of dimensionality": the number of PCE basis terms, and, therefore, of the associated PCE coefficients to be computed, exponentially increases with the dimension of $\boldsymbol{\chi}$ [Konakli and Sudret, 2016].
By adopting the formulation in Eq. (4.37), the output variable is the model response $\mathbf{v}$, which can be therefore approximated as follows:

$$\mathbf{v} \simeq \boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta}) = \begin{Bmatrix} \mathcal{M}_1^{PCE}(\boldsymbol{\theta}) \\ \vdots \\ \mathcal{M}_{n_{dof}}^{PCE}(\boldsymbol{\theta}) \end{Bmatrix} \tag{4.38}$$

where the multi-dimensional surrogate model $\boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta})$ gathers all the scalar surrogate models. According to this choice of input-output variables, a total number n_{dof} of PCE surrogate models would be required, thus making the computation unbearable. The combination of the surrogate model with dimensionality reduction strategies allows to overcome this problem.
The Principal Component Analysis (PCA) is a statistical tool for handling large datasets, first introduced by Pearson [Pearson, 1901] and Hotelling [Hotelling, 1933] and developed in different formulations [Liang et al., 2002, Wu et al., 2003] among different fields of research: Karhunen-Loeve Decomposition (KLD) [Karhunen, 1947, Loeve, 1941] in signal processing, Proper Orthogonal Decomposition (POD) [Kosambi, 1948] in mathematics, Singular Value Decomposition (SVD) in mechanical engineering [Mees et al., 1987]. Some examples of the application of the POD in structural health monitoring can be found in [Eftekhar Azam, 2014, Capellari et al., 2014, Mirzazadeh, 2017, Mirzazadeh et al., 2017] where it has been employed for the order reduction of dynamical models in Bayesian model

updating.
Considering the problem at hand, let us sample the model parameters (input variable of the surrogate model) from the prior pdf $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$, with $i = 1, ..., N^{PCE}$ and compute the associated response vectors (output variable of the surrogate model) $\mathbf{v}_i = \mathbf{v}(\boldsymbol{\theta}_i)$ through the FE model, building, in this way, the so-called experimental design of the surrogate model. The model response data is now supposed to be gathered in the matrix $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_{N^{PCE}} \end{bmatrix} \in \mathbb{R}^{n_{dof} \times N^{PCE}}$. $\mathbf{V}$ can be projected from the original space onto a new space of n_{dof} variables, which are uncorrelated to each other (i.e. with zero covariance):

$$\mathbf{T} = \mathbf{W}\mathbf{V} \tag{4.39}$$

where $\mathbf{W} \in \mathbb{R}^{n_{dof} \times n_{dof}}$ is a square orthogonal matrix, whose lines are the eigenvectors of the matrix $\mathbf{V}^{\mathrm{T}}\mathbf{V}$ and form an orthogonal basis, $\mathbf{T} \in \mathbb{R}^{n_{dof} \times N^{PCE}}$ is the matrix of the principal component scores, i.e. the representation of $\mathbf{V}$ in the principal component space.
The dimension of the original response matrix $\mathbf{V}$ can be reduced through the PCA by retaining only the first $l << n_{dof}$ components:

$$\mathbf{T}_l = \mathbf{W}_l\mathbf{V} \tag{4.40}$$

where $\mathbf{T}_l \in \mathbb{R}^{l \times N^{PCE}}$ is the reduced order response matrix. Knowing that $\mathbf{W}_l \in \mathbb{R}^{l \times n_{dof}}$ is an orthonormal matrix, the original full order matrix $\mathbf{V}$ can be computed by projecting it back to the full space according to the following equation:

$$\mathbf{V} = {\mathbf{W}_l}^{\mathrm{T}}\mathbf{T}_l \tag{4.41}$$

The discrepancy between the full and reduced order response and, thus, the accuracy of the model order reduction, naturally depends on the choice of l and it can be quantified through the so-called relative information content:

$$I^{PCA}(l) = \frac{\sum_{i=1}^{l} \lambda_i}{\sum_{i=1}^{n_{dof}} \lambda_i} \tag{4.42}$$

where λ_i are the eigenvalues of the matrix $\mathbf{V}^{\mathrm{T}}\mathbf{V}$.
The formulation defined in Eq. (4.38) is then modified by setting the output variable of the PCE surrogate model as the reduced-dimension response vector $\mathbf{t}_l$, as follows:

$$\mathbf{t}_l \cong \boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta}) = \begin{Bmatrix} \mathcal{M}_1^{PCE}(\boldsymbol{\theta}) \\ \vdots \\ \mathcal{M}_l^{PCE}(\boldsymbol{\theta}) \end{Bmatrix} \tag{4.43}$$

The resulting number of required PCE surrogates is reduced to $l << n_{dof}$, establishing a relation between the parameters $\boldsymbol{\theta}$ and the first principal components of $\mathbf{v}$.
In conclusion, by combining the PCE surrogate (Eq. (4.43)) and the PCA dimensionality reduction technique (Eq. (4.41)), the original response model $\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d})$ defined in Eq. (4.20) can be approximated through the following formulation:

$$\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d}) \cong \boldsymbol{\mathcal{M}}^{META}(\boldsymbol{\theta}, \mathbf{d}) = \mathbf{L}(\mathbf{d})\mathbf{W}_l{}^T \boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta}) \tag{4.44}$$

where $\boldsymbol{\mathcal{M}}^{META}(\boldsymbol{\theta}, \mathbf{d})$ refers to the metamodel which combines the PCA and the PCE surrogates.
From Eqs. (4.44) and (4.22), the resulting MC estimator is then:

$$\begin{aligned}\hat{U}(\mathbf{d}) = \frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}^i)\right]\right\} - \\ -\frac{1}{N}\sum_{i=1}^{N}\left\{\ln\left[\frac{1}{N}\sum_{j=1}^{N} p_{\boldsymbol{\epsilon}}\Big(\mathbf{L}(\mathbf{d})\mathbf{W}_l{}^T\boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta}^i) + \boldsymbol{\epsilon}^i - \right.\right. \\ \left.\left. -(\mathbf{L}(\mathbf{d})\mathbf{W}_l{}^T\boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta}^j)\Big)\right]\right\}\end{aligned} \tag{4.45}$$

4.4 Optimization procedure

In the previous sections, a numerical procedure for the estimation of the expected information gain $U(\mathbf{d})$, through the MC estimator $\hat{U}(\mathbf{d})$ of Eq. (4.15) and the use of PCE surrogate modeling, has been described. The procedure requires to draw N^{PCE} samples from the prior pdf $p(\boldsymbol{\theta})$, then to build the PCE surrogate according to the formulations described in Sections 4.3.2.1 and 4.3.2.2, and use it for computing the $N >> N^{PCE}$ model responses to be used in the MC estimator.
In [Papadimitriou, 2004], a sequential strategy for optimal sensor placement was proposed: at each iteration, only the position of one sensor is optimized, keeping all the other sensors fixed in the positions obtained through the previous algorithm steps. Two possible strategies have been highlighted:

- Forward Sequential Sensor Placement (FSSP) algorithm: in the initial configuration only one sensor is placed and, at each iteration, the

number of sensors is increased by one, such that the increment of information is maximized;

- Backward Sequential Sensor Placement (BSSP) algorithm: in the initial configuration all the model nodes are populated with n_{dof} sensors and, at each iteration, the number of sensors is decreased one by one, such that the reduction of information is minimized.

The algorithm iterations are stopped when the desired number $n_{\mathbf{y}}$ of sensors to be placed is reached.
Since the position of only one sensor at a time has to be optimized, the adoption of such strategies results in a low computational cost of the overall optimization algorithm. On the other hand, it cannot be guaranteed that the optimal solution, i.e., the global maximum of the objective function, is reached, resulting in possible sub-optimal sensors configurations.
An alternative approach for the maximization of the objective function would be to compute the MC estimator $\hat{U}(\mathbf{d})$ over a search grid in the design space $\mathcal{D}$, as a brute force optimization approach, and then to select the optimal solution $\mathbf{d}$ which guarantees the maximum expected Shannon information gain; this strategy is practically unfeasible, for $n_{\mathbf{y}} > 1$, as the number of required points, for exploring the design space, would increase exponentially with the space dimensions.
Since the estimator $\hat{U}(\mathbf{d})$ is based on a Monte Carlo sampling approximation, the resulting objective function is noisy, due to sampling of both the measurement error $\boldsymbol{\epsilon} \sim p_{\boldsymbol{\epsilon}}$ and of the parameters vector $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{d})$. As discussed in Section 4.2.2, the prior pdf can be assumed to be independent of the position of the sensors, i.e. $\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathbf{d}) = p(\boldsymbol{\theta})$. According to this non-restrictive assumption, the same batch of samples $\boldsymbol{\theta}$ can be used for each value of $\mathbf{d}$, resulting in a less noisy objective function. Moreover, since there is no need to re-sample $\boldsymbol{\theta}$ and re-compute the corresponding response for each different value of $\mathbf{d}$ and at each iteration of the optimization procedure, a consistent reduction in the overall computational cost is achieved.
Due to the noisy objective function, the application of deterministic optimization methods [Cavazzuti, 2013] to the problem at hand could lead to false local optima. For this reason, the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is here adopted: it is an iterative evolutionary derivative-free algorithm suitable for stochastic optimization problems, introduced in [Hansen et al., 1995, Hansen and Ostermeier, 2001]. The CMA-ES is suitable for non-linear problems and it was applied in many different research areas, such as energy grid optimization [Ramesh et al., 2012],

Algorithm 3 Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm.

Input : $\mathbf{m} \in \mathbb{R}^n$, $\sigma \in \mathbb{R}_+$, N^{opt}, $\mathbf{d}_0$
Output Optimal configuration $\mathbf{d}^*$
.
Initialize *parameters* c_c, c_σ, c_1, c_σ, d_σ, i
while *Stopping criteria in Eq. (4.46) is not met* **do**
  for $j = 1, ..., N^{opt}$ **do**
    Sample the design variable: $\mathbf{d}_j = \mathbf{m} + \sigma\gamma_j$, $\gamma_j \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$
    Evaluate the objective function through (Eq. (4.15): $\hat{U}_j = \hat{U}(\mathbf{d}_j)$)
  end
  Update parameters
  begin
    Sort samples: $\hat{U}(\mathbf{d}_{1:N^{opt}}) \leq ...\hat{U}(\mathbf{d}_{i:N^{opt}})... \leq \hat{U}(\mathbf{d}_{N^{opt}:N^{opt}})$
    Update mean: $\mathbf{m} = \sum_{i=1}^{\mu} w_i \mathbf{d}_{i:N^{opt}} = \mathbf{m} + \sigma\gamma_w$, with $\gamma_w = \sum_{i=1}^{\mu} w_i \gamma_{i:N^{opt}}$
    Cumulation for $\mathbf{C}$: ${\mathbf{p}_c}^i = (1 - c_c){\mathbf{p}_c}^{i-1} + \sqrt{1 - (1 - c_c)^2}\sqrt{\mu_w}\gamma_w$
    Cumulation for σ: ${\mathbf{p}_\sigma}^i = (1 - c_\sigma){\mathbf{p}_\sigma}^{i-1} + \sqrt{1 - (1 - c_\sigma)^2}\sqrt{\mu_w}(\mathbf{C}^{i-1})^{-1/2}$
    Update $\mathbf{C}$: $\mathbf{C}^i = (1 - c_1 - c_\mu)\mathbf{C}^{i-1} + c_1 + {\mathbf{p}_c}^{i-1}{{\mathbf{p}_c}^{i-1}}^T + c_\mu \sum_{i=1}^{\mu} w_i \gamma_{i:N^{opt}} {\gamma_{i:N^{opt}}}^T$
    Update σ: $\sigma_i = \sigma_{i-1} e^{\frac{c_\sigma}{d_\sigma}\left(\frac{||p_\sigma||}{E||\mathcal{N}(\mathbf{01})||} - 1\right)}$
    Update $i = i + 1$
  end
end

Bayesian uncertainty quantification [Hadjidoukas et al., 2015], groundwater management optimization [Bayer et al., 2009] and computer science [Bliss et al., 2014], where better performances than other existing optimization algorithms for stochastic problems were shown [Hansen et al., 2010].

The pseudo-code of the CMA-ES is listed in Algorithm 3. The algorithm is based on an evolutionary strategy, where, at each iteration i, a total number N^{opt} of samples $\mathbf{d}$ are drawn from a multivariate normal distribution $\mathbf{d}_j \sim \mathbf{m} + \sigma\mathcal{N}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C} \in \mathbb{R}^{n_\mathbf{d} \times \mathbf{d}}$ is the covariance matrix, $\mathbf{m} \in \mathbb{R}^{n_\mathbf{d}}$ is the mean and σ is the step size. Then, the values of $\mathbf{m}$, $\mathbf{C}$ and σ are updated in order for the population of new points $\mathbf{d}_{1:N^{opt}}$ to move towards the maximum of the objective function $\hat{U}(\mathbf{d})$ (computed through Eq. (4.15)). The search path evolution of the design variable (i.e. the sequence of consecutive steps of the mean $\mathbf{m}$) is performed through the so-called "cumulation" technique (as detailed in Algorithm 3). The vector $\mathbf{d}_0$ defines the initial

conditions of the optimization procedure, while μ and μ_w are parameters needed to control the update phase. The iterations are stopped whenever at least one of the two following criteria is fulfilled:

$$\begin{cases} \left|\hat{U}(\mathbf{d}_k) - \hat{U}(\mathbf{d}_{k-1})\right| \leq \rho_U \\ \|\mathbf{d}_k - \mathbf{d}_{k-1}\| \leq \rho_{\mathbf{d}} \end{cases} \tag{4.46}$$

where the symbol $|\cdot|$ stands for the absolute value of the argument, $\|\cdot\|$ represents an appropriately chosen norm (in the problem at hand, the L^2 norm has been used), ρ_U and $\rho_{\mathbf{d}}$ are parameters which allow to control the accuracy of the solution, respectively either in terms of objective function or design variable. These parameters cannot be chosen a priori, as they are dependent on the specific application, on the model discretization and on the desired accuracy.
For further details on the algorithm the interested reader may refer to [Hansen, 2006, Hansen, 2016].
The CMA-ES offers several advantages: first of all, it is not necessary to compute the gradient of the objective function with respect to the unknown variable, but only the function values. Moreover, thanks to the randomized search, it is possible to efficiently handle both noisy objective functions and the presence of local optima.

4.5 Algorithm

The procedure for computing the optimal sensor configuration is listed in Algorithm 4 and the corresponding flowchart is shown in Fig. 4.1.
First, the parameter vector $\boldsymbol{\theta}$ is sampled from the prior pdf $p(\boldsymbol{\theta})$, which is chosen a priori. For each sample $\boldsymbol{\theta}_i$ (with $i = 1, ..., N^{PCE}$), the corresponding response vector $\mathbf{v}_i = \mathbf{v}(\boldsymbol{\theta}_i)$ is computed through the FE model. Then, the dimension of the response vector is reduced from n_{dof} to l by performing the PCA of $\mathbf{V} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_{N^{PCE}}]$. A total number l of model surrogates is built, by considering $\boldsymbol{\theta}$ as the input variable and the components of the reduced-space vector $\mathbf{T}_l$ as the output variables. A fresh batch of $N >> N^{PCE}$ samples $\boldsymbol{\theta}$ is drawn from the prior $p(\boldsymbol{\theta})$ and the corresponding system response is computed through the PCE surrogates, according to Eq. (4.44). In the end, the optimal configuration is obtained through the CMA-ES optimization method (listed in Algorithm 3), where the evaluation of the objective function is performed through the MC estimator defined in Eq. (4.15).

Algorithm 4 Algorithm for the optimization of SHM sensor networks through Bayesian experimental design.

Input : Prior pdf $p(\boldsymbol{\theta})$
Output Optimal configuration $\mathbf{d}^*$
:
for $i = 1 : N^{PCE}$ **do**
  Sample parameters values: $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$
  Compute system response: $\mathbf{v}_i = \mathbf{v}(\boldsymbol{\theta}_i)$
end
Perform PCA: find $\mathbf{W}_l$, $\mathbf{T}_l$ s.t. $[\mathbf{v}_1 \ \ldots \ \mathbf{v}_{N^{PCE}}] = \mathbf{V} \cong {\mathbf{W}_l}^T \mathbf{T}_l$
Compute surrogate model $\mathcal{M}^{PCE}(\boldsymbol{\theta})$, which links the inputs $\boldsymbol{\theta}$ with the output $\mathbf{v}_l$
for $i = 1 : N$ **do**
  Sample parameters values: $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$
  Compute system response: $\mathbf{v}_i = {\mathbf{W}_l}^T \mathcal{M}^{PCE}(\boldsymbol{\theta}_i)$
end
Optimization: run Algorithm 3

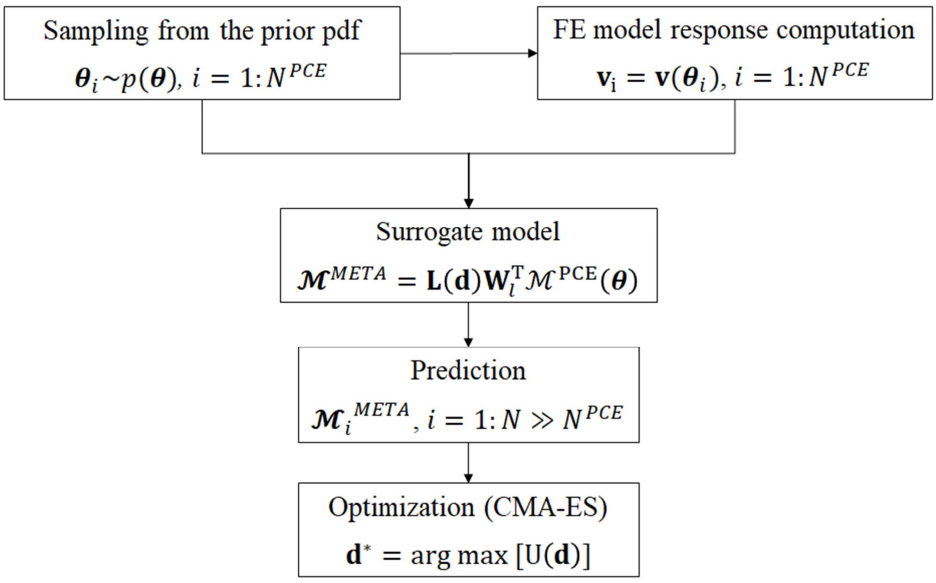


Figure 4.1. *Graphical flowchart of the proposed procedure.*

4.6 Optimal SHM system design

In the previous paragraphs, a stochastic strategy to optimally place sensors on a structure for estimating a set of parameters has been introduced and discussed. As stated in Section 4.2.2, the optimization problem consisted in determining the optimal spatial configuration $\mathbf{d}^*$ such that the associated expected Shannon information gain $U(\mathbf{d})$ is maximized. Up to this point, one main assumption has been put in place: the number of measurements $n_{\mathbf{y}}$ and the pdf $p_{\boldsymbol{\epsilon}}$ of the prediction error have been considered as fixed parameters. It is now assumed to take into account these settings as additional variables of the optimization problem.
$\boldsymbol{\epsilon}$ is supposed to be sampled from a zero mean Gaussian noise $p_{\boldsymbol{\epsilon}} = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix. The expected Shannon information gain would in general depend on both the sensor configuration, the number of measurements and the prediction error, i.e. $U = U(\boldsymbol{d}, n_{\mathbf{y}}, \boldsymbol{\Sigma})$.
The prediction error accounts for both the modeling error and the measurement error. Assuming independence between these two errors, the covariance matrix can be then written as:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_m + \boldsymbol{\Sigma}_n \tag{4.47}$$

where $\boldsymbol{\Sigma}_m$ and $\boldsymbol{\Sigma}_n$ respectively account for the model and the measurement error. In [Papadimitriou and Lombaert, 2012], it has been shown that the optimal sensor configuration can be affected by the spatial correlation among different measurements, which is taken into account in $\boldsymbol{\Sigma}_m$. In practice, supposing that the correlation among any couple of measurements decays exponentially with the distance between their locations, the spatial correlation length controls the optimal spatial configuration (i.e., the lower it is, the closer the sensors turn out to be placed). On the other hand, $\boldsymbol{\Sigma}_n$ can be related with the type of sensor to be employed in the SHM system, i.e. the instrumental noise which depends on its characteristics (such as signal-to-noise ratio). For the sake of simplicity, it is now assumed that there is no correlation between different measurements associated with the measurement noise and, therefore, $\boldsymbol{\Sigma}_n = \sigma^2\mathbf{I}$.
Let us assume now that the goal of the optimization procedure is the design of the SHM system in a more general setting, i.e., the spatial configuration $\mathbf{d}$, the number of sensors $n_{\mathbf{y}}$ and the type of sensors are unknown and have to be optimized. The model error $\boldsymbol{\Sigma}_m$ is assumed to be kept fixed and treated as a constant in the optimization problem; therefore, it can affect the optimal configuration, but it is not an object of the optimization procedure.

The optimization statement in Eq. (4.7), thus, becomes:

$$(\boldsymbol{d}^*, n_{\mathbf{y}}^*, \sigma^*) = \arg\max \left[U(\boldsymbol{d}, n_{\mathbf{y}}, \sigma)\right] \tag{4.48}$$

In order to better understand how these additional design variables can affect the objective function U, let us define the function $\bar{U} = U(\boldsymbol{d}^*, n_{\mathbf{y}}, \sigma)$ [Capellari et al., 2017b], which corresponds to the maxima values of the objective function for each value of $n_{\mathbf{y}}$ and σ. $\boldsymbol{d}^*$ is then the optimal configuration obtained by solving the relevant optimization statement, with fixed values of $n_{\mathbf{y}}$ and σ. Since $\boldsymbol{d}^*$ depends on the choice of $(n_{\mathbf{y}}, \sigma)$, then it is possible to conclude that the function $\bar{U} = \bar{U}(n_{\mathbf{y}}, \sigma)$ depends on $(n_{\mathbf{y}}, \sigma)$ only.
As analytically proven in [Papadimitriou et al., 2000] and numerically shown in Section 5.3 [Capellari et al., 2016c], the maximum of the objective function here considered increases as the number of sensors gets higher, since more information is provided by the SHM system. Moreover, it can be numerically shown that if the standard deviation σ of the measurement noise increases, then the information provided gets lower. Therefore, since the function U is a monotonically increasing function with respect to $n_{\mathbf{y}}$ and a monotonically decreasing function with respect to σ, additional constraints should be added to the unconstrained problem in Eq. (4.7).
Three types of constraints can be here highlighted:

(a) identifiability constraint: $n_{\mathbf{y}} > n_{iden}$, where n_{iden} is the minimum number of measurements which are required in order to guarantee identifiability of the parameters $\boldsymbol{\theta}$ (see [Katafygiotis and Beck, 1998, Yuen, 2010]);

(b) technological constraint: $\sigma > \sigma_{sens}$, where σ_{sens} is the lowest standard deviation of the measurement noise, associated with the sensors available on the market which can be used to measure the chosen physical quantities;

(c) cost constraint: $C(n_{\mathbf{y}}, \sigma) \leq B$, where $C(n_{\mathbf{y}}, \sigma)$ is the cost model of the SHM system and B is the maximum budget which can be spent for the SHM system.

The complete optimization problem can therefore be defined as:

$$\begin{aligned} &(\boldsymbol{d}^*, n_{sens}^*, \sigma^*) = \arg\max \left[U(\boldsymbol{d}, n_{sens}, \sigma)\right] \\ &\text{subject to } \begin{cases} n_{sens} > n_{obs} \\ \sigma > \sigma_{best} \\ C(n_{sens}, \sigma) \leq B \end{cases} \end{aligned} \tag{4.49}$$

The cost of the SHM network is here taken into account through the cost model $C(n_{\mathbf{y}}, \sigma)$. The simplest cost model consists in a constant sensor network cost C_0, which e.g. accounts for the data acquisition hardware, database, assemblage, etc., and a variable cost, which accounts for the sensor cost. The associated cost function is thus:

$$C(n_{\mathbf{y}}, \sigma) = C_0 + c(\sigma)\, n_{\mathbf{y}} \tag{4.50}$$

where $c(\sigma)$ is the unitary cost per sensor. In real applications or when a small set of sensors is available and a model cost cannot be established, the continuous optimization problem in Eq. (4.49) is simply re-interpreted as a discrete problem, with $\sigma \in \mathbf{S}$ and $\mathbf{S}$ being the set of possible standard deviations associated with the available sensors.
In order to solve the optimization problem, a possible approach would be to embed the unknown variables $n_{\mathbf{y}}$ and σ in the design variable $\mathbf{d}$; then, to obtain the optimal solution through the CMA-ES algorithm (see Algorithm 3), setting up the associated constraints.
An alternative procedure, which is particularly suitable for real applications where only a limited set of sensor types are available, is to explore the function $\bar{U} = U(\boldsymbol{d}^*, n_{\mathbf{y}}, \sigma)$, i.e. computing the maximum expected Shannon information gain over a search grid of points $\{n_{\mathbf{y}}, \sigma\}$. This approach is viable since only 2 dimensions, respectively associated with $n_{\mathbf{y}}$ and σ, need to be explored.
The optimization statement discussed so far aims at designing the SHM sensor network, such that the information provided is maximized, given a certain budget B. Following a classical decision making approach (see [Khoshnevisan et al., 2002, Parnell et al., 2011]), a procedure based on cost-benefit analysis can be alternatively followed. In the problem at hand the benefit is represented by the expected Shannon information gain. Despite $U(\boldsymbol{d}, n_{\mathbf{y}}, \sigma)$ cannot be directly traduced in an associated expected monetary gain (benefit), it is possible to define a relevant utility-cost index (UCI) [Capellari et al., 2018]:

$$\mathrm{UCI}(\boldsymbol{d}, n_{\mathbf{y}}, \sigma) = \frac{U(\boldsymbol{d}, n_{\mathbf{y}}, \sigma)}{C(n_{\mathbf{y}}, \sigma)} \tag{4.51}$$

where the associated unit of measure is [nat/€] ([nat] stands for natural unit of information).

The associated optimization problem would then be:

$$(\boldsymbol{d}^*, n_{\mathbf{y}}^*, \sigma^*) = \arg\max \left[\frac{U(\boldsymbol{d}, n_{\mathbf{y}}, \sigma)}{C(n_{\mathbf{y}}, \sigma)} \right]$$
$$\text{subject to } \begin{cases} n_{\mathbf{y}} > n_{obs} \\ \sigma > \sigma_{best} \\ C(n_{\mathbf{y}}, \sigma) \leq B \end{cases} \tag{4.52}$$

This optimization formulation allows to obtain the most efficient SHM design, i.e., to maximize the information per unitary cost.
Regarding the solution of the optimization problem, the same considerations previously presented holds also for the case in Eq. (4.52). The optimal solutions can be obtained by maximizing the associated objective function $\mathrm{UCI}(\boldsymbol{d}^*, n_{\mathbf{y}}, \sigma)$, where $\boldsymbol{d}^*$ is the optimal spatial configuration for each $\{n_{\mathbf{y}}, \sigma\}$ in the search grid.
The application of the two strategies defined in Eqs. (4.49) and (4.52) will be discussed in Section 5.4.
It is important to underline that, since the measurements $\mathbf{y}$ depend on the loads to which the structure is subjected, then the optimal sensor placement is necessarily dependent on the type, location and magnitude of the acting forces. Therefore, likewise the deterministic method, if one wants to obtain an optimal sensor placement which is robust with respect to the loading conditions, several optimizations with different loads should be performed and the solution should be chosen among them.

4.7 Conclusions

In the present chapter, a stochastic optimal sensor placement method for structural health monitoring has been presented.
The goal of the strategy is to determine the optimal sensor network configuration such that the Shannon information gain between the prior and the posterior pdfs of the parameters to be estimated is maximized. The sensor positions are defined through an appropriate design variable and the corresponding optimal value has to be obtained. The objective function can be numerically approximated through a Monte Carlo sampling approach: the resulting estimator turns out to be a double sum of terms, which depend on the likelihood function. Since a high number of model response evaluations is required, two procedures based on surrogate models and model order reduction strategies have been proposed. The PCE allows to build a computationally efficient meta-model in order to mimic the relation between its

input and output variables. One option is to define a joint input variable including both the model parameters and the design variable. This choice implies that only one surrogate model is needed, but it requires a large number of samples when complex geometries and multiple physical quantities can be measured by the sensors. An alternative approach is to exploit a model order reduction technique (PCA) and use the surrogate model to compute the reduced order response as a function of the model parameters. Since the resulting objective function is affected by noise, leading to possible undesired local maxima, an evolutionary strategy (CMA-ES) suitable for stochastic problems has been used.
If not only the sensor configuration, but also other SHM system network settings, i.e., number and type of sensors, have to be optimized, a comprehensive procedure for the optimal design has been proposed. In order to find the optimal solution, the cost, identifiability of the model parameters $\boldsymbol{\theta}$ and technological constraint have to be taken into account. Moreover, a different optimization problem can be established by maximizing the information gain per unitary cost, in a sort of cost-benefit analysis.
It should be highlighted that the whole strategy is completely non-intrusive, i.e., it does not require computation of the gradient, but exclusively relies on model responses. Moreover, the method is general and no restrictive assumptions, such as linearity or Gaussianity, are placed.

CHAPTER 5

Stochastic optimal sensor placement: numerical experiments

5.1 Introduction

In this chapter, the stochastic optimal sensor placement method introduced in Chapter 4 is applied to two structural instances.
First (see Section 5.2), a simple benchmark case, i.e., the clamped square plate earlier described in Section 3.1.1, is considered. The surrogate formulation introduced in Section 4.3.2.1 is used and its performances are analyzed with respect to the measurement noise and to the prior pdf of the parameters $\boldsymbol{\theta}$. The results are then compared with those obtained through the deterministic OSP method introduced in Chapter 2.
Then, a real-size structure, i.e., the Pirelli tower in Milan, is considered (see Section 5.3). The two meta-modeling formulations presented in Sections 4.3.2.1 and 4.3.2.2 are compared, both in terms of accuracy and computational cost. Next, the two estimators of the expected Shannon information gain introduced in Section 4.3, namely the Monte Carlo and the Kraskov estimators, are considered and the associated advantages are highlighted. In the end, the optimal solutions, in terms of sensor spatial

configurations, are computed through the CMA-ES algorithm (see Section 4.4).
In Section 5.4, a strategy for a comprehensive optimal design (see Section 4.6) of the SHM system is applied to the Pirelli tower structural system.

5.2 Optimal sensor placement on a thin square plate

The first application here considered is the structure described in Section 3.1.1, i.e. a clamped square plate subjected to a vertical force applied at the center point (as shown in Fig. 3.1). Please refer to Section 3.1.1 for the description of the mechanical properties of the mechanical system.
The plate is modeled through shell elements, whose nodes feature 6 DOFs, namely 3 displacements (u_{x_1}, u_{x_2} and u_{x_3}) and 3 rotations (φ_{x_1}, φ_{x_2} and φ_{x_3}). The same 20×20 FE discretization mesh used in Section 3.1.1 is here exploited. In order to simplify the problem, the plate is supposed to be divided in 4 regions (as shown in Fig. 5.1), hence reducing the number of damaged zones and relevant parameters to be handled: the parameters vector $\boldsymbol{\theta}$, which has to be estimated, includes the Young's moduli E_i, where the index $i = 1, ..., 4$ defines the associated region.
Two cases are here considered:

(a) the position of the damaged zone is assumed to be known and the goal of the SHM network is the estimation of the Young's modulus value E_2 at zone 2, where the damage is supposed to occur; therefore, the parameters vector $\boldsymbol{\theta}$ is defined as:

$$\boldsymbol{\theta} = \{\theta\} = \{E_2\} \tag{5.1}$$

(b) the position of the damaged zone is assumed to be unknown and the goal of the SHM system is to estimate the values of the Young's moduli at zones 1, 2, 3, 4; therefore, the parameters vector is defined as follows:

$$\boldsymbol{\theta} = \{E_1 \;\; E_2 \;\; E_3 \;\; E_4\}^T \tag{5.2}$$

In order to assess the OSP strategy for this benchmark problem, only the vertical displacement u_{x_3} is assumed to be measured by the SHM system: thus, the measurement orientation, defined through the parameter δ in Eq. (4.34), is not considered as an unknown value to be optimized. The choice of such a measurement will allow to easily check for the algorithm correctness, as compared to the physics of the problem. In Section 5.2.1, the results

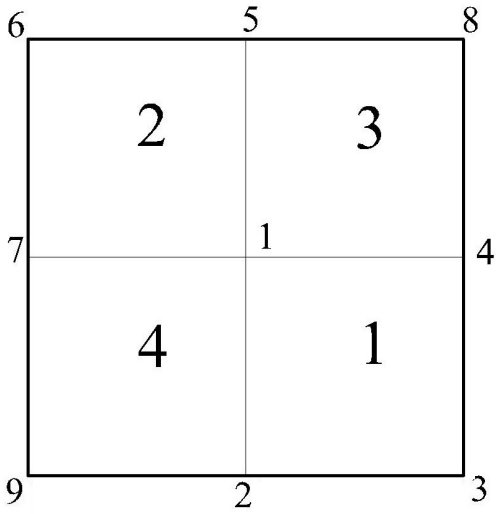


Figure 5.1. *Clamped plate: FE model nodes and elements numbering.*

obtained assuming to measure rotations φ_{x_1} or φ_{x_2} are shown, for sake of comparison with respect to the deterministic optimization scheme presented in Chapter 3 (see Section 3.1.1).
The associated design variable is defined as follows:

$$\mathbf{d} = \{{x_1}^1 \;\; {x_2}^1 \;\ldots\; {x_1}^{n_\mathbf{y}} \;\; {x_2}^{n_\mathbf{y}}\}^T \tag{5.3}$$

where $n_\mathbf{y}$ is the number of measurements. The resulting measurement vector is:

$$\mathbf{y} = \{{u_{x_3}}^1 \;\ldots\; {u_{x_3}}^{n_\mathbf{y}}\}^T \tag{5.4}$$

Since only one physical quantity (displacement u_{x_3}) in one direction is supposed to be measured, the formulation based on the joint input PCE described in Section 4.3.2.1 is applied.
Let us consider, first of all, that the position of only one sensor has to be determined ($n_\mathbf{y} = 1$): the resulting optimization problem is simply 2- dimensional, as the sensor can be placed only in the x_1-x_2 plane. The input variable of the surrogate model is defined as $\boldsymbol{\chi} = \{\boldsymbol{\theta} \;\; {x_1}^1 \;\; {x_2}^1\}^T$. The parameters $\boldsymbol{\theta}$ are supposed to be sampled from the uniform distribution $\mathcal{U}(0, E)$, where E is the Young's modulus of the undamaged material. No priority placement zones are considered and, therefore, the sensor can be equally likely placed on every location on the plate: consequently, the design variable is sampled from the uniform distribution $\mathcal{U}(\mathcal{D})$, with $\mathcal{D} = (0, \ell)\times(0, \ell)$, where $\ell = 0.2$ m is the side length of the plate. The pdf of the prediction

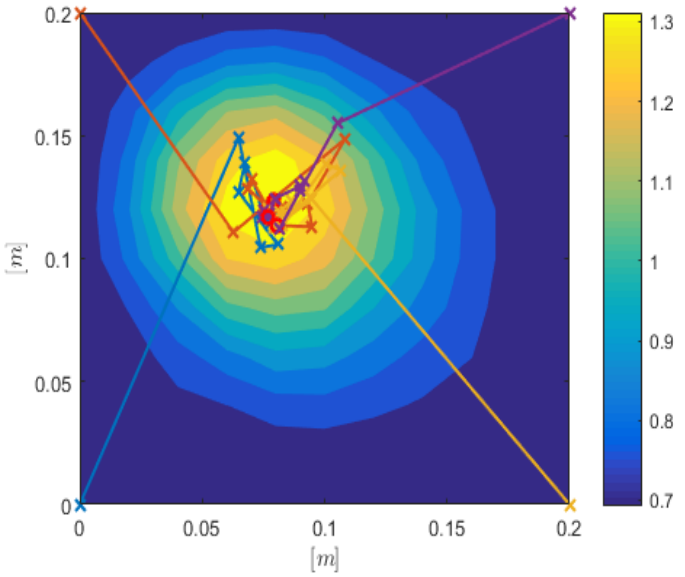


Figure 5.2. *Contour plot of the objective function $\hat{U}(\boldsymbol{d})$ with one sensor for case (a) and CMA-ES iteration points ($N^{PCE} = 10^4$, $p = 10$, $N = 5 \cdot 10^4$).*

error ϵ is further assumed to be Gaussian $\mathcal{N}(0, \sigma^2)$, with zero mean and standard deviation $\sigma = 10^{-5}$m, which is approximately equal to 1 % of the maximum displacement.

The objective function shape can be explored by computing it on a 20×20 search grid. The resulting contour plots of the expected Shannon information gain, calculated for each possible sensor position and evaluated through the MC estimator in Eq. (4.15), are shown in Figures 5.2 and 5.3, respectively for cases (a), i.e. only zone 2 is supposed to be damaged, and (b), i.e. all zones are supposed to be damaged. The corresponding optimal solution for case (a) is $\mathbf{d}^* = \{{x_1}^* \quad {x_2}^*\}^T = \{0.4\ \ell \quad 0.6\ \ell\}^T$, i.e. inside zone 2, where the material property has to be estimated. On the other hand, considering case (b), the maximum value of the objective function is at $\mathbf{d}^* = \{0.5\ \ell \quad 0.5\ \ell\}^T$, which reflects the symmetry of both the geometry and the unknown parameters distribution. This reflects the results obtained in Chapter 3, where a symmetric configuration has been achieved as well.

In the same graphs, the evolution of the solutions obtained by applying the CMA-ES optimization procedure described in Algorithm 3 is depicted. Each path corresponds to an algorithm run with a different initial condition $\mathbf{d}_0$, which is alternatively set at the 4 plate corners. The colored crosses

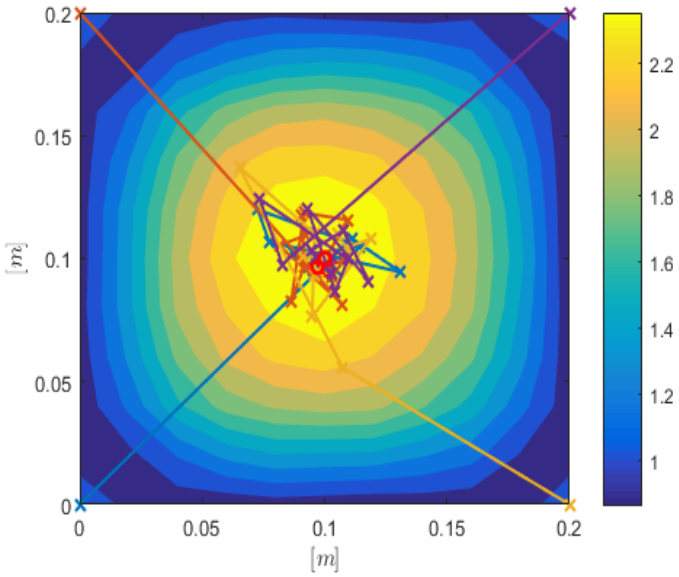
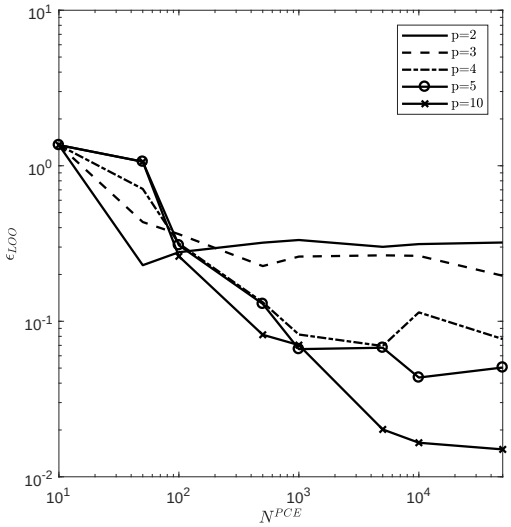
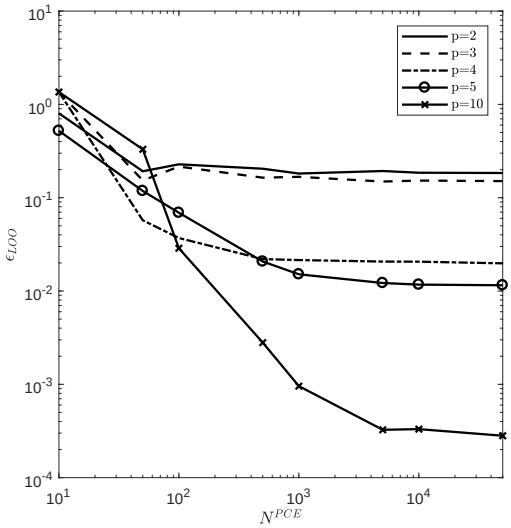


Figure 5.3. *Contour plot of the objective function $\hat{U}(\boldsymbol{d})$ with one sensor for case (b) and CMA-ES iteration points ($N^{PCE} = 10^4$, $p = 10$, $N = 5 \cdot 10^4$).*

represent the values of $\mathbf{d}$ at the end of each algorithm step, while the red circle corresponds to the optimal solution. It is shown that the solution is stable with respect to the choice of the initial settings and roughly 25 objective function evaluations are sufficient to reach the maximum.
In Figures 5.4a (case (a)) and 5.4b (case (b)), the values of ϵ_{LOO} defined in Eq. (5.5), for different number of samples N^{PCE} used for the PCE surrogate and degree p of the polynomial basis (see Appendix B), are shown. As expected, the error decreases as N^{PCE} and p increases. Moreover, the error is higher in case (b), since the dimension of the PCE surrogate input variable is greater than in case (a) (see Eqs. (5.1) and (5.2)).
As specified in Section 4.2.2, the formulation based on Bayesian experimental design naturally takes into account the uncertainties related with the problem and therefore the objective function is affected by the pdfs of $\boldsymbol{\epsilon}$ and $\boldsymbol{\theta}$. The effect of different prediction error pdfs $p_{\boldsymbol{\epsilon}}$ and parameters prior pdfs $p(\boldsymbol{\theta})$ on the objective function is here assessed.
In Fig. 5.5, the contour plot of the objective function $\hat{U}(\mathbf{d})$, for different values of the standard deviation σ, is shown. It can pointed out that as σ increases, the objective function gets noisier and several local minima and maxima appear.



(a)

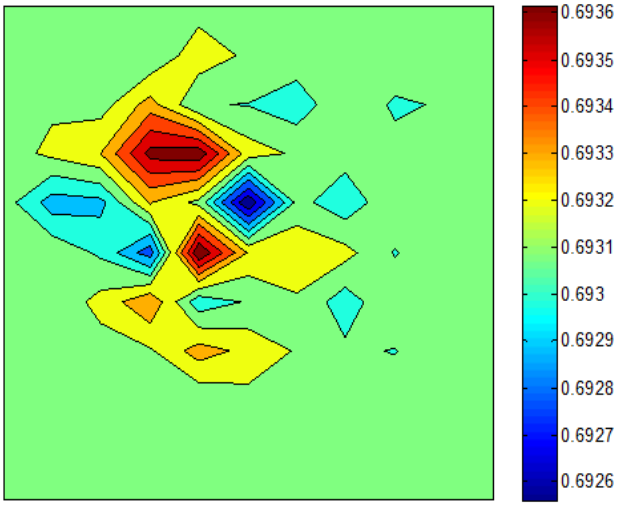


(b)

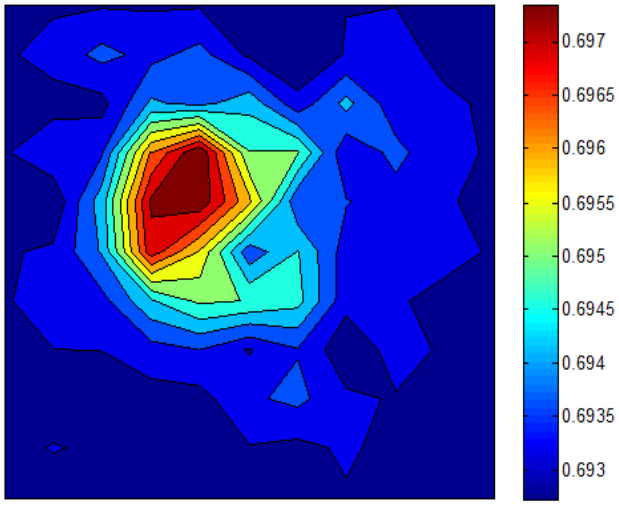
Figure 5.4. *LOO error ϵ_{LOO} (see Eq. (5.5)) associated to the PCE surrogate model for cases (5.4a) (a) and (5.4b) (b).*

Moreover, it is worth noting that the difference of $\hat{U}(\mathbf{d})$ between the global maxima, which corresponds to the optimal solution $\mathbf{d}^*$, and the global minima, which corresponds to the points near the clamped plate

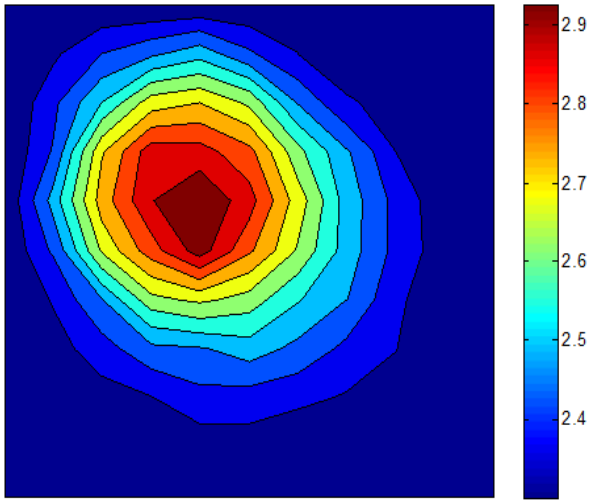
edges, decreases as σ increases. This means that if the prediction error is very high (see Fig. 5.6a), the system response is practically "hidden" by the noise and, therefore, the benefit of placing the sensor in the optimal position, rather than in a random point, gets progressively lower.
Regarding the choice of the prior pdf $p(\boldsymbol{\theta})$, in Fig. 5.6 the optimal solution $\mathbf{d}^*$ is shown, for $n_{\mathbf{y}} = 4$ measurements, obtained by running the optimization algorithm 10 times. The results are shown for $p(\boldsymbol{\theta}) = \mathcal{U}(0, E)$ and $p(\boldsymbol{\theta}) = \mathcal{U}(0.75\ E, E)$ respectively in Fig. 5.6a and Fig. 5.6b. It is evident that stability is largely affected by the choice of prior distribution $p(\boldsymbol{\theta})$, due to the fact that the support lower bound of the uniform distribution approaches the singularity condition $E = 0$.



(a)

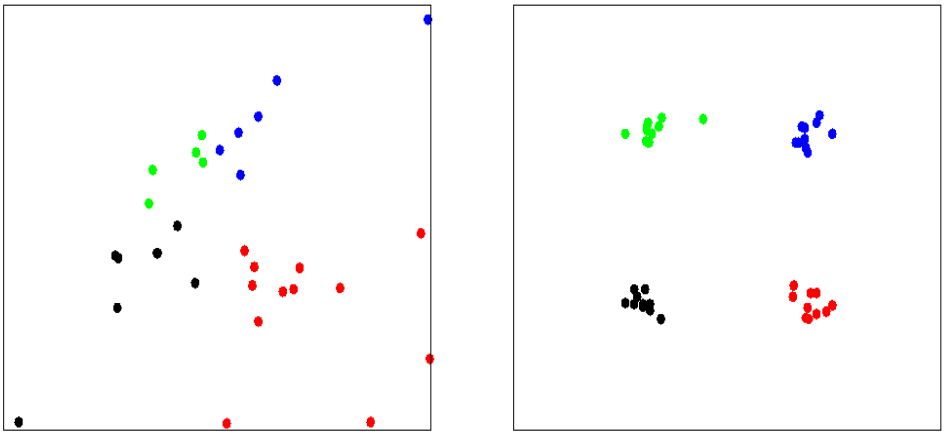


(b)



(c)

Figure 5.5. *Contour plot of the expected Shannon information gain with (a) $\sigma = 10^{-3}$ m, (b) $\sigma = 10^{-4}$ m and (c) $\sigma = 10^{-5}$ m.*



(a) (b)

Figure 5.6. *Optimal design* $\boldsymbol{d}^*$ *of* $n_{\mathbf{y}} = 4$ *measurements* ($N^{PCE} = 5 \cdot 10^3$, $p = 10$, $N = 5 \cdot 10^4$)*: results of 10 algorithm runs, with prior pdf (a)* $p(\boldsymbol{\theta}) \sim \mathcal{U}(0, E)$*, (b)* $p(\boldsymbol{\theta}) \sim \mathcal{U}(0.75\, E, E)$

5.2.1 Comparisons with the deterministic OSP method

In order to compare the stochastic procedure with the deterministic OSP method presented in Section 2.2, the values of the objective functions defined within the two algorithms are compared for the benchmark case described in Sections 3.1.1 and 5.2. It is here supposed to consider the clamped square plate, where the location of damage is assumed to be unknown (case (b)) and only the rotation φ_{x_2} about the axis x_2 is supposed to be observed through $n_{\mathbf{y}} = 1$ measurement.
In Fig. 5.7 the contour plot of the objective function ψ, defined in Eq. (2.2), is shown: as expected, the optimal sensor locations reflect the problem symmetry and the optimal measurement positions correspond to the points where the largest variations of rotations are experienced, i.e., at points $\mathbf{d}^* = \{0.25\ \ell \quad 0.5\ \ell\}^T$ and $\mathbf{d}^* = \{0.75\ \ell \quad 0.5\ \ell\}^T$. In Fig. 5.8a and 5.8b the contour plots of the expected Shannon information gain, respectively for $\sigma = 10^{-3}$ rad and $\sigma = 10^{-5}$ rad, are displayed, assuming $p(\boldsymbol{\theta}) \sim \mathcal{U}(0, E)$. In Fig. 5.9a and 5.9b the contour plots of the expected Shannon information gain, respectively for $p(\boldsymbol{\theta}) \sim \mathcal{U}(0, E)$ and $p(\boldsymbol{\theta}) \sim \mathcal{U}(0.75E, E)$, are demonstrated, supposing $\sigma = 10^{-7}$ rad.
It is possible to observe that both the deterministic and the stochastic meth-

ods approximately provide the same optimal solutions in terms of sensor placement. Nevertheless, the stochastic procedure allows to take into account additional SHM experimental settings, since the objective function depends on both the sensor location, the standard deviation σ of the prediction error and on the choice of the prior pdf $p(\boldsymbol{\theta})$. In particular, on one hand, the maximum expected Shannon information gain increases as σ gets lower, as more accurate measurements are provided by the sensors. On the other hand, it decreases if the bounds of the prior (uniform) pdf are closer, as this corresponds to a high information, which is assumed a priori taken into account, hence before performing the measurements. As shown in Fig. 5.8a, 5.8b, 5.9a and 5.9b, the contour plot of the expected Shannon information gain can be slightly non symmetric due to the noise introduced by sampling both from $p_{\boldsymbol{\epsilon}}$ and from $p(\boldsymbol{\theta})$.
Therefore, despite the higher computational cost of the stochastic method (at least N^{PCE} required FE model evaluations) with respect to the deterministic one ($n_{\boldsymbol{\theta}} + 1 = 5$ FE model analyses, see Sections 2.3), the first framework allows to take into account several experimental settings, i.e., $\mathbf{d}^*$, σ, $p(\boldsymbol{\theta})$, in the optimal design of the SHM network.

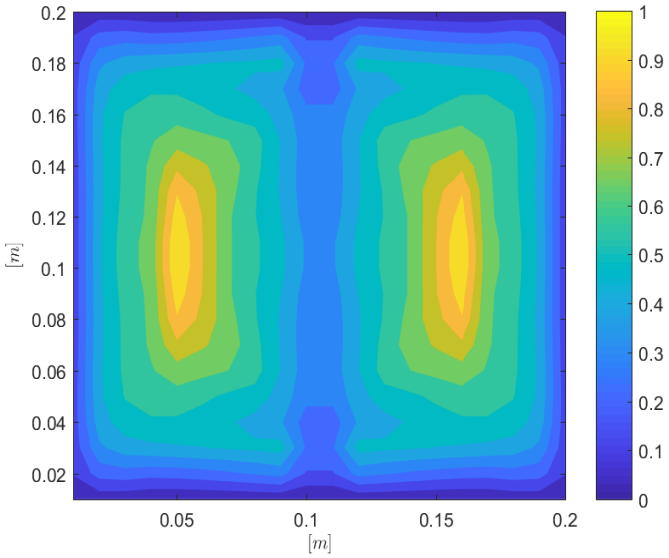
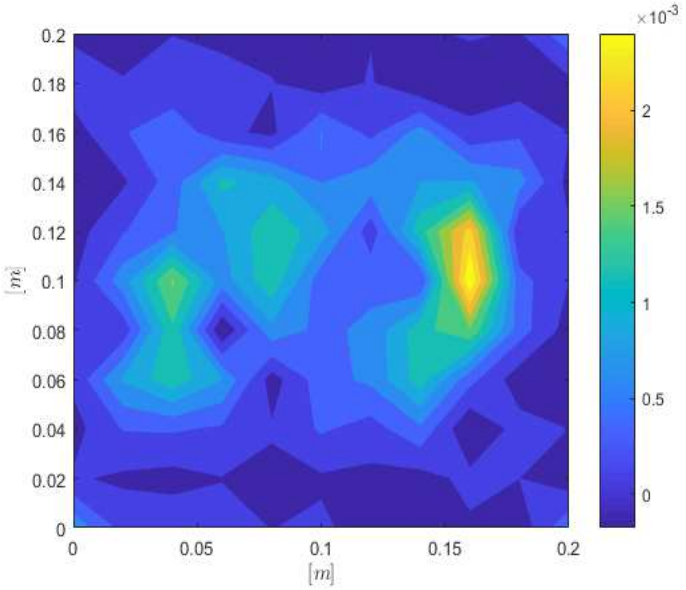
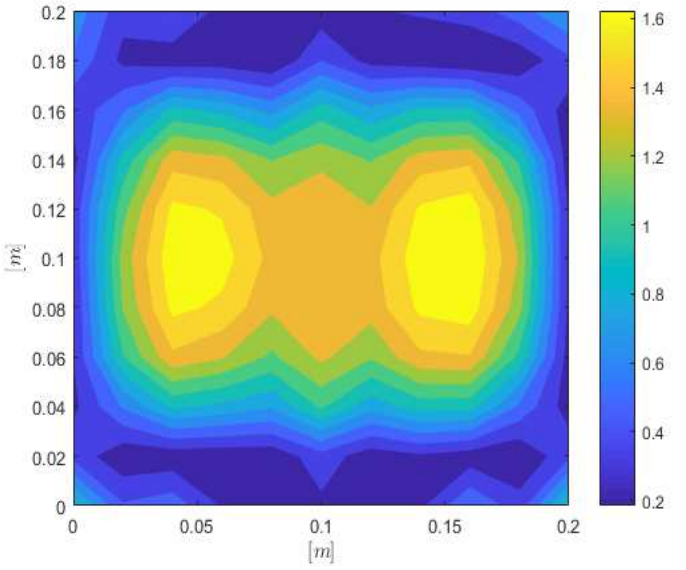


Figure 5.7. *Contour plot of the objective function ψ (Eq. (2.2)), where rotations u_{x_2} are supposed to be observed.*

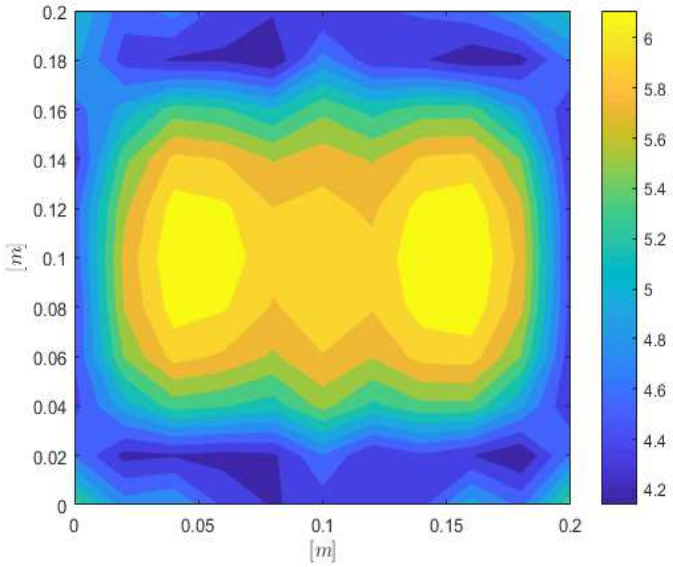


(a)

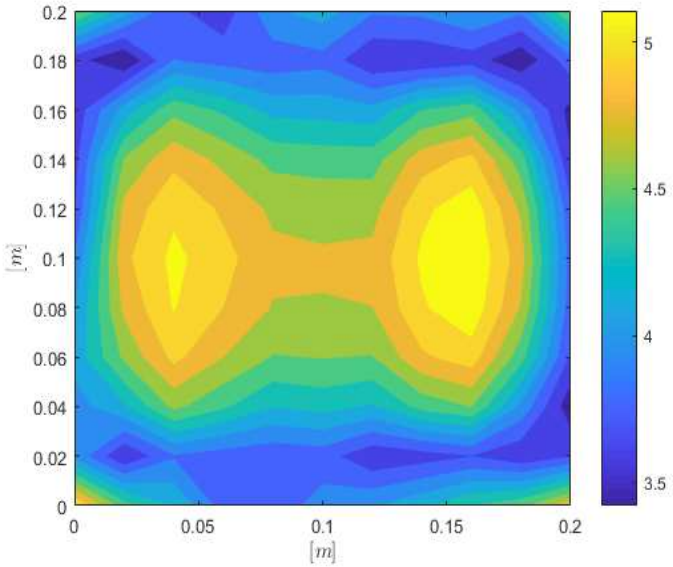


(b)

Figure 5.8. *Contour plot of the expected Shannon information gain with (a)* $\sigma = 10^{-3}$ *rad and (b)* $\sigma = 10^{-5}$ *rad (*$p(\boldsymbol{\theta}) \sim \mathcal{U}(0, E)$*).*



(a)



(b)

Figure 5.9. *Contour plot of the expected Shannon information gain with (a)* $p(\boldsymbol{\theta}) \sim \mathcal{U}(0, E)$ *and (b)* $p(\boldsymbol{\theta}) \sim \mathcal{U}(0.75\ E, E)$ *(*$\sigma = 10^{-7}$ *rad).*

5.3 Optimal sensor placement on a tall building

After the initial benchmark application considered in Section 5.2, a real large-scale structural problem, i.e., the Pirelli Tower, a known 130 m tall building in Milan (Italy), is here examined. The building consists of 35 stories out of ground, which are approximately 70 m long and 30 m wide. The structural system is entirely made of reinforced concrete; 4 symmetric triangular shaped cores at the two extremities are connected by T-shaped beams (see Fig. 5.10b). The structure has been modeled using the commercial software SAP2000 v19 (©Computer and Structures, Inc.) and the associated FE model (Fig. 5.10a) consists of 4106 nodes with 6 DOFs each, i.e. the 3 displacements $u_{x_1}, u_{x_2}, u_{x_3}$, and the 3 rotations $\varphi_{x_1}, \varphi_{x_2}, \varphi_{x_3}$, resulting in a total number of degrees of freedom $n_{dof} = 24,500$. The structure is supposed to be subjected to a horizontal force in the x_2 direction, applied at the top floor (see Fig. 5.10b), following the example introduced in [Eftekhar Azam, 2012, Eftekhar Azam and Mariani, 2018] where an online damage detection scheme has been applied to the Pirelli tower model. The force is assumed to be eccentric, as shown Fig. 5.10b, in order to induce a complex mechanical response of the tower (both displacement and torsional rotation) and, therefore, allow to better assess the employed OSP method. For further details on the structural characteristics and on the FE model, the reader may refer to [Barbella, 2009, Barbella et al., 2011, Eftekhar Azam, 2012].
The sensor placement is assumed to be optimized in order to estimate $n_{\boldsymbol{\theta}} = 6$ parameters, which are specified in Fig. 5.11. The parameters are chosen in order to make the example as general as possible: both mechanical and geometrical properties are selected, associated to both vertical and horizontal members. The chosen parameters are the Young's moduli of the column groups called LC and RC, the Young's moduli of the beam groups LB, CB and RB, the beam thickness of the group CB. The prior pdfs of each parameter are defined in Table 5.1. The prior pdfs of the concrete Young's modulus is assumed to be uniform distributions, with lower and upper bounds respectively equal to 24 GPa and 36 GPa. Regarding the beam thickness, its prior pdf is considered to be a uniform distribution, with lower and upper bounds respectively equal to 0.7 m and 0.9 m.
Unlike the previous application case (Section 5.2), it is here supposed that measurements could be either displacements or rotations in the 3 directions; therefore, the design variable has to be defined such that both the spatial position of the sensor and the physical quantity to be measured are taken into account. The associated design variable is thus defined as follows:

$$\mathbf{d} = \{x_1{}^1 \;\; x_2{}^1 \;\; x_3{}^1 \;\; \delta^1 \;\; \dots \;\; x_1{}^{n_\mathrm{y}} \;\; x_2{}^{n_\mathrm{y}} \;\; x_3{}^{n_\mathrm{y}} \;\; \delta^{n_\mathrm{y}}\}^T \tag{5.5}$$

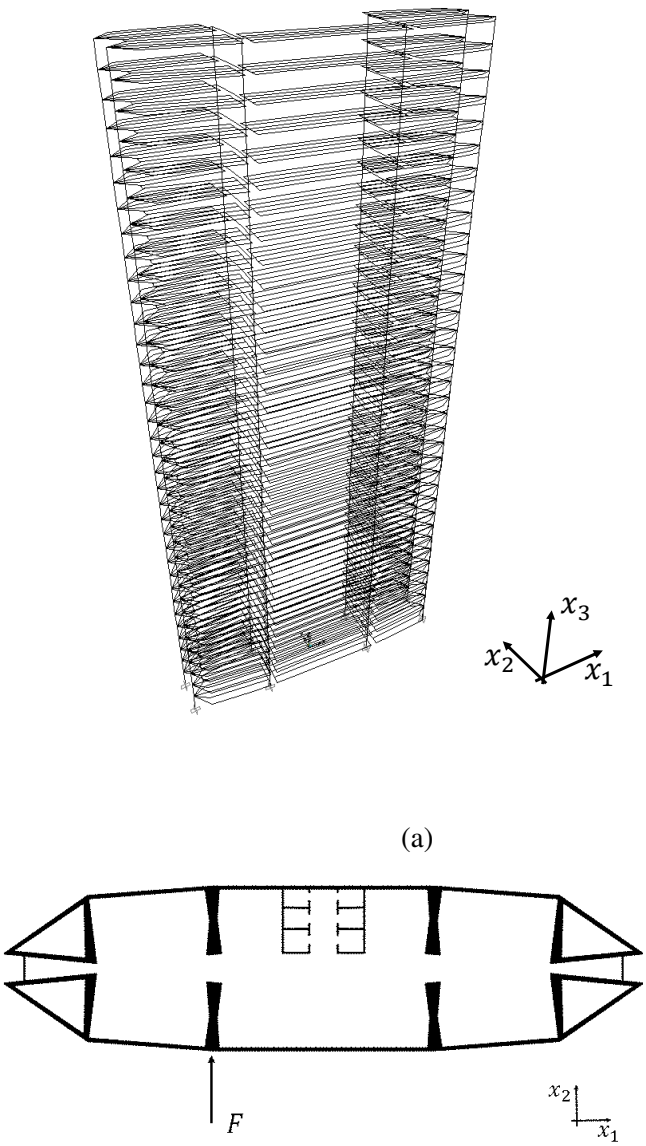
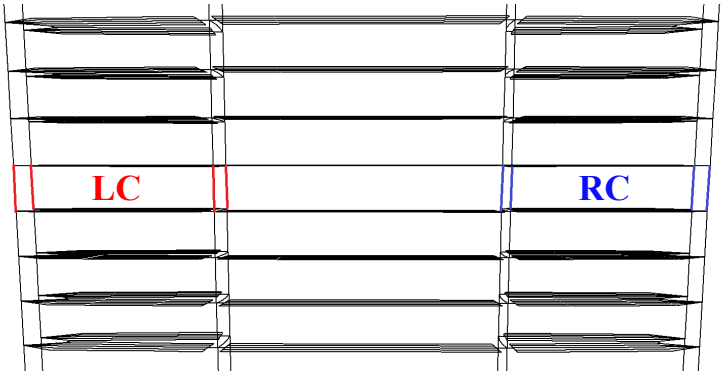
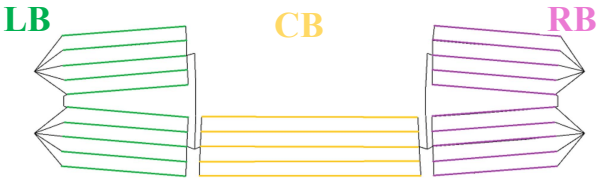


Figure 5.10. *Structural details of the Pirelli Tower: (a) FE model and (b) plan representation.*



(a)



(b)

Figure 5.11. *Locations of the $n_{\boldsymbol{\theta}}$ parameters at the 20^{th} floor: (a) front view, (b) plan view.*

Table 5.1. *Definition of parameters $\boldsymbol{\theta}$ (see Fig. 5.11) and related prior pdf $p(\boldsymbol{\theta})$.*

Position	Physical quantity	Prior pdf
Left columns (LC)	Young's modulus E [GPa]	$\mathcal{U}(24, 36)$
Right columns (RC)	Young's modulus E [GPa]	$\mathcal{U}(24, 36)$
Left beams (LB)	Young's modulus E [GPa]	$\mathcal{U}(24, 36)$
Right beams (RB)	Young's modulus E [GPa]	$\mathcal{U}(24, 36)$
Central beams (CB)	Young's modulus E [GPa]	$\mathcal{U}(24, 36)$
Central beams (CB)	Beam thickness t [m]	$\mathcal{U}(0.7, 0.9)$

For the sake of comparison, both the PCE surrogate formulations described in Sections 4.3.2.1 and 4.3.2.2 are here applied. Thus, first,

$N^{PCE} = 1000$ samples $(\boldsymbol{\theta}^i, \mathbf{d}^i)$ are respectively drawn from the pdfs $p(\boldsymbol{\theta})$ and $\mathcal{U}(\mathcal{D})$. Thereafter, the corresponding 1000 system responses $\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}^i, \mathbf{d}^i)$, supposing that $n_{\mathbf{y}} = 1$ (i.e. the monitoring system features only one measurement), are computed through the original FE model. In order to compare the FE model response and the surrogate model response, out of the 1000 input-output samples, 100 are retained for validation.
In Fig. 5.12 the FE response $\Upsilon = \mathcal{M}^{FE}(\boldsymbol{\theta}^i, \mathbf{d}^i)$ is plotted against the response $\Upsilon^{PC} = \mathcal{M}^{META}(\boldsymbol{\theta}^i, \mathbf{d}^i)$ computed through the surrogate model: Figs. 5.12a ($N^{PCE} = 20$), 5.12c ($N^{PCE} = 100$) and 5.12e ($N^{PCE} = 900$) are referred to the "joint input PCE" formulation defined in Section 4.3.2.1; Figs. 5.12b ($N^{PCE} = 20$), 5.12d ($N^{PCE} = 100$) and 5.12f ($N^{PCE} = 900$) are related to the "PCA-PCE" formulation defined in Section 4.3.2.2. In general, the more accurate a surrogate model is, the more points would cluster on the bisector of the positive x_1-x_2 quarter. It can be underlined that the PCA-PCE formulation outperforms the joint input PCE ones, as even with a low number of samples $N^{PCE} = 20$ there is a good agreement between real and surrogate response. Therefore, the application of the joint input PCE would be unfeasible for the structural case at hand, since a very large number of samples in the experimental design of the PCE surrogate would be required.
The error between real and surrogate model can be quantified by computing the Root Mean Square Relative Error (RMSRE), defined as:

$$\text{RMSRE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\Upsilon_i^{PCE} - \Upsilon_i}{\Upsilon_i}\right)^2} \tag{5.6}$$

The RMSRE is exploited in this case, instead of the more common Root Mean Square Error (RMSE) since the response magnitudes widely change between each DOF and, therefore, the errors associated with small amplitudes DOFs would be incorrectly disregarded. As shown in Fig. 5.13, the PCA-PCE clearly guarantees a low surrogate model error, which is approximately 6 orders of magnitudes lower than the one associated to the joint input PCE formulation. Due to the problem complexity, it is possible to conclude that the joint input PCE formulation cannot be here employed, as it was done for the benchmark problem in Section 5.2.
In Fig. 5.14, the total computational time required for the assembly of the surrogate models, using an Intel Core i7-4790 @ 3.60 GHz, with Microsoft Windows 10 64 bit, are shown. Although the PCA-PCE requires a further computational step (associated with the PCA analysis) than the joint input PCE, it turns out to be 5-6 times faster. It is worth noting that the computational complexity of the PCA-PCE is driven by the PCA ($O(N^{PCE})^3$),

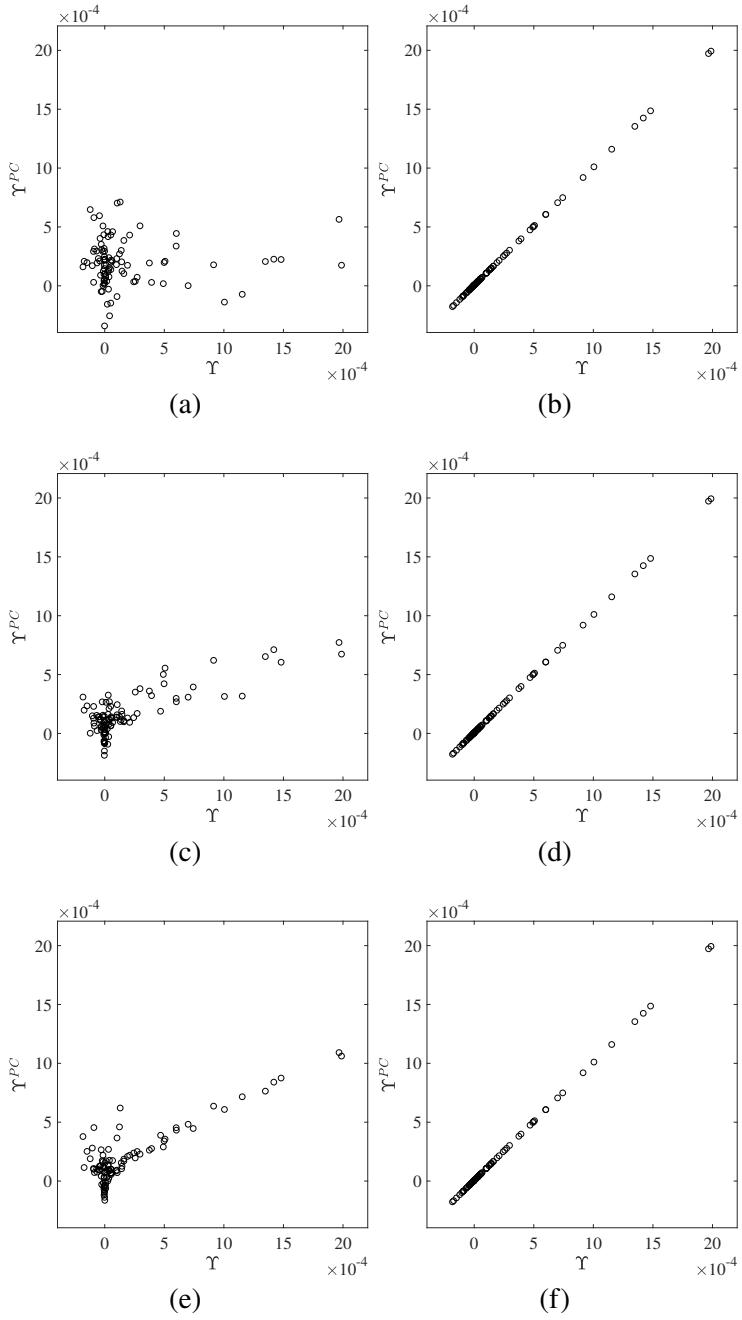


Figure 5.12. *Comparison between the FE model response* $\Upsilon = \mathcal{M}^{FE}(\boldsymbol{\theta}, \boldsymbol{d})$ *and the surrogate model response* $\Upsilon^{PC} = \mathcal{M}^{META}(\boldsymbol{\theta}, \boldsymbol{d})$*, adopting the formulation in Eq. (4.33), with (a)* $N^{PCE} = 20$*, (c)* $N^{PCE} = 100$ *and (e)* $N^{PCE} = 900$*, or the formulation in Eq. (4.44), with (b)* $N^{PCE} = 20$*, (d)* $N^{PCE} = 100$ *and (f)* $N^{PCE} = 900$*.*

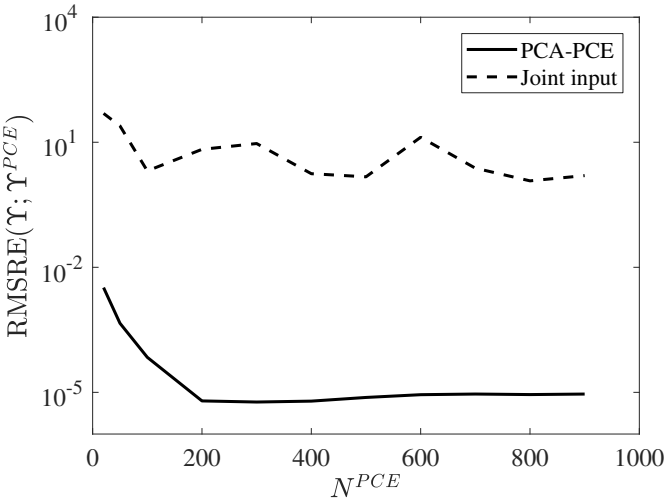


Figure 5.13. *Root mean squared relative error between* Υ *and* Υ^{PC}.

see [Trefethen and Bau III, 1997], preventing its application for high values of N^{PCE}. Nevertheless, the removal of $\mathbf{d}$ from the input PCE variable yields to a greater benefit.

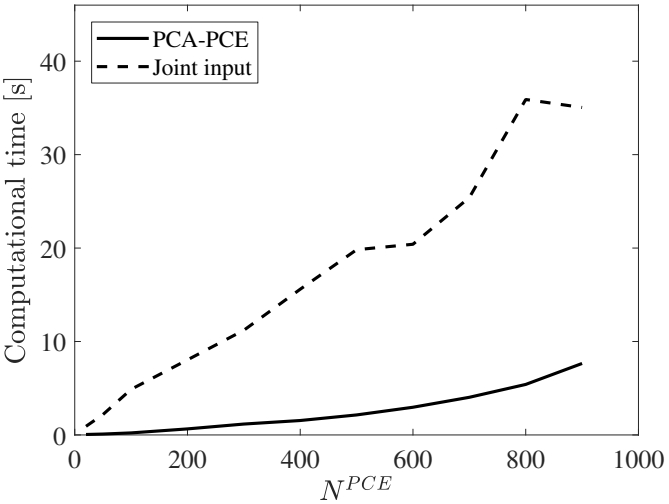


Figure 5.14. *Computational time required to build the surrogate model.*

While in Fig. 5.12 the response $\boldsymbol{\mathcal{M}}(\boldsymbol{\theta}, \mathbf{d})$, which depends on both $\boldsymbol{\theta}$ and $\mathbf{d}$, has been considered, the accuracy of the combined PCA-PCE surrogate model only is now assessed. This can be obtained by fixing the value

of the design variable $\mathbf{d}$ and comparing the response $\mathbf{W}_l^T \boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta})$ with the correspondent FE outcome. The results for the point $\mathbf{d} = \{0 \;\; 0 \;\; 130\,\text{m} \;\; 2\}$ (the index 2 identifies the displacement in the x_2 direction) at the top floor are shown in Fig. 5.15, respectively for $N^{PCE} = 20, 50, 100$. A good agreement between real and surrogate model responses is demonstrated, even for a relatively low number of samples $N^{PCE} = 50$.

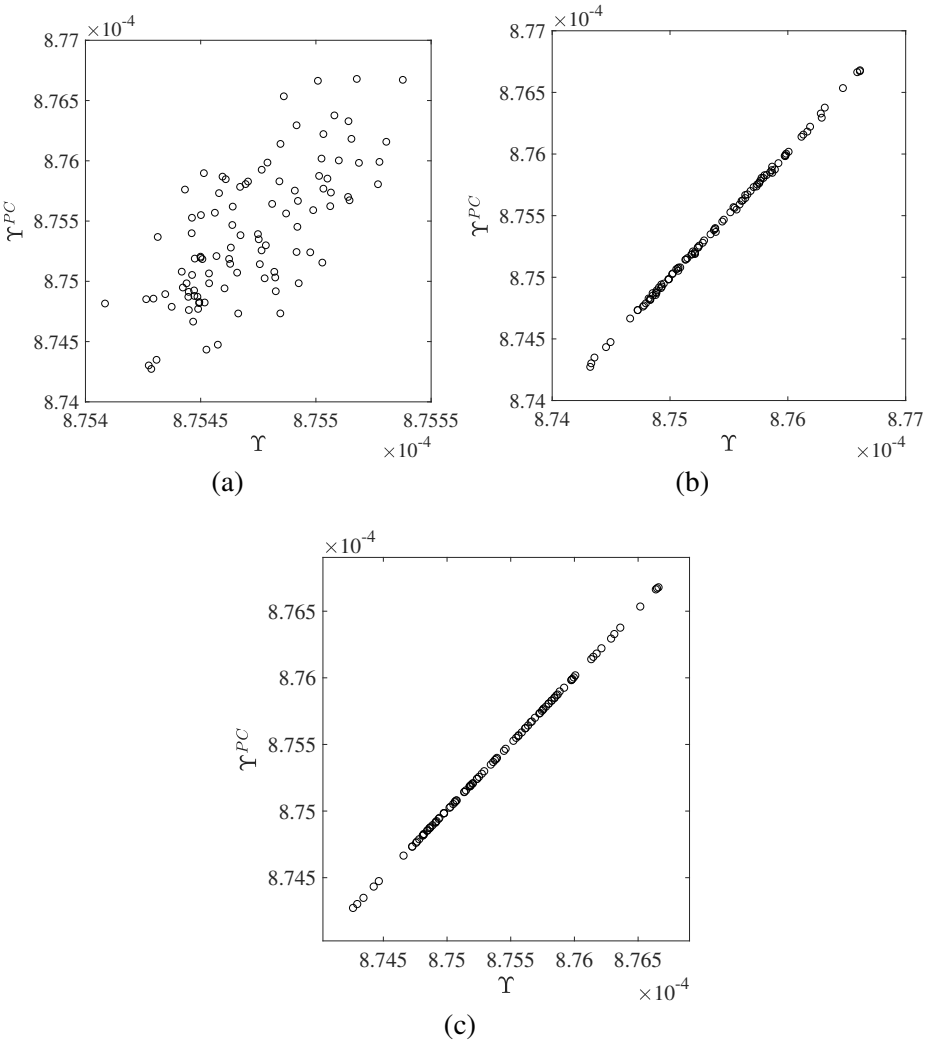


Figure 5.15. *Comparison between the FE model response* $\mathbf{v}$ *and the surrogate model response* $\mathbf{W}_l^T \boldsymbol{\mathcal{M}}^{PCE}(\boldsymbol{\theta})$*, according to the formulation in Eq. (*4.44*), with (a)* $N^{PCE} = 20$*, (b)* $N^{PCE} = 50$ *and (c)* $N^{PCE} = 100$*.*

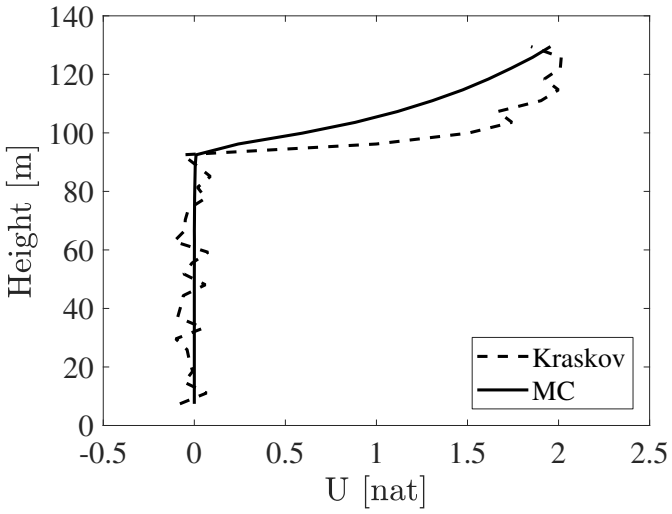


Figure 5.16. *Computation of the objective function $U(x_3^s)$: comparison between the MC estimator and the Kraskov estimator ($N = 2 \cdot 10^4$, $\sigma = 10^{-7}$).*

As discussed in Section 4.3, the expected Shannon information gain can be numerically computed either with the MC estimator $\hat{U}(\mathbf{d})$ defined in Eq. (4.15) or with the Kraskov estimator [Kraskov et al., 2004] defined in Eq. (4.16). The two estimators are now supposed to be compared, where the Kraskov estimator is implemented through the MATLAB toolbox proposed in [Szabó, 2014]. In order to effectively compare the estimated values, only one sensor, which can measure the displacement u_{x_2}, is supposed to be placed. The design variable is therefore $\mathbf{d} = \{0 \;\; 0 \;\; {x_3}^s \;\; 2\}^T$, with ${x_3}^s$ being the height at which the sensor is placed and $\delta = 2$ identifies the measured direction x_2. In Fig. 5.16 the objective function $U(x_3^s)$ computed with both the methods is shown. Despite the high number of samples $N = 2 \cdot 10^4$ here used, the Kraskov estimator results to be noisier than the MC estimator: as discussed in Section 4.3, this is due to the inherent different nature of the estimators. The computation of the Kraskov estimator relies, in fact, only on the batches of input and output samples, while in the MC estimator also the knowledge of the likelihood function is naturally taken into account.
In Fig. 5.17, the computational times required for computing the two estimators are reported. Since the MC estimator consists of a double sum (see Eq. (4.15)), it is less efficient than the Kraskov estimator. Nevertheless, on equal accuracy, the second estimator requires a higher number of samples in order to reduce the numerical noise, thus compensating the possible benefit in terms of computational cost. Therefore, the MC estimator is used in the

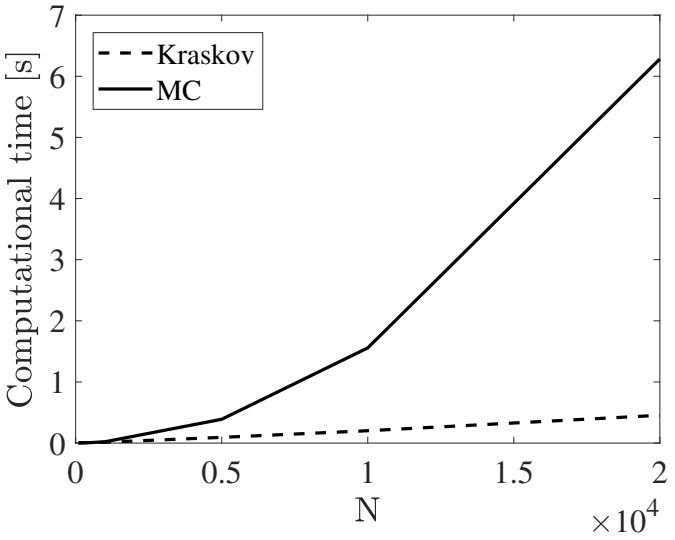


Figure 5.17. *Computational cost of the objective function $U(x_3^s)$: comparison between the MC estimator and the Kraskov estimator ($N = 2 \cdot 10^4$, $\sigma = 10^{-7}$).*

remainder of the shown results, as it allows to obtain accurate estimations of the objective function with a low number N of samples.

Next, the effect of the prediction error on the expected information gain $\hat{U}(x_3{}^s)$ is assessed, using the MC estimator. A zero-mean Gaussian noise is supposed to be added to the system response, with pdf $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. In Fig. 5.18, the values of $\hat{U}(x_3{}^s)$ for different levels of standard deviation σ and different measured directions, are shown. As demonstrated in Fig. 5.16, $\hat{U}(x_3^s) \simeq 0$ for the measured DOFs u_{x_3} (Fig. 5.18e), φ_{x_1} (Fig. 5.18b) and φ_{x_2} (Fig. 5.18d), as they are not affected by the parameters $\boldsymbol{\theta}$. On the other hand, for u_{x_1} (Fig. 5.18a), u_{x_2} (Fig. 5.18c) and φ_{x_3} (Fig. 5.18f), the related expected Shannon information gain is non-zero, as the associated measured DOFs do carry information on the parameters to be estimated. As already shown in Fig. 5.5, the information provided by the measured quantity increases as the standard deviation gets lower; on the other hand, for very high values of σ, then $\hat{U} \simeq 0$ as the system response is practically hidden by the noise. Of course, since the force is applied in the x_2 direction, any measurement in the same direction guarantees the highest possible information gain, compared to any other sensor orientation.

It can be underlined that by decreasing the standard deviation under a certain threshold which depends on the orientation of the sensor, the measurements become informative, i.e. $U > 0$, for $x_3 > x_{3\boldsymbol{\theta}}$, where $x_{3\boldsymbol{\theta}}$ is the

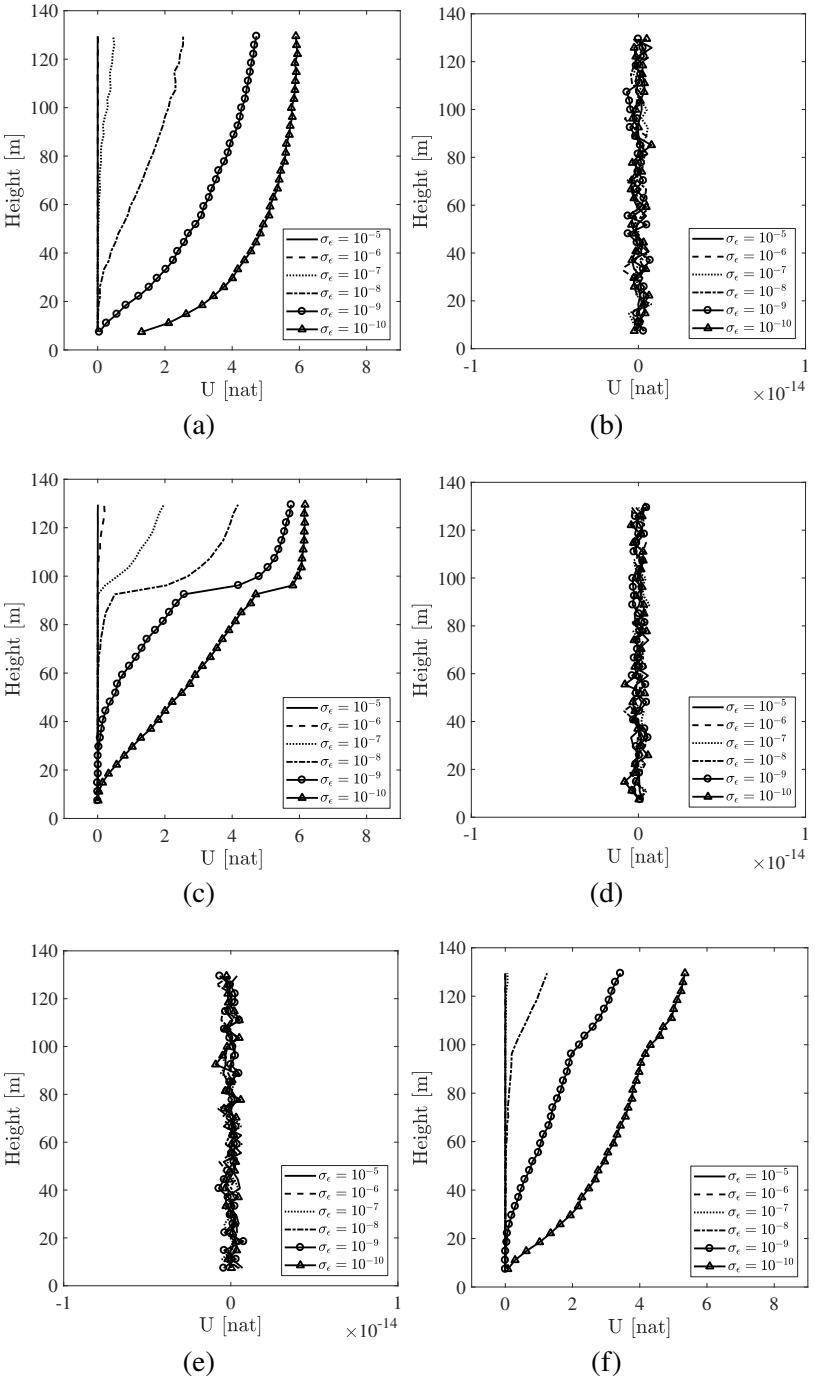
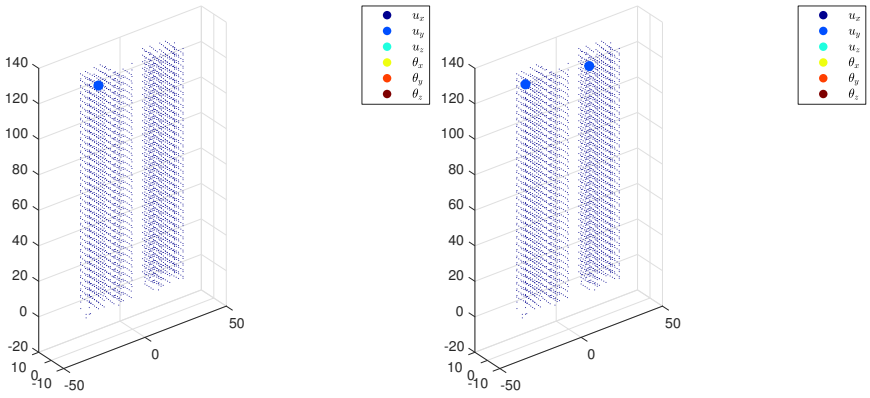
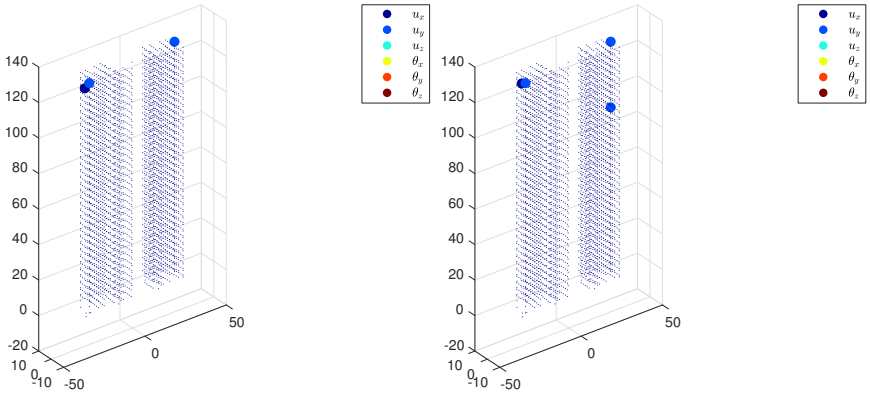


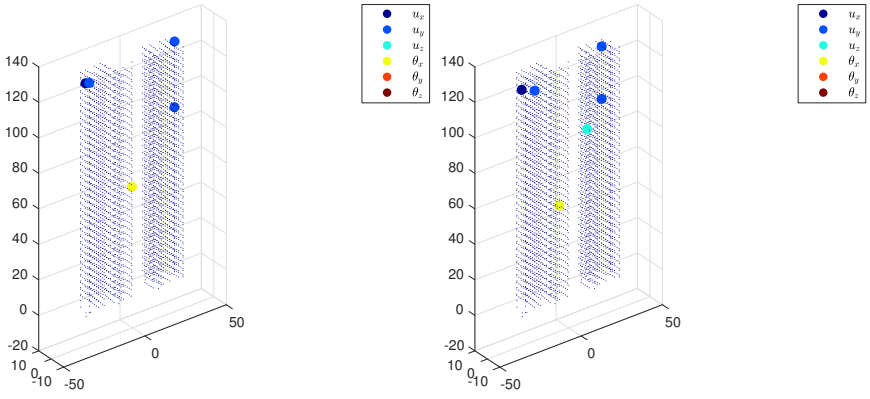
Figure 5.18. *Expected information gain $U(x_3)$, computed through the MC estimator, ($N = 2 \cdot 10^3$, $\sigma = 10^{-7}$) measuring (a) u_{x_1}, (c) u_{x_2}, (e) u_{x_3}, (b) φ_{x_1}, (d) φ_{x_2} and (f) φ_{x_3}.*


u_x
u_y
u_z
θ_x
θ_y
θ_z
140
120
100
80
60
40
20
0
-20
10
0
-10
-50
0
50

(a) (b)


u_x
u_y
u_z
θ_x
θ_y
θ_z
140
120
100
80
60
40
20
0
-20
10
0
-10
-50
0
50

(c) (d)


u_x
u_y
u_z
θ_x
θ_y
θ_z
140
120
100
80
60
40
20
0
-20
10
0
-10
-50
0
50

(e) (f)

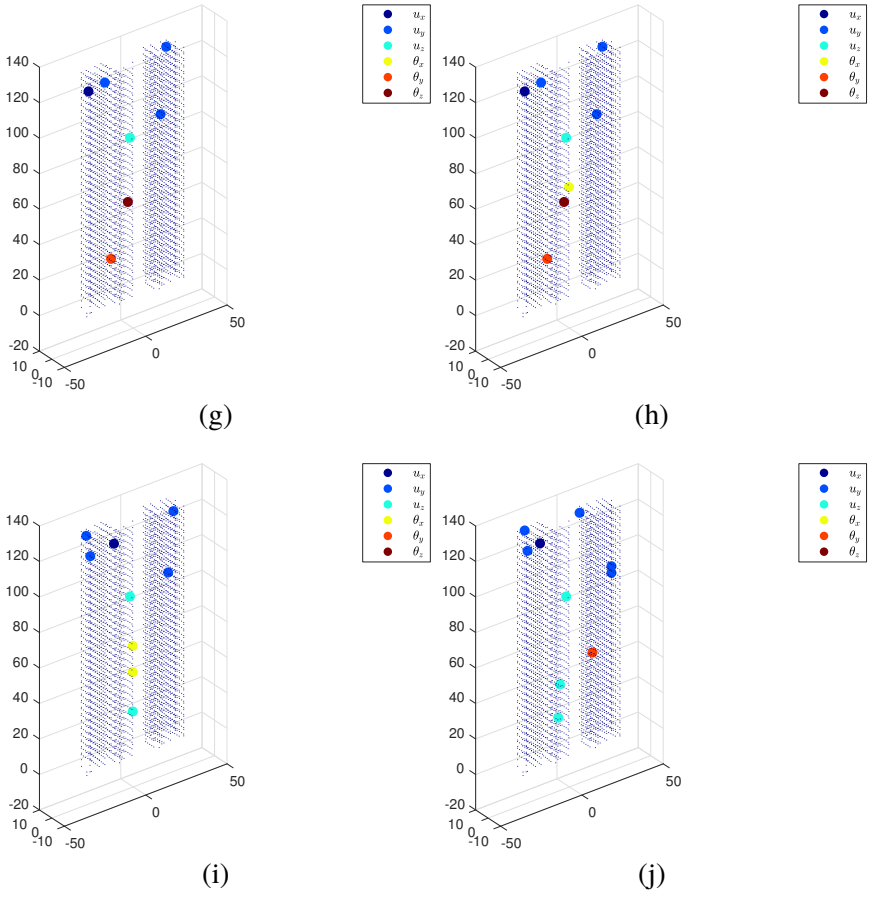


Figure 5.19. *Optimal sensor configuration* $\boldsymbol{d}^*$*, for* $\sigma = 10^{-7}$*, with (a)* $n_y = 1$*, (b)* $n_y = 2$*, (c)* $n_y = 3$*, (d)* $n_y = 4$*, (e)* $n_y = 5$*, (f)* $n_y = 6$*, (g)* $n_y = 7$*, (h)* $n_y = 8$*, (i)* $n_y = 9$*, (j)* $n_y = 10$*, .*

floor height referred to the parameters (see Fig. 5.11 and Table 5.1). This effect is higher for the measured DOFs u_{x_2} (Fig. 5.18c) and φ_{x_3} (Fig. 5.18f), as they are more affected by a variation of the parameters $\boldsymbol{\theta}$. For very low values of σ, practically unfeasible in real applications, $U > 0$ also for $x_3 < x_{3\boldsymbol{\theta}}$. In this particular application case, the information gain $\hat{U}(x_3{}^s)$ is monotonically increasing, as the response sensitivity gets higher when higher floors are considered.

Next, the most general case $n_{\mathbf{y}} > 1$, in which more than one sensor has to be optimally placed on the structure and all the DOFs can be measured ($\delta = 1, 2, 3, 4, 5, 6$), is considered. The resulting design variable $\mathbf{d}$ is defined as in Eq. (5.5). The optimization has been carried out by exploiting the CMA-ES algorithm (see Algorithm 3). Since it is not possible to a priori explore the shape of the objective function in the $n_{\mathbf{y}}$ multi-dimensional space and, therefore, to predict the presence of local maxima and minima, the algorithm has been run several times, with different values of $\mathbf{d}_0$ and then only the solution which guarantees the maximum value of $U\hat{(}\mathbf{d})$, among all the optimal solutions, has been retained.

The optimal configurations $\mathbf{d}^*$ are shown in Fig. 5.19, for (5.20a) $n_{\mathbf{y}} = 1$, (5.20b) $n_{\mathbf{y}} = 2$, (5.20c) $n_{\mathbf{y}} = 3$, (5.20d) $n_{\mathbf{y}} = 4$, (5.20e) $n_{\mathbf{y}} = 5$, (5.20f) $n_{\mathbf{y}} = 6$, (5.19g) $n_{\mathbf{y}} = 7$, (5.19h) $n_{\mathbf{y}} = 8$, (5.19i) $n_{\mathbf{y}} = 9$, (5.19j) $n_{\mathbf{y}} = 10$. Each bold dot indicates the presence of a sensor in that position, i.e. its spatial location, while the colors are related with the physical measured quantity, i.e. the value of δ^s. As it can be expected, for low $n_{\mathbf{y}}$, the sensors tends to cluster near the position of the external load (see Fig. 5.10b), and along the direction x_2. Then, by adding more sensors, also the other directions and the rotations are considered.

In order to sum up the results shown in Fig. 5.19, the resulting maximum values of the expected information gain $\hat{U}(\mathbf{d}^*)$ are reported in Fig. 5.20, as a function of the number of measurements $n_{\mathbf{y}}$ and the standard deviation σ. As it can be expected, the objective function $\hat{U}(\mathbf{d})$ increases as more measurements are considered and for lower values of the standard deviation. The objective function oscillations are due to possible local maxima. Moreover, it is worth noting that the general trend of U (not considering the numerical oscillations) appears to be a concave function, i.e. $\frac{\partial^2 U}{\partial^2 n_{\mathbf{y}}} \leq 0$: most of the information is provided by the first sensors placed, and, the more other sensors are added, the lower is the additional information gain. After a certain number of sensors added, which depends on σ, the increase in information becomes practically negligible.

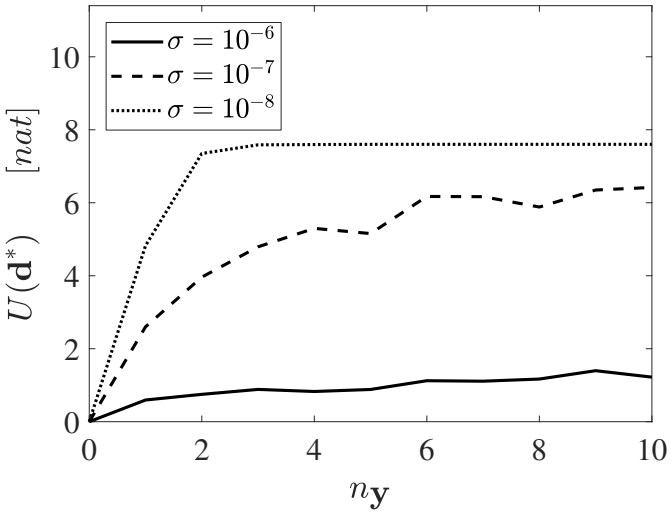


Figure 5.20. *Dependence of the expected information gain $U(\boldsymbol{d}^*)$ on the number $n_{\mathbf{y}}$ of sensors, at varying standard deviation σ of the prediction error.*

5.4 Optimal SHM system design

In the previous section, the optimization of a sensor network, only in terms of spatial configuration, has been performed: the number and type of sensors have been supposed to be fixed parameters and, therefore, they have not been considered as unknown variables in the optimization procedure. The optimization of the SHM system both in terms of spatial configuration $\mathbf{d}$, number of measurements $n_{\mathbf{y}}$ and standard deviation σ, is now considered. As discussed in Section 4.6, it is assumed that only the standard deviation σ, associated with the measurement error, can be varied and is object of the optimization procedure, while the model error is supposed to be constant. Moreover, σ is supposed to depend on the sensor characteristics: therefore, we aim at providing a procedure which allows to choose the optimal type of sensor.
The same structure described in Section 5.3 is here considered and the results obtained in the previous section, in terms of optimal sensor placement, are here employed.
The contour plot of the objective function $\bar{U}(n_{\mathbf{y}}, \sigma) = U(\mathbf{d}^*, n_{\mathbf{y}}, \sigma)$ is shown in Fig. 5.21. As expected, the maximum values of the expected Shannon information gain increase as the number of sensors gets higher and the standard deviations decreases.
As stated in the previous section, it can be observed that the increase in

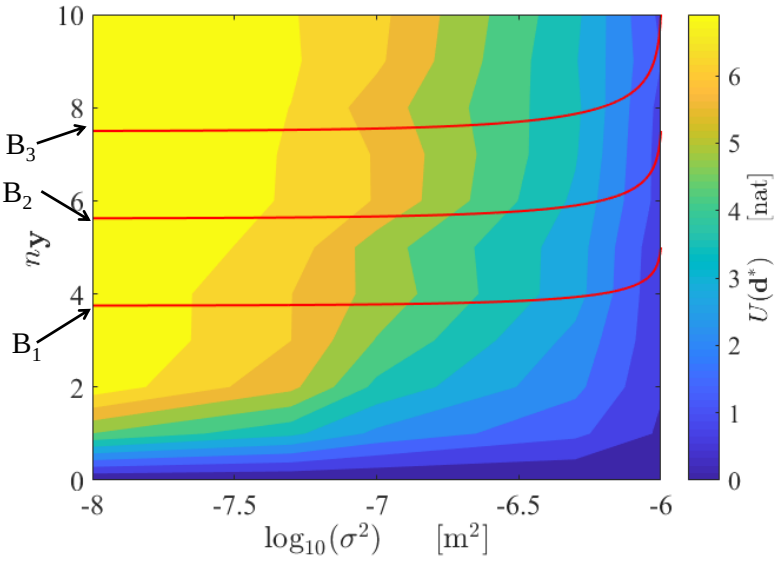


Figure 5.21. *Contour plot of $\bar{U}(n_{\mathbf{y}}, \sigma)$, where lines represents the budget constraints $B = C(\sigma, n_{\mathbf{y}})$, with $B_1 = 2000$ €, $B_1 = 2500$ €, $B_1 = 3000$ €.*

expected Shannon information gain due to each added measurement gets lower as more measurements are considered (see Fig. 5.20). In other words, the quantity $\frac{\partial U}{\partial n_{\mathbf{y}}}$ is a decreasing function of $n_{\mathbf{y}}$. From a decision-making perspective, it is interesting to underline that this behaviour corresponds to the so-called "law of diminishing marginal utility" (also known as Gossen's First Law [Gossen, 1983]), which is used in economics for the optimization of resource allocation. The law states that the marginal utility of each unit decreases as the supply of units increases. In the problem of optimal SHM system design, the utility of the sensor network is quantified by the expected Shannon information gain (see [Lindley, 1956] and Section 4.2.2) and the unit is represented by each measurement. Applications of this law to sensor network optimization in different engineering fields can be found in [Lee and Kulesz, 2008, Marbukh and Sayrafian-Pour, 2009, Tan and Zhang, 2015].
A simple linear cost model, as defined in Eq. (4.50), is assumed. The red lines in Fig. 5.21 represent different budget constraints, i.e. the solutions $\{\sigma\ n_{\mathbf{y}}\}$ of the equation $B = C_0 + c(\sigma)\ n_{\mathbf{y}}$, where B is the available budget (in this example $B_1 = 2000$ €, $B_2 = 2500$ €, $B_3 = 3000$ €). By using this chart, it is possible to optimally design the SHM network: the optimal

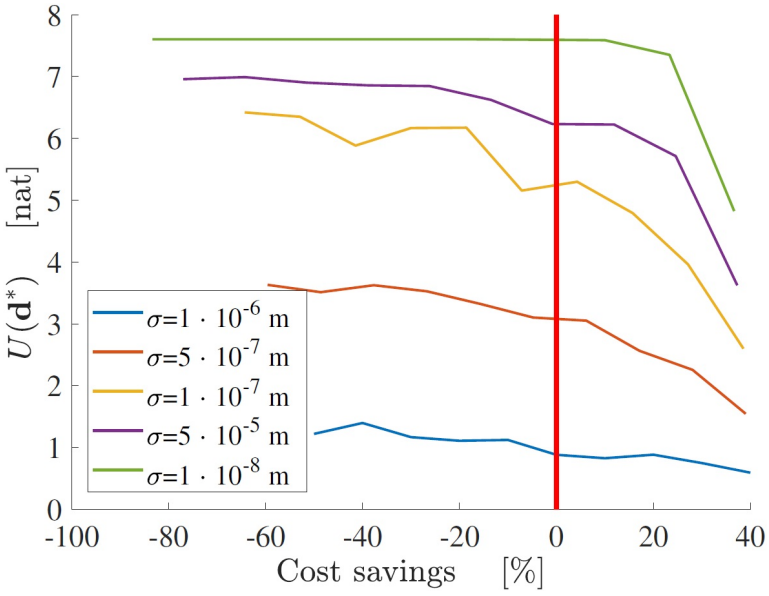


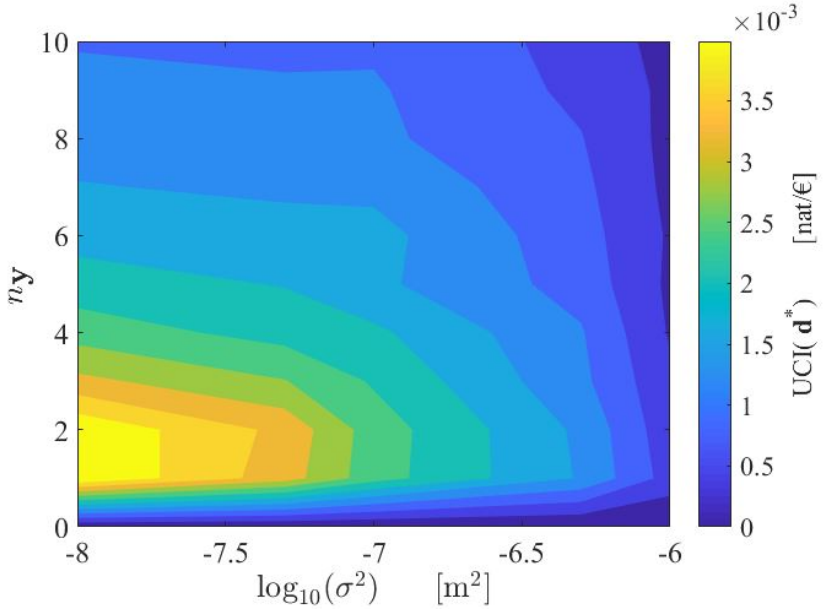
Figure 5.22. *Pareto fronts of the SHM sensor network optimization problem, for different values of standard deviation σ.*

point $\{\sigma^* \; n_{\mathbf{y}}^*\}$, for which $\bar{U}(n_{\mathbf{y}}, \sigma)$ is maximum, is ruled by the available budgetary constraint and it is uniquely associated with the corresponding optimal configuration $\mathbf{d}^*$.

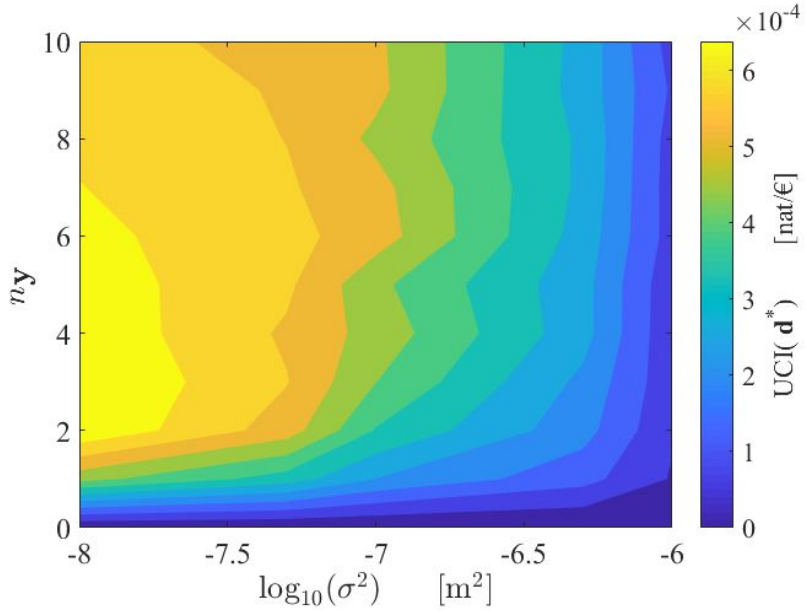
A different approach for decision making is to define a Pareto-like graph, as shown in Fig. 5.22: each line corresponds to the optimal design for a certain standard deviation, i.e. a certain type of sensors. The cost saving is defined in order to normalize the cost function with respect to the chosen budget. Any solution point located at the left-hand side of each line represents a non-optimal design solution, i.e. the associated cost does not correspond to the best choice of $\{\mathbf{d}, n_{\mathbf{y}}, \sigma\}$. The vertical straight line represents the budget B.

This graph can be particularly useful for the SHM designer in order to appropriately allocate the economic resources: for a certain chosen budget, it is possible to select the type of sensors, which turns out to be the most accurate one in Fig. 5.22, the number of sensors and their location, associated with the maximum possible expected information gain. The trend of each Pareto front provides an indication about the change of maximum utility due to a variation of budget and, thus, it helps to decide if an additional spending is justified. Moreover, fixed the value of U, it is possible to compare, from an economic point of view, different solutions in terms of number and type of

sensors.
The alternative design approach defined in Section 4.6 with Eq. (4.51) is based on the maximization of the ratio $UCI(n_{\mathbf{y}}, \sigma) = \frac{\bar{U}(\mathbf{d}^*, n_{\mathbf{y}}, \sigma)}{C(n_{\mathbf{y}}, \sigma)}$. The resulting optimal solution depends on the cost model: in Fig. 5.24a the SHM system is supposed to have a low initial cost, i.e. $C_0 = 500$ €; in Fig. 5.23b the SHM system is supposed to have a low initial cost, i.e. $C_0 = 1000$ €. In both cases, the most efficient employment of resources is reached if the best sensor, in terms of measurement noise, is chosen, while the optimal number of sensors depends on the cost model.
It is worth noting that while the objective function $\bar{U}(n_{\mathbf{y}}, \sigma)$ always increases with $n_{\mathbf{y}}$ and σ, the function $UCI(n_{\mathbf{y}}, \sigma)$ presents a maxima for $n_{\mathbf{y}} < \infty$. This is due to the fact that, as previously discussed, the increase in information associated with each additional sensor decreases as more sensors are considered. From a cost-benefit point of view, it is therefore worthless to add sensors, i.e. increase the SHM cost, if the resulting benefit (the additional expected Shannon information gain) is very low.



(a)



(b)

Figure 5.23. *Contour plot of* $UCI(n_y, \sigma) = \frac{\bar{U}(\boldsymbol{d}^*, n_y, \sigma)}{C(n_y, \sigma)}$, *with (a)* $C_0 = 500$ € *and (b)* $C_0 = 1000$ €.

5.5 Conclusions

In this chapter, the application of the stochastic optimal sensor placement method, introduced in Section 4, to two structural systems has been considered.

In the first benchmark case (Section 5.2), i.e., a clamped square plate subjected to a vertical force (Fig. 3.1), only the vertical displacement has been supposed to be measured, resulting in a 2-dimensional (coordinates x_1 and x_2) optimization problem. Since the sensor orientation is not part of the design variable to be optimized, the meta-modeling formulation based on the joint input PCE (see Section 4.3.2.1) can be employed. Due to the simplicity of the problem, it has been possible to compute the optimal solution by evaluating the objective function over a search grid. The outcome has been compared with the solutions provided by the stochastic optimization method CMA-ES, both for a known (Fig. 5.2) and unknown (Fig. 5.3) damage location. It is worth noting that just 25 objective function evaluations are sufficient, thanks to the low dimension of the problem. Nevertheless, if the standard deviation of the measurement noise is high, local minima

and maxima can appear and spoil the optimal solution search (Fig. 5.5). Moreover, the algorithm stability can be affected by the choice of the prior pdf $p(\boldsymbol{\theta})$ (Fig. 5.6), if near-zero values of the Young's modulus are sampled. By comparing the contour plots of the objective functions both for the deterministic (Fig. 5.7) and for the stochastic OSP method (Fig. 5.8a, 5.8b, 5.9a, 5.9b), it is possible to can conclude that similar resulting optimal sensor placements are obtained (Section 5.2.1). Nevertheless, the stochastic approach allows to take into account also the prediction error and the prior pdf, resulting in a more complete optimization procedure. Moreover, it is then possible to quantitatively compare different sensor network design through a theoretically sound approach, by taking into account both the number, location and pdf of the measurements.
In Section 5.3, a tall building has been considered: the problem of OSP for estimating 6 parameters (both mechanical and geometrical) has been analyzed (Fig. 5.11).
The accuracy of the two meta-modeling techniques described in Sections 4.3.2.1 and 4.3.2.2 has been considered: it has been shown that (Figures 5.12 and 5.13), due to the problem complexity, the PCA-PCE formulation outperforms the joint input PCE used in the previous case. Therefore, just a few samples (see Fig. 5.15), $N^{PCE} \simeq 50$, are sufficient to mimic the response of the FE model. Despite that the computational cost to build the PCA-PCE surrogate model increases as $O(N^3)$ (with N being the number of samples to be drawn for the evaluation of the objective function through the MC estimator), it results to be better than the joint PCE for low numbers of samples (Fig. 5.14).
Two estimators of the expected Shannon information gain have been considered, namely the MC estimator (Eq. (4.15)) and the Kraskov estimator (Eq. (4.16)). While the Kraskov estimator guarantees a lower computational cost (see Fig. 5.17), the estimated objective function is prone to a higher numerical noise (see Fig. 5.16). This is due to the type of approach: while the Kraskov estimator computation is only based on the sampled points, in the MC estimator the knowledge of the likelihood function is assumed.
The optimal solutions for different number of measurements $n_{\mathbf{y}}$ have been computed through the CMA-ES (Fig. 5.19). The optimal sensor configurations tends to cluster near the force location; nevertheless, the employment of the proposed optimization procedure allows to select also different orientations. It has been then shown (Fig. 5.20) that the maximum expected Shannon information gain of the SHM system increases as more sensors are added to the system and lower standard deviations of the prediction error

are considered.
In order to optimally design the sensor network, both in terms of spatial configuration, number of measurements and type of sensors, a cost-benefit analysis has been proposed. It has been shown that the optimal solution, in terms of maximum information gain, does not necessarily correspond to the most efficient one (see Fig. 5.21). The main reason is that the increase in information gain, due to additional sensors, gets lower as more measurements are considered. A Pareto-front approach can also be followed in order to choose the best solution, both in terms of maximum information and minimum cost (Fig. 5.22).
An alternative procedure based on the maximization of the utility-cost ratio can be used to optimally allocate the available resources. In this case, the optimal solution depends on the variation of the sensor network cost with respect to the number of measurements and to the sensor type (see Figures 5.23b and 5.24a). It is worth noting that the same considerations holds also if only a few type of sensors are available and, therefore, if it is not possible to establish a cost model. The optimization can be performed in the same way, i.e., by computing the maximum values of the objective function (which correspond to the optimal spatial configurations) over the discrete search grid.

CHAPTER 6

Practical identifiability

6.1 Introduction

In Chapter 4, a method for optimally design sensor networks for SHM, based on Bayesian experimental design, has been proposed. Once the sensor spatial configuration has been optimized, a key factor to be taken into account in order to obtain an effective monitoring system, in terms of capability of estimating the structural parameters, is the identifiability of the quantities to be estimated.

The concept of identifiability was first introduced in [Bellman and Åström, 1970] and further discussed in [Godfrey and DiStefano III, 1987]: it concerns the uniqueness of the estimation problem, i.e. whether a set of parameters can be uniquely identified, given the experimental data obtained through the sensors.

The definition of identifiability has been specialized for the problem of Bayesian model updating in structural mechanics in [Katafygiotis and Beck, 1998]. From a Bayesian perspective, a set of parameters $\boldsymbol{\theta}$ are said to be: globally identifiable if it exists a unique solution of the estimation problem $\boldsymbol{\theta}^* = \arg\ \max[p(\boldsymbol{\theta}|\mathbf{y})]$; locally identifiable if a finite number of solutions exists, i.e. the posterior pdf presents several local maxima; non-identifiable

if there is an infinite number of solutions $\boldsymbol{\theta}^*$. In other words, the parameters are said to be non-identifiable if they cannot be uniquely determined by the measurement data.
It is possible to define two different types of identifiability: structural (or model) identifiability and practical identifiability.
The first one is related with the model only, and, thus, it depends on the relations between parameters and model outputs. As model identifiability is basically associated only to the mathematical relation between observables and parameters, i.e. on the locations and number of sensors, it could happen that a model identifiable parameter cannot be practically identified due to the lack of information provided by the measurements [Raue et al., 2009].
Several approaches for studying the model identifiability have been proposed, using different methods: differential algebra [Glad and Ljung, 1990, Diop and Fliess, 1991, Ljung and Glad, 1994], power series expansion [Pohjanpalo, 1978], similarity transformation [Walter and Lecourtier, 1981]. A thorough review on the methods can be found in [Miao et al., 2011] and in [Chatzis et al., 2014]. These methods cannot be exploited for assessing practical identifiability since only noise-free experimental data are considered. In [Udwadia and Sharma, 1978, Katafygiotis and Beck, 1998, Mukhopadhyay et al., 2014], some specific algorithms for structural systems have been proposed. In particular, in [Katafygiotis and Beck, 1998], the idea was to approximate the posterior pdf using asymptotic approximation and to study its local curvature by exploiting the Hessian matrix. Following the same concept, the use of the Fisher information matrix has been introduced in [Jacquez and Greif, 1985, Papadimitriou, 2004]. These methods can, sometimes, fail to assess practical identifiability, as they only rely on local (Gaussian) approximations of the posterior pdf, and, thus, the exact shape of the posterior pdf cannot be considered.
In [Raue et al., 2009] a method for studying the practical identifiability, which is based on profile likelihood, was proposed.
In the optimal design of SHM systems, the problem of practical identifiability cannot be ignored: the optimization of the experimental settings (in terms of location, number and type of sensors) does not guarantee that all the parameters can be individually identified or distinguished from each other.
We aim now at proposing a method for investigating the practical identifiability of the parameters by means of information-based measures. Information theory has been employed in [Nienaltowski et al., 2015] and [Pant and Lombardi, 2015] to study identifiability problems in dynamical models; in [Kam et al., 1987], the concepts of stability and observability have been

highlighted within an information theory approach.
It can be pointed out that practical non-identifiability can basically arise due to two causes [Brun et al., 2001, Raue et al., 2009]: lack of sensitivity of the measured quantities with respect to the parameters and compensation of the effects of the parameters on the measurements. In order to address these two issues and measure their occurrences, the employment of the Mutual Information (MI) and Conditional Mutual Information (CMI) is proposed [Capellari et al., 2017c].
The chapter is organized as follows: first, a brief overview of the information measures and the relevant definitions will be provided in Section 6.2.1. Then, the problem of practical identifiability within the information theory will be discussed in Section 6.2.2. A method to numerically approximate the aforementioned indexes will be explained in Section 6.2.3. In the end, the application of the proposed approach to a non-linear structural problem will be given is Section 6.3.

6.2 Theoretical background

6.2.1 Preliminary definitions

In this section, the definitions of MI and CMI will be recalled: all the variables here employed are purely general and they do not refer to a specific physical problem. The defined quantities will be then specialized for SHM applications in Section 6.2.2.
Let us consider two random variables $\mathbf{Q} \in \mathbb{R}^{n_\mathbf{Q}}$ and $\mathbf{R} \in \mathbb{R}^{n_\mathbf{R}}$, the associated marginal pdfs $p(\mathbf{q})$ and $p(\mathbf{r})$ and the joint pdfs $p(\mathbf{q}, \mathbf{r})$. The MI between the random variables $\mathbf{Q}$ and $\mathbf{R}$ is defined as follows [Cover and Thomas, 2012]:

$$I(\mathbf{Q}; \mathbf{R}) = \int_\mathbf{Q} \int_\mathbf{R} p(\mathbf{q}, \mathbf{r}) \ln \left[\frac{p(\mathbf{q}, \mathbf{r})}{p(\mathbf{q}) p(\mathbf{r})} \right] d\mathbf{q} d\mathbf{r} \tag{6.1}$$

The MI can also be also expressed as the KLD (defined in Eq. (4.8)) from the product of the marginal distributions $p(\mathbf{q})p(\mathbf{r})$ to the joint pdf $p(\mathbf{q}, \mathbf{r})$, as:

$$I(\mathbf{Q}; \mathbf{R}) = D_{KL}[p(\mathbf{q}, \mathbf{r}) || p(\mathbf{q})p(\mathbf{r})] \tag{6.2}$$

Following the interpretation of the KLD, the MI can therefore be viewed as the difference (or similarity) in information between the joint and the marginal pdfs, or, in other words, the degree of correlation between the two random variables $\mathbf{Q}$ and $\mathbf{R}$. Accordingly, it can be underlined that, if $\mathbf{Q}$ and $\mathbf{R}$ are independent, then the joint pdf is $p(\mathbf{q}, \mathbf{r}) = p(\mathbf{q})p(\mathbf{r})$ and, thus,

the MI becomes $I(\mathbf{Q};\mathbf{R}) = 0$.
The amount of information associated with the random variable $\mathbf{Q}$ can be measured through the Shannon entropy, which is defined as:

$$H(\mathbf{Q}) = -\int_{\mathbf{Q}} p(\mathbf{q})\ln[p(\mathbf{q})]d\mathbf{q} \tag{6.3}$$

Moreover, the information associated with $\mathbf{Q}$, given $\mathbf{R}$, can be quantified through the conditional Shannon entropy, defined as:

$$H(\mathbf{Q}|\mathbf{R}) = -\int_{\mathbf{Q},\mathbf{R}} p(\mathbf{q},\mathbf{r})\ln[p(\mathbf{q}|\mathbf{r})]d\mathbf{q}d\mathbf{r} \tag{6.4}$$

As shown in Appendix A, the MI can therefore be related to the Shannon entropies through the following expression:

$$I(\mathbf{Q};\mathbf{R}) = H(\mathbf{Q}) - H(\mathbf{Q}|\mathbf{R}) \tag{6.5}$$

Thus, the mutual information $I(\mathbf{Q};\mathbf{R})$ can also be interpreted as the difference of Shannon entropy between $p(\mathbf{q})$ and $p(\mathbf{q}|\mathbf{r})$.
An additional random variable $\mathbf{S}$ is now considered. The Conditional Mutual Information (CMI) is defined as follows [Cover and Thomas, 2012]:

$$\begin{aligned} I(\mathbf{Q};\mathbf{R}|\mathbf{S}) &= \mathbb{E}_{\mathbf{S}}[I(\mathbf{Q};\mathbf{R})|\mathbf{S}] = \\ &= \int_{\mathbf{S}} p(\mathbf{s}) \int_{\mathbf{Q}} \int_{\mathbf{R}} p(\mathbf{q},\mathbf{r}|\mathbf{s})\ln\left[\frac{p(\mathbf{q},\mathbf{r}|\mathbf{s})}{p(\mathbf{q}|\mathbf{s})p(\mathbf{r}|\mathbf{s})}\right] d\mathbf{q}d\mathbf{r}d\mathbf{s} \end{aligned} \tag{6.6}$$

where the expression $\mathbb{E}_{\mathbf{S}}[\square]$ represents the conditional expectation of the argument $\square$ with respect to the variable $\mathbf{S}$. The CMI of two random variables $\mathbf{Q}$ and $\mathbf{R}$ can be interpreted analogously to the MI, whenever the pdfs of the same variables are conditioned on the third variable $\mathbf{S}$.
As for the MI (Eq. (6.2)), the CMI can be defined through the KLD as:

$$I(\mathbf{Q};\mathbf{R}|\mathbf{S}) = \mathbb{E}_{\mathbf{S}}[D_{KL}[p(\mathbf{s},\mathbf{r}|\mathbf{s})||p(\mathbf{q}|\mathbf{s})p(\mathbf{r}|\mathbf{s})]] \tag{6.7}$$

The definition of the Interaction Information (II) is now recalled as [Cover and Thomas, 2012]:

$$I(\mathbf{Q};\mathbf{R};\mathbf{S}) = I(\mathbf{Q};\mathbf{R}) - I(\mathbf{Q};\mathbf{R}|\mathbf{S}) \tag{6.8}$$

Considering Eqs. (6.1) and (6.6), the II can be interpreted as the difference between the MI and the CMI, i.e. the difference between the information shared by $\mathbf{Q}$ and $\mathbf{R}$ when $\mathbf{S}$ is given, and the same quantity when $\mathbf{S}$ is not given. Whenever the II is positive, it can be concluded that the knowledge of $\mathbf{S}$ enhances the correlation between $\mathbf{Q}$ and $\mathbf{R}$ (redundancy), while if the II is negative, the corrlation decreases (synergy) [Pearl, 1988].

6.2.2 Assessment of the practical identifiability through information theory

Let us consider the model which describes the relation between the parameters $\boldsymbol{\theta}$ and the measurements $\mathbf{y}$, which are obtained through the SHM system, as defined in Eq. (4.20). Supposing to make explicit the loading vector $\mathbf{f} \in \mathbb{R}^{n_{dof}}$ as a model input, Eq. (4.20) can be written as:

$$\mathbf{y} = \mathbf{L}(\mathbf{d})\mathbf{v}(\boldsymbol{\theta}, \mathbf{f}) + \boldsymbol{\epsilon} = \boldsymbol{\mathcal{M}}(\mathbf{d}, \boldsymbol{\theta}, \mathbf{f}) + \boldsymbol{\epsilon} \tag{6.9}$$

The model identifiability has been defined within the Bayesian framework in [Beck and Katafygiotis, 1998]:

- a parameter θ_i is said to be locally identifiable (where i identifies the i-th parameter) if θ_i^* is uniquely specified by $\mathbf{f}$ and $\mathbf{y}$, within a neighborhood of each of its possible values;
- a parameter θ_i is said to be globally identifiable if it is uniquely identified by $\mathbf{f}$ and $\mathbf{y}$;
- a parameter θ_i is model-unidentifiable if none of the previous conditions holds.

In other words, the parameters are non-identifiable if the parameter estimation problem is characterized by an infinite number of solutions [Yuen, 2010].
As specified in Section 6.1, while model identifiability allows to mathematically assess whether the parameters can be estimated, as a function of the structural model only, we are now interested in studying the sources of practical non-identifiability, which can compromise the effectiveness of the sensor network, even if it has been optimized using the procedures presented in Chapters 2 and 4. It can be pointed out that practical non-identifiability can be basically due to two main causes [Brun et al., 2001, Raue et al., 2009]:

(a) compensation of the effects of a parameter by others (also known as collinearity): this can happen whenever some parameters have the same effect on the model response and, thus, their discrimination (i.e. the ability of estimating them separately) can be troublesome.

(b) lack of sensitivity of the measurements with respect to a parameter; this can occur, for instance, if the amplitude of the model input $\mathbf{f}$ is so small that the dependency of the measurements with respect to a parameter is basically negligible.

As regards case (a), it is possible to underline that any couple of parameters $\{\theta^i, \theta^j\}$ can be assumed to be a common cause of the measurement $\mathbf{y}$ if they are highly correlated, given the measurements $\mathbf{y}$. In other words, the higher the difference in information between the conditional joint pdf $p(\theta^i, \theta^j|\mathbf{y})$ and the product of the conditional marginal pdfs $p(\theta^i|\mathbf{y})p(\theta^j|\mathbf{y})$, the harder it is to estimate these parameters separately. As suggested in [Pant and Lombardi, 2015], a way to measure this occurrence is to employ the CMI $I(\theta^i; \theta^j|\boldsymbol{\mathcal{Y}})$ between any parameter couple $\{\theta^i, \theta^j\}$: the higher this is, the more correlated the parameters are, given the measurements, and therefore the less they are identifiable together. In other words, the CMI quantifies, after having observed $\mathbf{y}$, to which extent the knowledge of one parameter provides information about the other one.
It is interesting to underline that the II $I(\theta^i; \theta^j; \boldsymbol{\mathcal{Y}})$ could be employed to assess the extent of correlation between the parameters which may be attributed to the measurements or, in other words, if the correlation increases or decreases as the model response is measured. Since in all the applications here presented the parameters are supposed to be not correlated, before performing the measurements, it can be concluded that $p(\theta^i, \theta^j) = p(\theta^i)p(\theta^j)$ and, thus, the MI is $I(\theta^i; \theta^j) = 0$. Following Eq. (6.8), it can be concluded that $I(\theta^i; \theta^j; \boldsymbol{\mathcal{Y}}) = -I(\theta^i; \theta^j|\boldsymbol{\mathcal{Y}})$ and, therefore, the II yields the same results as the CMI.
Considering case (b), the MI $I(\theta^i; \boldsymbol{\mathcal{Y}})$ between each parameter and the measurements is exploited: if $I(\theta^i; \boldsymbol{\mathcal{Y}}) \to 0$, then the dependency of $\boldsymbol{\mathcal{Y}}$ to θ^i is low and, therefore, the measurements carry a low amount of information about parameter θ^i. In Section 4.2.2, the same quantity has been exploited in order to optimize the sensor locations, such that the information provided by the measurements is maximized with respect to the quantities to be estimated. While the latter approach allows to increase the practical identifiability of all the parameters, i.e. θ_i for $i = 1, ..., n_{\boldsymbol{\theta}}$, we are here seeking to assess the practical identifiability of each single parameter and the relations between them.
It is important to underline that, while the method proposed in [Beck and Katafygiotis, 1998] is valid only for linear structural models and for a large number of measurements, no assumptions are required for the application of MI and CMI.

6.2.3 Numerical approximations of the mutual information and the conditional mutual information

Since, for general cases, neither the MI $I(\theta^i; \mathcal{Y})$ nor the CMI $I(\Theta^i; \Theta^j|\mathcal{Y})$ can be computed analytically, respectively through Eqs. (6.1) and (6.6), a strategy to numerically evaluate them is now described.
In Section 4.3, two estimators of the MI have been introduced: the MC estimator [Huan and Marzouk, 2013] in Eq. (4.15) and the Kraskov estimator (based on the k-Nearest Neighbors estimator [Kraskov et al., 2004]) in Eq. (4.16). While many different methods can be exploited for the evaluation of the MI, the high computational complexity of the CMI prevents the extension of the same strategies for its evaluation. For this reason and in order to use comparable estimators for both the MI and the CMI, a Kernel Density Estimation (KDE) [Moon et al., 1995] based method will be here exploited.
Considering a generic random variable $\mathbf{Q} \in \mathbb{R}^{n_\mathbf{q}})$, with an associated pdf $p(\mathbf{q}$, the kernel density estimator $\hat{p}(q)$ is defined as:

$$\hat{p}(q) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{K}(\mathbf{q} - \mathbf{q}_i) \tag{6.10}$$

where N is the number of samples $\mathbf{q}_i$ to be drawn. $\mathcal{K}(\mathbf{q} - \mathbf{q}_i)$ is the multivariate Gaussian kernel, defined as:

$$\mathcal{K}(\mathbf{q} - \mathbf{q}_i) = (2\pi)^{-\frac{n_\mathbf{q}}{2}} |\mathbf{H}|^{-\frac{1}{2}} \exp^{-\frac{1}{2}(\mathbf{q}-\mathbf{q}_i)^\mathrm{T}\mathbf{H}^{-1}(\mathbf{q}-\mathbf{q}_i)} \tag{6.11}$$

where $\mathbf{H} \in \mathbb{R}^{n_\mathbf{q} \times n_\mathbf{q}}$ is called bandwidth matrix and it allows to control the smoothness of the estimator.
According to [Steuer et al., 2002], the MI $I(\mathbf{Q}, \mathbf{R})$ can therefore be estimated as:

$$\hat{I}(\mathbf{Q}, \mathbf{R}) = \frac{1}{N} \sum_{i=1}^{N} \ln \left[\frac{\hat{p}(\mathbf{q}_i, \mathbf{r}_i)}{\hat{p}(\mathbf{q}_i)\hat{p}(\mathbf{r}_i)} \right] \tag{6.12}$$

where $\hat{p}(\mathbf{q}_i, \mathbf{r}_i)$, $\hat{p}(\mathbf{q}_i)$ and $\hat{p}(\mathbf{r}_i)$ are computed according to Eq. (6.10). In order to reduce the allocated memory required for the computation, an ensemble estimator has been used, as suggested in [Kybic, 2004]. The N samples are thus divided into M groups and the MI is simply computed as:

$$\hat{I}(\mathbf{Q}; \mathbf{R}) = \frac{1}{N} \sum_{j=1}^{M} \hat{I}^j(\mathbf{Q}; \mathbf{R}) = \frac{1}{N} \sum_{i=j}^{M} \frac{M}{N} \left\{ \sum_{i=(j-1)N/M+1}^{jN/M} \ln \left[\frac{\hat{p}(\mathbf{q}_i, \mathbf{r}_i)}{\hat{p}(\mathbf{q}_i)\hat{p}(\mathbf{r}_i)} \right] \right\} \tag{6.13}$$

where $\hat{I}^i(X;Y)$ is the MI estimation related to the $i-$th group of samples. The same KDE-based approach can be used for the computation of the CMI. Using the definition in Eq. (6.6) and the marginalization rule, the CMI can be re-written as:

$$I(\mathbf{Q};\mathbf{R}|\mathbf{S}) = \int_{\mathbf{S}}\int_{\mathbf{Q}}\int_{\mathbf{R}} p(\mathbf{q},\mathbf{r},\mathbf{s})\ln\left[\frac{p(\mathbf{s})p(\mathbf{q},\mathbf{r},\mathbf{s})}{p(\mathbf{q},\mathbf{s})p(\mathbf{r},\mathbf{s})}\right] d\mathbf{q}d\mathbf{r}d\mathbf{s} \tag{6.14}$$

Thus, the KDE estimator is:

$$\hat{I}(\mathbf{Q};\mathbf{R}|\mathbf{S}) = \frac{1}{N}\sum_{i=1}^{N}\ln\left[\frac{\hat{p}(\mathbf{s})\hat{p}(\mathbf{q},\mathbf{r},\mathbf{s})}{\hat{p}(\mathbf{q},\mathbf{s})\hat{p}(\mathbf{r},\mathbf{s})}\right] \tag{6.15}$$

As previously highlighted, despite a faster convergence rate of the MC estimator with respect to the KDE estimator, the approach used in Section 4.3 for the optimal sensor placement, based on the MC estimator, would be practically unsuitable for the computation of the CMI, because of the high computational cost of the multi-dimensional numerical integration.

6.3 Numerical application

In this section, a non-linear structural example is chosen so that the practical non-identifiability causes described in Section 6.2.2 are clearly manifested and, hence, the use of the aforementioned information theory based indices can be validated.
The approach discussed in the previous section is now applied to a shear-type 8-storey building model (Fig. 6.1), which is a slightly modified version of the model provided in [De Callafon et al., 2008]. It is assumed that the flexural rigidity of all the horizontal members to be much higher than that of the column elements, so that the only relevant degrees of freedom are the horizontal displacements of each floor. According to this assumption, the floor stiffness at the $i-$th storey is $k = c\frac{12E_iI_i}{h^3}$, where c is the number of columns per floor, E_i is the material elastic modulus, I_i is the moment of inertia in the storey columns and h the floor height.

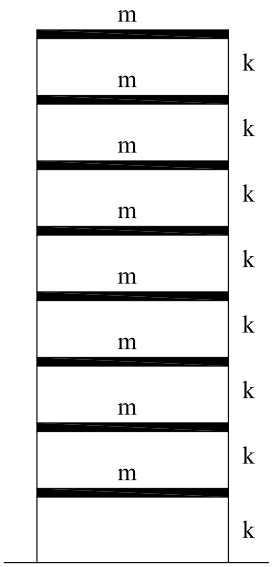


Figure 6.1. *Shear-type 8-storey building [De Callafon et al., 2008].*

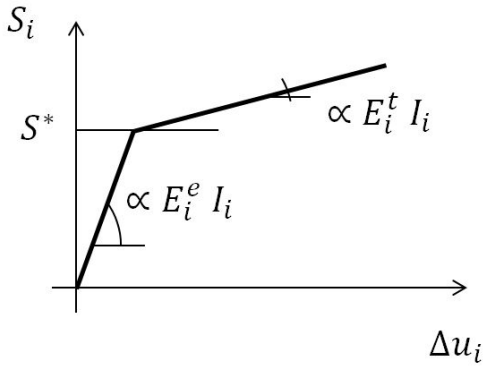


Figure 6.2. *Bi-linear relation between non-linear inter-storey drift and shear force, as defined in Eq. (6.16).*

The inter-story drifts Δu_i depend on the shear force S_i according to the following relation (Fig. 6.2):

$$\Delta u_i = \begin{cases} \frac{h^3}{c12E_i^e I_i} S_i & \text{if } S_i < S_i^* \\ \frac{h^3}{c12E_i^e I_i} S_i^* + \frac{h^3}{c12E_i^t I_i}(S_i - S_i^*) & \text{if } S_i \geq S_i^* \end{cases} \tag{6.16}$$

where a simple bi-linear rule for the modulus E_i is assumed:

$$E_i = \begin{cases} E_i^e & \text{if } S_i < S_i^* \\ E_i^t & \text{if } S_i \geq S_i^* \end{cases} \tag{6.17}$$

where E_i^e is the elastic modulus, $E_i^t < E_i^e$ is the tangent modulus and S_i^* is the shear force value for which the material behavior is no longer linear elastic.
The building is supposed to be subjected to a horizontal force applied at the top floor. For the sake of simplicity, only the displacement at the top floor u_8 is supposed to be measured and, in order to compare only a limited number of parameters, the aim is to study the practical identifiability of the parameters relevant to the first two floors, i.e. $\boldsymbol{\theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$. The parameters are assumed to be uniformly distributed, as detailed in Table 6.1.

Table 6.1. *Definition of parameters $\boldsymbol{\theta}$ and related prior pdf $p(\boldsymbol{\theta})$.*

Physical quantity	Prior pdf
E_1^e [GPa]	$\mathcal{U}(24, 36)$
E_1^t [GPa]	$\mathcal{U}(8, 12)$
I_1 [m^4]	$\mathcal{U}(0.018, 0.022)$
E_2^e [GPa]	$\mathcal{U}(24, 36)$
E_2^t [GPa]	$\mathcal{U}(8, 12)$
I_2 [m^4]	$\mathcal{U}(0.018, 0.022)$

First, the case in which $S_i < S_i^* \ \forall i$, i.e. the behavior of the material in each column remains in the linear elastic domain (see Fig. 6.2), is considered: from Eq. (6.16), it can be pointed out that the displacements and, hence, the measurement, do not depend on E_i^t. The practical non-identifiability of E_i^t shows up in the values of the MI $I(\Theta_i; \mathcal{Y})$ in Table 6.2: $I(E_1^t; \mathcal{Y})$ and $I(E_2^t; \mathcal{Y})$ are one order of magnitude lower than the other MI value. The MI is not exactly zero as one may expect from the definition in

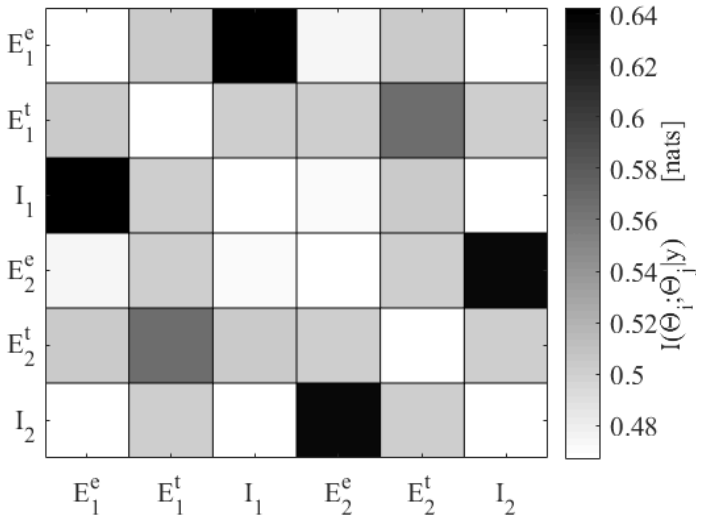
Table 6.2. *Mutual Information* $I(\Theta_i; \mathcal{Y})$ *of each parameter in* $\boldsymbol{\Theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$ *and the measured top-floor displacement* $\mathcal{Y}$*, considering the cases* $S_i < S_i^*$ *and* $S_i > S_i^*$*.*

	$I(E_1^e; \mathcal{Y})$	$I(E_1^t; \mathcal{Y})$	$I(I_1; \mathcal{Y})$	$I(E_2^e; \mathcal{Y})$	$I(E_2^t; \mathcal{Y})$	$I(I_2; \mathcal{Y})$
$S_i < S_i^*$	0.2280	0.0419	0.2256	0.2240	0.0427	0.2233
$S_i > S_i^*$	0.1040	0.1336	0.2238	0.1052	0.1353	0.2239

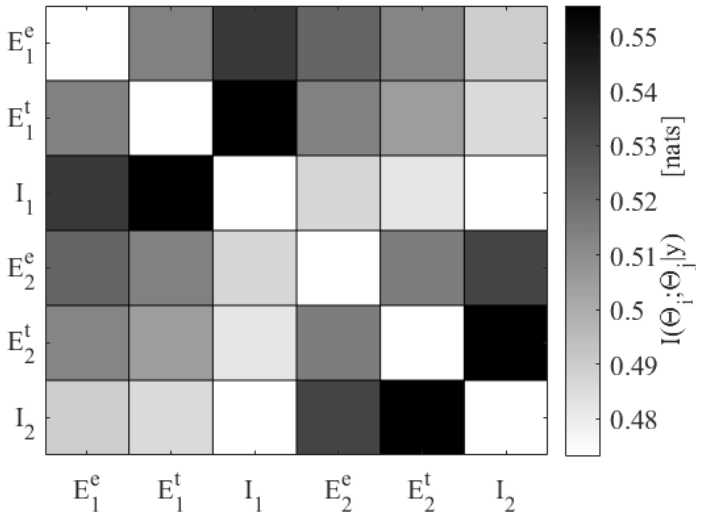
Eq. (6.1), because of the estimation error of the KDE method. The measured displacement depends only on the flexural stiffness $E_i^e I_i$: these two parameters cannot be estimated separately, since they offer a joint influence to the model response. This is highlighted by the CMI values $I(\Theta_i; \Theta_j|\mathcal{Y})$ reported in Figure 6.3a: as expected, the maximum values are reached for the couples $\{E_1^e, I_1\}$ and $\{E_2^e, I_2\}$.
On the other hand, if $S_i > S_i^* \ \forall i$, the non-linear mechanical behavior described in Eq. (6.16) affects the solution. As noted from Table 6.2, there are no parameters for which $I(\Theta_i; \mathcal{Y}) \simeq 0$. However, from Eq. (6.16), it can be pointed out that the model responses Δu_i depends with the same relationship on E_i^e / E_i^t and I_i, and therefore this prevents identifiability. The related CMI values stem from the latter fact: in Figure 6.3b, $I(\Theta_i; \Theta_j|\mathcal{Y})$ is maximum for the couples $\{E_1^e, I_1\}$ and $\{E_2^e, I_2\}$. Moreover, $I(E_{1,2}^t; I_{1,2}|\mathcal{Y}) > I(E_{1,2}^e; I_{1,2}|\mathcal{Y})$ as, since $E_i^t I_i < E_i^e I_i$, the resulting displacement is heavily dependent on E_i^t. The same fact can be underlined in Table 6.2, as $I(E_i^e; \mathcal{Y}) < I(E_i^t; \mathcal{Y})$.
From this simple example, it is, therefore, possible to check the validity of the proposed information theory based indexes to assess the practical non-identifiability, as its sources (lack of sensitivity and collinearity) can be easily recognized in the model. These measures provide a way to check the effectiveness of the monitoring system, in terms of practical identifiability, before performing the measurements.
Suitable remedies to cope with this occurrence can be:

- increase of the number of measurements $n_{\mathbf{y}}$;
- employment of accurate sensors (increase of σ);
- reduction of the number of parameters to be estimated $n_{\boldsymbol{\theta}}$ or change in their selection.

It should be highlighted that while the method here presented does not allow to choose the minimum number of measurements needed to guarantee identifiability, it allows to investigate the causes of practical non-identifiability.



(a)



(b)

Figure 6.3. *Conditional Mutual Information* $I(\Theta_i;\Theta_j|\mathcal{Y})$ *of each couple of parameters in* $\boldsymbol{\Theta} = [E_1^e; E_1^t; I_1; E_2^e; E_2^t; I_2]$ *and the measured top-floor displacement* $\mathcal{Y}$*, considering the cases (a)* $S_i < S_i^*$ *(b) and* $S_i > S_i^*$ *.*

6.4 Conclusions

In Chapters 2 and 4, two methods for optimally design a sensor network for SHM have been proposed: while these methods guarantee the maximization of the measurements information content and, hence, the accuracy of the estimates, they do not guarantee practical identifiability, i.e., that the estimation problem has a unique solution.
In the present chapter, the problem of practical identifiability in mechanical systems have been addressed through information theory based indexes (Section 6.2.1). The practical non-identifiability of parameters can arise due to either lack of sensitivity or compensations in the dependency of measurements on the parameters. Within an information theoretic approach, identifiability is here detected in terms of these two occurrences, using respectively the mutual information between each parameter and the measurements, and the conditional mutual information between each couple of parameters, conditioned on the measurements (Section 6.2.2).
In order to numerically evaluate both the mutual information and the conditional mutual information, two estimators based on the Kernel Density Estimation are introduced in Section 6.2.3. In Section 6.3, the methodology is applied to a non-linear mechanical model, namely a shear-type 8-storeys building model, where the former causes of non-identifiability are easily recognizable. It is shown that both the MI and the CMI allow to detect and quantify practical identifiability.
It should be highlighted that these methods do not allow to select a minimum number of sensors to be deployed in order to guarantee identifiability. Conversely, it is possible to investigate the causes which may lead to practical non-identifiability, which is critical for the design of an effective SHM sensor network.

CHAPTER 7

Conclusions

7.1 Summary of contributions

The main objective of the thesis lies in the development of new strategies for optimally design a Structural Health Monitoring (SHM) system, such that its effectiveness, i.e., the ability to estimate either damage, or the geometrical or mechanical parameters of a structural system, is maximized.
Two different approaches have been presented and evaluated: a deterministic Optimal Sensor Placement (OSP) method, which relies on the maximization of the sensitivity of the measurements, with respect to the damage parameters to be estimated, and a stochastic OSP strategy, based on the maximization of the expected information provided by the measurements.
The main contributions of this research work may be summarized as follows:

1. The deterministic OSP method has been introduced in Chapter 2. It relies on the main assumption of disregarding all the uncertainties related with the SHM process, i.e., the error associated with the measurements and the uncertainty related with the parameters to be estimated. The rationale of the method is the maximization of the sensitivity of the structural response with respect to the quantities to be estimated, e.g. the damage indices relating to a reduction of the

material stiffness. The optimal sensor configuration is obtained by adopting a topology optimization framework, which allows to select the most sensitive regions, in which the structural model is supposed to be discretized. In order to reduce the computational cost of the optimization algorithm, whenever large structural models have to be considered, a multi-scale optimization strategy has been introduced. In addition to the associated computational speed-up, it allows to account for the different scales of the problem, i.e., the characteristic size of the structure, the damage dimensions and the sensor sizes, while retaining the accuracy of the solution. Such an approach is particularly beneficial when micro-sensors have to be placed in structures characterized by complex geometries, as it permits to finely tune their positions. Thanks to the very low computational cost of the deterministic OSP approach, the optimal sensor spatial configuration can be obtained by considering that damage can occur anywhere in the structure.

2. Two applications of the deterministic OSP method have been shown in Chapter 3. First, the single-scale strategy has been adopted in a benchmark structural model, i.e., a clamped square plate, subjected to a vertical force applied at its center. Assuming to measure the rotations about the in-plane axes, the resulting optimal sensor configuration reflects the geometric conditions of the structure and, therefore, turns out to be symmetric. It can be underlined that, due to the particular formulation of the topology optimization problem, some post-processing techniques, such as the filtering of the objective function, is needed, for certain sensors spatial configurations, in order to obtain physically sound solutions. It can moreover be demonstrated that the optimal sensor locations depend on the damage scenarios which are assumed to occur: thus, if a robust OSP has to be achieved, all the optimal solutions, associated to different damage cases, should be considered. The multi-scale OSP method is applied to a real-size structural model, i.e., a stiffened composite fuselage section, which is supposed to undergo two different damage scenarios: a part-trough damage, simulating aging, and delamination. The optimal sensor locations turn out to be highly affected by the position of the external load, as it drives the zones where the maximum variations of response occur. It is important to underline that the discretization mesh at each length-scale allows to appropriately tune both the resolution of the damage to be estimated and the position of the sensors. Moreover, it can be concluded

that the computational speed-up obtained by adopting the multi-scale strategy, as compared to the single-scale one, is approximately 50, 000 in the plate benchmark structure and 200 in the fuselage case.

3. In order to take into account the uncertainties inherent in the measurement process, i.e., the prediction error (due to the model and measurement errors) and the uncertainties related with the quantities to be estimated, in Chapter 4 a stochastic OSP method has been proposed. Within the Bayesian framework, the sensors spatial configuration can be optimized by maximizing the expected Shannon information gain, which quantifies the amount of information provided by the measurements, in order to estimate a set of mechanical parameters of the structural model. The mentioned objective function is basically a measure of the increase in information between the prior and the posterior pdfs of the parameters, i.e. respectively associated to their knowledge before and after having performed the measurements. Two numerical estimators of the expected information gain have been discussed, i.e. a Monte Carlo (MC) based estimator and a k-Nearest Neighbors (kNN) based estimator. While the second one guarantees a lower computational cost with respect to the first one, the MC estimator allows to obtain a less noisy objective function, as the knowledge of the likelihood function is naturally taken into account. In order to numerically evaluate the MC estimator, which relies on the repeated computation of the structural model response, the use of surrogate models, based on Polynomial Chaos Expansion (PCE), allows to mimic the input-output relations and efficiently reduce the overall computational cost. Two different strategies have been proposed: one assumes that the model inputs do take into account both the parameters to be estimated and the design variable, which defines the sensor positions (joint input PCE); the second one is based on the synergy between the PCE and the Principal Component Analysis (PCA), which is a model order reduction technique. Although in the first formulation the computation of only one PCE surrogate is needed, the application of the joint input PCE based surrogate becomes troublesome if numerical structural models with large number of nodes or multiple physical quantities to be measured have to be taken into account. Conversely, the PCA-PCE approach allows to handle such problems, as the design variable is not part of the surrogate input and, therefore, its variability has not to be considered.
   Since the numerically computed objective function is noisy, the adop-

tion of standard optimization algorithms could lead to false optima, i.e., local optima which are only due to the numerical noise. To this end, the adoption of a stochastic optimization algorithm, namely the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), proves to be an efficient way to overcome the mentioned problem.
The main benefit of the stochastic approach, based on the Bayesian experimental design, is that the expected Bayesian information gain depends on both the sensor positions, the measurement noise (and hence the type of sensor) and the number of sensors. Thus, by accounting for all these experimental settings, it is possible to comprehensively optimally design the SHM monitoring system. Moreover, following a cost-benefit approach, also the cost of the SHM system can be taken into account by maximizing the ratio between the expected information gain and the cost model and, thus, efficiently allocating the economic resources.

4. In Chapter 5 two structural models have been considered in order to assess the stochastic OSP method. First, the strategy is applied to a benchmark structural case, i.e. the same clamped square plate earlier introduced in Chapter 3. Two cases have been considered: a damage located in a known region of the structure, or a damage which can occur anywhere. It has been shown that, as in the deterministic case, the resulting optimal spatial configurations depend both on the loading conditions and on the assumed damage scenarios. The effectiveness of the CMA-ES is assessed by computing the values of the objective function on a fine mesh, resulting in about 25 objective functions evaluations. Moreover, it has been shown that, as the measurement error increases, the resulting objective function presents several local optima due to the numerical approximation introduced by the estimator. Thus, it has been shown that, despite the resulting optimal sensor placement is equal to the one obtained through the deterministic OSP method, the stochastic method allows to consider additional experimental settings in the optimal design, i.e., the prediction error and the prior pdf of the parameters.
In order to compare the joint input PCE and the PCA-PCE strategies, a real-size structural model, i.e., the Pirelli tower in Milan, has been considered, where 6 model parameters (both mechanical and geometrical) have been supposed to be estimated. It has been demonstrated that the first approach requires an unbearable number of samples in order an accurate surrogate model to be built, making therefore its

employment useless for large structural models. Moreover, the computational cost associated to the PCE-PCE turns out to be 5-6 times lower than the one related with the joint input PCE, guaranteeing at the same time a high accuracy in the evaluation of the structural response. Regarding the numerical computation of the expected Shannon information gain, even if the kNN estimator is characterized by a lower computational cost than the MC estimator (approximately 10 times faster), its applicability is prevented due to the higher numerical noise which is introduced in the objective function.
As concerns the optimal sensor spatial configuration, the information-based approach proves to be dependent on the loading conditions and on the choice of the parameters to be estimated. It should be highlighted that, although the CMA-ES allows to handle the objective function noise which is due to the numerical approximation, multiple optimal solutions, arising from the presence of several optima, can be detected by running the algorithm repeatedly with different initial conditions.
Then, it has been shown that the expected Shannon information gain, associated with the optimal solution, increases as more accurate sensors are employed (i.e. with decreasing measurement noise) and as more sensors are placed, since more information is provided by the SHM system. Moreover, it has been demonstrated that, fixed the type of sensors, the additional information provided by each sensor added to the network, gets progressively lower.
In order to comprehensively design the sensor network, both in terms of sensor locations, types and number, the Pareto frontiers of the multi-dimensional optimization problem allow to maximize the information provided by the sensor network and minimize the overall cost of the monitoring system.
It should be underlined that, while the deterministic OSP method allows to take into account damages potentially located anywhere in the structure, the stochastic OSP method can only be applied when a limited number of damage parameters have to be estimated, due to the high computational cost needed for taking into account the inherent variability of the parameters.

5. Despite the optimization of the sensor network allows to maximize the information provided by the measurements and, hence, the ability of estimating the parameters, its effectiveness can be prevented whenever the parameters result to be practically non-identifiable. For this

reason, two measures to assess and quantify the sources of practical non-identifiability of the parameters have been proposed in Chapter 6. By considering a simple non-linear structural case, it has been shown that the mutual information between each parameter and the measurements allows to quantify the lack of sensitivity of the data with respect to the parameters, while the conditional mutual information provides an indication on possible common effects of the parameters on the measured data, which can prevent them from being estimated separately.

To summarize, it can be remarked that the deterministic OSP method allows to efficiently provide a physically sound optimal sensor spatial configuration and, thanks to the low computational cost, a damage located anywhere in the structure can be taken into account. On the other hand, the stochastic OSP allows to design the SHM monitoring system in a more comprehensive way, i.e. determining both the location, the number and the types of sensors, and taking into account the prior knowledge about the parameters' distributions; nevertheless, only a limited number of parameters, hence damage locations, can be considered.

7.2 Limitations

In Section 7.1, the main contributions introduced in the present thesis are highlighted. The main limitations of the methods presented in the thesis are now highlighted:

- Regarding the deterministic OSP method, the main limitation of such an approach regards the linearity of the structural problems which have to be considered, since the employed topology optimization framework can be applied to linear problems only. By non-linearity, we refer here to models where the model response varies non-linearly with respect to a variation of the external load, e.g. geometric non-linearities; the non-linearity related to the damage, defined as a variation of stiffness, is instead taken into account.
- Regarding the stochastic approach, its applicability is mainly ruled by the number $n_{\boldsymbol{\theta}}$ of parameters which are supposed to be estimated through the sensor network. Due to the curse of dimensionality of the surrogate models and to the high computational cost of the MC estimator, it cannot be applied when the parameters are of the order of tens, or whenever the location of the damage is unknown and several parameters have to be taken into account.

- Another main limitation of the stochastic method concerns the locations of the sensors to be placed: the optimal solutions in terms of spatial configuration can only be obtained among the nodes of the discretization mesh; therefore, the resolution of the optimal sensor placement cannot be freely set, but it is ruled by the model mesh.
- As regards the approach for the assessment of the practical non-identifiability of the parameters, a main limitation is that it does not provide a yes/no binary outcome, i.e., whether the parameters are identifiable or not; thus, it is not possible to select a minimum number of sensors to be deployed in order to guarantee practical identifiability.

7.3 Suggestions for future research

Based on the presented work, some ideas for future research on this topic are here summarized:

- The applicability of the deterministic OSP method has been here limited to linear structural models; its extension to non-linear cases could be taken into account by applying suitable topology optimization algorithms [Maute et al., 1998, Jung and Gea, 2004].
- The results provided by the two OSP methods in terms of sensor locations could be experimentally validated both on a simple benchmark structure (e.g. the clamped square plate here used) and on a real-size civil structure. Different sensor configurations could be considered and their effectiveness could be checked by estimating the structural parameters and the associated uncertainties through, e.g., Bayesian model updating.
- The proposed stochastic OSP method is valid only for cases where the inertial effects can be disregarded. The extension of the stochastic OSP method to structural dynamics could be considered by, e.g., using appropriate surrogate models [Mai et al., 2016].
- The applicability of the stochastic OSP method is limited by the number of parameters, due to the curse of dimensionality: high number of parameters or damage located anywhere in the structure could be considered by employing the stochastic finite element method [Der Kiureghian and Ke, 1988, Sudret and Der Kiureghian, 2000], where the spatial variability of the material properties can be taken into account through random fields [Chowdhury and Adhikari, 2010].

- The stochastic OSP method could be extended by taking into account the robustness of sensor networks [Hoblos et al., 2000, Kakamanshadi et al., 2015], in order to obtain efficient fault tolerant optimal SHM designs.
- A method for determining the minimal instrumentation, i.e., the minimum number of sensors to be deployed for guaranteeing identifiability of parameters, could be studied by combining information theory and existing methods [Mukhopadhyay et al., 2013, Mukhopadhyay et al., 2014].

APPENDIX $\mathcal{A}$

Expected information gain and mutual information

In this Appendix, some details of the relationship between the expected gain in Shannon information $U(\mathbf{d})$, as defined in Eq. (4.11), and the mutual information, defined in Eq. (6.1), are summarized. Further details can be found in [Cover and Thomas, 2012].
The expected information gain is defined as follows:

$$\begin{aligned} U(\mathbf{d}) &= \int_{\mathcal{Y}} \int_{\Theta} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) \ln \frac{p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})}{p(\boldsymbol{\theta})} p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = \\ &= \int_{\mathcal{Y}} \int_{\Theta} \{\ln[p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})]\} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \\ &\quad - \int_{\mathcal{Y}} \int_{\Theta} \{\ln[p(\boldsymbol{\theta})]\} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \end{aligned} \tag{A.1}$$

where, as $p(\mathbf{y}|\mathbf{d})$ is independent of $\boldsymbol{\theta}$, the first term of the last representation of $U(\mathbf{d})$ may be re-written as:

$$\begin{aligned}\int_{\boldsymbol{\mathcal{Y}}} p(\mathbf{y}|\mathbf{d}) \left\{ \int_{\boldsymbol{\Theta}} \{\ln[p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})]\} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) d\boldsymbol{\theta} \right\} d\mathbf{y} = \\ = -\int_{\boldsymbol{\mathcal{Y}}} p(\bar{\mathbf{y}}|\mathbf{d}) H(\boldsymbol{\theta}|\mathbf{y}=\bar{\mathbf{y}},\mathbf{d}) d\mathbf{y} = \\ = -H(\boldsymbol{\theta}|\boldsymbol{\mathcal{Y}},\mathcal{D})\end{aligned} \tag{A.2}$$

$H(\boldsymbol{\theta}|\boldsymbol{\mathcal{Y}},\mathcal{D})$ being the conditional entropy, i.e., the Shannon entropy of the conditional probability $p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})$.
Recalling that $p(\boldsymbol{\theta},\mathbf{y}|\mathbf{d}) = p(\mathbf{y}|\mathbf{d})p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d})$, and that $p(\boldsymbol{\theta}|,\mathbf{d}) = \int_{\boldsymbol{\mathcal{Y}}} p(\boldsymbol{\theta},\mathbf{y}|\mathbf{d})d\mathbf{y}$, the second term in Eq. (A.1) may be re-stated as:

$$\begin{aligned}\int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} \{\ln[p(\boldsymbol{\theta})]\} p(\boldsymbol{\theta}|\mathbf{y},\mathbf{d}) p(\mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = \\ = -\int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} \{\ln[p(\boldsymbol{\theta})]\} p(\boldsymbol{\theta}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = \\ = -H(\boldsymbol{\theta}|\mathcal{D})\end{aligned} \tag{A.3}$$

$H(\boldsymbol{\theta}|\mathcal{D})$ being the Shannon entropy of the pdf $p(\boldsymbol{\theta}|\mathbf{d})$. We then arrive at:

$$U(\mathbf{d}) = H(\boldsymbol{\theta}|\mathcal{D}) - H(\boldsymbol{\theta}|\boldsymbol{\mathcal{Y}},\mathcal{D}) \tag{A.4}$$

i.e., the expected information gain is equal to the difference in Shannon entropy between the prior and the posterior pdfs, which is therefore equal to the MI:

$$I(\boldsymbol{\Theta},\boldsymbol{\mathcal{Y}}) = \int_{\boldsymbol{\mathcal{Y}}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\theta},\mathbf{y}|\mathbf{d}) \ln \frac{p(\boldsymbol{\theta},\mathbf{y}|\mathbf{d})}{p(\boldsymbol{\theta})p(\mathbf{y}|\mathbf{d})} d\boldsymbol{\theta} d\mathbf{y} \tag{A.5}$$

APPENDIX B

PCE polynomials bases

The polynomial bases employed in the PCE models are here described. The multivariate polynomials $\mathbf{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\chi})$ introduced in Eq. (4.24) are computed as the tensor product of the related univariate polynomials $\psi^{(i)}_{\alpha_i}(\chi_i)$ [Marelli and Sudret, 2014]:

$$\mathbf{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\chi}) = \prod_{i=1}^{M} \psi^{(i)}_{\alpha_i}(\chi_i) \tag{B.1}$$

In Table B.1, the univariate polynomial families, which are orthonormal to the most common input probability distributions, are listed [Xiu and Karniadakis, 2002, Sudret, 2007].

Table B.1. *List of polynomial functions commonly used in PCE, for each type of pdf.*

Type of pdf	Orthonormal polynomials	
Uniform	Legendre	$\psi_p^{(i)}(\chi_i) = \frac{1}{2^p p!}\frac{d^p}{d\chi_i^p}[(\chi_i - 1)^p]$
Gaussian	Hermite	$\psi_p^{(i)}(\chi_i) = (-1)^p e^{\frac{x^2}{2}}\frac{d^p}{d\chi^p}e^{-\frac{x^2}{2}}$
Gamma	Laguerre	$\psi_p^{(i)}(\chi_i) = \sum_{i=1}^{p}\frac{(-1)^i}{i!}\binom{p}{i}\chi^i$
Beta	Jacobi	$\psi_p^{(i)}(\chi_i) = \frac{(-1)^p}{2^p p!}(1-\chi)^{-\alpha}(1+\chi)^{-\beta}$ $\frac{d^p}{d\chi^p}\left\{(1-\chi)^{\alpha}(1+\chi)^{\beta}\left(1-\chi^2\right)^p\right\}$

Bibliography

[Abdelgawad and Yelamarthi, 2016] Abdelgawad, A. and Yelamarthi, K. (2016). Structural health monitoring: internet of things application. In *Circuits and Systems (MWSCAS), 2016 IEEE 59th International Midwest Symposium on*, pages 1–4. IEEE.

[Abdelgawad and Yelamarthi, 2017] Abdelgawad, A. and Yelamarthi, K. (2017). Internet of Things (IoT) platform for structure health monitoring. *Wireless Communications and Mobile Computing*, 2017.

[Altman, 1992] Altman, N. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

[Annamdas et al., 2017] Annamdas, V. G. M., Bhalla, S., and Soh, C. K. (2017). Applications of structural health monitoring technology in Asia. *Structural Health Monitoring*, 16(3):324–346.

[Arcadius et al., 2017] Arcadius, T. C., Gao, B., Tian, G., and Yan, Y. (2017). Structural health monitoring framework based on internet of things: a survey. *IEEE Internet of Things Journal*, 4(3):619 – 635.

[Asgarieh et al., 2014] Asgarieh, E., Moaveni, B., and Stavridis, A. (2014). Nonlinear finite element model updating of an infilled frame based on identified time-varying modal parameters during an earthquake. *Journal of Sound and Vibration*, 333(23):6057–6073.

[Association, 2000] Association, T. A. (2000). Aluminum standards and data.

[ASTM International, 2014] ASTM International (2014). *ASTM Standard B209, Standard Specification for Aluminum and Aluminum-Alloy Sheet and Plate*.

[Balageas et al., 2006] Balageas, D., Fritzen, C. P., and Güemes, A. (2006). *Structural Health Monitoring*. Wiley-ISTE.

[Barbella, 2009] Barbella, G. (2009). *Frequency domain analysis of slender structural systems under turbulent wind excitation*. PhD thesis, Politecnico di Milano, Milano.

[Barbella et al., 2011] Barbella, G., Perotti, F., and Simoncini, V. (2011). Block Krylov subspace methods for the computation of structural response to turbulent wind. *Computer Methods in Applied Mechanics and Engineering*, 200(23):2067–2082.

[Barbero and Damiani, 2003] Barbero, E. J. and Damiani, T. M. (2003). Phenomenological prediction of tensile strength of E-glass composites from available aging and stress corrosion data. *Journal of reinforced plastics and composites*, 22(4):373–394.

[Basto et al., 2017] Basto, C., Pelà, L., and Chacón, R. (2017). Open-source digital technologies for low-cost monitoring of historical constructions. *Journal of Cultural Heritage*, 25:31–40.

[Bayer et al., 2009] Bayer, P., Duran, E., Baumann, R., and Finkel, M. (2009). Optimized groundwater drawdown in a subsiding urban mining area. *Journal of Hydrology*, 365(1):95–104.

[Beck and Au, 2002] Beck, J. and Au, S. (2002). Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *Journal of Engineering Mechanics*, 128(4):380–391.

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

[Beck and Yuen, 2004] Beck, J. and Yuen, K.-V. (2004). Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics*, 130(2):192–203.

[Bellman and Åström, 1970] Bellman, R. and Åström, K. J. (1970). On structural identifiability. *Mathematical biosciences*, 7(3-4):329–339.

[Bendsøe, 1989] Bendsøe, M. (1989). Optimal shape design as a material distribution problem. *Structural and multidisciplinary optimization*, 1(4):193–202.

[Bendsøe and Sigmund, 2003] Bendsøe, M. and Sigmund, O. (2003). *Topology optimization: theory, methods and applications*. Springer.

[Bendsøe and Sigmund, 2004] Bendsøe, M. and Sigmund, O. (2004). *Topology optimization by distribution of isotropic material*. Springer.

[Bendsøe and Kikuchi, 1988] Bendsøe, M. P. and Kikuchi, N. (1988). Generating optimal topologies in structural design using a homogenization method. *Computer methods in applied mechanics and engineering*, 71(2):197–224.

[Bernardo, 1979] Bernardo, J. M. (1979). Expected information as expected utility. *The Annals of Statistics*, 7(3):686–690.

[Berveiller et al., 2006] Berveiller, M., Sudret, B., and Lemaire, M. (2006). Stochastic finite element: a non intrusive approach by regression. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15(1-3):81–92.

[Bimpas et al., 2011] Bimpas, M., Trapani, D., Zonta, D., Amditis, A., Frondistou-Yannas, S., Kalidromitis, V., Stratakos, Y., and Gkaretsos, A. (2011). MEMSCON project: wireless sensor network for post-earthquake evaluation of concrete buildings. *Proceedings of the 5th International Conference on Structural Health Monitoring of Intelligent Infrastructure*.

[Blatman, 2009] Blatman, G. (2009). *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Blaise Pascal University, Clermont-Ferrand.

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

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

[Bleistein and Handelsman, 1975] Bleistein, N. and Handelsman, R. (1975). *Asymptotic expansions of integrals*. Dover Publications.

[Bliss et al., 2014] Bliss, C., Frank, M., Danforth, C., and Dodds, P. (2014). An evolutionary algorithm approach to link prediction in dynamic social networks. *Journal of Computational Science*, 5(5):750–764.

[Bresciani, 2013] Bresciani, L. (2013). Analytical and numerical modelling of ballistic impacts on composite and ceramic composite targets. Master's thesis, Politecnico di Milano, Milano.

[Brownjohn and Moyo, 2001] Brownjohn, J. and Moyo, P. (2001). Monitoring of Singapore-Malaysia second link during construction. In *Second international conference on experimental mechanics*, volume 4317, pages 528–534. International Society for Optics and Photonics.

[Bruggi and Mariani, 2013] Bruggi, M. and Mariani, S. (2013). Optimization of sensor placement to detect damage in flexible plates. *Engineering Optimization*, 45:659–676.

[Bruggi and Venini, 2008] Bruggi, M. and Venini, P. (2008). Eigenvalue-based optimization of incompressible media using mixed finite elements with application to isolation devices. *Computer Methods in Applied Mechanics and Engineering*, 197(13-16):1262–1279.

[Brun et al., 2001] Brun, R., Reichert, P., and Künsch, H. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37(4):1015–1030.

[Bruns and Tortorelli, 2001] Bruns, T. and Tortorelli, D. (2001). Topology optimization of non-linear elastic structures and compliant mechanisms. *Computer Methods in Applied Mechanics and Engineering*, 190(26-27):3443–3459.

[Buildings Performance Institute Europe, 2011] Buildings Performance Institute Europe, B. (2011). Europes's buildings under the microscope. Technical report.

[Caimmi et al., 2014] Caimmi, F., Mariani, S., De Fazio, M., and Bendiscioli, P. (2014). Investigation of the effectiveness and robustness of a MEMS-based structural health monitoring system for composite laminates. *IEEE Sensors Journal*, 14:2208–2215.

[Capellari et al., 2016a] Capellari, G., Caimmi, F., Bruggi, M., and Mariani, S. (2016a). A multiscale approach to the smart deployment of micro-sensors over flexible plates. In *Multidisciplinary Digital Publishing Institute Proceedings*, volume 1, page 40.

[Capellari et al., 2017a] Capellari, G., Caimmi, F., Bruggi, M., and Mariani, S. (2017a). A multiscale approach to the smart deployment of micro-sensors over lightweight structures. *Sensors*, 17(7):1632.

[Capellari et al., 2016b] Capellari, G., Chatzi, E., and Mariani, S. (2016b). An optimal sensor placement method for SHM based on Bayesian experimental design and Polynomial Chaos Expansion. *Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering*, 3:6272–6282.

[Capellari et al., 2016c] Capellari, G., Chatzi, E., and Mariani, S. (2016c). Optimal sensor placement through bayesian experimental design: effect of measurement noise and number of sensors. In *Multidisciplinary Digital Publishing Institute Proceedings*, volume 1, page 41.

[Capellari et al., 2017b] Capellari, G., Chatzi, E., and Mariani, S. (2017b). Optimal design of sensor networks for damage detection. *Procedia Engineering*, 199:1864–1869.

[Capellari et al., 2017c] Capellari, G., Chatzi, E., and Mariani, S. (2017c). Parameter identifiability through information theory. *Proceedings of the 2nd ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering, UNCECOMP*.

[Capellari et al., 2018] Capellari, G., Chatzi, E., and Stefano, M. (2018). Cost-benefit optimization of sensor networks for SHM applications. *Proceedings*, 2(3):132.

[Capellari et al., 2014] Capellari, G., Eftekhar Azam, S., and Mariani, S. (2014). Hybrid reduced-order modeling and particle-Kalman filtering for the health monitoring of flexible structures. *International Electronic Conference on Sensors and Applications*.

[Capellari et al., 2015a] Capellari, G., Eftekhar Azam, S., and Mariani, S. (2015a). Online damage detection in plates via vibration measurements. In *Model Validation and Uncertainty Quantification, Volume 3*, pages 85–91. Springer.

[Capellari et al., 2015b] Capellari, G., Eftekhar Azam, S., and Mariani, S. (2015b). Towards real-time health monitoring of structural systems via recursive Bayesian filtering and reduced order modelling. *International Journal of Sustainable Materials and Structural Systems*, 2(1-2):27–51.

[Capellari et al., 2016d] Capellari, G., Eftekhar Azam, S., and Mariani, S. (2016d). Damage detection in flexible plates through reduced-order modeling and hybrid particle-Kalman filtering. *Sensors*, 16:2.

[Cavazzuti, 2013] Cavazzuti, M. (2013). Deterministic optimization. In *Optimization Methods*, pages 77–102. Springer.

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

[Chang, 2014] Chang, G. (2014). On Kalman filter for linear system with colored measurement noise. *Journal of Geodesy*, 88(12):1163–1170.

[Chatzi et al., 2010] Chatzi, E., Smyth, A., and Masri, S. (2010). Experimental application of on-line parametric identication for nonlinear hysteretic systems with model uncertainty. *Structural Safety*, 32:326–337.

[Chatzis et al., 2014] Chatzis, M., Chatzi, E., and Smyth, A. (2014). On the observability and identifiability of nonlinear structural and mechanical systems. *Structural Control and Health Monitoring*, 22(3):574–593.

[Chisari et al., 2017] Chisari, C., Macorini, L., Amadio, C., and Izzuddin, B. (2017). Optimal sensor placement for structural parameter identification. *Structural and Multidisciplinary Optimization*, 55(2):647–662.

[Cho et al., 2008] Cho, S., Yun, C., Lynch, J., Zimmerman, A., Spencer, B., and Nagayama, T. (2008). Smart wireless sensor technology for structural health monitoring of civil structures. *Steel Structures*, 8:267–275.

[Chowdhury and Adhikari, 2010] Chowdhury, R. and Adhikari, S. (2010). High dimensional model representation for stochastic finite element analysis. *Applied Mathematical Modelling*, 34(12):3917–3932.

[Christensen and Klarbring, 2008] Christensen, P. and Klarbring, A. (2008). *An introduction to structural optimization*, volume 153. Springer.

[Corigliano et al., 2013] Corigliano, A., Dossi, M., and Mariani, S. (2013). Domain decomposition and model order reduction methods applied to the simulation of multiphysics problems in MEMS. *Computers & Structures*, 122:113–127.

[Corigliano et al., 2015] Corigliano, A., Dossi, M., and Mariani, S. (2015). Model order reduction and domain decomposition strategies for the solution of the dynamic elastic–plastic structural problem. *Computer Methods in Applied Mechanics and Engineering*, 290:127–155.

[Corigliano and Mariani, 2004] Corigliano, A. and Mariani, S. (2004). Parameter identification in explicit structural dynamics: performance of the extended Kalman filter. *Computer Methods in Applied Mechanics and Engineering*, 193(36):3807–3835.

[Costa et al., 2013] Costa, A., Guedes, J., and Varum, H. (2013). *Structural rehabilitation of old buildings*. Springer.

[Cover and Thomas, 2012] Cover, T. and Thomas, J. (2012). *Elements of information theory*. John Wiley & Sons.

[De Callafon et al., 2008] De Callafon, R., Moaveni, B., Conte, J., He, X., and Udd, E. (2008). General realization algorithm for modal identification of linear dynamic systems. *Journal of Engineering Mechanics*, 134(9):712–722.

[De Groot, 1962] De Groot, M. (1962). Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419.

[Der Kiureghian and Ke, 1988] Der Kiureghian, A. and Ke, J.-B. (1988). The stochastic finite element method in structural reliability. *Probabilistic Engineering Mechanics*, 3(2):83–91.

[Diop and Fliess, 1991] Diop, S. and Fliess, M. (1991). Nonlinear observability, identifiability, and persistent trajectories. *Proceedings of the 30th IEEE Conference on Decision and Control*, pages 714–719.

[Doebling, 1995] Doebling, S. (1995). *Measurement of structural flexibility matrices for experiments with incomplete reciprocity*. PhD thesis, University of Colorado, Boulder.

[Eftekhar Azam, 2012] Eftekhar Azam, S. (2012). *Dual estimation and reduced order modelling of damaging structures*. PhD thesis, Politecnico di Milano, Milano.

[Eftekhar Azam, 2014] Eftekhar Azam, S. (2014). *Online Damage Detection in Structural Systems*. Springer.

[Eftekhar Azam et al., 2012] Eftekhar Azam, S., Bagherinia, M., and Mariani, S. (2012). Stochastic system identification via particle and sigma-point Kalman filtering. *Scientia Iranica*, 19:982–991.

[Eftekhar Azam and Mariani, 2018] Eftekhar Azam, S. and Mariani, S. (2018). Online damage detection in structural systems via dynamic inverse analysis: a recursive Bayesian approach. *Engineering Structures*, 159:28–45.

[European Commission, 2014] European Commission (2014). Infrastructure in the EU: developments and impact on growth. Technical report, Directorate-General for Economic and Financial Affairs.

[Fan and Qiao, 2011] Fan, W. and Qiao, P. (2011). Vibration-based damage identification methods: a review and comparative study. *Structural Health Monitoring*, 10:83–110.

[Farmaga et al., 2011] Farmaga, I., Shmigelskyi, P., Spiewak, P., and Ciupinski, L. (2011). Evaluation of computational complexity of finite element analysis. In *CAD Systems in Microelectronics (CADSM), 2011 11th International Conference The Experience of Designing and Application of*, pages 213–214. IEEE.

[Farrar and Doebling, 1997] Farrar, C. and Doebling, S. (1997). An overview of modal-based damage identification methods. *Structural Damage Assessment Using Advanced Signal Processing Procedures*.

[Farrar and Jauregui, 1998] Farrar, C. and Jauregui, D. (1998). Comparative study of damage identication algorithms applied to a bridge: II. Numerical study. *Smart Materials and Structures*, 7:720–721.

[Farrar and Worden, 2007] Farrar, C. and Worden, K. (2007). An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A*, 365:303–315.

[Fedorov and Hackl, 1994] Fedorov, V. and Hackl, P. (1994). Optimal experimental design: spatial sampling. *Calcutta Statistical Association Bulletin*, 44(1-2):57–82.

[Ferrari et al., 2016] Ferrari, R., Pioldi, F., Rizzi, E., Gentile, C., Chatzi, E., Serantoni, E., and Wieser, A. (2016). Fusion of wireless and non-contact technologies for the dynamic testing of a historic rc bridge. *Measurement Science and Technology*, 27(12):124014.

[Fix and Hodges Jr, 1951] Fix, E. and Hodges Jr, J. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, California Univ Berkeley.

[Fleury, 1979] Fleury, C. (1979). Structural weight optimization by dual methods of convex programming. *International Journal for Numerical Methods in Engineering*, 14(12):1761–1783.

[Furuta et al., 2014] Furuta, H., Frangopol, D., and Akiyama, M. (2014). *Life-Cycle of Structural Systems: Design, Assessment, Maintenance and Management*. CRC Press.

[Geubelle and Baylor, 1998] Geubelle, P. and Baylor, J. (1998). Impact-induced delamination of composites: a 2D simulation. *Composites Part B: Engineering*, 29(5):589–602.

[Ghanem and Spanos, 1990] Ghanem, R. and Spanos, P. (1990). Polynomial chaos in stochastic finite elements. *Journal of Applied Mechanics*, 57(1):197–202.

[Ghanem and Spanos, 2003] Ghanem, R. and Spanos, P. (2003). *Stochastic finite elements: a spectral approach*. Courier Corporation.

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

[Girolami et al., 2017] Girolami, A., Brunelli, D., Benini, L., et al. (2017). Low-cost and distributed health monitoring system for critical buildings. *Proceedings of the 2017 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems*, pages 1–6.

[Glad and Ljung, 1990] Glad, S. and Ljung, L. (1990). Model structure identifiability and persistence of excitation. In *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 3236–3240. IEEE.

[Godfrey and DiStefano III, 1987] Godfrey, K. and DiStefano III, J. (1987). Identifiability of model parameters. *Identifiability of parametric models*, 1:1–20.

[Gossen, 1983] Gossen, H. (1983). *The laws of human relations and the rules of human action derived therefrom*. Mit Press.

[Green and Worden, 2014] Green, P. and Worden, K. (2014). Bayesian system identification of nonlinear systems: Informative training data through experimental design. In *Proceedings of IMAC XXXII, Conference and Exposition on Structural Dynamics*. Sheffield.

[Guida and Marulo, 2014] Guida, M. and Marulo, F. (2014). Partial modeling of aircraft fuselage during an emergency crash landing. *Procedia Engineering*, 88:26–33.

[Hadjidoukas et al., 2015] Hadjidoukas, P., Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P. (2015). π4U: A high performance computing framework for Bayesian uncertainty quantification of complex models. *Journal of Computational Physics*, 284:1–21.

[Hansen, 2006] Hansen, N. (2006). The CMA evolution strategy: a comparing review. In *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, pages 75–102. Springer.

[Hansen, 2016] Hansen, N. (2016). The CMA evolution strategy: a tutorial. *arXiv*.

[Hansen et al., 2010] Hansen, N., Auger, A., Ros, R., Finck, S., and Pošík, P. (2010). Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009. In *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, pages 1689–1696. ACM.

[Hansen and Ostermeier, 2001] Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.

[Hansen et al., 1995] Hansen, N., Ostermeier, A., and Gawelczyk, A. (1995). On the adaptation of arbitrary normal mutation distributions in evolution strategies: The generating set adaptation. *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 57–64.

[He et al., 2006] He, X., Moaveni, B., Conte, J., and Elgamal, A. (2006). Comparative study of system identification techniques applied to new carquinez bridge.

[Heo and Jeon, 2016] Heo, G. and Jeon, J. (2016). An experimental study of structural identification of bridges using the kinetic energy optimization technique and the direct matrix updating method. *Shock and Vibration*, 2016.

[Heo et al., 1997] Heo, G., Wang, M., and Satpathi, D. (1997). Optimal transducer placement for health monitoring of long span bridge. *Soil Dynamics and Earthquake Engineering*, 16(7-8):495–502.

[Heredia-Zavoni and Esteva, 1998] Heredia-Zavoni, E. and Esteva, L. (1998). Optimal instrumentation of uncertain structural systems subject to earthquake ground motions. *Earthquake Engineering & Structural Dynamics*, 27(4):343–362.

[Herzog et al., 2008] Herzog, M., Gilg, A., Paffrath, M., Rentrop, P., and Wever, U. (2008). Intrusive versus non-intrusive methods for stochastic finite elements. *From Nano to Space*, pages 161–174.

[Hoblos et al., 2000] Hoblos, G., Staroswiecki, M., and Aitouche, A. (2000). Optimal design of fault tolerant sensor networks. In *Control Applications, 2000. Proceedings of the 2000 IEEE International Conference on*, pages 467–472. IEEE.

[Hoffman, 2007] Hoffman, D. (2007). Prognostics and health management (PHM)/condition based maintenance (CBM). *IEEE Reliability Society 2007 Annual Technology Report*.

[Hotelling, 1933] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417.

[Huan and Marzouk, 2013] Huan, X. and Marzouk, Y. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

[Imamovic, 1998] Imamovic, N. (1998). *Model validation of large finite element model using test data*. PhD thesis, Imperial College, London.

[Jacquez and Greif, 1985] Jacquez, J. and Greif, P. (1985). Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences*, 77:201–228.

[Jeffreys, 1973] Jeffreys, H. (1973). *Scientific inference*. Cambridge University Press.

[Jia, 2017] Jia, J. (2017). Structural health monitoring and earthquake insurance. In *Modern Earthquake Engineering*, pages 795–807. Springer.

[Jung and Gea, 2004] Jung, D. and Gea, H. (2004). Topology optimization of nonlinear structures. *Finite Elements in Analysis and Design*, 40(11):1417–1427.

[Kakamanshadi et al., 2015] Kakamanshadi, G., Gupta, S., and Singh, S. (2015). A survey on fault tolerance techniques in wireless sensor networks. In *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*, pages 168–173. IEEE.

[Kalman, 1960] Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, pages 35–45.

[Kam et al., 1987] Kam, M., Cheng, R., and Kalata, P. (1987). An information-theoretic interpretation of stability and observability. *American Control Conference, 1987*, pages 1957–1962.

[Kammer, 1991] Kammer, D. (1991). Sensor placement for on-orbit modal identification and correlation of large space structures. *Journal of Guidance, Control and Dynamics*, 14:251–259.

[Kammer, 1996] Kammer, D. (1996). Optimal sensor placement for modal identification using system-realization methods. *Journal of Guidance, Control, and Dynamics*, 19(3):729–731.

[Karhunen, 1947] Karhunen, K. (1947). Über lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae Mathematica*, 37:1–79.

[Katafygiotis and Beck, 1998] Katafygiotis, L. and Beck, J. (1998). Updating models and their uncertainties. II: Model identifiability. *Journal of Engineering Mechanics*, 124(4):463–467.

[Khoshnevisan et al., 2002] Khoshnevisan, M., Bhattacharya, S., and Smarandache, F. (2002). Utility of choice: An information theoretic approach to investment decision-making. *Arxiv*.

[Kim, 2008] Kim, D.-H. (2008). A fiber-optic tiltmeter system based on the moiré-fringe effect. *Measurement Science and Technology*, 20(2):025203.

[Konakli and Sudret, 2016] Konakli, K. and Sudret, B. (2016). Polynomial meta-models with canonical low-rank approximations: numerical insights and comparison to sparse polynomial chaos expansions. *Journal of Computational Physics*, 321:1144–1169.

[Kosambi, 1948] Kosambi, D. (1948). Statistics in function space. *Journal of the Indian Mathemathical Society*, 7:76–88.

[Krajcinovic, 1996] Krajcinovic, D. (1996). *Damage Mechanics*. North Holland (1849).

[Kraskov et al., 2004] Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical review E*, 69(6):066138.

[Kuhlmann, 2003] Kuhlmann, H. (2003). Kalman-filtering with coloured measurement noise for deformation analysis.

[Kullback, 1959] Kullback, S. (1959). *Statistics and Information theory*.

[Kullback and Leibler, 1951] Kullback, S. and Leibler, R. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

[Kurata et al., 2005] Kurata, N., Spencer, B., and Ruiz-Sandoval, M. (2005). Risk monitoring of buildings with wireless sensor networks. *Structural Control and Health Monitoring*, 12(3-4):315–327.

[Kybic, 2004] Kybic, J. (2004). High-dimensional mutual information estimation for image registration. *International Conference on Image Processing ICIP '04*, pages 1779–1782.

[Le Maître and Knio, 2010] Le Maître, O. and Knio, O. M. (2010). *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer.

[Le Maitre et al., 2002] Le Maitre, O., Reagan, M., Najm, H., Ghanem, R., and Knio, O. (2002). A stochastic projection method for fluid flow: II. random process. *Journal of computational Physics*, 181(1):9–44.

[Lee and Kulesz, 2008] Lee, R. W. and Kulesz, J. J. (2008). A risk-based sensor placement methodology. *Journal of hazardous materials*, 158(2):417–429.

[Lemaitre, 1996] Lemaitre, J. (1996). *A Course on Damage Mechanics*. Springer-Verlag Berlin Heidelberg.

[Leyder et al., 2015] Leyder, C., Ntertimanis, V., Chatzi, E., and Frangi, A. (2015). Optimal sensor placement for the modal identification of an innovative timber structure. *Proceedings of the 1st International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2015)*, pages 467–476.

[Liang et al., 2002] Liang, Y., Lee, H., Lim, S., Lin, W., Lee, K., and Wu, C. (2002). Proper orthogonal decomposition and its applications, part I: Theory. *Jounal of Sound and Vibration*, 252:527–544.

[Lindley, 1956] Lindley, D. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, pages 986–1005.

[Liu and Cao, 2017] Liu, X. and Cao, J. (2017). Smart sensor networks for building safety. *Big Data Analytics for Sensor-Network Collected Intelligence*, pages 241–255.

[Ljung and Glad, 1994] Ljung, L. and Glad, T. (1994). On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265–276.

[Lo Iacono et al., 2017] Lo Iacono, F., Navarra, G., and Oliva, M. (2017). Structural monitoring of himera viaduct by low-cost MEMS sensors: characterization and preliminary results. *Meccanica*, 52(13):3221–3236.

[Loeve, 1941] Loeve, M. (1941). *Asymptotical study of dependent random variables*. PhD thesis, Universite de Paris, Paris.

[Łuczak, 2011] Łuczak, S. (2011). Single-axis tilt measurements realized by means of MEMS accelerometers. *Engineering Mechanics*, 18(5-6):341–351.

[Lynch et al., 2003] Lynch, J., Sundararajan, A., Law, K., Kiremidjian, A., Carryer, E., Sohn, H., and Farrar, C. (2003). Field validation of a wireless structural monitoring system on the alamosa canyon bridge. 5057:267–279.

[Mai et al., 2016] Mai, C., Spiridonakos, M., Chatzi, E., and Sudret, B. (2016). Surrogate modeling for stochastic dynamical systems by combining nonlinear autoregressive with exogenous input models and polynomial chaos expansions. *International Journal for Uncertainty Quantification*, 6(4).

[Mallardo and Aliabadi, 2013] Mallardo, V. and Aliabadi, M. (2013). Optimal sensor placement for structural, damage and impact identification: A review. *Structural Durability Health Monitoring*, 9(4):287–323.

[Marbukh and Sayrafian-Pour, 2009] Marbukh, V. and Sayrafian-Pour, K. (2009). Mobile sensor networks self-organization for system utility maximization: work in progress. *Fifth International Conference on Wireless and Mobile Communications, 2009. ICWMC '09.*, pages 416–419.

[Marelli and Sudret, 2014] Marelli, S. and Sudret, B. (2014). Uqlab: a framework for uncertainty quantification in matlab. *2nd International Conference on Vulnerability, Risk Analysis and Management, ICVRAM2014*, pages 2554–2563.

[Mariani et al., 2013a] Mariani, S., Bruggi, M., Caimmi, F., Bendiscioli, P., and De Fazio, M. (2013a). Sensor deployment over damage-containing plates: a topology optimization approach. *Journal of Intelligent Material Systems and Structures*, 24:1105–1122.

[Mariani et al., 2014] Mariani, S., Caimmi, F., Bruggi, M., and Bendiscioli, P. (2014). Smart sensing of damage in flexible plates through MEMS. *International Journal of Mechanisms and Robotic Systems*, 2:67–95.

[Mariani and Corigliano, 2005] Mariani, S. and Corigliano, A. (2005). Impact induced composite delamination: state and parameter identification via joint and dual extended Kalman filters. *Computer methods in applied mechanics and engineering*, 194(50):5242–5272.

[Mariani et al., 2013b] Mariani, S., Corigliano, A., Caimmi, F., Bruggi, M., Bendiscioli, P., and De Fazio, M. (2013b). MEMS-based surface mounted health monitoring system for composite laminates. *Microelectronics Journal*, 44:598 – 605.

[Mariani and Ghisi, 2007] Mariani, S. and Ghisi, A. (2007). Unscented Kalman filtering for nonlinear structural dynamics. *Nonlinear Dynamics*, 49(1):131–150.

[MarketsandMarkets, 2017] MarketsandMarkets (2017). Structural health monitoring market by solutions (hardware: Sensors, data acquisition system; software & services), technology (wired and wireless), end users and geography - global forecast to 2022.

[Maute et al., 1998] Maute, K., Schwarz, S., and Ramm, E. (1998). Adaptive topology optimization of elastoplastic structures. *Structural and Multidisciplinary Optimization*, 15(2):81–91.

[Mees et al., 1987] Mees, A., Rapp, P., and Jennings, L. (1987). Singular-value decomposition and embedding dimension. *Physical Review A*, 1;36(1):340–346.

[Meo and Zumpano, 2005] Meo, M. and Zumpano, G. (2005). On the optimal sensor placement techniques for a bridge structure. *Engineering Structures*, 27(10):1488–1497.

[Miao et al., 2011] Miao, H., Xia, X., Perelson, A. S., and Wu, H. (2011). On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM review*, 53(1):3–39.

[Mirzazadeh, 2017] Mirzazadeh, R. (2017). *Micromechanical characterization of polysilicon films: on-chip testing, multi-uncertainty quantification and Bayesian inverse modelling*. PhD thesis, Politecnico di Milano, Milano.

[Mirzazadeh et al., 2017] Mirzazadeh, R., Azam Eftekhar, S., Jansen, E., and Mariani, S. (2017). Uncertainty quantification in polysilicon MEMS through on-chip testing and reduced-order modelling. *18th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, pages 1–8.

[Moaveni and Asgarieh, 2012] Moaveni, B. and Asgarieh, E. (2012). Deterministic-stochastic subspace identification method for identification of nonlinear structures as time-varying linear systems. *Mechanical Systems and Signal Processing*, 31:40–55.

[Moaveni and Conte, 2007] Moaveni, B. and Conte, J. (2007). *System and damage identification of civil structures*. Springer.

[Monnier et al., 2000] Monnier, T., Jayet, Y., Guy, P., and Baboux, J. (2000). Aging and damage assessment of composite structures using embedded piezoelectric sensors. *AIP Conference Proceedings*, 509(1):1269–1276.

[Moon et al., 1995] Moon, Y.-I., Rajagopalan, B., and Lall, U. (1995). Estimation of mutual information using kernel density estimators. *Physical Review. E*, 52(3):2318–2321.

[Moser and Moaveni, 2013] Moser, P. and Moaveni, B. (2013). Design and deployment of a continuous monitoring system for the dowling hall footbridge. *Experimental Techniques*, 37(1):15–26.

[Mouzakis et al., 2014] Mouzakis, D., Dimogianopoulos, D., and Zaoutsos, S. (2014). Damage assessment of carbon fiber reinforced composites under accelerated aging and validation via stochastic model-based analysis. *International Journal of Damage Mechanics*, 23(5):702–726.

[Mukhopadhyay et al., 2013] Mukhopadhyay, S., Betti, R., and Lus, H. (2013). Output only structural identification with minimal instrumentation. In *Special Topics in Structural Dynamics, Volume 6*, pages 593–602. Springer.

[Mukhopadhyay et al., 2014] Mukhopadhyay, S., Luş, H., and Betti, R. (2014). Modal parameter based structural identification using input–output data: minimal instrumentation and global identifiability issues. *Mechanical Systems and Signal Processing*, 45(2):283–301.

[National Concrete Pavement Technology Center, 2011] National Concrete Pavement Technology Center, N. (2011). A feasibility study on embedded microelectromechanical sensors and systems MEMS for monitoring highway structures. *Iowa State University Report*.

[Ni and Wong, 2012] Ni, Y. and Wong, K. (2012). Integrating bridge structural health monitoring and condition-based maintenance management. In *Civil Structural Health Monitoring Workshop (CSHM-4). ISHMII*.

[Nienaltowski et al., 2015] Nienaltowski, K., Wlodarczyk, M., Lipniacki, T., and Komorowski, M. (2015). Clustering reveals limits of parameter identifiability in multiparameter models of biochemical dynamics. *BMC Systems Biology*, 9(65).

[Noel et al., 2017] Noel, A., Abdaoui, A., Badawy, A., Elfouly, T., Ahmed, M., and Shehata, M. (2017). Structural health monitoring using wireless sensor networks: a comprehensive survey. *IEEE Communications Surveys & Tutorials*, 19(3):1403–1423.

[Pant and Lombardi, 2015] Pant, S. and Lombardi, D. (2015). An information-theoretic approach to assess practical identifiability of parametric dynamical systems. *Mathematical Biosciences*, 268:66–79.

[Papadimitriou, 2004] Papadimitriou, C. (2004). Optimal sensor placement methodology for parametric identification of structural systems. *Journal of Sound and Vibration*, 278(4):923–947.

[Papadimitriou et al., 2000] Papadimitriou, C., Beck, J., and Au, S.-K. (2000). Entropy-based optimal sensor location for structural model updating. *Journal of Vibration and Control*, 6(5):781–800.

[Papadimitriou and Lombaert, 2012] Papadimitriou, C. and Lombaert, G. (2012). The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mechanical Systems and Signal Processing*, 28:105–127.

[Parnell et al., 2011] Parnell, G., Driscoll, P., and Henderson, D. (2011). *Decision making in systems engineering and management*, volume 81. John Wiley & Sons.

[Pearl, 1988] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc.

[Pearson, 1901] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

[Pioldi et al., 2016] Pioldi, F., Ferrari, R., and Rizzi, E. (2016). Output-only modal dynamic identification of frames by a refined fdd algorithm at seismic input and high damping. *Mechanical Systems and Signal Processing*, 68:265–291.

[Pioldi et al., 2017] Pioldi, F., Ferrari, R., and Rizzi, E. (2017). Earthquake structural modal estimates of multi-storey frames by a refined frequency domain decomposition algorithm. *Journal of Vibration and Control*, 23(13):2037–2063.

[Pioldi and Rizzi, 2016] Pioldi, F. and Rizzi, E. (2016). A full dynamic compound inverse method for output-only element-level system identification and input estimation from earthquake response signals. *Computational Mechanics*, 58(2):307–327.

[Pohjanpalo, 1978] Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Mathematical biosciences*, 41(1-2):21–33.

[Raiffa and Schlaifer, 1961] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Wiley-Interscience.

[Ramesh et al., 2012] Ramesh, S., Kannan, S., and Baskar, S. (2012). Application of modified NSGA-II algorithm to multi-objective reactive power planning. *Applied Soft Computing*, 12(2):741–753.

[Raue et al., 2009] Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929.

[Ryan, 2003] Ryan, K. (2003). Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12(3):585–603.

[Sabato et al., 2017] Sabato, A., Niezrecki, C., and Fortino, G. (2017). Wireless mems-based accelerometer sensor boards for structural vibration monitoring: a review. *IEEE Sensors Journal*, 17(2):226–235.

[Salawu, 1997] Salawu, O. (1997). Detection of structural damage through changes in frequency: a review. *Engineering Structures*, 19:718–723.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

[Sharma et al., 2013] Sharma, G., Agarwala, A., and Bhattacharya, B. (2013). A fast parallel gauss jordan algorithm for matrix inversion using cuda. *Computers & Structures*, 128:31–37.

[SIMULIA, 2013] SIMULIA (2013). *Abaqus 6.13 Documentation*. Dassault Systemes.

[Sobczyk, 1987] Sobczyk, K. (1987). Theoretic information approach to identification and signal processing. *Reliability and Optimization of Structural Systems*, 33:373–383.

[Sohn et al., 2004] Sohn, H., Farrar, C. R., Hemez, F., and Czarnecki, J. (2004). A review of structural health monitoring literature 1996-2001. Technical report, Los Alamos National Laboratory.

[Spiridonakos et al., 2016] Spiridonakos, M. D., Chatzi, E., and Sudret, B. (2016). Polynomial chaos expansion models for the monitoring of structures under operational variability. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 2(3):B4016003.

[Stallings et al., 2000] Stallings, J., Tedesco, J., El-Mihilmy, M., and McCauley, M. (2000). Field performance of frp bridge repairs. *Journal of Bridge Engineering*, 5(2):107–113.

[Steuer et al., 2002] Steuer, R., Kurths, J., Daub, C., Weise, J., and Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18(2):S231–S240.

[STMicroelectronics, 2014] STMicroelectronics (2014). Tilt measurement using a low-g 3-axis accelerometer. *AN4509 Application note*.

[Sudret, 2007] Sudret, B. (2007). Uncertainty propagation and sensitivity analysis in mechanical models–contributions to structural reliability and stochastic spectral methods. Technical report, Habilitation á diriger des recherches, Université Blaise Pascal.

[Sudret and Der Kiureghian, 2000] Sudret, B. and Der Kiureghian, A. (2000). *Stochastic finite element methods and reliability: a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California Berkeley.

[Svanberg, 1982] Svanberg, K. (1982). An algorithm for optimum structural design using duality. *Applications*, pages 161–177.

[Svanberg, 1987] Svanberg, K. (1987). The method of moving asymptotes–a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24(2):359–373.

[Swann and Chattopadhyay, 2006] Swann, C. and Chattopadhyay, A. (2006). Optimization of piezoelectric sensor location for delamination detection in composite laminates. *Engineering Optimization*, 38(5):511–528.

[Szabó, 2014] Szabó, Z. (2014). Information theoretical estimators toolbox. *Journal of Machine Learning Research*, 15(1):283–287.

[Tan and Zhang, 2015] Tan, L. and Zhang, Y. (2015). Optimal resource allocation with principle of equality and diminishing marginal utility in wireless networks. *Wireless Personal Communications*, 84(1):671–693.

[Taylor, 1997] Taylor, J. (1997). *Introduction to error analysis, the study of uncertainties in physical measurements*. University Science Books.

[Trefethen and Bau III, 1997] Trefethen, L. N. and Bau III, D. (1997). *Numerical linear algebra*, volume 50. SIAM.

[Ubertini et al., 2016] Ubertini, F., Comanducci, G., and Cavalagli, N. (2016). Vibration-based structural health monitoring of a historic bell-tower using output-only measurements and multivariate statistical analysis. *Structural Health Monitoring*, 15(4):438–457.

[Ubertini et al., 2013] Ubertini, F., Gentile, C., and Materazzi, A. L. (2013). Automated modal identification in operational conditions and its application to bridges. *Engineering Structures*, 46:264–278.

[Udwadia and Sharma, 1978] Udwadia, F. and Sharma, D. (1978). Some uniqueness results related to building structural identification. *SIAM Journal on Applied Mathematics*, 34(1):104–118.

[U.S. Department of Transportation, 2001] U.S. Department of Transportation (2001). Federal Highway Administration. Highway bridge inspection: state-of-the-practice survey.

[Vanlier et al., 2014] Vanlier, J., Tiemann, C. A., Hilbers, P. A., and van Riel, N. A. (2014). Optimal experiment design for model selection in biochemical networks. *BMC systems biology*, 8(1):20.

[Vergara and Estévez, 2014] Vergara, J. R. and Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, 24(1):175–186.

[Walter and Lecourtier, 1981] Walter, E. and Lecourtier, Y. (1981). Unidentifiable compartmental models: What to do? *Mathematical biosciences*, 56(1-2):1–25.

[Walters-Williams and Li, 2009] Walters-Williams, J. and Li, Y. (2009). Estimation of mutual information: a survey. *Rough Sets and Knowledge Technology*, pages 389–396.

[Wenzel, 2009] Wenzel, H. (2009). From structural health monitoring to risk based infrastructure management. *Conference of the International Society for Structural Health Monitoring of Intelligent Infrastructure*, pages 22–24.

[Wiener, 1938] Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936.

[Wong et al., 2000] Wong, K., Lau, C., and Flint, A. (2000). Planning and implementation of the structural health monitoring system for cable-supported bridges in Hong Kong. *Nondestructive Evaluation of Highways, Utilities, and Pipelines IV*, 3995:266–275.

[Wu et al., 2003] Wu, C., Liang, Y., Lin, W., Lee, H., and Lim, S. (2003). A note on equivalence of proper orthogonal decomposition methods. *Journal of Sound and Vibration*, 265(5):1103–1110.

[Xiu and Karniadakis, 2002] Xiu, D. and Karniadakis, G. (2002). The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644.

[Xiu and Karniadakis, 2003] Xiu, D. and Karniadakis, G. (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137–167.

[Xiu et al., 2002] Xiu, D., Lucor, D., Su, C.-H., and Karniadakis, G. (2002). Stochastic modeling of flow-structure interactions using generalized polynomial chaos. *Journal of Fluids Engineering*, 124(1):51–59.

[Yang and Lu, 2017] Yang, C. and Lu, Z. (2017). An interval effective independence method for optimal sensor placement based on non-probabilistic approach. *Science China Technological Sciences*, 60(2):186–198.

[Yang et al., 2006] Yang, J., Lin, S., Huang, H., and Zhou, L. (2006). An adaptive extended Kalman filter for structural damage identification. *Structural Control Health Monitoring*, 13:849–867.

[Yang et al., 2015] Yang, Y., Cattaneo, A., and Mascareñas, D. (2015). Potential structural health monitoring tools to mitigate corruption in the construction industry associated with rapid urbanization. *International Conference on Sustainable Development, ICSD*.

[Yao et al., 1993] Yao, L., Sethares, W., and Kammer, D. (1993). Sensor placement for on-orbit modal identification via a genetic algorithm. *AIAA journal*, 31(10):1922–1928.

[Yi and Li, 2012] Yi, T.-H. and Li, H.-N. (2012). Methodology developments in sensor placement for health monitoring of civil infrastructures. *International Journal of Distributed Sensor Networks*, 8(8):612726.

[Yin et al., 2016] Yin, R.-C., Wu, Y.-M., and Hsu, T.-Y. (2016). Application of the low-cost MEMS-type seismometer for structural health monitoring: A pre-study. *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International*, pages 1–5.

[Yu et al., 2017] Yu, H., Mohammed, M., Mohammadi, M., Moaveni, B., Barbosa, A., Stavridis, A., and Wood, R. (2017). Structural identification of an 18-story RC building in Nepal using post-earthquake ambient vibration and lidar data. *Frontiers in Built Environment*, 3:11.

[Yuen, 2010] Yuen, K.-V. (2010). *Bayesian methods for structural dynamics and civil engineering*. John Wiley & Sons.

[Zak et al., 2012] Zak, A., Radzieński, M., Krawczuk, M., and Ostachowicz, W. (2012). Damage detection strategies based on propagation of guided elastic waves. *Smart Materials and Structures*, 21(3).

[Zhang and Aktan, 1998] Zhang, Z. and Aktan, A. (1998). Application of modal flexibility and its derivatives in structural identification. *Research in Nondestructive Evaluation*, 10:43–61.